Three Essays on Health and Environmental Economics: Applications of Spatial Econometrics and Spatial Analysis

Mohammed Syedul Islam

Dissertation submitted to the Davis College of Agriculture, Natural Resources and Design at West Virginia University in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Natural Resource Economics

Alan R. Collins, Ph.D., Committee Chairperson

Michael P. Strager, Ph.D.

Daniel S. Grossman, Ph.D.

Donald J. Lacombe, Ph.D.

Xiaoli Etienne, Ph.D.

Division of Resource Economics and Management

Morgantown, West Virginia

2023

Keywords: Lung and Bronchus Cancer Incidence, COVID-19, Greenhouse Gases Emissions, Drinking Water and Air Quality Non-Compliance, Non-Pharmaceutical Interventions, Urbanization, Energy Use, Spillover Effects, Spatial Econometrics, Dynamic Panel Model.

ABSTRACT

THREE ESSAYS ON HEALTH AND ENVIRONMENTAL ECONOMICS:

APPLICATIONS OF SPATIAL ECONOMETRICS AND SPATIAL ANALYSIS

Mohammed Syedul Islam

Spatial interaction and the locational structure between observations play key roles in the field of econometrics for both cross-sectional and panel data analysis. Compared to a non-spatial econometric model, a spatial model relaxes the assumption of independency in observations. This research applies spatial and non-spatial econometrics in three different fields of applied economics: (1) drinking water and air quality violations impacts on lung and bronchus cancer incidence in the contiguous United States (U.S.); (2) spillover effects of non-pharmaceutical interventions (NPIs) on COVID-19 cases across the contiguous U.S. counties; and (3) urbanization impacts on carbon dioxide (CO_2) emissions in selected 119 countries.

In Chapter 2, ordinary least squares (OLS) and Spatial Durbin Model (SDM) are applied to data from 48 states plus Washington D.C. in the contiguous U.S. for the period of 2006-2016 to examine the impacts of population-level exposures to environmental quality standards non-compliance on the lung and bronchus cancer incidence. The SDM reveals statistically significant impacts of population-level exposures to violations of environmental pollution standards on lung and bronchus cancer incidence. While impacts are statistically significant (direct effect for water and total effect for air), they are small relative to smoking behavior. Example calculations show that a 10% reduction in population exposure rate to drinking water quality violations across the state of Oklahoma results in a decrease of five cancer cases annually with an estimated annual monetary benefit of $20.6 million. A 10% reduction in population exposure to air quality violations in the state of Utah results in six fewer cancer cases annually with a $24.7 million annual monetary benefit in Utah and three-neighboring states.

Chapter 3 examines the spatial spillover effects of NPIs policies on reductions of COVID-19 cases in the contiguous U.S. Using annual cross-sectional data for the year 2020, I apply a spatial Durbin model (SDM) to find statistically significant spillover effects from stay-at-home mandatory orders and mask mandates on COVID-19 cases per 100,000 people. Nationally, on average, 4 cases per 100,000 people can be reduced within the mandate county while an additional 9 cases per 100,000 people can be reduced in 6-nearest neighboring counties as the spillover effect from implementing a mask mandate policy for a month. On the other hand, the direct effect of mandatory stay-at-home orders is 15 cases per 100,000 people while the spillover (indirect) effect is 38 cases per 100,000 people when this NPI is implemented for one month. Thus, mandatory stay-at-home orders show a greater impact on reducing COVID-19 case rates than mask mandates. Further, the indirect effect in neighboring counties is larger than the direct effect for both NPI policies, showing the importance of accounting for spillover effects when determining NPI policy benefits. Based on the SDM model, an example of mask mandate and mandatory stay-at-home policies in Powder River County in Montana (as a mandate state) with three neighboring non-mandate counties in the state of Wyoming is examined. A month-long implementation of mask

mandate, stay-at-home order, and both these policies contributes \$0.01 million, \$0.04 million, and \$0.05 million of direct benefits, respectively from reduced COVID-19 costs in Powder River County in the state of Montana. The spillover benefits of about \$0.70 million, \$2.95 million, and \$3.65 million, respectively are estimated based on reduced COVID-19 costs in the three neighboring counties of Wyoming (Sheridan, Campbell, Crook) from one-month implementation of public mask mandate, mandatory stay-at-home orders and both these policies in Powder River County in the state of Montana. Most of this benefit is the result of reduced hospitalizations in the states of Montana and Wyoming. The results of this research will help to design cost-efficient interventions to tackle future pandemics.

Finally, Chapter 4 relates urbanization to greenhouse gases (GHG) emissions based on data from 105 countries in 1990-2018. Using pooled mean group (PMG) estimation technique, significant nonlinear relationships are found between urbanization and both carbon dioxide (CO_2) intensity and CO_2 emissions. For CO_2 intensity, urbanization impacts hit a turning point at 59.16 percent such that urbanization prior to this percentage increases CO_2 intensity while rates of urbanization rate above this percentage decrease CO_2 intensity. No such turning point is observed for CO_2 emissions. Based on the PMG model, renewable energy consumption significantly reduces both CO_2 intensity and emissions. However, at a given rate of renewable energy consumption, urbanization only reduces CO_2 emissions. Empirical results from this study highlight the importance of global scale action on urban buildings and transportation to reduce GHG emissions in both developing and under-developed countries with technical support from developed economies.

DEDICATION

My family (wife and two kids), who made my life meaningful and happy during this journey.

My father, who enlightened many youngsters in my village but could not see my doctoral degree because of his early demise.

My mother, who devoted most of her life to us.

For my elder brother and elder sister, who always inspired me to pursue higher degrees by means of their all efforts.

And, in absence of me, my younger sister who is taking care of my mother both mentally and physically.

ACKNOWLEDGMENTS

Firstly, I would like to express my profound gratitude to my advisor Professor Alan Collins for his continued support of my Ph.D. study and related research, for his patience, inspiration, and versatile knowledge. His guidance helped me in shaping my research and writing this dissertation. I could not think of completing this dissertation without his competent supervision. In addition to my advisor, I would like to thank other members of my dissertation committee: Professor Donald Lacombe, Professor Daniel Grossman, Professor Xiaoli Etienne, and Professor Michael Strager. I appreciate their thoughtful criticisms, encouragement, and questions, which motivated me to broaden my research from a number of angles. I would like to thank two anonymous referees for their insightful comments on second-year mentored research paper that I included as first essay in this dissertation.

Several courses helped me to shape my dissertation essays. I should be thankful to those course instructors, particularly Professor Levan Elbakidze (Economics of Water and Energy), Professor Heather Stephens (Advanced Applied Econometrics), Professor Brad Humphreys (Advanced Urban Economics), who helped me to develop my research questions, offered comments, and asked questions on my course paper. Later, I entertained those comments and questions in my dissertation.

I would like to thank comments from seminar participants at the Division of Resource Economics and Management where I presented my first essay. Also, I am grateful to participants at the Agricultural and Applied Economics Association (AAEA) Annual Meeting 2021, Southern Economic Association (SEA) Annual Meeting 2022, Social Costs of Water Pollution Workshop 2022, Spatial Econometrics Association Paelinck Seminar 2023, and the WVU Davis College of

Agriculture, Natural Resources, and Design Annual Conferences in 2021, and 2022 for their questions and comments.

I should acknowledge the help of Professor Brian Hendricks (School of Public Health, WVU) during the workshop on spatial cross-sectional econometrics. Special thanks to Professor Jacquelyn Strager who helped me in dealing with spatial data.

I appreciate the help from my fellow classmates, officemates, and my friend Maher Mekky for all their support during my Ph.D. journey.

Last but not the least, I would like to thank Lisa Lewis for all her support at the Division of Resource Economics and Management.

TABLE OF CONTENTS

Contents

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1:

BACKGROUND AND SIGNIFICANCE

1.1 Introduction

For both cross-sectional and panel data studies, spatial interaction and the locational structure between observations have recently drawn increased attention in the field of econometrics. Spatial econometrics specifies, estimates, and tests how the magnitude of a variable of interest would be determined by the value of the same variable at other locations in the system, given a location to any observation in the system (Anselin, 2001). A spatial economic model relaxes the assumption of observational independence when compared to a non-spatial economic model. In this regard, zip codes, cities, municipalities, counties, states, and countries could be considered spatially correlated observations (Elhorst, 2014). In this dissertation, spatial spillovers are estimated for topics related to water, air, and health by applying different level of spatially correlated observations. The impact of urbanization on carbon dioxide (CO_2) emissions is estimated by applying dynamic panel data models.

1.2 Purpose of this study

The overarching aim of this study is to empirically exhibit, at the state, county, and country levels, how spatial and dynamic panel data analysis may help in finding more accurate results when evaluating the effects of public policies, shocks, and exogenous variables on health and environmental outcomes. This study consists of three essays.

Aim of Essay 1: Examine the impacts of water and air quality standards non-compliance on lung and bronchus cancer incidence in the contiguous United States by applying a spatial econometric model.

The first essay describes the policy-related environmental factors that explain lung and bronchus cancer incidence in 48 contiguous states plus Washington D.C.

The objectives for this essay are listed below:

1. The main objective of this research is to examine the impacts of non-compliance for drinking water and air quality standards (as policy-related environmental factors) on lung and bronchus cancer incidence in the contiguous United States.

2. As part of state-level calculations, this research assesses the impacts of percentage reduction in population exposure to drinking water and air quality violations on cancer cases within the state boundary and neighboring states.

3. Finally, example calculations are made for annual monetary benefits from increased compliance for drinking water and air quality.

Aim of Essay 2: Examine spillover effects of NPIs on the reduction of COVID-19 cases in the contiguous U.S. counties.

The second essay focuses on state and county-level variations in adopting NPIs and therefore the COVID-19 spread across the contiguous U.S. Among the NPIs, stay-at-home orders, mask mandate, restaurant and bar closure, and reopening were dominant interventions during the COVID-19 pandemic. While many states and counties adopted those policies to restrict the spread of COVID-19, other states and counties did not implement those policies. This essay investigates the spatial spillover effects of NPIs on the reduction of COVID-19 cases in contiguous U.S. counties. Since COVID-19 is an infectious disease, the prevalence of COVID-19 in one county may affect its neighboring counties because of the cross-border mobility of people (WHO, n.d.).

Therefore, this essay applies a spatial econometric model with objectives of:

1. Estimate the spillover effects of implementing NPIs such as stay-at-home orders, mask mandates, restaurant, and bar closure, and reopening for one month on the reduction of COVID-19 cases in the contiguous U.S. counties.

2. Calculate monetary benefits from reducing COVID-19 spread in the county where policies are in place as well as neighboring counties.

Aim of Essay 3: Examine the relationship between country-level urbanization and GHG emissions: Application of dynamic panel data approach.

The third essay investigates the non-linear relationship between urbanization and GHG emissions in selected 105 countries. Over the period of 1990-2018, this essay applies pooled mean group (PMG) estimate technique and find a statistically significant long-run relationship between urbanization and CO_2 intensity in sample countries. This study also finds an inverted U-shaped relationship between urbanization and CO_2 emissions. While many researchers investigate the effects of short-term and long-term relationships and non-linear relationships, this study estimates the interaction term of urbanization and consumption of renewable energy. Therefore, using PMG model, the objectives are to:

1. Estimate the short-term and long-term impacts of urbanization on GHG emissions across countries.

2. Test the non-linear relationship between urbanization and GHG emissions in the sample countries

3. Estimate the interaction effect of urbanization and renewable energy use on GHG emissions in the sample countries.

References

Anselin, L. (2001). *Spatial econometrics. A companion to theoretical econometrics*, 310330.

Elhorst, J. P. (2014). *Spatial econometrics: from cross-sectional data to spatial panels* (pp. 20-25). Heidelberg: Springer.

WHO (n.d.). *Coronavirus disease (COVID-19)*. Retrieved from https://www.who.int/health-topics/coronavirus#tab=tab_1

CHAPTER 2:

ESSAY 1: THE IMPACT OF NON-COMPLIANCE FOR AIR AND WATER QUALITY STANDARDS ON LUNG AND BRONCHUS CANCER INCIDENCE: A SPATIAL PANEL DATA ANALYSIS

2.1 Introduction

Cancer is the second-leading cause of death among Americans (Heron, 2019). In 2020, the projected number of deaths from cancer was over 600,000 in the United States (U.S.) (American Cancer Society, 2020). Based on 2020 estimated new cases of cancer, the second most common type of cancer (after breast cancer) is lung and bronchus cancer. Although each type of cancer has its own set of risk factors, certain age groups, usage of alcohol and tobacco, exposure to UV radiation and excessive sunlight, chronic inflammation, obesity, and a history of cancer among family members make carcinogenic compounds more likely to develop into cancer in some people (National Cancer Institute, 2015). Empirically, factors identified as risk factors for lung cancer have included genetic factors (age, gender, family and ethnic background), behavioral factors (smoking, alcohol, diet and food supplements, poverty), and environmental factors (passive smoking, air pollution, radon, asbestos, occupational exposure, urban/rural residence). (De Groot and Munden, 2012; Doll, 1978; Grant, 2009; Speizer, et al., 1999).

Lung and bronchus cancer involve both direct and indirect costs to the individual suffering from this cancer as well as society (Yabroff, et al., 2011; Kaye, et al., 2018; Lokhandwala, et al., 2017; Mausbach, et al., 2018). Park and Look (2019) found a $16,346 annual U.S. mean healthcare expenditure per person for cancer patients which is four times higher than non-cancer patients. They also found lung cancer to be the most expensive type of cancer out of the four most common cancers. Based on Surveillance, Epidemiology and End Results (SEER)-Medicare data for the period of 1991-2003, it was estimated that a 72-year-old lung cancer patient pays $2,687 to $9,360 per month for the first 6 months of care depending on the stage of diagnosis and histologic type (Cipriano, et al., 2011). Conducting survey among patients aged 66-99 years, Kaye et al. (2018) estimated the annual Medicare payments for lung cancer at over $20,000 for patient.

In dealing with environmental factors related human health impacts like cancer, historically the U.S. federal government has adopted drinking water and air quality policies to improve the health status of Americans. For example, the Safe Drinking Water Act (SDWA) was enacted in 1974 and later amended in 1996 to protect people from naturally occurring and/or man-made contaminants that may be found in drinking water. As one example, in 2001, the U.S. Environmental Protection Agency (U.S. EPA) strengthened the standards, lowering the allowable amount of arsenic in drinking water from 50 ppb (parts per billion) to 10 ppb (or 0.01 milligram per liter). Under the SWDA, the U.S. EPA has set maximum contaminant level (MCL) for asbestos at 7 million fibers which is longer than 10 μm (Office of Federal Register, 2011), and for lead at 15 μg/L in public water systems (CDC, 2019).

Similarly, for air quality, under the Clean Air Act (CAA) of 1970, National Ambient Air Quality Standards (NAAQS) were established and set by the U.S. EPA. For outdoor air quality standards, the U.S. EPA determines both primary and secondary standards of fine particulate matter ($PM_{2.5}$). The primary annual $PM_{2.5}$ standard is designed to protect public health from short-term and long-term exposures while the 24-hour $PM_{2.5}$ standard is designed to work with annual $PM_{2.5}$ standard to supplement health protection from short-term exposure. The primary annual $PM_{2.5}$ standard is currently set at 12 $\mu g/m^3$ and the 24-hour $PM_{2.5}$ standard to 35 $\mu g/m^3$.

Both water and air quality non-compliance depend on several factors. Using data on drinking water quality non-compliance by public water system, Rahman et al. (2010) identified firm size, ownership type, and the location where the non-compliance occurred as major causes of non-compliance. Similarly, violating air quality standard has been shown to result in a higher number of deaths (Crosignani et al., 2021). Thus, controlling pollution non-compliance is important to improve human health outcomes.

Previous studies have focused on cancer rates and the link with air pollution and other non-environmental risk factors (Molina et al., 2008). There are fewer studies on the linkages between water contamination and cancer incidence (Evans et al., 2019; Morris, 1995; Stoiber et al., 2019). Some studies have highlighted the effect of arsenic contaminant in water on lung cancer or other types of cancer, however, they have not considered other water contaminants in their studies (Shiber, 2005; Heck et al., 2009). There are, however, no studies which have examined the human health implications of non-compliance with air and water quality standards. In addition, excess pollution moves from one location to another location when violation occurs over the NAAQS for air, and the MCL for water. Moreover, spatial distribution of population exposure to air and water quality non-compliance follows cluster patterns across the contiguous U.S. Based upon our review of the literature, prior studies have applied strictly non-spatial econometric techniques to estimate the impact of environmental factors on lung and bronchus cancer incidence.

The aim of this research is the impacts of water and air quality standards non-compliance on lung and bronchus cancer incidence in the contiguous United States by applying a spatial econometric model. As the first objective, this research examines the impacts of non-compliance for drinking water and air quality standards (as policy-related environmental factors) on lung and bronchus cancer incidence in the contiguous United States. Secondly, as part of state-level calculations, this research assesses the impacts of percentage reduction in population exposure to drinking water and air quality violations on cancer cases within the state boundary and neighboring states. need to estimate spillover effects of water and air quality non-compliance in neighboring states. Finally, example calculations are made for annual monetary benefits from increased compliance for drinking water and air quality. Utilizing spatial economic techniques, this research will fill this gap in the literature by estimating the local and regional effects of non-compliance with air and water quality standards on lung and bronchus cancer incidence in the U.S.

