Convergence of U.S. State Greenhouse Gas Emissions and Galton's Fallacy

by

Soheil Shaeri
MA Economics, University of Tehran, 2016

An Extended Essay Submitted in Partial Fulfillment
of the Requirements for the Degree of

MASTER OF ARTS
in the Department of Economics

We accept this extended essay as conforming
to the required standard

Ken Stewart

Dr. Kenneth Stewart, Supervisor (Department of Economics)

Dr. Emma Hutchinson, Member (Department of Economics)

University of Victoria

Abstract

In this study, I examine the convergence of state-level emissions data from the United States. I begin by replicating the green Solow model established by Brock and Taylor (hereafter, BT) in 2010. I then employ a second test of convergence, namely, an analysis based on coefficients of variation, first proposed by Hotelling and codified by Friedman (1992) and Quah (1993). Unlike SO2, which has a clear convergent pattern across the states, my findings for CO2 and CO do not reflect convergence paths but rather diverging patterns. The coefficient of variation analysis of the state emissions data contradicts BT's interpretation of the regression coefficients, suggesting that interpreting the regression results as a sign of convergence could lead to Galton's fallacy.

Table of Contents

1. Introduction

As climate policies are at the center of the attention of scientists, activists, and citizens worldwide, it is crucial to predict emission trends. One of the major factors in climate treaties is the prospect of convergence of polluting emissions across countries—or states in the case of my study. Convergence occurs when states with high initial per capita emissions have a lower emission growth rate than those with low initial per capita emissions. Having some evidence of convergence is important as it provides a rationale for developing states to support climate policies which often focus on limiting emissions. Otherwise, in the case of divergence, developed states continue to have higher emission levels, which makes it difficult to convince the growing states to join climate pledges and limit their emissions. I used U.S. state data for this study, which makes it an interesting topic for U.S. climate policies regarding emissions.

In this study, I analyze whether state-level emissions trajectories in the United States are converging, focusing on CO2, CO, and SO2 emissions. I use two different approaches to test for convergence. The first one is based on the green Solow model which was proposed and developed by Brock and Taylor (2010). BT regarded the regression coefficient as a reliable inference for testing the possible convergence (they even referred to regression equations as "convergence equations"). However, the emissions data may manifest Galton's fallacy[1] (*i.e.,* an incorrect interpretation of the "regression toward the mean" phenomenon) as some relevant literature has suggested, and if that is the case, the presence of Galton's fallacy would put such interpretations into question. Therefore, for my second approach, I test the convergence by analyzing the coefficient of variation approach, which is a more reliable tool in this context, according to relevant research by Friedman (1992) and Quah (1993). In Section 4, I explain Galton's fallacy in detail and show how it applies to this topic. I focus on both approaches while hinging my trend analysis on the results from the coefficient of variation rather than the potentially misleading sign of the regression coefficients.

A wide portion of the relevant literature has focused on interpreting the sign of the coefficient of regression equations and has primarily followed the framework provided by Brock and Taylor, using different datasets and some modified estimations. In contrast, considering the probable fallacy mentioned above, the coefficients of variation and their corresponding results are presented as an alternative method for convergence prediction. In addition to carbon dioxide, I also included other emissions, such as CO and SO2, in this study. Subsequently, I check whether there is any discrepancy in the long-run patterns.

[1] We have described this phenomenon in detail in the section 4 of this essay.

The Green Solow model was derived from Solow's (1956) model. The model was first proposed by BT in 2005, and since then, other researchers have frequently used it in the green growth literature to predict emission trends. I evaluate the coefficient of variation (CV) to cross-examine the convergence pattern derived from BT regression estimation. I obtain the dataset by collecting and combining several regional and national datasets. Namely, CO2, CO and SO2 emissions, saving rates, and population growth rates across states. The main sources of data are the U.S. Census Bureau, Bureau of Economic Analysis, U.S. Energy Administration and Environmental Protection Agency (EPA).

My results show that the long (*conditional*) specification does not have an edge over the short (*unconditional*) specification, mainly because it does not offer a higher capacity than the short estimation to explain the variations in the dependent variable. Therefore, based on my estimation, it looks like savings rates and population growth rates do not play important roles. This can be anticipated, as there doesn't seem to be a direct link between these conditional variables and the average annual emissions growth rate, which contradicts the findings of the green Solow model. According to my findings, in the case of CO2 and CO, there are some indications of moderate divergence rather than expected convergence across states, which differs from BT's findings. This is due to the fact that their CV graphs are increasing, revealing divergent patterns over the last three decades across states. However, the results of the SO2 estimations and the declining trend of CV over time suggest a possible convergent pattern for these emissions.

2. Literature Review

My paper draws on four separate but related strands of the literature. First, studies based on Environmental Kuznets Curve (hereafter, EKC) tried to analyze the link between development and environmental degradation. Many of these studies have suggested an inverted U-shaped relationship between environmental degradation and GDP (and/or other similar variables). For instance, Grossman and Krueger (1990) suggested that when economies are developing there is a subsequent period of environmental improvements after experiencing the expected deterioration of environmental indicators. Also, it is believed that poor or underdeveloped countries do not hurt the environment to a large degree because they do not have a high capacity for production Mäler (2013). However, once economic expansion and industrialization begin, they tend to damage the environment and exacerbate environmental indicators more severely. After reaching their peak, these countries have sufficient resources to invest in the environment with the aim of preserving it. Similarly, Dinda (2004) suggested that at this stage there is a tendency for the general public to care about having higher environmental quality. Thus, the country's emissions would drop by having a higher level of development—a higher income per capita.

The second category of studies is those with their roots in macro, which examines whether key economic variables, including emissions, tend to converge over time. One of the essential studies is Mankiw, Romer, and Weil (1992) (hereafter, MRW). They have examined the implication of the Solow model in the context of convergence of per capita income. They concluded that per capita incomes converge across countries if we control for macroeconomic variables such as saving rate, population growth rate, and rate of technological progress. That means poor countries tend to grow faster on average than rich countries.

This literature has been extended to test for convergence in environmental indicators, including emissions. The question of convergence is important because if we are diverging and we have transfers from rich countries to poor countries, those transfers would be increasing over time––because of the increasing gap. As you may guess, in such a situation, it is hard to reach a climate agreement. Conversely, if we have convergence, we may have a better chance of achieving a climate treaty. One of the studies that discussed the abovementioned issue in detail is Aldy (2006), which suggests that the allocation of carbon emissions across countries is a key element in the current state of international climate agreements. They believe in such climate treaties which aim to decrease emissions over time, the allocation of carbon emission should be based on the per capita approach than the income of the countries—which is favourable for developed countries. Accordingly, there should be some compensation from wealthy countries with high levels of per capita emissions to undeveloped countries.

Delgado (2013) points out that both parties would engage in climate change treaties only if per capita emissions were converging; otherwise, such climate policies would not be feasible since developed countries have to compensate at higher rates each year.

As I discuss the background literature, I should define both absolute and conditional convergence. Absolute convergence is where countries with lower initial per capita levels of emissions would experience a faster growth rate of per capita emissions relative to countries with a higher initial level of per capita emissions. Thus, they will reach the same steady state per capita emissions Delgado (2013). However, Conditional convergence is where the convergence happens only if we control for key variables like population growth and investment rates (this is analogous to the MRW approach). The same has started to be interpreted in the environmental economics literature as well.

Strazicich and List (2003) tested the expected convergence of CO2 emissions among industrialized countries for the period 1960-1997. First, they examined cross-sectional regressions conditional on income, gasoline prices, population density, and average temperature. Then, in the time-series framework, they conducted a panel unit root test developed by Im, Pesaran, and Shin (2002) to prove the expected conditional convergence pattern in per capita CO2 emissions. The advantage of their study, compared to others, such as BT (2010), is the use of two different methods for testing their main hypothesis of convergence in CO_2 emissions. That being said, future investigation into other pollutants rather than CO_2 seems to be necessary.

One of the relevant papers in this regard by Van (2005) concludes that there is a convergence pattern among industrial countries—with similar conditions—but not much evidence was found for the whole sample[2]. One significant finding of this study is that countries with low initial levels of per capita emissions tend to stay at roughly the same rate, whereas countries with high initial levels of per capita emissions tend to decline. This can be interpreted as typical steady-states, where it would fall into the gap between the countries with high and low levels of per capita emissions. This is also something that both sides disagree on because it can be used as a benchmark for policy proposals to set some goals at the level of per capita emissions that is feasible. The author pointed out that income convergence may not be the only factor and introducing other pollutants with a longer data duration would be beneficial.

The third and most relevant category of studies is the one focusing on the Green Solow model. This topic was introduced and developed by Brock and Taylor (2010) and it was followed in other studies such as Delgado (2013), Hao and Wei (2015) and Chen (2017).

Based on the forces introduced by the Solow model (1956), Brock and Taylor created a new framework known as the "Green Solow Model". It reveals a major connection between the EKC and the Solow model in macroeconomic terms. In addition, based on their conclusion, they found a significant result for the convergence of per capita carbon emissions across 165 countries throughout 1960–1998. They tested their model for both absolute and conditional convergence (defined as the regression equations), in which the conditional specification was based on variables, including the saving rate and population growth rate.

