The Blind Men and the Elephant

The Fourth reached out an eager hand,
And felt about the knee:
"What most this wondrous beast is like
Is mighty plain," quoth he;
"Tis clear enough the Elephant
Is very like a tree!"

John Godfrey Saxe's (1816-1887)
version of the famous Indian legend.

CHAPTER 5
DATA SOURCES AND METHODS

Statistical exploratory methods were used to answer the research questions for cross-sectional analysis for the years 1990 and 1997, and analysis of changes from 1990 to 1997. Quantitative models of progress and levels of human development were used to establish the importance of each aspect of interest of the global digital divide, for 174 countries (out of 231 territories) representing 93% of the world's population.

THE DEPENDENT VARIABLE: HUMAN DEVELOPMENT

Quantitative measures of development are varied and numerous. From a modernization perspective, development is usually measured in terms of the percentage of labor force employed in industry and the service sector (Moaddel 1994), gross national product (GNP) per capita, energy consumption per capita, and the percentage of economically active males in non-agricultural employment (Roberts and Hite 2000). Alternative indicators from a World Systems perspective express development as a result of trade globalization measured in terms of the sum of world exports divided by the sum of GDP on a world scale. Trade globalization is used for determining the a country's

position (whether it belongs to the core, the semi-periphery or the periphery) within the World System (Wallerstein 1991). Finally, development economics and economic growth models traditionally measure development by economic indicators such as gross domestic product (GDP) per capita, or GDP growth per capita.

Not directly associated with development theory[1], but from experience of agencies involved in development projects, an early composite indicator of development that included non-economic indicators was the Physical Quality of Life Index (PQLI). This index was developed in 1979 by Morris D. Morris (Morris 1979) of the Overseas Development Council. Morris averaged rankings on three indicators – life expectancy at age 1, infant mortality, and literacy – to form the PQLI, which ranged from 1 (for the poorest performance) to 100 (for the best performance). However, the PQLI was abandoned since it captured only the health and education dimensions, leaving out the economic dimension.

In 1987, the Population Crisis Committee of Washington DC (now known as Population Action International) developed the International Human Suffering Index. This index was a composite of ten indicators – income, inflation, demand for new jobs, urban population pressures, infant mortality, nutrition, clean water, energy, adult literacy, and personal freedom. Largely as a result of the inclusion of a notion of personal freedom and focus on population, the Human Suffering Index proved controversial and was not continued.

---

[1] For more background on development theories see Appendix E.

With the release of the first 1990 Human Development Report, the United Nations Development Program (UNDP) acknowledged a more comprehensive development concept than just that of economic development measurement: the concept of *human development*. Like the PQLI, the HDI is a composite of the rankings on three variables[2]: *longevity*, as measured by life expectancy at birth; *knowledge*, as measured by a weighted combination of adult literacy (two-thirds weight) and the combined gross primary, secondary, and tertiary enrollment ratio (one-third weight); and *standard of living*, as measured by GDP per capita in purchasing power parity (PPP) in U.S. dollars.

Since 1990, the HDI has been published for the majority (174) of United Nations (UN) member countries as well as basic social indicators for other UN member countries (13). The Human Development Report publishes the HDI compiled through a collective effort of original sources of data. These range from national census and surveys to international data series collected and harmonized by international organizations. All the

---

[2] The estimates for "life expectancy at birth" are obtained using a linear interpolation based on five-year averages published by the United Nations Population Division. The latest survey data available for 1998 makes significant adjustments to incorporate the demographic impact of HIV/AIDS, especially significant in Sub-Saharan Africa. Adult literacy rates, as well as gross primary, secondary and tertiary enrolment for year 2000 are projections using UNESCO February 2000 literacy assessment. GDP per capita is provided by the World Bank and based on the latest International Comparison Programme (ICP) surveys. For comparison purposes, process and expenditure data from the surveys are linked to a standard classification scheme to compile internationally comparable purchasing power parity (PPP) data. The year for the PPP data is 1996; data for 1998 was extrapolated using relative price movements over time between each country and the United States, the base country. For countries not covered by the World Bank, the UNDP uses PPP estimates for the construction of the HDI taken from data produced by the University of Pennsylvania UN. 1990. "Human Development Report." New York: United Nations (UN), —. 1996. "Human Development Report." New York: United Nations (UN), —. 1998. "Human Development Report." New York: United Nations (UN), —. 1999. "Human Development Report." New York: United Nations (UN),, —. 2000. "Human Development Report." New York: United Nations (UN), —. 2001. "Human Development Report: making new technologies work for human develoment." New York: United Nations (UN)..

statistical tables in the report are based on internationally standardized data, collected and processed by sister agencies in the international system. These organizations, whether collecting data from national sources or through their own surveys, harmonize definitions and collection methods to make their data as internationally comparable as possible. The data produced by these agencies may sometimes differ from data produced by national sources, often because of adjustments to harmonize data.

