Have greenhouse gas emissions from US energy production peaked? State level evidence from six subsectors

Michael Cary

Abstract Analyses of the Environmental Kuznet's Curve (EKC) hypothesis have largely focused on economy level data with occasional analyses exploring sector level data. This paper exploits a new data set which contains sector level data on greenhouse gas emissions from the US energy sector as well as subsector data from six disjoint subsectors which together comprise the entire energy sector. The data contained in this data set is annual data at the state level from 1990 through 2011. By using differenced data we specify an econometrically sound EKC model and compare it against a model containing only a linear GDP per capita term. We find that by using a subsector level modelling approach, evidence for the EKC hypothesis is virtually nonexistent. Moreover, we find that aggregated subsector level estimates outperform sector level estimate on in-sample accuracy. These estimated models are then used to forecast emissions for the energy sector. We find evidence that US greenhouse gas emissions from energy production are at or near a peak.

Keywords Emissions · Energy · Environmental Kuznet's Curve · Panel Data · Subsector Analysis

JEL classification Q53, Q54, Q56, Q57

1 Introduction

The primary source of greenhouse gas (GHG) emissions from an economy is the energy production sector. Marrero (2010) determined that greater than 65% of the world's GHG emissions come from energy production with many OECD nations generating greater than 80% of their total GHG emissions from energy production. According to data from the WRI CAIT Climate Data Explorer (2014), the data set used for this paper, during the period from 1990 through 2011 the energy production sector accounted for more than 88% of total GHG emissions in the United States. Considering the contribution to total GHG emissions from the energy sector of the US economy, any analysis or forecast of GHG emissions in the United States must seriously consider the dynamics of GHG emissions due to energy production.

With new data from the WRI CAIT Climate Data Explorer data set, we explore for the first time the dynamics of six subsectors of the energy production sector of the US economy. By order of magnitude of GHG emissions, these six subsectors are: electric power, transportation, industrial, residential, commercial, and fugitive emissions. By not only studying the energy sector as a whole, but also by studying the behavior of its subsectors, we can improve our understanding of GHG emissions from energy production and thereby enhance forecasts of GHG emissions. By showing that this forecasting approach leads to greater accuracy when dividing data into training and testing periods, aggregated subsector forecasts are shown to be superior to sector level forecasts.

Since the quantity and composition of GHG emissions varies across sectors of the economy, and similarly across subsectors of the energy sector, no two subsectors of energy production should expect to benefit equally from any given emissions mitigation program. Moreover, technological advancement does not affect the energy production sector ubiquitously. By breaking the energy sector down into subsectors, we can increase our ability to estimate the effects of subsector specific tech-

M. Cary
Division of Resource Economics and Management
West Virginia University
E-mail: macary@mix.wvu.edu

nological advancements or emissions mitigation policies. As energy consumption habits are known to change with income, see, e.g., Caron and Fally (2018), energy production portfolios should vary across income levels and thus across states. For all of these reasons, finding an overall trend in GHG emissions from energy production will not be as accurate as finding an aggregated trend of the subsectors of energy production.

In order to study energy subsector GHG emissions, we use an econometrically sound variant of the Environmental Kuznet's Curve (EKC) approach. Under this approach, GHG emissions are differenced and these differences are estimated primarily as a function of GDP per capita. The EKC hypothesis states that there is an inverted "U" shape relationship between GHG emissions and GDP per capita. To test this hypothesis, models using the EKC approach include both GDP per capita and GDP per capita squared. The ecological consequences of this hypothesis are that, if true, GHG emissions are not a monotonically increasing function of production. That is to say, as production continues to increase, overall levels of emissions may actually decrease. Possible explanations for this behavior include the notion that the adaption of greener technologies, the switch to renewable sources of energy, and the implementation of other abatement strategies are all monotonically increasing functions of income.

The two overwhelming drivers of emissions from the energy production sector are electric power generation and transportation. Both electric power generation and transportation are areas which are experiencing rapid growth in the adaptation of greener technologies, the switch to renewable sources of energy, and in the growth of regulation which imposes, through various means, emissions limitations (Fox et al., 2017). However, the methods employed to achieve emissions reductions in these two subsectors are not identical, so our method of isolating these subsectors and studying and forecasting them independently of each other allows us to determine not only which factors influence emissions from energy production, but also how these factors influence emissions at the subsector level. Similarly, we can test not only whether the EKC hypothesis is likely for the energy production sector as a whole, but how valid the EKC hypothesis is for each of its subsectors. By studying the determinants of GHG emissions at the subsector level, better policy proposals and decisions regarding emissions and emissions abatement strategies can be made through this more in depth targeting of the energy production sector.

The remainder of this paper follows the following structure. After providing a review of the relevant literature, we describe our data set and the methodologies we employ to determine which factors influence emissions from the energy sector and its subsectors in the United States. At this point the results from the various regression models are presented and their results and implications for the EKC hypothesis are discussed. In particular, we show that when analyzing the subsectors of the energy production sector, no support for the EKC hypothesis exists. Moreover, we show through use of the Akaike Information Criterion (AIC) that a model specified to test for the EKC hypothesis is effectively not different from a model which includes only a linear term for GDP per capita. Once we have an understanding of these results, we describe our forecasting technique which comes from Selden and Song (1994) and forecast state level and total US GHG emissions for the energy production sector both independently and as an aggregate of the forecasts for each of its subsectors. To determine the validity of our aggregated forecasting technique, we show that aggregated fitted values from subsector models have greater in-sample accuracy in the energy sector level data by using a secondary data set from the EPA which includes national level GHG emissions data for an additional five years beyond what is offered by the WRI data set. Consequentially, we see that US GHG emissions from the energy production sector have likely peaked.

