Statistical Analysis of the U.S. CO_2 Emissions Using State-Level Data

by

Peng-Yu Kuo
Dr. Song S. Qian, Advisor
December 2010

Masters project submitted in partial fulfillment of the requirements for the Master of Environmental Management degree in the Nicholas School of the Environment of Duke University

2010

Approved

____
Dr. Song S. Qian

____
Date

Abstract

Past statistical modeling of carbon dioxide (CO_2) emissions has primarily followed the reduced-form specification in the Environmental Kuznets Curve literature. This traditional approach relies on the assumption of a homogenous slope relationship between CO_2 emissions and economic growth across different geographical/political regions. This study uses a panel data comprised of the fifty U.S. states and the District of Columbia and challenges the homogenous slope assumption. By an innovative multilevel modeling approach, it is possible to model heterogeneous slopes for each state. Moreover, the multilevel models can partially explain the variation in slopes by incorporating state economic compositions as state-level explanatory variables.

The modeling results of this paper show that six out of ten of the major economic sectors are significant state-level explanatory variables of CO_2-GDP slope variations with their explanatory capability ranked from high to low as: services, transportation, finance, agriculture, energy-intensive manufacture, and mining. Two of these six sectors, finance and services, have negative effects on the CO_2-GDP slopes, while the other four sectors have positive effects. This study demonstrates the feasibility of studying CO_2 emission trends using multilevel models; however, this study also has some weaknesses in that it fails to account for certain state-level covariates and temporal changes of economic compositions. Future research should aim to overcome these shortcomings.

Introduction

Studies of emission trends of greenhouse gases (GHG) are of central importance to the discussion of climate change for three major reasons. First, GHG emission forecasts are used as an input in climate models to assess the physical impacts of global warming. The quality of this input would affect the quality of modeled outputs. Second, the expected costs of climate change mitigation are calculated based on predicted levels of future emissions. Expected cost is a main factor influencing each country's domestic and international mitigation decisions. If great uncertainty exists in predictions of emissions and hence in expected costs, then the undesirable consequences of under- or over-abatement are likely to occur. Third, simulations of climate change policies demand knowledge of the interaction between GHG emissions and their underlying drivers. For instance, assessing the impacts of a potential cap-and-trade system on a country's economy and its carbon emissions requires understandings of the relationship between economic development and emission patterns.

Among all the anthropogenic greenhouse gases, carbon dioxide (CO_2) is the most important one with the highest radiative forcing[i] (RF) value. In 2005, the RF value of carbon dioxide was 1.66 W m^{-2}—1.7 times greater than the combined values of methane, nitrous oxide, and halocarbons [18]. Fossil fuel consumption and land-use change are the main sources of increasing CO_2 concentration in the atmosphere. In the 1990s, global emissions of fossil CO_2 increased 23.5 gigatonne (Gt) per year on average and CO_2 emissions associated with land-use change grew 5.9 Gt per year [18]. In the United States, the world's second largest carbon emitter, CO_2 from fossil fuel combustion has accounted for about 79 percent of total global warming

[i] The IPCC defines radiative forcing as "a measure of the influence that a factor has in altering the balance of incoming and outgoing energy in the Earth-atmosphere system and is an index of the importance of the factor as a potential climate change mechanism. Positive forcing tends to warm the surface while negative forcing tends to cool it." Its unit is in watts per square metre (W/m^2) [18].

potential (GWP) weighted emissions since 1990 [19]. Since fossil CO_2 is the dominant source affecting the U.S. GHG emission trends, it is chosen as the analysis object of this paper.

Changes in CO_2 emissions from fossil fuel combustion are influenced by various factors, including population and economic growth, energy price fluctuations, weather conditions, technological changes, and the type of fuel consumed [19]. The mechanisms by which these factors interact with each other are rather complex and have not been discussed much in the literature. More research on these mechanisms is needed in order to improve the accuracy of policy simulations and GHG emission forecasts. This study attempts to explore the impacts of economic factors on fossil CO_2 emission trends. The following section will provide a brief review of the existing literature on the subject of CO_2 emissions and then discuss the research question of this paper.

Literature Review and Critiques

A common methodology of modeling CO_2 emissions is the structural approach, which constructs computable general equilibrium models (CGE) based on national input-output matrices. Structural models provide details of variation at the sector level and are able to embody responses to varying conditions such as technological change, capital accumulation, and shifting demands. Thus, this approach is suitable for simulating the impacts of policy instruments on the economy and changes in emissions. For example, CGE models are used by both Jorgenson and Wilcoxen (1991) and Garbaccio et al. (1999) to assess the impacts of carbon taxes on CO_2 reduction goals in the U.S. and China's economies, respectively [12, 6]. A more recent example is the National Energy Modeling System (NEMS) designed by the Energy Information Administration (EIA) of the U.S. Department of Energy (DOE). The NEMS is composed of several submodules and is capable of projecting the energy, economic, environmental, and security impacts on the U.S. of alternative energy policies and different assumptions about energy markets.[ii]

Although the structural approach is suited to policy simulation, it has a few disadvantages. Firstly, in order to explore variation at the sector level, structural models demand an enormous amount of very detailed data inputs. It is not only that collecting the data poses a burden on researchers, but often these data are only available at infrequent intervals. For example, the input-output accounts and the census profiles of the U.S. economy are only released every five years. Furthermore, structural models have to make a large number of parametric and functional assumptions. These assumptions may not always be subject to scrutiny and thus the credibility of the models may be doubtful. Lastly, structural models are not calibrated according to their out-of-sample predictive ability, and thus they may not be the best tools to use for forecasting purposes [1].

