University of Texas at Dallas (2023), 1-8

ARTICLE TYPE

Time Series Analysis of Air Pollution in the United States

Manoj Mareedu,[1] Prriyamvradha Parthasarathi,[2] Premi Jawahar Vasagam,[3] Mira Radhakrishnan,[4] Sofia Rajan,[5] Martin Navarro,[6] Vyshnavi Gangineni,[7] and Siva Renuka Chowdary Nandigam[8]

[1] MXM220069, The University of Texas at Dallas, Dallas, 75080, TX, Richardson
[2] PXP220005, The University of Texas at Dallas, Dallas, 75080, TX, Richardson
[3] PXJ220007, The University of Texas at Dallas, Dallas, 75080, TX, Richardson
[4] MXR220049, The University of Texas at Dallas, Dallas, 75080, TX, Richardson
[5] SXR220034, The University of Texas at Dallas, Dallas, 75080, TX, Richardson
[6] MXN180019, The University of Texas at Dallas, Dallas, 75080, TX, Richardson
[7] VXG220046, The University of Texas at Dallas, Dallas, 75080, TX, Richardson
[8] SXN220083, The University of Texas at Dallas, Dallas, 75080, TX, Richardson
Author for correspondence: M. Mareedu, Email: manoj.mareedu@utdallas.edu

Abstract

This paper examines the identifying key contributors to air pollution and understanding their dynamics in the United States. The analysis focuses on four major pollutants carbon monoxide (CO), nitrogen dioxide (NO2), ozone (O3), and sulfur dioxide (SO2). The study draws on data from 2000 to 2015 to uncover patterns of the air pollutants in each state. Overall, this paper contributes to a better understanding of the complexities of air pollution over a specific time period with a holistic examination of air quality dynamics across various U.S. states.

Keywords: Air Pollution, Time Series Analysis, United States, Environmental Economics

INTRODUCTION

The United States, with its diverse geography, industrial activities, and population density, presents a unique landscape for studying the fluctuation of air quality. Over the years, regulatory measures and technological advancements have influenced the levels of pollutants emitted into the atmosphere. Institutions like the Environmental Protection Agency (EPA) periodically review and update the National Ambient Air Quality Standards (NAAQS) to protect and regulate air quality. Moreover, it also documents various units of measurement for multiple pollutants, such as carbon monoxide (CO), nitrogen dioxide (NO_2), ozone (O_3), and sulfur dioxide (SO_2), that affect all of us. The atmosphere we breathe is susceptible to various pollutants originating from diverse sources. These pollutants, including carbon monoxide (CO), nitrogen dioxide (NO_2), ozone (O_3), and sulfur dioxide (SO_2), arise from industrial emissions, vehicular exhaust, and natural processes. Understanding their sources and potential health impacts is crucial for comprehending the complexities of air quality. Recent technological advancements have significantly transformed our ability to monitor and analyze air quality. Integrating satellite monitoring, sensor networks, and sophisticated modeling techniques has provided unprecedented insights into pollutant concentrations and their spatial distribution. Using big data and machine learning algorithms has empowered researchers to handle vast datasets, facilitating more accurate and comprehensive analyses. Hence, having a constant and compelling analysis that tracks such pollutants can lead to casual inferences, which can be a critical piece of information that can benefit all and help implement new regulations and laws. Due to its importance, we will conduct a study to undertake a comprehensive time series analysis to gain insights into the patterns and trends shaping air quality across

the United States. By doing so, we hope to uncover the intricate fluctuations of air pollution, contribute to a more profound understanding of its complexities, and reveal causal inferences that can assist in implementing regularity measures.

RESEARCH OBJECTIVE

In our time series analysis, we have established study objectives to understand better the complexities of air quality in the United States. Achieving our research objectives holds substantial societal and environmental benefits. By identifying key contributors to air pollution and understanding their dynamics, we can inform evidence-based policies. This has the potential to lead to tangible improvements in air quality, reducing the burden of respiratory diseases and fostering a healthier living environment for communities across the United States. As we give top priority to conducting a thorough investigation, we will focus on the following goals:

- Analyze Time Series Patterns: Analyze time series data in detail to learn about how the quality of the air has changed in various regions.
- Study Pollutants of Interest: Conduct a thorough analysis of specific contaminants to understand air quality dynamics comprehensively.
- State-wise Analysis: Analyze the best states, looking for differences and patterns to gain an essential understanding of the local air quality scene.

From the objectives mentioned above, we will elaborate on interpreting the results and investigate causal inferences when applicable.

