An Algorithm for Predicting Carbon Dioxide Emissions from Economic Related Indicators

Final report prepared for the Data Analysis and Interpretation Specialization

March 8, 2017

Introduction and Motivation to Research the Project

The purpose of this research is to measure the best predictors of a country's carbon dioxide (Co2) emission contribution from several economic indicators such as energy depletion rate, export and import of goods and services, natural resource depletion rate, food production, urbanization rate, industry, working population rate, and gross domestic product (GDP) per annum.

We live in an age where the planetary boundaries (such as climate change, ocean acidification, ozone depletion, heavy agriculture loading of nitrogen and phosphorus, overuse of flesh water resources, and biodiversity loss) are getting stressed in addition to increasing poverty in many parts of the world. Global economic activities have pushed the subject of climate change and sustainability to the mainstream where individuals, businesses and governments are facing increasing pressure to take action to curve greenhouse gas emissions in order to avoid catastrophic impacts from global warming. Co2 emissions accounted for 78% of total emissions in 2010 (Center for Sustainable Systems, University of Michigan. 2016). The intergovernmental panel on climate change (IPCC) is currently sure that 95 % of global warming is caused by activities (IPCC, 2014). The starting point of tackling the problem of global carbon emissions is to measure and understand which economic predictors are associated most Co2, the most abundant green house gas that endangers the climate system so that appropriate adaptation and mitigation strategies can be developed towards a sustainable future of our planet.

According to the United Nation's sustainable development goal (UN SDGs, n.d) #13, climate change will affect every country on every continent with the poorest and most vulnerable being affected the most.

As an environmentalist and a climate change scientist, I feel the urge to contribute to the growing knowledge of global effort of transitioning into low carbon economy and thereby curbing future climate change impacts. Investigating the relationship between economic activities and climate change will help low income countries to channel their scarce resources to help develop resilience strategies. This study will help middle income countries to develop a balance approach between economic activity and emission control initiatives. In addition, this study will contribute to the glowing need for high income countries to develop energy efficiency technologies and a shift towards renewable energy economy.

Methods

Sample

The sample (N=248) is from the world bank development indictors (WDI), compiled from officially-recognized international sources in 2013. It presents the most accurate global development data available, and included 163 developmental indicators measured from a panel of 248 countries and regions of the world. Some of the WDI includes gross domestic product, total employment rate, and estimated HIV prevalence in 2012 and 2013.

Measures

The Co2 emissions was measured in terms of carbon dioxide damage for each country by multiplying $20 per ton of carbon (the unit damage in 1995 U.S dollars) and the number of tons of carbon emitted and expressed in units of GNI. GNI (gross national income) is the sum of value added by all resident producers plus any product taxes (minus subsidies) not included in the valuation of output plus net receipts of primary income (compensation of employees and property income) from abroad (World bank, n.d)

Predictors included 1) energy depletion expressed as the ratio of the value of the stock of energy resources to the remaining reserve lifetime (capped at 25 years). It covers coal, crude oil, and natural gas, 2) export of goods and services and other markets provided to the rest of the world (include the value of merchandise, freight, transport, travel, business and government services, 3) import of goods and services and other markets received from the rest of the world (include the value of merchandise, freight, transport, travel, business and government services 4) natural resources depletion expressed as the sum of net forest depletion, energy depletion, and mineral depletion, 5) food production activities covering food crops that are considered edible and that contain nutrient, which involves the use of fertilizer and mechanized agriculture, 6) urbanization rate referring to people living in urban areas as defined by national statistical offices , 7) industry added as a separate net output of the industrial sector, 8) working population between 15 and 64 age and 8) an annual percentage growth of GDP at market price based on constant local currency. This presents the total value of all goods and services produced. There are inherent activities in all the predictors that can spur climate change through the emission of more Co2 into the atmosphere.

Analyses

The distributions for the predictors and the carbon dioxide damage response variable were evaluated by performing descriptive statistics for quantitative variables, including calculating the mean, standard deviation and minimum and maximum values for quantitative variables.

