Spatial Patterns of Carbon Emissions in the U.S.: A Geographically Weighted Regression Approach

Julio Videras

Hamilton College

January 2012

Work in progress

Abstract

This paper uses U.S. county-level data to examine the extent of geographical variability in the process liking emissions of carbon dioxide to measures of population, affluence, and technology. Results from geographically-weighted regression models show that there is strong evidence of geographical heterogeneity and that the magnitude, and in some cases, the direction, of the effects vary within and across the 48 contiguous states in the U.S. These results suggest that one ought to be cautious of policy recommendations based on global models that ignore or account imperfectly for spatial dependence.

1. Introduction

Since the late 1960s, researchers and policymakers have debated how economic and population growth influence environmental quality and the stock of natural resources, and whether technological process may be a mechanism to reduce the ecological impacts of growth. Much of this debate has focused on the IPAT identity (and its stochastic formulation, the STIRPAT model) and the environmental Kuznets curve (EKC) hypothesis as approaches that allow for the identification of the determinants of environmental quality.[1]

The IPAT identity describes environmental impacts (I) as the outcome of three variables: population (P), affluence (A), and technological factors (T). The stochastic formulation of the identity – the Stochastic Impacts by Regression on Population, Affluence, and Technology (STIRPAT) model – provides a testable parsimonious model to estimate the contribution of each factor on impacts (Dietz and Rosa, 1997). The environmental Kuznets curve (EKC) hypothesis was first examined by Grossman and Krueger (1991) who proposed a non-linear relationship between affluence and environmental pollution. In recent practice, these two conceptual models have converged as researchers working within the framework of the IPAT identity typically include a quadratic term for affluence to test the EKC hypothesis.

As Maddison (2006) points out, the continuing interest on the EKC hypothesis is understandable since decision-makers need to evaluate policies that may cause a trade-off between economic growth and environmental quality, and must consider the extent to which rising levels of affluence, institutional responses, and technological progress may curve or even reverse the impact of economic growth on the natural environment.

[1] Carson (2010) discusses the IPAT identity in the context of the controversy generated by Ehrlich's *The Population Bomb*, of 1968, and the Club of Rome's *Limits to Growth* of 1972. Brander (2007) argues that these issues can be traced back to Malthus and the "Malthusian trap" hypothesis.

However, the results in the literature vary according to the measure of environmental impact (such as air and water pollutants, biodiversity risk, ecological footprint, and others), the sample used for the estimation, and the choice of econometric specification. While researchers have examined how several econometric issues might drive this lack of robust results, less, but growing, attention has been paid to the statistical issues due to the spatial nature of the data used to estimate EKC and STIRPAT models.[2]

In empirical applications of the EKC and STIRPAT models, many researchers have estimated linear regression models and made the assumption that observations are spatially independent, that is, it is assumed that environmental outcomes in one region are independent of impacts in other regions. This assumption may not hold if there are unobserved contextual effects that make impacts in nearby locations to be similar: geographical, historical, administrative, and cultural factors that adjacent areas share due to technological diffusion, similar paths of industrialization and de-industrialization, cultural factors that influence lifestyles and consumption patterns, and the implementation of policies that react to decisions by neighboring political units.

There are several approaches to modeling spatial dependence (Maddison, 2006). Suppose we want to estimate emissions of an air pollutant at the state level in the U.S. In a spatial lag model, researchers would model state *i*'s emissions as a function of the state's characteristics and its neighbors' average emissions, where the neighbors' emissions are weighted by a matrix of spatial weights. Alternatively, a spatial regression approach would model state *i*'s emissions as a function of state-level characteristics, such as affluence and technological factors, and a

[2] Rupasingha et al. (2004) and Maddison (2006) are two seminal papers examining spatial issues in the EKC literature.

spatially-weighted average of its neighbors' affluence and technology. These two approaches address the issue of whether spatial dependence may bias estimates and, in both cases, researchers test the hypothesis of whether the characteristics of neighboring observations influence, on average, own outcomes. It is also possible to model spatial dependence in the error term by defining the error as a weighted average of neighboring errors plus a random element. In this case, researchers aim to account for the spatial autocorrelation of observations.

While these models provide insights about the influence of spatial dependence on environmental impacts on average, these approaches still assume that the relationship between emissions and its determinants is constant (stationary) over space. There are reasons to examine this assumption. A puzzling result in the literature is that global results do not necessarily hold for specific countries (List and Gallet, 1999; Stern and Common, 2001). If the process linking environmental performance to affluence, population, and technology is not constant over space and thus depends on where the data are taken from, global models might not be useful policy tools. Policies that ignore spatial heterogeneity are likely to have unexpected consequences as factors that have a large impact in a given region may not have a meaningful effect, or even have an opposite effect, in another area; in that case, accounting for local differences, if politically feasible, might result in larger aggregate welfare. Ignoring the local context may also have normative implications. For example, structural and natural factors might increase the level of greenhouse gas emissions in some locations so that policies requiring uniform abatement across space would generate unfair outcomes (Neumayer, 2002).

This paper addresses the issue of spatial dependence as well as the issue of parameter heterogeneity. This paper uses U.S. county-level data to examine the extent of geographical variability in the process liking carbon emissions to measures of population, affluence, and technology. I apply geographically-weighted regression (GWR) models to data on carbon emissions. GWR allows the relationship between dependent and independent variables to vary across space by estimating a spatially weighted model at every location in the area of study so that, for each location, nearby observations are weighted more heavily than distant observations. Thus, in addition to addressing the issue of spatial dependence, GWR models generate local estimates, that is, the distributions of coefficient estimates over the study area. These coefficient estimates, as well as diagnostic statistics, can be displayed to understand better the nature of spatial heterogeneity. The analysis shows that linear regression residuals are spatially correlated while GWR modeling solves this problem; in addition, there is strong evidence of geographical heterogeneity and patterns across counties. The results suggest that while the processes for carbon emissions are similar across adjoining counties, the magnitude, and in some cases, the direction, of the effects can vary within and across the 48 contiguous states in the U.S. While this research cannot reveal why these spatial patterns exist, it provides a starting point for a more detailed investigation. These results show that one ought to be cautious of policy recommendations based on global models that ignore or account imperfectly for spatial dependence.

This paper examines emissions of carbon dioxide in the context of a STIRPAT model that includes a quadratic term for affluence in order to test the EKC hypothesis. Thus, the next section discusses research on the STIRPAT model and on spatial econometric approaches to the EKC hypothesis. Section 3 presents the empirical models and data while Section 4 outlines GWR methods. Section 5 discusses the results and Section 6 concludes.

2. Literature Review

The literature on the IPAT identity and the STIRPAT model is extensive. Researchers have used the IPAT identity as an accounting equation to formalize the dependencies between environmental impacts, population, affluence, and technological factors. As York, Rosa, and Dietz (2003) point out, the IPAT identity has several strengths: it is based on ecological principles, it is parsimonious, and it highlights the multiplicative contribution of each factor. However, the identity assumes proportional effects (unitary elasticities) for population, affluence, and technology. Dietz and Rosa (1997) have proposed an alternative specification that allows for non-unitary elasticities and hypothesis testing. The Stochastic Impacts by Regression on Population, Affluence, and Technology (STIRPAT) model follows the basic multiplicative formulation of the IPAT identity but assumes the effects of each factor on impacts may not be proportional and that there is a random component v :

$$I = cP^{\beta_1} A^{\beta_2} T^{\beta_3} v \,.$$

After taking logarithms, the model can be expressed as:

$$\log I = \beta_0 + \beta_1 \log P + \beta_2 \log A + \beta_3 \log T + \varepsilon$$

where the beta coefficients represent the elasticity of emissions with respect to population, affluence, and technology; and epsilon is the error term. Researchers have used this framework to test the influence of different measures of the basic determinants. For example, York, Rosa, and Dietz (2003) apply this model to a cross-section of countries to estimate carbon emissions and energy footprint including a non-linear term for GDP, percent of urban population, and a control for latitude.