The model did not consider the possibility of heterogeneity across countries. That said, potential heterogeneity may lead to inconsistent estimations with a simple ordinary regression model (OLS). Some, like Delgado (2013), argue that due to the abovementioned problem, it is better to integrate heterogeneity by letting the model's parameters differ from country to country. He claims that this generalization uses a less restrictive approach in structural theory, even though the convergence result appears to be identical to that of the standard green Solow model. The other point, answered by BT (2010), is to utilize a panel data approach instead of a cross-sectional approach to deal with unobserved heterogeneity. They concluded that this approach would also have some disadvantages, which have been shown in the prior literature (e.g., Durlauf and Quah 1999). In this regard, some scholars (e.g., Azomahou et al., 2006; Nguyen-Van, 2010) have moved

[2] The paper used a sample of 100 countries during 1966-1996.

forward and used advanced specifications or methodologies such as nonparametric and semiparametric panel data (see Hao and Wei, 2015).

Both BT (2010) and Hao and Wei (2015) argue about the shortcomings of the panel data method, claiming that using sophisticated models does not solve these problems. Thus, I used the original Green Solow model first proposed by Brock and Taylor (2010). Besides, Delgado (2013) suggests that "BT drives the convergence equations from a solid theoretical foundation with roots in both macroeconomics and environmental economics."

Finally, we have the CV literature which serves as a critique of the mainstream approach used in the Green Solow Model literature. They believe the typical regression-based approach cannot distinguish between the "true" convergence and a simple reversion to the mean-type convergence. The essential examples of this view are Friedman (1992) and Quah (1993) that reject the notion of determining emission patterns using the sign of the regression coefficient. They suggested that using the coefficient's sign as a test of convergence could be misleading because of the common fallacy of the regression toward the mean. Instead, they recommended using coefficients of variation as "the real test of a tendency to convergence," first introduced by Hotelling.

Based on Friedman's (1992) elaboration, I use coefficients of variation (CV) instead of regression coefficients to determine the possible pattern of convergence. Thus, I avoid Galton's fallacy as a result of this. One of the contributions of this study is the implementation of the Green Solow model across all 50 states in the U.S. I use a new and different dataset, with probably different characteristics, compared to the data used by BT (2010). Moreover, collecting and conducting analyses on additional sets of pollutant data, such as carbon monoxide (CO) and sulfur dioxide (SO2), provides a more complete picture of how greenhouse gases change in different states. While I use both conditional and unconditional specifications in my regression analysis, as MRW (1992) and BT (2010) do, I do not interpret the results to determine the convergence patterns—which is the mainstream approach in some background studies.

3. Data and Descriptive Evidence

In this section, I explore two different datasets: national and state. The main inferential analysis and regression equations were conducted based on the state dataset. The coverage of national data is beneficial and relevant for several reasons, such as providing general background information, understanding the national trends of emissions across the U.S., and possible asymmetrical trends for the whole country vs. individual states. An analysis of these trends could be practical for future studies.

3.1. National Data

National data is obtained from U.S. Energy and Information Administration (EIA) and U.S. Environmental Protection Agency (EPA)[3]. Figure 1 illustrates the general trends related to US CO2 emissions for the period (1990-2019) for carbon dioxide, carbon monoxide, and Sulphur dioxide respectively. In panel (a) of Figure 1, the plot shows the trends of several different emission variables using 1990 as a reference year. The first thing to point out here is that after the 2007 recovery, the emission level path seems to have changed because of the use of less polluting energy resources such as natural gas and renewable energy (EIA Report, 2019). The population trend in the U.S. is upward with a slightly declining growth rate. Although the U.S. economy is growing, both in aggregate (measured by population) and per capita (measured by GDP/capita), the "energy intensity" (measured by energy/GDP) ratio has decreased over this period which is suggesting a less energy-dependent economy for the U.S. This ratio demonstrates the type of economy—post-industrialized in this case—and its efficiency. Total CO2 emissions have declined since around 2005, and the "CO2/intensity" of energy (measured by CO2/energy) has remained fairly stable or has declined since 2005. Hence, not only are we becoming more efficient in our energy use, but if anything, we are becoming even more so regarding the CO2 intensiveness of production.

Panel (b). provides some insights into the shares of the different sectors that are releasing carbon emissions. This suggests that residential-related emissions increased until approximately 2004, plateaued for some years, and decreased afterward. This could be the result of using more efficient ways of providing residential energy. This graph shows that relative to the industrial, commercial, and residential sectors, transportation has been a growing source of CO2 emissions.

Panel (c) is a representation of carbon emission by the type of fuel used in the process which leads to emission. Total CO2 emissions decreased after the 2007 recession. Emissions from petroleum and liquid usage were fairly constant over time, and emissions from natural gas use expanded after the recession occurred. Another interesting trend is coal consumption emissions. The associated CO2 emissions have decreased over the last 15 years.

Figure 2. represents the long-run trends of CO emissions in the U.S. from 1990 to 2019 analogous to the trends shown in Figure 1. The datasets used for CO and SO2 in Figures 2 and 3 are different from those used for CO2 in Figure 1. As a result, it has different categories, which led to slightly different graphs compared to CO2; these categories were defined by the EPA in the relevant annual reports. Panel (a) is similar to the first panel in Figure 1. One of the differences that should be pointed out is that CO emissions have a faster diminishing trend at the national

[3] The data source is provided in detail in the "Data Appendix" section at the end of this paper.

level than CO2 emissions. The same comparison holds for the CO/GDP ratio. This offers a more promising pace for CO reduction than CO2 reduction. The second and third panels categorize CO emissions by sector and their respective shares over the study (1990-2019). It should be noted that these graphs are based on different datasets and categories defined by the EPA. Based on the graphs, the total emissions of carbon monoxide were reduced over the study period. The transportation sector has the largest role in releasing CO, and it decreases with a pattern similar to the total emissions. On the other hand, miscellaneous emissions increased over the period and had the same share as transportation after 2018. Finally, industrial and other processes and stationary fuel combustion have a small stable share of emissions.

Figure 3. shows the long-run trend of SO2 in the U.S. over the same period. The general pattern in panel (a) is similar to that of the analogous panel in Figure 2. Needless to say, there is a key difference: the downward patterns of SO2-related emissions are steeper and eventually heading toward zero in the latest year of study. This shows a more significant reduction in SO2 emissions compared to the other two pollutants (especially by 2006, we are experiencing a sudden but steady decrease in SO2 emission levels, which could be a result of new environmental policies or standards by government agencies). Panel (b) shows the same concept as the analogous panel in the previous graphs. The total amount has increased rapidly over the last three decades. Based on the last panel, the major source of SO2 pollution is stationary fuel combustion. The transportation sector had lower contributions after 2008. The amount of emissions from industry-related sectors is also quite minor compared with the total level of SO2 emissions.

I now have a general perception of the national trends for all three emissions. This is useful because I can first compare these three emissions with each other and be aware of their differences—the general behaviour of the series. This is crucial to my judgment and conclusions regarding these emission trends. For instance, the level of national CO2 emissions is fairly stable in comparison with CO and SO2, which experienced significant drops in the study. Now, I move the focus to state-wide datasets that have been used for inference and regression equations.

3.2. State Data

State data is obtained from U.S. Energy and Information Administration (EIA) and U.S. Environmental Protection Agency (EPA)[4]. I obtained state per capita CO2, CO and SO2 emissions from the EPA. I used CO and SO2 emissions in levels for each state, and then used the population series to calculate the emissions per capita, as needed.

Table 1. contains the summary statistics of the state dataset. The first column represents the summary statistics for the full sample, which includes all 50 states plus D.C. for the period (1990-

[4] The data source is provided in detail in the "Data Appendix" section at the end of this paper.

2018). Column (2) contains the summary statistics for all 50 states from 1996 to 2018, where all emissions are available.

The shorter span of the dataset in column (2) is due to the lack of earlier data for CO and SO2 emissions as well as the lack of data on the state-saving rate, which has been added to the analysis in the next section. Based on the dataset, there is a significant gap in the initial emissions among the five highest pollutant states and the next highest pollutant states in my sample. Any of these five states with the highest initial emissions levels have at least twice the emissions as the next pollutant state in the ranking. Accordingly, for the subsample presented in column (3), I exclude the states with the highest pollution–Wyoming, North Dakota, Alaska, West Virginia, and Louisiana–in the case of CO2. It should be mentioned that this subsampling method aligns with BT's 2010 paper.

Table 1 indicates that the subsample of the states in column (3), on average, has almost equal growth rates of CO2, CO, and SO2 emissions, but a lower level of initial CO2 emissions compared to the full sample of 50 states. Nevertheless, this means they have comparable values across samples and are not significantly different since the mean values are within two standard deviations of each other across samples Delgado (2013).

Figure 4. provides the dispersion of CO2, CO, and SO2 emissions across states. These panels show the disparity in emission levels across the states for each type of pollutant. In panel (a), I have fairly constant averages for the U.S., with a slight decrease over the last decade. This suggests that the magnitude of changes in emissions for states with downward trends prevails over that for states with an upward trend. In addition, the graphs suggest a more significant decline in CO and SO2 emissions for the U.S. across the states relative to CO2 (thus, in general, the emission reduction efforts for CO and SO2 were relatively more successful). In addition, one distinctive point regarding SO2 is the shrinking distribution of data points over the years (i.e., states have less dispersion in SO2 emissions over time).

