Inference of carbon dioxide emissions by economic growth and energy mix through hierarchical clustering and statistical approach

Tung-Hung Chueh
Green Energy and Environment Research Laboratories
Industrial Technology Research Institute
Hsinchu, R.O.C
tunghung@itri.org.tw

ABSTRACT

The global greenhouse gas emission was increasing rapidly recently because of the rapid development of technology and economic. In 2011 global emissions of carbon dioxide has reached to 31.3 billion tons, with the growth more than doubled compared to 1971. The growth of global carbon dioxide emission is the results of 326% increase in the global gross domestic product and 237% increase on energy consumption. This paper provides a theoretical analysis of CO2 emissions and economic growth, energy growth and energy structure based on a real dataset over the period from 1990 to 2011. First, we use the hierarchical clustering approach to cluster 36 countries by their behavior of carbon dioxide emissions, gross domestic product (GDP) and primary energy supply. Then, we explore the relationship between carbon dioxide emission, GDP, primary energy supply and low carbon energy contribution for each group. The result show that the growth rate of per capita carbon dioxide emission depends positively on the growth of primary energy supply. However, the emissions of carbon dioxide may not depend on the growth of per capita GDP and energy structures significantly in some groups.

Keywords: greenhouse gas emissions, data mining, energy demand, hierarchical clustering

I. INTRODUCTION

With the rapid development of technology and economics, the global greenhouse gas emissions have been increasing rapidly. In 1970 to 2000, the global greenhouse gas emissions average annual growth rate was around 1.3%. However, in 2000 to 2010, global greenhouse gas emissions have increased by 9 billion tons with an average annual growth rate of 2.2%. In 1970 the global greenhouse gas emissions was 27 billion tons, of which the contribution of greenhouse gas emission by carbon dioxide was around 55%. By 2010, the global greenhouse gas emissions were increased to 49 billion tons, of which the proportion of carbon dioxide emission has increased to 65%. Therefore, the rapid growth of carbon dioxide has been a key factor for the increase of greenhouse gas emission.

According to the database of Organization for Economic Co-operation and Development (OECD), the global carbon dioxide emission was around 14.1 billion tons in 1971. Among which the total carbon dioxide emissions in OECD countries were amounts to 67%. In non-OECD countries, China and India were amounted to 5.8% and 1.4%, respectively. However, in 2011 global emissions of carbon dioxide has reached to 31.3 billion tons, with the growth more than doubled compared to 1971. In 2011, the OECD countries accounted for the overall carbon dioxide emission have dropped to 39% and the contribution of emission by China and India have grown up to 25.5% and 5.6%, respectively. The growth of global carbon dioxide emission is the results of 326% increase in the global gross domestic product and 237% increase on energy consumption.

Several authors have studied the problem and estimation of carbon dioxide emission in literatures. Ang used the dynamic causal relationship to analyze the relationship between energy consumption and carbon dioxide emissions in France [2]. Narayan and Smyth are using the technique of panel unit root test and structural analysis to estimate the relationship between energy consumption and GDP in G7 countries [3]. Hossain is exploring the relationship between carbon dioxide emissions, energy consumption, economic growth, trade openness and urbanization of newly industrialized countries [4]. Arouri et al. were studied the energy consumption, economic growth and carbon dioxide emission for the Middle East and North African countries [5]. Stolyarova used the Hierarchical Clustering on Principal Components with Dynamic Panel Data and Within model to explain the growth rate of per capita CO2 emissions for 93 countries [6]. Salahuddin and Gow analyzed the relationship between economic growth, energy consumption and carbon dioxide emissions for the Gulf Cooperation Council countries [7].

In this study, we will focus on the research and analysis of the global carbon dioxide emissions for 36 countries. In general, we believe that there exists closely relation between CO2 emissions and economic growth of a country. In most case, the carbon dioxide emissions will increase with the rapid growth of economics for the developing countries. However, if the growth of economic to a certain size, the carbon dioxide emissions may reverse to decreasing directions because of the environmental awareness. Hence, it may lead to a serious bias if we estimate the carbon dioxide emissions with the same model for the countries with different economic sizes, energy structures, and populations.