Shi (2003) uses the STIRPAT model to estimate the elasticity of carbon emissions with respect to population, GDP per capita, and measures of industrial composition. Shi hypothesizes that population changes in low-income countries are likely to have a more detrimental effect on environmental quality than population changes in high-income countries. Shi finds evidence for this hypothesis using a panel of 93 countries and estimating models with country and time fixed effects that include three interaction terms for population and income category (high, medium, and low income countries). While Shi's work highlights the need to account for heterogeneity in the elasticity of emissions with respect to population, this paper extends the analysis to allow for spatial heterogeneity in the elasticity of emissions with respect to each determinant.

Neumayer (2002) also uses the STIRPAT framework but focuses on the influence of natural factors on emissions, after controlling for affluence, in a sample of countries. In particular, Neumayer controls for maximum and minimum temperatures, stock of renewable resources, stock of oil and gas reserves, and land area impacted by human activities. Neumayer estimates linear regression models with a time trend and finds that natural factors do matter, although affluence has the largest effect on emissions. Following Neumayer's argument that natural factors matter, I estimate models that account for cooling and heating requirements due to climate.

In these papers and related literature, efforts to account for the effects of geographical context typically take the form of country fixed-effects with panel data (Shi, 2003) or some indicator of geography such as a control for latitude (York, Rosa, and Dietz, 2003) or controls for natural factors (Neumayer, 2002) . These approaches might be sufficient to address the issue of spatially correlated residuals and spatial nonstationarity when examining a sample of countries. A more direct and sophisticated approach to spatial heterogeneity is likely needed when examining variability within a country since adjacent political units within a country likely share unobservable contextual factors that may generate spatially correlated observations and non-stationary relationships. Roberts (2011) estimates the STIRPAT model using carbon dioxide emissions at the county level as the dependent variable for 755 counties in nine southeastern states. Roberts estimates spatial lag regression models and finds that affluence is either negatively correlated with emissions or statistically insignificant. Although the negative sign on affluence might be the result of misspecification, it is also possible that the relationship between impacts and affluence for these nine states differs from the relationship for a larger sample of states and the relationship researchers have found for cross-sections or panels of counties. I extend Roberts's research by accounting for spatially heterogeneity in a larger sample (all counties in contiguous 48 states and Washington, D.C.) and estimating the spatial distributions of coefficient estimates.

Researchers have also examined spatial econometric issues in the EKC literature. Maddison (2006) examines the EKC hypothesis in a sample of 135 countries and estimates spatial lag and spatial regression models, as well as models with spatial errors, for different measures of air pollution. Maddison finds that neighbors' per capita emissions of sulfur dioxide and nitogren oxides explain countries' own emissions of these pollutants. Tevie, Grimsrud, and Berrens (2011) also estimate spatial lag and spatial error models to text the environmental Kuznets curve hypothesis for biodiversity risk at the state level in the U.S. The authors do not find evidence that the EKC hypothesis holds for biodiversity but do find that a state's biodiversity risk depends on the biodiversity risk of adjacent states. The authors argue that this is expected as biodiversity risk depends on geographical and natural factors that are not contained within predetermined political units. These spatial lag and spatial regression models estimate a global relationship between environmental impacts and its determinants, augmenting the model with spatially-related variables that capture the influence of nearby observations. The findings from these models are insightful and, as Maddison (2006) points out, may help explain why EKC models that do not account for spatial data exhibit little stability. What these approaches do not test is whether the parameters of the model are heterogeneous across space. Estimating GWR models addresses the issue of spatial dependence and allows for the examination of the distribution of coefficients over space.

Finally, the work by Pizer, Sanchirico, and Batz (2010) is similar to this paper in terms of methods and policy implications. Pizer, Sanchirico, and Batz use non-publicly available household data from the U.S. Consumer Expenditure Survey to estimate energy use. Because sampled households are not located in all counties (and urban areas are over-sampled) the authors use non-parametric kernel regression models to estimate county-level energy use as an average of energy use in nearby sampled counties, an approach that is similar to a geographically weighted regression. Pizer, Sanchirico, and Batz (2010) find substantial spatial variability in fuel oil, electricity, and natural gas usage, and argue that it is important to account for spatial variability when designing fair energy policy. In addition, the authors point out that "states are not necessarily the most interesting geographic unit" because of variation within states and the fact that some large urban areas cross state lines. The findings in this paper are consistent with the results and implications in Pizer, Sanchirico, and Batz (2010).

3. Empirical Models and Data

The IPAT identity defines environmental impacts (I) as the multiplicative outcome of population (P), wealth or affluence (A), and technological factors (T) as follows: *I=PAT*. The Stochastic

Impacts by Regression on Population, Affluence, and Technology model (STIRPAT) preserves the multiplicative nature of the identity but allows for empirical testing. Letting *i* design the unit of observation, the STIRPAT follows the form:

$$I_i = cP_i^{\beta_1} A_i^{\beta_2} T_i^{\beta_3} v_i \,.$$

After taking logarithms:

$$\log I_i = \beta_0 + \beta_1 \log P_i + \beta_2 \log A_i + \beta_3 \log T_i + \varepsilon_i$$

where the beta coefficients represent the elasticity of emissions with respect to population, affluence, and technology; and epsilon is the error term that we can interpret as unobserved technological factors that influence environmental impacts.

I use Vulcan Project data to measure impacts as county-level total emissions of carbon dioxide in 2002 (Gurney et al, 2009). Population comes from the 2000 Census.[3] To measure affluence, I use 2000 Census data on median household income. I also estimate models that include a quadratic term for affluence following the insights from the Environmental Kuznets Curve literature.

In terms of technology measures, researchers typically use proxies of industrial composition and, in particular, variables for manufacturing intensity. I use the proportion of workers in the manufacturing sector as a technology variable. In addition, I estimate models that include the proportion of workers in the transportation, warehousing, and utilities sectors to control for on-the-road emissions and emissions from electric production; and the proportion of

[3] I also estimated models with alternative measures of population such as number of households and urban and rural population. The coefficients of other variables barely change, the insights regarding spatial heterogeneity are the same, and the models with these alternative measures tend to fit the data worse than the model with population.

workers in extraction (mining, oil and gas production) industries as a control for economic activity contributing to fossil fuel emissions.

Climate is a factor that might generate spatial correlation if it is omitted from the models. As Neumayer (2002) argues, heating and cooling requirements influence the demand for energy and, consequently, emissions of carbon dioxide. To account for the effects of climate on emissions via energy consumption, I estimate models that include county average temperatures, in particular, average temperature of three coldest months of 2002 year and average temperature of three warmest months of 2002.[4] Table 1 presents summary statistics for the variables and their logarithmic transformation.

I estimate OLS models and GWR models for each specification. The OLS models include state dummy variables to control for state-specific effects. GWR models do not include state dummies since the approach itself accounts for spatially-specific factors.