Forecasts of CO_2 emissions have often been conducted using *reduced-form* models. The advantages of reduced-form models are that they have lower data requirements and make less parametric assumptions than structural models [17]. A central question in the reduced-form literature is the Environmental Kuznets Curves (EKCs) hypothesis, which claims an inverse U-shaped relation between per capita GDP and pollutant emissions. In the context of CO_2 emissions, it is called the Carbon Kuznets Curve (CKC). The implication of the CKC is that if such a relationship exists and if peak emissions occur at low enough levels, then instead of there being a trade-off, policies inducing economic growth could actually be desirable as they will

[ii] Details of the NEMS model can be found at: http://www.eia.doe.gov/oiaf/aeo/overview/index.html (accessed August 15, 2010).

ultimately lead to emission reductions [9]. Because of this interest in the CKC, most of the reduced-form models in the literature used the general form:

$$log(c_{it}) = \beta_1\ state_i + \beta_2\ time_t + \beta_3\ log(income_{it}) + \beta_4\ log(income_{it})^2 + \varepsilon_{it}$$

Where c_{it} are per capita carbon dioxide emissions for state/country[iii] i in time period t, $state_i$ and $time_t$ are state and time fixed effects, $income_{it}$ are per capita income for state i in time period t, and ε_{it} is the error term. Notice that per capita GDP is the only explanatory variable included in the model. As mentioned earlier, there are many other variables affecting CO_2 emissions, and if these omitted variables are correlated with per capita GDP, then the coefficient estimates from the model above are likely to be biased. The state and time fixed effect terms are thus included in the model to control for potential omitted variable bias. The state fixed effect reflects differences across states that are time invariant such as geography, resource endowments, and industry mix. The time fixed effect controls for time-varying factors shared by all states like changes in technologies, environmental policies, and macroeconomic conditions [9, 17].

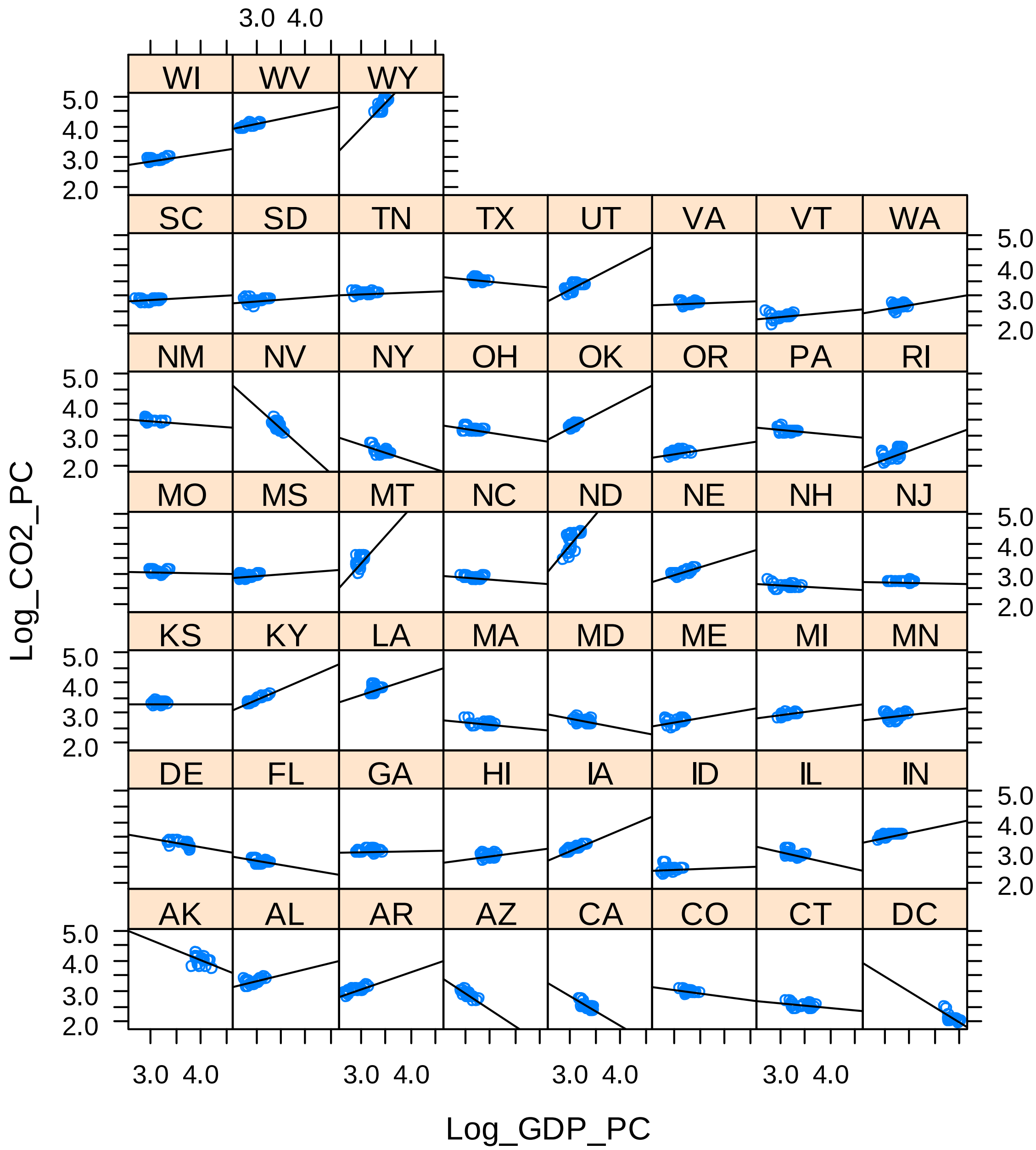
The model specification above, however, rests on a problematic assumption—homogeneity of the slope. The fixed effect terms allow for state-specific heterogeneous intercepts, but not for heterogeneous slopes. Given that states' natural and social conditions span across a wide spectrum, it would be unreasonable to assume the same CO_2-GDP slope for each state. Dijkgraaf and Vollebergh (2001) tests the assumption of slope homogeneity with a panel data of OECD countries on CO_2 emissions for the period 1960-1997 and finds this hypothesis strongly rejected at any conventional level of significance [5]. Müller-Fürstenberger and Wagner (2007) discusses in details some other controversial applications of econometric techniques in the EKCs literature [14].