2 Literature Review

Selden and Song (1994) popularized the EKC hypothesis as a means for studying various GHG emissions. After testing the EKC hypothesis, which they found evidence for in the case of several greenhouse gases, they forecasted future emissions and found that GHG emissions were likely to taper off, though not decrease, over the next several decades. Since then, several papers have studied and tested the EKC hypothesis in various contexts.

Early support for the confirmation for the EKC hypothesis was given by List and Gallet (1999). From there, numerous case studies on various nations, US states, and global regions were performed. Friedl and Getzner (2003) studied the EKC hypothesis in Austria and found evidence for the standard EKC model with a cubic term for GDP per capita included. Similarly, Shittu et al. (2018) studied the EKC hypothesis in Malaysia, Kunnas and Myllyntaus (2007) studied the case of Finland, and Soytas and Sari (2009) did the same in Turkey. At the state level in the United States, Konan and Chan (2010) performed a case study on Hawai'i and Yang et al. (2009) performed a case study on California. On a much larger geographic scale, Zoundi (2017) studied the EKC hypothesis in Africa.

This analysis found ambiguous evidence, showing that in Africa CO_2 levels rise monotonically with rises in income.

However, not every study has provided the same level of support for the EKC hypothesis. Azomahou et al. (2006) found that the EKC hypothesis has more or less support in different countries and under different statistical approaches. The standard model for EKC hypothesis testing did exhibit evidence for the EKC hypothesis, but a non-parametric model found only limited evidence which varied widely by country. In the context of sulfur emissions specifically, Stern and Common (2001) found that there was support for the EKC hypothesis, but only for high income countries. The standard EKC model found monotonically increasing sulfur emissions when using a global data set.

There is much concern over the specification of EKC models due to the inclusions of multiple powers of GDP per capita. When a cubic term is included, it is possible to see an 'N' shaped model due to the effects of the cubic term. While this appears to lend some support to the EKC hypothesis over what could be considered a realistic income domain, Canas et al. (2003) pointed out that the 'N' shaped curve can perhaps be better thought of as an 'augmented' EKC. Hüttler et al. (1998) went further and asserted that the 'N' shaped curve supports not the EKC hypothesis, but a perverted variant in which emissions reduction with economic growth is merely a transitory phase that precedes a final phase in which further economic growth drives a rapid growth in emissions. Because of this, Canas et al. (2003) recommended that any inference of support for EKC models relying on significant effects from the cubic term should be taken cautiously.

Many more robust approaches for EKC models have been developed. Spatial models were first used as a tool for assessing the EKC hypothesis by Rupasingha et al. (2004). Extending the use of more sophisticated econometric approaches, Maddison (2006) studied a spatial variant of the EKC hypothesis for national emissions of SO_2 and NO_x and found evidence that national emissions are lowered by proximity to nations with high per capita income, contradicting the notion that nations must achieve increased environmental quality at the expense of their neighbors. Burnett et al. (2013a) also studied several spatial panel data models in the context of CO_2 emissions in the United States, and Ordás Criado (2008) performed a case study on Spain using non-parametric estimation techniques.

In addition to testing the EKC hypothesis with semi- and non-parametric and spatial econometrics models, recent emphasis has been placed on validating estimations and on the importance of the data itself. Chow and Li (2014) established a method using t-tests for validating conclusions supporting the EKC hypothesis using standard econometric approaches. However, Millimet et al. (2003) demonstrated that semi-parametric techniques tend to be superior to standard econometric approaches. As for the choice of data set on the part of the researcher, both Harbaugh et al. (2002) and Galeotti et al. (2006) showed that the data set chosen does not significantly impact any conclusions drawn from statistical analyses.

In lieu of testing the EKC hypothesis itself, the topic of convergence of per capita GHG emissions has also garnered recent attention in the literature. Aldy (2006) studied the convergence of per capita GHG emissions in OECD countries and found evidence supporting this idea. Criado and Grether (2011) found with non-parametric methods that convergence in per capita emissions existed only within groups of similar countries. Aldy (2007) found no evidence, in the context of the United States, for state-wise convergence in per capita emissions, while Burnett (2016) found a group of 26 states for which there is support for convergence in emissions. Relatedly, Tol et al. (2006) studied the EKC hypothesis for emissions intensities and Burnett et al. (2013b) found the economic growth drives not absolute emissions but emissions intensities.

Not all sectors of the economy are equally responsible for producing emissions. For this reason several papers have explored the EKC hypothesis for certain sectors of the economy. Using data from the industrial sector, Edelenbosch et al. (2017) forecasted decreasing energy intensities for this sector in non-OECD countries following a brief increase in industrial energy demand in the short term. Suri and Chapman (1998) first studied the EKC hypothesis in the context of energy consumption. Expanding on their results, Soytas et al. (2007) investigated energy consumption in the United States and found that increases in energy use, not increases in income, cause carbon emissions in the United States by using a Granger causality test. This result is noteworthy because it implies that economic growth alone is not a solution for long term emissions reductions.

Furthering these sector level analyses, Bruvoll and Medin (2003) looked at potential economic factors driving air pollution and found that changes in the relative size of different sectors of the economy helps drive changes in emissions. Mazzanti et al. (2008) used data on value added and capital stock per employee as covariates and found that the inclusion of these measures of labor productivity led to mixed support at the sector level for the EKC hypothesis in the case of Italy.