For the purposes of this study, development –the dependent variable of interest- will be measured as defined by the UN in terms of the Human Development Index (HDI). The human development index classification is also used by the UN to classify countries into three clusters: high human development (with an HDI of 0.800 or above), medium human development (0.500-0.799) and low human development (less than 0.500). Other non mutually exclusive world classifications used for the study are: 1) OECD, 2) All developing countries, 3) Eastern Europe and CIS, and 4) Least developed countries. Regional analysis are carried out using the following UN regions: 1) Arab States, 2) East Asia, 3) Latin America and the Caribbean, 4) South Asia, 5) South-East Asia and the Pacific, 6) Southern Europe and 7) Sub-Saharan Africa.

As stated before, the HDI is a composite index constructed for countries for a particular year. The maximum and minimum values that define the maximum distance for each indicator included in the index is specific for that year. Due to changes in the values of the maximum and minimum values of these variables, measuring progress in human development is not a mere subtraction of indexes. Thus, if we are measuring progress between 1990 and 1997, the minimum and maximum of all values for all

countries are used for the construction of the index of progress of human development. Multiyear averages of growth rates are expressed as compound annual rates of change, whereas year to year growth rates are expressed an annual percentage changes. These considerations were taken into account for the construction of the indexes of progress of human development between 1990 and 1997. The HDI for 1999 was not used due to further comparability complications[3] produced by a variation in the methodology for the construction of the GDP per capita (PPP US$) for 1999.

Even though in 1997 the HDI is reported for 174 countries, in 1990 the HDI was not reported for 76. Missing values for 1990 were estimated using the median value of six values generated by different statistical estimation techniques[4]. Following UN methodology[5], the estimated values were then used to compute the rate of progress between 1990 and 1997 for these 76 countries. See Appendix H for full listing of reported and calculated values. All the analysis were carried out for computed rates of progress in human development for 174 countries, as well as reported rates of progress in human development for 98 countries.

---

[3] Personal communication with the research and statistics team in the UN confirmed that comparison between the HDI in 1990 and 1999 were not possible Stewart, David. 2001. "HDI Trend." edited by Author. Washington, D.C.: Research Associate. Human Development Report Office. UNDP..

[4] Statistical estimation techniques used for estimating missing values were: regression of HDI in 1990 based on HDI 1998 values, mean of four nearby points, median of five nearby points, linear interpolation, linear trend at a point, and average per region of development.

[5] Progress in human development is measured as percentual increases in what a country was missing in achieving full development in which HDI=1. This is why progress in human development is called “reduction in shortfall” by the UN.

As one would expect, a composite indicator of development does not capture the full range of dimensions of development. The "variations in levels of human development disaggregated by region, gender, ethnic group, or rural or urban areas reveal significant disparities" (UN 1999). When the HDI is disaggregated along rural and urban populations, it reflects more progress in human development for urban populations than for rural populations. Similarly, when the HDI is disaggregated along regions, ethnic, language or gender lines it shows wide variation. Even though sub-national variations are not captured in this level of aggregation, the HDI has proven to be appropriate for cross-national comparisons (Baker 2001; Crenshaw and Robison 2000; Gage 2000; ITU 1999; Jha and Bhanu Murthy 2001; Kirkman 1999; Mody and Dahlman 1992; Norris in press; Wilson and Rodriguez 2000).

DATA SOURCES AND DATA REDUCTION METHODS FOR INDEPENDENT VARIABLES

Country-level statistics for analysis in this study comes mainly from reports published by four multilateral and international agencies: the United Nations (UN), the National Science Foundation (NSF), the World Bank (WB), and the International Telecommunication Union[6] (ITU).