While statistically significant impacts are found for both water and air quality standard non-compliance, their effects on lung and bronchus cancer incidence are small. Based on spatial Durbin model (SDM), this research finds that a 10% decrease in annual exposure rate to drinking water quality violations from arsenic, lead, or asbestos contamination reduces the annual number of cancers by only five cases within the entire state of Oklahoma. A similar calculation is made for the state of Utah for a 10% reduction in annual exposure to $PM_{2.5}$ contamination violations results in a regional (Utah and three neighboring states) number of cancers being reduced by only six cases. The estimated annual monetary benefits from this reduced population exposure are $20.6 million and $24.7 million for water and air, respectively. In comparison, a similar magnitude of reduction in smoking in the population results in cancer incidence reductions of 16 to 46 times larger than non-compliance with environmental standards. These results show the need for improvement in compliance with drinking water and air quality standards as well as policies to reduce smoking.

In what follows, literature is reviewed in section 2, theory and econometric model are described section 3, data and methods are discussed in section 4, results are discussed in section 5, concluding words and policy recommendations are provided in section 6.

2.2 Literature Review

Previous literature has identified both environmental and non-environmental factors that lead to the development of lung cancer (Molina et al., 2008; Doll, 1978; Grant, 2009). While smoking was found to be the dominant risk factor for lung cancer, this type of cancer is also affected by other environmental as well as non-environmental factors. Water and air pollutants as policy-related physical environmental factors are discussed in the following sub-sections.

2.2.1 Water Pollution

Source-specific (e.g. agricultural chemicals and hazardous waste) water contamination is a contributing factor to increased risk of cancer, particularly due to arsenic, disinfection byproducts, lead, and nitrate contamination (Ebenstein, 2012; Frederick et al., 2016; Steenland and Boffetta, 2000). Epidemiological studies suggest that arsenic shows the strongest evidence (over all water pollutants) of substantial risk of lung cancer (Wu, et al. 1989). Based on the spatial distribution of cancer rates and drinking water contamination, Morris (1995) also claimed arsenic as the strongest cause of lung, liver, bladder, and kidney cancers in the US. For example, Shiber (2005) found 1 in 333 individuals who were exposed to water arsenic contamination at the MCL of 10 ppb had an increase in their lifetime risk of bladder and lung cancer. Based on cumulative risk analysis of contamination over the period of 2010–2017, several studies show that over 100,000-lifetime cancer cases are due to carcinogenic chemicals in tap water in the US (Evans et al., 2019; Stoiber, et al., 2019). Based on cohort studies, it is evident that lead cause an increased risk of lung cancer though the link is weak (Steenland and Boffetta, 2000; Lundstrom et al. 1997). Lead chromate is also responsible for bronchial cancer (Xie et al. 2005). Asbestos equally has a latency effect in causing human cancer, especially lung cancer (Torato et al. 2019). Although these studies utilized long-term average contaminant concentration in a community water system (CWS) and lifetime cancer risk from each contaminant concentration, what they lack is a consideration of spatial spillover effect in developing lung and bronchus cancer incidence as a result of less exposure to water contamination violations.

2.2.2 Air Pollution

Smaller particulate matter such as $PM_{2.5}$ has a strong association with lung cancer (Vinikoor-Imler et al., 2011; Raaschou-Nielsen et al., 2016). Using various type of models, it also has been documented that people get more exposed to $PM_{2.5}$ than any other air pollutant (Kersey and Yin, 2020; Diao et al., 2019). Based on residential exposure to three major air pollutants O_3, NO_2, and $PM_{2.5}$ from Canadian population-based case-control study, Hystad, et al (2013) found $PM_{2.5}$ as the strongest risk factor of lung cancer incidence. Because $PM_{2.5}$ particles are so small (compare to other pollutants) the human body cannot filter it out (Li et al., 2018; Huang et al., 2017). Particles even smaller than $PM_{2.5}$ in size enter the alveoli in the lung and negatively affects gas exchange within the lung (Hu and Jiang, 2014). Moreover, the interaction among those pollutants is complex and varies across seasons (Ito et al., 2007).

Setting ambient air quality standards is crucial for both benefits and costs to the society (Currie and Walker, 2019). Studies show that air pollution causes direct and indirect costs to the society due to increased hospitalizations and pre-mature deaths (Alberini et al., 2004; Chen et al., 2018; Zhu et al., 2019). Wind speed and direction further exacerbate pollution that accelerates the economic costs (Deryugina et al., 2019; Anderson, 2019; Yang and Chou, 2018). Pollutants such as ground-level ozone (O_3), fine particulate matter ($PM_{2.5}$ and PM_{10}), nitrogen dioxide (NO_2), and sulfur dioxide (SO_2) have been found to increase the incidence of cancers (Kim et al., 2018; Grant, 2009; Molina et al., 2008; Eckel et al., 2016; Fann et al., 2011). Study shows that about 129 to 354 lives can be saved from lung cancer each year by improving air quality primarily $PM_{2.5}$ concentration (via reduced fracking activity) which is equivalent to $1.2 to $3.3 billion (Johnsen et al., 2019). Thus, the air quality plays a crucial role in determining lung cancer incidence in the U.S.

2.2.3 Non-compliance

Theoretically, optimal pollution limits and non-compliance have been examined in prior literature (Arguedas, 2008; Arguedas et al., 2016; Arguedas et al., 2020). Optimal pollution level depends on a standard that minimizes the sum of abatement costs and external damages, the minimum probability required to induce compliance and a linear gravity sanction (Arguedas, 2008). Any highly stringent pollution standard and progressive penalties lead to policy non-compliance. When a fine is introduced for non-compliance, both pollution and non-compliance decrease as polluter's capital stock increases (Arguedas, 2016). However, discounting this fine for polluter's capital investment is socially desirable. Similarly, no effect is found on polluter's behavior if a linear and progressive fine is imposed on stock pollution (Arguedas et al., 2020).

Empirically, environmental non-compliance has been identified as a cause of increased health hazards in neighboring communities. A study on gold mining pollution in Ghana, where lack of environmental regulation exists, found a positive association between population exposure to NO_2 concentration from mining and healthcare expenditure (Akpalu and Normanyo, 2017). Mu, Y., Rubin, E. A., and Zou, E. (2021) constructed a framework to identify whether local governments avoid air pollution monitoring when they expect air quality to deteriorate. Based on all 1,359 monitors in Jersey City Firehouse in NJ, they found a 33% reduction of this monitor's sampling rate on pollution-alert days. The majority of literature has focused on theoretical aspects of non-compliance so that no single study has highlighted the impacts of water and air quality non-compliance on human health.

2.3 Theory and Econometric Model

'Health status' depends upon four broad categories of factors, health behaviors, clinical care, social and economic factors, and physical environment (UWPHI, 2022). Early research by Grossman (1972) and Wagstaff (1986) proposed a health production function (i.e., individual health behavior) from a microeconomic viewpoint, where individual implicitly engages in tradeoffs between 'health' against all other aims. The tradeoff between objectives of 'consumption' and 'health' can be described by the conventional utility maximization problem. By solving this problem, 'health' serves as an input to generate income, therefore helps a consumer to purchase more consumption goods. While the Grossman model states that age, education, and income determine the production of health through the demand for health capital, other economists estimate exposure-response functions linking different environmental compounds to health outcomes.

A health production function, H, is defined following Phaneuf and Requate (2017):

$$H = f(Q, X, A; S) \quad \ldots\ldots (2.1)$$

where, H is a binary variable describing individual health status (e.g. good or bad), Q is physical environment (e.g. ambient environmental quality), X represents a vector of clinical care (e.g. medical service), A is individual's health behavior (e.g. exercise), and the function is conditioned on individual's socioeconomic characteristics S. Based on this health production function, individuals can maximize utility given his/her earned income.

The aggregate health production can be derived by adding health production functions of all individuals in a society. In fact, physical environment such as ambient environmental pollution is exposed to a group of population; rather than individual. It is also assumed that individuals within the state are not equally exposed to water and air pollution which depends on many other factors. Therefore, moving from an individual health production function to a population perspective, an aggregate health production function can be represented by the following equation (Peter et al., 2009):

AH= f(PQ, Q, X, A, S, R)…………………………….……………………………………(2.2)

where AH denotes aggregate health status (e.g. age-adjusted lung and bronchus cancer incidence), PQ represents variables related to compliance environmental policies dealing with water and air quality standards, Q stands for non-policy-related environmental impacts on the population health (e.g. precipitation and other climatic factors), X represents clinical care available to the population (e.g. access to healthcare, quality of healthcare), A is behavioral factors across the population (e.g. cigarette consumption, alcohol consumption), S represents socioeconomic status within populations (e.g. education, income), and R represents demographic (e.g. age, gender) along with genetic factors (e.g. family and ethnicity) of the population.

This research utilizes an aggregate health production function approach to derive an econometric model to estimate the influence of environmental policy non-compliance as a determinant of health. Age-adjusted lung and bronchus cancer incidence (LUNCAN) is utilized to measure aggregated health among the population with the production function as follows:

LUNCAN = f(WATER QUALITY, AIR QUALITY, PRECIPITATION, SMOKE, ALCOHOL, GENDER, ETHNICITY, INCOME, OBESITY, UNINSURED)……………………….….(2.3)

In equation (2.3), WATER QUALITY and AIR QUALITY represent policy-related measures which show the level of non-compliance with drinking water and air quality standards, precipitation (PRECIPITATION) represents non-policy-related environmental factors, OBESITY, SMOKE and ALCOHOL represents behavioral factors, GENDER and ETHNICITY represents demographic and genetic factors, household median income (INCOME) represents socioeconomic factors, and UNINSURED measures the clinical care.

While family genetic background has become a growing concern in regard to lung and bronchus cancer incidence, genetic backgrounds are not available so that these backgrounds will be assumed to remain randomly distributed throughout the states over time (Wakelee, H.A. et al. 2007). Although data on per capita expenditures for all healthcare are available, per capita healthcare expenditures specifically for lung and bronchus cancer are not available. Per capita healthcare expenditures are not included due to potential correlation with household median income. State and year fixed effects are included in the estimation process to account for unobservable and/or lacking data factors in the model.

2.4 Methodology

2.4.1 Data

Data include the years from 2006 to 2016 for forty-eight contiguous states plus District of Columbia (D.C.) Non-compliance for drinking water quality is measured by population exposure to three dominant water contaminants such as arsenic, lead and asbestos that causes lung and bronchus cancer incidence. These data are publicly available for free in U.S. EPA Freedom of Information Act (FOIA) website (tracking ID: EPA-HQ-2018-006943). Number of violations by community water systems (CWS) and population served count by CWS can be found from this dataset. Violations data were filtered out based up on individual contaminant such as arsenic (contamination code 1005), asbestos (contamination code 1094), and lead (contamination code 1030) at the state level. Irrespective of health-based violations, the U.S. EPA FOIA published drinking water quality data for quarter 3. Within the state, data for populations exposed to the above contaminants from each violation are summed by:

Annual Exposure to Water Quality Violations (arsenic, lead, asbestos)$_{i,s,t}$ = $\sum_{1}^{n}$*population served count*... *(2.4)*

In equation (2.4), i stands for episode 1, 2,…n, s stands for state, t stand for year. For two years, however, (2004 and 2005), these exposure data are not available. Thus, in these two years, the number of violations per CWS are multiplied by the CWS populations exposed to those violations and then summed those population exposures for entire states to derive annual exposure to water quality violations at each state. Table 2.1 exhibits the variables and their sources considered in this study.

Table 2.1: Definition of Variables, 2006-2016, by State, USA.[a]

Variable	Definition	Source of data
(a) Dependent variable:		
LUNCAN	Age-adjusted incidence rate of lung and bronchus cancer per 100,000 population	National Environmental Public Health Tracking Network (NEPHTN)
(b) Environmental Policy Non-Compliance factors:		
AIR QUALITY[b]	Population exposure rate per 100,000 people to $PM_{2.5}$ over NAAQS[c], as a measure of population exposure to $PM_{2.5}$ level	National Environmental Public Health Tracking Network (NEPHTN)
WATER QUALITY[b]	Population exposure rate per 100,000 people to water quality[d] violations[e] by CWS over MCL[f], as a measure of drinking water contamination	US EPA Freedom of Information Act (FOIA)[g]
(c) Physical Environment factor:		
PRECIPITATION	Average number of precipitation days above 0.01 inches	National Environmental Public Health Tracking Network (NEPHTN)
(d) Behavioral factors:		
SMOKE	% of current or former smokers	BRFSS Prevalence & Trends data
ALCOHOL	% of adults who have had at least one drink of alcohol within the past 30 days	BRFSS Prevalence & Trends data
OBESITY	% of adults aged 18 years and over who were obese (BMI=30.0 to 99.8)	National Environmental Public Health Tracking Network
(e) Demographic & Genetic factors:		
GENDER	% of male	BRFSS Prevalence & Trends data
ETHNICITY	% of Non-White	BRFSS Prevalence & Trends data
(f)Socioeconomic factor:		
INCOME	Real median household income (2020 dollars)	US Census Bureau
(g) Accessibility to healthcare:		
UNINSURED	% of adults aged 18-65 who does not have any kind of healthcare coverage	BRFSS Prevalence & Trends data

Note:
[a]This study is based on 48 contiguous U.S. states plus D. C.
[b]WATER QUALITY and AIR QUALITY have been derived from 3-year moving average.
[c]NAAQS stands for National Ambient Air Quality Standard.
[d]Water quality contaminants represents Arsenic, lead, asbestos contamination.
[e]For some states in some years, no water quality violations were shown in the dataset, therefore, I considered zero '0' observations for those years in the sense that population did not expose to water quality violations above the MCL in those years. This does not mean that the population exposure below the MCL have not been occurred in those states for those years.
[f]MCL stands for Maximum Contamination Level.
[g]For water quality data, please check the following Tracking# EPA-HQ-2018-006943

Over time, population size from one state to another varies. Instead of using crude incidence rate which is influenced by specific age group of the state population, this study has used the age-adjusted incidence rate because the variation in this rate over time or across geographical locations is not influenced by the changes in age-distribution of populations being compared. Thus, the dependent variable LUNCAN stands for age-adjusted lung and bronchus cancer incidence per 100,000 people. Moreover, prior studies also used 'age-adjusted incidence/mortality rate' instead of 'number of cases' as a dependent variable (Hendryx et al., 2008; Okunade and Karakus, 2003). Since the outcome variable is population-weighted, 'population size' has not been used as a regressor in this study.

The main variables of interest showing environmental policy non-compliance are measures of drinking water and air quality violations. Although many contaminants (e.g. arsenic, disinfection byproducts, lead, nitrate etc.) are found in drinking water, three contaminants (arsenic, lead, and asbestos) have been identified as the leading causes among water contaminants in the development of lung and bronchus cancer in the U.S. (Morris 1995; Steenland and Boffetta, 2000; Lundstrom et al. 1997; Totaro et al. 2019). Therefore, three major water contaminants over the MCL are included to measure drinking water standards non-compliance. The variable WATER QUALITY stands for population exposure rate to MCL violations per 100,000 people served by CWS.

Similar to water pollution, multiple contaminants, such as fine particulate matter ($PM_{2.5}, PM_{10}$), ground-level ozone, and nitrogen dioxide, are found in the air with $PM_{2.5}$ being the dominant indicator of lung and bronchus cancer incidence (Vinikoor-Imler et al., 2011; Raaschou-Nielsen et al., 2016). In terms of population exposure to air pollution, data utilized are person-days exposed to $PM_{2.5}$ over national ambient air quality standard (NAAQS) based on monitored $PM_{2.5}$ concentration. County level population exposure data were accumulated from the National Environmental Public Health Tracking Network (NEPHTN) database to derive a state level population exposure. Populations located outside the monitoring activities included in NEPHTN were not included in this analysis. The variable AIR QUALITY stands for population exposure rate per 100,000 people to $PM_{2.5}$ violations over the NAAQS.

Since the distributions of population exposure to drinking water and air quality violations data are right-skewed, the following data generating process was applied. At the first step, population exposure rate to water and air quality violations were calculated. Since about thirty-eight percent and thirty-three percent of total observations are found zero '0' values in water and air quality variables, respectively in the data set, a three-year moving average is applied to both variables. Figures 2.1A and 2.2A in appendix shows how right-skewness can be reduced by taking three-year moving average for water and air quality variables. Since population exposure due to excess pollution from drinking water and air quality violations is a long-term issue, three-year moving averages can also capture some of this long-term exposure.

Data are collected on control variables of smoking, alcohol consumption, gender, ethnicity, income, obesity, and uninsured population. For smoking, the Behavioral Risk Factor Surveillance System (BRFSS) provides a prevalence and trend database with four types of smokers: smokes every day, someday, former smoker, and never smoked etc. In this research, the smoking variable is derived by subtracting the percentage of never smoked from total respondents (out of 100 percent) in each year for each state. For drinking alcohol, there are four types of drinking habits reported in the BRFSS prevalence & trend database: alcohol consumption (adults who have had at least one drink of alcohol within the past 30 days), binge drinkers (males having five or more drinks on one occasion, females having four or more drinks on one occasion), chronic drinkers (adult respondent having more than 60 alcoholic beverages in the past month), heavy drinkers (adult men having more than 14 drinks per week and adult women having more than 7 drinks per week). The alcohol variable is based on 'respondents who have had at least one drink of alcohol last month' to test if moderate drinking brings any benefits to health. BRFSS prevalence & trend database provides percentage of both male and female population. The gender variable represents the percentage of male population (crude prevalence) in this research. Based on the question 'what is your race/ethnicity?', BRFSS prevalence & trends database classifies race into several categories: white, black, Hispanic, multiracial, other up to 2014, since then classifies as white (non-Hispanic), Black (non-Hispanic), Hispanic, American Indian or Alaskan native (non-Hispanic), Asian (non-Hispanic), Native Hawaiian or Other Pacific Islander (non-Hispanic), Other race (non-Hispanic), multiracial (non-Hispanic) etc. The ethnicity variable is derived by subtracting the percentage of white or white (non-Hispanic) category from total respondents (out of 100 percent) in each year for each state which is called as percentage of non-White population in this study. For the income variable, real median household income measured in dollars is used. The obesity variable reflects the percentage of adults whose BMI ranges from 30 to 99.8. Finally, for health insurance coverage of the population, I used the percentage of people not covered by any plan to examine how uninsured non-elderly population are impacted by lung and bronchus cancer incidence.