LITERATURE REVIEW

A time series analysis is a statistical technique utilized to analyze and interpret records over time. The objective is to understand the data's patterns, trends, and behaviors to make predictions, draw conclusions, or reveal causal inferences. Such a technique has been used before to predict, understand, and regulate air pollution in the world. An example of this study is expressed in Nina Sidneva Jones' dissertation essay written jointly with Eric Zivot, titled Trends in U.S. Air Pollution: A Time Series Analysis, in which her goal is to determine whether emissions of nitrogen oxides and volatile organic compounds are trend-stationary or difference-stationary and whether there was a break in the trends of these emissions around the same time the Clean Air Act of 1970 was passed. The study results show that volatile organic compound emissions are trend-stationary with a break when the Clean Air Act of 1970 was passed.

Furthermore, the article, the worst times for air quality: Understanding air pollution patterns and trends, by Sienna Bishop illustrates air quality fluctuations across time, trends over time in specific air pollutants, and an explanation of air pollution patterns. In her article, Bishop shows the decrease of carbon monoxide, nitrogen dioxide, sulfur dioxide, and ozone over the last 20 years worldwide. She mentions such a decrease is due to the implementation of various air quality policies worldwide to reduce pollution and further explains patterns by continents due to climate change.

Our study addresses significant gaps in the existing literature by undertaking a comprehensive analysis of multiple pollutants across various U.S. states over a specific time -period. While previous research has provided valuable insights, the need for a holistic examination of air quality dynamics in the U.S. context remains, especially considering regional variations, policy nuances, and diverse pollution sources. While global trends in air quality, as elucidated by Sienna Bishop, offer valuable insights into broader patterns, our U.S.-focused study recognizes the importance of understanding regional variations and policy-specific influences. By concentrating on the unique dynamics within the United States, we aim to contribute nuanced findings that complement and extend the existing global perspective.

With the mentioned papers in mind, our study will contribute to the field by uncovering patterns of air pollutants across the states from 2000 to 2015, disclosing how much each state contributes to air

pollution, discovering which pollutant contributes to maximum air pollution, and tracking trends in air pollution over time. Nevertheless, such findings will be interpreted to determine causal inferences and provide relevant information that can be used to implement new regulations for air quality in the United States.

DATA COLLECTION

The data used for our analysis is a panel dataset comprising records from 2000 to 2015 in the United States. The data is arranged in 24 columns. Four of them provide the location of the measurements, such as the address and city. For each pollutant, information is provided in the remaining 20 columns, including the Air Quality Index (AQI), the units, average and maximum values, and the recording date. It is important to acknowledge the challenges inherent in our dataset. Instances of missing data, particularly for certain periods or regions, may introduce limitations to our analysis. By transparently communicating these challenges, we ensure a nuanced interpretation of our findings and encourage future researchers to address these limitations. The utilization of open data sources, such as data World, plays a pivotal role in fostering collaborative and transparent research efforts. Access to comprehensive datasets enables a community-driven approach, encouraging researchers worldwide to engage in the exploration of complex issues such as air quality dynamics. This collaborative spirit enhances the reliability and reproducibility of scientific findings. The following variables are implemented in our analysis:

- State – State monitoring site.
- Date Location: Date of monitoring.
- NO_2 Mean (yn_mean) – The arithmetic means of concentration of nitrogen dioxide within a given day.
- NO_2 AQI (yn) – The calculated air quality index of nitrogen dioxide within a given day.
- CO Mean (yc_mean) – The arithmetic means of concentration of carbon monoxide within a given day.
- CO AQI (yc) – The calculated air quality index of carbon monoxide within a given day.
- O_3 Mean – The arithmetic means of concentration of ozone within a given day.
- O_3 AQI (yo) – The calculated air quality index of ozone within a given day.
- SO_2 Mean (ys_mean) – The arithmetic means of concentration of sulfur dioxide within a given day.
- SO_2 AQI (ys) – The calculated air quality index of sulfur dioxide within a given day.

Source: data. World

EXPLORATORY DATA ANALYSIS

To gain initial insights into air pollution trends across different states in the United States from 2000 to 2015, we conducted exploratory data analysis (EDA), focusing on total pollutants' concentration, identifying states with significant contributions, and pinpointing contaminants driving air pollution, with the aim of discerning reasons behind the observed increase in pollution.