In the United States, another interesting way to subdivide emissions is at the state level. Aldy (2005) studied per capita emissions at the state level, finding that the estimated EKC varied by state. Using a spatial autoregressive model, Burnett and Bergstrom (2010) found further evidence for the EKC hypothesis in the United States. Auffhammer and Steinhauser (2012) found that, while standard EKC models did well with in sample accuracy (R^2, Akaike Information Criterion, Schwartz Information Criterion, and others), they were not the top performing models for out of sample testing. Combining the state level and sector specific analyses, Zhou and Gurney (2011) used a spatial econometric model to study sector specific emissions and the state and county level in the United States. Using surface temperature data, they found that emissions due to space cooling in warmer climates should more than offset any future emissions reductions due to reduced needs for space heating. A study by Apergis et al. (2017) found limited evidence for the EKC hypothesis at the state level (10 of 48 states fit the hypothesis) using a standard EKC model without a cubic term.

Using a similar basis, namely province level EKC testing on province level data from China, Auffhammer and Carson (2008) forecasted aggregate Chinese emissions. Of particular interest in this paper was the response by the forecasted Chinese emissions to the Kyoto Protocol. More generally, the theme of increased emissions mitigation techniques, ranging from legislation to greener technologies to socioeconomic inequality and beyond, to combat global climate change and increased air pollution have also been studied in the context of the EKC hypothesis. Magnani (2000) studied how income inequality affects investment into emissions mitigation R&D and found that income inequality leads to a gap between the willingness to pay of a country and its ability to pay.

The negative externalities associated with GHG emissions and how they will impact the economy are currently attracting much attention from interdisciplinary researchers. Fankhauser and Tol (2005) studied the effects of climate change on capital accumulation, finding that forward looking agents are likely to change their savings behavior which will lead to reduced economic growth. A growth model with endogenous emissions reduction was created by Criado et al. (2011), and by using data from European nations they were able to validate this theoretical model. Simulations of the impact of automation on emissions at the household level were performed by Ringel et al. (2019), demonstrating evidence that automation of lighting, heating, and cooling in a household can help reduce emissions. Al-Mulali et al. (2015) studied contributing factors to air pollution in Vietnam and Yates and Strzepek (1998) found that climate change had a minimal impact on agricultural production in Egypt. Beyond agricultural production, the negative externalities due to emissions are vast and include depletion of the ozone layer, primarily due to N2O emissions (Ravishankara et al., 2009). Additionally, impacts on human health due to emissions, including increased risks of various chronic respiratory diseases and cancers (Kampa and Castanas, 2008), are also concerns which emissions mitigation policies attempt to address. The economic impact of increased health risks due to emissions are well documented (West et al., 2013).

We conclude the literature review section of this paper by acknowledging several well written survey papers on the EKC hypothesis which the reader may find helpful, including (Dinda, 2004), (Gill et al., 2018), (Stern, 2004), and (Stern, 2017).

3 Data

The basis for our data set is from the WRI CAIT Climate Data Explorer (2014). This data set is the only data set known to the author which includes GHG emissions data not only for the energy production sector, but for its comprising subsectors (electric power generation, transportation, industrial, residential, commercial, and fugitive emissions). This data considers total GHG emissions in Mt CO_2 equivalent rather than by Mt of individual chemical species, i.e., it is an aggregate measure of emissions. This data is at the state level, includes estimates of GDP per capita and population, and covers the time period from 1990 to 2011. While 22 years is a relatively short period of time, the novelty and level of specification of this data set allows us to be the first to analyze GHG emissions at the state level for subsectors of the US energy production sector.

To clarify the scope of the subsectors, they are defined as follows. The electric power generation subsector reflects emissions from the primary generation of electricity. The transportation subsector reflects emissions from passenger vehicles, trucks, ships, trains, and plains, typically from the combustion of fossil fuels in combustion engines. The industrial subsector reflects emissions from production processes for the production of goods. The residential subsector reflects emissions from residential structures, such as emissions from heating or cooling. The commercial subsector reflects emissions from non-production aspects of industry, e.g., emissions from heating or cooling retail stores. Finally, fugitive emissions are emissions that escape into the atmosphere due to leaks, mechanical failures, or other unintentional causes, i.e., these are emissions that were intended to be captured but were not.

The socioeconomic data used in this study includes GDP per capita and population density. The WRI data set included GDP per capita and population. To derive population density, the given population of each state was divided the area of the state in square kilometers. To address the various state level portfolios of energy sources (i.e., clean versus fossil fuels), we use data on this blend from the United States Energy Information Administration (2019). Because climate change is a significant concern in forecasting future energy consumption and emissions, we include data on heating degree days and cooling degree days from United States National Climate Data Center (2019). Heating degree days are defined as the sum of the number of degrees that average temperature is above 65 degrees Fahrenheit over a given time period (in our case annually). Cooling degree days are defined similarly by using the sum of the number of degrees the average temperature is below 65 degrees. These data are the standards used in the literature. We differentiate between these because they often use different fuel sources. Summary statistics for the data are presented below in Table 1.

Table 1: Summary statistics for the data

Variable	Mean	St. Dev.
GHG Emissions	121	118
GDP per capita	52,355	12,850
Energy Blend	0.877	0.246
Heating Degree Days	5,280	2,084
Cooling Degree Days	1,065	804
Population Density	61.14	76.30

The data for the subsectors exhibit several important quantitative and qualitative properties. First, the sum of GHG emissions over all subsectors in a fixed year and state are equal to the total GHG emissions from the energy sector in that same state and year. Qualitatively, there are two primary drivers of the energy production sector (electric power generation and transportation), and the three largest subsectors account for nearly 90% of GHG emissions from US energy production. The average contribution of each subsector to total GHG emissions from energy production over all states and years in the data set is summarized in the table below.

In order to validate the forecasts using the WRI data, we use national level data from the United States Environmental Protection Agency (2019) which extends through 2017.