Table 2.2 shows the summary statistics of variables. Out of 539 observations, values of water quality, and air quality are highly fluctuating, while values of gender, obesity, smoking, uninsured vary only moderately. Since the minimum value of water quality and air quality are zero, a special data generating process has been applied to each observation of those variables as described before. For lung and bronchus cancer incidence equation, positive signs are expected for population exposure to water and air quality violations along with greater population exposure to water and air quality violations (Hystad, et al. 2013; Steenland and Boffetta, 2000; Shiber, 2005; Totaro et al. 2019). Similarly, positive signs are expected for variables representing: (a) the average number of precipitation days as more precipitation leads to higher rate of lung and bronchus cancer, (b) the percentage of smokers as the more years that an individual smokes along with the more cigarettes smoked per day, the higher the risk of lung cancer, (c) the percentage of male population, and (d) percentage uninsured among the population as the use of lung cancer screening test is associated with insurance coverage (Shah, et al. 2019; Tabatabai, et al. 2016; Slatore, et al., 2010; Doria-Rose, et al., 2012). On the contrary, a negative sign is expected for median household income (Hajizadeh, et al., 2020). Since prior literature has found both positive and negative impacts on cancer incidence, definite expectations for the impacts from variables on alcohol consumption and ethnicity, and adult obesity are not made (Freudenheim, et al., 2005; Lee, 2013; DeSantis, et al., 2019; Mavridis and Michaelidou, 2019). The following section describes the spatial distribution of the quality of air, water, and the lung & bronchus cancer incidence in the contiguous U.S. states.

Table 2.2: Summary Statistics, 2006-2016, 48 Contiguous States plus Washington D.C.

Variable	Mean	Std. Dev.	Min.	Max.	Expected sign of coefficient
Lung and bronchus cancer incidence age-adjusted rate (per 100,000)	64.22	12.11	25.10	101.90	
Population exposure rate (per 100,000 people) to water quality violations by CWS over MCL[a]	1432.76	4281.12	0.00	36571.67	+
Population exposure rate (per 100,000 people) to $PM_{2.5}$ over NAAQS[b]	125913	229908.9	0.00	1533453	+
Average number of precipitation days (above 0.01 inches)	134.66	27.86	39.00	204.00	+
Current or former smoker (%)	44.28	4.79	23.40	85.10	+
Alcohol consumption within the past 30 days (%)	52.93	9.07	25.00	68.80	+/-
Gender (% of male)	48.71	0.84	46.00	51.30	+
Ethnicity (% of non-white)	24.95	14.35	3.70	64.70	+/-
Median household income (in 2020 dollars)	60295.11	9306.02	38876	87590.00	-
Adult Obesity (%)	27.87	3.56	18.20	37.70	+/-
Uninsured (%)	17.16	5.63	5.00	35.80	+

Notes:
[a]Water Quality contaminants represent Arsenic, lead, asbestos contamination. MCL stands for Maximum Contamination Level.
[b]NAAQS stands for National Ambient Air Quality Standard.

2.4.2 Spatial Dependency in Cancer Incidence across States

The geographical distance is a key motivation for spatial spillover effects (LeSage, 2014). Before analyzing spatial dependency by statistical technique, let us identify clusters of lung and bronchus cancer incidence in the map, and link it with the clusters of water, and air quality in other maps. Figure 2.1 shows the population exposure rate (per 100,000 people) to $PM_{2.5}$ in micrograms per cubic meter ($\mu g/m^3$) above the NAAQS for the period of 2006-2016 in the contiguous U.S. In this figure, population exposure is higher in Southeastern and Ohio River valley portions of the U.S. while lower in the central and southern part of the U.S. Figure 2.2 shows the population exposure rate per 100,000 people (14-years average from 2004-2017) to arsenic, lead, and asbestos concentration in drinking water above the MCL in the contiguous U.S. In this figure, population exposure is mostly prevalent in south central and southwest portions of the country. Figure 2.3 exhibits lung and bronchus cancer incidence (5-year estimate from 2013-2017) per 100,000 people in the contiguous U.S. This figure shows that lung and bronchus cancer incidence is higher in the states along the lower Mississippi River, and lower in the western part of the country.

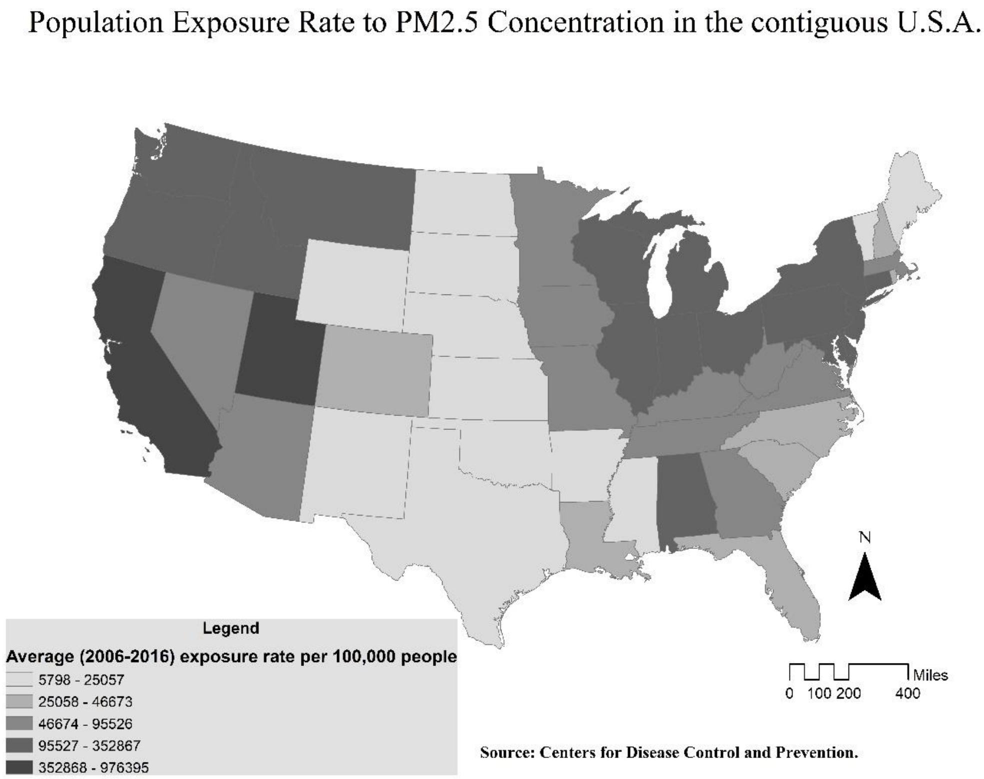


Figure 2.1: Population Exposure Rate by State to Airborne $PM_{2.5}$ Concentrations in the Contiguous U.S.

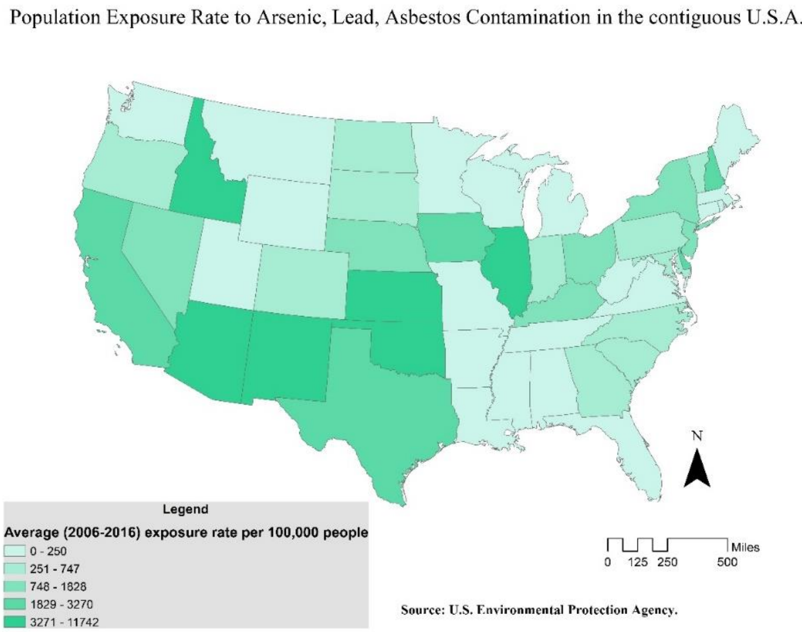


Figure 2.2: Population Exposure Rate to Water Contaminants of Arsenic, Lead, and Asbestos in the Contiguous U.S.

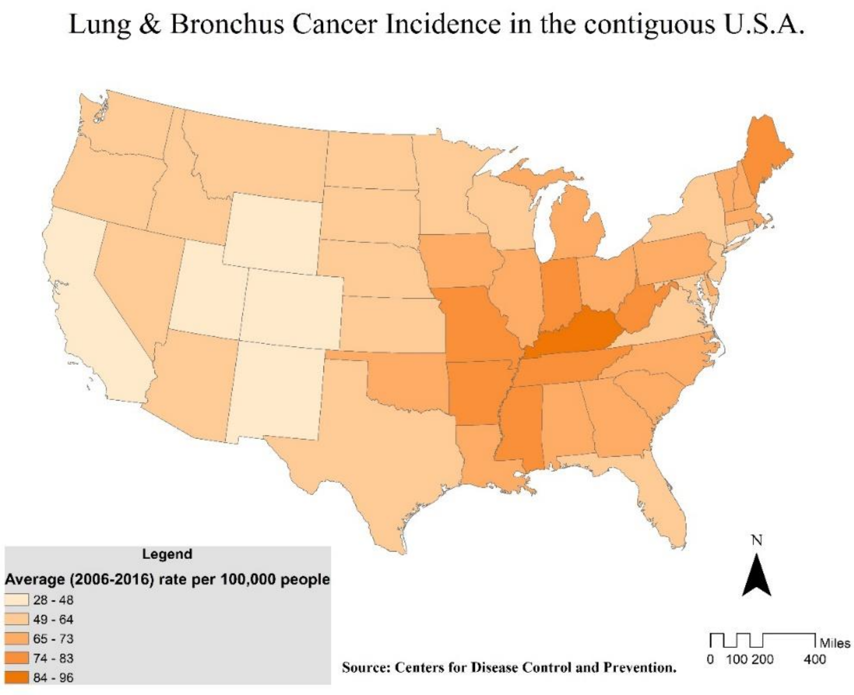


Figure 2.3: Lung and Bronchus Cancer Incidence in the Contiguous U.S.

Given the visual evidence of spatial relation between air quality and the incidence of lung and bronchus cancer in the U.S., the next step is to detect spatial autocorrelation. In this case, the spatial autocorrelation measures the interrelationship of lung and bronchus cancer incidence across neighboring states. The Global Moran's I index is a common measure that is used to detect spatial autocorrelation, which is based upon null hypothesis of spatial randomness, i.e. global Moran's I statistic is (Moran, 1948; Cliff and Ord, 1973):

$$I = \frac{\sum_i \sum_j w_{ij} z_i . z_j / S_0}{\sum_i z_i^2 / n} \quad \text{……………………………………………………………………(2.5)}$$

where, z_i is the deviations from the mean of the variable for any observation at location i, w_{ij} is the elements of spatial weights matrix, representing the spatial relationship between locations i and j, S_0=$\sum_i \sum_j w_{ij}$ is the sum of all the weights, and n is the number of observations. In this research, statistically significant and positive z-values for Moran's I index indicates a consistent pattern of positive spatial autocorrelation for lung and bronchus cancer incidence in forty-eight contiguous states plus D.C. for each year during the period of 2006-2016 (Table 2.3).

Table 2.3: Moran's I index for State Level Lung and Bronchus Cancer Incidence Rates.

Year	Moran's I	Pseudo p-value
2006	0.470	0.001
2007	0.511	0.001
2008	0.512	0.001
2009	0.574	0.001
2010	0.574	0.001
2011	0.533	0.001
2012	0.565	0.001
2013	0.543	0.001
2014	0.551	0.001
2015	0.569	0.001
2016	0.559	0.001

Figure 2.4 provides further evidence of spatial autocorrelation based on Moran scatter plots of lung and bronchus cancer incidence in 2006 and 2016. This scatter plots shows observations in four quadrants, where high value observations are surrounded by high value observations (i.e., Q1: HH). Figure 2.4 also shows that most of the states with high lung and bronchus cancer incidence

rate are adjacent to states with high incidence rates. This also happens for states with low lung and bronchus cancer incidence rates. Given this spatial autocorrelation, a three-nearest neighbors weight matrix is utilized based upon prior spatial econometric modeling (Lacombe and Flores, 2017).

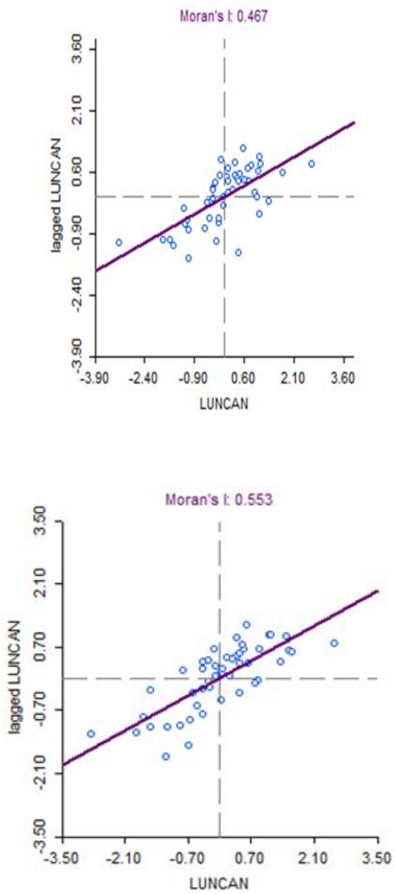


Figure 2.4: Moran's scatter plot of state-level Lung & Bronchus Cancers Incidence rates in 2006 and 2016.

As the Moran's I indices show statistically significant, positive spatial autocorrelation exists among states, performing the OLS estimations (non-spatial models) may lead to biased estimates of our regression results. Therefore, it is appropriate to apply spatial models to estimate the effects of environmental policy non-compliance on the incidence of lung and bronchus cancer across states in the U.S.

2.4.3 Spatial Econometric Analysis

There are five different spatial models (Elhorst, 2014). The first one is the spatial autocorrelation model (SAC) where the dependent variable as well as error term in neighbor *j* influences the dependent variable and the error term in neighbor *i* and vice versa. Second, the spatial lag model (SAR) assumes that only the dependent variable in neighbor *j* play a role in determining dependent variable in neighbor *i*. The spatial Durbin model (SDM) assumes that both dependent and control variables have spatial influence across neighbors, and error terms has no spatial influence at all, however, the spatial Durbin Error Model (SDEM) assumes that control variables along with error terms have spatial effect across neighbors. The spatial lag of X model (SLX) or spatial lag of control variables assumes that dependency is only exist in the control factors (WX). Finally, a spatial error model (SEM) assumes dependency in the error term only.

In prior literature, some researchers have applied spatial econometrics tools to capture the spillover effects of health outcomes across neighbors (Hoffer et. al. 2019; Chen et al., 2017; Weng et al., 2017). Thus, there are likely spillover effects from environmental policy non-compliance for both state and temporal variation on the incidence of lung and bronchus cancer. To evaluate the effects of drinking water and air quality violations on the incidence of lung and bronchus cancer in the U.S., SDM with spatial and time fixed effects is the spatial model applied here using a three-nearest neighbors weight matrix following previous state level health economics research (Alzahrani, et al., 2020). Equation (2.6) shows the age-adjusted lung and bronchus cancer incidence model using SDM state and year fixed effects model (Elhorst, 2014):

$$LUNCAN_{s,t} = \rho W * LUNCAN_{s,t} + \alpha l_N + \beta_1 WATERQUALITY_{s,t} + \beta_2 AIRQUALITY_{s,t} + \beta_3 PRECIPITATION_{s,t} + \beta_4 SMOKE_{s,t} + \beta_5 ALCOHOL_{s,t} + \beta_6 GENDER_{s,t} + \beta_7 ETHNICITY_{s,t} + \beta_8 INCOME_{s,t} + \beta_9 OBESITY_{s,t} + \beta_{10} UNINSURED_{s,t} + \theta_1 W * WATERQUALITY_{s,t} + \theta_2 W * AIRQUALITY_{s,t} + \theta_3 W * PRECIPITATION_{s,t} + \theta_4 W * SMOKE_{s,t} + \theta_5 W * ALCOHOL_{s,t} + \theta_6 W * GENDER_{s,t} + \theta_7 W * ETHNICITY_{s,t} + \theta_8 W * INCOME_{s,t} + \theta_9 W * OBESITY_{s,t} + \theta_{10} W * UNINSURED_{s,t} + \xi + \psi + \varepsilon_{s,t} \ldots\ldots\ldots\ldots(2.6)$$

In equation (2.6), ρ, and θ shows the spatial interactions of dependent variable, and variables of interest, ξ, ψ, and ε represent state fixed effects, year fixed effects, and disturbance term, respectively. The subscripts s and t stands for state and year, respectively. MATLAB software is utilized to estimate the SDM routines (LeSage, 2021).

2.5 Spatial Results

As discussed in the previous section, it is important to estimate spillover effects of population exposures to drinking water and air quality violations on the incidence of lung and bronchus cancer. Table 2.4 reports spatial Durbin model (SDM) estimates while Table 2.1A in the appendix reports the coefficient estimates for the OLS model. While the OLS model shows statistically significant coefficients for the variables AIR QUALITY and six of the control variables, with statistically significant rho (ρ) value in the SDM with spatial and time fixed effects model, these OLS coefficient estimates are potentially biased. OLS model results are presented in Table 2.1A to provide a comparison with the magnitude of direct effects from the SDM model. For example, the OLS coefficient for WATER QUALITY is statistically insignificant and about eight times smaller than the SDM direct effect which has a statistically significant p-value at the 10 percent level. On the other hand, the OLS estimate coefficient for AIR QUALITY is statistically significant and close to the magnitude of the SDM direct effect.