[iii] The data could comprise of a panel of countries, states, or geographical regions depending on the subject of research. Because this paper focuses its analysis on the U.S. states, I would use the word "state" in my discussion with the understanding that the word may be replaced by "country" or other geographical unit as needed.

Steinhauser and Auffhammer (2005) employ a panel dataset covering the fifty U.S. States and Washington D.C. during the years 1960-2000 and find that the out-of-sample predictive ability of models following the general form above is outperformed by a large universe of reduced-form models. They use mean square forecast error (MSFE) as a selection criterion to choose the best-performing model and make U.S. CO_2 emission forecasts to year 2010 [17]. Despite their systematic approach in selecting the best forecasting models, the final model they adopt still rely on the questionable assumption of homogenous slope across states. One would naturally doubt the homogeneity assumption by looking at graphs of per capita CO_2 plotted versus per capita GDP for each state. *Graph 1* plots the observed data points and shows the least-square estimated linear CO_2-GDP relationship for each state. The varied pattern in the graph seems to support heterogeneous slopes rather than a homogenous slope across states.

Because past studies have built their reduced-form models on a disputable assumption, it casts doubts on the reliability of their forecasting results. Furthermore, the method of using fixed effects to account for the influence from various drivers of CO_2 emissions does not enhance our understanding of the underlying mechanisms by which these drivers affect emissions. This paper, therefore, attempts to apply a different statistical technique—*multilevel modeling*—to investigate the CO_2-GDP slope variations among the U.S. states and how differences in states' economic compositions would affect the slope variations. The section below will discuss the rationale behind the multilevel approach, the model specification and the data used in this analysis.

Graph 1: Per capita CO_2 emissions plotted against per capita GDP in logarithm scale for each U.S. state. The linear lines are the slopes estimated from simple linear least-square regressions.



Model Description and Data

Multilevel models have been used for decades in agriculture and education research [7]. More recently, multilevel models have become popular in various disciplines, including demography, sociology, epidemiology, criminology and health economics [16]. Multilevel applications in social sciences have been reviewed in Diez-Roux (2000), Leyland and Goldstein (2001), Hox (2002), Raudenbush and Bryk (2002) and therefore a review is not attempted here [7, 4, 13, 10, 15].

Multilevel models can be used when the data has a hierarchical structure with observation units grouped at different levels. For instance, students are clustered within different schools and patients are allocated to different clinics. The students and patients would be the level 1 units and the schools and clinics would be the level 2 units. Characteristics of level 2 groups may affect individual outcomes at level 1: different teaching styles of schools or treatment practices of clinics can induce varying student performance or patient health. Multilevel modeling allows researchers to examine the effects of level 2 variables on level 1 outcomes while traditional regression analyses often ignore the group-level effects [8].

The hierarchical structure of the data can take more than two levels and also various forms. For example, the data can be longitudinal with repeat observations nested within individuals over time, can contain multivariate responses with multiple outcomes nested within individuals, and can be results of repeat cross-sectional surveys with multiple observations nested within time periods [4]. This study uses a longitudinal dataset with yearly CO_2 and GDP data as the level 1 observations and the U.S. states as the level 2 groups. Detailed discussions of theories and applications of multilevel modeling on longitudinal data can be found in Goldstein (1999), Hox, and Bates [8, 11, 2].