3.3. Are State Emissions Converging? Descriptive Evidence

The Coefficient of variation (CV) shows the degree of cross-sectional dispersion of the data points compared to the mean of the population. For example, a lower CV indicates less dispersion relative to the mean. One benefit of the CV is that it enables us to compare different series with different units of measurement because it is a ratio of units; thus, it is unitless.

As shown in Figure 5, none of the CVs exhibited strong trends. For instance, in the case of CO2 and CO, panels (a) and (b) show a modest increase in CVs for per capita emissions over decades. This suggests increasing differences (i.e., a likely diverging pattern) among the states.

On the other hand, Panel (c) shows a declining trend of CV for SO2 per capita emissions, which suggests diminishing dispersion (or a converging pattern) among states in the case of SO2.

Figure 6 is analogous to Figure 5, but for a subset of the 46 least-polluting states. By excluding the five "outlier" states with the highest initial emissions per capita, I test the sensitivity of the patterns in Figure 5 to this omission. The basic patterns in Figure 5 are insensitive to the excluded states, but CO and SO2 show stronger evidence in this case. In other words, there were stronger indications of divergence in CO emissions per capita and stronger evidence of convergence in SO2 emissions per capita.

In conclusion, the nationwide level of per capita CO2 emissions has been fairly stable with a modest decrease over the last decade. For CO and SO2, a substantial decrease over time was observed. State-wide evidence suggests that the emission patterns for CO2 and CO are modestly diverging, and the omission of the most polluting states suggests a stronger divergence path in the case of CO. However, SO2 emissions converge, and the omission of the most polluting states creates stronger evidence of this. These CV graphs are a proper way to utilize descriptive evidence based on the panel datasets that I have composed. Even though these plots are not the usual inference tests, Friedman (1992) and other authors have called such interpretation of CVs the "real test of convergence."

4. Galton's Fallacy and Regression Equations

In this section, I present the concept of "regression fallacy" and I explain one famous example of this phenomenon, introduced by Galton. Then, I discuss its implications for my study, if any. I continue by introducing the regression model, which is a replication of BT's (2010) model and provide estimates from the regression equations.

Galton found that sons were typically shorter than their fathers when he regressed the heights of tall dads on the heights of their sons. Initially, this was puzzling because, in the real world, there were considerable counter-examples—notably tall sons from fathers with average heights, according to Quah (1993). However, this misunderstanding arose from the phenomena of regression to mediocrity, often known as "regression toward the mean." To put it simply, this concept indicates that the extreme random variable (*the son's height* in his study) tends to be closer to the average value in the following period [5]. The "regression fallacy" refers to the idea of wrongly attributing such a trend to an external cause, such as a policy that has been implemented.

---

[5]. Similarly, complimenting someone after a very good performance does not result in decreased future performances—he or she simply returns to their ordinary level of performance (Daniel Kahneman, 2002 Nobel prize winner in economic sciences). Another well-known example in psychology is the "second-year symptom," in which students who perform well in the first year tend to perform worse in the second year.

Similarly, Friedman and Quah argue that the sign of the regression coefficient of the initial-year data (*initial emissions* in my case) has nothing to do with the probable convergence pattern that I am looking for as the major question in this essay. Friedman believes that naively determining the long-run trends using the sign of the regression coefficients is another example of the aforementioned fallacy. He proposes using coefficients of variation to assess the long-run trends (*the big picture*) that I am looking for. This would help to avoid the regression fallacy by accounting for variable averages, and this approach utilizes the cross-sectional nature of the dataset, unlike the convergence regressions of BT. To determine if the Galton fallacy is present in my study, I compare the results of the CV analysis from section 3.3 with the graphical representation of the regression equations.

My regression model replicates the "Green Solow Model" proposed by Brock and Taylor (2010). As the name suggests, this foundation is analogous to the Solow Model (1956). Parallel to BT's (2010) work, I have two key empirical predictions of emission trends. Using the Green Solow Model, I expect to see an absolute convergence of emissions in the set of states with different initial levels of emissions but similar Solow forces, such as saving rates and the rate of technological progress. To achieve conditional convergence of emissions, I need to condition my regressions on the characteristics of the states that I am analyzing.

The short specification (*unconditional regression*) of the regression equation is a linear regression with the error term μ_{it}.

$$\frac{1}{N} ln\left(\frac{e_{it}^{c}}{e_{it-N}^{c}}\right) = \beta_0 + \beta_1 ln\left[e_{it-N}^{c}\right] + \mu_{it}. \tag{1}$$

And the long specification (*conditional regression*) of the regression equation is

$$\frac{1}{N} ln\left(\frac{e_{it}^{c}}{e_{it-N}^{c}}\right) = \beta_0 + \beta_1 ln\left[e_{it-N}^{c}\right] + \beta_2 ln\left[s_i\right] + \beta_3 ln\left[(n + g_B + \delta)_i\right] + \mu_{it}. \tag{2}$$

The dependent variable is the average annual growth rate of emissions per capita across the study period. $\left[e_{it-N}^{c}\right]$ refers to the per capita emission in the first year. The other variables are recognizable from the Solow Model (1956), where (s_i) is the average saving to GDP ratio for the period of study, and $(n + g_B + \delta)_i$ is the summation of the average population growth rate, rate of technological change, and depreciation over the years. It is worth mentioning that, for simplicity, I would consider the abovementioned summation term equal to $(n+0.05)$[6].

[6] I assume their values are constant because g_B and δ are not directly observable. MRW (1992), Brock & Taylor (2010) have a similar approach.

I obtained regional statistics from the Bureau of Economic Analysis, including the state disposable income, GDP, and private consumption expenditure (PCE). Using these three variables, I determine the state-saving rate. The saving rate is calculated by subtracting consumption expenditure from disposable income expressed as a percentage of GDP[7]. To make the conditional regression symmetrical, I used the average population growth rate series (n) for the same period. I prefer the long specification because it relaxes the requirement that all the states have the same steady state.

Ordinary least squares (OLS) was used to estimate the possible convergence in the dataset. This result is analogous to the BT estimation model. In addition, they concluded that heterogeneity is not an issue in this case study (BT 2010, page 144). Their argument indeed holds here as well, even with greater possibility, because I expect less dispersion within states than among countries in their dataset.

Figure 7 illustrates the relationship reflected by the unconditional regressions between the initial emissions and the subsequent growth rate of emissions. The graphs in panel (a) reflect the entire sample of 51 states, whereas panel (b) depicts their counterparts for the subsample of 46 states, from which the states with the five highest initial levels are excluded. Some might mistakenly assume that an inverse relationship between the initial value of a variable (*i.e., emissions*) and its subsequent growth rate implies that the variable converges over time (see SO2 trends in Figure 7). In contrast, a direct relationship could be an indication that the variable is diverging, such as CO in Figure 7. These interpretations are examples of Galton's fallacy in the context of this study. Friedman (1992) highlighted that this interpretation is false. Quah (1993) demonstrated that there is no fundamental link between the signs of these slope coefficients (in unconditional or conditional regression) and the existence or lack of convergence. Equivalently, true convergence or lack thereof (as manifested in a declining or increasing CV) is compatible with the slope coefficients of any sign for either conditional or unconditional regressions. As an example, consider CO2 for a subset of 46 states. It is clearly diverging in Figure 6, yet has an inverse relationship in Figure 7—indicating a convergence trend. This contradictory conclusion reinforces Friedman and Quah's viewpoint that the sign of the slope coefficient is unreliable, and that further inference from this sign leads to Galton's fallacy. Thus, as I follow the relevant literature, specifically BT (2010), I should emphasize that although their estimation equations (*i.e.,* equations 1 and 2) are mathematically correct, they cannot be interpreted as a sign of convergence.

[7] I reduced the length of my conditional estimations based on the fact that the earliest data available on consumption was for 1997. Also, 2018 is the last year for which the data on carbon dioxide emissions has been provided.

Some might question the necessity of conditional specification using this dataset. They may assume that all states seem to have the same saving and population growth rates. If this is the case, why should I be concerned about conditioning the key explanatory variables? Figure 8 attempts to address this matter. Based on this, I conclude that the two “conditioning” variables of saving, and population growth rates vary significantly across states. Hence, U.S. state data provide a reasonable basis for estimating the conditional version of the regression. The disparity among states demonstrated in Figure 8 is not surprising given the rate at which the population is growing in some states. For example, California has undergone considerable expansion over the years. However, certain states in the Midwest do not have similar growth, if any. This is another example of why we may need to pay more attention to the conditional specification and assess the results.

Table 2 summarizes my estimation results, with the short specification for the full sample and subsample of the 46 states presented in columns 1 and 3, respectively. Similarly, Columns 2 and 4 are associated with the long specifications. The short specification (*i.e.,* unconditional regression) for both the full sample and subsample ranges from 1990 to 2018. Similarly, the long specification (*i.e.,* conditional regression) covers the period 1997-2018, with 1997 being the earliest year for which data on the explanatory variables are available. Thus, the conditional regression spans over 21 years, from 1997 to 2018, indicating that the population and saving rate were treated symmetrically.