In this analysis, the direct effect of drinking water quality non-compliance and the total effect of air quality non-compliance on the total number of lung and bronchus cancers are computed and interpreted with examples. Based on 11 years (2006-2016) average population exposure to water quality violations, the highest population exposure to water quality violations was found in the state of Oklahoma. Assuming a 10% reduction in population exposure rate to arsenic, lead, or asbestos violations in this state, the number of lung and bronchus cancer cases decreases by five cases annually based upon the 2016 Oklahoma population. The percentage reduction in total age-adjusted lung and bronchus cancer incidence is 0.146% due to a 10% reduction in population exposure to arsenic, lead, or asbestos violations.

For air quality, Utah has the second highest proportion of population exposed to $PM_{2.5}$ violations during the data period. A 10% decrease in annual population exposure to $PM_{2.5}$ violations over the NAAQS in the state of Utah has spillover effects across the three-nearest neighbor states (Colorado, Idaho, Wyoming) reducing lung and bronchus cancer incidence by 0.77%. Based upon 2016 populations of these states, lung and bronchus cancer cases decline by six cases annually. Thus, based upon the SDM model, it is observed that drinking water quality non-compliance has a local effect while air quality non-compliance has the global effect (LeSage, 2014).

Table 2.4: Coefficients of Spatial Durbin Model (SDM) for 48 Contiguous U.S., 2006-2016.

Variables	**Dep. Variable: Lung and Bronchus Cancer Incidence**		
	Direct Effect	Indirect Effect	Total Effect
WATER QUALITY (pop. exposure rate)	0.000062*	-0.000064	-0.000002
	(0.054)	(0.206)	(0.961)
AIR QUALITY (pop. exposure rate)	0.000002	0.000001	0.000002**
	(0.128)	(0.603)	(0.034)
PRECIPITATION (days)	0.018	-0.00003	0.018*
	(0.238)	(0.999)	(0.083)
SMOKE (% of smokers)	0.092*	0.016	0.108
	(0.063)	(0.807)	(0.149)
ALCOHOL (% of adults)	0.129**	-0.161*	-0.031
	(0.036)	(0.070)	(0.716)
GENDER (% of Male)	2.191***	3.765***	5.956***
	(0.002)	(0.000)	(0.000)
ETHNICITY (% of non-white)	0.159***	-0.042	0.118
	(0.008)	(0.654)	(0.238)
INCOME ($)	0.00007*	0.00001	0.00008
	(0.057)	(0.857)	(0.189)
OBESITY (% of adults)	-0.055	0.033	-0.023
	(0.593)	(0.830)	(0.881)
UNINSURED (% of adults)	-0.043	0.022	-0.021
	(0.468)	(0.788)	(0.802)
Rho		-0.236***	
		(0.000)	
State FE		Yes	
Year FE		Yes	
R-squared		0.970	
Obs.		539	

Note: *** for 0.01, ** for 0.05, * for 0.1 Numbers in the parentheses represent p-values. Although state and year fixed effects were included in the SDM model, these coefficients are not reported.

Smoking as a behavioral factor leading to cancer has a positive, statistically significant direct effect on cancer incidence with p-value 0.063 in the SDM model, although indirect and total effects are statistically insignificant (Table 2.4). To compare smoking behavior with the impacts from environmental policy non-compliance for drinking water and air quality, I use a consistent 1% reduction in the magnitude of each of these explanatory variables. Thus, a 1% reduction in the percentage of former and current smokers (which translates into a 0.44% reduction) will result in a 0.041 person reduction in lung and bronchus cancer incidence per 100,000 population. Compared 1% reduction impacts from drinking water (0.00089 per 100,000) and air (0.0025 per

100,000), reductions in smoking have 46 to 16 times larger impacts on lung and bronchus cancer incidence than do reductions in exposure to environmental pollutants created by non-compliance with drinking water and air quality standards.

In terms of the other control variables, the INCOME variable does not fulfill our expectation of a negative impact based upon its direct effect in the spatial model. While most prior research finds a negative impact of household median income on cancer incidence, I find a statistically significant, positive direct effect from INCOME (though it is small in magnitude), which in part may be explained by life-style effects among high income groups. That is, less healthy eating habits (e.g. eating fast foods) among high income people leads to higher incidence of lung and bronchus cancer than among low-income people (Sugar, 2018; Scutti, 2018). The PRECIPITATION variable (total effects in the SDM model) fulfills our expectation that the lung and bronchus cancer incidence is positively impacted by the average number of precipitation days (above 0.01 inches). Although alcohol consumption has statistically significant, negative direct effects which is supported by Korte et al. (2002) and Freudenheim et al. (2005), the total effect is not statistically significant. Finally, OBESITY and UNINSURED have no statistically significant effects on cancer incidence.

The estimation results for the GENDER and ETHNICITY variables are notable. These variables have statistically significant estimates across models which switch signs moving from the non-spatial model OLS to the spatial SDM. This result is indicative of the potential bias in OLS coefficient estimates with spatial autocorrelated data. From the SDM model estimates, both male and non-white populations are more likely to experience lung and bronchus cancers, although the impact is much larger in magnitude for gender compared to ethnicity. These demographic category results conform with prior research by Tabatabai et al. (2016).

Prior literature on willingness to pay (WTP) estimate for cancer risk reduction is utilized to estimate the annual monetary benefits of reducing population exposures and preventing lung and bronchus cancer cases. First, an estimate of 2.6 million Euro per prevented case of cancer in Italy obtained from Tonin, et al., (2012) is adjusted to U.S. dollar values and then adjusted for inflation to 2016. A $4.1 million monetary value per cancer case prevented is then multiplied by the five and six annual reductions of cancer case. Thus, based on WTP estimates for reduced cancer cases from Tonin, et al. (2012), the annual monetary benefit of reducing population exposure to drinking water quality violations (arsenic, lead, asbestos) by 10% is an estimated $20.6 million within the state of Oklahoma. Also, for a 10% reduction in population exposure to air quality violations in the state of Utah, $24.7 million of annual benefits are estimated for this state plus the three-neighboring states.

Finally, the examination of lung and bronchus cancer incidence variable resulted in positive Moran's I indices, which explains spatial autocorrelation of this variable across neighboring states. However, a negative, statistically significant rho (ρ) value in the SDM model indicates that lung and bronchus cancer incidence in one state is negatively influenced by the incidence in neighboring states. Although positive, statistically significant rho values are well accepted in prior literature, a few studies have addressed the presence of a negative rho coefficient (Griffith, 2019; Kao, et al. 2016). For instance, Pavlyuk (2011) found a negative rho in SLM model, and Basdas (2009) found a negative lambda in SEM model, while Garret and Marsh (2002) found negative rho and lambda in SLM and SEM models, respectively. It is noteworthy that factors likely to result in positive spillover effects (such as violations of air and water pollution standards) have much smaller impacts on cancer incidence than other factors.

2.6 Conclusions and Policy Implications

While prior studies have focused on non-spatial econometric techniques and ignored spatial and non-compliance issues related to environmental policies' impacts on human health, this research incorporates spatial and time influences in order to estimate the impact of drinking water and air quality violations on the incidence of lung and bronchus cancer in the U.S. Based on a Spatial Durbin Model (SDM) with time fixed effects, this research finds positive, statistically significant intra-state (direct effect) impacts on cancer incidence from drinking water quality non-compliance as opposed to regional (total effects) impacts for air quality non-compliance. While statistically significant, the impacts from reductions in water and air quality non-compliance on cancer incidence are dwarfed in magnitude (16 to 46 times smaller) by a similar sized reduction in smoking behavior.

Example calculations are utilized to show that a 10% reduction in population exposure rate to drinking water quality violations within the state of Oklahoma results in a decrease of five cancer cases annually with an estimated annual monetary benefit of $20.6 million. A 10% reduction in population exposure to air quality violations in the state of Utah results in six fewer cancer cases annually with a $24.7 million annual monetary benefit in Utah and three-neighboring states. These monetary benefits are rough estimates since no WTP literature is available which is related directly to cancer risk reductions from reduced population exposure to drinking water and air quality violations.

The regulatory behavior of government can determine the aggregate health status of the population. Environmental justice issues arise from our results given that environmental quality non-compliance occurs more commonly in communities of color (Rahman et al. 2010; Switzer and Teodoro, 2017). In terms of $PM_{2.5}$ exposure, while the CAA contributed to reduce racial disparity between black and white communities over the past 20 years, this study confirms the continued existence of racial disparity in terms of lung and bronchus cancer incidence (Currie, et al. 2020). Besides ensuring water quality, environmental justice across race and class to CAA is thus equally important because loose enforcement of CAA (particularly to $PM_{2.5}$) may cause death (Konisky, 2009; Clay and Muller, 2019).

Since environmental enforcement involves costs, it is imperative to consider the resources utilized for enforcing drinking water and air quality standards versus resources allocated to preventing smoking among children and adolescents (Currie and Walker, 2019; Krutilla, 1999). The much larger impact of smoking on lung and bronchus cancer incidence shows the importance of more stringent policies to encourage non-smokers from ever starting smoking in order to ensure lower rates of lung and bronchus cancers. A recent study shows that some cancers (including lung cancer) are preventable through changing behavior such as control smoking, drinking alcohol and body mass index (Global Burden of Disease, 2022). Preventable cancers include those that stem from non-compliance with air and water quality standards- thus this research adds to the preventable cancers list that includes behavioral changes.

Anselin and Rey (1991) have argued that an appropriate choice of spatial weight matrix is important to estimate coefficients properly. As I mentioned that this research employs a three-nearest neighbors weight matrix to estimate the effect of drinking water and air quality violations on the lung and bronchus cancer incidence in the U.S. Instead of using a general homogenous spatial weight matrix, future research may apply a heterogenous spatial weight matrix based upon the annual prevailing wind direction map (Erfanian and Collins, 2020; Cheng et al. 2014; Chen and Ye, 2018).

In the U.S., CWS supplies drinking water that comes from surface as well as ground sources. Debate arises when CWS collects water from surface sources such as lakes and rivers as an appropriate spatial neighborhood does not always depend on geographical distance. Rather, neighborhoods may depend on hydrological pattern which is called hydrological distance between states, where some states are in upstream, and other states are in downstream (Peterson and Ver Hoef, 2010). Since drinking water quality violations has been included as a variable of interest in this study, future research may also use hydrologic unit code (HUC) to measure the location of states whether state i should be considered as a neighbor of state j to construct a heterogenous spatial weight matrix to get better results.

The causes and nature of cancer incidence is very complex in terms of population exposure, genetics, and other unobservable factors. In this regard, time lag considerations between population exposure to environmental pollution and the incidence of lung and bronchus cancer could not be included into the model. While prior study utilized longitudinal data to test the impact of air pollution (particularly $PM_{2.5}$ and nitrogen dioxides) on other health outcomes, future researchers could use longitudinal data and investigate the relationship between long-term exposure to environmental pollution, family genetic background and the incidence of lung and bronchus cancer in the U.S. (Ekstrom, et al. 2022).

References

Akpalu W, Normanyo AK (2017). Gold mining pollution and the cost of private healthcare: The case of Ghana. *Ecological Economics* 142: 104-112.

Alberini A, Cropper M, Krupnick A, Simon NB (2004). Does the value of a statistical life vary with age and health status? Evidence from the US and Canada. *Journal Environmental Economics and Management* 48(1): 769-792.

Alzahrani F, Collins AR, Erfanian E (2020). Water quality impact on health care expenditure in the United States. *Water Resources and Economics 32*: 100162.

American Cancer Society (2020). American Cancer Society: Cancer Facts and Figures 2020. Atlanta, Ga: American Cancer Society. https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2020/cancer-facts-and-figures-2020.pdf

Anselin L, Rey S (1991). Properties of tests for spatial dependence in linear regression models. *Geographical Analysis* 23 (2): 112-131.

Anderson ML (2019). As the wind blows: The effects of long-term exposure to air pollution on mortality. *Journal of European Economic Association* jvz051. https://doi.org/10.1093/jeea/jvz051

Arguedas C (2008). To comply or not to comply? Pollution standard setting under costly monitoring and sanctioning. *Environmental and Resource Economics* 41: 155-168.

Arguedas C, Cabo F, Martin-Herran G (2016). Optimal pollution standards and non-compliance in a dynamic framework. *Environmental and Resource Economics* 68: 537-567. DOI: 10.1007/s10640-016-0031-5

Arguedas C, Cabo F, Martin-Herran G (2020). Enforcing regulatory standards in stock pollution problems. *Journal of Environmental Economics and Management* 100: 102297.

Basdas, U. (2009). Spatial Econometric Analysis of the Determinants of Location in Turkish Manufacturing Industry. *NBER Working Paper Series*.

CDC (2014). *Best Practices for Comprehensive Tobacco Control Programs-2014.* Atlanta: US Department of Health and Human Services, Centers for Disease Control and Prevention. https://www.cdc.gov/tobacco/stateandcommunity/best_practices/pdfs/2014/comprehensive.pdf

CDC (2019). What are U.S. standards for lead levels? Centers for Disease Control and Prevention. https://www.atsdr.cdc.gov/csem/leadtoxicity/safety_standards.html

Chen X, Shao S, Tian Z, Xie Z, Yin P (2017). Impacts of air pollution and its spatial spillover effect on public health based on China's big data sample. *Journal of Cleaner Production* 142 (2): 915-925.

Chen X, Ye J. (2018). When the wind blows: spatial spillover effects of urban air pollution. International Association of Agricultural Economists 2018 Conference, July 28-August 2, 2018, Vancouver, British Columbia (No. 277146).

Chen S, Li Y, Yao Q (2018). The health costs of the industrial leap forward in China: Evidence from the sulfur dioxide emissions of coal-fired power stations. *China Economic Review* 49: 68-83.

Cheng T, Wang J, Haworth J, Heydecker B, Chow A (2014). A dynamic spatial weight matrix and localized space-time autoregressive integrated moving average for network modeling. *Geographical Analysis* 46 (1): 75-97.

Cipriano LE, Romanus D, Earle CC, et al. (2011). Lung cancer treatment costs, including patient responsibility, by stage of disease and treatment modality, 1992-2003. *Value Health* 14 (1): 41-52.

Clay K, Muller NZ (2019). Recent Increases in Air Pollution: Evidence and Implications for Mortality. National Bureau of Economic Research, Inc, *NBER Working Papers: 26381*.

Cliff A, Ord JK (1973). *Spatial Autocorrelation*. London: Pion.

Crosignani P, Nanni A, Pepe N, Pozzi C, Silibello C, Poggio A, Conte M (2021). The effect of non-compliance of diesel vehicle emissions with Euro limits on mortality in the city of Milan. *Atmosphere* 12 (3). https://doi.org/10.3390/atmos12030342

Currie J, Walker R (2019). What do Economists have to say about the Clean Air Act 50 years after the establishment of the Environmental Protection Agency? *Journal of Economic Perspectives* 33 (4): 3-26.

Currie J, Voorheis J, Walker R (2020). What caused racial disparities in particulate exposure to fall? New evidence from the Clean Air Act and satellite-based measures of air quality. *NBER Working Paper No. 26659.*

De Groot P, Munden RF (2012). Lung cancer epidemiology, risk factors, and prevention. *Radiologic Clinics of North America* 50 (5): 863–876.

Deryugina T, Heutel G, Miller NH, Molitor D, Reif J (2019). The mortality and medical costs of air pollution: Evidence from changes in wind direction. *American Economic Review* 109 (12): 4178-4219.

DeSantis CE, Miller KD, Sauer AG, Jemal A, Siegel RL (2019). Cancer Statistics for African Americans, 2019. *CA: A Cancer Journal of Clinicians* 69 (3): 211-233.

Diao M, Holloway T, Choi S, et al. (2019). Methods, availability, and applications of $PM_{2.5}$ exposure estimates derived from ground measurements, satellite, and atmospheric models. *Journal of the Air & Waste Management Association* 69 (12):1391-1414.

Doll R (1978). Atmospheric pollution and lung cancer. *Environmental Health Perspectives*, 22: 23-31.

Doria-Rose VP, White MC, Klabunde CN, et al., (2012). Use of lung cancer screening tests in the United States: Results from the 2010 National Health Interview Survey. *Cancer epidemiology, biomarkers & prevention* 21 (7): *1049–59.*

Ebenstein A (2012). The consequences of industrialization: evidence from water pollution and digestive cancers in China. *The Review of Economics and Statistics* 94 (1): 186-201.

Eckel SP, Cockburn M, Shu Y, et al. (2016). Air pollution affects lung cancer survival. *Thorax* 71: 891-898.

Ekstrom IA, Rizzuto D, Grande G, et al. (2022) Environmental Air Pollution and Olfactory Decline in Aging. *Environmental Health Perspectives* 130 (2).

Elhorst JP (2014). *Spatial Econometrics: From Cross-sectional Data to Spatial Panels*. Springer Briefs in Regional Science.

Erfanian E, Collins AR (2020). Air Quality and Asthma Hospitalization: Evidence of PM2.5 Concentrations in Pennsylvania Counties. *Journal of Regional Analysis & Policy* 50 (1): 1-15.

Evans S, Campbell C, Naidenko OV (2019). Cumulative risk analysis of carcinogenic contaminants in United States drinking water. *Heliyon* 5 (9): e02314.

Fann N, Lamson AD, Anenberg SC, et al. (2011). Estimating the national public health burden associated with exposure to ambient PM2.5 and ozone. *Risk Analysis* 32(1):81-95.