---

[6] The International Telecommunication Union (ITU) is a United Nations (UN) specialized agency that maintains an extensive collection of statistics on communication and information. The data are collected from an annual questionnaire sent out by the Telecommunication Development Bureau (BDT) of the ITU. Additional data are obtained from reports of telecommunication ministries, regulators and operators and from ITU staff reports. The World Telecommunication Indicators Database contains time series data for over 200 countries for the years 1960, 1965, 1970 and annually from 1975-1999 for more than 100 communication statistics. Indicators include telephone network size and dimension, mobile services, quality of service, traffic, staff, tariffs, revenue and

National Core Computing and Networking Capacity and National Info Tech Infrastructure

Core national and computing infrastructure is measured using Wilson and Rodriguez' Core Index of Technological Progress (Core-ITP) (Wilson and Rodriguez 2000). In spite of the availability of other indexes such as the Index of Technological Progress[7], the Broad-ITP[8], and the Forward Looking Core ITP[9], among all the indexes, the Core ITP was selected because it represented the *core* of the digital information and communication technologies at national level by adding the number of personal computers per capita and internet hosts per capita in a country over the last decade[10]. So instead of constructing an index to measure national core computing and networking capacity, Wilson and Rodriguez' Core Index of Technological Progress (Core-ITP) index was selected as proxy measure of national core computing and networking capacity. Another advantage of using the Core-ITP is that Wilson and Rodriguez used an alternative source[11] to validate the statistics offered in number of personal computers per

---

investment, as well as selected demographic, macro-economic, broadcasting and information technology statistics.

[7] The ITP includes television sets, mobile phones, personal computers, internet hosts, fax machines, R&D as percentage of GDP, technicians, scientists and telephone mainlines.

[8] Expands the ITP to include newspapers and radios.

[9] Includes the Core ITP and telephone mainlines and televisions.

[10] See Appendix M for a discussion on how the core components of the index of national computing and networking capacity vary in importance over 1990 and 1997 depending whether the absolute indicator or the relative indicator per capita is used for the construction of the index. Appendix M also includes a presentation of constructing alternative indexes to measure the core computing and networking national capacity.

[11] WTIA's Digital Planet produces and sells a yearly computer industry almanac that includes statistics for the number of personal computers purchased aggregated by countries around the world based on industry reports.

capita included in the ITU[12] database. The Core-ITP ranges from 0 (Niger) to 100 (United States). Finally, selecting Wilson and Rodriguez' Core ITP instead of constructing another index based on the same variables allows future comparisons among studies. See Appendix I for full listing of country values for Core-ITP.

As stated before, national computing and networking capacity is one of the many components of an info tech infrastructure. The info tech infrastructure (IT-infrastructure) requires an underlying "wired" infrastructure provided by electrical and telephonic grids. National computing and networking capacity for 120 countries was measured using Wilson and Rodriguez' Core Index of Technological Progress (Core-ITP) (Wilson and Rodriguez 2000), whereas a national telecommunication index was constructed using main telephone lines per 10,000 habitants[13] taken from those published in the ITU database. Likewise, electrical consumption per capita in 1990 and 1997 was taken from the World Bank 1999 Report to construct an index of electrical consumption per capita.

A correlation matrix illustrated in Table 5.1. illustrates, wired variables[14] and national computing and networking capacity are highly correlated among each other.

---

[12] The International Telecommunication Union (ITU) database is much stronger in statistics for the telecommunication sector such as telephone network and dimension, mobile services, quality of services, etc. For example, the ITU database includes estimates for the number of personal computers for only 63 countries in 1990 (mostly developed countries).

[13] Also known as teledensity.