Table 2:
Contribution of each subsector to GHG emissions from energy production

Subsector	Contribution
Electric Power	36.23%
Transportation	32.16%
Industrial	18.60%
Residential	6.18%
Commercial	3.90%
Fugitive Emissions	2.91%

4 Methodologies

This paper has two primary goals: to test the EKC hypothesis for GHG emissions from the US energy production sector and its subsectors, and to forecast GHG emissions from US energy production using these results. In this section we detail the regression models used to test the EKC hypothesis on the data set.

As detailed in the literature review section of this paper, several regression techniques, including fixed effects models, random effects models, autoregressive models, semi- and non-parametric techniques, and spatial regression models have been used on various data sets to test the EKC hypothesis. In this paper we use two different models applied to each of the six subsectors of the energy production sector as well as the energy production sector itself. These models are both variants of autoregressive models with covariates, with one model testing the EKC hypothesis (i.e., it includes GDP^2) and one which does not include the GDP^2 term to serve as an econometrically sound comparison for the EKC model. Eventually, the estimated coefficients from both models will be used in our forecasts.

Model misspecification is a common problem faced when testing the EKC hypothesis which causes correlated residuals Burnett et al. (2013a). In order to test the EKC hypothesis for GHG emissions in a statistically sound manner, we consider a model which includes the lagged dependent variable (in our case this is emissions) as a determinant of the current level of emissions. The inclusion of the lagged dependent variable helps eliminate any potential endogeneity without concern of model misspecification Achen (2000). Given this, the EKC model takes on the following specification in which the variable $GHG_{s,t}$ represents total GHG emissions from state s in year t, $y_{s,t}$ represents GDP per capita in state s in year t, $p_{s,t}$ represents the population density of state s in year t, $h_{s,t}$ represents the number of heating degree days in state s in year t, and $c_{s,t}$ represents the number of cooling degree days in state s in year t.

$$GHG_{s,t} = \alpha GHG_{s,t-1} + \beta_1 y_{s,t} + \beta_2 y_{s,t}^2 + \beta_4 p_{s,t} + \beta_5 h_{s,t} + \beta_6 c_{s,t} + \epsilon_{i,t} \quad (1)$$

If the emissions generation process exhibits a unit root, then testing the EKC hypothesis on emissions levels becomes equivalent to testing the EKC hypothesis on changes in emissions. That is, if the estimated parameter $\hat{\alpha}$ from Equation 1 is not statistically different from unity, then Equation 1 may be replaced with a model for the change in GHG emissions. We test this with the following null hypothesis.

$$\mathcal{H}_0 : \hat{\alpha} = 1 \quad (2)$$

When running this regression we found that we could not reject the null hypothesis $\mathcal{H}_0$ and so we conclude that we are justified in using a model of the difference in GHG emissions to test the EKC hypothesis. The form our model takes is given below where $\Delta_{s,t} = GHG_{s,t} - GHG_{s,t-1}$.

$$\Delta_{s,t} = \beta_1 y_{s,t} + \beta_2 y_{s,t}^2 + \beta_4 p_{s,t} + \beta_5 h_{s,t} + \beta_6 c_{s,t} + \epsilon_{i,t} \quad (3)$$

In the second model which serves as a comparison for the EKC model, we simply remove the $y_{s,t}^2$ term and its coefficient from the first model.

After estimating each model, the estimated coefficients are used to find the point at which GHG emissions peak according to the EKC hypothesis. This is done by using the parameters β_1 and β_2 by checking that $\beta_1 > 0$ and $\beta_2 < 0$. If so, the estimated peak, or turning point as it is often called, is given by $\frac{-\beta_1}{2\beta_2}$. In instances in which the coefficients reject the EKC hypothesis, i.e., they do not fit the above form and thus monotonically increase, we denote the estimated turning point by —.

The forecasts performed using this data are validated with national level data from the EPA. To ensure that we have a reliable basis for comparing the performance of the forecasts using each model, we transform the EPA data to fit the national level WRI GHG emissions data using a standard univariate OLS regression approach. By using GHG emissions data from 2012 through 2017, this allows us to use mean squared error (MSE) as a metric for assessing which forecasting approach is superior.

5 Results

Since we are analyzing results from not only the energy sector as whole, but for six subsectors of this sector, we break this section down into seven subsections, one for the energy production sector and one for each of its six subsectors. In each subsection we will present the estimates from the regression models, provide the estimated turning point ($\frac{-\beta_1}{2\beta_2}$) representing an estimated peak emissions level, and discuss whether the results fit the EKC hypothesis.

Overall we do not find evidence for the EKC hypothesis for the energy sector, even when we restrict our analysis to the subsector level. This is evidence that any claims supporting the EKC are quite possibly spurious due to misspecified econometric models. Moreover, we find that the linear GDP model outperforms the EKC model in terms of forecast accuracy.

The tables for the results discussed in this section can be found in Appendix A. In those tables, the symbol * will denote that a variable is significant at the 10% level, ** will denote significance at the 5% level, and *** will denote significance at the 1% level.

We begin by examining the US energy production sector as a whole. The data set used for this study broke down the energy production sector into six disjoint subsectors: electric power, transportation, industrial, residential, commercial, and fugitive emissions. An initial survey of the estimates from the regression models, presented in Tables 4 and 5 in Appendix A, immediately dissuades us from making any conclusions in support of the EKC hypothesis. In none of the models using the EKC specification was the coefficient β_1 positive. This means that the EKC hypothesis was not supported at either the level of the energy production sector or for any of its subsectors. When comparing this against the linear GDP model by using the Akaike Information Criterion (AIC), we see no substantial difference between the two approaches. Considering that the coefficient β_1 behaves the sames under both model specifications and that the GDP squared term in the EKC model was only statistically significant in one instance (for the energy sector as a whole), the AIC statistics suggest that the two models are effectively the same, further eroding support for the EKC hypothesis.

