PAA Paper 2013

Title: Integrating economic geography and relative cohort size into a panel interaction model of carbon dioxide emissions in the US, 2000-2009

Tyler D. Roberts
Department of Geography, University of Colorado at Boulder

Abstract:

The nexus of age-structure and economic growth has received long-running attention in various social science fields of inquiry. Positing that different age cohorts possess exogenously- and endogenously-driven levels of productivity, this line of inquiry argues that aggregate economic growth is a consequence of changing demographics and aging societies. This research leverages these concepts in order to better understand the human dynamics of carbon dioxide emissions. Using a panel model of US state-level data, these estimates illustrate that an aging of the working-age population is a positive determinant of CO_2 emissions. At the same time, the size of industrial employment relative is a positive correlate of carbon dioxide emissions. Cross-product interaction effects contrast this picture, where states with younger, more industrialized labor forces possess growing volumes of per capita carbon emissions. This research confirms hypotheses that posit greenhouse gas emissions growth to partially be a consequence of changing age-structures.

Keywords: IPAT, STIRPAT, age-structure, economic base, carbon dioxide

Research Questions and Methods

With the above discussion of demography, migration, age-structure, and industrial-economic growth in mind, this research constructs a panel STIRPAT model at the US state level with the following hypotheses:

H1.0: Economic base and industrial employment supplant income as the metric of affluence (A) for industrially-oriented emissions sectors (industrial, electric power generation, and transportation). Household and personal income account for consumption and emissions from a final-demand perspective, failing to account for emissions throughout the production chain. This will be most apparent in export-oriented states where a high proportion of CO_2 emissions are the result of goods produced for non-local demand.

H1.1: Total (all-sector) and transport CO_2 emissions will be understated using only income as the STIRPAT metric of A. The volume of emissions will be under-predicted in export-oriented states by the volume of industrialization as a percent of the total economy, and over-predicted in service-oriented or import-reliant states.

H1.2: Income as the metric of affluence represents the most appropriate specification for household or residential CO_2 emissions.

H2: The relative age-structure of the labor force exerts influence on carbon dioxide emissions. Different age cohorts possess differing levels of capital access, educational attainment, fertility, and other demographic correlates which either leverage or divert resources from the aggregate pool of production, and by extension, the propensity to generate greenhouse gasses (GHG).

H3.0: Both age-structure and industrial-economic base work in concert with one another to produce a complex landscape of age-specific economic activities. These age-economy profiles in the past have been investigated under the umbrella of Easterlin, the demographic dividend, knowledge spillovers and innovation in economic geography, and the (R.) Florida's rise of the 'creative class.' In this research, propose that younger, more industrialized cohorts will be correlated with greater carbon emissions borne out by the greater levels of productivity hypothesized by the above theories of economic growth.

Critical to this analysis is the idea that common STIRPAT variables are germane to sector specific emissions totals; although income is nearly always included in cross-national emissions estimates, the effects of income are likely to be unobserved in an analysis of industrial emissions at more local levels of analysis. Roberts [2011], for example, illustrates that household income exhibits a negative relationship with CO_2 at the county level in the southeastern US. The implication is not that STIRPAT modelers have it backwards, but that industrialized places are not located in high income, high consumption areas, as the theory suggests. To suppose it does contravenes long-held understandings about the inherently spatial nature of production chains. The above panel model explicitly addresses this shortcoming.

This research project is concerned with integrating complementary theories from demography and economic geography into a STIRPAT model of carbon dioxide emissions. The premise of this particular research design is that CO_2 estimates that account for interaction effects between specific demographic components of a population and common industrial correlates offer a potential window into emissions profiles that 'static' STIRPAT estimates do not. The statistical signal from these interaction effects is potentially strong enough that modeling efforts which ignore economically-driven age-structures will potentially produce biased estimates of the original coefficients.

