Olin College of Engineering

Computational Probability and Statistics

Searching for a Global Environmental Kuznet's Curve

Examining the Relationship between Economic Development and the Environment

Author:
Jared Kirschner

Instructor:
Dr. Allen Downey

October 25, 2011

1 Abstract

The Environmental Kuznet's Curve (EKC) is a proposed relationship between human economic development and negative environmental behaviors. Initially, development leads to an increase in the negative environmental behavior. However, the trend starts to reverse when a certain level of development is reached, forming an inverted-U shaped curve. The proposed mechanism for the reversal is that at a certain level of empowerment, individuals will be able to influence the behavior if they seek a change. In this paper, we explore the possibility of a global EKC—where each country in the world is a data point on a general, inverted-U shape for a given environmental behavior according to its level of development. This is accomplished by examining the relationship between the Human Development Index (HDI) and components of the Environmental Performance Index (EPI). The evidence also suggests that there is no global EKC for the EPI as a composite statistic nor for any components of the EPI. However, linear correlations between the HDI and components of the EPI are found herein. Strong correlations exist between the HDI and reductions of environmental burden of disease and human impacts from water and air pollution. Moderate correlations exist between improvement in forest health and an increase in contribution to climate change. These correlations are consistent with the predictions of the EKC mechanism.

2 Background

Simon Kuznet, a famous economist, proposed a relationship between income per capita and income inequality. His hypothesis stated that economic inequality will increase as a country develops, but that at a certain average income level, inequality will begin to decrease—forming an inverted-U shape. The idea of an Environmental Kuznet's Curve (EKC) appeared in the 1990s, popularized by the World Bank in its 1992 World Bank Development Report. The EKC proposes that the relationship between income per capita and environmental behaviors follows the same inverted-U shape as the Kuznet's Curve.

To explore the proposed mechanism for this effect, let us examine SO_2 and NO_x air pollution. The major source of SO_2 and NO_x air pollution is electricity generation by coal-fired power plants. At very low levels of income per capita, low demand and lacking infrastructure will prevent the existence of a coal-fired power plant, and levels of SO_2 and NO_x air pollution will be quite low. As the country develops, demand and infrastructure will reach a level which can support a basic coal-fired power plant, leading to an increase in air pollution levels. The effects of the pollution will be felt by the citizens, but they will lack the resources to change their circumstances. As the average income level per capita continues to increase, the citizens will eventually possess the time and money necessary to advocate for a reduction in air pollution. For example, citizens may successfully advocate for a law requiring coal scrubbers on all power plants. This mechanism works best if the environmental impact is "dirty, local, obvious, and now." Otherwise, the people responsible for the impact are less likely to notice or even care about the problem enough to strive for change.[1]

Some versions of the EKC use the "Human Development Index", or HDI, which accounts for literacy rate, school enrollment, and life expectancy in addition to GDP per capita, rather than GDP per capita individually. This is in an attempt to indirectly account for other factors that are necessary to trigger the mechanism of the EKC described in the above example (e.g. an empowered citizenry, a responsive government).

The EKC is generally applied to understand the dynamics of one area over a period of time. For example, one might study indoor air pollution in rural India or water pollution of the Yangtze River in China. By only studying one specific context, the effect of many other variables is controlled for. Some have applied the EKC to a snapshot in time over a region. For example, one might study deforestation across frontier regions in Brazil in the year 2000.[2] This analysis assumes that the behavior of each region of study will have the same functional dependence on GDP per capita or HDI as all other regions.

In this paper, I examine the possibility of a global EKC, where each data point represents the environmental behavior of a single country. This involves first determining which variable—GDP per capita or HDI—is most appropriate for an EKC. After this determination is made, I investigate the relationships between the appropriate variable and various environmental performance indicators.

[1] Information from lectures in a "Social Causes and Consequences of Environmental Problems" course taught by Dr. Elizabeth DeSombre, Frost Associate Professor of Environmental Studies and Associate Professor of Political Science at Wellesley College.

[2] Rodrigues, Ana S, *et al.* Boom-And-Bust Development Patterns Across the Amazon Deforestation Frontier. *Science* 324, 1435 (2009).

3 Description of Data

For this analysis, I will be using two data sets. The first data set is the Hybrid-HDI Data Report 2010 released by the United Nations Development Programme.[3] It contains the information used to compute the Hybrid Human Development Index (henceforth referred to as simply "HDI") from 1970 to 2010 for 135 countries. The HDI for a given country is a number between 0.0 and 1.0 representing the geometric mean between transformed versions of the GDP, GDP_x, life expectancy, LIFE_x, and education, EDU_x (literacy rate, LIT_x, and enrollment, GER_x):

$$\mathrm{HDI} = (\mathrm{GDP}_x \, \mathrm{LIFE}_x \, \mathrm{EDU}_x)^{1/3} \tag{1}$$

$$\mathrm{EDU}_x = (\mathrm{LIT}_x \, \mathrm{GER}_x)^{1/2} \tag{2}$$

The second data set is the Environmental Performance Index (EPI) Data Set for 2010 released by Yale University.[4] It contains the information used to compute the EPI for 163 countries in 2010. The EPI is a complicated, composite statistic which ranges between 0.0 and 100.0 and accounts for many aspects of environmental performance. The process for calculating the EPI is as follows:

1. Gather raw data for each sub-component of the EPI where available and applicable.
2. If the raw data follows an exponential distribution, perform a logarithmic-transformation.
3. Winsorize the data to reduce the effect of outliers on analysis. (discussed in detail below)
4. Set the targets for each sub-component of the EPI and perform a logarithmic transform if the same was done to the sub-component.
5. Compute the proximity-to-target of each subcomponent, and then combine by a weighted average to give a proximity-to-target value for the component. If one of the sub-components is missing, adjust the weighting to compensate. (discussed in detail below)
6. Compute the EPI by a weighted average of the components as shown in Figure 1.