Scatter plots were also examined, and Pearson correlation analysis was used to test bivariate associations between individual predictors and the carbon dioxide damage response variable.

Least absolute shrinkage and selection operator (LASSO) regression with the least angle regression selection algorithm was used to identify the subset of variables that best predicted carbon dioxide damage. The LASSO regression model was estimated on a training data set consisting of a random sample of 70% of the total dataset (N=174), and a test data set included the other 30% of the total dataset (N=74). All predictor variables were standardized to have a mean=0 and standard deviation=1 prior to conducting the lasso regression analysis. Cross validation was performed using k-fold cross validation specifying 10 folds. The change in the cross validation mean squared error rate at each step was used to identify the best subset of predictor variables. Predictive accuracy was assessed by determining the mean squared error rate of the training data prediction algorithm when applied to observations in the test data set.

Results

Descriptive Statistics

Table 1 shows descriptive statistic for Carbon dioxide emissions calculated in terms of carbon dioxide damage and the quantitative predictors., including one categorical variable (GDP). The mean carbon dioxide damage was 0.5 % of GNI (sd=0.4), with a minimum damage of 0.05 % of GNI and a maximum of 2.49 % of GNI.

Table 1. Descriptive Statistics for Data Analytic Variables					
Analysis Variables	N	Mean	Std Dev	Minimum	Maximum
Carbon Dioxide Damage (X12_203)	170	0.5	0.40	0.05	2.49
Energy Depletion (X18_2013)	170	3.60	7.32	0.00	53.91
Natural Resources Depletion (X21_2013)	170	6.27	8.2	0.00	53.96
Export of Goods and Services (X121_203)	170	40.71	23.52	6.31	195.57
Food Production (x129_2013)	170	120.98	21.48	63.53	196.38
GDP (X140_2013)	170	3.24	4.19	-36.05	14.04
Import of Goods and Services (X154_2013)	170	46.50	21.71	13.00	161.91
Industry Value Added (X161_2013)	170	29.11	11.51	7.74	72.02
Working Population (X222_2013)	170	63.59	6.59	47.10	84.01
Urbanization (X283_2013)	170	56.33	21.51	8.67	99.06

Bivariate Analyses

Scatter plots for the association between the carbon dioxide damage response variable and quantitative predictors (Figure 1; Table 2) revealed that carbon dioxide damage increased when industry value added increased (Pearson $r=0.38$, $p<.0001$), working population increased (Pearson $r=0.34$, $p<.0001$) and energy depletion increased (Pearson $r=0.29$, $p=0.0002$). Carbon dioxide damage was not significantly associated with urbanization rate, food production, export of goods and services and import of goods and services.