Frederick L, VanDerslice J, Taddie M, et al. (2016). Contrasting regional and national mechanisms for predicting elevated arsenic in private wells across the United States using classification and regression trees. *Water Research* 91: 295-304.

Freudenheim JL, Ritz J, Smith-Warner SA, et al., (2005). Alcohol consumption and risk of lung cancer: A pooled analysis of cohort studies. *American Journal of Clinical Nutrition* 82 (3): 657-667.

Garret, T. A. and Marsh, T. L. (2002). The revenue impacts of cross-border lottery shopping in the presence of spatial autocorrelation. *Regional Science and Urban Economics 32*: 501-519.

Global Burden of Disease (2022). The global burden of cancer attributable to risk factors, 2010–19: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet* 400: 563–91. https://www.thelancet.com/action/showPdf?pii=S0140-6736%2822%2901438-6

Grant WB (2009). Air Pollution in Relation to U.S. Cancer Mortality Rates: An Ecological Study; Likely Role of Carbonaceous Aerosols and Polycyclic Aromatic Hydrocarbons. *Anti-Cancer Research* 29:3537-3546.

Griffith, D. A. (2019). Negative Spatial Autocorrelation: One of the Most Neglected Concepts in Spatial Statistics. *Stats* 2: 388-415. https://doi.org/10.3390/stats2030027

Grossman M (1972). On the Concept of Health Capital and the Demand for Health, *Journal of Political Economy* 80: 223-250.

Hajizadeh M, Johnston GM, Manos D (2020). Socioeconomic inequalities in lung cancer incidence in Canada, 1992-2010: Results from the Canadian Cancer Registry. *Public Health* 185: 189-195.

Heck JE, Andrew AS, Onega T, et al. (2009). Lung cancer in a U.S. population with low to moderate arsenic exposure. *Environmental Health Perspective* 117 (11): 1718-1723.

Hendryx M, O'Donnell K, Horn K (2008). Lung cancer mortality is elevated in coal-mining areas of Appalachia. *Lung Cancer* 62 (1): 1-7.

Heron M. (2019). Deaths: Leading causes for 2017. *National Vital Statistics Reports*, 68(6). Hyattsville, MD: National Center for Health Statistics. https://www.cdc.gov/nchs/data/nvsr/nvsr68/nvsr68_06-508.pdf.

Hoffer A, Humphreys BR, Ruseski JE (2019). State cigarette taxes and health expenditures: Evidence from dynamic spatial lag panel models. *Papers in Regional Science* 98 (2): 925-950.

Hollingsworth, A. J., Konisky, D.M., Zirogiannis, N. (2021). The health consequences of excess emissions: Evidence from Texas. *Journal of Environmental Economics and Management, 108*: 102449.

Hu D, Jiang J (2014). $PM_{2.5}$ pollution and risk for lung cancer: a rising issue in China. *Journal of Environmental Protection* 5: 731-738.

Huang F, Pan B, Wu J, et al. (2017). Relationship between exposure to $PM_{2.5}$ and lung cancer incidence and mortality: a meta-analysis. *Oncotarget* 8 (26): 43322-43331.

Hystad P, Demers PA, Johnson KC, et al. (2013). Long-term residential exposure to air pollution and lung cancer risk. *Epidemiology* 24: 762-772.

Ito K, Thurston GD, Silverman RA (2007). Characterization of $PM_{2.5}$, gaseous pollutants, and meteorological interactions in the context of time-series health effects models. *Journal of Exposure Science and Environmental Epidemiology*. 17: S45-S60.

Johnsen R, LaRiviere J, Wolff H. (2019). Fracking, coal, and air quality. *Journal of the Association of Environmental and Resource Economists* 6 (5): 1001-1037.

Kao, S. Y. et al. (2016). Spatial regression: The curious case of negative spatial dependence. Retrieved from http://www.econ.uiuc.edu/~hrtdmrt2/Teaching/SE_2016_19/References/Neg.pdf

Kaye DR, Min HS, Herrel LA, et al. (2018). Costs of cancer care across the disease continuum. *Oncologist* 23 (7): 798-805.

Kersey J, Yin J (2020). Does $PM_{2.5}$ contribute to the incidence of lung and bronchial cancers in the United States? Chapter 3: Case Study In: *Spatiotemporal Analysis of Air Pollution and its application in Public Health* pp. 69-89.

Kim H. Shim J, Park B, et al. (2018). Long-term exposure to air pollutants and cancer mortality: A meta-analysis of cohort studies. *International Journal Environmental Research and Public Health* 15 (11).

Konisky DM (2009). The Limited Effects of Federal Environmental Justice Policy on State Enforcement. *The Policy Studies Journal* 37 (3): 475-496.

Korte JE, Brennan P, Henley SJ, et al. (2002). Dose-specific meta-analysis and sensitivity analysis of the relation between alcohol consumption and lung cancer risk. *American Journal of Epidemiology* 155 (6): 496-506.

Krutilla K (1999). Environmental policy and transactions costs. In J. C. J. M. van den Bergh, (Ed.). *Handbook of Environmental and Resource Economics* (pp. 249-264). Edward Elger, Cheltenham.

Lacombe DJ, Flores M (2017). A hierarchical SLX model application to violent crime in Mexico. *Annals of Regional Science* 58: 119-134.

Lee J (2013). The influence of alcohol consumption on income and health: empirical evidence from a panel of OECD countries. *Seoul Journal of Economics* 26 (2): 255-282.

LeSage JP (2014). What Regional Scientists Need to Know about Spatial Econometrics? *The Review of Regional Studies* 44: 13-32.

LeSage JP (2021). Spatial econometrics toolbox. https://www.spatial-econometrics.com/

Li R, Zhou R, Zhang J (2018). Function of $PM_{2.5}$ in the pathogenesis of lung cancer and chronic airway inflammatory diseases. *Oncology Letters* 15 (5): 7506-7514.

Lokhandwala T, Bittoni MA, Dann RA, et al. (2017). Costs of diagnostic assessment for lung cancer: A Medicare claims analysis. *Clinical Lung Cancer* 18 (1): e27-e34.

Lundstrom NG, Nordberg G, Englyst V, et al. (1997). Cumulative lead exposure in relation to mortality and lung cancer morbidity in a cohort of primary smelter workers. *Scandinavian Journal of Work, Environment & Health* 23: 24-30.

Mausbach BT, Yeung P, Bos T, et al. (2018). Health care costs of depression in patients diagnosed with cancer. *Psycho-Oncology* 27 (7): 1735-1741.

Mavridis K, Michaelidou K (2019). The obesity paradox in lung cancer: Is there a missing biological link? *Journal of Thoracic Disease* 11 (suppl 3): S363-S366.

Molina JR, Yang P, Cassivi SD, et al. (2008). Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship. *Mayo Clin Proc*. 83 (5): 584-594.

Moran PAP (1948). The interpretation of statistical maps. *Journal of the Royal Statistical Society, B 10*: 243–51.

Morris RD (1995). Drinking water and cancer. *Environmental Health Perspectives* 103 (Suppl 8): 225-232.

Mu, Y., Rubin, E. A., & Zou, E. (2021). What's missing in environmental (self-) monitoring: evidence from strategic shutdowns of pollution monitors (No. w28735). National Bureau of Economic Research.

National Cancer Institute, (2015). Risk factors for cancer. https://www.cancer.gov/about-cancer/causes-prevention/risk.

Office of Federal Register (2011). Code of Federal regulations 141.62. https://www.govinfo.gov/content/pkg/CFR-2011-title40-vol23/pdf/CFR-2011-title40-vol23-sec141-62.pdf

Okunade AA, Karakus MC (2003). Mortality from breast carcinoma among US women: the role and implications of socioeconomics, heterogenous insurance, screening mammography, and geography. *Health Care Management Science* 6: 237-248.

Park J, Look KA (2019). Health care expenditure burden of cancer care in the United States. *Inquiry* 56. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6778988/

Pavlyuk, D. (2011). Spatial Analysis of Regional Employment Rates in Latvia. *Scientific Journal of Riga Technical University 2*: 56-62.

Peter Z, Breyer F, Kifmann M. (2009). *Health Economics* (second edition). Berlin, Germany.

Peterson EE, Ver Hoef JM (2010). A mixed-model moving-average approach to geospatial modeling in stream networks. *Ecology* 91 (3): 644-651.

Phaneuf DJ, Requate T (2017). *A Course in Environmental Economics: Theory, Policy, and Practice*. Cambridge, United Kingdom.

Raaschou-Nielsen O, Beelen R, Wang M, et al (2016). Particulate matter air pollution components and risk for lung cancer. *Environmental International* 87: 66-73.

Rahman T, Kohli M, Megdal S, et al. (2010). Determinants of environmental noncompliance by public water systems. *Contemporary Economic Policy* 28 (2): 264-274.

Scutti S (2018, September 18). Eating junk food tied to higher risk of numerous cancers. CNN. https://www.cnn.com/2018/09/18/health/junk-food-nutri-score-study/index.html

Shah V, Rieger RH, Pan LX (2019). Precipitation and climate zone explains the geographical disparity in the invasive cancer incidence rates in the United States. *Environmental Engineering Science* 36 (12).

Shiber JG (2005). Arsenic in domestic well water and health in central Appalachia, USA. *Water, Air, and Soil Pollution* 160: 327-341.

Slatore CG, Au DH, Gould MK (2010). An official American Thoracic Society Systematic Review: Insurance status and disparities in lung cancer practices and outcomes. *American Journal of Respiratory and Critical Care Medicine* 182 (9): 1195-1205.

Speizer FE, Colditz GA, Rosner HB, et al. (1999). Prospective study of smoking, antioxidant intake, and lung cancer in middle-aged women (USA). *Cancer causes and Control* 10: 475-482.

Steenland K, Boffetta P (2000). Lead and cancer in humans: Where are we now? *American Journal of Industrial Medicine* 38 (3): 295-299.

Stoiber T, Temkin A, Andrews D, et al. (2019). Applying a cumulative risk framework to drinking water assessment: a commentary. *Environmental Health* 18: 37.

Switzer D, Teodoro MP (2017). The color of drinking water: Class, race, ethnicity, and safe drinking water act compliance. *Journal of American Water Works Association* 109 (9): 40-45.

Sugar R (2018, Oct 24). The more money you make, the more fast food you eat. Vox. https://www.vox.com/the-goods/2018/10/24/18018544/fast-food-cdc-class-rich-people

Tabatabai MA, Kengwoung-Keumo J, Oates GR, et al. (2016). Racial and Gender Disparities in Incidence of Lung and Bronchus Cancer in the United States: A Longitudinal Analysis. *PLoS ONE* 11 (9): e0162949.

Tonin S, Alberini A, Turvani M (2012). The value of reducing cancer risks at contaminated sites: Are more knowledgeable people willing to pay more? *Risk Analysis* 32 (7).

Totaro M, Giorgi S, Filippetti E, et al. (2019). Asbestos in drinking water and hazards to human health: A narrative synthesis. *Ig Sanita Pubbl.* 75 (4): 303-312.

UWPHI (2022). County Health Rankings Model. University of Wisconsin Population Health Institute. https://www.countyhealthrankings.org/explore-health-rankings/measures-data-sources/county-health-rankings-model.

Vinikoor-Imler LC, Davis JA, Luben TJ (2011). An ecologic analysis of county-level $PM_{2.5}$ concentrations and lung cancer incidence and mortality (Brief Report). *International Journal of Environmental Research and Public Health* 8 (6): 1865-1871.

Wagstaff A (1986). The demand for health: A simplified Grossman model, *Bulletin of Economic Research* 38: 93-95.

Wakelee HA, Chang ET, Gomez SL, et al. (2007). Lung cancer incidence in the never-smokers. *Journal of Clinical Oncology* 25 (5): 472- 478.

Weng M, Pi J, Tan B, et al. (2017). Area deprivation and liver cancer prevalence in Shenzhen, China: A spatial approach based on social indicators. *Social Indicators Research* 133: 317-332.

Wu MM, Kuo TL, Hwang YH, et al. (1989). Dose-response relation between arsenic well water and mortality from cancer. *American Journal of Epidemiology* 130: 1123-1132.

Xie H, Wise SS, Holmes AL, et al. (2005). Carcinogenic lead chromic induces DNA double-strand breaks in human lung cells. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis* 586 (2): 160-172.

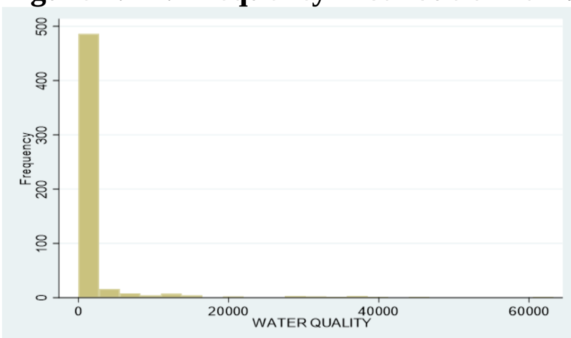
Yabroff KR, Lund J, Kepka D, et al. (2011). Economic burden of cancer in the US: Estimates, projections, and future research. *Cancer Epidemiol, Biomarkers & Prevention* 20 (10): 2006-2014.

Yang M, Chou S (2018). The impact of environmental regulation on fetal health: Evidence from the shutdown of a coal-fired power plant located upwind of New Jersey. *Journal of Environmental Economics and Management.* 90: 269-293.

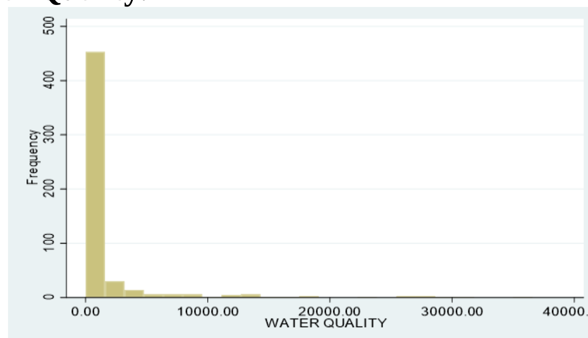
Zhu B, Pang R, Chevallier J, et al. (2019). Including intangible costs into the cost-of-illness approach: a method refinement illustrated based on the $PM_{2.5}$ economic burden in China. *European Journal of Health Economics* 20 (4): 501-511.

APPENDIX

Figure 2.1A: Frequency Distribution for Water Quality.

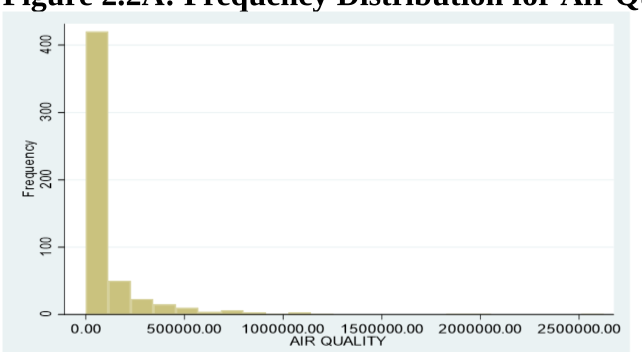


Before taking 3-year moving average.

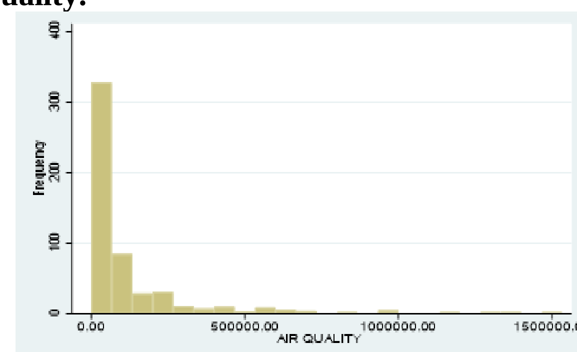


After taking 3-year moving average.

Figure 2.2A: Frequency Distribution for Air Quality.



Before taking 3-year moving average.



After taking 3-year moving average.

Table 2.1A: Coefficients of Ordinary Least Squares (OLS) Model and Spatial Error Model (SEM) for 48 Contiguous States plus Washington D.C. 2006-2016.

Variables	**Dep. Variable: Lung and Bronchus Cancer Incidence**	
	OLS	**SEM**
WATER QUALITY (pop. exposure rate)	0.000008	0.00003
	(0.903)	(0.267)
AIR QUALITY (pop. exposure rate)	0.000003**	0.000002**
	(0.030)	(0.033)
PRECIPITATION (days)	0.007	0.0121**
	(0.593)	(0.180)
SMOKE (% of smokers)	1.218***	0.1004***
	(0.000)	(0.0432)
ALCOHOL (% of adults)	-0.140**	0.112**
	(0.006)	(0.056)
GENDER (% of Male)	-7.243***	3.163***
	(0.000)	(0.000)
ETHNICITY (% of non-white)	-0.230***	0.143***
	(0.000)	(0.012)
INCOME (household median income in $)	0.00003	0.000081***
	(0.506)	(0.031)
OBESITY (% of adults)	0.513***	-0.028
	(0.000)	(0.785)
UNINSURED (% of adults)	0.302***	0.0192
	(0.000)	(0.737)
Constant	353.275***	
	(0.000)	
Lambda		-0.03
		(0.598)
State FE		Yes
Year FE		Yes
R-squared	0.71	0.97
Obs.	539	539

Note: *** for 0.01, ** for 0.05, * for 0.1 Numbers in the parentheses represent p-values. Although state and year fixed effects were included in the SDM model, these coefficients are not reported.