Winsorization is a process which reduces the effect of spurious outliers on further analysis. For a winsorization at percent w, the values below the $\frac{w}{2}^{\mathrm{th}}$ percentile are coerced to the value of the $\frac{w}{2}^{\mathrm{th}}$ percentile. The same is done for the $\left(100 - \frac{w}{2}\right)^{\mathrm{th}}$ percentile. For this data set, w is not given, but appears to be 5%—a fairly standard threshold.

The "Proximity-to-Target" (PTT) values range between 0.0 and 100.0 representing how close the value is to the target. According to Yale, where high values equate to good performance (e.g. protected areas coverage),

$$\mathrm{PTT} = 100 \times [1 - (\text{target value} - \text{winsorized value})/(\text{target value} - \text{minimum winsorized value})]$$

Where high values equate to bad performance (e.g. air pollution emissions),

$$\mathrm{PTT} = 100 \times [1 - (\text{winsorized value} - \text{target value})/(\text{maximum winsorized value} - \text{target value})]$$

The exact formulation for the EPI can be seen in Figure 1. For the purposes of analysis, I use the composite proximity-to-target values. The proximity-to-target data is easier to analyze because it provides a consistent context in that 100.0 and 0.0 represent the best and worst state

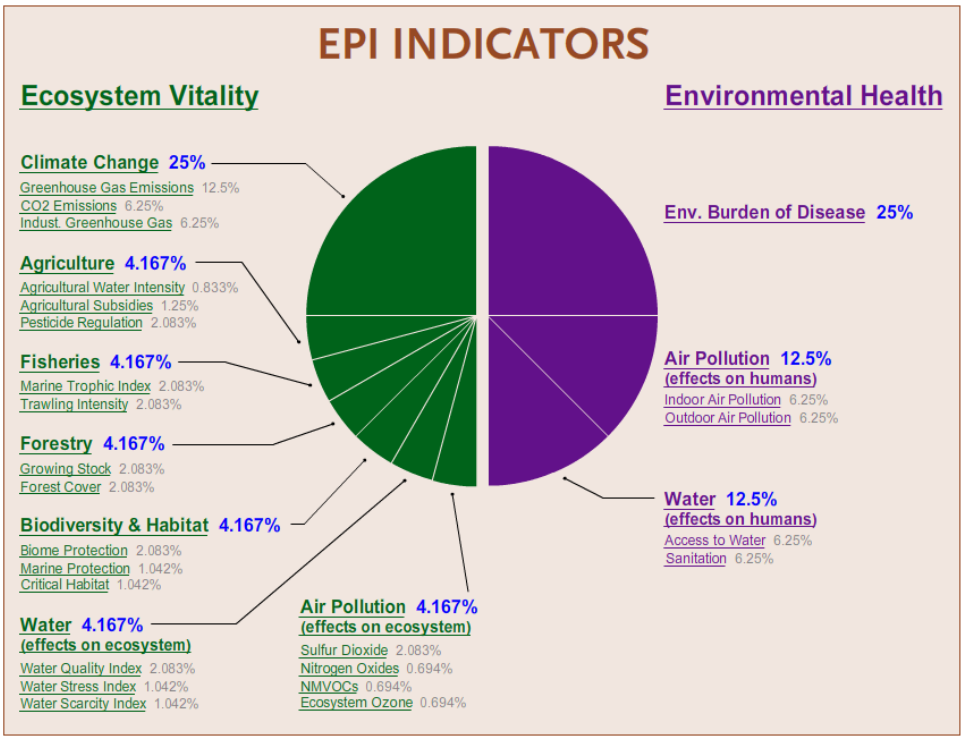


Figure 1: Formulation of the Environmental Performance Index (EPI) from Yale University. This index is used to evaluate countries around the world on many aspects of the integrity of their environment.

for a given environmental condition. The raw data is sometimes difficult to interpret because the units are not reported and, as such, it is difficult to contextualize the information.

I use only the information from the Hybrid-HDI Report for the year 2010. This information is then merged with the EPI data. I accomplish this by matching the keys which both data sets have in column: country names. Unfortunately, not all of the country names are consistent between the two tables (e.g. United States vs. Unites States of America, Russia vs. Russian Federation). Matching country names by the best possible match between the two data sets does not work in all cases, as many country names have certain elements in common (e.g. "Democratic", "Republic"). Some country names in the data files were manually edited to assure a match between the files. Any country names for which matches did not exist between the data files are thrown out for analysis. I will later (Section 6) analyze this process to determine whether it introduces bias to the results.

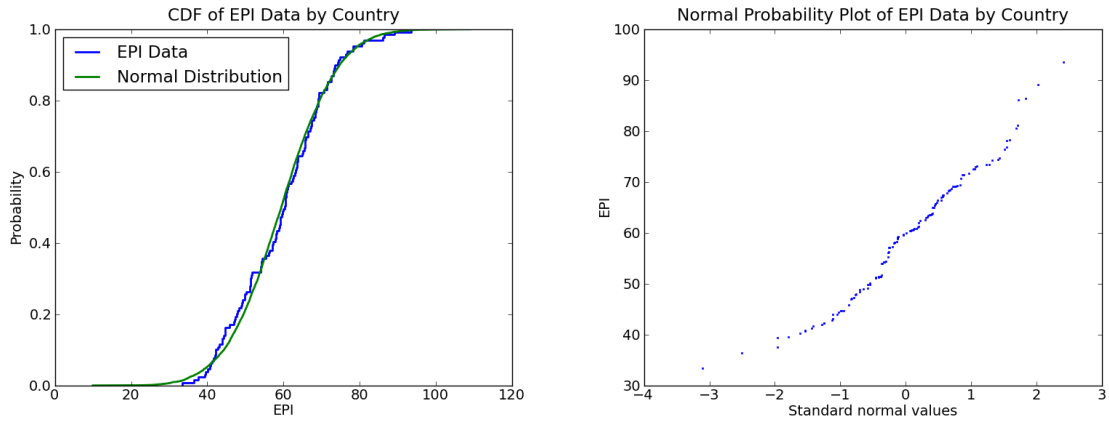
4 Examining the Data

Before deciding which questions to ask and how to ask them, it is important to first examine and visualize the available data. Applying tests without first understanding the underlying distributions of data can result in incorrect or misleading results.