I estimate the effects of personal/household income, the age-structure of the working-age population (ages 15-64), and industrial economic base with a panel model of the following form:

$$Y_{\mathrm{ijt}} = a_{\mathrm{i}} + b_1 A_{\mathrm{it}} + b_2 M_{\mathrm{it}} + b_3 D_{\mathrm{it}} + b_3 (M_{\mathrm{it}} \cdot D_{\mathrm{it}}) + e_{\mathrm{ijt}}$$

where Y is per capita carbon dioxide emissions in state i at in sector j at time t, A is the state median income, M is the percent of the state labor force employed in production, D is relative cohort size [pampel, brunello], and $M_{\mathrm{it}} \cdot D_{\mathrm{it}}$ is an interaction variable designed to test for multiplicative effects hypothesized by scholars to be integral to economic change. States i include all US states, less Alaska and Hawaii; time t are years; and sectors are per capita emissions classifications defined by the EPA. These sectors include a total, or all-sectors, volume for per capita emissions.

In this research, I do not comprehensively model all relevant theories of economic geography and economic demography into the STIRPAT framework of GHG analysis. Instead, I isolate two reasonably representative variables of both theoretical backgrounds that effectively establish the effect hypothesized to impact carbon emissions. Both the size of industry and the age-structure of the population have been analyzed and shown to have positive impacts on emissions at the national level. My aim here is to use prior investigations as a starting point for employing metrics that provide greater theoretical leverage and a more accurate assessment of the anthropogenic drivers of GHG emissions. Using the industrial share of employment and parsing the working-age population into cohort units is one straightforward method for which there are many potential alternatives.

Data and Variables

This research specifies a three variable panel regression with an additional interaction term for forty-eight US states and Washington, DC, for the years 2000-2009. Alaska and Hawaii are excluded from the analysis due to issues that arise in estimating spatial models with non-contiguous units. Although earlier demographic and carbon emissions data are available, employment data that provide the basis for the economic base variable are unavailable in reliably similar units due to the changeover from SIC to NAICS employment classifications.

Dependent variable: per capita CO_2 emissions. The dependent variable in this research is state-level per capital carbon dioxide emissions. In addition to the natural log of total (all sectors) carbon dioxide emissions per capita, I also estimate each regression model using sector-specific data for industrial, electric power, transport, and residential emissions. Hypotheses addressed below illustrate how the expected relationship between age-structure, industrial, and economic base variables differ between different sectors of emissions.

Income. Affluence is a first-order variable in IPAT/STIRPAT modeling, as greater incomes are hypothesized to lead to greater volumes of consumption, which both directly and indirectly lead to greater use of energy and, by extension, GHG emissions. The volume of research that confirms a positive relationship with CO_2 and other trace atmospheric gasses is substantial and too numerous to reference individually in this article.

In these estimates, I represent affluence by median household income (US Census Table H-08). Income as affluence is typically hypothesized to be positively correlated with CO_2 emissions. In the demography-economic base STIRPAT model I estimate here, however, it is entirely possible that in the event that age-structure and industrialization variables are the correct model specification, income will be non-significant or even negatively correlated with per capita emissions. Potentially, this would potentially be particularly likely for regression estimates that utilize industrial, electric power generation, and transportation emissions as the dependent variables. Emissions metrics for these sectors are theoretically accounting for economic activity exogenous to the unit of analysis—i.e. industrial production driven by state-level exports—and income is therefore an inappropriate determinant for the portion of the emissions that are non-local. The expected sign for residential emissions is positive and would confirm hypothesis H1.2.

Percent Production Employment.

Having an industrial economic base is hypothesized to positively correlate with carbon dioxide emissions. Put broadly, places with greater energy-intensive economic activity are expected to have greater volumes of GHG emissions borne out of the processes of both production and high volume of vehicle-miles necessary to transport goods produced. Several STIRPAT studies have confirmed this positive relationship [cite here….Roberts 2011]. Although there are many ways to represent the relative magnitude of heavy industry as a component of the economy, in this analysis I use the percent employed in production (NAICS "Good Producing" 2-digit group 07) for each state *i* in time *t*.