The convergence coefficient associated with the initial level of emissions, log(e0), is the main variable of interest in this table. The coefficient for the initial carbon emission per capita is positive in the case of the unconditional regression in column (1) and negative in the other three regressions, although none of them are statistically significant. This contradicts the BT results, which show that countries with low emissions have higher growth rates of emissions, which means that convergence is anticipated across the different models. They reported that in both conditional and unconditional regressions, the estimated coefficients are negative and statistically significant. In addition, the estimated results in Table 1 are inconsistent with the results obtained by the CV analysis (the true test of convergence). The CV analysis shows a divergence pattern in both models (*i.e.,* positive trends). However, because of the insignificant results, the regression findings fall short of concluding such a trend. Interpreting the possibility of convergence by relying on so-called “convergence coefficients,” obtained from the regression results, could be misleading. The concept of the potential Galton’s fallacy was elaborated in detail by Friedman (1992) and Quah (1993). Thus, they advocated such interpretations regarding the convergence of trends, which need to be assessed using CV analysis.

Having a higher $\bar{R}^2$ (Adjusted R^2) for the conditional specifications versus the unconditional ones reveals whether adding the conditional variables to my specification is practically beneficial or not. In the case of both the conditional and unconditional specifications, the higher $\bar{R}^2$ suggests that adding the conditional variables helps to improve the explanatory power of the regression model for the $CO2$ emissions.

Additionally, in parallel to BT's analysis of the regression coefficients, I calculate the implied share of capital using the β coefficients in Table 2. To do so, based on equation (24) of BT, the ratio of beta coefficients $\{\beta_2/\beta_1\}$ is equal to $\{\alpha/(1-\alpha)\}$, where α is the capital's factor share. Therefore, using the estimated coefficients in Column 4, for example, the implied share of capital would be equal to 0.95, which is too high. The estimated implied share of capital is higher than that of BT, which is too large at around 0.7. Moreover, based on the coefficients that I found, I can determine how the rising rates of savings and population growth in emerging states impact the growth rate of their emissions. For instance, based on Column 2 of Table 2, if the state-saving rate increases by 1 %, this is associated with a 1.45% increase in the annual growth rate of emissions. Similarly, if the population growth rate increases by 1%, it implies a slightly less than one percent decrease in the state's annual growth rate of emissions per capita.

Similar to the estimation results for carbon dioxide, Table 3 displays the regression results for carbon monoxide. First, in both the conditional and unconditional regressions, the estimated coefficient for the initial level of emissions is positive and statistically significant. Based on BT's interpretation of the data, I should expect a divergence pattern for CO. Although BT's analysis is only limited to CO2 emissions—I do not have their CO conclusion with which to compare it—the CV analysis of CO is consistent with the regression analysis's finding that the emissions follow a divergent pattern.

Second, based on the $\bar{R}^2$ for the full sample, contrary to the evidence reported for CO2, I have a lower level of explanatory power after adding the conditional variables. This is consistent with the low coefficient t-ratios, and it can be interpreted that the conditional variables do not contribute to the explanatory power of the model, at least based on the result for the full sample. The low-level t-ratios (indicating statistical insignificance) for the estimated coefficients suggest that conditional variables seems unimportant and adds little or nothing to the low explanatory power of the regressions.

Third, based on the outcome shown in Column 4, the implied share of capital is 0.87. Furthermore, if the saving rate increases by 1%, the CO emissions growth rate increases by 2.3% (this direct correlation is consistent with the BT findings). The annual growth rate of CO emissions increased by approximately 0.1 percent for every 1 percent increase in the population growth rate.

Table 4 displays the sulphur dioxide regression results. In all versions of the regression model, the coefficient of the initial amount of SO2 emissions is negative but not statistically significant. The corresponding CV results, however, demonstrate that the emissions are gradually converging.

As indicated above, BT's study only looks at CO2 data, so I don't have a comparable set of results.

By including the conditional variables in the specification, the proportion of variation in the emission growth rate described by the explanatory variables does not increase (although it decreases in the case of the subsample). Again, this reinforces the notion that conditional variables do not play a significant role, based on the overall results. Similar to the interpretation of the other emissions results, the implied share of capital is roughly 0.9 when the coefficients in column 4 are taken into account, which is too high.

5. Conclusions

In this study, I aim to check the findings of the green Solow model, suggested and developed by Brock and Taylor (2010), using U.S. state data. I present my estimation regarding the possibility of convergence based on state-wide pollutant emissions of CO2, CO, and SO2. The green Solow model was developed based on a strong theoretical foundation. I considered the counterargument by Friedman and Quah regarding the possible Galton's fallacy, so I evaluated the coefficients of variation for all emissions to test for this phenomenon and explain the results based on the descriptive evidence that Friedman and Quah found to be a more reliable approach.

First, I can conclude that adding more conditional variables, such as state-saving rates and population growth rates, does not significantly increase the power of the regression model. This might be expected because I probably do not think that emissions are directly linked to variables such as savings rates, although the green Solow model suggests that they might be. Based on the overall results, the explanatory power of the regressions is pretty low, and the long specification does little to address this issue. That said, adding two conditional variables in the case of CO2 is helpful to some degree for both the full sample and the subsample of less-polluting states. Thus, focusing on the unconditional regression model for interpretation would be sufficient. The explanatory power of the regression models is not significantly increased by using the subsample of the 46 states with the least pollution. The obtained regression results, strengthen the overall argument that CVs provide the real evidence of convergence, not regressions.

Second, by comparing the results from the CV analysis and regression results, I can conclude that there are discrepancies between these two methods across pollutants, especially for CO2 and SO2, due to committing Galton's fallacy. Although the estimated coefficients are insignificant, the CO2 regression plot shows a pattern of convergence for the subsample. Nonetheless, their CV counterparts show a pattern of divergence. The CV analysis for CO2 and CO is showing a modest divergence across the states, which is not in line with the Brock and Taylor conclusion of convergence. In the case of SO2, a declining CV suggests a possible convergent pattern. Choosing the subsample reinforces this divergent trend. However, the estimated coefficients from the regression models for SO2 are not significant and therefore are not sufficiently informative. These inconsistencies among the results of the regression and CV analyses were anticipated and covered by Friedman (1992) and Quah (1993) in their relevant papers. They inform the reader that there is a potential Galton fallacy in interpreting the sign of the coefficient of initial levels as a sign of convergence. Based on their recommendation, the interpretation of the long-run trends should be deduced from CV analysis rather than misleading regressions.

Lastly, the analysis of CO2 emissions has been the focus of many scientists in relevant papers because of its huge share and impact on climate change compared to other pollutants. The lack of support for possible convergence across states is suggesting a challenging future for policymakers in the U.S. due to the discussions about the fair contribution of each state to climate agreements. As discussed in the introduction to this study, lack of convergence means that we have two types of states: some developed and some developing. Those that are already developed continue to pollute more, which makes it harder for states in their important stages of development to pledge to reduce pollution (by signing climate agreements). More efforts in the grassroots movement are needed to bring both local and state governments to the realization that paying attention to climate preservation policies is important. This requires more spending (both as private subsidies and federal investments in renewable energies instead of fossil fuels) as well as pledging to make major areas of the economy carbon neutral in the next decade. These can be perceived as beneficial sets of legislation for their constituencies in rural parts of the country by investing more and creating new jobs. Because of the irreversible nature of climate change and the aforementioned conflicts of interest, this issue requires more coordinated efforts than other economic matters.