[3] `http://hdr.undp.org/en/media/2010\_Hybrid-HDI-data.xls`
[4] `http://epi.yale.edu/file\_columns/0000/0158/2010epi\_data.xls`

4.1 EPI

All elements of the investigation will analyze the relationship between a variable and the EPI, or the EPI and a variable. As discussed above, the EPI is composed of a weighted sum of many factors which have been transformed through an unknown method. Before working with the EPI data, it is important to understand the distribution of the EPI. To quickly visualize the distribution, we compute the compute the cumulative distribution function (CDF) of the EPI data, as shown in Figure 2a.



(a) CDF of EPI data compared to a normal distribution.

(b) Normal probability plot of EPI data.

Figure 2: To visualize the distribution of the EPI data, we compute the CDF, shown in Figure 2a. The shape of the distribution is well modeled by a normal distribution with mean and variance equal to that of the EPI data. We can more directly compare the shape of the EPI distribution to that of a normal probability plot by creating a normal probability plot, as shown in Figure 2b, which plots the EPI data against a normal distribution. If the distribution of the data is normal, the normal probability plot will show a straight line.

A normal distribution with mean of 59.5 and standard deviation 12.1 models the distribution of EPI data reasonably well across most of the distribution but with greater deviations in the left-tail, as evidenced by the normal probability plot shown in Figure 2b. The Central Limit Theorem offers an explanation as to why the EPI, a sum of samples from unknown distributions, would follow a normal distribution. According to the Central Limit Theorem, the sum of n values pulled from any similar set of distributions with finite variances will converge to a normal distribution as n increases. The actual distribution is below the normal distribution near 100.0 because of the truncation of distributions that occurs at 100.0 for the components of the EPI. At the left-tail of the distribution, economic mechanisms are likely to limit the downside of the composition of components which forms the EPI.

4.2 GDP per Capita

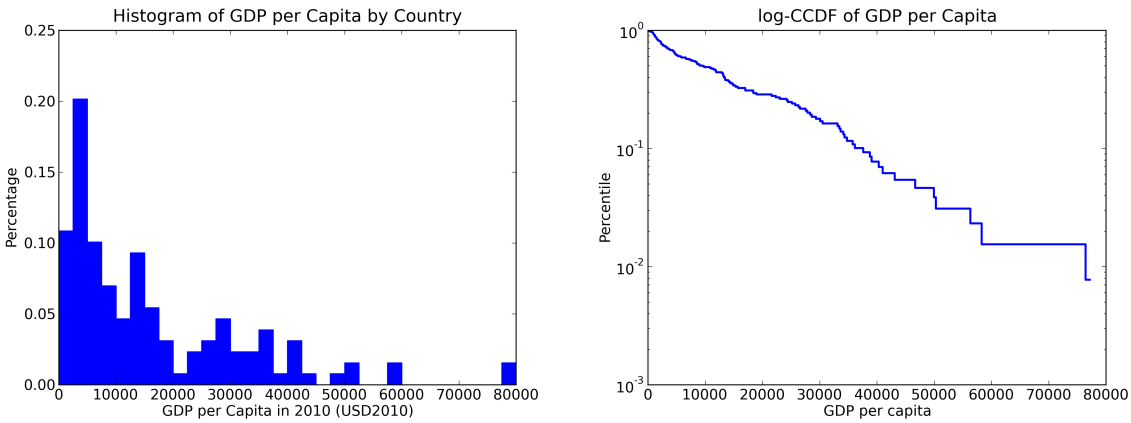
The Environmental Kuznet's Curve traditionally compares the average wealth (GDP per capita) of a country to an environmental factor. To perform such analysis properly requires an understanding of the GDP per capita data. Figure 3 shows visualizations of the data. The binned PMF of the data (Figure 3a) shows a distribution which appears to be exponential. To determine whether or not the data follows an exponential distribution, we can investigate the CDF. For an exponential distribution, the CDF is defined as:

$$\mathrm{CDF}(x, \lambda) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

One way to test whether or not a CDF follows an exponential distribution is to plot its complementary CDF (CCDF) on a log-y scale. $\mathrm{CCDF}(x)$ is defined as $1 - \mathrm{CCDF}(x)$. Thus,

$$\begin{aligned} \mathrm{CCDF}(x, \lambda) &= 1 - \mathrm{CDF}(x, \lambda) \\ &= e^{-\lambda x} \\ \log\left(\mathrm{CCDF}(x, \lambda)\right) &= -\lambda x \end{aligned}$$

Figure 3b shows the CCDF of the GDP per capita data plotted on a log-y scale. If the distribution is exponential, we would expect Figure 3b to show a line with a slope of $-\lambda$. The plot confirms that the distribution of GDP per capita is, in fact, from an exponential distribution. Therefore, the log-transformed GDP per capita should approximate a discrete uniform distribution. A log-transformed GDP per capita will be more amenable to analysis because it is more reflective of differences in income levels between countries.



(a) Binned PMF of GDP per capita by country.

(b) CCDF of on a log-y scale of GDP per capita by country.

Figure 3: To visualize the distribution of the GDP per capita data, we can visualize a binned PMF, shown in Figure 3a. The shape of the distribution appears to be somewhat exponential. To test this thought, we can compute the CCDF of GDP per capita and then visualize it on a log-y scale, shown in Figure 3b. The relationship is straight for the majority of the distribution, further suggesting an exponential distribution.