There are two modeling objectives in this paper. First is to model for heterogeneous CO_2-GDP slopes for each U.S. state including Washington D.C. Second, the models will include state level variables as an attempt to explain what causes the slopes to be heterogeneous. The economic composition of each state is chosen here as the explanatory variable for slopes. It seems reasonable that if a state's economy is made up by high percentages of energy intensive sectors, its CO_2-GDP slope would be steeper than a state whose economy is primarily composed of less energy intensive sectors. That is, the more energy intensive a state's overall economy is, the higher the CO_2 emission would be per unit of GDP generated.

In theory, the two objectives above could be achieved by a two-step analysis. The first step is to run classical regressions for each state separately and obtain 51 different CO_2-GDP slope estimates. Next, another regression would be run using the 51 slope estimates as the "data" of the independent variable, which would be explained by the state level variable of economic composition. This two-step approach, however, is not always viable if there are only a small number of observations in each state. Multilevel modeling, on the other hand, allows us to achieve both objectives simultaneously in one step, and it is more resistant to small sample sizes. Furthermore, multilevel models would yield *partial-pooling* coefficient estimates that are balanced between the *complete pooling* estimates (as in the traditional reduced-from approach with a homogenous slope) and the *no-pooling* estimates (as the 51 estimates from the first step of the two-step analysis). In simple terms, *partial-pooling* estimates offer more flexibility and information than *complete pooling* estimates and perform no worse than *no-pooling* estimates; therefore, some researchers have suggested to always using multilevel models whenever possible [7].

The models constructed in this paper follow the form of:

$$log(C_{it}) = \alpha_i + year_i + \beta_i \;\; log(income_{it}) + \varepsilon_{it}$$

$$\alpha_i = \gamma_0^{\alpha} + \gamma_{1x}^{\alpha} \; composition_{xi} + \eta_i^{\alpha}$$

$$\beta_i = \gamma_0^{\beta} + \gamma_{1x}^{\beta} \; composition_{xi} + \eta_i^{\beta}$$

Where $log(C_{it})$ and $log(income_{it})$ are respectively per capita carbon dioxide emissions and gross domestic products in natural-logarithm scale for state *i* in year *t*. α_i and β_i are the estimated intercepts and slopes for state *i*. $year_i$ is a year effect for state *i*, reflecting the overall time trend of emissions in that state. $composition_{xi}$ is the average percentage (over time) of a certain economic sector *x* (out of a state's total GDP) in state *i* or a vector of several different economic sectors $\boldsymbol{x}$. ε_{it}, η_i^{α}, η_i^{β} are the error terms. The hierarchical structure is two-level with level 1 observations of $log(C_{it})$ and $log(income_{it})$ nested within states over time and level 2 variables $composition_{xi}$ explaining the level 1 parameters of α_i and β_i.

A panel dataset is compiled with 1071 observations covering the 50 U.S. states and Washington D.C. during the years 1977-1997. The CO_2 emission and population data is from Blasing et al. (2004). They calculated CO_2 emissions using state consumption data of coal, petroleum, and natural gas obtained from the EIA State Energy Data Report [3]. Per capita CO_2 emissions are in unit of metric tons.

The aggregate GDP data is taken from the Bureau of Economic Analysis and per capita GDP is calculated using population data from Blasing et al. (2004). Per capita GDP is in unit of thousands of year 2000 real dollars. Per capita GDP data points are then log-transformed and mean-centered around 3.3 (in log-scale), which corresponds to the national average (across states and across time) per capita GDP of 26 (thousands of dollars). This is done for more meaningful interpretation of the model intercepts. The economic composition variables ($composition_{xi}$) are in units of percent (%) and the average percentage of sector x in state i is calculated as:

$$\frac{\sum_{t=1977}^{t=1997} [(\mathrm{GDPxit}/\mathrm{GDPit}) * 100]}{21}$$

Where GDP_{xit} is the amount of gross domestic product contributed by sector x of state i in year t and GDP_{it} is the total GDP produced by state i in year t. Because average values are used here, it is assumed that states' economic compositions remain rather steady over the two decades. Current techniques of multilevel modeling allow the level 2 explanatory variables to have different values from state to state, but they must be time-invariant within each state, and this is why the $composition_{xi}$ variables are not allowed to fluctuate from year to year but take the average values instead.

State GDP accounts before 1998 follow the Standard Industrial Classification (SIC) system and can be divided into ten major sectors: agriculture, mining, construction, manufacture, transportation, wholesale trade, retail trade, finance, services, and government.[iv] With these sectors in mind, a central issue of this study's modeling approach is to determine which sectors should be included as the state-level explanatory variables. The following section will discuss the modeling process and the results.

Modeling Process and Results

As mentioned before, this paper aims to explore the question that whether states' economic composition is a reasonable explanatory variable of heterogeneous CO_2-GDP slopes across states. The magnitudes and significance of the estimates of γ_{1x}^{β}, therefore, are of central importance. The magnitudes of γ_{1x}^{β}, in some sense, suggest the carbon intensity of each

[iv] Definitions of these sectors can be found at: http://www.bea.gov/regional/definitions/

economic sector. A greater positive value of γ_{1x}^{β} means a higher percentage of sector *x* will lead to steeper CO_2-GDP slopes, and hence higher carbon intensity per dollar GDP.