Figure 1. Association Between Quantitative Predictors and Carbon Dioxide Damage

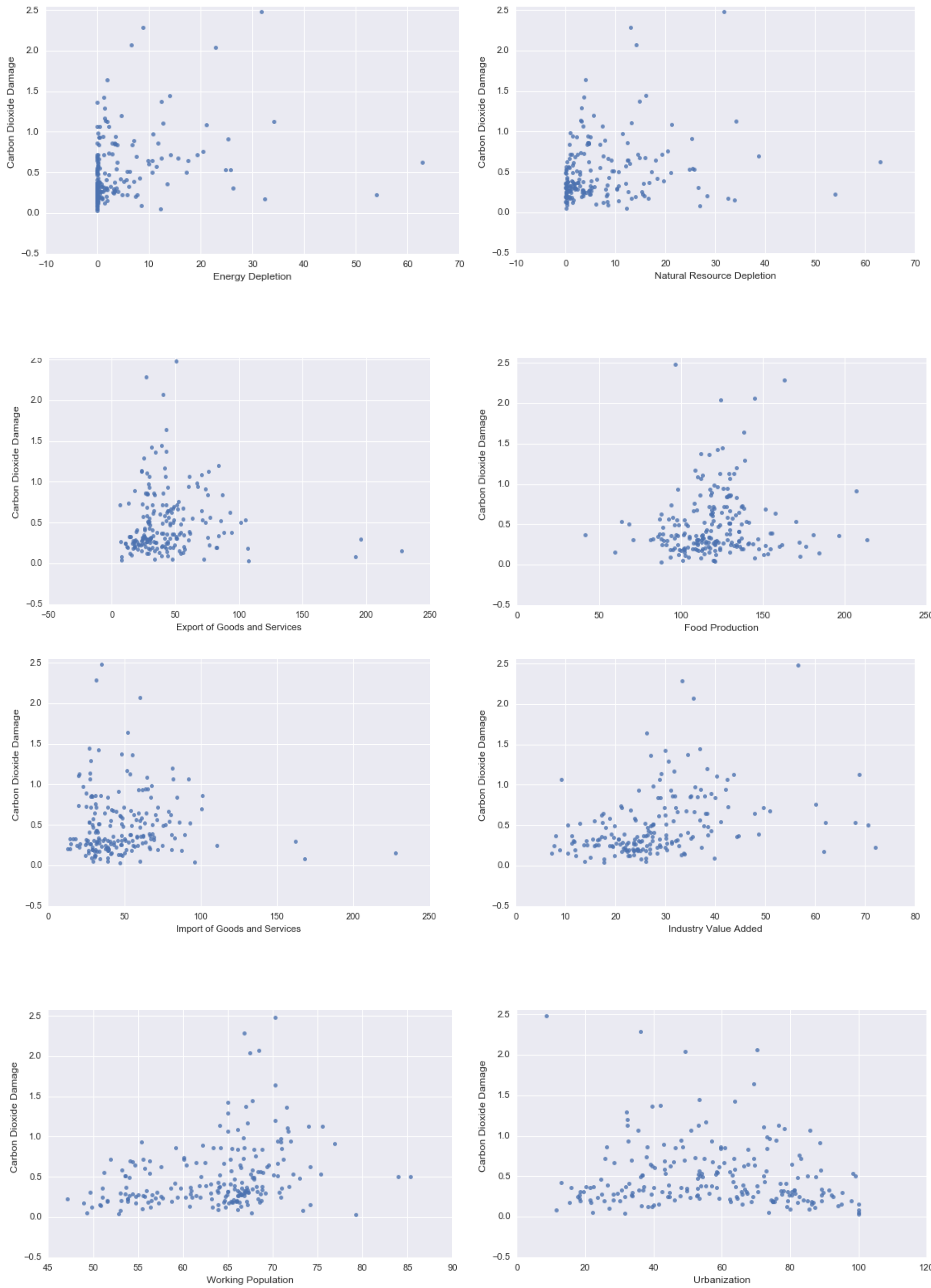


Table 2. Pearson Correlation Results					
Response - predictors	r - value	p-value/ significance	Direction	Form	Strength
Carbon dioxide damage – Energy Depletion	0.2855	0.0002	Positive	Linear	Weak
Carbon dioxide damage – Natural Resources Depletion	0.2049	0.0073	Positive	Linear	Weak
Carbon dioxide damage – Export of Goods and Services	0.0533	0.4898	Positive	Linear	Very weak
Carbon dioxide damage – Food Production	0.0885	0.2513	Positive	Linear	Very weak
Carbon dioxide damage – Import of Goods and Services	0.0277	0.7196	Positive	Linear	Very weak
Carbon dioxide damage – Industry Value Added	0.3767	4.1380e-7	Positive	Linear	Week
Carbon dioxide damage – Working population	0.3383	6.4324e-6	Positive	Linear	Week
Carbon dioxide damage – Urbanization	-0.1234	0.1090	Negative	Linear	Very weak