4. Methods: Geographically weighted regression

The standard approach to estimating the STIRPAT model is to fit the regression model:

$$\log I_i = \beta_0 + \beta_1 \log P_i + \beta_2 \log A_i + \beta_3 \log T_i + \varepsilon_i \, .$$

[4] To calculate these variables I use weather station data and geoprocessing tools to calculate average temperatures in counties with multiple weather stations and to assign the temperature of the weather station closest to those counties without weather stations. To convert these variables to logs, temperatures below zero are set to zero.

Geographically weighted regression extends the typical specification to allow the parameters to vary across locations. Let (u_i, v_i) denote the longitude and latitude coordinates of location i.[5] The standard model can be generalized to:

$$\log I_i = \beta_0(u_i, v_i) + \beta_1(u_i, v_i)\log P_i + \beta_2(u_i, v_i)\log A_i + \beta_3(u_i, v_i)\log T_i + \varepsilon_i \; .$$

GWR assumes that observations near to location i have a larger influence in the estimation of the parameters in location i than observations farther from i have. In this way, each observation is weighted according to its proximity to each area i.

Let $\mathbf{X}$ denote the matrix of independent variables. The coefficient estimates of the population parameters are estimated as:

$$\mathbf{b}(u_i, v_i) = (\mathbf{X}'\mathbf{W}(u_i, v_i)\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}(u_i, v_i)\mathbf{y} \; ,$$

where $\mathbf{W}(u_i, v_i)$ is a matrix with off-diagonal elements equal to zero and diagonal elements equal to the weight of each observation for location i, w_{ij}.

To calculate spatial weights, a standard approach is to weight observations based on distance from each location j to each location i (d_{ij}) according to a Gausssian function with lower weight as the distance increases up to b, where b is the bandwidth or distance beyond which the weight is zero. The bandwidth can be constant across the entire area of analysis or it can vary. In this application, because the density of counties varies across the area of study, it is better to assume the bandwidth changes over space. In particular, the weighting function for adaptive bandwidth is:

[5] For the following discussion of GWR models I follow closely Fotheringham, Brunsdon, and Charlton (2002). In the case of polygon data, such as counties or census tracts, the coordinates correspond to the geometric center, or centroid, of the polygon.

$$w_{ij} = [1-(\frac{d_{ij}}{b_i})^2]^2 \text{ if } d_{ij} < b_i,$$
$$w_{ij} = 0 \text{ if } d_{ij} \geq b_i,$$

where d_{ij} is the distance between locations i and j, b_i is the Nth nearest neighbor distance from i, and the weight is zero if j is not one of the Nth nearest neighbors to i. Under this formulation, a location that is surrounded by many small-area neighbors will have a smaller bandwidth than a location surrounded by few large-area neighbors. The value of neighbors (or data points) N that is used to estimate each local model is that that minimizes the Akaike Information Criterion (AIC).[6] In this application, the AIC is minimized when the models are estimated with 300 neighbors, approximately ten percent of the data. I also discuss results when N is set to 150 and 450 neighbors.

As with linear regression models, GWR estimation is sensitive to collinearity problems and extreme values. O'Sullivan and Unwin (2010) discuss the inference problems these issues may cause and suggest the method is best used as an exploratory technique. These issues recommend estimating a parsimonious model. In addition, the variables are transformed into logarithms so that the problem of extreme values is less likely to influence the results. Another challenge with spatial data is the modifiable areal unit problem (MAUP). The issue is that the unit of analysis may be arbitrary (administrative units) and a different level of aggregation may generate different results. In this application, data availability determines the area of analysis, counties. Nonetheless, as Fotheringham, Brunsdon, and Charlton (2002) point out, when using GWR models, the data points included in the estimation of each local model are determined

[6] The results are identical if the bandwidth and number of neighbors are selected through the method of cross-validation.

endogenously through the process that selects the number of neighbors, and consequently the bandwidth, that minimizes the AIC.

5. Results

I first estimate linear regression models with state dummy variables. The results are consistent with the predictions and the findings of the STIRPAT model and EKC literature. However, there is strong statistical evidence that the residuals of the linear models are not randomly distributed across space. I then estimate GWR models and show that this approach resolves the issue of spatially correlated residuals. There is also strong statistical evidence to reject the null hypothesis that the parameters of the model are stationary across counties.

Linear Regression Models

Table 2 presents the estimated elasticity of emissions with respect to population, median household income, and measures of technology, as well as goodness-of-fit statistics (R-squared and AIC values). The table also presents values of the Moran's I statistic. This statistic measures spatial autocorrelation such that a value of zero implies a random distribution. The associated z-score indicates whether to reject the null hypothesis that the residuals are uncorrelated versus the hypothesis that the residuals exhibit spatial correlation.

Regarding the coefficient estimates, the elasticity of emissions with respect to population is, as expected, positive. The hypothesis that the elasticity equals one can be rejected. This result implies that, on average, emissions change less than proportionally with changes in population. The elasticity of emissions with respect to median household income is also positive and strongly significant. The models than include a second-order effect for income indicate that, on average, there is evidence of an inverted-U relationship between emissions and median household income. The measures of technology have all a positive and statistically significant impact on emissions. Everything else equal, warmer temperatures during the summer months (increasing on average cooling requirements) correlate with more emissions while warmer temperatures during the winter months (reducing on average heating requirements) decrease emissions.

Goodness-of-fit statistics show that the model that includes proportion of workers in the manufacturing, transportation, warehousing, utilities, mining and oil and gas extraction sectors fits the data best. Adding controls for cooling and heating requirements does not change the fit much, although the coefficient estimates are statistically significant at the 10 percent level.

The results from the linear regression models represent the estimated average or global relationship between carbon emissions and population, affluence, and technology for counties in the U.S. However, the values of the Moran's I statistic and its associated z-score show that, in all models, the null hypothesis that the residuals are not spatially correlated can be rejected at the one percent level in favor of the hypothesis that the residuals exhibit spatial correlation. It is important to note that the models include state dummy variables that account for state-specific factors that might influence emissions in the same way across all counties in a given state.

Geographically Weighted Regression Models

Table 3 presents the results from GWR models for the model specification that includes a quadratic term for median household income, three measures of technology, and climate variables. Since GWR estimation generates a distribution of local coefficients, the table presents the estimated median elasticity as well as the 25^{th}- and 75^{th}-percentiles. The table also displays R-squared and AIC values, Moran's I statistic and its associated z-score. To compare these results to the estimates from linear regression models, the first column in Table 3 shows linear regression estimates and goodness-of-fit statistics from Table 2 (note that while the linear regression models include state dummy variables, as explained above the GWR models do not). The model's AIC is minimized when the number of nearest neighbors that are used to estimate the local models is set to 300 (approximately ten percent of the data points).

First, comparing AIC values across linear and GWR models, GWR models have lower values indicating a better fit (after accounting for the number of parameters estimated). In addition, the values of the Moran's I statistic indicate that the null hypothesis that the residuals are not spatially correlated cannot be rejected at any of the conventional significance levels.

The medians of the estimated elasticity of emissions with respect to population, income, and technology have the same sign than the linear regression estimates and are similar in magnitude. However, there is substantial variability in the distributions of local coefficients of affluence and technology measures, and the null hypothesis that each parameter is stationary can be rejected at the one percent significance level for all independent variables.[7] These results then present strong evidence that the global estimates are not a valid representation of the relationship between emissions and its determinants at all locations.