Expected hypotheses for percent production employment theoretically are antitheses of income—the variable is intended to capture emissions from the production side of commodity chains, rather than the consumption side. The argument that underpins this approach is that using only income as a way of representing consumption will fail to account for the volume of GHG emissions that are a product of export-driven economic activity. Middle- or low-income states with high levels of industrialization—common in the US southeast and Midwest—will theoretically have per capita emissions overstated in a model where income is the only determinant of economic activity. The hypothesized sign for this variable is positive for total, industrial, electric power generation, and transport emissions models. Percent production employment is expected to be non-significant for residential emissions, since there is no theoretical basis for industrial activity to be directly driving higher or lower aggregate emissions in the domestic sphere.

Relative Cohort Size

Prior STIRPAT investigations illustrate the importance of changes in population age-structure for future carbon emissions scenarios [cite papers here]. These investigations largely employ the dependency ratio as the metric of age-structure, measuring the ratio of the working age population (ages 15-64) to that of the non-working age (ages 0-15 and 65+). While this metric is excellent for measuring comparing places with aging populations, or places that have undergone rapid changes in age-structure over preceding decades, it gives little information about the how age-structure effects function within the working-age population. The efficacy of compositional changes in age-structure of the working-age population have been debated by economic demographers extensively, and offer a potential window into the population dynamics of economic growth.

In order to address this issue, I use relative cohort size (RCS) to measure potential age-structure effects of the working population on carbon dioxide emissions. Though

Pampel's RCS is commonly used for RCS estimates [cite pampel], I employ Brunello's more 'balanced' ratio [cite brunello], given by the following:

$$\frac{Pop.\ Ages\ 35-50}{Pop.\ Ages\ 20-34}$$

Two competing hypotheses are at work. First, is a scenario working populations weighted towards younger populations are more productive; younger laborers have fewer dependents, lesser status in the place of employment, and are thereby hypothesized to have greater productivity borne out of this disadvantageous position in the workforce and a dis-incentive towards non-work activities. Under this scenario, states with younger populations would have greater emissions, indicated by a negative coefficient for the variable.

The second RCS hypothesis is one of greater capitalization and productivity through an aging, wealthier working population. Under this scenario, places with older workers receive their hypothesized higher economic intensity from educated, highly-capitalized workers driving production rates through the ability to command capital and make large-scale investments. Positive secondary effects on emissions are also borne out through higher consumption rates made available from higher personal incomes. A positive coefficient for RCS confirms this latter hypothesis.

Interaction of RCS and Percent Industrial

The crux of this investigation is that demographic and economic-base metrics work in concert with one another to illustrate how complex population and economic geographies drive differential carbon dioxide emissions from place to place. While these ideas have been explored by other STIRPAT scholars at different scales [see, citations here], this investigation seeks to establish how the product of RCS and industrial employment correlate with CO_2 emissions though a multiplicative interaction term.   A positive coefficient for this term would confirm H3.1, where older, industrialized working populations drive carbon emissions, while a negative coefficient supports H3.2, where younger laborers in heavy industry and goods-producing occupations are associated with increases in emissions.

Estimate Results

Model estimates are presented in Tables 1 through 3. Per panel data analysis procedures, a variety of diagnostic statistical tests were performed on each of the specified models in order to mitigate potential bias problems [Millo and Piras 2012, Croissant and Millo XXXX, Franzese and Hays 2007]. This involved estimating the models through several iterations. First, I eliminated a pooled OLS approach by applying a Lagrange Multiplier test of the Gourieroux, Holly and Monfort method and estimated significant values for all regression models. Second, significant Hausman tests indicated that fixed-effects procedures are more appropriate than a random effects model. This applied for most regressions in the analysis. Third, the strength of serial correlation in fixed-effects models was established via Wooldridge; a significant test for Wooldridge indicates that a first-differenced model is the preferable procedure. Finally, a Wooldridge First Difference test delineated whether there were further serial correlation problems in the first differenced models themselves. Serial correlation of first-differenced estimates are notably problematic and difficult to deal with, indicating a broader mis-specification issue.