Moving beyond the income terms, we see that the significant factors were heating degree days and cooling degree days. An increase in the number of heating or cooling degree days corresponds to an increase in emissions; however, cooling degree days were responsible for roughly triple the increase in emissions, suggesting that, at least in the United States, temperature increases could have a slightly net negative effect on GHG emissions.

Interestingly enough, these results suggest that GHG emissions from the energy production sector have likely peaked.

Moving on to the subsector level analyses, we see a nearly identical story. Heating degree days and cooling degree days are the most significant covariates included in this study, with both of these variables being statistically significant in the majority of models.

We do take the time to mention that some subsectors could benefit from perhaps a more targeted set of covariates. For example, transportation is relatively elastic in how it is consumed when compared to the other subsectors. For this reason, transportation is a common target of emissions mitigation policies, legislation, and technologies (Fox et al., 2017). Furthermore, emissions from transportation often occur in heavily populated areas. These emissions contribute directly to numerous negative health consequences (Kagawa, 2002), further incentivizing reductions in emissions from this subsector. Alternatively, results from fugitive emissions are due to faulty equipment, faulty processes, and corporate corruption, among other distinct factors. For this reason, results for fugitive emissions indicate that the factors which contribute to changes in emissions in other subsectors are not as impactful on fugitive emissions.

6 Forecasting GHG Emissions

Forecasts are a valuable tool for policy making, in part because they allow for us to attempt to answer the question of whether or not emissions have likely peaked. Using the results from our econometric models, we forecast GHG emissions from the US energy production sector using the forecasting method used by Selden and Song (1994). We then complement these forecasts using aggregated subsector emissions forecasts following the same methodology. This dual forecasting approach allows us to achieve several novel results. First, it allows for us to forecast GHG emissions at the subsector level. Additionally, this allows for state level forecasts of emissions at the subsector level. But most importantly, we can compare aggregated forecasted subsector emissions against forecasted emissions from the energy sector as a whole. By doing so, we can potentially improve upon forecasts of emissions from the energy production sector.

Since the EKC hypothesis asserts that GHG emissions are a function of economic growth, we can forecast GHG emissions using the estimated parameters from our regression models and forecasted economic growth. Because economic growth in the United States is relatively consistent, we can forecast GDP per capita on a state by state basis with the following model specified by Selden and Song (1994) in which $y_{i,t}$ represents GDP per capita in state s in year t, and $\theta_{s,t}$ denotes the error term.

$$\ln\left(\frac{y_{s,t}}{y_{s,t-1}}\right) = \gamma_0 + \gamma_1 \ln(y_{s,t-1}) + \gamma_2 \ln(y_{s,t-1}^2) + \theta_{s,t} \quad (4)$$

Because we used a series of covariates in our regression models, and because these covariates can be easily forecasted with reasonable accuracy, we include forecasts for each covariate in our forecasts using a linear trend on a state by state basis. We then use our forecasted GDP per capita and forecasted covariates along with the estimated parameters from each of our models to create a raw forecast of emissions for each model and (sub)sector combination over the period from 2012 through 2017, the time period for which we have additional national level GHG emissions data from the EPA.

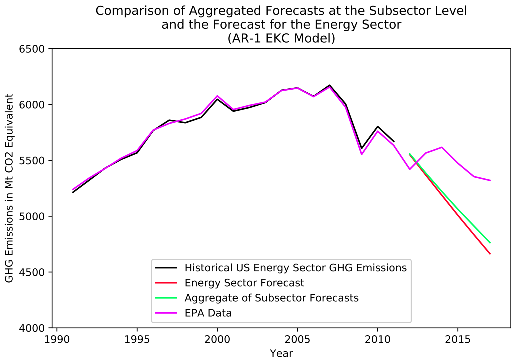
These raw emissions are then transformed using a Tobit functional form. This approach flattens forecasts, creating more gradual increases and decreases. Selden and Song (1994) preferred this method for asymptotic properties compared to the standard EKC and because it does not affect the turning points estimated by the EKC model. The Tobit transformation used here is given by the following equation in which $\hat{m}_{s,t}$ denotes the Tobit transformed forecasted emissions for state s in year t, $m_{s,t}$ denotes the raw forecasted emissions for state s in year t, Φ denotes the standard normal cumulative distribution function, ϕ denotes the standard normal probability distribution function, and σ denotes the standard deviation of the error terms in the original regression models from which the forecasts are based.

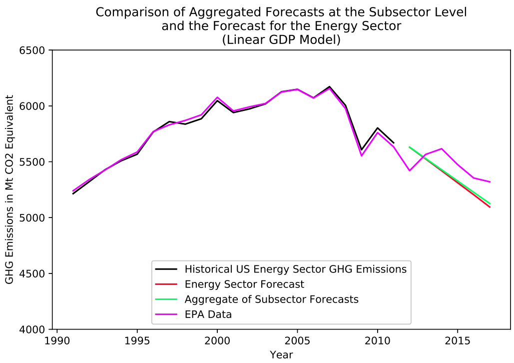
$$\hat{m}_{s,t} = \Phi\left(\frac{m_{s,t}}{\sigma}\right) * m_{s,t} + \sigma * \phi(m_{s,t}) \quad (5)$$

Once the Tobit transformations are complete, we aggregate the state level forecasts for each model into a single national level forecast of GHG emissions.