It might be useful to display the spatial pattern of coefficient estimates and of t-statistics as a first step to understand the nature of regional differences. Figures 1, 2, 5, 7, 9, 11, and 13 present the spatial distribution of the elasticity of emissions with respect to population, affluence, and technology and climate measures. In each map, the classes correspond to the quartiles of the distribution of coefficient estimates (darkest grey corresponds to the upper quartile and the

[7] The test is computed using a Monte Carlo approach under which the sampling distribution of the observed standard deviation of each coefficient estimate is compared to the distribution obtained from randomly shuffling the data in space and estimating GWR models on the rearranged data (Fotheringham, Brunsdon, and Charlton, 2002).

lightest gray corresponds to the bottom quartile). Because the estimates might not be statistically significant in all local models, it is also important to display associated t-statistics. Figures 3, 4, 6, 8, 10, 12, and 14 display the significance levels of local estimates for income, income squared, and measures of technology and temperature (the coefficients on population are statistically significant in all local models).

Figure 1 presents the spatial distribution of the estimated elasticity of total carbon emissions with respect to population. The local coefficient estimates are positive and statistically significant across all counties. The range of estimates goes from .674 to 1.369. The lowest estimates are for counties in Texas, Montana, North Dakota, California, and New York. On the other hand, all counties in Missouri, Iowa, and Arkansas exhibit estimates above one.

While the linear regression model and median estimates of the GWR model show an inverted-U relationship between income and emissions, the local estimates suggest that the impact of affluence on emissions varies substantially across the area of study. Figure 2 displays the estimates of income elasticity based on the coefficients on income and income squared, Figure 3 shows the distribution of t-statistics for the coefficient on income, and Figure 4 displays the t-statistics for the coefficient on income squared. Counties where income elasticity is negative and statistically significant are located in Alabama, Florida, Georgia, Louisiana, and Mississippi (but not all local models in these states show statistically significant coefficients on income). In these cases, the coefficient estimates on income are negative and the coefficient estimates on income squared are positive. These results are consistent with Roberts (2011) who, using a different methodology in a study of nine southern states, finds that the coefficient estimates are either statistically insignificant or negative. Roberts points out that these states have received an influx of industry and employment as manufacturing and skilled blue-collar jobs have shifted from the urban Northeast to the South. Large positive and statistically significant estimates of income elasticity are distributed across counties in Texas, Colorado, and Wyoming, and from Kansas to Pennsylvania and South Carolina. In sum, there is substantial spatial variability in the distribution of the estimated income elasticity with high and low values of the income elasticity across counties in several adjacent states.

The next set of maps display the estimates of elasticity of emissions with respect to the different measures of technology. Higher manufacturing employment is correlated with higher emissions of carbon dioxide across counties in the Midwest and South (Figures 5 and 6). In the case of employment in the transportation sector, the estimates are statistically significant and positive across most counties (Figures 7 and 8). Employment in the extraction sector is also positively correlated with emissions across the West and Midwest, Florida, Georgia, Virginia, New York, and Pennsylvania (figures 9 and 10). Higher temperatures during the winter months are negatively correlated with emissions in most counties in the West and positively correlated with emissions across counties in the South (Figures 11 and 12). Regarding summer temperatures, Figures 13 and 14 show areas with positive and negative statistically significant coefficients across the U.S. without a clear pattern.

Finally, Figure 15 maps the distribution of local R-squared statistics that range from 48.5 percent to almost 90 percent of variability explained by the model. The model predicts more than 80 percent of the variability in emissions in the Pacific region, New England, West North Central region, and Florida. The model predicts 65 percent or less of the variability in counties in South East Central and West Central regions.

Model results depend on the model's bandwidth. Although there is no theoretical reason to fix the number of neighbors to a specific value, I also estimate the GWR models with 150 and 450 neighbors to explore how the results change with the bandwidth. As expected, the fewer the number of neighbors the greater the variability in estimates. The distribution of coefficient estimates on population is remarkably stable. For the other estimates, median estimates barely change expect for the coefficient on COOL (the medians for the models with 150, 300, and 450 neighbors are, respectively, -.02, -.08, and -.11). Qualitatively, upper and lower quartiles are consistent across models with different number of neighbors. In terms of magnitude, upper and lower quartiles are generally of similar magnitude with two main exceptions: the coefficient on MHI is .4 in the model with 150 neighbors but 5.3 and 6.8 in the models with 300 and 450 neighbors, respectively; while the coefficient on MANUF is .008 in the model with 150 neighbors but .17 and .22 in the models with 300 and 450 neighbors, respectively. Importantly, even with 450 neighbors the null hypothesis that the parameters are stationary can be rejected at the one percent level (at the four percent level for the control for transportation and utilities).

6. Conclusions

This paper uses U.S. county-level data to account for spatial correlation and geographical variability in the process liking carbon emissions to measures of population, affluence, and technology. The results show that linear regression models may result in flawed statistical inference as there is evidence that the residuals of linear models are spatially correlated across the U.S., even after including state dummy variables. GWR models solve the problem of correlated residuals and also provide evidence that the parameters of the STIRPAT model are not stationary across space. Displaying the local coefficient estimates and their t-statistics shows that there are spatial patterns of relationships between emissions and its determinants. This research cannot explain what causes these patterns over space but it may provide a useful starting point to aim for a more accurate specification or to examine in detail the nature of variation within and across counties.

It is possible that these models are misspecified and that spatial heterogeneity would be resolved within the framework of a global linear regression model if all factors that determine the spatial distribution of emissions were included. In practice, there are likely to be unobservable local geographical, historical, and cultural factors that may influence environmental impacts.[8] One important insight from this research is that compared to standard regression models, parsimonious GWR models account for spatial heterogeneity and generate output that allows mapping and examining the spatial distribution of coefficients; this could help deciding which omitted variables should be included and also address the question of how to design and implement policy across political units.

These results suggest that policymakers should account for regional differences that are not necessarily contained within predefined political units. Clearly, federal policy that does not account for the local context may result in less than optimal outcomes both in terms of overall impact and in terms of a fair distribution of costs. Local estimates are more stable within states but there is still evidence that the magnitude of the impacts vary within states and that there are patterns that cross state lines.

In general, if researchers and policy-makers are interested exclusively on average relationships, then a global model that accounts for spatial autocorrelation might be sufficient. If local variability matters for efficiency and fairness, then models that generate the spatial

[8] In addition, the models in this paper explain a large proportion of the variability in the dependent variable.

distribution of coefficient estimates are a useful exploratory tool to test the extent of spatial heterogeneity and to identify spatial patterns.

References

Brander, J.A., 2007, "Sustainability: Malthus revisited?" *Canadian Journal of Economics*, 40(1): 1-38.

Carson, Richard T., 2010, "The environmental Kuznets curve: Seeking empirical regularity and theoretical structure," *Review of Environmental Economics and Policy*, 4(1): 3-23.

Fotheringham, Stewart A., Brunsdon, Chris, and Martin Charlton, 2002, *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*, John Wiley & Sons, West Sussex, England

List, John A., and Craig A. Gallet, 1999, "The environmental Kuznets curve: Does one size fit all?" *Ecological Economics,* 31:409-423.

Maddison, David, 2006, "Environmental Kuznets curves: A spatial econometric approach," *Journal of Environmental Economics and Management,* 51:218-230.

Neumayer, Eric, 2002, "Can natural factors explain any cross-country differences in carbon dioxide emissions?, *Energy Policy* 30: 7-12.

Pizer, William, Sanchirico, James, M., and Michael Batz, 2010, "Regional patterns of U.S. household carbon emissions," *Climatic Change* 99: 47-63.

Rupasingha, Anil; Goetz, Stephan J.; Debertin, David L.; and Angelos Pagoulatos, 2004, "The environmental Kuznets curve for US counties: A spatial econometric analysis with extensions," *Papers in Regional Science* 83:407-424.