We could further characterize the distribution of GDP per capita by estimating the parameter λ for the distribution. To perform this estimation, I will proceed as follows:

1. Define a suite of hypotheses for the value of λ, each with an associated prior probability.
2. Compute the posterior probability of each hypothesis in the suite after observing the available evidence (GDP per capita data).
3. Normalize the suite of probabilities. The distribution of the curve should have a peak representing the most likely hypothesis for the value of λ.

For my suite of hypothesis, I defined a list of 1000 evenly spaced guesses for λ between 1×10^{-6} and 1×10^{-4}. Because I have no reason to suspect a particular value of λ over another, I will use a discrete uniform distribution for my hypothesis, giving them each a value of $\frac{1}{n} = 0.001$. After computing the posterior probabilities and normalizing, I obtain the distribution shown in Figure 4.

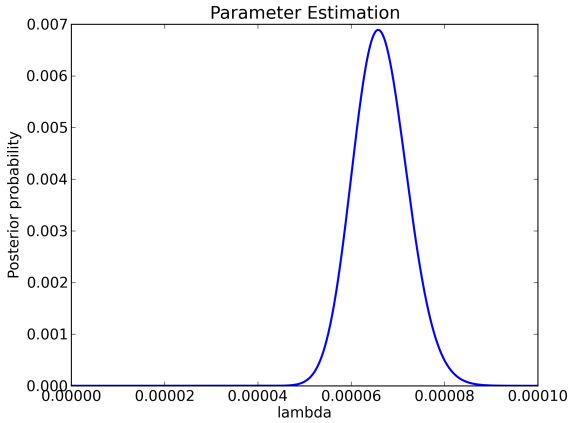


Figure 4: This PMF is a distribution of the posterior probabilities of a suite of hypotheses for the value of lambda for an exponential model of GDP per capita. The GDP per capita numbers are used as evidence to update the probabilities of a prior discrete uniform distribution.

The maximum likelihood estimator for lambda is the peak of the distribution—6.6×10^{-7}—with a 90% credible interval of 5.9×10^{-7} and 7.4×10^{-7}. The 90% credible interval is found by converting the posterior distribution into a CDF and searching for the 5th and 95th percentile.

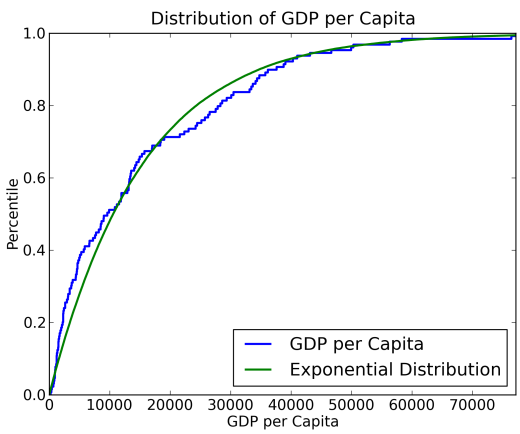


Figure 5: Comparison of the actual CDF of the GDP per capita data against an exponential model with $\lambda = 66 \times 10^{-6}$. The model fits the data fairly well across the range of the distribution.

If we compare an exponential model with $\lambda = 6.6 \times 10^{-7}$ to the actual observed distribution of GDP per capita (as seen in Figure 5), we see that the model represents the data fairly well. Some error in the model is introduced by the fact that there is a non-zero maximum probability for GDP per capita (refer to Figure 3a). In an exponential distribution, values closer to zero always have a higher probability than values farther from zero.

4.3 HDI

The HDI is sometimes used instead of GDP per capita with Environmental Kuznet's Curves in an attempt to include important non-monetary aspects of the human condition. Figure 6 shows a visualization of the distribution of HDI data by country. This particular distribution does not appear to follow a common continuous model (normal, log-normal, exponential, Weibull). Above the 30th percentile, the distribution appears approximately uniform. Below the 30th percentile, the distribution appears sigmoidal.

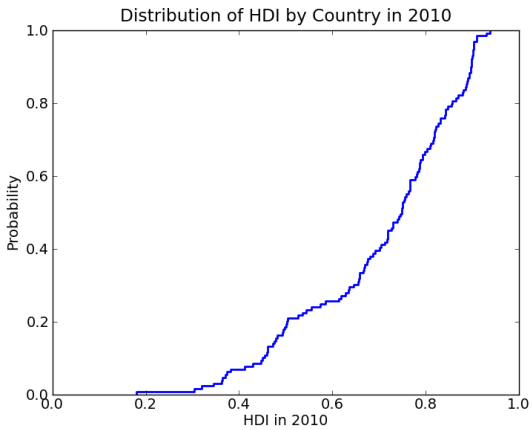


Figure 6: CDF of the HDI by country. This particular distribution does not appear to follow a common continuous model (normal, log-normal, exponential, weibull).