Table 4 visually illustrates differences in sign and significance for independent variables between each of the models and procedures. Table 5 shows that there are no substantive differences between significant independent variables in terms of sign from model-to-model. Generally, positive changes in RCS and greater percent production employment are positive determinants of CO_2 in US states during the 2000s, while the product of these two metrics are significant and negative for most model procedures. Income was not significant for any regression procedure except the basic and spatial interaction panel for electric power, a result that makes little substantive sense in light of the non-significance of income in all other procedures and dependent variable specifications.

A primary goal of this research was to test whether more specific metrics of economic productivity and demography were better able to capture the impacts of human activity on carbon dioxide emissions. The models investigated here utilize the percent of the state-level employment base involved production activities as a proxy for the level of industrialization. Theoretically, higher levels of industrial employment should correlate with higher emissions than income alone can account for. Put differently, in high-export states income alone will fail to account for carbon emissions, as emissions created in the production process are generated by non-local incomes. This is the obverse case of the Netherlands Fallacy, where high-import countries appear ‘greener’ than they really are owing to the tendency of exporting environmental problems [cite NF here].

Estimates for sector-by-sector and total (all-sectors) CO_2 emissions support this hypothesis. States with high or growing levels of production employment have greater per capital levels of carbon dioxide emissions. Significant estimates for percent production are positive and significant for total, industrial, and transportation emissions models. Percent production was non-significant for electric power for interaction and spatial panel models, while also non-significant for residential emissions in the non-interaction model. The strength of coefficients for the percent production variables are modest—in the industrial models, where the hypothesized effect would be greatest, a ten percent increase in production employment as a percent of the overall state economy is correlated with ~ 1.3 to 1.5% increase in industrial CO_2 emissions. More modestly, a ten percent increase in production employment is correlated with a ~1% increase in transportation CO_2 emissions, and a ~0.7 - 0.8% increase in total (all-sector) emissions per capita. Importantly, estimates for the spatial panel procedures were substantively similar to the estimates in non-spatial models in sign, significance, and magnitude, indicating an estimate that is robust when accounting for a variety of factors as biases.

Estimates for median personal income contrast with percent production employment, remaining largely non-significant for the preponderance of procedures and model specifications. Income is a significant and negative determinant for electric power emissions, an estimate that makes little theoretical sense. Production employment would be a more likely (positive) correlate for electric power emissions, but the geography of energy production in the US—weighted heavily towards rural, mountain regions suitable for both coal extraction and hydroelectric production—indicates that the more likely explanation for this negative result is that the regressions are merely capturing the positive relationship between population density and income. That emissions from energy production are sharing in this inverse relationship is hardly surprising when the largely rural nature of energy services is considered. I do wish to be perfectly clear, however, that the negative coefficient observed here is more like the result of this sub-national geography of energy production rather than a true inverse relationship with carbon emissions observed between states in the US.

Economic geography and economic demography specify an age-structure dependent theory of economic development. If these theories are indeed true and observable, it is potentially feasible to also observe an increase in carbon dioxide emissions as a result of increased levels of productivity. In the above sections, I highlighted how competing theories specified potential origins of increased economic growth. The regression coefficients estimated for relative cohort size (RCS) in this research are consistent with the second hypothesis of age-structure, namely state labor forces weighted towards growing older working populations have higher levels of carbon dioxide emissions. This effect is considerably strong for both the total per capita carbon emissions and for transportation emissions. In these dependent variable categories, a ten percent increase in the state-level ratio of older workers to younger workers is correlated with an ~ 10.0 – 17.0% increase in per capita carbon dioxide emissions. Put another way, aging labor force populations go hand-in-hand with increases in CO_2 emissions for total and transportation sectors. When spatial effects are accounted for, RCS is also significant, positive, and near unit-elastic for industrial emissions as well. No significant effect is observed for electric power or residential emissions in models with spatial and interaction effects.