A. Shi, 2003, "The impact of population pressure on global carbon dioxide emissions, 1975-1996: Evidence from pooled cross-country data," *Ecological Economics* 44: 29-42.

Tevie, Justin, Grimsrud, Kristine M., and Robert P. Berrens, 2011, "Testing the environmental Kuznets curve hypothesis for biodiversity risk in the U.S.: A spatial econometric approach," *Sustainability* 3: 2182-2199; doi:10.3390/su3112182

Tyler D. Roberts, 2011, "Applying the STIRPAT model in a post-Fordist landscape: Can a traditional econometric model work at the local level?" *Applied Geography* 31: 731-739

York, Richard, Rosa, Eugene A., and Thomas Dietz, 2003, "STIRPAT, IPAT, and ImPACT: Analytical tools for unpacking the driving forces of environmental impacts," *Ecological Economics* 46: 351-365

Table 1: Summary Statistics: mean and standard deviation in parenthesis

	Log(1+variable)	Original values
Emissions	11.85 (1.53)	511,542.6 (1,205,261)
Population	10.23 (1.23)	89,954.68 (293,558)
Median Household Income	10.44 (.23)	$35,266.95 ($8,836.6)
Proportion workers in Manufacturing	.145 (.077)	.159 (.091)
Proportion workers in Transportation, Utilities	.053 (.017)	.055 (.018)
Proportion workers in Extraction	.011 (.025)	.012 (.027)
Temperature Summer months*	6.75 (.060)	267.69 (89.37)
Temperature Winter months*	5.52 (.47)	858.42 (50.73)

*Temperatures are recorded such that average of 858.42 means 85.8 Fahrenheit

Table 2: Linear Regression Model Results (variables defined as log of 1 plus raw value)

	(1)	(2)	(3)	(4)	(5)
VARIABLES					
Population	0.886***	0.891***	0.909***	0.918***	0.923***
	(0.015)	(0.014)	(0.014)	(0.014)	(0.014)
Median Household Income	0.508***	25.008***	21.864***	22.271***	21.861***
	(0.096)	(4.148)	(4.105)	(4.169)	(4.167)
Median Household Income, squared		-1.168***	-1.021***	-1.039***	-1.020***
		(0.196)	(0.194)	(0.197)	(0.197)
Proportion workers in Manufacturing	0.440*	0.156	0.651**	0.973***	1.014***
	(0.259)	(0.257)	(0.254)	(0.257)	(0.258)
Proportion workers in Transportation, Utilities			11.422***	11.062***	11.043***
			(1.435)	(1.393)	(1.396)
Proportion workers in Extraction				4.359***	4.352***
				(1.049)	(1.042)
Temperature Summer months					0.632*
					(0.372)
Temperature Winter months					-0.106**
					(0.051)
Constant	-2.313**	-130.737***	-114.804***	-117.350***	-118.919***
	(0.917)	(21.895)	(21.662)	(22.014)	(22.119)
Observations	3108	3108	3108	3108	3108
AIC	7653.251	7611.194	7458.264	7416.586	7418.022
R-squared	.714	.718	.732	.736	.736
Moran's I (z-score)	0.033*** (6.573)	.024*** (4.866)	.016*** (3.282)	.015*** (2.929)	.014*** (2.831)

Models include state dummy variables

Table 3: OLS and GWR results (variables defined as log of 1 plus raw value)

	OLS	GWR	GWR	GWR
		Median	Lower quartile	Upper quartile
Population	0.923***	1.006	0.911	1.064
	(0.014)			
Median Household Income	21.861***	18.148	5.346	31.482
	(4.167)			
Median Household Income, squared	-1.020***	-0.854	-1.481	-0.251
	(0.197)			
Proportion workers in Manufacturing	1.014***	1.087	0.165	2.601
	(0.258)			
Proportion workers in Transportation, Utilities	11.043***	6.119	10.338	15.482
	(1.396)			
Proportion workers in Extraction	4.352***	7.859	2.806	14.120
	(1.042)			
Temperature Summer months	0.632*	0.3259	-1.611	1.879
	(0.372)			
Temperature Winter months	-0.106**	-0.083	-0.373	0.527
	(0.051)			
Constant	-118.919***	-97.157	-174.076	-23.275
	(22.119)			
	STATE Dummies			
Observations	3108	3108		
AIC	7418.022	7232.639		
R-squared	.736	0.777		
Moran's I (z-score)	.014*** (2.831)	-0.00466 (-0.853)		

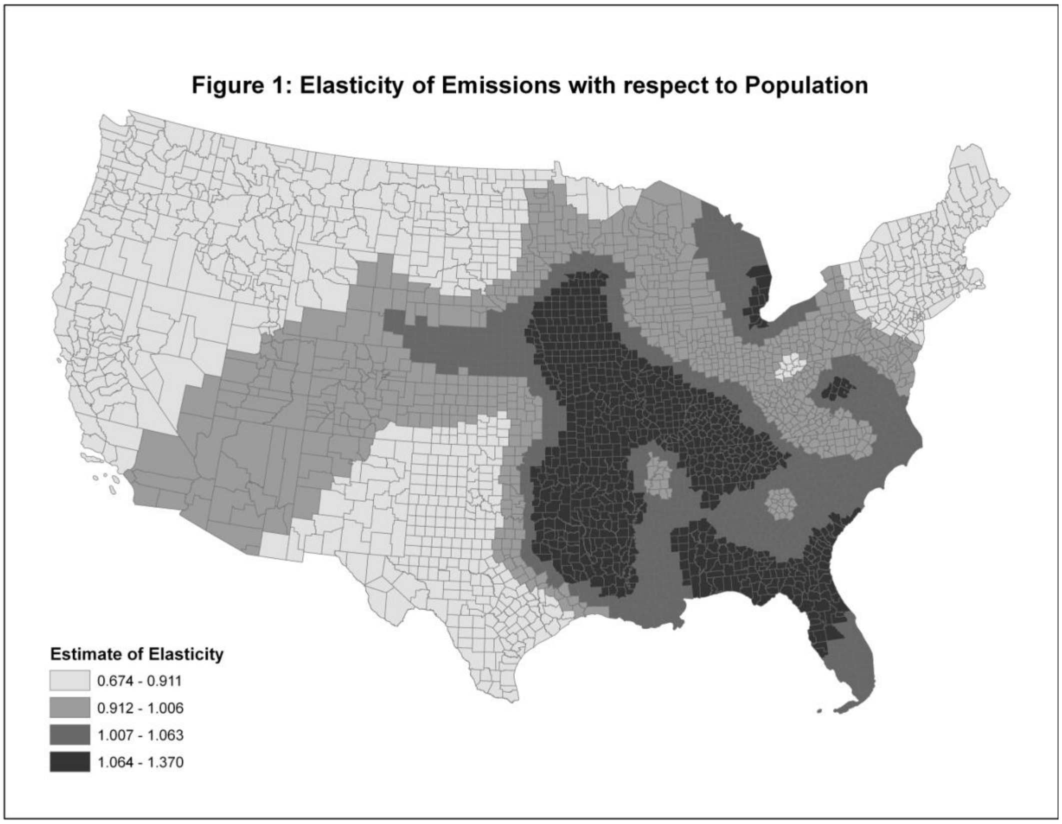
Figure 1: Elasticity of Emissions with respect to Population
Estimate of Elasticity
0.674 - 0.911
0.912 - 1.006
1.007 - 1.063
1.064 - 1.370