1. Total Pollutants Concentration Across States:

To initiate our exploratory data analysis (EDA), we plotted the total concentration of pollutants across various states in the United States. This visual representation allowed us to observe general patterns and identify states with consistently high pollution levels. The graph showed that New Jersey, Massachusetts, New York, Arizona, and Colorado were the top states contributing significantly to air pollution.

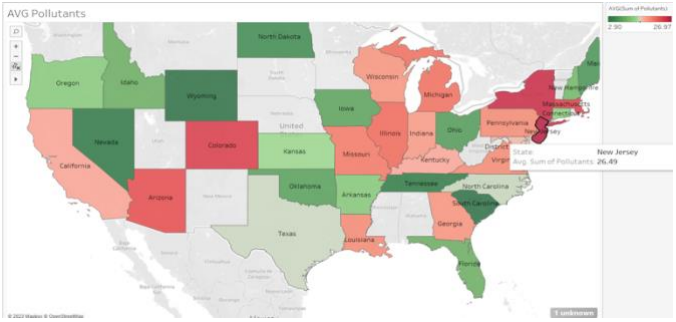


Figure 1. Total Pollutants Concentration Across U.S. States

2. Average Pollutant Concentrations

Subsequently, we plotted the average values of the four pollutants' concentrations—carbon monoxide (CO), nitrogen dioxide (NO_2), sulfur dioxide (SO_2), and ozone (O_3). This analysis revealed that ozone (O_3) had the highest average concentration among the pollutants, signifying its significant contribution to overall air pollution.

To refine our analysis further, we chose ozone (O_3) as the base pollutant for our study. The decision was based on its consistently high average concentration compared to other pollutants. By selecting ozone (O_3) as the base pollutant, we aimed to explore the dynamics of this pollutant in greater detail.

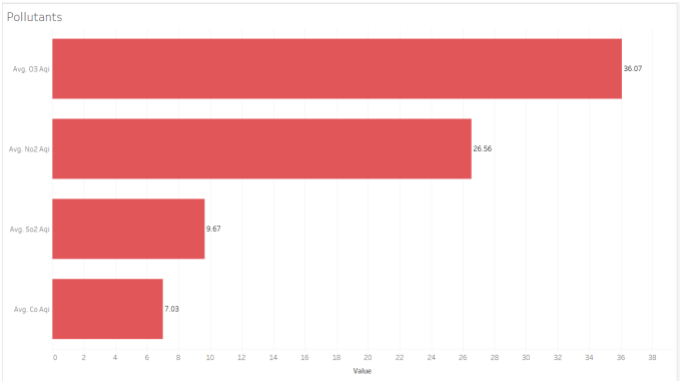


Figure 2. Average Concentrations of Key Air Pollutants

3. Trend Analysis Across Years

To analyze the trends of the four pollutants over the years, we plotted the average pollutant concentration across the study period (2000-2015). The results indicated a stable ozone (O_3) concentration across the years, with a noticeable increase towards the end of the study period. In contrast, the other pollutants—carbon monoxide (CO), nitrogen dioxide (NO_2), and sulfur dioxide (SO_2)—exhibited a downward trend, indicating a decline in their concentrations over the years.

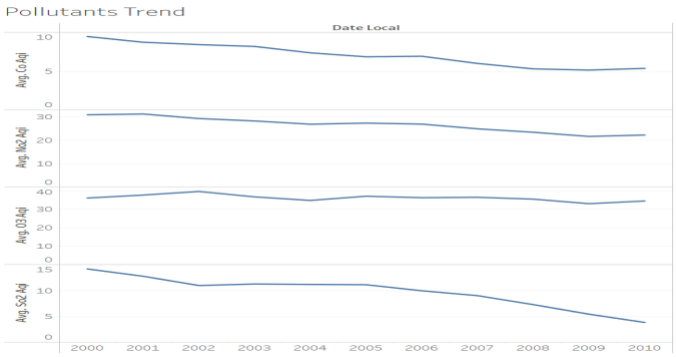


Figure 3. Temporal Trends in Air Quality: A Comprehensive Analysis of Ozone and Associated Pollutants

Our EDA has not only offered a panoramic view of pollutant concentrations but has been further refined by homing in on ozone (O_3) as the cornerstone. This choice, substantiated by its persistent high concentration and stable trend, paves the way for an in-depth time series analysis, poised to unravel the subtleties of ozone dynamics and its evolving impact on air quality.

Significance of Ozone:

Ozone, with its high average concentration, emerges as a critical pollutant with significant health implications and regulatory importance. Understanding the dynamics of ozone is paramount for formulating effective air quality management strategies. Our choice of ozone as the base pollutant reflects its vital role in shaping the overall air quality landscape.