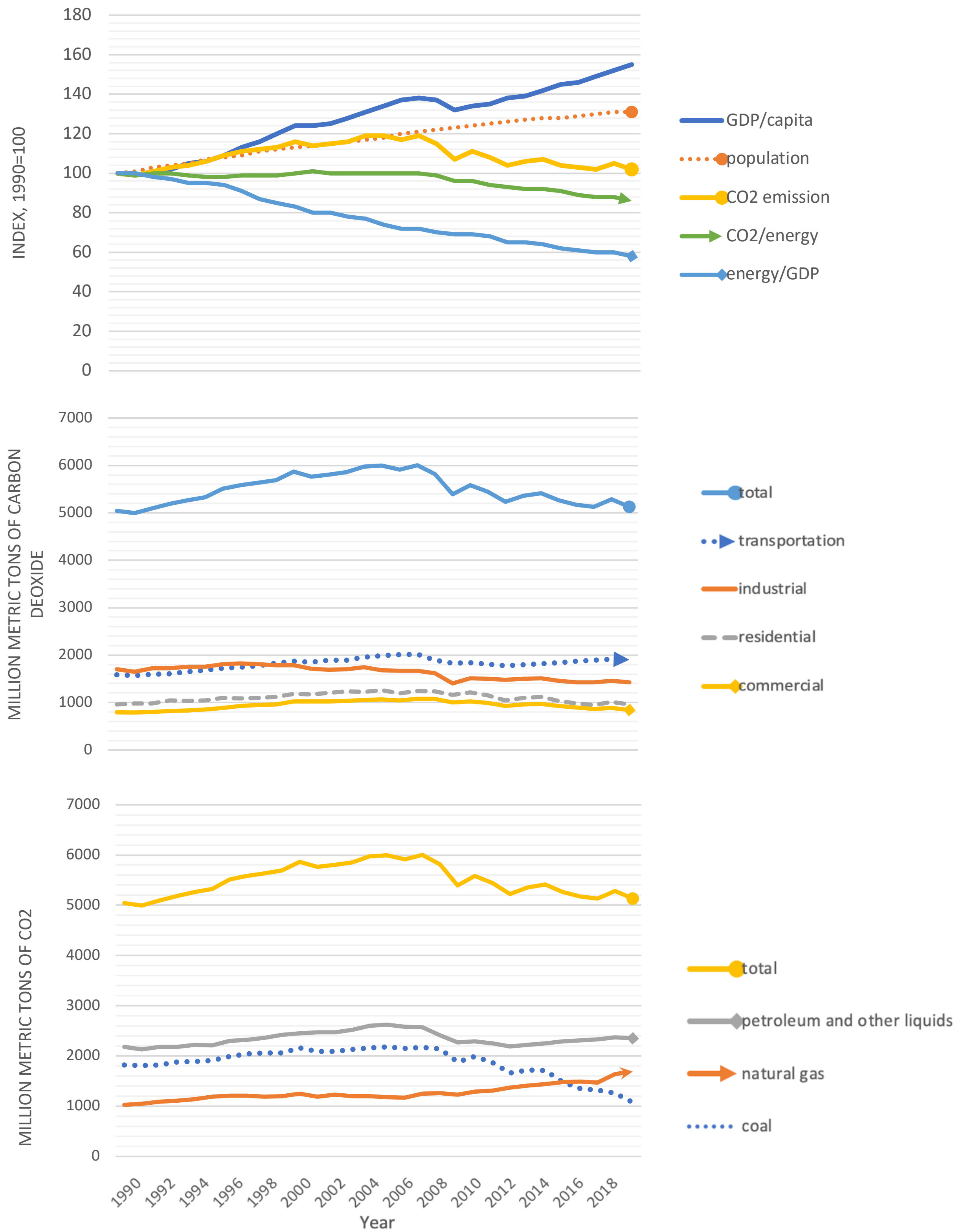


Figure 1. a) CO2 emissions in comparison with GDP per capita, population, and the energy intensity ratio and carbon intensity ratio. **b)** CO2 emissions by the fuel sector. **c)** CO2 emissions by fuel source, U.S. (1990-2019). Energy/GDP is called energy intensity, which measures the energy used (measured by Btu) to produce the total outputs in the economy. CO2/energy is carbon intensity, which is the weight of carbon emitted per unit of energy consumed.

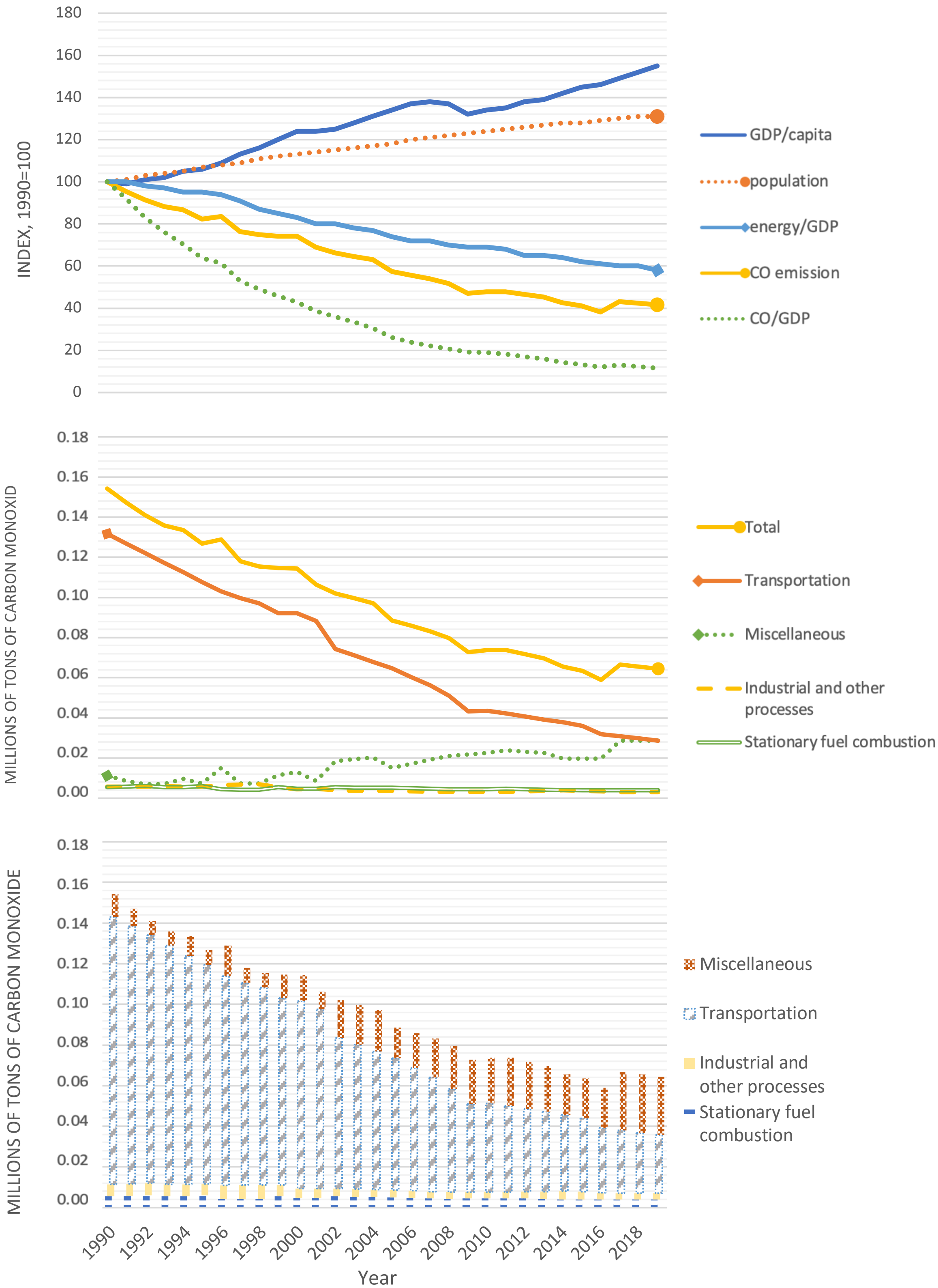


Figure 2. **a)** CO emissions in comparison with GDP per capita, population, and the energy intensity ratio. **b)** CO emissions trends by sector. **c)** CO emissions shares by sector, U.S. (1990-2019). Stationary fuel combustion is defined as the stationary combustion of fuels to produce heat, power and energy. Miscellaneous is defined as other sources of emissions that are not included in other categories mentioned.

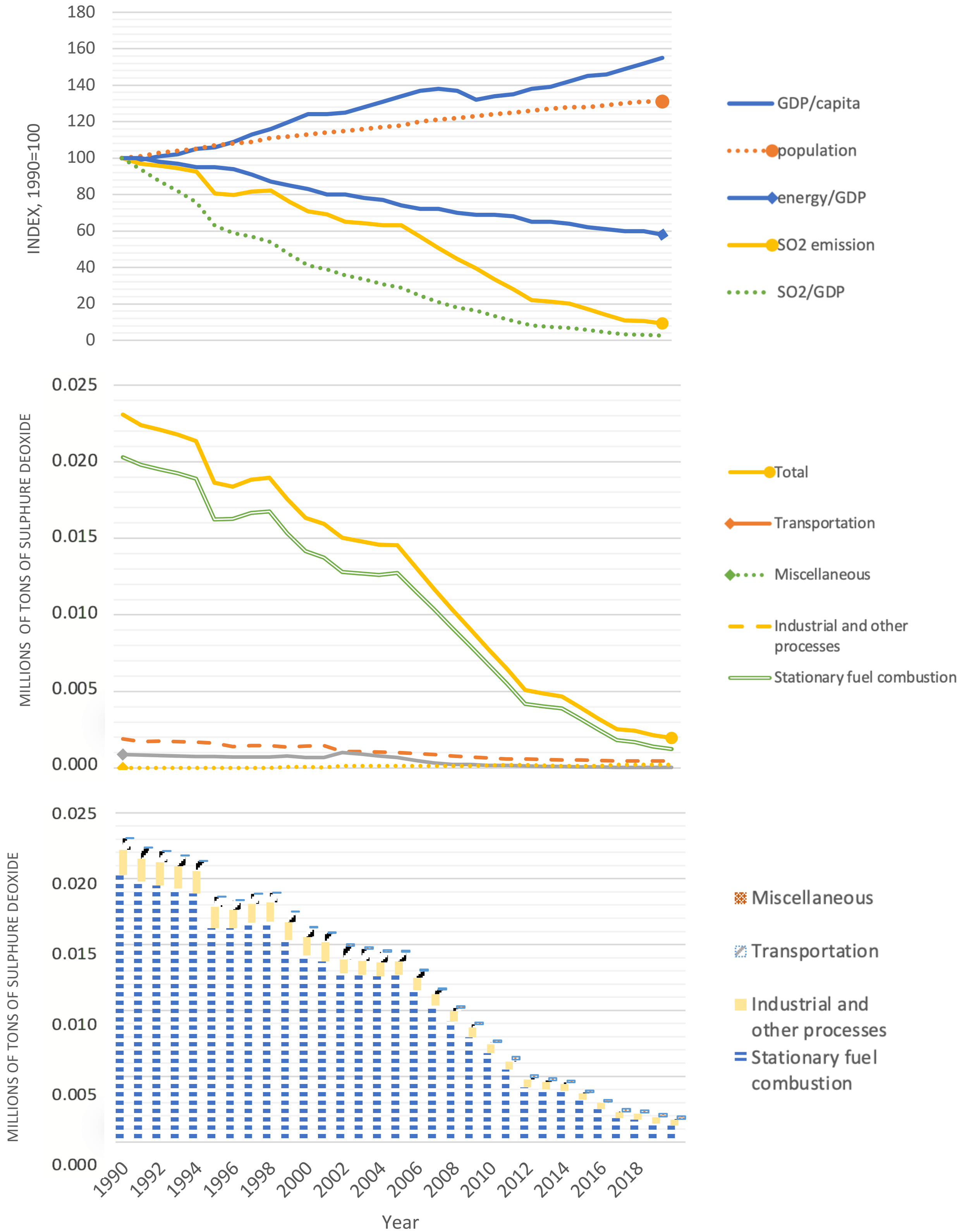


Figure 3. **a)** SO2 emissions in comparison with GDP per capita, population, and the energy-GDP ratio. **b)** SO2 emissions trends by sector. **c)** SO2 emissions shares by sector, U.S. (1990-2019). Stationary fuel combustion is defined as the stationary combustion of fuels to produce heat, power and energy. Miscellaneous is defined as other sources of emissions that are not included in other categories mentioned.