A primary concern of this article was whether the failure to recognize the substantive significance of economic base and age-structure interactions would result in bias estimates. Testing for this involved simply including an interaction term that attempted to account for a more dynamic relationship at the nexus of age-structure and economic base, as hypothesized by social science scholars investigating the 'creative class' [Florida], innovation and universities [anselin et al], and other more macro-level investigations of changing demographies and economic performance [cite here]. The estimates presented here fail to confirm this hypothesis; the addition of the interaction term does not substantially change the primary regression coefficients in terms of sign or significance. Changes in magnitude are apparent, per expectations in a regression that includes one or more interaction terms.

Conclusions

Regression estimates presented in this article paint a consistent picture across different procedures and specifications—growth and aging in the American labor-force, as well as increasing industrialization, are positive drivers of carbon dioxide emissions at the state-level for the most recent decade. Working in conjunction with one another, however paints a different story. Cross-product interaction estimates illustrate that industrial growth in states with younger labor forces experienced higher growth in carbon emissions during the decade of the 2000s. This estimate indicates that the relationship between economic-demographic phenomena and state-level carbon dioxide emission is complex.

In this research I have endeavored to analyze the STIRPAT model using frameworks of economic growth from geography and demography. In doing so, I aimed not only to illustrate the ways that metrics such as relative cohort size and industrial employment drive carbon dioxide emissions both directly and indirectly, but also to open up the conversation to metrics beyond the common STIRPAT variables of GDP, median income, and other 'consumption-side' variable used to represent affluence. While these have served cross-national analyses very well in prior works, understanding local- and sub-national level CO_2 emissions requires greater attention to the mechanisms of the production chain and economic geographies of labor. Walker's dictum that there is often a very good deal of within-country variation in terms of labor when compared to between-country variation is particularly relevant here [cite Walker]; just as global GHG emissions trends mirror core-periphery rubrics of prior decades, similar sub-national geographies of production and demographic change are apparent within many nation-states.

Table 1
Panel Regressions, US States, 2000-2009, per capita CO_2 emissions

	1	2	3	4	5
Dep Var:	ln(Total)	ln (Industrial)	ln(Electric Power)	ln(Trans)	ln (Resid't'l)
	First-difference model	First-difference model	Random Effects Model	First-difference model	First-difference model
log(Income)	-0.045	0.071	-0.783 **	0.086	-0.121
(t-value)	-0.987	0.674	-2.958	1.747	-1.757
PctProduction	0.029 ***	0.063 ***	0.034 ***	0.024 ***	0.005
	7.965	8.058	3.921	6.926	1.708
RCS	0.378 **	0.030	0.034 ***	0.570 ***	0.005 ***
	3.135	0.076	3.921	5.272	1.708
(intercept)	0.001	-0.004	8.939 **	0.007 **	-0.030 ***
	0.407	-0.598	3.099	2.828	-8.793
Total Sum Sq.	1.007	5.5759	26.887	0.976	2.550
Resid. Sum Sq.	0.863	4.9623	24.761	0.838	2.446
Adj. R-squared	0.142	0.109	0.078	0.140	0.040
F	24.277	18.011	13.910	24.032	6.173
F-sig	0.000	0.000	0.000	0.000	0.000
idiosyncratic share			0.030		
individual share			0.970		
theta			0.944		
Hausman	28.167	14.385	5.993	11.127	33.056
p-value	0.000	0.002	0.112	0.011	0.000
Wooldridge FE	214.937	118.918	12.272	67.063	210.64
p-value	0.000	0.000	0.000	0.000	0.000
Wooldridge 1st Diff	2.410	0.023	3.487	10.741	17.285
p-value	0.121	0.879	0.062	0.001	0.000
GHM-ML	2096.61	2045.15	2039.53	1958.81	2049.9
p-value	0.000	0.000	0.000	0.000	0.000