Figure 2: Elasticity of Emissions with respect to Income

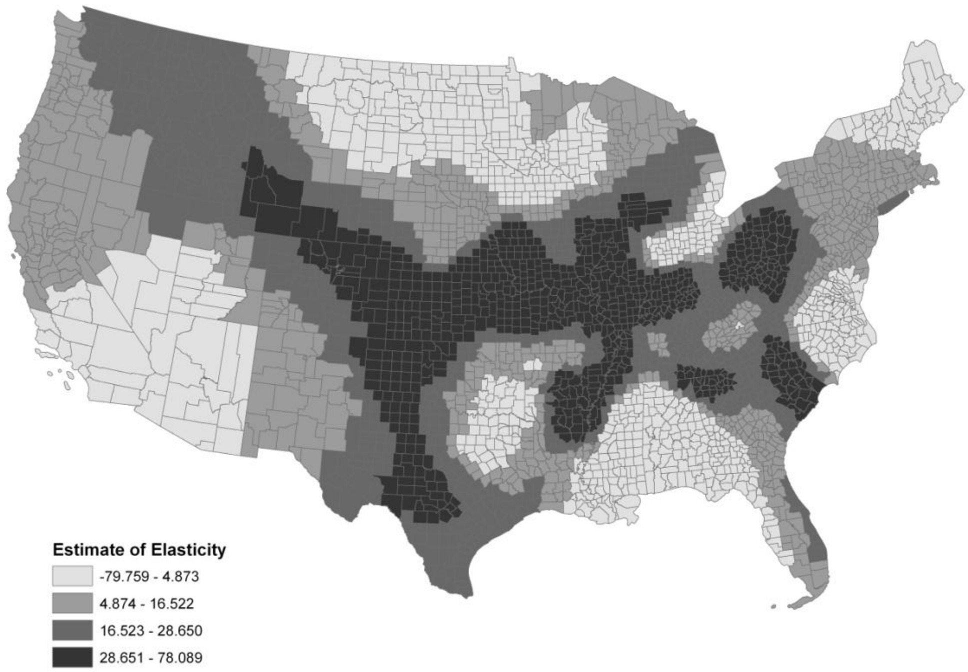

Figure 3: Significance Level of Coefficient on Income

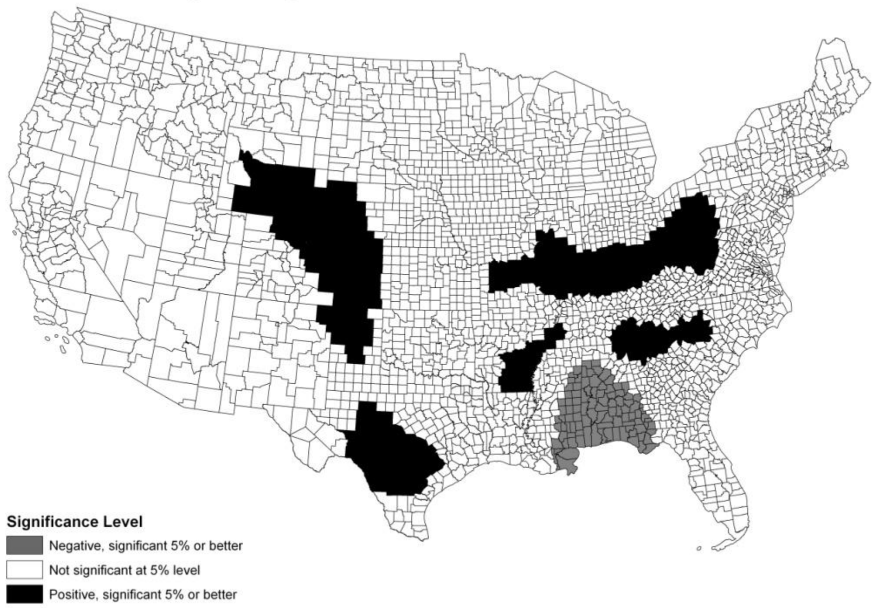

Figure 4: Significance Level of Coefficient on Income squared

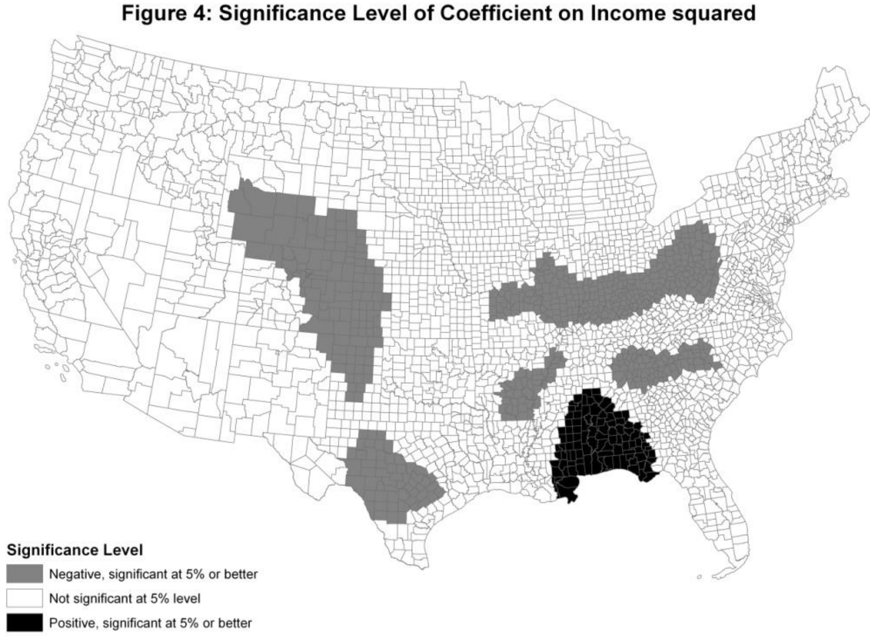

Figure 5: Elasticity of Emissions with respect to Manufacturing

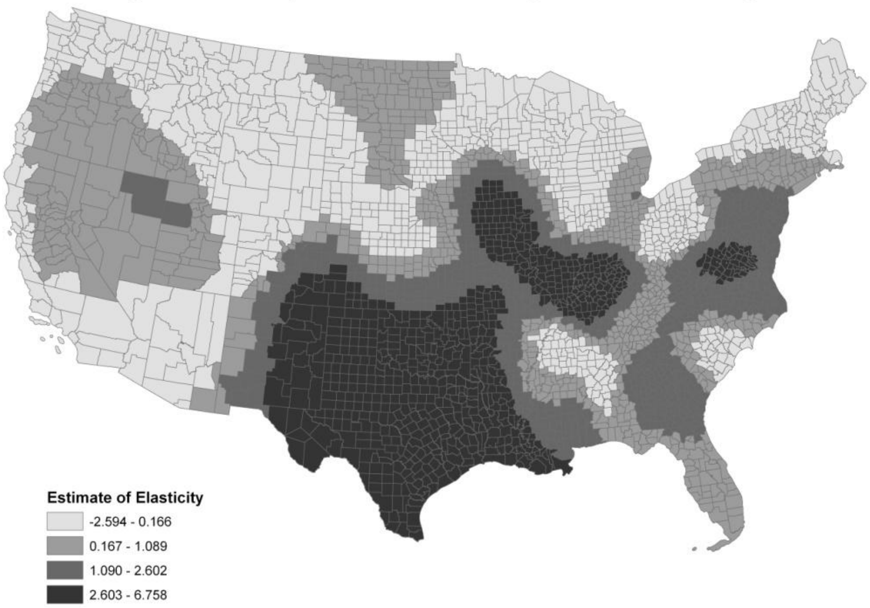

Figure 6: Significance Level of Coefficient on Manufacturing