5 Choosing a Variable for Analysis: HDI or log-GDP per Capita

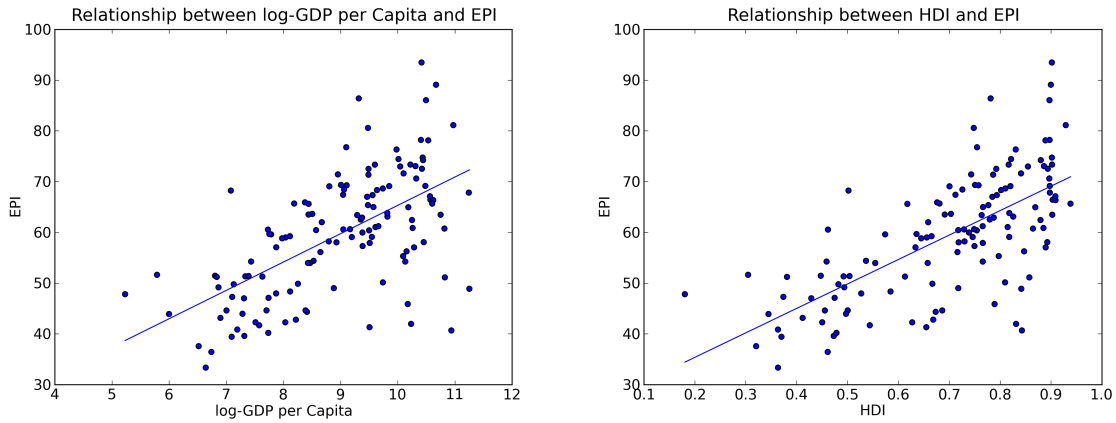
In order to perform the analysis of various environmental factors using the method proposed by the Environmental Kuznet's Curve, we must decide which independent variable to use. In this case, the choice is between the Human Development Index (HDI) and log-GDP per capita. This decision can be made statistically by calculating the correlation between the candidate variables and the EPI and selecting the variable with the highest correlation.

In deciding which correlation statistic to compute, I consider the distribution of the variables. Neither the HDI nor log-GDP have significant outliers. The EPI is approximately normally distributed, but neither HDI nor log-GDP is normally distributed. The Pearson correlation is not robust to distributional differences or outliers. When a difference in distribution is present, the Pearson correlation coefficient tends to understate the correlation. Thus, I use the Spearman correlation to compute the correlation.

Scatter plots comparing HDI and log-GDP with EPI are shown in Figure 7. HDI ($s = 0.678$) has a slightly higher Spearman correlation coefficient than log-GDP ($s = 0.596$). For this reason, I will use HDI as the independent variable in investigation of the relationship between economic development and environmental factors.

For both HDI and log-GDP, there is a positive relationship between the variables and the EPI. The EKC predicts that for a particular negative environmental behavior, the curve will follow an inverted U-shape. As the EPI is a positive indicator, the EKC would predict a U-shaped graph. However, the scatter plots in Figure 7 show no indication of a U-shaped relationship. There are several possible explanations for this:

Option 0: The EKC trajectory of individual countries is different enough through time that



(a) log-GDP per capita versus EPI ($s = 0.596$).

(b) HDI versus EPI ($s = 0.678$).

Figure 7: Comparison of the strength of the relationship between log-GDP or HDI and EPI. A correlation test shows that HDI correlates more strongly than log-GDP with EPI.

a global EKC will not result from data collected at one time from countries at different HDI levels.

Option 1: Countries with low HDI values have been excluded from the data set. If the inflection point occurs near the HDI values of most excluded countries, then it is possible that half of the U-shape would be missing.

Option 2: The EKC does not apply to the EPI as a composite statistic, but may apply to its components. As the EKC is a combination of many environmental factors, each with potentially different points of inflection, a weighted summation of U-shaped relationships may result in an overall linear relationship across the majority of the distribution.

I examine the plausibility of options 1 and 2 in Sections 6 and 7 respectively. If neither option is plausible, I will conclude that the null option (0) is the most likely explanation for the lack of an EKC behavior for the EPI.

6 Examining Bias Introduced by Data Selection Process

The data set I used for analysis merged the two data sets discussed in section 3. Rows from the tables were combined based on a match between countries in the two tables. If a particular country was not present in both data sets, it would not be present in the final data set used for analysis. Unfortunately, the country selection process is not given for either data set. Factors such as small populations, great political instability, and lower GDP per capita or HDI may contribute to a country being excluded from my data set.

The EPI data set contains data from many countries (231), but not enough in all cases to compute the EPI. The majority of countries excluded from my data set (67%) were present in EPI data set, but lacked EPI scores. Therefore, in order to search for sources of bias, we focus on the exclusion process resulting from missing EPI values.

The EPI countries can be separated into two groups—those included in and excluded from the merged set. Unfortunately, those countries lacking EPI scores also lack GDP per capita and HDI data. Instead, I can indirectly test for a bias in the data associated with lower HDI.

For the indirect method, I use the EPI region categorization that is provided for all countries in the EPI data set (including those without an EPI value). If we assume that HDI is similar across countries within EPI region categories, we expect excluded countries to have HDI values within the distribution of values for the associated EPI region. Hence, if there exists an unequal geographical distribution of excluded countries, one possible causal factor is the relative level of wealth or development of the different EPI regions.

To test for an unequal geographical distribution of excluded countries, I perform a Chi-Squared (χ^2) Goodness of Fit test. The hypothesis for the test are:

Null Hypothesis, H_0: Countries without a computed EPI value are equally distributed across all EPI regions.

Alternative Hypothesis, H_1: Countries without a computed EPI value are not equally distributed across EPI regions.

I perform the test in the following series of steps:

1. I define the set of cells each country can fall into. With two types of countries (excluded and included) and eight EPI regions, the test will use sixteen cells.

2. I compute the expected number of countries in each cell. Under the null hypothesis, the expected number of countries in each cell is equal to the total number of countries in that region multiplied by the percent of included or excluded countries across all regions.

3. I compute the Chi-Squared statistic, which is a measure of the deviation of observed values O_i from expected values E_i:
$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

4. I use a Monte Carlo simulation to compute the p-value, which is the probability of seeing a χ^2 value as high as I observed under the null hypothesis.

Using this process I calculated a χ^2 test statistic of 37.4, which has a p-value of < 0.0001. I reject the null hypothesis and accept the result as evidence in favor of the alternative hypothesis.

Now, to see whether or not the uneven distribution is associated with HDI, I sort the EPI regions by their median HDI and then compare their EPI value inclusion to the average inclusion across all regions. Table 1 shows this comparison. EPI inclusion does not appear to correlate with increasing HDI. After examining the excluded countries, it appears that most are small, island nations. As such, exclusion is likely caused by low GDP or small population rather than low HDI. The uneven distribution is likely the result of an uneven distribution of islands across the 8 EPI regions. Therefore, I conclude that option 1 in Section 5 is implausible.