In order to validate our model and its forecasts, we use data from the United States Environmental Protection Agency (2019). The reason for using this data is two fold. First, it is an independent source which allows not only for model verification but also serves to validate our primary emissions data. Consistency between the EPA and WRI data sets confirms the quality of the data used in this paper. Second, the most recent year of data available in the WRI data set was from 2012. The EPA data set contained data for as recently as 2017. Since we want to determine whether or not US GHG emissions have likely peaked, we use the EPA data to confirm the result from our forecasts that US GHG emissions have indeed likely peaked. In order to directly compare the EPA data to our forecasts, we transformed the EPA data to fit the historical WRI data using a simple linear model. This allows us to account for any discrepancies in emissions accounting while ensuring that the year over year emissions levels are consistent.

The forecasted models are presented below and are grouped by model. For legibility purposes, the forecasts are separated into two plots. Before discussing the results, we again mention that both models predict decreasing GHG emissions going forward, supporting the conclusion that US GHG emissions may have reached their peak.





In order to justify this forecasting approach, we first present a comparison of the in-sample accuracy of the models for the energy sector as a whole versus the accuracy of aggregated subsector models. We find that, in accordance with the 'sum of the parts is more than the whole' approach espoused by Ferreira and Santa-Clara (2011), the aggregated subsector models are more accurate than the energy sector models. The following table contains the comparisons between the mean squared error (MSE) values for the independent energy sector models and the aggregated energy subsector models within the energy sector data. They show not only that the aggregated energy subsector models are more accurate, but also that the linear growth models are more accurate than the EKC models.

Table 3: Comparison of MSE Values

	Mean Squared Error (MSE)	
Model	Aggregate	Independent
EKC	1.47e5	1.93e5
Linear	6.07e3	3.07e4

Throughout all forecasts we see that the aggregated forecasts, which we have demonstrated to be more accurate within the data set, universally predict a slower decrease in future emissions than the forecasts using energy sector level data. This indicates that forecasts based on the EKC are relatively inaccurate; in this case we see that forecasts based on the standard EKC consistently forecast below aggregate forecast levels. Given that the aggregated subsector level models outperform the energy sector level models for in-sample accuracy, we conclude that the 'sum of the part is greater than the whole' concept applies to emissions estimation and forecasting at the sector and subsector level.

7 Conclusion

In this paper we tested the EKC hypothesis for GHG emissions from the US energy production sector and its subsectors using an econometrically rigorous variant of the autoregressive EKC regression model which models GHG emissions as a function of economic growth. We found no evidence for the EKC hypothesis at either the energy production sector level or at the subsector level. The behavior of the linear GDP term was the same (negative coefficients) for a linear GDP model, further supporting that the EKC hypothesis is not the best approach to modelling GHG emissions.

After testing the EKC hypothesis for the energy production sector of the US economy and its six subsectors, we forecasted future emissions for the energy production sector both independently and as the aggregate of forecasts of its subsectors. We found that the aggregated forecasts were sightly higher than the independent forecasts, demonstrating a lessened decrease in emissions during the forecast. This result was validated with in-sample testing of the accuracy of these two approaches. Both models suggest that US GHG emissions due to energy production have peaked.

To extend this research, we suggest that this approach of subsector level emissions based analysis and forecasting be applied elsewhere, beyond the United States. In-sample accuracy testing may not necessarily offer the same results, so any subsector level forecasting should pay special attention to how it is validated.

References

Achen CH (2000) Why lagged dependent variables can suppress the explanatory power of other independent variables. In: annual meeting of the political methodology section of the American political science association, UCLA, vol 20, pp 07–2000

Al-Mulali U, Saboori B, Ozturk I (2015) Investigating the environmental kuznets curve hypothesis in vietnam. Energy Policy 76:123–131

Aldy JE (2005) An environmental kuznets curve analysis of us state-level carbon dioxide emissions. The Journal of Environment & Development 14(1):48–72

Aldy JE (2006) Per capita carbon dioxide emissions: convergence or divergence? Environmental and Resource Economics 33(4):533–555

Aldy JE (2007) Divergence in state-level per capita carbon dioxide emissions. Land Economics 83(3):353–369

Apergis N, Christou C, Gupta R (2017) Are there environmental kuznets curves for us state-level co2 emissions? Renewable and Sustainable Energy Reviews 69:551–558

Auffhammer M, Carson RT (2008) Forecasting the path of china's co2 emissions using province-level information. Journal of Environmental Economics and Management 55(3):229–247

Auffhammer M, Steinhauser R (2012) Forecasting the path of us co2 emissions using state-level information. Review of Economics and Statistics 94(1):172–185

Azomahou T, Laisney F, Van PN (2006) Economic development and co2 emissions: a nonparametric panel approach. Journal of Public Economics 90(6-7):1347–1363

Bruvoll A, Medin H (2003) Factors behind the environmental kuznets curve. a decomposition of the changes in air pollution. Environmental and Resource Economics 24(1):27–48

Burnett J, Bergstrom JC (2010) Us state-level carbon dioxide emissions: a spatial-temporal econometric approach of the environmental kuznets curve. Tech. rep.