It would be ideal if one could incorporate all sectors into a single model and examine the resulting estimates of the vector $\boldsymbol{\gamma_{1x}^{\beta}}$ to compare the carbon intensity of each sector. However, the fact that the percentages of the ten economic sectors must sum up to 100% indicates that these sectors are highly correlated with each other. As it is not clear how the correlation would affect the modeling results, it is prudent to start with the simplest model form with only one economic sector as the state-level explanatory variable.

Separate models were run for each economic sector, and five out of the ten major sectors were found to be significant state-level explanatory variables: agriculture, mining, transportation, finance, and services. The manufacture sector, which makes up on average 15% of a state's economy, appears to be insignificant. This is a somewhat surprising result as manufacturing is generally believed to be energy intensive. A possible explanation is that the manufacture sector is composed of twenty subsectors, and thus even if states have the same percentage of manufacture, the underlying production activities can look quite different. If a state's manufacture were primarily composed of less energy intensive industries, then a higher percentage of manufacture would not necessarily lead to a steeper GDP-CO_2 slope. To test the validity of this explanation, the manufacture sector is divided into the energy-intensive sector[v] and the non-energy-intensive sector, and two models were fitted with these two subsectors. *Table 1* shows the estimated γ_{1x}^{β}'s and the corresponding standard errors for the overall manufacture and its two subsets.

---

[v] The energy intensive subsectors include: 1. Stone, clay, glass; 2. Primary metals; 3. Food & kindred products; 4. Paper products; 5. Chemicals; 6. Petroleum products; 7. Rubber & plastics.

Table 1. Comparison of γ_{1x}^{β} estimates for the overall manufacture sector and its subsectors of less energy intensive industries and more energy intensive industries

	Non-energy Intensive Man.	Manufacture	Energy Intensive Man.
Estimate of γ_{1x}^{β}	0.0057	0.0097	0.0364
Std. Error of γ_{1x}^{β}	0.0079	0.0079	0.0149

The results in *Table 1* meet the expectation that the coefficient estimate is the most positive for the energy-intensive manufacture sector, the coefficient for the non-energy intensive manufacture is smaller in comparison, and the estimate for the overall manufacture falls in between these two ends. The energy intensive γ_{1x}^{β} is significant at the 5% level while the other two estimates are not significant. Thus, the energy intensive manufacture sector will be considered as a potential state-level explanatory variable in the analysis follows and it will be referred as the intensive manufacture sector for abbreviation.

Referring back to prior modeling results, *Table 2* summarizes the intercept and slope coefficient estimates from the six models each using agriculture, mining, transportation, finance, services or intensive manufacture sector as its state-level explanatory variable. *Graph 2* is a graphical presentation of the modeling results, with estimates $\pm$standard errors for the state slopes β_i plotted versus $composition_{xi}$, along with the estimated multilevel regression line $\beta = \gamma_0^{\beta} + \gamma_{1x}^{\beta}\ composition_{xi}$.

Table 2. Summary of coefficients estimates and the corresponding standard errors from six models each fitted with a different economic sector as the state-level explanatory variable.

	Agriculture	Mining	Transportation	Finance	Services	Int. Man.
Estimate of γ_{1x}^{α}	0.0561	0.0469	0.1814	-0.0705	-0.0534	0.0544
Std. Error of γ_{1x}^{α}	0.0483	0.0073	0.0458	0.0128	0.0132	0.0191
Estimate of γ_{1x}^{β}	0.1051	0.0161	0.1225	-0.0350	-0.0472	0.0364
Std. Error of γ_{1x}^{β}	0.0324	0.0075	0.0359	0.0120	0.0107	0.0149
Explained Change	0.69	0.49	1.18	0.73	1.35	0.56

Table 2 shows that higher percentages of agriculture, mining, transportation, and intensive manufacture are associated with steeper CO_2-GDP slopes and greater percentages of

finance and services are correlated with flatter CO_2-GDP slopes.[vi] This result seems to make intuitive sense as the former four sectors are generally believed to involve more energy intensive industrial activities than the finance and services sectors. It is noteworthy that while agriculture is traditionally not regarded as an energy intensive sector in developing countries, the same cannot be said about the U.S. because American agriculture employs mechanical farming and other more modernized techniques that usually require greater energy inputs than the conventional labor-based practice.

To compare the relative importance of each economic sector as an explanatory variable of CO_2-GDP slopes, a primitive indicator is used to compare the amount of slope variations explained by each sector. If a multilevel model is fitted without any state-level explanatory variable, the estimated CO_2-GDP slopes for each state would fall between -0.835 and 1.512 yielding a range of 2.347. This number of 2.347 would serve as the common base of comparison for the indicator numbers in the last row of *Table 2*.[vii] For example, agriculture explains 0.69 and mining explains 0.49 out of the total variation of 2.347, respectively. Thus, the sectors ranked from the highest to the lowest in their explanatory capability of slope variations are in the order of: services, transportation, finance, agriculture, intensive manufacture, and mining.