7 Searching for a Global Environmental Kuznet's Curve

I now test the plausibility of option 2 from Section 5. To search for the presence of an Environmental Kuznet's Curve in the individual components of the EPI, I create scatter plot of HDI and said components. Scatter plots showing the relationship between the HDI and the 10 components of the EPI are shown in Appendix A. A quick examination of the curves reveals that there is no inverted-U shape. I now accept option 0 from Section 5—that there is no global EKC.

EPI Region	# of Countries	Median HDI	% Diff. from Avg. EPI Inclusion
Sub-Saharan Africa	50	0.456	16.2
South Asia	8	0.502	24.0
East Asia & the Pacific	44	0.686	-35.6
Eastern Europe & Central Asia	18	0.743	41.7
Latin America & Caribbean	45	0.755	-18.1
Middle East & North Africa	21	0.766	28.2
Europe	39	0.885	9.0
North America	6	0.902	-52.8

Table 1: EPI regions sorted by median HDI. There does not appear to be a correlation between EPI inclusion and HDI.

Environmental Issue	Pearson Coeff. (ρ)	Coeff. of Determination (R^2)	Slope
Environmental Burden of Disease	0.945	0.892	141.2
Human Effects of Water	0.884	0.782	146.7
Human Effects of Air Pollution	0.779	0.607	124.8
Climate Change	-0.583	0.340	-57.1
Forest Health	0.491	0.241	62.0
Ecosystem Effects of Air Pollution	-0.301	0.091	-20.5
Agriculture	0.158	0.025	17.4
Ecosystem Effects of Water	0.129	0.017	15.1
Biodiversity	-0.122	0.015	-19.1
Fisheries Health	-0.113	0.013	-14.7

Table 2: Results of correlation and linear least squares testing between HDI and the components of the HDI. Results are grouped horizontally by relative strength of the correlation (strong, moderate, and weak or non-existent).

It does, however, appear that there are linear relationships between HDI and many of the components of the EPI. In order to investigate the nature of these relationships, I perform correlation and linear least squares fit analysis on the scatter plot data. A linear least squares fit will created a linear model between the variables which minimizes the mean squared error between the model and the data.

Table 2 shows the results of the correlation and linear least squares fit analysis. The Pearson coefficient ρ is within the range of [-1.0,1.0] and shows the strength of the correlation between the two variables. The sign of ρ signifies the direction of the relationship and the magnitude of ρ signifies how strong the correlation is. The coefficient of determination R^2 shows the proportion of variability in the data reduced by the linear least squares fit, and is equal to ρ^2. The slope, as calculated by the linear least squares fit, shows how strong of a change is induced in the EPI coefficient by a change in HDI.

The results in the table are grouped into three categories by the strength of the relationship (strong, moderate, and weak or non-existent). From the mechanisms of EKC, we expect the strongest correlations for components of the EPI with impacts which are "dirty, local, obvious, and now." If an issue impacts people that is close in a time and space to an identified cause of the problem, the affected will be more likely to act for change.

The three strong positive correlations all fit this description. Environmental burden disease measures the impact of human health problems which can be contributed to environmental factors. Human effects of water measures access to clean water, and human effects of air

pollution measures the direct impact of pollution on humanity. The only moderate positive correlation is forest health. As deforestation primarily affects people local to the problem and interacts with other problems such as soil erosion, it could be expected to follow EKC mechanisms. Note that the strength of the correlation with forest health may be understated, as the many countries have surpassed the target (refer to Figure 10b).

The only moderate negative correlation is for climate change, which is a combined measure of greenhouse gas emissions per capita, CO_2 emissions per unit electricity produced, and industrial greenhouse production intensity. Climate change is the least likely issue to follow the mechanisms of the EKC because the impacts of an individual's contribution to climate change are globally distributed, difficult to pinpoint, and distant in both time and space.

The remaining issues are generally environmental components with impacts which are not felt directly by humanity in a short time horizon. The decline of fish stocks and diversity (fisheries health) and land problems caused by certain agricultural practices have impacts which are distant in time from the causal actions. Negative impacts on the ecosystem and biodiversity tend to have little impact on humanity, at least on the short term.

An examination of the slopes calculated by the linear least squares fit reveals that the EPI components with a greater correlation also have much higher slopes. This indicates both a strong correlation and a strong relationship between the HDI and these components. As a point of reference, consider that a slope of 100.0 would connect the bottom-left and top-right corners of the scatter plots. A slope of greater than 100.0 indicates that the ratio of incremental gain in proximity to the target of the EPI component is greater than the associated incremental gain in HDI.

8 Effects of the Environment on Human Development

We have already seen that human development influences environmental factors. However, it is also possible that the environment influences human development, as human activity must fundamentally interact with the environment. I explore this possibility here.

8.1 Examining the Effect of Being Landlocked

A country which is landlocked has no access to open seas. Access to open seas is important for international maritime trade. For a landlocked country, all imports or exports by ship must go through at least one other country first. The tariffs and land transportation costs associated with trade for a landlocked country could conceivably reduce trade opportunities.

A visual comparison of the CDFs of the HDI values for the two groups (landlocked and not landlocked countries) reveals an apparent difference between the groups, as shown in Figure 8. To test whether the apparent effect that being landlocked affects the human development of a country, I perform a difference in means test with the following hypothesis:

Null Hypothesis, H_0: The distribution of HDI values for all countries is the same, regardless of whether or not said countries are landlocked.

Alternative Hypothesis, H_1: The distributions of HDI values between countries which are and are not landlocked is different.