Table 1. Summary Statistics for the state dataset

Variable	50 states (1990-2018)		50 states (1996-2018)		45 states (1996-2018)	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
CO2 growth	-0.0061	0.05	-0.0061	0.05	-0.0063	0.05
Initial CO2	24.85	19.28	25.32	19.21	20.29	7.74
CO growth			0.010	0.86	0.011	0.72
Initial CO			0.68	0.97	0.61	0.45
SO2 growth			-0.049	0.216	-0.035	0.216
Initial SO2			0.10	0.15	0.09	0.11
Sample size	51*		51*		46*	

Notes: CO2 is measured in metric tons per capita, and CO and SO2 are measured in metric kilograms per capita. Column (3) is a subsample of 45 states, excluding 5 states with extremely high initial emissions. (*) District of Columbia has been assumed as an additional state in the dataset.

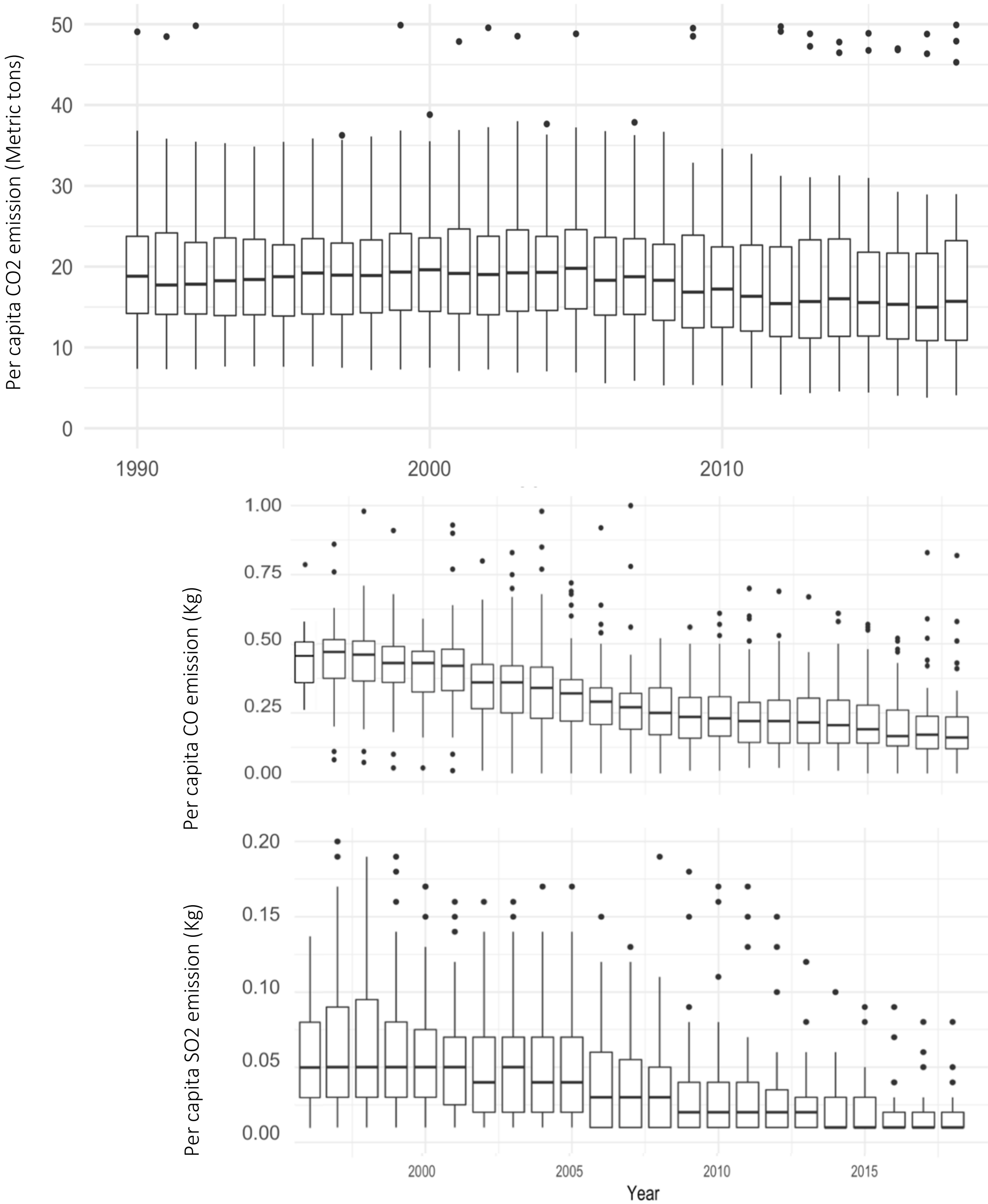


Figure 4. Per capita CO2 emission series across states (1990-2018)**.** Per capita CO emissions series across states (1996-2018). Per capita SO2 emissions series across states (1996-2018). The boxes are showing the data points from the 1st quartile (*25th percentile*) to the 3rd quartile (*75th percentile*) in my sample. The horizontal line within the box shows the median value. Each vertical line spans from the *minimum* (-1.5*IQR) to the *maximum* (+1.5*IQR) beyond the boxes (this gives us around a 95% confidence interval). Interquartile Range (IQR) is the range of boxes, which is the data from Q1 to Q3. The dots indicate potential outliers based on the sample