EMPIRICAL METHOD

We have evidence of the variables being stationary before being fetched into the models by the Dicker-Fuller test, a critical factor in time series data analysis. Due to accuracy, the data is considered for analysis in a month-wise order. It is justifiable because pollutants' effects cannot be studied day-wise. Only with a more extended period will it show some environmental impacts. So, analysis is performed monthly. The Dicker-Filler test is done after inputting the null values with the mean. The Dicker-Fuller test holds practical significance in ensuring the stationarity of our time series data.

Stationarity is a crucial assumption for accurate time series analysis, and this test plays a vital role in confirming the robustness of our models. Non-stationary data can lead to unreliable predictions, emphasizing the importance of this preliminary testing phase.

Our choice of conducting a monthly analysis aligns with the environmental impact considerations of pollutants. Studying pollutants monthly provides a more comprehensive view of their effects, as daily variations may not fully capture their environmental impact. This approach allows us to identify broader patterns that contribute to a more nuanced understanding of air quality dynamics.

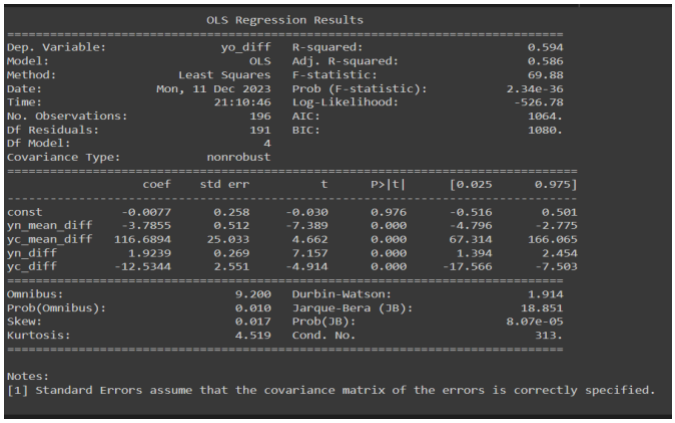
After checking the stationarity of the dependent and independent timeseries variables, initially they the p-value of Dicker-Fuller test is greater than significant level (> 0.05) which denotes statistically not significant. This means that our time series variables are not stationary. So, we used first order differencing method to transform the time series variables. After the transformation we made sure that our time series variables are stationary through Dicker-Fuller test. After the transformation all the time series variables are stationary (i.e., p-value <0.05).

After applying fist-order differencing to the time series variables, trend and seasonality is removed from the time series and made the time series variables stationary. Now all the differenced time series variables represent change in them.

After running an OLS regression, we got a model that predicted the calculated air quality index of ozone within a given month based on the concentration and air quality index of nitrogen dioxide and carbon monoxide in the air.

The model equation is estimated as,

$$\Delta\,\text{O3 AQI} = \beta_0 + \beta_1 \Delta\text{NO2 Mean} + \beta_2 \Delta\text{CO Mean} + \beta_3 \Delta\text{NO2 AQI} + \beta_4 \Delta\text{CO AQI} + u$$



OLS Regression Results

Dep. Variable:	yo_diff	R-squared:	0.594
Model:	OLS	Adj. R-squared:	0.586
Method:	Least Squares	F-statistic:	69.88
Date:	Mon, 11 Dec 2023	Prob (F-statistic):	2.34e-36
Time:	21:10:46	Log-Likelihood:	-526.78
No. Observations:	196	AIC:	1064.
Df Residuals:	191	BIC:	1080.
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-0.0077	0.258	-0.030	0.976	-0.516	0.501
yn_mean_diff	-3.7855	0.512	-7.389	0.000	-4.796	-2.775
yc_mean_diff	116.6894	25.033	4.662	0.000	67.314	166.065
yn_diff	1.9239	0.269	7.157	0.000	1.394	2.454
yc_diff	-12.5344	2.551	-4.914	0.000	-17.566	-7.503

Omnibus:	9.200	Durbin-Watson:	1.914
Prob(Omnibus):	0.010	Jarque-Bera (JB):	18.851
Skew:	0.017	Prob(JB):	8.07e-05
Kurtosis:	4.519	Cond. No.	313.