The highest ratio between these indicators and the base number 2.347 is 0.58 (from services), and this clearly indicates that much variation is left unexplained using only economic compositions as the explanatory variable. It can be seen from *Graph 2* that while these six models are able to explain in some degree the heterogeneity of slopes, the amount of variation

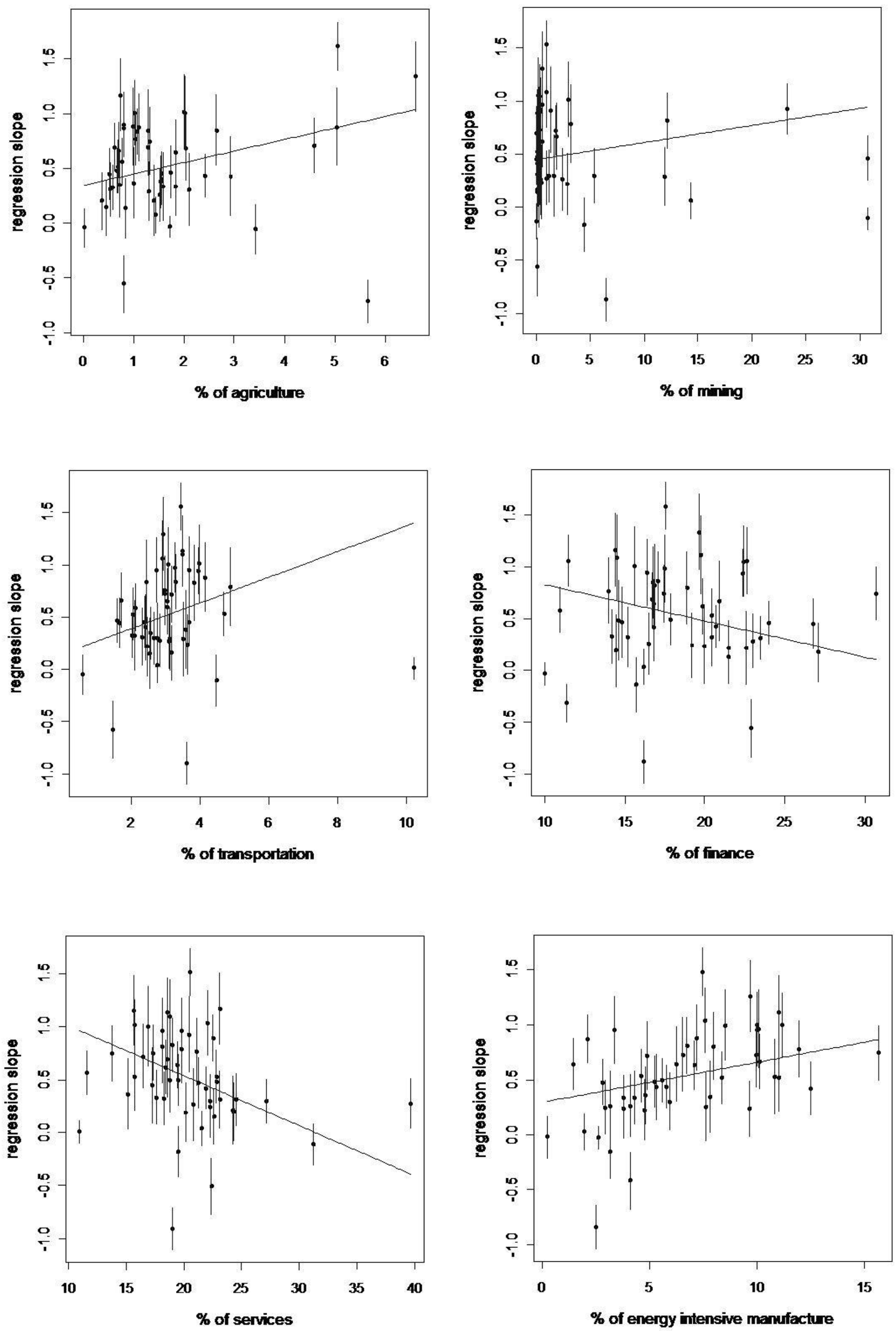
---

[vi] A 1% increase in agriculture, mining, transportation, finance, services, or intensive manufacturing will correspond to changes in slopes (β_i) by 0.1051, 0.0161, 0.1225, -0.0350, -0.0472, and 0.0364, respectively.

[vii] The indicator numbers are the ranges of the predicted slopes from the multilevel regression line: $\beta = \gamma_0^{\beta} + \gamma_{1x}^{\beta}$ $composition_{xi.}$

explained is actually limited as the slope estimates are scattered around and do not closely follow the predicted model regression lines.

Graph 2: Graphical Summary of the results from the six models



After fitting the models using a single economic sector as the state-level explanatory variable, the next step is to see if it is possible to construct models with multiple sectors as a matrix of state-level explanatory variables. When a model was attempted to incorporate altogether the six significant sectors in the previous models, all sectors except agriculture were shown to be insignificant. As mentioned already, this result is likely due to the correlation among sectors; therefore, more close examinations of the correlations is needed to choose the appropriate group of state-level explanatory variables. *Table 3* shows the correlation coefficients between the six sectors.