I perform the difference in means test as follows:

1. Compute the actual difference in means between landlocked and not landlocked countries.

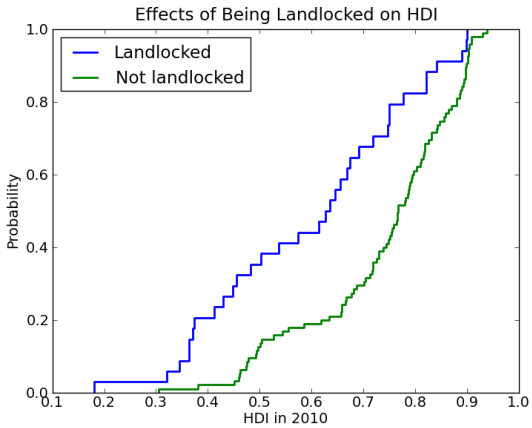


Figure 8: Comparison of the CDFs of the HDI values for landlocked and not landlocked countries. The apparent difference between means in the distributions can be tested statistically by a difference in means hypothesis test.

2. Run a series of n simulations to determine the difference in means of sample sets equal in length to the sets of landlocked and not landlocked countries. These sample sets are created under the assumption that the null hypothesis is true. This is accomplished by generating samples with replacement from the set of all countries.

3. Count the number of times c the difference in means for the simulated data was greater than or equal to the difference in means of the observed data. As a p-value signifies the probability of observing an effect as large as the observed effect under the Null Hypothesis, $p = \frac{c}{n}$.

The resulting difference in means for the HDI between countries which are and are not landlocked is -0.141 with a p-value of < 0.0001. This result favors the alternative hypothesis, indicating a strong relationship between a country being landlocked and a country having a low HDI. This is likely caused by the additional expense associated with participation in international maritime trade.

9 Conclusion

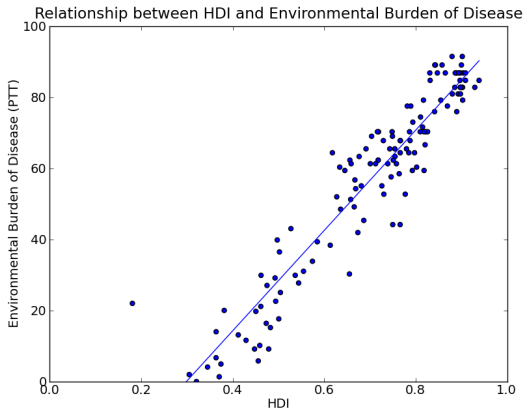
An initial examination of the behavior of the EPI in relation to GDP per capita and HDI suggested that a global EKC does not exist, at least in the context of the EPI as a composite statistic. Two reasons for this were proposed: (1) a bias in the data which unevenly excluded countries with a lower HDI, or (2) EKC behavior existed in the individual components of the EPI, but differed enough by component to create an additive result without EKC behavior. Both of these proposals were eliminated as possibilities in the analysis. As such, I conclude that there is no global EKC. This is likely due to the falsehood of the fundamental assumption that all countries will have approximately the same functional dependence on HDI in relation to indicators of environmental performance. While this could be a limitation of the HDI as a measure of human development, it seems more likely that the EKC model is limited to studies of homogenous regions.

Though a global EKC was not observed, linear correlations between the HDI and components of the EPI were found. These correlations were consistent with the theory behind

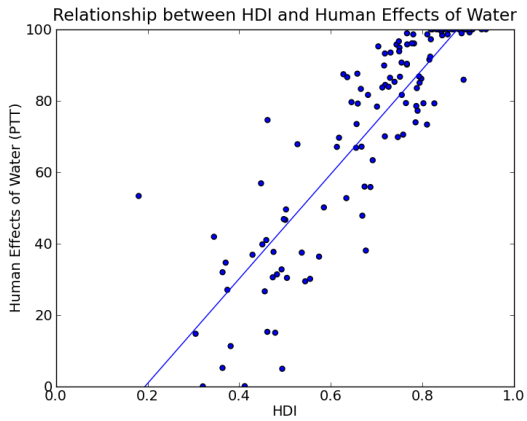
the mechanisms of the EKC. Strong correlations exist between the HDI and reductions of environmental burden of disease and human impacts from water and air pollution. Moderate correlations exist between the HDI and the improvement of forest health and increased contribution to climate change. In general, the EKC mechanism works well with environmental issues that directly impact humanity in a close spatial and temporal range from the source of the problem. Among the components of the EPI, contribution to climate change is the least likely to follow an EKC mechanism, which is probably why it is the only issue that continues to get worse with increasing development at a high level of correlation.

A Scatter Plots of HDI and Environmental Aspects

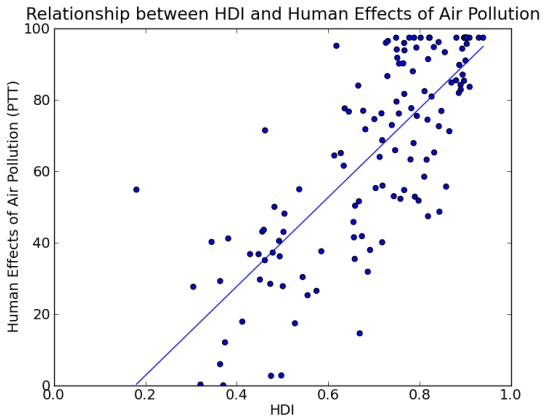
All figures in this section are scatter plots comparing HDI to components of the EPI. Figures are grouped according to the strength of their correlations. Pearson correlation values (ρ) are given for each graph in the caption. The line of best fit as determined by a linear least squares algorithm is shown as a blue line on each graph.