Figure 7: Elasticity of Emissions with respect to Transportation

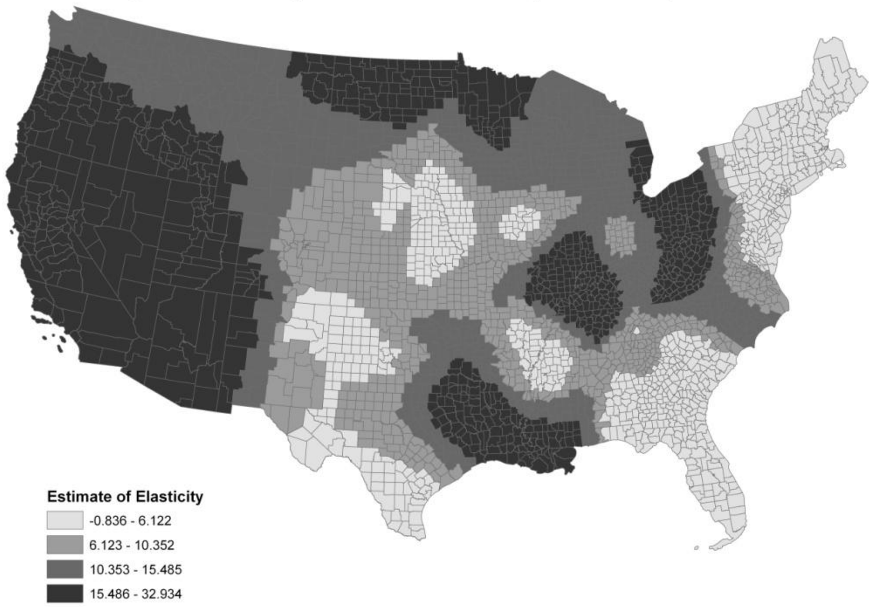

Figure 8: Significance Level of Coefficient on Transportation

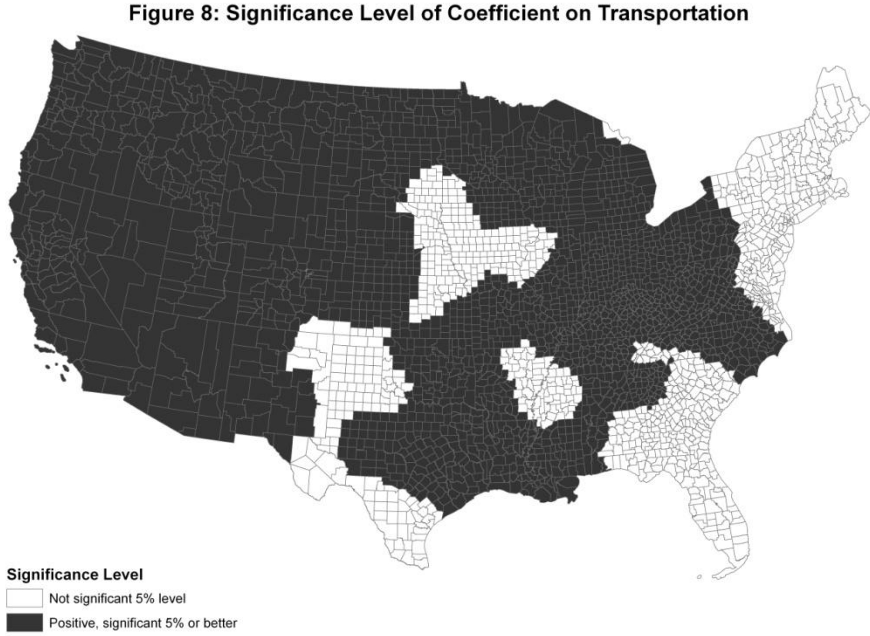

Figure 9: Elasticity of Emissions with respect to Extraction

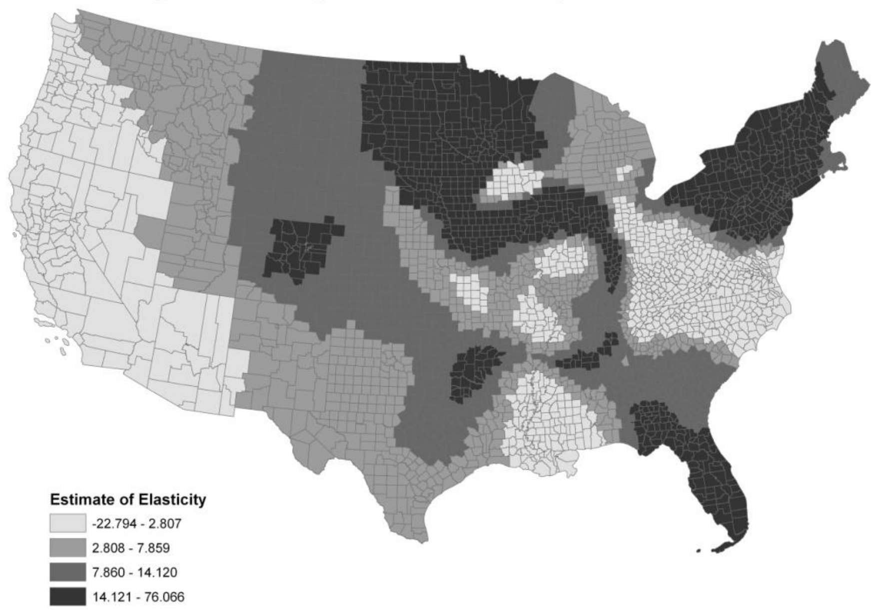

Figure 10: Significance Level of Coefficient on Extraction

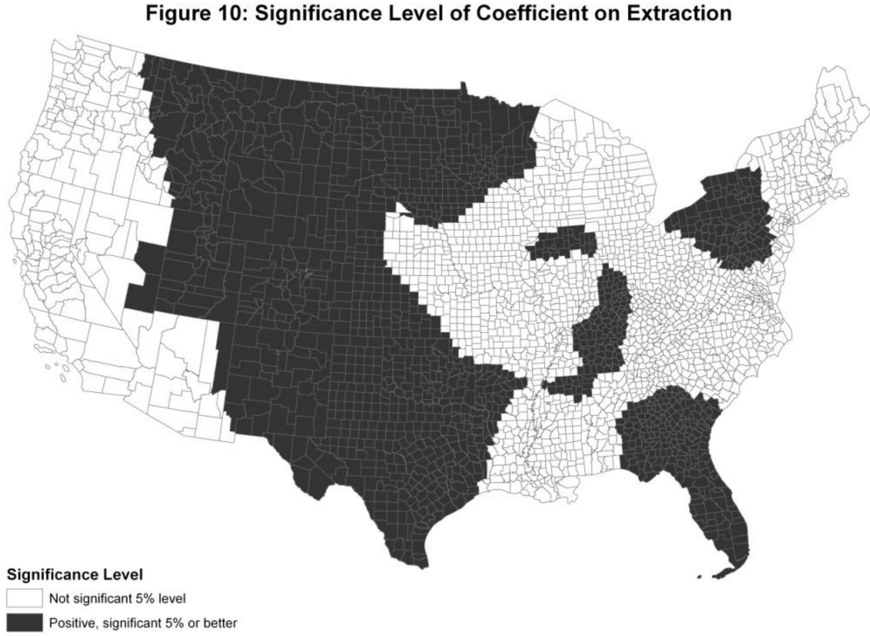