CHAPTER 3:

ESSAY 2: SPILLOVER EFFECTS OF NON-PHARMACEUTICAL INTERVENTIONS ON REDUCING COVID-19 CASES: EVIDENCE FROM THE CONTIGUOUS U.S. COUNTIES

3.1 Introduction

More than 100 million people have been reported as coronavirus disease (COVID-19) cases and over a million people have died from COVID-19 in the U.S. since 29 February 2020 (CDC, May 5, 2023). The World Health Organization (WHO)[1] identified COVID-19 as an infectious disease caused by the SARS-CoV-2 virus, and the Centers for Disease Control and Prevention (CDC) labeled this disease as a pandemic[2]. Within a pandemic landscape, it is imperative to investigate the spatial pattern of COVID-19 spread across counties, and state and local governments should adopt appropriate planning and policies based on evidence[3]. COVID-19 transmission had become widespread in New York City (NYC) along with urban-areas of northeast and northwest by mid-March 2020, while rural areas of the country were not affected until later. CDC recommended various non-pharmaceutical interventions (NPIs) to slow down the spatial spread of COVID-19 pandemic. These measures include social distancing, business closure and reopening, school closing, public gathering and travel restrictions, stay-at-home and shelter-in-place orders. While some states and counties implemented those NPIs very strictly, other states and counties did not adopt those policies. The variation in adopting NPIs across the country thus may have affected the transmission of COVID-19 across these different areas.

COVID-19 involved a significant economic burden in the U.S. Nationally, $150 billion was allocated under the Coronavirus Aid, Relief, and Economic Security (CARES) Act to support public health emergency during COVID–19 pandemic (IRS, 2023). In 2020, each COVID-19

---

[1] https://www.who.int/health-topics/coronavirus#tab=tab_1
[2] CDC identified this disease as outbreak, epidemic, and pandemic simultaneously, See, https://www.cdc.gov/coronavirus/2019-ncov/science/about-epidemiology/identifying-source-outbreak.html
[3] Center for Spatial Data Science, University of Chicago. See, https://theuscovidatlas.org/

patient paid, on average, \$41,611 for hospitalization including out-of-pocket expenses of \$1,280 for those with significant employment coverage (Wager, et al. 2022).

This essay aims to examine spillover effects of NPIs on the reduction of COVID-19 cases in the contiguous U.S. counties. In this research, the first objective is to estimate the spillover effects of implementing stay-at-home orders, mask mandates, restaurant, and bar closure, and reopening for one month on the reduction of COVID-19 cases in the contiguous U.S. counties. Second and final objective of this research is to calculate monetary benefits from reducing COVID-19 spread in the county where policies are in place as well as neighboring counties. By utilizing spatial econometric models, this paper finds a statistically significant negative effect of mandatory stay-at-home order, mask wearing mandate on COVID-19 spread. With a significant ρ (rho) value in the SDM model, ordinary least squares (OLS) coefficients are likely biased estimates.

From the SDM model based on annual county-level cross-sectional data for 2020, a one-month implementation of mandatory stay-at-home order and mask mandate results in 19 fewer cases per 100,000 people in the mandate county and 47 fewer cases per 100,000 people in six-nearest neighboring counties. By separating out the effect of mask mandate, the SDM model gives an estimate of 4 fewer cases as direct effect within the county and 9 fewer cases as indirect effect in six-nearest neighboring counties. Based upon the SDM model, spillover benefits of about \$0.70 million, \$2.95 million, and \$3.65 million are calculated from three nearest neighboring Counties (Sheridan, Campbell, Crook)[4] of Powder River[5] County by implementing mask mandate, mandatory stay-at-home order, and both these policies, respectively for one month. Thus, the indirect effect in neighboring counties is larger than the direct effect of a these NPIs implemented

[4] Sheridan, Campbell, and Crook are three counties in the state of Wyoming, a non-mask mandate state.
[5] Powder River County is in the state of Montana, a mandate state.

in that particular county. Restaurant and bar limitations and closures did not have any statistically significant impacts in the model. The SEM model provides similar estimates but also provide evidence that some significant spillover effects of variables are not explicitly modeled in this study. So far, my knowledge goes, major contribution of this study is the estimates of spatial spillover effects of NPIs on the COVID-19 spread that nobody has estimated yet.

Rest of this paper is organized in the following manner: section 2 reviews literature, section 3 describes data and methods, section 4 discusses empirical results, and section 5 offers conclusions and policy implications.

3.2 Literature Review

3.2.1 Non-Pharmaceutical Interventions

Among the NPIs, the mask-wearing mandate is the key initiative that has been implemented by many states in the U.S. Using county-level data for Kansas, Zambrana and Ginther (2020) applied difference-in-differences methods to investigate the effect of mask mandate on COVID-19 cases. Based on 7-day rolling average data for the COVID-19 cases per 100,000 people, they find the mask mandate working for controlling COVID-19 spread in the state of Kansas. When the mask mandate was in effect, the COVID-19 case rates in counties with mask mandates were three times higher than those in counties without them (15 cases per 100 000 people compared to 5 cases per 100,000 people). These patterns changed by October 26, cases in counties without mask mandates were 2.1 times more (44 cases per 100 000 people compared to 21 cases per 100,000 people) than cases in mandate counties (Ginther and Zambrana, 2021). Through December 4, there were 20.33 fewer cases per day in counties with mask mandates than in counties without mask mandate.

From 412 U.S. counties between March 21 and October 20, 2020, Huang, et al. (2022) estimated an additional benefit of county-level public mask mandates. In masked counties as compared to unmasked counties, the daily case incidence per 100,000 individuals decreased on average by 25% at four weeks, 35% at six weeks, and 18% throughout the course of six weeks postintervention. Based on natural experiment in two school districts in greater Boston area, Tori et al. (2022) finds additional 44.9 COVID-19 cases per 1,000 students and staffs during 15 weeks after lifting the statewide masking policy. Cui, et al (2020) also used mask and shelter-in-place (SIP) mandates to examine the performance of those policies as a response to COVID-19 pandemic in the U.S. Based on Probit, Logit, and LPM models, they claimed that wearing mask is not necessarily reducing COVID-19 cases, rather it is a decision motivated by politics. Although Rader, et al. (2021) found highest chance of transmission control was achieved by wearing mask and maintaining physical distance, but no significant change was observed after adopting the state-wide mask-wearing mandate. At the national level, Chernozhukov, et al. (2021a) found that nationally mask-wearing mandate for employees during early pandemic reduced the weekly growth rate of cases and deaths by more than 10 percentage points in late April that led to as much as 19 to 47 percent less deaths by the end of May. Similar results were found by Zhang, et al. (2020) and Eikenberry, et al. (2020). Although there is a discrepancy between mask-wearing behavior and compliance to the mask mandate, an appropriate intervention can increase mask wearing and reduce symptomatic COVID-19 infections (Kim, et al. 2022; Abaluck et al. 2022).

State and local governments' actions such as business and school closures contributed to reducing COVID-19 spread in the U.S. (Chernozhukov, et al. 2021b) For instance, Gupta et al. (2021) examines the determinants of social physical distancing during shutdown phase of the COVID-19 pandemic and found that about 55 percent of COVID-19 growth were responsible for not following those policies, thus suggesting that the NPIs were effective in reducing human mobility. Chernozhukov, et al. (2021a) mentioned that COVID-19 cases would have been larger by 17 to 78 percent without business closures but no certain effect was found for school closures decision due to lack of cross-sectional variation. Based on daily county-level data, Isphording et al. (2021) found no positive effect of school re-openings on COVID-19 cases in Germany. On the other hand, using county-level data in Texas, Courtemanche, et al. (2021) found that community spread of COVID-19 accelerated substantially due to re-opening schools. More specifically, school reopening led to at least 43,000 additional COVID-19 cases and 800 additional deaths within first two months of school reopening. Under three alternative scenarios, a 22% higher rate of infections was found if 50% capacity with 100% face-mask adherence in schools than the rate if schools operating remotely in Indiana (Espana et al, 2021). Since the association between K-12 schools' in-person reopening and case growth was found stronger in counties where the mask mandate has not been implemented (Chernozhukov et al. 2021b). Thus, schools reopening does not a have robust effect on COVID-19 spread in the U.S.

3.2.2 Political Stringency and Trust

Political stringency and trust in politicians are of great importance in the endeavor of controlling the COVID-19 pandemic (Bargain and Aminjonov, 2020). The debate arises among researchers on whether political affiliation at the state level impact wearing masks. Cui, et al. (2021) indicated ideology[6] (i.e., conservatism and support for President Trump) to be the dominant predictor of low rates of mask-wearing. Kahane (2021) also finds a significant influence of

[6] Cui, et al. (2021) measured ideology by support for President Trump, Evangelicals per 1000 people, Catholics per 1000 people, religious diversity, belief in climate change, percentage of white population, percentage of latino population, percentage of male population.

political ideology in wearing masks. Even, public confidence in medical scientists has declined since COVID-19 in terms of the efficacy of NPIs on COVID-19 spread (CNN, March 9, 2023).

3.2.3 Spatial Interaction

Finally, spatial interaction plays a vital role in case of infectious disease transmission such as COVID-19. Based on Spearman *rho* correlation and Moran's I test, Jackson, et al. (2021) find regional clustering patterns of COVID-19 cases and fatalities in the U.S. counties. These COVID-19 cumulative experiences were initially evident in the Great plains, Southwestern and Southern regions. Using county-level data on COVID-19 community vulnerability index (CCVI) for U.S. counties, Ulimwengu and Kibnonge (2021) applied spatial Durbin model (SDM) and find an existence of spatial spillover effects where indirect effects (from neighboring counties) dominate the direct effects (from county-own vulnerability). As a variable of interest, this study utilized seven-types of vulnerability such as socioeconomic status, minority status & language, transportation, household composition & disability, epidemiological factors, healthcare system, high-risk environment, and population density whose are either socioeconomic or infrastructural factors and used cumulative data. While most literature has found that mask-wearing mandates reduce COVID-19 spread (Zambrana and Ginther 2020; Rader, et al. 2021; Zhang, et al. 2020; Eikenberry, et al. 2020; Chernozhukov, et al. 2021a; Courtemanche, et al. 2021; Gupta et al. 2021), variation in adopting NPIs across jurisdictions provides an opportunity to examine the spatial spillover effects from mandates (Olmo and Sanso-Navarro, 2021). Thus, this paper investigates the spatial spillover effects of NPIs policies on the reduction of COVID-19 cases within the contiguous U.S.

3.3 Data and Methods

3.3.1 Data

This study investigates the spatial spillover effects of NPIs in reducing COVID-19 cases in the contiguous U.S. As a pre-vaccination period, this study collects annual cross-sectional data from 3045 contiguous counties for the year 2020 to avoid the potential effect of vaccination on the spread of COVID-19 cases. Table 3.1 provides descriptions of the outcome variable, the variables of interest, and other control variables. In this study, COVID-19 cases per 100,000 people serve as the dependent variable, and the independent variables of interest are mandates for: stay-at-home order, mask-wearing, restaurant open with limitations/curbside delivery, bar open with limitations/curbside delivery, and bar fully closed. Data on the outcome variable have been collected from the USA Facts website[7], and that on independent variables of interest have been collected from National Environmental Public Health Tracking Network (NEPHTN). Data on the mask-wearing mandate was available from April 2020, and that on stay-at-home order, restaurant and bar closure were available from March 2020. Before the initial implementation of NPIs in the U.S., no similar restrictions existed, thus I input values of zero for mask mandates for the months of January, February, and March. Similarly, I put the value zero for the month of January and February 2020 for restaurant and bar as variables of interest for those counties considered in this study.

For a dependent variable, cumulative COVID-19 cases (confirmed) for the year 2020 are collected and divided by county population (which is also available in USAFacts website) and then multiplied by 100,000 to obtain COVID-19 cases per 100,000 people as shown in equation 3.1.

[7] https://usafacts.org/visualizations/coronavirus-covid-19-spread-map

$$COVID\text{-}19 \text{ cases per } 100{,}000 \text{ people} = \frac{COVID-19\ Cases\ in\ County\ i}{Population\ in\ County\ i} \times 100{,}000 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.1)$$

Dummy variables are created for the variables of interest on a daily basis. Data on mask mandate started from 10 April 2020 through end of the year. For the mask-wearing mandate, the value =1 if mask-wearing is required in public place, and value =0 if otherwise. Data on state orders on business (restaurant and bar) closing and reopening started from 15 March 2020 through end of the year. For restaurant closing and reopening, two separate binary variables are created: (1) no restrictions or authorized to fully reopen (1 if restriction was found or authorized to fully reopen, 0= otherwise); (2) open with limitations or curbside delivery only (=1 if open with limitation/curbside delivery only, 0=otherwise). Since no county mandated that all restaurants remain completely closed, no dummy variable is created for restaurant closing. For bar closing and reopening, three separate binary variables are created: no restrictions or authorized to fully reopen (=1 if restriction was found or authorized to fully reopen, =0 otherwise), open with limitations or curbside delivery only (=1 if open with limitation/curbside delivery only, =0 otherwise), and closed (=1 if bars were mandated to remain closed, =0 if otherwise). Data included primarily counties in forty-eight contiguous states plus Washington D.C. and excluded counties in Alaska and Hawaii. Since four counties: Dukes, and Nantucket counties (Massachusetts), Richmond County (New York), and San Juan County (Washington) have no neighbors, they are excluded from the analysis.

Although stay-at-home orders were executed between 15 March- 31 May 2020, data are publicly available for only 15 March- 5 May 2020 (Moreland, et al. 2020). Therefore, daily dummy variables are used for stay-at-home order between 15 March- 5 May 2020 for annual cross-sectional data with a value =1 if the advisory order is placed and =0 if otherwise. Similarly, a value =1 is placed if mandatory order is placed and =0 if otherwise. In both scenarios, ‘no order found’ is used as the base. Since the initial implementation of stay-at-home order started on 15 March 2020 in the U.S., daily values of zero are for stay-at-home order policy from January to 14 March. Similarly, I input values of zero for the same policy after 5 May 2020.

NPI variables utilized in the analyses are measured as the percentage of days where a NPI is in place on an annual basis. Daily NPI dummy variables are aggregated across the entire year 2020 and divided by 365 to obtain percentage variables for an annual data model.

In this study, cross-sectional data were collected on other control factors such as non-compliance factor, demographic factors, socioeconomic factors, specific medical factors, and lifestyle factors. Trump's votes as a percentage of total votes represent the non-compliance factor that comes from the 2020 presidential election results. The reason behind Trump support as a non-compliance factor is President Trump's skepticism over the effectiveness of mask and he denied the adoption of such national mandate[8]. I used dummy for 'urban' as a proxy of rurality factor where value =1 if county is in metro area or adjacent to metro area, and value =0 if county is not adjacent to metro area. Data on socioeconomic factors such as income measured as 'median household income', poverty measured as 'percentage of people of all ages in poverty in 2020', and unemployment comes from the USDA ERS website while that on housing measured as 'percentage of housing units with more people than rooms' is collected from NEPHTN database. The COVID-19 Risk Index collected from PolicyMap measures specific medical conditions that include chronic obstructive pulmonary disease (COPD), heart disease, high blood pressure, diabetes, and obesity. Finally, data on lifestyle factors such as smoking measured as 'percentage of current smokers aged

[8] Victor, et al. (2020). *In His Own Words, Trump on the Coronavirus and Masks*. NYTimes. Accessed: https://www.nytimes.com/2020/10/02/us/politics/donald-trump-masks.html

18 years and over' and remote work measured as 'percentage of people worked from home aged 16 years and over' comes from NEPHTN and ACS, respectively.

Table 3.1 shows expected signs of coefficients based on prior literature. NPIs such as stay-at-home orders, mask-mandate, restaurant, and bar closures and reopening are expected to lower COVID-19 infections (Castillo, et al. 2020; Chernozhukov, et al. 2021a). The expected sign of Trump votes is positive (Cui, et al. 2021; Kahane, 2021). Older adults, the male population, people of color, poor, unemployed, and low-income communities, people with specific medical conditions, adult current smokers, and the population living in dense areas and dense houses are expected to be at higher risk of COVID-19 symptoms (Abedi, et al. 2020). On the other hand, people with better education, and who work remotely are expected to be at less risk of COVID-19 symptoms. Prior studies show mixed results for people who live in urban areas (Lee, et al. 2021). While COVID-19 infection rates were higher in urban counties in the early period of 2020, the rate becomes higher in rural counties in the later period of 2020 (Nelson and Cromartie, 2022).

Table 3.1: Description of Variables.