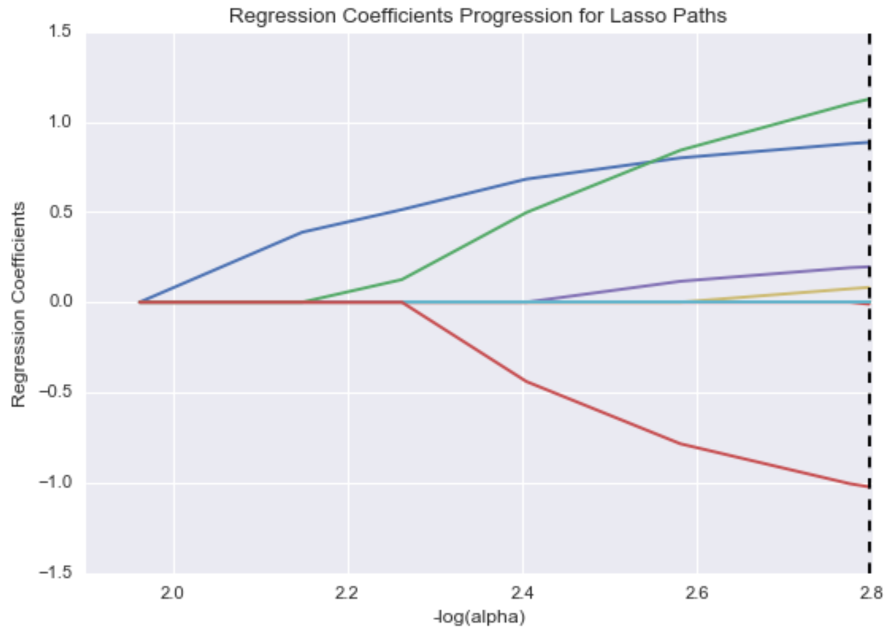
LASSO Regression Analysis

The regression coefficients on Table 3 below show that on applying penalty to the predictor variables; three of them become zero (Energy depletion, natural resource depletion and import of goods and services), and one variable was almost zero (Export of goods and services). Therefore, using Lasso regression, five of the variables have some effect on carbon dioxide damage (Table 3)

Table 3. Lasso Regression Results	
Predictor	Regression Coefficient
Energy Depletion	0.0000
Natural Resources Depletion	0.0000
Export of Goods and Services	-4.274e-5
Food Production	0.0009
GDPGRP	0.01622
Import of Goods and Services	0.000
Industry Value Added	0.0087
Working population	0.0173
Urbanization	-0.0046

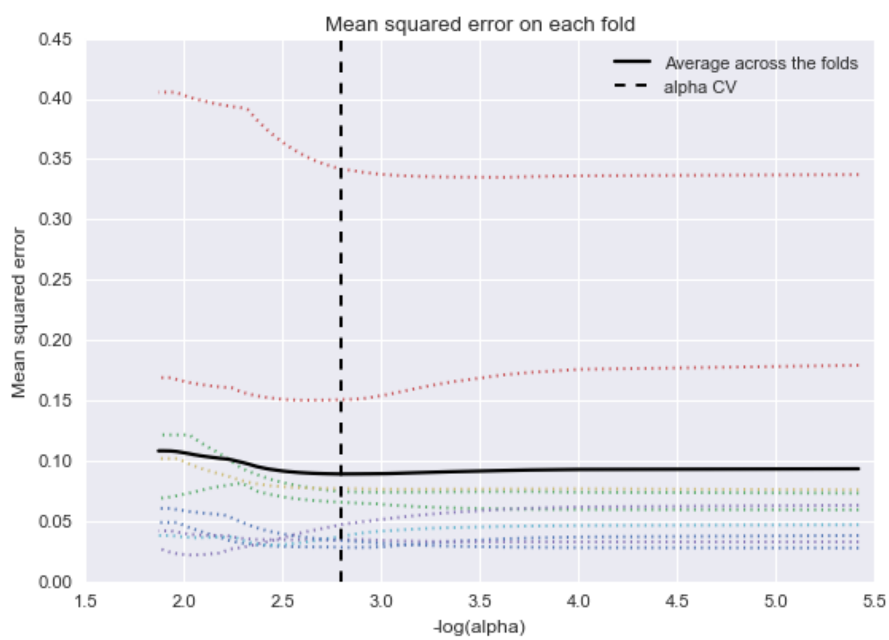
Figure 2 below shows the relative importance of the predictors and the steps in which each variable enters the model. Based on the regression coefficients stated above, the working population rate and GDP growth were most associated with the carbon dioxide damage, followed by industry value added. Urbanization and export of goods and services were negatively associated with carbon dioxide damage (Table 3; Figure 2).