Figure 11: Elasticity of Emissions with respect to Climate, Cold Months

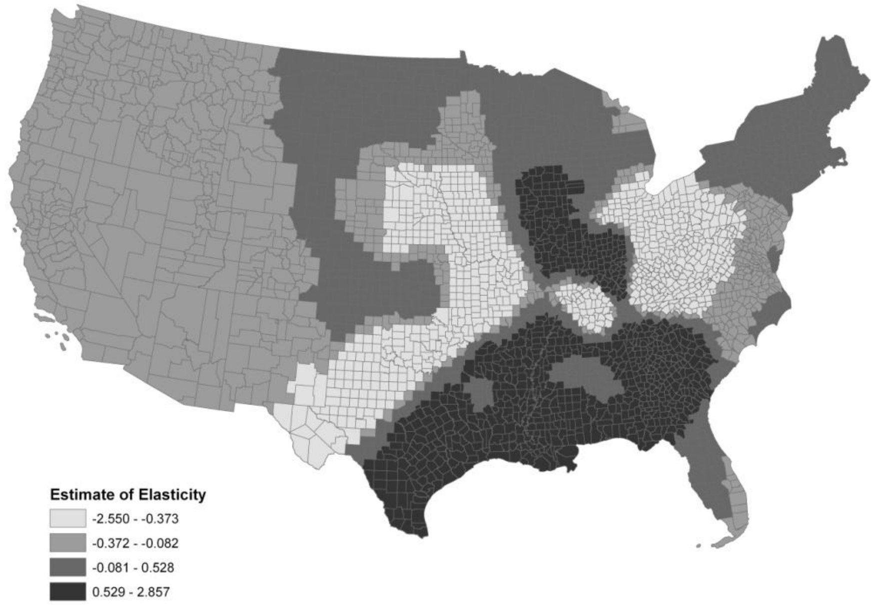

Figure 12: Significance Level of Coefficient on Cold Months

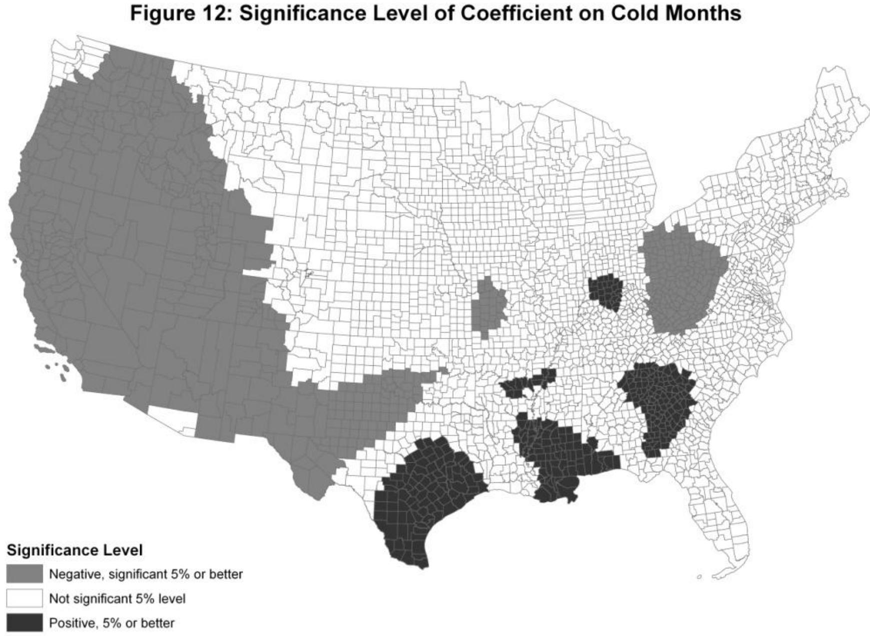

Figure 13: Elasticity of Emissions with respect to Climate, Warm Months

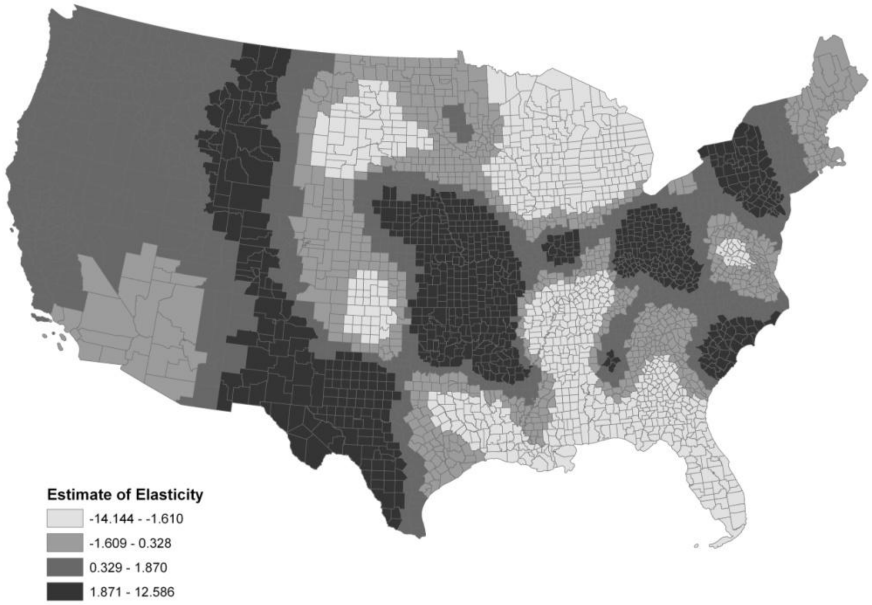

Figure 14: Significance Level of Coefficent on Warm Months

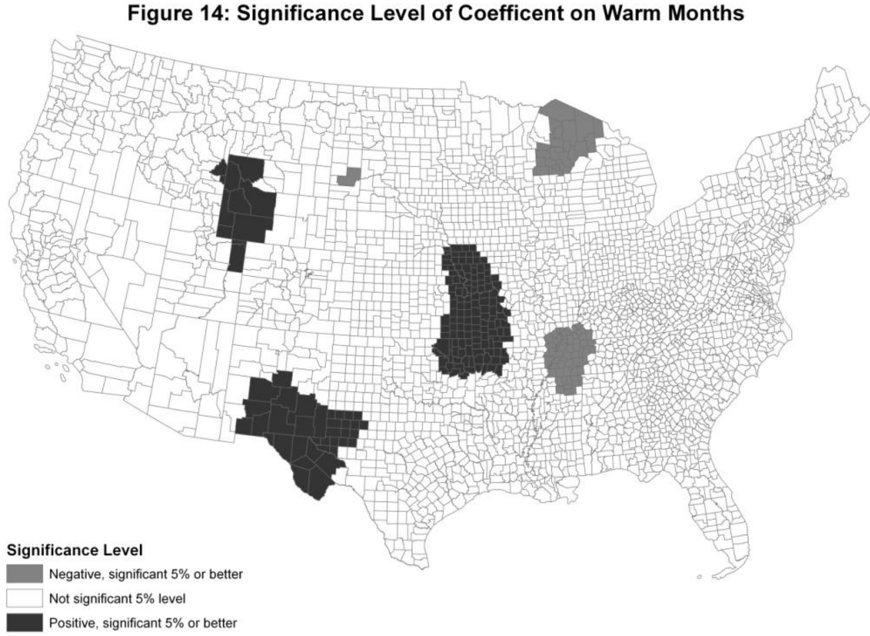

Figure 15: Local R-Square Statistics