In this paper, we will explore the model of carbon dioxide emissions estimation with a two-step approach. First, we will use the technique of data mining: hierarchical clustering to classify 36 countries from their similarity of historical trend of energy consumption, economic growth and carbon dioxide emissions. Finally, we explore the relationship between carbon dioxide emission, energy demand and economic for each classification group.

The structure of this paper was described as follows. In section 2, we will describe the raw panel data we used and illustrate the analysis of descriptive statistics from our selected countries. Then, the characteristic of features and specific approach we used for grouping countries will be discussed in section 3. In section 4, we will describe the classification result, and explore the model of carbon dioxide emissions by the statistical approach. Finally, we will conclude our analysis in section 5.

II. DATA COLLECATION AND DESCRIPTIVE STATISTICS

In this paper, we perform the data collection from 36 countries worldwide. These countries were selected based on the land area, population, economics scale in the world. We collect these countries with their carbon dioxide emissions, gross domestic product (GDP), primary energy supply and low carbon energy percentage. The period of the panel data was selected from 1990 to 2011 and all of the data was collected from the database of Organization of Economic Cooperation and Development (OECD).

TABLE I. 36 COUNTRIES WE USED IN THIS STUDY

Argentina	Demark	Japan	Russian Federation
Australia	Finland	South Korea	Spain
Austria	France	Malaysia	Sweden
Belgium	Germany	Mexico	Taiwan
Brazil	Greece	Netherlands	Thailand
Canada	India	New Zealand	Turkey
Chile	Indonesia	Norway	United Kingdom
China	Israel	Poland	United States
Czech Republic	Italy	Portugal	Switzerland

Thirty-six countries gathered in this study accounted for 81% of carbon dioxide emissions of the world's total emissions. The descriptive statistics of per capita carbon dioxide emissions, per capita GDP and per capita primary energy consumption were illustrated in Table 2. In 2011, the top 5 countries of highest per capita carbon dioxide emissions were Australia(17.43), United States (16.94), Canada (15.37), South Korea (11.81) and Russia (11.65), and the lowest emissions country is India (1.41).

The five countries with the highest per capita GDP were Norway (64.57), Switzerland (55.42), Denmark (46.70), Ireland (45.83), and Sweden (44.08), while the county with the lowest per capita GDP was India (1.06). The level of per capita primary energy supply varies from 0.60 to 7.3 ton oil equivalent. The five highest per capita primary energy supply were Canada (7.3), United States (7.02), Finland (6.45), Norway (5.68), and Australia (5.4), and India was also the lowest country (0.60).

TABLE II. DESCRIPTIVE STATISTICS

	Carbon Dioxide Emission	GDP	Primary Energy Supply
Mean	7.62	25.67	3.52
Medium	7.23	26.35	3.23
Maximum	17.43	64.57	7.30
Minimum	1.41	1.06	0.60
Standard deviation	3.85	17.04	1.70

III. HIERARCHICAL CLUSTERING

It may lead to a significant bias if we assume that the estimation of carbon dioxide emission from economic growth and primary energy supply have the same model because of the large variety of economics size and energy consumption for different countries. Hence, we have to perform the clustering for the 36 countries before model estimating in this study.

In this paper, we will use the hierarchical clustering method to perform the countries grouping. A hierarchical method creates hierarchical nested partitions of the data objects such that instances in one group have more similar expression than instance in different groups. Based on the hierarchical decomposition, the hierarchical method can be further classified as agglomerative clustering method and divisive clustering method. The processes of hierarchical clustering are usually presented in a tree structure called a dendrogram.

In an agglomerative clustering method, the decomposition of the dendrogram can be bottom-up, starts with N groups, each group containing only one instance. Then, we measure the distance between any two groups and agglomeratively progressing upwards by merging similar groups to form larger groups until all subjects in a single cluster and divide the cluster into smaller groups until each group containing only one instance.