[14] Wired variables include: electrical consumption per capita and main telephone lines per 10,000 (also known as teledensity)

Correlations between electrical consumption per capita, main telephone lines per 10,000 habitants and Core-ITP

		Electricty consumption per capita	Main telephone lines per 10,000 hab	Core ITP suggested by Wilson & Rodriguez
Electrical consumption per capita	Pearson Correlation	1.000	.783**	.833**
	Sig. (2-tailed)	.	.000	.000
	N	118	118	118
Main telephone lines per 10,000 hab	Pearson Correlation	.783**	1.000	.811**
	Sig. (2-tailed)	.000	.	.000
	N	118	120	120
Core ITP suggested by Wilson & Rodriguez	Pearson Correlation	.833**	.811**	1.000
	Sig. (2-tailed)	.000	.000	.
	N	118	120	120

**. Correlation is significant at the 0.01 level (2-tailed).

Table 5.1. Correlations among Core-ITP, electrical consumption per capita and main telephone lines per 10,000

Therefore, an overall index for national info tech infrastructure was constructed giving equal weight to the Core-ITP, the teledensity index and the index of electrical consumption[15]. Three indexes were used for the analysis: the Core-ITP (representing the core computing and networking national capacity), the "wired" index (representing the

[15] The index was constructed from the results obtained from exploratory factor analysis using Principal Components Analysis (PCA) that identified the factors or latent variables underling each measure, and generated factors that exhibit no covariance among them. The generation of indexes based on PCA for each group of indicators is a solution to three problems: 1) the inherent multidimensionality of what we are trying to measure, i.e. info tech infrastructure, 2) the high level of collinearity (multi-collinearity) among different indicators, and 3) simultaneity. As stated before, the first problem is due to the multidimensional nature of the info tech infrastructure that is expressed by a series of indicators measuring different dimensions of the same phenomenon. PCA is based on linear transformation of the explanatory variables such that they are orthogonal to each other by design producing principal components that are an exact mathematical transformation of the raw variables. The second and third problems are especially important to solve since the different indicators of information tech infrastructure may be too correlated among themselves not allowing the use of their coefficients as explanatory variables. Varimax rotations were used for all cases.

underlying wired capacity in electrical power and telephonic grids) and the national info tech infrastructure (representing the core national computing and networking capacity and the wired index). Therefore, an index representing national info tech infrastructure was constructed by adding each variable as an index[16] with equal weight. The IT-infrastructure index was used to represent national computing and networking infrastructure *as well as* the underlying "wired" infrastructure as illustrated in Figure 5.1. Appendix I includes values for each index for 120 countries.

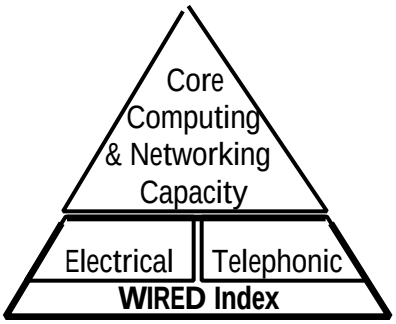

Figure 5.1. National info tech infrastructure composed of:
core national computing and networking capacity,
electrical and telephonic capacity (wired index)

[16] The indexes were constructed ranging from 0 to 100.

Human Capital and Intellectual Index

Human and Intellectual Capital indicators were taken from those published by the United Nations[17] and the National Science Foundation[18] to produce the Human and Intellectual Capital Index[19]. Human Capital and Intellectual indicators include published indicators for 57 countries for scientific and engineering labor force, number of scientific and technical publications, and publishing productivity (number of articles per GDP). Values for the Human and Intellectual Capital index are included in Appendix J.

Cultural Indexes

The dataset for analyzing the cross-national impact of the values and beliefs of mass publics on political and social life was taken from the World Value Surveys[20].

---

[17] The United Nations publishes human capital indicators provided by UNESCO (United Nations agency for Education and Science) UN. 1999. "Human Development Report." New York: United Nations (UN),.

[18] The indicators from the National Science Foundation were taken from the "Science and Engineering Indicators 2000" NSF, and NSB. 2000. "Science and Engineering Indicators 2000." Washington DC.: National Science Foundation (NSF),

National Science Board (NSB).

[19] This index was also constructed from the results obtained from exploratory factor analysis using Principal Components Analysis (PCA). See footnote 7.

[20] Starting with the 1990 survey, the survey has been carried out by a global network of social scientists participating in the design, fieldwork, analysis and interpretation of the data. The 1990-93 survey covered 42 countries, whereas the 1995-97 survey covered 53

Cultural indexes were constructed by using the mean value of aggregating individual responses by countries included in the World Value Survey. The original questions and their scales are included in Appendix K. Ten cultural zones were used to classify countries (1=Protestant Europe, 2=Catholic Europe, 3=North America, 4=Latin America, 5=Confucian, 6=South Asia, 7=Eastern Europe, 8=Africa, 9=Ex-Communist, 10=English Speaking). Cultural zones use Huntington's cultural zones as guide (Huntington 1971). Changes in societies were calculated as a difference between values for a "rational society" in 1990-91 and 1995-96 reported in the literature (Inglehart 2000). Values for means were recoded and recalculated to produce cultural indexes that range from 0 to 1. All values are reported in Appendix L.