Table 3. Correlation coefficients of composition$_{xi}$ between six sectors

	Agriculture	Mining	Transportation	Finance	Services	Int. Man.
Agriculture	1.0000	-0.0091	0.31	-0.03186	-0.24	-0.08835
Mining	-0.0091	1.0000	0.60	-0.53791	-0.43	-0.19996
Transportation	0.3109	0.5958	1.00	-0.38565	-0.41	-0.12096
Finance	-0.0319	-0.5379	-0.39	1.00000	0.30	0.00020
Services	-0.2372	-0.4327	-0.41	0.29778	1.00	-0.46391
Int. Man.	-0.0883	-0.2000	-0.12	0.00020	-0.46	1.00000

Table 3 indicates that transportation and services are highly correlated with every other sector, and mining is also highly correlated with the other sectors except agriculture; therefore, problems of multicollinearity could arise when these three sectors are modeled together with the other sectors and lead to insignificant explanatory variables. On the other hand, agriculture, intensive manufacture, and finance are weakly correlated with each other as the absolute values of the correlation coefficients between them are all smaller than 0.1. This information suggests that these three sectors may be combined together as a matrix of state-level explanatory variables. A model is thus fitted accordingly and the results of its coefficient estimates are summarized in *Table 4*. Just to clarify and emphasize again that the results are from a single model fitted with

three state-level explanatory variables, instead of three separate models fitted for each economic sector. This model will be referred to as Model 7.

Table 4. Summary of coefficients estimates and the corresponding standard errors from Model 7 fitted with a matrix of three state-level explanatory variables.

	Agriculture	Int. Man.	Finance
Estimate of γ_{1x}^{α}	0.0619	0.0449	-0.0671
Std. Error of γ_{1x}^{α}	0.0366	0.0153	0.0120
Estimate of γ_{1x}^{β}	0.1146	0.0399	-0.0316
Std. Error of γ_{1x}^{β}	0.0291	0.0125	0.0097

Model 7 seems to be the ultimate desirable model as all three of its state-level explanatory variables for β_i appear to be significant. The magnitude of Model 7 estimates of $\boldsymbol{\gamma_{1x}^{\beta}}$ do not alter much in comparison to those obtained in earlier models (the differences are all under 0.01). As expected, higher percentages of agricultural and energy-intensive manufacturing sectors are associated with steeper CO_2-GDP slopes and greater percentages of financial composition are associated with flatter slopes. A 1% increase in agriculture, intensive manufacturing, or finance will correspond to changes in slopes (β_i) by 0.1146, 0.0399, and -0.0316, respectively. The estimated β_i's ($\beta_i = \gamma_0^{\beta} + \gamma_{1x}^{\beta}\ composition_{xi} + \eta_i^{\beta}$) from Model 7 range from -0.7670 to 1.6878. Because the CO_2 and GDP variables are in natural-logarithm scale, the interpretations of the slope coefficients are in percentage terms. That is, a 1% increase in states' gross domestic product can lead to CO_2 emission reduction by as low as 0.77% or emission growth by as high as 1.69%. Most of the states have positive CO_2-GDP slopes except North Dakota, Rhode Island, Washington D.C., Montana, and Alaska. See the appendix for a complete table of the estimated β_i for each state.

It has been demonstrated so far that the multilevel modeling approach seems to be a reasonable technique in studying CO_2 emission trends and that the economic composition variable is explanatory of heterogeneous CO_2-GDP slopes. A cautionary note has to be raised here, however, as there are certain shortcomings in this study. One may have noticed that the results of Model 7 suggest that the agriculture sector is more carbon intensive than the energy-intensive manufacture sector as the coefficient estimates of γ_{1x}^{β} is more positive for agriculture than for intensive manufacture. This result does not fit with the common understanding that manufacture is more energy intensive than agriculture and makes one question whether the coefficients estimated from the above multilevel approach were biased.

As mentioned already, various covariates other than economic compositions may affect CO_2 emission trends. The varying-intercept specification adopted in the multilevel models of this paper could take into account the effects of time invariant variables like geography and resource endowments. However, variables that change over time, such as technology improvement and environmental regulations, are not accounted for in the models of this study. Another weakness of this study is that it fails to consider temporal changes of states' economic compositions. In the data description section, it stated that the average values of economic compositions are used because multilevel modeling requires the state-level explanatory variable to be time invariant. While this assumption may be applicable for certain sectors in some states, it certainly does not hold universally. For example, the agriculture sector in Delaware increased only slightly from 0.52% in 1977 to 0.67% in 1997 but its finance sector grew significantly from 19% to 41% during the same period. For the reasons discussed above, it is clear that the multilevel modeling approach utilized in this study is far from perfect and further research is needed to overcome the deficiencies.