Variable	Definition	Source of Data	Expected Sign
(a) Dependent variable:			
COVIDCASE	COVID-19 Case per 100,000 people	USA Facts	
(b) Variable of Interest:			
STAYHOMEADV	Stay-At-Home Order Advisory[a]	NEPHTN[d]	Negative
STAYHOMEMAN	Stay-At-Home Order Mandatory[a]	NEPHTN[d]	Negative
MASKMANDATE	Mask wearing mandate	NEPHTN[d]	Negative
RESTLIM	Restaurant open with limitations/curbside delivery[b]	NEPHTN[d]	Negative
BARLIM	Bar open with limitations/ curbside delivery[c]	NEPHTN[d]	Negative
BARCLOSED	Bar closed[c]	NEPHTN[d]	Negative
(c) *Non-compliance Factor:*			
TRUMPVOTES	Trump Votes (%)	github[e]	Positive
(d) *Demographic Factors:*			
ADULT	Age (% of adult ≥ 65 years)	ACS[f]	Positive
MALE	Gender (% of Male)	ACS[f]	Positive
EDUCATION	Education (% people w/ bachelor's degree or higher)	ACS[f]	Negative
URBAN	Urban (metro or adjacent counties)	USDA ERS[g]	Positive /Negative
NON-WHITE	Non-White (% of Non-White)	ACS[f]	Positive
POPDENSITY	Population Density (Population per square mile)	Census Bureau	Positive
(e) *Socioeconomic Factors:*			
INCOME	Income (Median Household Income in dollars)	USDA ERS[h]	Positive
POVERTY	Poverty (% of people of all ages in poverty)	USDA ERS[h]	Positive
UNEMPLOYMENT	Unemployment (% of unemployed people)	USDA ERS[h]	Positive
HOUSING	Housing Condition (% of Housing Units w/ more people than Rooms)	NEPHTN[d]	Positive
(f) *Specific Medical Conditions:*			
COVIDRISK	COVID-19 Risk Index (% of population ≥ 18 years)	PolicyMap[i]	Positive
(g) Lifestyle Factors:			
SMOKING	Smoking (% of current smokers ≥ 18 years)	NEPHTN[d]	Positive
REMOTEWORK	Remote Work (% of people worked from home ≥ 16 years)	ACS[j]	Negative

Note:

(a) No Stay-at-Home order is the base.
(b) Restaurant with no restrictions/ authorized to fully reopen" is the base. No county mandated that all restaurants remain completely closed.
(c) "Bar with no restrictions/ authorized to fully reopen" is the base.
(d) National Environmental Public Health Tracking Network.
(e) https://github.com/tonmcg/US_County_Level_Election_Results_08-20/blob/master/2020_US_County_Level_Presidential_Results.csv
(f) American Community Survey
(g) USDA Economic Research Service: https://www.ers.usda.gov/data-products/rural-urban-continuum-codes.aspx
(h) USDA Economic Research Service: https://www.ers.usda.gov/data-products/county-level-data-sets/county-level-data-sets-download-data/
(i) PolicyMap: https://www.policymap.com/solutions/covid-19
(j) American Community Survey: https://data.census.gov/cedsci/table?t=Employment&g=0100000US%240500000&y=2020&tid=ACSDP5Y2020.DP03&moe=false&tp=false

Table 3.2 depicts summary statistics on annual cross-sectional data on dependent, independent, and control variables for 3045 contiguous counties. On average, counties mandated 4.77 percent of days out of the entire period of 2020 with advisory stay-at-home orders and 12.49 percent of days with mandatory stay-at-home orders. For the same period, counties on average implemented 36.03 percent of days with mask mandate. To the highest, counties in New Jersey implemented mask mandate for 75 percent of the entire year (equivalent to 274 days) followed by counties in New York. Counties in New Jersey implemented mandatory stay-at-home for 22.06 percent of the year (equivalent to 81 days) which is the second highest. The average COVID-19 risk index is 51.57 where the index ranges between 0 and 100.

Table 3.2: Summary Statistics, by County, Annual, 2020.

Variable	Obs.	Mean	Std. dev.	Min	Max
Dependent variable:					
COVID-19 Case per 100,000 people	3,049	555	231	35	2264
Variables of Interest:					
Stay at Home- No Order (% of days)	3,049	7.74	5.80	0.49	25.00
Stay at Home- Advisory Order (% of days)	3,049	4.77	6.57	0.00	24.51
Stay at Home- Mandatory Order (% of days)	3,049	12.49	8.28	0.00	24.02
Mask Mandate (% of days)	3,049	36.03	24.70	0.00	75.00
Restaurant with No Restrictions/authorize to fully reopen (% of days)	3,049	9.57	18.36	0.00	63.93
Restaurant Open with limitations/curbside delivery (% of days)	3,049	73.76	18.36	19.40	83.33
Bar with No restrictions/authorize to fully reopen (% of days)	3,049	9.74	18.30	0.00	63.93
Bar Open with limitations/curbside delivery (% of days)	3,049	66.51	21.25	0.00	83.33
Bar Closed (% of days)	3,049	7.09	14.94	0.00	78.92
Non-compliance Factor:					
Trump Votes (%)	3,049	65.04	16.04	8.73	96.18
Demographic Factors:					
Age (% of adult aged 65 years and over)	3,049	19.30	4.78	3.00	57.80
Gender (% of Male)	3,049	50.07	2.45	42.00	70.90
Education (% of adults with a bachelor's degree or higher)	3,049	22.56	9.68	0.00	79.14
Urban dummy	3,049	0.70	0.46	0.00	1.00
Ethnicity (% of Non-White)	3,049	17.85	16.38	0.00	95.20
Population Density (Population per square mile)	3,049	240.75	1225.16	0.10	36004.6
Socioeconomic Factors:					
Income (Median Household Income in dollars)	3,049	57290	14469	22901	160305
Poverty (% of people of all ages in poverty)	3,049	13.76	5.42	3.00	43.90
Unemployment (% of unemployed people)	3,049	6.71	2.27	1.50	22.80
Housing Condition (% of Housing Units with more people than Rooms)	3,049	2.28	1.92	0.00	39.00
Specific Medical Conditions:					
COVID-19 Risk Index (% of population aged 18 years and over)	3,049	51.57	27.96	2.85	100.00
Lifestyle Factors:					
Smoking (% of current smokers aged 18 years and over)	3,049	20.36	4.14	7.30	38.20
Remote Work (% of people worked from home aged 16 years and over)	3,049	6.01	3.45	0.00	36.70

Based on 12-month average data, Figure 3.1 shows the COVID-19 case per 100,000 people in the contiguous U.S. counties. Out of 3104 counties, counties on both the east and west coasts have lower number of cases while counties in the north and south regions have higher number of COVID-19 cases per 100,000 people in the country. Figure 3.2 shows that northeastern region follows a more stringent mask-wearing mandate, while the middle and southeast regions apply loose mask-wearing mandate. Figure 3.3 shows whether restaurants were open with limitations or delivered at curbside while Figure 3.4 shows whether any restaurants were fully reopened or even no restrictions were imposed across counties. Finally, Figures 3.5, 3.6, and 3.7 exhibit the execution of bar restrictions across counties of the U.S.

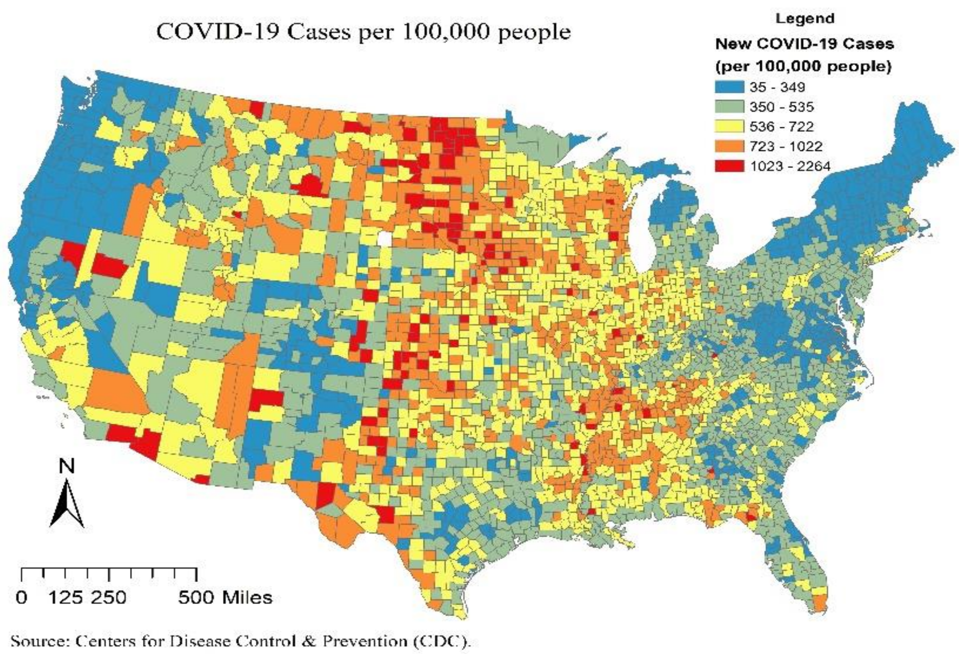


Figure 3.1: COVID-19 Case (12-months average) in the Contiguous U.S.A, by County in 2020.

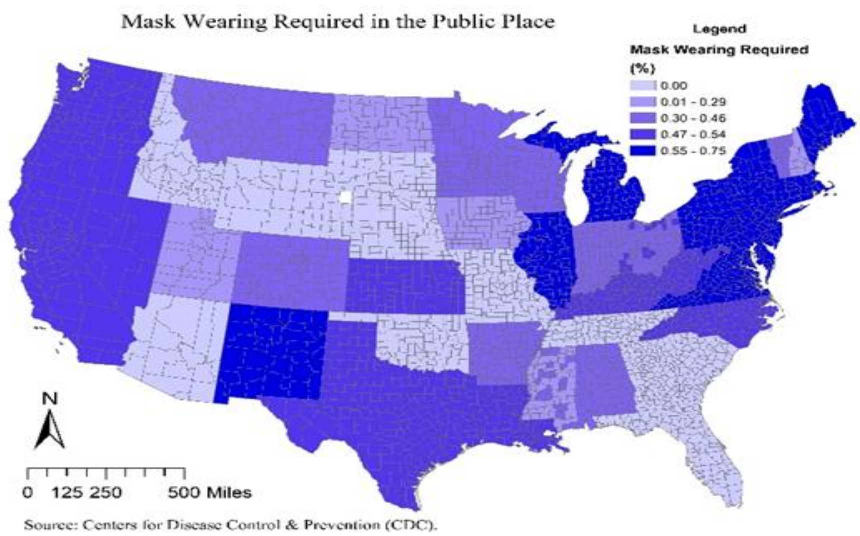


Figure 3.2: Mask Wearing Mandate (12-months average) in the Contiguous U.S.A, by County in 2020.

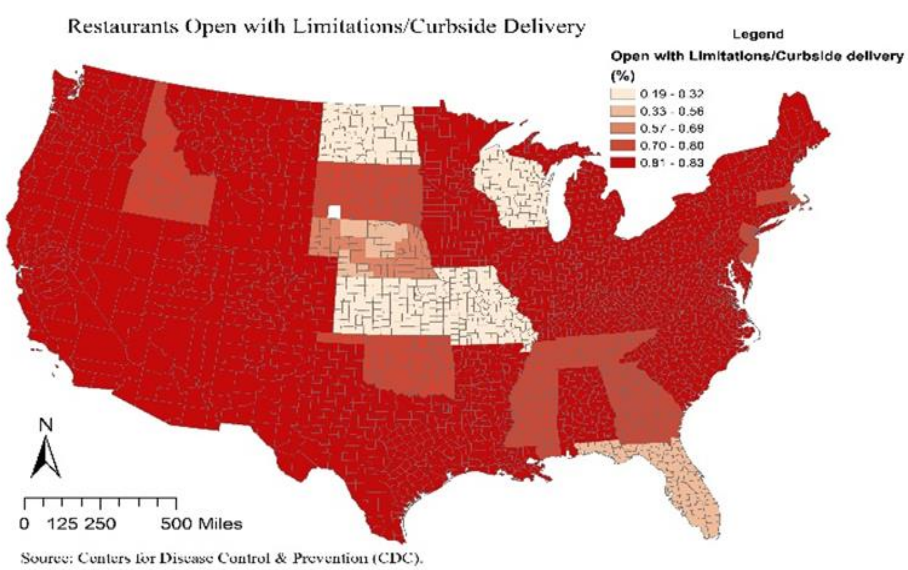


Figure 3.3: Restaurants Open with Limitations/Curbside Delivery (12-months average) in the Contiguous U.S.A, by County in 2020.

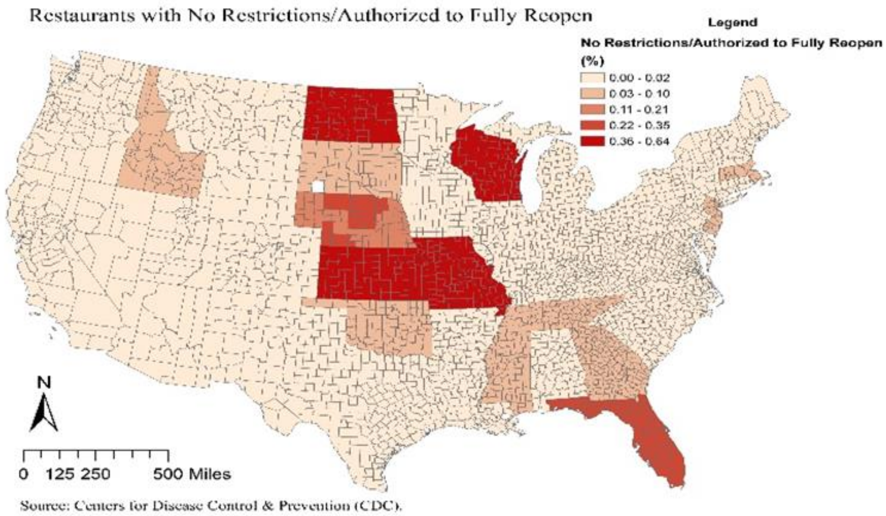


Figure 3.4: Restaurants with No Restrictions/Authorized to fully reopen (12-months average) in the Contiguous U.S.A, by County in 2020.

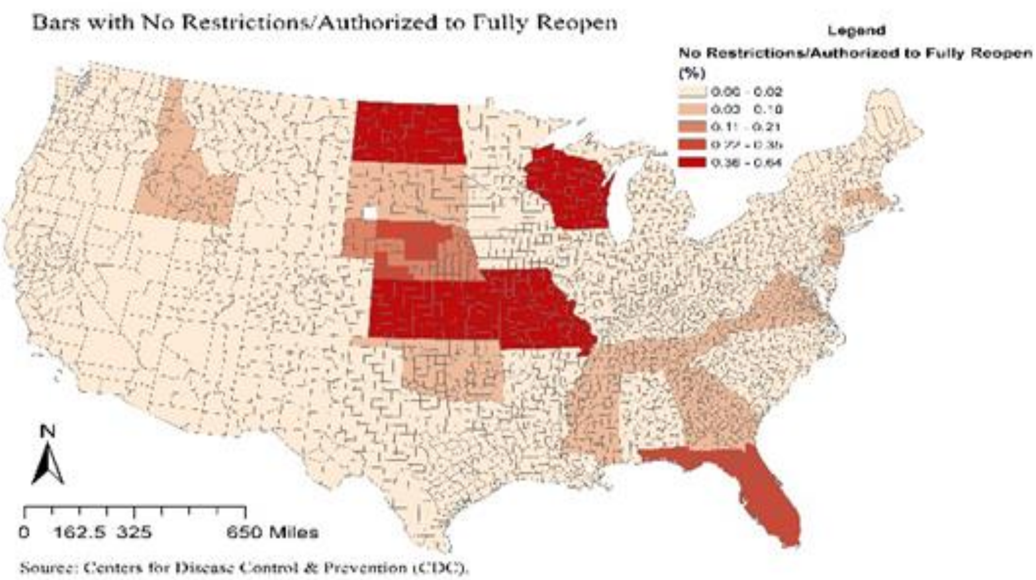


Figure 3.5: Bars with no restrictions/authorized to fully reopen (12-months average) in the contiguous U.S.A, by County in 2020.

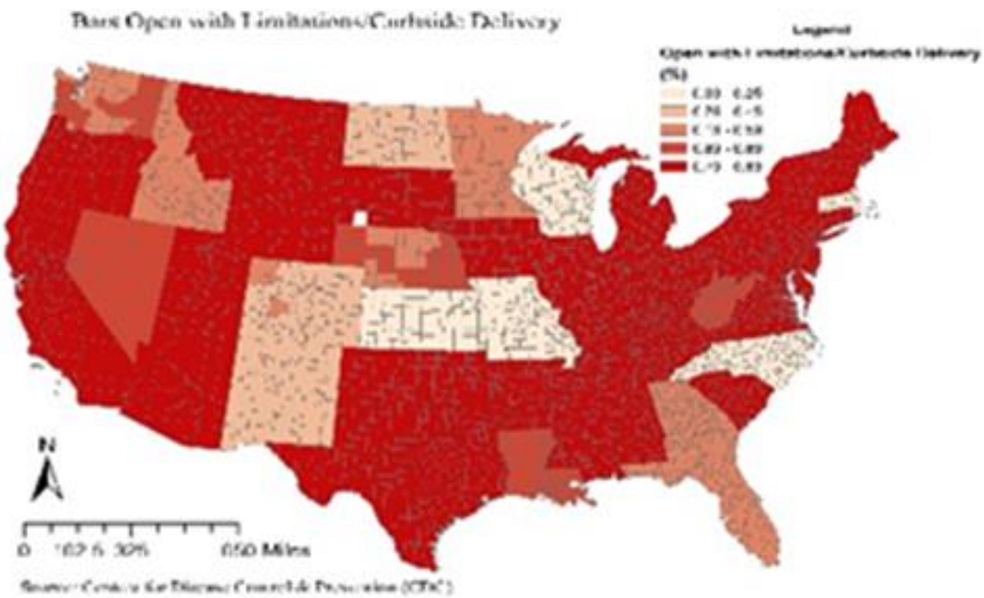


Figure 3.6: Bars open with limitations/curbside delivery (12-months average) in the contiguous U.S.A, by County in 2020.

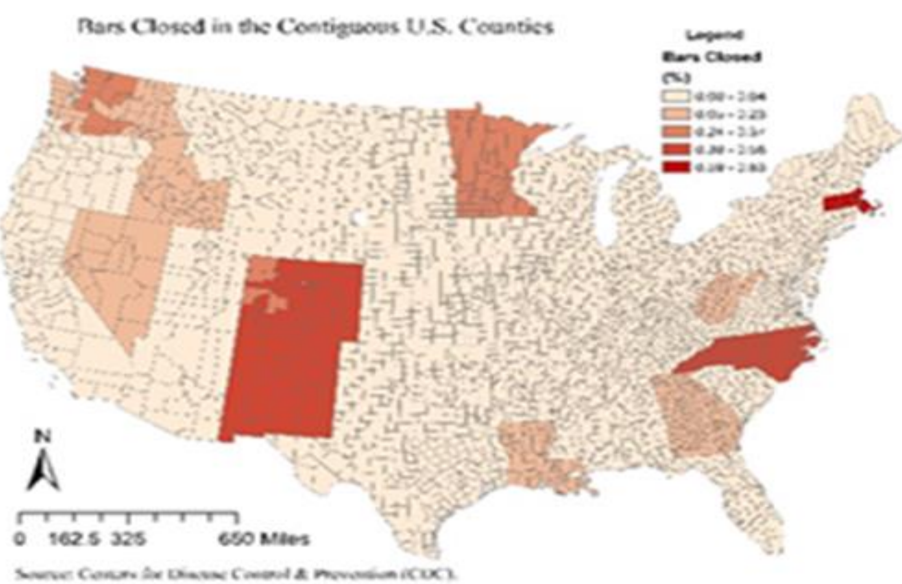


Figure 3.7: Bars fully closed (12-months average) in the contiguous U.S.A, by County in 2020.