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Figure 4. OLS Regression Results

The output shows the regression results with the change calculated air quality index of ozone as the dependent variable and the change in concentration and change calculated air quality index of nitrogen dioxide and carbon monoxide in the air as independent variables. From the result, we can understand that all these variables are statistically significant, with their p-values being less than 0.05. On observing the coefficient, the change in mean concentration of carbon monoxide contributes more to increasing change in ozone. This is evident because when carbon monoxide mixes with other gases or pollutants, it increases ozone, proven by our model output's higher coefficient and carbon monoxide combined with nitrogen dioxide. This model explains 59.4% variation in the target variable. The Durbin-Watson value of 1.914 indicates no compelling evidence of autocorrelation in the residuals, which is a positive aspect for causal analysis. From the model Analogously, change in NO_2 Mean, NO_2 AQI and CO AQI may have causal relationships with change in O_3 AQI based on the coefficients.

We performed the Granger causality test to study the existence of past value influences in the present values in a regression analysis. Here, it is performed for all the independent variables. This is like studying for the F test. This result shows its significance with its F stat value, which is more significant in lag 2 with a model accuracy of 60%. Here, the result for one independent variable is added. Similarly, results are obtained for all other independent variables.

```
Test results for yo_diff, yc_diff

Granger Causality
number of lags (no zero) 1
ssr based F test:         F=13.2134 , p=0.0004  , df_denom=190, df_num=1
ssr based chi2 test:   chi2=13.4221 , p=0.0002  , df=1
likelihood ratio test: chi2=12.9759 , p=0.0003  , df=1
parameter F test:         F=13.2134 , p=0.0004  , df_denom=190, df_num=1

Granger Causality
number of lags (no zero) 2
ssr based F test:         F=15.8529 , p=0.0000  , df_denom=187, df_num=2
ssr based chi2 test:   chi2=32.5535 , p=0.0000  , df=2
likelihood ratio test: chi2=30.0708 , p=0.0000  , df=2
parameter F test:         F=15.8529 , p=0.0000  , df_denom=187, df_num=2
```

Figure 5. Granger causality test result 1

```
Test results for yo_diff, yn_diff

Granger Causality
number of lags (no zero) 1
ssr based F test:         F=12.2342 , p=0.0006  , df_denom=190, df_num=1
ssr based chi2 test:   chi2=12.4274 , p=0.0004  , df=1
likelihood ratio test: chi2=12.0437 , p=0.0005  , df=1
parameter F test:         F=12.2342 , p=0.0006  , df_denom=190, df_num=1

Granger Causality
number of lags (no zero) 2
ssr based F test:         F=21.1976 , p=0.0000  , df_denom=187, df_num=2
ssr based chi2 test:   chi2=43.5288 , p=0.0000  , df=2
likelihood ratio test: chi2=39.2329 , p=0.0000  , df=2
parameter F test:         F=21.1976 , p=0.0000  , df_denom=187, df_num=2
```

Figure 6. Granger causality test result 2

```
Test results for yo_diff, yn_mean_diff

Granger Causality
number of lags (no zero) 1
ssr based F test:         F=17.7463 , p=0.0000  , df_denom=190, df_num=1
ssr based chi2 test:   chi2=18.0265 , p=0.0000  , df=1
likelihood ratio test: chi2=17.2336 , p=0.0000  , df=1
parameter F test:         F=17.7463 , p=0.0000  , df_denom=190, df_num=1

Granger Causality
number of lags (no zero) 2
ssr based F test:         F=18.3842 , p=0.0000  , df_denom=187, df_num=2
ssr based chi2 test:   chi2=37.7516 , p=0.0000  , df=2
likelihood ratio test: chi2=34.4646 , p=0.0000  , df=2
parameter F test:         F=18.3842 , p=0.0000  , df_denom=187, df_num=2
```

Figure 7. Granger causality test result 3

```
Test results for yo_diff, yc_mean_diff

Granger Causality
number of lags (no zero) 1
ssr based F test:         F=13.8210 , p=0.0003  , df_denom=190, df_num=1
ssr based chi2 test:   chi2=14.0392 , p=0.0002  , df=1
likelihood ratio test: chi2=13.5521 , p=0.0002  , df=1
parameter F test:         F=13.8210 , p=0.0003  , df_denom=190, df_num=1

Granger Causality
number of lags (no zero) 2
ssr based F test:         F=14.2398 , p=0.0000  , df_denom=187, df_num=2
ssr based chi2 test:   chi2=29.2412 , p=0.0000  , df=2
likelihood ratio test: chi2=27.2175 , p=0.0000  , df=2
parameter F test:         F=14.2398 , p=0.0000  , df_denom=187, df_num=2
```

Figure 8. Granger causality test result 4

This test results we observed that the past values of NO_2 mean, NO_2 AQI, CO mean, and CO.