Before taking hierarchical clustering, we need to define the similarity, or distance between two subjects or two groups. In this study, we used the distance of correlation which is defined as 1 minus the correlation of two points:

$$d(x,y)=1-corr(x,y) \tag{1}$$

where corr(x,y) was the correlation coefficient for two instances. After defining the measured distance, we used the average-linkage clustering to determine the similarity for any two groups. In average-linkage clustering, the distance is defined as the average distance of all possible distance between two groups:

$$d_{average}(G_i, G_j) = \frac{1}{|G_i||G_j|} \sum_{x \in G_i} \sum_{y \in G_j} d(x, y) \tag{2}$$

When we performed countries grouping, the index of per capita carbon dioxide emission, GDP, and primary energy supplies were used to assess the similarity between two countries or groups. To perform the hierarchical clustering, we use the statistical language R with package hclust. The clustering results with the tree structure were illustrated in Figure 1. Based on the tree structure, we classify the 36 countries as 4 groups. Most countries in group1 belong to developed country which has slow increasing in carbon dioxide emission, economic growth and primary energy supply. For group 2, the countries have rapid growth of economic with a high growth of per capita carbon dioxide emission and primary energy supply. For group 3, most countries have a decreasing trend in per capita carbon dioxide emission and primary energy supply while their annual economic growth rate are very slow. The countries of group 4 come from East Europe, their carbon dioxide emission and energy supply have rapid decreasing from 1990-1998 because of the effect of policy and economics.

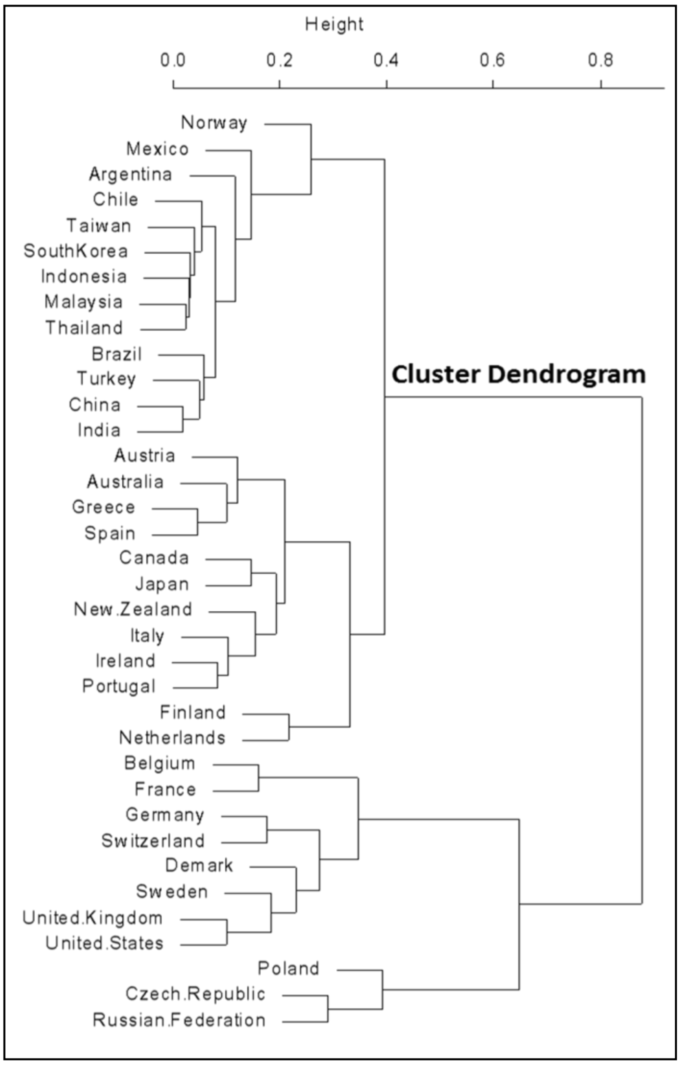

Figure 1. Clustering dendogram

IV. MODEL ESTIMATION AND RESULTS

After using hierarchical method to classify the 36 countries as 4 groups, we explore the relationship between carbon dioxide emission, GDP, primary energy supply and Emergy-Mix, where. Because the economic size of different counties has large variety, we use the growth rate as our inference index. We assume the growth rate of per capita carbon dioxide emissions was affected by per capita GDP and primary energy supply and use the model as follows:

$$gCO2=\alpha+\beta*gGDP+\gamma*gEnergy+\delta*Energy\text{-}Mix+\varepsilon \qquad (3)$$

where gCO2 is the growth rate of per capita carbon dioxide emission, gGDP is the growth rate of per capita of GDP, gEnergy is the growth rate of per capita of primary energy supply and Energy-Mix was the percent of low carbon energy in a country. In this study, we treat the nuclear, hydropower, renewable energy was the low carbon energy, whereas the coal, nature gas and oil was the traditional energy with high carbon energy. The results of model estimation for different groups were illustrated in table 3.