1) **Pro Technology Index**– This index refers to a country's pro-technology cultural values. Societies that consider the "effects of technology of our way of life to be beneficial" will rank higher on the Pro Technology Index. The Pro Technology Index is constructed for 1990-91 and 1995-96. Values are available for 41 countries in 1990-91 and 49 countries in 1995-96.

2) **Pro Science Index**– This index refers to a country's pro-scientific cultural values. Societies that consider the "effects of scientific advances to be

---

countries including more developing countries than previously. All of theses surveys were carried out through face to face interviews with a sampling universe consisting of adult citizens, ages 18 and older. In the usual sampling design, within each country, a stratified multi-stage random selection of sampling points was used. First, a random selection of sampling locations was made ensuring that all types of locations were represented in the portion of their population. Next, a random selection of individuals was drawn up. In spite of having access to some sub-national data, only the national samples were selected for the study. The dataset was weighed to correct for undersampling the illiterate portion of the population and oversampling the urban areas and more educated strata. Individual responses were aggregated to produce indicators acting as proxies for public opinion ideology at macro level.

beneficial for mankind in the long run" will rank higher on the Pro Science Index. The index is constructed for 1990-91 and 1995-96. Values are available for 41 countries in 1990-91 and 49 countries in 1995-96.

3) **Openness to New Ideas Index** – This index refers to cultural values that may be affected by increases in information flows produced by the information revolution. The Openness to Ideas Index is constructed for 1990-91 and 1995-96. Values are available for 41 countries in 1990-91 and 49 countries in 1995-96. Percentual changes relative to the original value in the index are used to construct an index of progress in openness to new ideas available for 27 countries.

4) **Global Sense of Belonging Index** – This index refers to cultural values that may be affected by increased levels of information. Increases in access to global information will theoretically affect the geographical sense of belonging (local, state/region, nation, continent or world). The Global Sense of Belonging Index is constructed for 1990-1991 and 1995-1996. Values are available for 41 countries in 1990-91 and 46 countries in 1995-96. Percentual changes relative to the original value in the index are used to construct index of progress in the global sense of belonging available for 26 countries.

Control Variables

The quantitative model was tested with control demographic, economic, and political variables illustrated in Table 5.2.

Control Variable[21]	Data Source
Economic Freedom[22] ***Measures: Corruption, Non-tariff barriers to trade, Fiscal burden of government, Rule of law, Regulatory burdens, Restriction on banks, Labor market regulations, and Black market activities***	Heritage Foundation[23]
Political Freedom (Civil Liberties) ***Measures: Freedom of expression and belief, Association and organizational rights, Rule of law and human rights, Personal autonomy, and Economic rights***	The Freedom House[24]
Annual percentage of population growth	United Nations
Percentage of productive population	World Bank Reports
Foreign Direct Investment (in mill US)	World Bank Reports
National Investment as percentage of GDP	World Bank Reports
Loss in number of lives per disasters	United Nations
Percentage of urban population	United Nations
Percentage of the service sector in the economy	United Nations

Table 5.2. Control variables and data sources

---

[21] Data is for 1997 (otherwise noted).

[22] Data is for 1995.

[23] The Heritage Foundation and The Wall Street Journal carry out a survey for 155 countries rating them by the index of economic freedom since 1995. Ratings on the "Index of Economic Freedom" are based on an analysis of 50 different economic variables grouped into 10 categories: banking and finance, capital flows and foreign investment, monetary policy, fiscal burden of government, trade policy, wages and prices, government intervention in the economy, property rights, regulation, and black-market activity. Countries are rated one to five in each category, one being the best, five the worst. The index also is used to rate the world's economies as: unfree, free, mostly free, mostly unfree or repressed. Most of the freest economies are concentrated in North America and Europe, while a majority of the world's most repressed economies are in Asia and Africa.