Burnett JW (2016) Club convergence and clustering of us energy-related co2 emissions. Resource and Energy Economics 46:62–84

Burnett JW, Bergstrom JC, Dorfman JH (2013a) A spatial panel data approach to estimating us state-level energy emissions. Energy Economics 40:396–404

Burnett JW, Bergstrom JC, Wetzstein ME (2013b) Carbon dioxide emissions and economic growth in the us. Journal of Policy Modeling 35(6):1014–1028

Canas A, Ferrao P, Conceicao P (2003) A new environmental kuznets curve? relationship between direct material input and income per capita: evidence from industrialised countries. Ecological Economics 46(2):217–229

Caron J, Fally T (2018) Per capita income, consumption patterns, and co2 emissions. Tech. rep., National Bureau of Economic Research

Chow GC, Li J (2014) Environmental kuznets curve: conclusive econometric evidence for co2. Pacific Economic Review 19(1):1–7

Criado CO, Grether JM (2011) Convergence in per capita co2 emissions: A robust distributional approach. Resource and Energy Economics 33(3):637–665

Criado CO, Valente S, Stengos T (2011) Growth and pollution convergence: Theory and evidence. Journal of Environmental Economics and Management 62(2):199–214

Dinda S (2004) Environmental kuznets curve hypothesis: a survey. Ecological Economics 49(4):431–455

Edelenbosch O, Kermeli K, Crijns-Graus W, Worrell E, Bibas R, Fais B, Fujimori S, Kyle P, Sano F, van Vuuren D (2017) Comparing projections of industrial energy demand and greenhouse gas emissions in long-term energy models. Energy 122:701–710

Fankhauser S, Tol RS (2005) On climate change and economic growth. Resource and Energy Economics 27(1):1–17

Ferreira MA, Santa-Clara P (2011) Forecasting stock market returns: The sum of the parts is more than the whole. Journal of Financial Economics 100(3):514–537

Fox J, Axsen J, Jaccard M (2017) Picking winners: Modelling the costs of technology-specific climate policy in the us passenger vehicle sector. Ecological Economics 137:133–147

Friedl B, Getzner M (2003) Determinants of co2 emissions in a small open economy. Ecological Economics 45(1):133–148

Galeotti M, Lanza A, Pauli F (2006) Reassessing the environmental kuznets curve for co2 emissions: A robustness exercise. Ecological Economics 57(1):152–163

Gill AR, Viswanathan KK, Hassan S (2018) The environmental kuznets curve (ekc) and the environmental problem of the day. Renewable and Sustainable Energy Reviews 81:1636–1642

Harbaugh WT, Levinson A, Wilson DM (2002) Reexamining the empirical evidence for an environmental kuznets curve. Review of Economics and Statistics 84(3):541–551

Hüttler W, Schandl H, Weisz H (1998) Are industrial economies on the path of dematerialization? material flow accounts for austria 1960-1996: indicators and international comparison. In: ConAccount workshop Ecologizing Societal Metabolism, p 23

Kagawa J (2002) Health effects of diesel exhaust emissions—a mixture of air pollutants of worldwide concern. Toxicology 181:349–353

Kampa M, Castanas E (2008) Human health effects of air pollution. Environmental Pollution 151(2):362–367

Konan DE, Chan HL (2010) Greenhouse gas emissions in hawai'i: Household and visitor expenditure analysis. Energy Economics 32(1):210–219

Kunnas J, Myllyntaus T (2007) The environmental kuznets curve hypothesis and air pollution in finland. Scandinavian Economic History Review 55(2):101–127

List JA, Gallet CA (1999) The environmental kuznets curve: does one size fit all? Ecological Economics 31(3):409–423

Maddison D (2006) Environmental kuznets curves: A spatial econometric approach. Journal of Environmental Economics and Management 51(2):218–230

Magnani E (2000) The environmental kuznets curve, environmental protection policy and income distribution. Ecological Economics 32(3):431–443

Marrero GA (2010) Greenhouse gases emissions, growth and the energy mix in europe. Energy Economics 32(6):1356–1363

Mazzanti M, Montini A, Zoboli R (2008) Environmental kuznets curves for air pollutant emissions in italy: evidence from environmental accounts (namea) panel data. Economic Systems Research 20(3):277–301

Millimet DL, List JA, Stengos T (2003) The environmental kuznets curve: Real progress or misspecified models? Review of Economics and Statistics 85(4):1038–1047

Ordás Criado C (2008) Temporal and spatial homogeneity in air pollutants panel ekc estimations. Environmental and Resource Economics 40(2):265–283

Ravishankara A, Daniel JS, Portmann RW (2009) Nitrous oxide (n2o): the dominant ozone-depleting substance emitted in the 21st century. Science 326(5949):123–125

Ringel M, Laidi R, Djenouri D (2019) Multiple benefits through smart home energy management solutions—a simulation-based case study of a single-family-house in algeria and germany. Energies 12(8):1537

Rupasingha A, Goetz SJ, Debertin DL, Pagoulatos A (2004) The environmental kuznets curve for us counties: A spatial econometric analysis with extensions. Papers in Regional Science 83(2):407–424

Selden TM, Song D (1994) Environmental quality and development: is there a kuznets curve for air pollution emissions? Journal of Environmental Economics and Management 27(2):147–162

Shittu WO, Musibau H, Hassan S (2018) Revisiting the environmental kuznets curve in malaysia: the interactive roles of deforestation and urbanisation. International Journal of Green Economics 12(3-4):272–293

Soytas U, Sari R (2009) Energy consumption, economic growth, and carbon emissions: challenges faced by an eu candidate member. Ecological Economics 68(6):1667–1675

Soytas U, Sari R, Ewing BT (2007) Energy consumption, income, and carbon emissions in the united states. Ecological Economics 62(3-4):482–489

Stern DI (2004) The rise and fall of the environmental kuznets curve. World Development 32(8):1419–1439

Stern DI (2017) The environmental kuznets curve after 25 years. Journal of Bioeconomics 19(1):7–28

Stern DI, Common MS (2001) Is there an environmental kuznets curve for sulfur? Journal of Environmental Economics and Management 41(2):162–178

Suri V, Chapman D (1998) Economic growth, trade and energy: implications for the environmental kuznets curve. Ecological Economics 25(2):195–208

Tol RS, Pacala SW, Socolow R (2006) Understanding long-term energy use and carbon dioxide emissions in the usa