TABLE III. REUSLT OF CO_2 ESTIMATION MODEL FOR 4 GROUPS

	Group 1	Group 2	Group 3	Group 4
(Intercept)	-0.003	0.005	-0.003	-0.006
gGDP	-0.07	0.349***	-0.089	0.013
gEnergy	1.158***	0.630***	1.217***	0.928***
Energy-Mix	-0.011	-0.013	-0.008	-0.013
R^2	0.645	0.626	0.676	0.849

Note: *** and ** respect 0<p-value<0.001 and 0.001<p-value<0.01, respectively.

According to the model analysis as showed in Table 3, we found that the relationship between GDP, per capita carbon dioxide emission and primary energy supply have different expression with different groups. The majority countries of group 2 belong to developing counties, carbon dioxide emission not only driven by the primary energy supply, but also the have a significant affected by rapid economic growth. However, most countries in group 1 and 3 are among the developed countries, the carbon dioxide emission only affect by the primary energy supply significantly because of the economic growth have slowed and most citizen have aware and concern the environmental issue.

V. CONCLUSION

In this paper, we investigate the relationship between per capita carbon dioxide emission, per capita gross domestic product and per capita primary energy supply. Our data include the major 36 countries worldwide and their carbon dioxide emission, GDP and primary energy supply from 1990 to 2011. In order to avoid the difference from socio-economic conditions between countries, we use the technique of data mining: hierarchical clustering to classify 36 countries. We use the countries' historical trend to decide the similarity of two countries or two groups. Finally, we classified the 36 countries to 4 groups and most countries in group 2 are belonging to developing country.

To further investigate the relationship between per capita carbon dioxide emissions, per capita GDP and per capita primary energy supply for each group. We use the statistical approach of linear regression to establish the model, among which the carbon dioxide emission was the dependent variable while GDP and primary energy supply were the independent variables. The results indicated that the relationship between per capita carbon dioxide emission, gross domestic product and primary energy supply have different expression for different countries. The majority countries of group 2 belong to developing counties, carbon dioxide emission not only driven by the primary energy supply, but also the have a significant affected by rapid economic growth. However, most countries in group 1 and 3 are among the developed countries, the carbon dioxide emission only affect by the primary energy supply significantly because of the economic growth have slowed and most citizen have aware and concern the environmental issue.

ACKNOWLEDGEMENT

The author would like to acknowledge support from the Bureau of Energy, Ministry of Economic Affairs, Taiwan, under Grant no. 103-A0101 (D455CG4100).

REFERENCES

[1] IPCC, "Climate Change 2014 Mitigation of Climate Change," Intergovernmental Panel on Climate Change, 2014.

[2] J. B. Ang, "CO2 emission, energy consumption, and output in France," Energy Policy, vol. 35, pp. 4772-4778, 2007.

[3] P. K. Narayan, R. Smyth, "Energy consumption and real GDP in G7 countries: New evidence from panel cointegration with structural breaks," Energy Economics, vol. 30, pp. 2331-2341, 2008.

[4] M. S. Hossain, "Panel estimation for CO2 emissions, energy consumption, economic growth, trade openness and urbanization of newly industrialized countries," Energy Policy, vol. 39, pp. 6991-6999, 2011.

[5] M. E. H. Arouri, A. B. Youssef, H. M'henni, C. Rault, "Energy consumption, economic growth and CO2 emissions in Middle East and North African countries," Energy Policy, vol. 45, pp. 341-349, 2012.

[6] E. Stolyarova, "Carbon dioxide emissions, economic growth and energy mix: empirical evidence from 93 countries," 2012.

[7] M. Salahuddin, J. Gow, "Economic growth, enegy consumption and CO2 emissions in Gulf Copperation Council countries," Energy, vol. 73, pp. 44-58, 2014.