Conclusion

This paper attempts a multilevel approach which is capable of modeling heterogeneous CO_2-GDP slopes across different states. Moreover, multilevel modeling allows the researcher to test the hypothesis that the different economic compositions of each state can be explanatory of the heterogeneous slopes. The modeling results support the hypothesis and rank the significant explanatory sectors according to their explanatory capability from high to low as: services, transportation, finance, agriculture, intensive manufacture, and mining. However, the amount of slope variations explained by the fitted models is quite limited as other time-varying drivers of CO_2 emissions, such as technology advancement and environmental regulations, are not incorporated into the models. Omissions of relevant variables together with the failure to account for temporal changes of economic compositions are likely to cause the fitted models to generate biased coefficient estimates. Therefore, future research should endeavor to overcome these shortcomings.

Reference

[1] Maximilian Auffhammer and Richard T. Carson. Forecasting the path of China's CO_2 emissions using province-level information. *Journal of Environmental Economics and Management*, 55:229–247, 2008.

[2] Douglas Bates. lme4: Mixed-effects modeling with R. Available at: http://lme4.r-forge.r-project.org/book/ (accessed June 30, 2010).

[3] T.J. Blasing, C.T. Broniak, and G. Marland. Estimates of annual fossil-fuel CO_2 emitted for each state in the U.S.A. and the District of Columbia for each year from 1960 through 2001. Technical report, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, 2004.

[4] Ana V. Diez-Roux. Multilevel analysis in public health research. *Annual Review of Public Health*, 21:171–92, 2000.

[5] Elbert Dijkgraaf and Herman R.J. Vollebergh. A note on testing for environmental kuznets curves. In *OCFEB Research Memorandum 0103*, 2001.

[6] Richard F. Garbaccio, Mun S. Ho, and Dale W. Jorgenson. Controlling carbon emissions in China. *Environment and Development Economics*, 4:493–518, 1999.

[7] Andrew Gelman and Jennifer Hill. *Data Analysis Using Regression and Multilevel/Hierarchical Models*, chapter 12, pages 251–277. Cambridge University Press, 2009.

[8] Harvey Goldstein. Multilevel statistical models, April 1999.

[9] Douglas Holtz-Eakin and Thomas M. Selden. Stoking the fires? CO_2 emissions and economic growth. *Journal of Public Economics*, 57:85–101, 1995.

[10] J. Hox. *Multilevel Analysis: Techniques and Applications*. Mahwah, N.J.: Lawrence Erlbaum Associates, 2002.

[11] Joop J. Hox. Multilevel analysis of grouped and longitudinal data. Available at: joophox.net/publist/mpib2000.pdf (accessed September 10, 2010).

[12] Dale Jorgenson and Peter Wilcoxen. Reducing US carbon dioxide emissions: The cost of different goals. In *CSIA Discussion Paper 91-9*. Kennedy School of Government, Harvard University, October 1991.

[13] A.H. Leyland and H. Goldstein, editors. *Multilevel Modelling of Health Statistics*. Chichester: Wiley, 2001.

[14] Georg Müller-Fürstenberger and Martic Wagner. Exploring the environmental Kuznets hypothesis: Theoretical and econometric problems. *Ecological Economics*, 62:648–660, 2007.

[15] S. W. Raudenbush and A. S. Bryk. *Hierarchical Linear Models*. Thousand Oaks, Calif.: Sage, 2002.

[16] Nigel Rice and Andrew Jones. Multilevel models and health economics. *Health Economics*, 6:561–575, 1997.

[17] Ralf Steinhauser and Maximilian Auffhammer. Forecasting US CO_2 emissions using state-level data. *Breslauer Symposium, University of California*, 2005.

[18] Working Group I to the Fourth Assessment Report of the IPCC. *Climate Change 2007: The Physical Science Basis*, chapter Summary for Policymakers, pages 1–18. Cambridge University Press, 2007.

[19] Climate Change Division US EPA, OAR. *Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990-2008*, chapter Trends. 2010.

Appendix: Slope ($\beta_i = \gamma_0^{\beta} + \gamma_{1x}^{\beta}\ composition_{xi} + \eta_i^{\beta}$) estimates from Model 7

Rank	State	Slope
1	ND	-0.76696
2	RI	-0.59786
3	DC	-0.2419
4	MT	-0.08897
5	AK	-0.0579
6	NM	0.023768
7	NY	0.08587
8	UT	0.098065
9	NJ	0.160862
10	FL	0.164368
11	CO	0.208915
12	AZ	0.239699
13	OK	0.274279
14	MA	0.287336
15	CA	0.331854
16	VT	0.337979
17	MN	0.348334
18	NC	0.361712
19	VA	0.372304
20	WY	0.383609
21	WV	0.39798
22	CT	0.414662
23	NH	0.442604
24	OR	0.448091
25	MD	0.474987
26	ME	0.53961
27	NV	0.560122
28	MI	0.59379
29	SC	0.622281
30	WA	0.7093
31	DE	0.73281
32	TX	0.741557
33	AR	0.768211
34	IN	0.770993
35	IA	0.775064
36	HI	0.779119
37	KY	0.783852
38	MO	0.826407
39	KS	0.852801
40	NE	0.854615
41	GA	0.925019
42	LA	0.967001
43	IL	0.975725
44	OH	1.013824
45	AL	1.036606
46	TN	1.052129
47	WI	1.152777
48	MS	1.18435
49	PA	1.201776
50	SD	1.212247
51	ID	1.687769