United States Energy Information Administration (2019) State energy data system (seds): 1960-2017 (complete). Available online at: https://www.eia.gov/state/seds/seds-data-complete.php?sid=US

United States Environmental Protection Agency (2019) Inventory of u.s. greenhouse gas emissions and sinks. Available online at: https://www.epa.gov/ghgemissions/inventory-us-greenhouse-gas-emissions-and-sinks-1990-2017. Accessed on July 17, 2019

United States National Climate Data Center (2019) Heating and cooling degree data Available online at: `http://www.ncdc.noaa.org`. Accessed on August 14, 2019

West JJ, Smith SJ, Silva RA, Naik V, Zhang Y, Adelman Z, Fry MM, Anenberg S, Horowitz LW, Lamarque JF (2013) Co-benefits of mitigating global greenhouse gas emissions for future air quality and human health. Nature Climate Change 3(10):885

WRI CAIT Climate Data Explorer (2014) Climate analysis indicators tool: Wri's climate data explorer. Washington, DC: World Resources Institute Available online at: http://cait.wri.org. Accessed on April 2, 2019

Yang C, McCollum D, McCarthy R, Leighty W (2009) Meeting an 80% reduction in greenhouse gas emissions from transportation by 2050: A case study in california. Transportation Research Part D: Transport and Environment 14(3):147–156

Yates DN, Strzepek KM (1998) An assessment of integrated climate change impacts on the agricultural economy of egypt. Climatic Change 38(3):261–287

Zhou Y, Gurney KR (2011) Spatial relationships of sector-specific fossil fuel co2 emissions in the united states. Global Biogeochemical Cycles 25(3)

Zoundi Z (2017) Co2 emissions, renewable energy and the environmental kuznets curve, a panel cointegration approach. Renewable and Sustainable Energy Reviews 72:1067–1075

A Regression Tables

Table 4: Estimation Results for the EKC Model

Variable	Energy Sector	Electric Power Subsector	Transportation Subsector	Industrial Subsector	Residential Subsector	Commercial Subsector	Fugitive Emissions
GDP	-0.0001***	-6.827e-5**	-9.52e-6	-2.766e-5	-1.849e-5**	-6.701e-6	-9.557e-6
	(4.79e-5)	(2.74e-5)	(1.72e-5)	(2.12e-5)	(7.29e-6)	(4.61e-6)	(6.84e-6)
GDP^2	8.138e-10***	4.118e-10	-3.795e-11	1.685e-10	1.152e-10	4.212e-11	1.141e-10
	(4.64e-10)	(2.65e-10)	(1.67e-10)	(2.05e-10)	(7.05e-11)	(4.46e-11)	(6.62e-11)
Population Density	0.0016	0.0006	0.0010	-1.058e-5	0.0003	9.33e-5	-0.0005
	(0.003)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)
Energy Blend	0.1978	0.3004	0.0242	-0.0833	-0.1215	-0.0596	0.1377
	(0.734)	(0.419)	(0.264)	(0.324)	(0.112)	(0.071)	(0.105)
Heating Days	0.0006***	0.0002***	6.828e-5	0.0001	9.469e-5***	3.754e-5**	1.393e-5
	(0.000)	(9.43e-5)	(5.93e-5)	(7.29e-5)	(2.51e-5)	(1.59e-5)	(2.36e-5)
Cooling Days	0.0019***	0.0010***	0.0004**	0.0002	0.0002***	7.015e-5	3.103e-5
	(0.000)	(0.000)	(0.000)	(0.000)	(6.48e-5)	(4.1e-5)	(6.08e-5)
Adjusted R^2	0.032	0.035	0.022	0.003	0.013	0.001	0.002
AIC	6335	5206	4272	4688	2537	1615	2409
Turning Point	—	—	—	—	—	—	—

Table 5: Estimation Results for the Linear GDP Model

Variable	Energy Sector	Electric Power Subsector	Transportation Subsector	Industrial Subsector	Residential Subsector	Commercial Subsector	Fugitive Emissions
GDP	-5.949e-5***	-2.743e-5***	-1.328e-5***	-1.094e-5*	-7.073e-6***	-2.523e-6*	1.763e-6
	(1.35e-5)	(7.74e-5)	(4.86e-6)	(5.98e-6)	(2.06e-6)	(1.3e-6)	(1.93e-6)
Population Density	0.0011	0.0004	0.0011	-0.0001	0.0002	6.929e-5	-0.0005
	(0.003)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)
Energy Blend	-0.0054	0.1976	0.0336	-0.1254	-0.1503	-0.0701	0.1092
	(0.726)	(0.414)	(0.260)	(0.320)	(0.110)	(0.000)	(0.104)
Heating Days	0.0004***	0.0001**	7.766e-5	8.48e-5*	6.623e-5***	2.713e-5**	-1.428e-5
	(0.000)	(6.79e-5)	(4.27e-5)	(5.25e-5)	(1.81e-5)	(1.14e-5)	(1.7e-5)
Cooling Days	0.0014***	0.0008***	0.0004***	9.596e-5	0.0001***	4.268e-5	-4.34e-5
	(0.000)	(0.000)	(0.000)	(0.000)	(4.56e-5)	(2.89e-5)	(4.29e-5)
Adjusted R^2	0.030	0.034	0.023	0.004	0.012	0.001	0.000
AIC	6336	5207	4270	4686	2537	1614	2410

B Code and Data Source Availability

The python script and csv files containing data from WRI CAIT Climate Data Explorer (2014) are available online at:

`https://github.com/cat-astrophic/energy_subsector_emissions`

Please cite the data source if using the data used in this paper, but feel free to use, edit, and redistribute any scripts related to this project freely!