[24] The data for Civil Liberties was taken from the survey published by the Freedom House since 1974. Civil liberties are rated on a seven-category scale, 1 representing the most free and 7 the least free. (The scale was recoded to 1 representing the least free and 7 the most free).

MISSING VALUES AND DROPPED CASES

Sixty-nine countries were dropped from the analysis due to missing values for all dependent variables (See Appendix M for full listing) leaving a sample of 174 countries (out of 231 territories) that account for more than 93% of the world's population. Dropped cases included those countries for which war or civil strife affected the uninterrupted reporting of data. These countries are: Afghanistan, Bosnia and Herzegovina, The Democratic Republic of Korea, Liberia, and Somalia. Other cases such as Monaco, San Marino and Liechtenstein with populations ranging from 26,000 to 33,000 habitants for which no Human Development Indexes (HDI) are available were also dropped from the sample. Finally, territories (mostly islands) that are under the influence of an OECD country in commonwealth-type agreement were also dropped.

Listwise treatment was used for all other missing values, except for cases in which missing values represented a group worth exploring. In such cases, a missing values group was constructed. This was particularly important for some cases in which scanty data was most frequent, i.e. as in less developed countries that lack mechanisms to collect and report information to international agencies.

OTHER STATISTICAL METHODS

The following levels of significance were selected to report findings: $p < 0.05$ (*), $p < 0.01$ (**) and $p < 0.001$ (***). Other levels of significance were considered statistically insignificant. Adjusted R^2 are reported for models with more than one predictor, R^2 are reported for analysis with only one predictor.

Correlations were used to test the existence of pair-wise linear relations among variables when the data set was reduced[25] like the case of cultural values or when the data set was divided by regions or development levels.

Scatter plots were used to observe the distribution of the variances of the dependent variable (levels of human development and progress in human development), and to examine what type of mathematical functions would be appropriate to describe the association between the variables.

Since we had information for almost all countries in the world, we assumed a fixed-effects model that allowed the use of ANOVA. One way ANOVAs were used to test whether the means of the dependent variables[26] for different groups or categories of independent variables[27] were statistically significant from each other. This was also tested with values obtained from Kendall's *tau b* between categories of dependent and independent variables.

Special care was given to meet the assumptions of the regression model which are:

1) *Independence*. The independent variables are statistically independent of each other. That is, observations are in no way influenced by other observations. (This criteria was met by the characteristics of the data).

---

[25] When the number of observations was less than 50 ($N<50$).

[26] Dependent variables were: levels of human development, and progress in human development.

[27] Independent variables such as: core national computing and networking infrastructure, or info tech infrastructure.

2) *Linearity* . The mean values of $\mu_{Y/X}$ all lie in a straight line, which is the regression line. (See further discussion on how models of best fit were selected for non-linear relations).
3) *Homoskedasticity*[28]. For any fixed value of the independent variable *X*, the distribution of the variable *Y* is normal with mean $\mu_{Y/X}$ (the mean of *Y* for a given *X*) and a constant variance of σ^2. Levene's test of equality in variances was used to test for homogeneity of variances.
4) *Behavior of the residuals*. The residuals are to be assumed normally distributed, independent, with a mean is equal to 0, and a variance of of σ^2. Normal probability P-P plots were used to test for residual normality. Levene's test of equality in variances was used to test for homogeneity of variances.

When evidence of violation of assumption appeared, the variables were transformed to stabilize the variance, achieve normality and linearize the relationship. The transformations were carried out by Y^P where *P* was the suggested power of transformation closest to 0.5. So, if the power of transformation suggested was 5.241, instead of using *Y* as dependent variable, Y^5 was used. Once the variance was stabilized, when the associations were non-linear, models of best fit were tested to describe the associations between the transformed dependent variable and the independent variables. Once the assumptions of the regression models were met, stepwise regressions were used to test the model with the independent control variables.

---

[28] See Appendix P for more on heteroskedasticity, power of transformation, variable transformations, models of best fit, and analysis of residuals.

In sum, the research strategy was designed to overcome the difficulties posed by using varied data sources. By progressively exploring the core relations of the global digital divide and development, and adding more variables to the model we could distinguish the differences that the digital component was making in overall levels of development and changes over the last decade. While such strategy has certain risks, it also provides insight as to what kind of data and information might be needed for future research in this area.