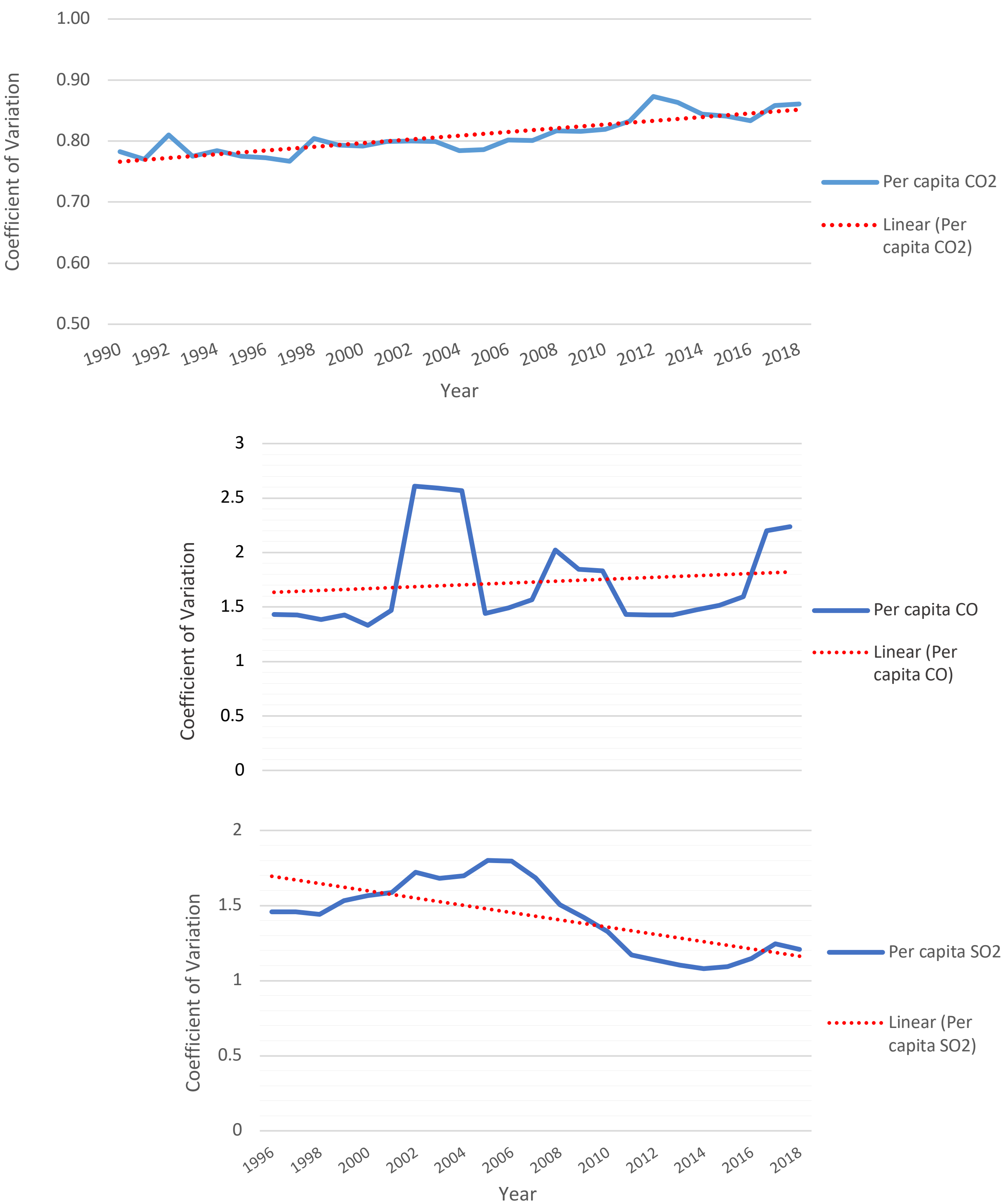
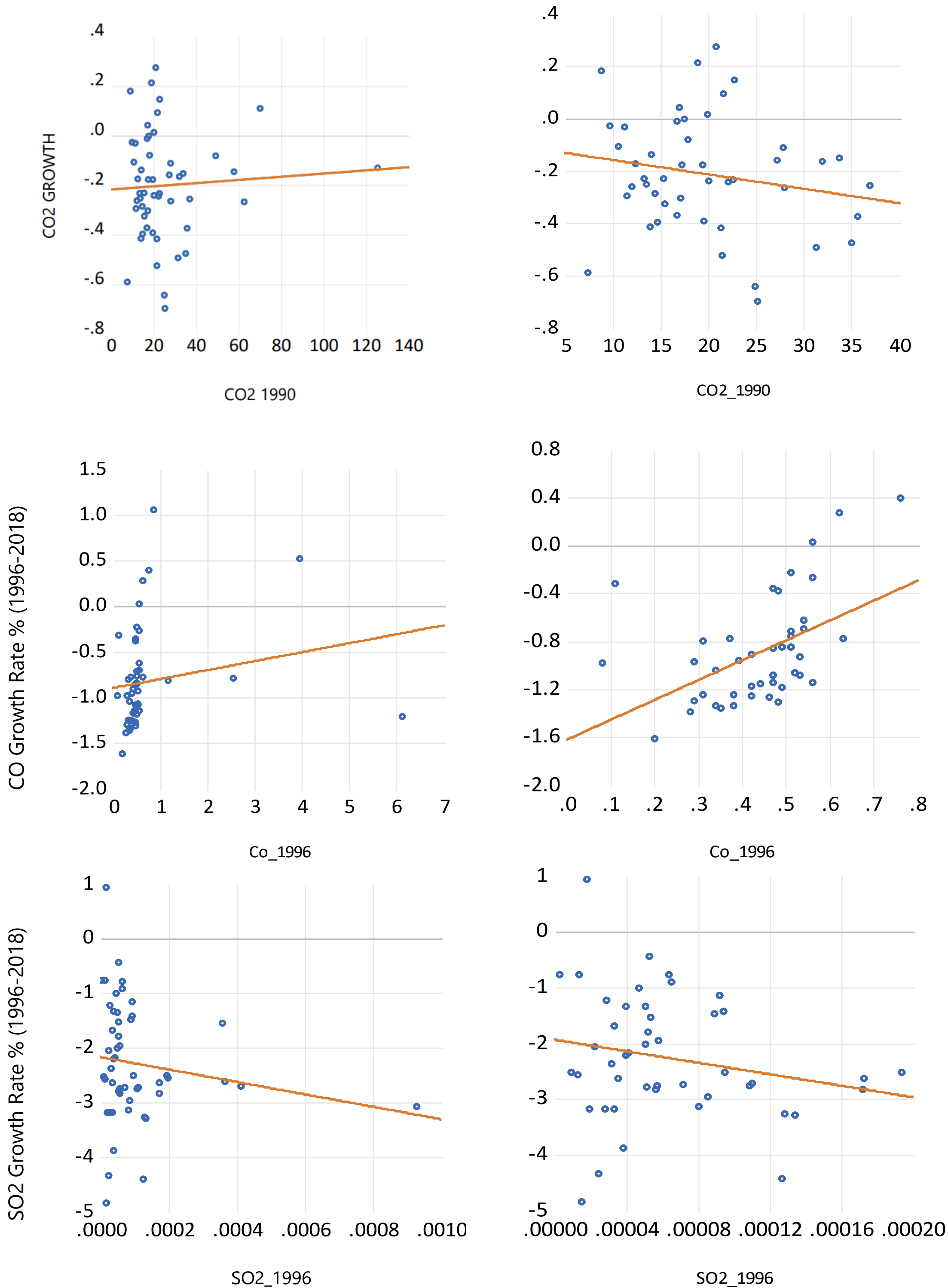


Figure 5. Coefficient of variation for emissions per capita across 51 states, for (a) CO2 (1990–2018), (b) CO (1997–2018), and (c) SO2 (1997–2018).

Figure 6. Coefficient of variation (CV) for emissions per capita across a subset of 46 states (states with the five highest initial levels are excluded). For (a) CO2 (1990–2018), (b) CO (1997–2018), and (c) SO2 (1997–2018).



a) Full Sample (51 States) b) Sub-sample of 46 States with lower initial emissions

Figure 7. The initial CO2 level (1990) versus the subsequent average growth rates (1990-2018), and the initial CO & SO2 levels (1996) versus their subsequent average growth rates, respectively. Each data point corresponds to a state. This is a demonstration of the unconditional regression, the short regression model in my study.[8]

[8] This graph is analogous to the Figures 1. presented by MRW (1992) and Quah (1993).

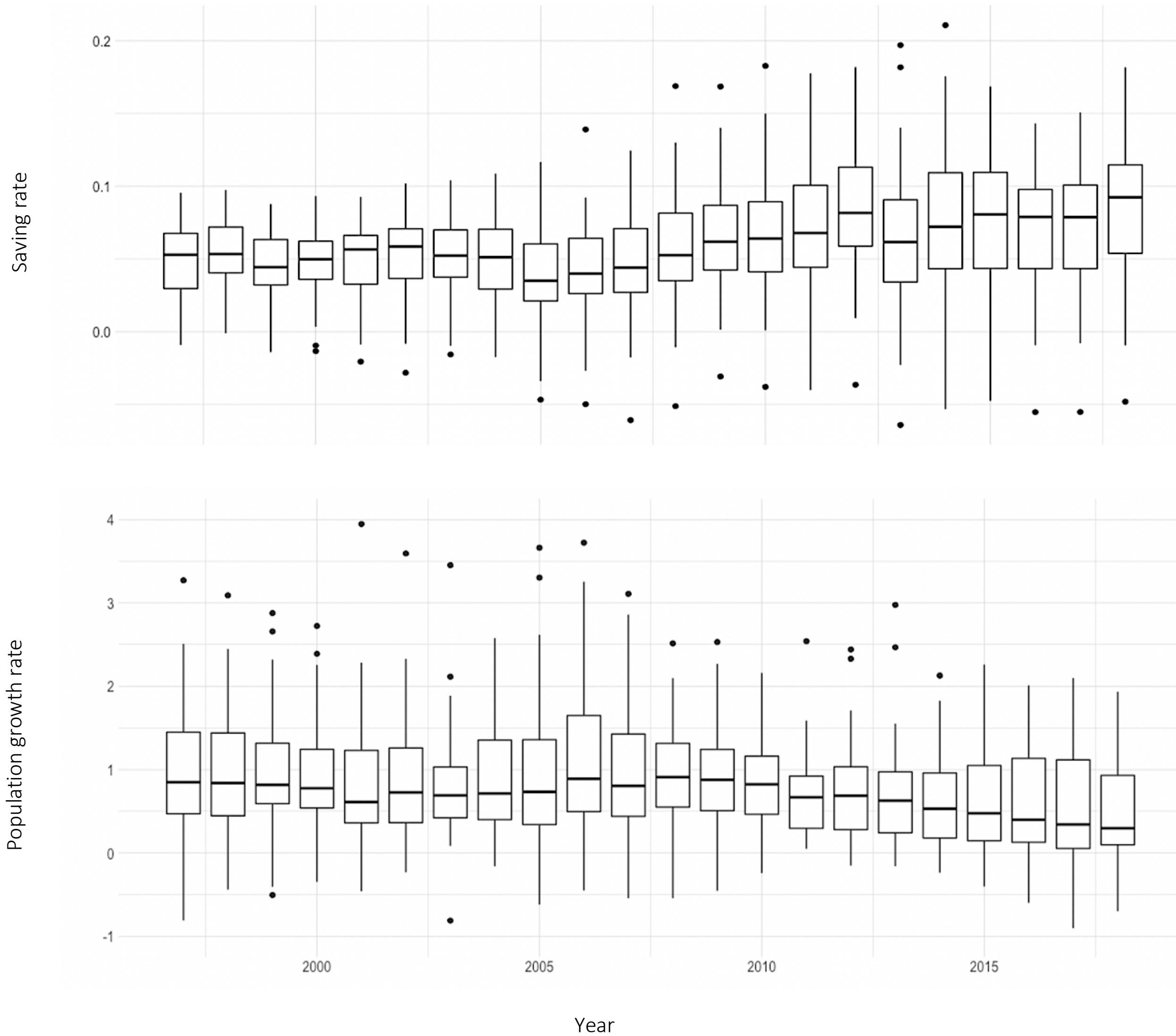


Figure 8. Saving rate and population growth rate across states (1997-2018).