Figure 2: Regression Coefficient Progression for LASSO Paths



As shown in Figure 3 below, the change in the cross-validation average (mean) squared error at each step was used to identify the best subset of predictor variables. Initially it decreases rapidly and gets to a plateau or levels off at which point adding more predictors doesn't lead to much reduction in the mean square error. The behavior was expected as model complexity increases

Figure 3: k-Fold Mean Square Errors



The mean squared error (MSE) for the test data (MSE=0.16) was doubled from the MSE for the training data (MSE=0.08), which suggests that predictive accuracy did decline when the LASSO regression algorithm developed on the training data set was applied to predict lead carbon dioxide damage in the test data set.

Also, training data R-square value signifies that 27% of the variability can be explained for carbon dioxide damage using the predictor variables. Similarly, for test data, the error was 26 % implying that 26% of the variability of carbon dioxide damage in test data can be explained using predictor variables.

Conclusions/Limitations

This project used LASSO regression analysis to identify a subset of world bank economic related indicator variables that best predicted carbon dioxide emissions measured in terms of carbon dioxide damage rate in N=248 world bank countries and regions in 2013. Carbon dioxide damage ranged from 0.05 to 2.49 % of GNI, indicating that there was considerable variability in the amount of carbon dioxide emitted throughout the year.

The LASSO regression analysis indicated that 5 of the 9 developmental related predictor variable were selected in the final model. The strongest predictors of carbon dioxide damage were working population, industry value added and annual GDP. Carbon dioxide damage increased when the working population, industrial activities and GDP growth increased.

The MSE doubled when the training set LASSO regression algorithm was used to predict carbon dioxide damage in the test data set. This suggests that the predictive accuracy of the algorithm may be unstable in future samples, biased and variance in different samples.

The results of this project indicate that controlling population growth rate, transitioning into low carbon economy, energy efficiency and industrial activates based on low carbon input sources are priorities for achieving consistently low carbon dioxide damage (emissions).

Although these three predictors were positively associated with carbon dioxide damage (climate change), the associations were weak, as the correlation coefficients were all below 0.02 (Table 3). Moreover, neither of these predictors was significantly associated with carbon dioxide damage in the bivariate analyses (Pearson r <0.4; Table 2). However, there are some limitations that should be taken into account when considering changes in the carbon dioxide damage indicator based on the results of this project. First, I analyzed only data from a single year, but changes in the carbon dioxide emissions are ongoing. So, it is important to test this algorithm in carbon dioxide emissions in multiple years to determine whether the current algorithm remains relatively biased and unstable despite these ongoing changes. Second, the analysis conducted was limited to the predictor variables available in the dataset. This may lead to misspecification resulting from the exclusion of important variables from the regression model. There is a large number of economic related factors in the carbon dioxide emission process that could impact climate change, but the current project examined only a few of these factors. It is possible that the factors identified as important predictors of carbon dioxide damage among the set of predictors analyzed in this project are confounded by other factors not considered in this analysis. As a result, these same factors may not emerge as important factors when other factors are taken into consideration. Finally, it appears the dataset was not specifically collected to test the research question. Climate change trends are best observed and evaluated over a time series dataset. Therefore, future efforts to develop a solid predictive algorithm for carbon dioxide emissions should expand the algorithm by adding more developmental indicator predictors to the statistical model, and evaluating the applicability of the algorithm over time series dataset.

References

Center for Sustainable Systems, University of Michigan. 2016. "Greenhouse Gases Factsheet." Pub. No. CSS05-21.

IPCC, 2014: Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland, 151 pp

United Nations Sustainable Development Goals (UN SDGs), n.d: http://www.un.org/sustainabledevelopment/climate-change-2/ . The website was accessed on March 6, 2017.

World bank, n.d: http://data.worldbank.org/indicator/NY.GNP.PCAP.PP.CD. The website was accessed on March 6, 2017.