AQI are useful in predicting the present value of O_3 AQI. It is determined by the p-vales of F-test. All the p-values are less than 0.05 which denotes that there is statistically significant evidence to reject null hypothesis. Which means the past values of NO_2 Mean, NO_2 AQI, CO Mean and CO AQI time series can be used to predict future values of O_3 AQI time series. In other words, there is causal relationship between the target variable O_3 AQI and each of the independent variables mentioned above based on observed data.

CONCLUSION

In conclusion, our comprehensive time series analysis has provided valuable insights into the fluctuations and dynamics of air quality across the United States from 2000 to 2015. Through a meticulous exploration of the data and the application of empirical methods, we aimed to achieve several key objectives.

Our study began by acknowledging the unique environmental landscape of the United States, marked by diverse geography, industrial activities, and population density. Recognizing the influence of regulatory measures and technological advancements on air quality, we embarked on a journey to uncover the intricate patterns of pollutants such as carbon monoxide (CO), nitrogen dioxide (NO_2), ozone (O_3), and sulfur dioxide (SO_2). This endeavor's importance lies in its potential to contribute critical information for the implementation of regulatory measures and laws aimed at safeguarding air quality.

The research objectives guided our exploration, leading to a focused analysis of time series patterns, a detailed examination of specific pollutants, and a state-wise analysis to understand regional variations. Building on existing literature, particularly studies such as Nina Sidneva Jones' dissertation on U.S. air pollution trends, and Sienna Bishop's exploration of global air quality patterns,

our study aimed to make a distinct contribution to the field.

Our data collection involved a meticulous assembly of a panel dataset spanning 15 years, encompassing 24 columns of information. Through exploratory data analysis (EDA), we identified New Jersey as a base state for further investigation and selected ozone (O_3) as the primary pollutant of interest due to its consistent contribution to overall air pollution.

The empirical method employed involved rigorous testing for stationarity, OLS regression, and Granger causality tests. The results revealed statistically significant relationships among key variables, emphasizing the substantial contribution of carbon monoxide to ozone levels. The monthly analysis, supported by the Dicker-Fuller test, provided a nuanced understanding of pollutant dynamics over time.

Our findings carry implications for policymakers, offering insights into pollutant trends, regional variations, and the influence of specific contaminants on air quality. As we move forward, it is essential to interpret these results judiciously, considering the potential impact on regulatory frameworks and public health.

From an econometric perspective, stressing the importance of ongoing monitoring efforts is essential for maintaining the accuracy and relevance of econometric models. The dynamic nature of air quality necessitates continuous data collection, allowing for the recalibration of econometric models in response to changing economic and environmental conditions. Incorporating real-time data into econometric analyses ensures that policy recommendations derived from such models remain timely and effective.

Our research contributes to environmental economics and emphasizes the interdisciplinary nature of econometrics. Collaborations with public health experts, policymakers, and environmental scientists are critical for enhancing the validity and applicability of econometric models. This interdisciplinary approach ensures that the economic implications derived from our findings align with broader societal and environmental goals. Encouraging future researchers to explore such collaborations emphasizes the need for a comprehensive econometric framework in addressing complex challenges associated with air pollution and environmental economics.

Our findings suggest that reducing air pollution can potentially decrease instances of respiratory diseases, leading to lower healthcare costs for individuals and governments. This health benefit is associated with increased productivity, positively impacting the overall labor force. Businesses in pollution-contributing regions can leverage our study to comply with regulations and drive innovation in technologies targeting specific pollutants, with long-term economic advantages. Furthermore, our research supports the growth of green industries by identifying key pollutants, promoting the development of emission-reducing technologies, and fostering the creation of sustainable jobs, thereby contributing to economic growth.

In essence, this research not only contributes to the academic understanding of air quality dynamics but also holds practical relevance for shaping policies that address the complex challenges posed by air pollution in the United States. As we conclude this study, we recognize the ongoing importance of monitoring air quality and refining regulatory measures to ensure a healthier and sustainable environment for current and future generations.

REFERENCES

https://data.world/data-society/us-air-pollution-data
https://www.proquest.com/openview/06ad39d9bbc604fe02fa58d37d228eab/1?pq-origsite=gscholar&cbl=18750
https://www.clarity.io/blog/worst-times-for-air-quality-understanding-air-pollution-patterns-and-trends