(a) Environmental Burden of Disease ($\rho = 0.945$)

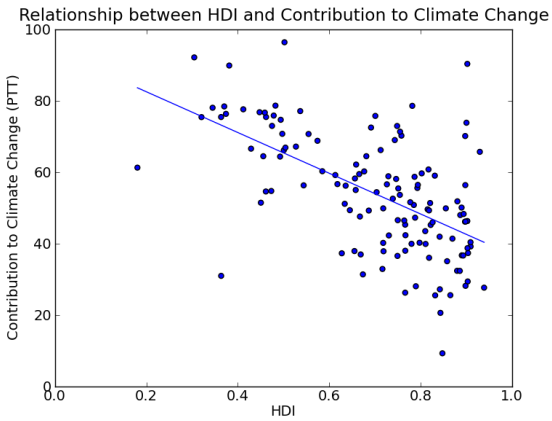


(b) Human Effects of Water ($\rho = 0.884$)

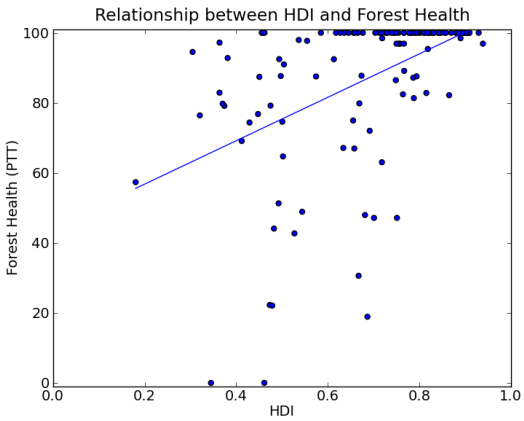


(c) Human Effects of Air Pollution ($\rho = 0.779$)

Figure 9: Strong correlation between HDI and EPI components.

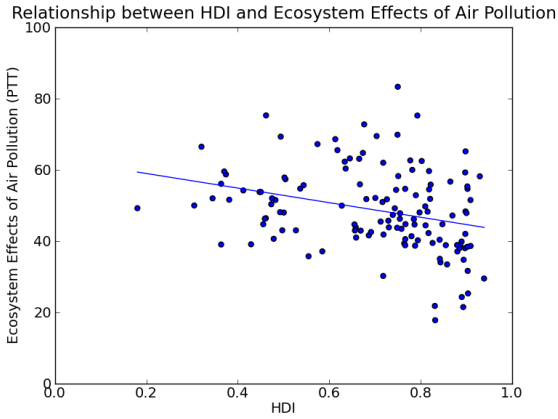


(a) Climate Change ($\rho = -0.583$)

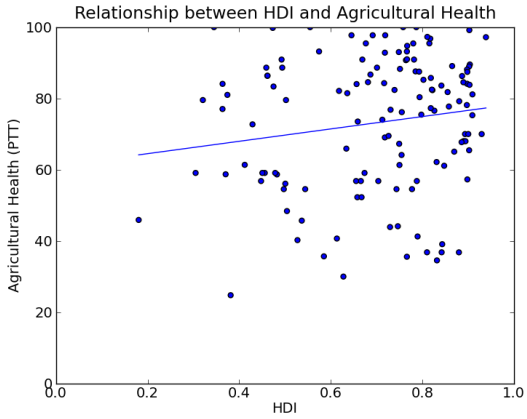


(b) Forest Health ($\rho = 0.491$)

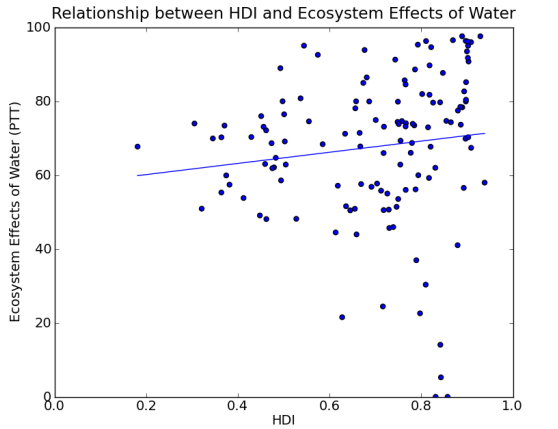
Figure 10: Moderate correlation between HDI and EPI components.



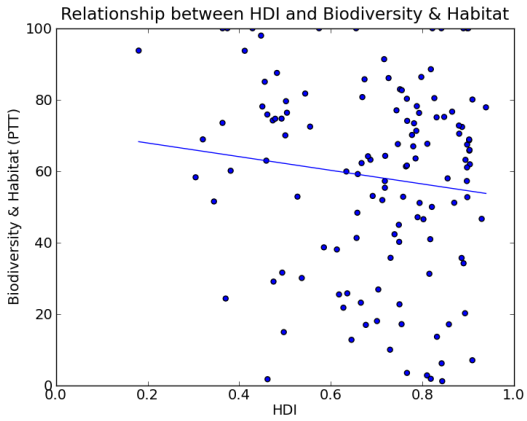
(a) Ecosystem Effects of Air ($\rho = -0.301$)



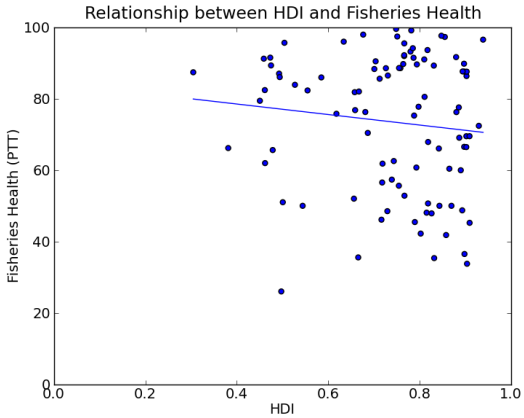
(b) Agricultural Health ($\rho = 0.158$)



(c) Ecosystem Effects of Water ($\rho = 0.129$)



(d) Biodiversity ($\rho = -0.122$)



(e) Fisheries Health ($\rho = -0.113$)

Figure 11: Weak or no correlation between HDI and EPI components.