Note: The boxes are showing the observations from the 1st quartile (*25th percentile*) to the 3rd quartile (*75th percentile*) in my sample. The horizontal line within the box shows the median value. Each vertical line spans from the *minimum* (-1.5*IQR) to the *maximum* (+1.5*IQR) beyond the boxes (this gives us around a 95% confidence interval). Interquartile Range (IQR) is the range of boxes, (Q1 to Q3). The dots indicate potential outliers based on the sample.

Table 2. Carbon dioxide regression model estimates across the states using ordinary least squares.

Variable	(1)	(2)	(3)	(4)
log(e0)	0.0050	-0.0057	-0.086	-0.0559
	(0.055)	(0.049)	(0.084)	(0.077)
log(S)		1.4533		1.6454
		(0.835)		(0.883)
log (N + 0.05)		-0.967		-0.9172
		(0.258)		(0.280)
Constant	-0.215	-3.007	-0.041	-2.734
	(0.172)	(0.751)	(0.249)	(0.869)
Observations	51	51	46	46
R^2	0.0001	0.026	0.023	0.27
Adjusted R^2	-0.020	0.021	0.00009	0.21
F Statistic	0.008	5.565	1.043	5.201

Notes: standard errors are in parentheses. Column (1) is the unconditional regression model, while column (2) is the conditional estimation model for the full sample. Columns (3) and (4) are the regressions for the subsample of 46 States with lower initial emissions. The dependent variable is the growth rate in log CO2 emissions per capita for 1990-2018. (e0) is the emission level in 1990, and (S) is the average saving to GDP ratio over 1997-2018. Similarly, (N+0.05) is the average population growth over the same period plus 0.05. Statistically significant coefficient estimates at the 5% level are denoted with asterisk.

Table 3. Carbon monoxide divergence across states

Variable	(1)	(2)	(3)	(4)
log (e0)	0.254	0.241	0.327	0.310
	(0.105)	(0.109)	(0.154)	(0.152)
log (S)		-0.471		2.321
		(2.256)		(1.902)
log (N + 0.05)		0.623		0.0911
		(0.807)		(0.650)
Constant	-0.643	1.153	-0.598	1.850
	(0.104)	(2.296)	(0.149)	(1.858)
Observations	51	51	46	46
R^2	0.105	0.117	0.092	0.163
Adjusted R^2	0.086	0.060	0.072	0.102
F Statistic	5.762	2.079	4.505	2.709

Notes: standard errors are in parentheses. Column (1) is the unconditional regression model, while column (2) is the conditional estimation model for the full sample. Columns (3) and (4) are the regressions of the subsample of 46 states with lower initial emissions. The dependent variable is the growth rate in log of CO emissions per capita for 1996-2018. (e0) is the emission level in 1996, and (S) is the average saving to GDP ratio over 1997-2018. Similarly, (N+0.05) is the average population growth over the same period plus 0.05.

Table 4. Sulphur dioxide regression model across states

Variable	(1)	(2)	(3)	(4)
log (e0)	-0.2002	-0.197	-0.246	-0.245
	(0.145)	(0.150)	(0.196)	(0.201)
log (S)		6.252		5.569
		(4.650)		(5.051)
log (N + 0.05)		0.718		0.741
		(1.686)		(1.804)
Constant	-4.229	-2.250	-4.707	-2.926
	(1.430)	(5.367)	(1.965)	(5.865)
Observations	51	51	46	46
R^2	0.037	0.077	0.034	0.067
Adjusted R^2	0.017	0.018	0.012	0.00005
F Statistic	1.884	1.309	1.570	1.000

Notes: standard errors are in parentheses. Column (1) is the unconditional regression model, while columns (2) is the conditional estimation model for the full sample. Columns (3) and (4) are the regressions of the subsample of 46 states with lower initial emissions. The dependent variable is the growth rate in log of SO2 emissions per capita for 1996-2018. (e0) is the emission level in 1996, and (S) is the average saving to GDP ratio over 1997-2018. Similarly, (N+0.05) is the average population growth over the same period plus 0.05.

Data Appendix

I) National Data

I have collected data from different U.S. agencies. The national data set of CO_2 was obtained from U.S. Energy and Information Administration (EIA). For CO and SO_2, the national data set was collected by U.S. Environmental Protection Agency (EPA).

In Figure 1. panel a), the plot has combined the trends of several emission variables in the same plot considering 1990 as the benchmark year[9]. In Figure 2 I obtain the data from different sources for CO and SO_2. This data collection is based on EPA reports [10] on national air pollution. It should be noted that these national series are only available with annual frequency starting in 1990—some data elements were not measured prior to 1990. Thus, I considered this year as the initial year of plot analysis which gives us a proper overview of the last three decades[11].

II) State Data

One of the main sources of the U.S. state emission data set is U.S. Energy Administration. The data on per capita carbon emissions was obtained under the section of state energy-related carbon dioxide emission (1990-2018)[12]. For the additional emissions like CO and SO_2, I have used the data gathered by U.S. Environmental Protection Agency[13] which was reported as National Emission Inventory documentation. I have used this extended dataset to gather CO and SO_2 emissions in levels for each state and then used population series to calculate the emissions per capita which were needed. For the population series, the source is the U.S. Census Bureau[14].

[9] U.S. Energy Information Administration, U.S. Energy-Related Carbon Dioxide Emission, (2019). https://www.eia.gov/environment/emissions/carbon/ https://www.eia.gov/environment/emissions/carbon/pdf/2019_co2analysis.pdf. The data on U.S. emission for the Figures 3-5 is taken from the above report by EIA (2019).

[10] https://www.epa.gov/air-emissions-inventories/2020-national-emissions-inventory-nei-documentation

[11] https://www.epa.gov/air-emissions-inventories/air-pollutant-emissions-trends-data

[12] https://www.eia.gov/environment/emissions/state/

[13] https://www.epa.gov/air-emissions-inventories/2020-national-emissions-inventory-nei-documentation. And https://www.epa.gov/air-emissions-inventories/air-pollutant-emissions-trends-data

[14] https://data.census.gov.

References

Aldy, J. E. (2006), Per capita carbon dioxide emissions? Convergence or divergence? Environ Res Econ, 33: 533–555

Azomahou, T., F. Laisney, and P. Nguyen Van (2006), 'Economic development and CO2 emissions: A nonparametric panel approach, Journal of Public Economics

Brock, W. A., & Taylor, M. S. (2004). The Green Solow model. Social Sciences Research Institute Working Paper No. 2004-16, Economics Department, University of Wisconsin.

Brock, W.A. and M.S. Taylor (2010), 'The Green Solow Model', Journal of Economic Growth 15(2): 127 153.

Chen, WJ. (2017). Is the Green Solow Model Valid for CO2 Emissions in the European Union? Environ Resource Econ 67, 23–45.

Delgado Michael, 2013. "A smooth coefficient model of carbon emissions," Empirical Economics, Springer, vol. 45(3), pages 1049-1071.

Dinda, S., 2004. Environmental Kuznets curve hypothesis: a survey. Ecological economics, 49(4), pp.431-455.

Durlauf, S., & Quah, D. (1999). The new empirics of economic growth. In J. Taylor & M. Woodford, Handbook of macroeconomics. Amsterdam: North Holland.
Encyclopedia of Biodiversity (Second Edition), Academic Press, 25–30.

Friedman, M. (1992). "Do old fallacies ever die?" Journal of Economic Literature 30, 2129-2132.

Grossman, G. M., & Krueger, A. B. (1994), Environmental impacts of a North American Free Trade Agreement. In P. Garber, The US-Mexico Free Trade Agreement. Cambridge, MA: MIT Press.

Grossman, G. M., & Krueger, A. B. (1995), Economic growth and the environment. Quarterly Journal of Economics, 110, 353–377.

Hao, Y., & Wei, Y. (2015). When does the turning point in China's CO2 emissions occur? Results based on the Green Solow model. *Environment and Development Economics, 20*(6), 723- 745. doi:10.1017/S1355770X15000017

Hotelling, H., "Review of The triumph of mediocrity in business, by Horace Secrist," J. Amer. Statist. Assoc., Dec. 1933, 28(184), pp. 463-65.

Im, K., M. Pesaran and Y. Shin (2002), 'Testing for Unit Roots in Heterogeneous Panels', Journal of Econometrics, forthcoming.

Mäler, K. G. (2013), Economic Growth and the Environment, Editor(s): Simon A Levin,

Mankiw, N. G., Romer, D., & Weil, D. N. (1992). A contribution to the empirics of economic growth. Quarterly Journal of Economics, 107, 407–437.

of Economics 95, 427-443.

Quah, D. (1993). “Galton's fallacy and tests of the convergence hypothesis” Scandinavian Journal

Solow, 1956. "A Contribution to the Theory of Economic Growth," The Quarterly Journal of Economics, Oxford University Press, vol. 70(1), pages 65-94.

Strazicich, M.C., List JA (2003), Are CO2 emissions levels converging among industrial countries? Environ Res Econ 24:263–271

U.S. Energy Information Administration, U.S. Energy-Related Carbon Dioxide Emission, (2019). https://www.eia.gov/environment/emissions/carbon/pdf/2019_co2analysis.pdf.

Van, P.N. (2005). Distribution dynamics of CO2 emissions. Environ Res Econ, 32: 495–508

Van, P.N. (2010). Energy consumption and income: A semiparametric panel data analysis. Energy Economics, 32(3), 557-563.