White's Covariance Matrix Applied (White's SE's)
n=49, T=10, N=490

Table 2
Panel Regressions, US States, 2000-2009, per capita CO_2 emissions

	1	2	3	4	5
Dep Var:	ln(Total)	ln (Industrial)	ln(Electric Power)	ln(Trans)	ln (Resid't'l)
	First-difference model	First-difference model	Random Effects Model	Random Effects Model	First-difference model
log(Income)	-0.047	0.069	-0.806	-0.038	-0.122
(t-value)	-1.018	0.654	-1.630	-0.571	-1.754
PctProduction	0.088 ***	0.152 **	0.093	0.119 **	0.065 *
	5.157	3.086	0.464	2.964	2.481
RCS	1.207 ***	1.299	1.528	1.706 **	-0.159
	4.363	1.390	0.512	2.624	-0.431
RCS*PctProduction	-0.055 ***	-0.084	-0.052	-0.095 **	-0.056 *
	-3.652	-1.911	-0.275	-2.622	-2.313
(intercept)	0.000	-0.007	8.271	0.179	-0.031 ***
	-0.157	-0.988	1.281	0.243	-8.687
Total Sum Sq.	1.007	5.576	27.290	1.604	2.550
Resid. Sum Sq.	0.848	4.926	25.067	1.138	2.430
Adj. R-squared	0.156	0.115	0.081	0.287	0.046
F	20.472	14.373	10.752	49.617	5.376
F-sig	0.000	0.000	0.000	0.000	0.000
idiosyncratic share			0.035	0.049	
individual share			0.965	0.951	
theta			0.940	0.928	
Hausman	25.840	39.397	8.810	0.5549	44.107
p-value	0.000	0.000	0.066	0.968	0.000
Wooldridge FE	166.668	88.339	11.816	241.077	199.446
p-value	0.000	0.000	0.001	0.000	0.000
Wooldridge 1st Diff	3.155	0.016	3.220	29.705	18.144
p-value	0.076	0.899	0.073	0.000	0.000
GHM-ML	2055.25	1932.24	1927.41	1855.56	1994.71
p-value	0.000	0.000	0.000	0.000	0.000

White's Covariance Matrix Applied (White's SE's)
n=49, T=10, N=490

Table 3
Spatial Panel Regressions, US States, 2000-2009, per capita CO_2 emissions

	1		2		3		4		5	
DV:	ln(Total)		ln (Industrial)		ln(Electric Power)		ln(Trans)		ln (Resid't'l)	
log(Income)	-0.089		-0.233		-0.793	**	-0.055		0.039	
(t-value)	-1.503		-1.605		-3.008		-1.025		0.454	
PctProduction	0.076	***	0.135	***	0.087		0.106	***	0.049	**
	6.700		4.811		1.689		10.461		3.063	
RCS	1.066	***	0.972	*	1.484		1.493	***	0.348	
	6.244		2.309		1.928		9.589		1.381	
RCS*PctProduction	-0.049	***	-0.060	*	-0.047		-0.083	***	-0.011	
	-4.967		-2.449		-1.038		-9.292		-0.795	
(intercept)	2.457	***	1.272		8.189	**	0.599		-1.257	
	3.666		0.777		2.755		0.995		-1.285	
phi	102.851	***	45.325	***	32.422	***	24.809	***	76.519	***
	4.658		4.608		4.624		4.590		4.628	
rho	0.159	**	0.089		0.021		0.264	***	0.392	***
	3.017		1.587		0.339		5.542		9.047	
Hausman	1.305		8.7507		3.844		7.209		4.836	
p-value	0.861		0.068		0.428		0.125		0.305	

n=49, T=10, N=490
Regression procedure: spatial panel random effects

Table 4

Cross-model comparison of estimate results

Dep. Var.:	ln(Total)				ln(Industrial)			ln(Elec Pwr)			ln(Trans)			ln(Residential)		
	Basic	IRE	IS	FI	Basic	IRE	IS	Basic	IRE	IS	Basic	IRE	IS	Basic	IRE	IS
ln(Income)								-		-						
Pct Production	+	+	+		+	+	+	+			+	+	+		+	+
RCS	+	+	+				+	+			+	+	+	+		
RCS*Pct Production		-	-				-					-	-		-	

Lists only variables that are used across multiple regression procedures.

Basic = basic panel, IRE = interaction random effects, IS = interaction spatial panel, FI = full interaction specification

+	Positive and sig. at the 95% level
-	Negative and sig. at the 95% level
	Not significant
	N/A