3.3.2 Methods

Possible control factors for COVID-19 cases that determine the severity of COVID-19 outcomes include: non-compliance variable (trump votes), demographic variables (age, gender, race, ethnicity, population density), socioeconomic variables (income, education, unemployment, poverty, social deprivation, housing condition, mode of transportation), biological/specific medical conditions (cancer, chronic kidney disease, COPD, diabetes, fever, pneumonia, serious heart conditions), geographic factors (rurality, number of highway miles, number of airports), lifestyle factors (smoking, remote working), environmental factors (air quality, water quality, temperature, humidity, occupational exposure) (Crankson, et al. 2021; Benitez, et al. 2020; CDC 2020a; CDC 2020b; Correa-Areneda, et al. 2021). As a dependent variable, logarithm of COVID-19 cases per 100,000 people is used in this cross-sectional model (annual, county-level). In the following sections, I presented SDM model to estimate spillover effects of NPIs on reducing COVID-19 cases.

3.3.2.1 Spatial Durbin Model

Equation 3.2 represents the spatial Durbin model (SDM) based on annual cross-sectional data and shows the functional relationship between the outcome variable and independent variables (Elhorst, 2014).

$$log(COVIDCASE)_c = \rho W^* log(COVIDCASE)_c + \alpha I_N + \beta NPI_c + \gamma NONCOMPLIANCE_c + \delta DEMOGRAPHIC_c + \zeta SOCIOECONOMIC_c + \eta MEDICAL_c + \mu LIFESTYLE_c + \theta W^* NPI_c + \kappa W^* NONCOMPLIANCE_c + \varsigma W^* DEMOGRAPHIC_c + \sigma W^* SOCIOECONOMIC_c + \tau W^* MEDICAL_c + \varphi W^* LIFESTYLE_c + u_c \quad (3.2)$$

where, log(COVIDCASE) represents the logarithmic form of COVID-19 cases per 100,000 people, the vector *NPI* represents non-pharmaceutical interventions that include stay-at-home order, mask-wearing mandate, restaurants and bars closure and reopening, the vector *NONCOMPLIANCE* represents non-compliance categories of variables (e.g. Trump Votes), the vector *DEMOGRAPHIC* represents demographic categories of variables such as age, gender, education, ethnicity, rurality, and population density, the vector *SOCIOECONOMIC* represents socioeconomic categories of variables such as income, poverty, unemployment, housing condition, and the vector *MEDICAL* represents specific medical conditions to develop serious health complications from COVID-19, the vector *LIFESTYLE* represents lifestyle categories of variables such as smoking and remote working. In equation (3.2), ρ, θ, and W shows the spatial interactions of dependent variable, variables of interest, and k-nearest neighbor spatial weight matrix, respectively. The term u stands for disturbance term. The subscript c represents county. Equation (3.2) is estimated using annual cross-sectional models for the entire period of 2020.

Since COVID-19 is an infectious disease that might have interaction effects across counties, spatial econometric models can suitably model the spatial spillovers of this disease by using appropriate spatial weight matrices (Krisztin, et al., 2020; Guliyeb 2020). Prior literature also focuses the importance of spatial econometrics for quantifying the spatial spillovers (Emch et al. 2012; Wang et al. 2015; Chagas et al. 2016). Since spatial dependence varies across geographic locations in case of corona pandemic, Krisztin, et al. (2020) applied spatial weight matrices based on trade intensity, common borders, flight connection, and free movement of people and found different coefficient estimates under the spatial autoregressive (SAR) model and Poisson spatial error model (SEM). A Global Moran's I is computed to test spatial dependence across counties.

Statistically significant values for Moran's I indicate the need to use spatial econometric regressions (LeSage, 2014).

Equation 3.3 gives a full model with all types of interaction effects (Elhorst, 2014).

$$COVIDCASE=\rho W * COVIDCASE+\alpha I_N+(NPI)\beta+(Z)\beta+W^*(NPI)\theta+W^*(Z)\theta+u\ldots\ldots\ldots(3.3a)$$

$$u=\lambda W^*u+\varepsilon\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(3.3b)$$

where W*COVIDCASE represents the endogenous interaction effects among the dependent variable (COVID-19 case per 100,000 people), W*NPI shows the exogenous interaction effects among the independent interest variables (non-pharmaceutical interventions such as stay-at-home order, mask mandate, restaurant, and bar closure and reopening), W*Z represents the exogenous interaction effects among the covariates. In equations 3.3a and 3.3b, ρ and θ show the spatial interactions of dependent, and independent variables, respectively, while λ shows the spatial interactions of unobservable disturbance term. As covariates, non-compliance, demographic, socioeconomic, specific medical conditions, and lifestyle factors are included in the model. Finally, W*u shows the interaction effects among the disturbance term of the different units. Equations 3.3a and 3.3b thus refer to the general nesting spatial (GNS) model.

As part of spatial econometric technique, Spatial Durbin Model (SDM) and Spatial Error Model (SEM) are estimated. Using these models, both direct (i.e., effect within the county) and indirect (i.e., effect in neighboring counties which is called spillover effect) effects are estimated for NPIs' impacts on the reduction of COVID-19 in the contiguous U.S. Prior literature also applied these methods in the field of health economics (Orea and Alvarez, 2021; Chen et al. 2017; Weng et al. 2017). R software is utilized for annual cross-sectional data (Bivand, et al. 2021).

3.4 Empirical Results

Spatial spillover effects of NPIs are estimated for the spread of COVID-19 cases per 100,0000 people in the contiguous U.S. Using county-level data, impacts of stay-at-home order, mask-wearing mandates, restaurant and bars closure and reopening as part of NPIs are estimated by applying spatial econometric models.

3.4.1 Spatial Autocorrelation Test

Using annual cross-sectional data, the Global Moran's I from OLS residuals are based on the 6-nearest neighbors' weight matrix[9] (Arbia, 2014). In Table 3.3, Moran's I statistic is 0.55 which is highly statistically significant (p-value-0.000) implying that spatial relationships exist for COVID-19 cases per 100,000 people i.e., COVID-19 cases in one county (county i) depends on the COVID-19 cases in neighboring county (county j), thus need to apply spatial econometric models.

Table 3.3: Moran's I test under randomization for regression residuals for OLS model; alternative hypothesis: greater.

	OLS
Moran's I statistic	0.551
Expectation	-0.000
Variance	0.000
Standard Deviate	54.392
P-value	0.000

3.4.2 Lagrange Multiplier (LM) Tests

Using annual cross-sectional data, Lagrange Multiplier (LM) tests are used to compare across models based on residuals of the OLS model and follow a chi-squared distribution (Elhorst 2014; Bivand et al. 2021). The first test is whether SDM is restricted to the OLS model and the

[9] Here, 6 neighbors are the appropriate number of nearest neighbors.

second test is whether SDM is restricted to the spatial error Durbin model (SEDM). The likelihood ratio (LR) test of the SDM model versus the OLS model takes the value of 1903.8 with 1 degrees of freedom (df), and 1% p-value (Table 3.6). Similarly, the LR-test of the SDM model versus the SEDM model takes the value of -245.04 with 20 df and 1% p-value (Table 3.4). On the basis of these LM tests, the SDM is more appropriate.

Table 3.4: Likelihood ratio for spatial linear models.

	SDM vs OLS	SDM vs SEDM
Likelihood Ratio	1903.8	-245.04
p-value	0.000	0.000
df = 20	1	20

3.4.3 Spatial Regression

Using annual cross-sectional data for 3045 counties, Table 3.5 reports the results for the SDM model. From the SDM model, a statistically significant spatial coefficient of dependent variables is found (ρ= .75036 with 1% significance level) indicating that an increase/decrease in COVID-19 cases in the county i increase/decrease in COVID-19 cases in county j. Thus, any non-spatial model including OLS is likely to be biased. For comparison, Table 3.1A in the appendix provides OLS and SEM coefficient estimates. The SEM has a positive, statistically significant lambda, thus some factors not included in this model have significant effects on COVID-19 cases.

From the SDM model, statistically significant effects are found for mandatory stay-at-home orders and mask mandates within and beyond the county on the COVID-19 case per 100,000 people. In terms of mandatory stay-at-home orders, a one-month implementation in one county results in a 2.78 percent decrease in COVID-19 cases per 100,000 people within the same county, and a 6.80 percent decrease in COVID-19 cases per 100,000 people in 6-nearest neighboring counties. In terms of mask mandate, if a county implements a mask mandate for one month, then

COVID-19 case per 100,000 people, on an average, reduces by 0.64 percent in the mandate county and 1.58 percent in 6-nearest neighboring counties. Nationally, on average, four cases per 100,000 people are reduced within the mandate county while nine cases per 100,000 people are reduced in 6-nearest neighboring counties by implementing mask mandate for one-month. Thus, the spillover benefit is greater in neighboring counties than in mandate county when mask mandates are put in place. On the contrary, fifteen cases per 100,000 people (about four times larger effect than the effect from mask mandate) can be reduced in the county where mandatory stay-at-home order put in place, and thirty-eight cases per 100,000 people (over four times larger effect than the effect from mask mandate) can be reduced in six nearest neighboring counties, respectively. Other NPIs (restaurant and bar limitations and closures and stay-at-home advisories) did not have any statistically significant impacts in the SDM model. Thus, mandates for staying at home are much more effective at preventing COVID-19 cases than advisories. Finally, the SEM model in Table 3.1A provides similar estimates but also provide evidence that some significant spillover effects of variables are not explicitly modeled in this study.

Table 3.5: Fitted Coefficients for Spatial Durbin Model (SDM), by County, 2020.

Variables	Dep. Variable: log(COVID-19 Case per 100,000 people)		
	Direct	Indirect	Total
(a) Variable of Interest:			
STAYHOMEADV (% days)	-.001473	-.003603	-.005086
	(.0013)	(.0033)	(.0046)
STAYHOMEMAN (% days)	-.003379***	-.008269***	-.011648***
	(.0011)	(.0028)	(.0039)
MASKMANDATE (% days)	-.000783***	-.001917***	-.002700***
	(.00028)	(.00069)	(.00097)
RESTLIM (% days)	-.002382	-.005828	-.008211
	(.0074)	(.018)	(.025)
BARLIM (% days)	.000148	.000363	.000511
	(.0074)	(.018)	(.026)
BARCLOSED (% days)	-.000666	-.001631	-.002300
	(.0073)	(.018)	(.025)
(b) Non-compliance Factor:			
TRUMPVOTES (%)	-.000051	-.000124	-.000174
	(.0006)	(.0015)	(.0022)
(c) Demographic Factors			
ADULT (% aged≥ 65 years)	-.02366***	-.057891***	-.08155***
	(.0016)	(.0052)	(.0063)
MALE (%)	.016031***	.039229***	.055260***
	(.0025)	(.0069)	(.0093)
EDUCATION (% with bachelor's degree or higher	-.00260**	-.006369**	-.00897**
	(.0012)	(.0029)	(.0041)
URBAN (Dummy)	-.34679***	-.084859***	-.11954***
	(.014)	(.034)	(.048)
NON-WHITE (%)	.00155***	.003782***	.005327***
	(..00063)	(.0016)	(.0022)
POPDENSITY (per square mile)	.000001	.0000034	.000005
	(.000005)	(.000012)	(.000017)
(d) Socioeconomic Factors:			
INCOME ($)	-.000005***	-.000013***	-.000018***
	(.000000)	(.00000)	(.00000)
POVERTY (% of all ages)	-.006554***	-.016037***	-.022591***
	(.0023)	(.006)	(.008)
UNEMPLOYMENT (%)	.005915**	.0144743**	.020389**
	(.003)	(.008)	(.012)
HOUSING (% Housing Units with more people than Rooms)	.0017984	.004401	.0061994
	(.003)	(.008)	(.012)
(e) Specific Medical Conditions:			
COVIDRISK (% ≥ 18 years)	-.000075	-.000184	-.000259
	(.0003)	(.0008)	(.0012)
(f) Lifestyle Factors:			
SMOKING (% current smokers ≥ 18 years)	-.006600***	-.016150***	-.022750***
	(.003)	(.0068)	(.009)
REMOTEWORK (% aged≥ 16 years)	-.009795***	-.023968***	-.033762***
	(.002)	(.005)	(.007)
Rho		.75036***	
Obs.		3045	

Note: *** for 0.01, ** for 0.05, * for 0.1. Numbers in the parentheses represent standard error estimates.

Using SDM model, Table 3.6 demonstrates the direct, indirect, and total effects on COVID-19 spread from one-month implementations of a mask mandate and a combined mask mandate and stay-at-home order. The direct effect of public mask mandate is 4 fewer cases per 100,000 people in the mandate county while the indirect effect of the same policy is 9 fewer cases per 100,000 people in the six nearest neighboring counties. Finally, the total effects of the mask mandate policy result in 13 fewer cases per 100,000 people. On the other hand, the direct, indirect, and total effects of mandatory stay-at-home order are 15, 38, and 53 fewer cases per 100,000 people, respectively. Thus, the impact of mandatory stay-at-home order is larger than that of mask mandate. One limitation of the SDM model is that the temporal effects on COVID-19 cases could not be captured by this model.

Table 3.6: Effect of One-Month Implementation of NPIs on COVID-19 Case per 100,000 people.

	Mask Mandate (per 100,000 people)	**Mandatory Stay-At-Home Order (per 100,000 people)**
Direct Effect	4 fewer cases	15 fewer cases
Indirect Effect	9 fewer cases	38 fewer cases
Total Effect	13 fewer cases	53 fewer cases

The monetary benefits from reduced COVID-19 cases are calculated based upon the value per nonfatal statistical case saved (Robinson et al. 2021). Table 3.7 shows monetary benefits from Powder River County in the state of Montana (a mandate state) and the three-nearest neighboring counties in the state of Wyoming (which is a non-mandate state) [10]. Benefit calculations use SDM

[10] Here, benefits are calculated from Wyoming counties due to no mask or stay-at-home mandates being implemented in this state while the state of Montana implemented both these policies. One caveat to these benefit calculations is the relatively small size of population in Montana and Wyoming, which is why, monetary benefits are small in size.

model results and compute to direct monetary benefits of $0.01 million, $0.04 million, and $0.05 million from Powder River County, Montana due to one-month implementation of mask mandate, mandatory stay-at-home order, and both these policies, respectively. As a non-mandate state, spillover benefits to Wyoming from this one county are $0.70 million, $2.95 million, and $3.65 million from three-nearest neighboring counties (Sheridan, Campbell, Crook) by implementing mask mandate, mandatory stay-at-home order, and both these policies, respectively for one month in Powder River County, Montana. The bulk of these reduced COVID-19 cases benefits come in the form of reduced hospitalizations given the high rate of hospitalizations (8% in the state of Montana, and 6% in the state of Wyoming) from COVID-19 cases in both states.

Table 3.7: Monetary Benefits from One-Month Implementation of NPIs.

Type of NPI	**Direct Benefits from Powder River County, Montana (Million $)**	**Spillover Benefits from three-nearest neighboring Counties in Wyoming (Million $)**
Mask Mandate	0.01	0.70
Mandatory stay-at-home order	0.04	2.95
Mask mandate and Mandatory stay-at-home order	0.05	3.65

Note: The spillover benefits calculations are based upon the three-nearest neighboring Counties (Sheridan, Campbell, and Crook in the state of Wyoming) of Powder River County in the state of Montana, where Powder River County is a mandate County.

From the cross-sectional model in Table 3.5, interpretations of control variable results are presented below. While few studies claimed that political stringency and trust in a county and state influence COVID-19 case incidence, no statistically significant impact from Trump vote percentage is found, which contradicts prior findings of Kahane (2021). Similarly, population density at the county level does not influence COVID-19 case per 100,000 people. The male population is disproportionately affected by COVID-19 cases compared to its counterpart because the male population tends to move outside more frequently than the female population. Similarly, communities of color are more vulnerable to COVID-19 infections than white counterpart within the county and its neighboring counties, thus supports racial disparity in COVID-19 infections.

People holding a higher rate of bachelor's degrees and above in a county have less chance of COVID-19 cases compared to a county with less literacy rate which supports the findings of Yoshikawa and Asaba (2021). Remote work by the able-bodied labor force helps to reduce COVID-19 cases within the county and its nearest counties which strengthen the argument for shelter at home policy. Based on annual cross-sectional data, a statistically significant negative coefficient of the urban variable means that rural communities are affected more by COVID-19 because of limited medical care (Dobis and McGranahan, 2021). Trump votes and healthcare inequities[11] (e.g. insufficient testing facilities) might influence COVID-19 in the rural communities. Finally, a significant negative coefficient for poverty indicates that low income communities are less infected by COVID-19. Explanation behind this sign coefficient is two folds: one possible explanation is poor compliance of NPIs in low-income communities, other explanation is the less availability of testing resources in disadvantaged communities (Finch and Finch, 2020).

3.5 Conclusions and Policy Implications

This study aims to investigate the spatial spillover effects of NPIs on the spread of COVID-19 cases in the contiguous U.S. counties using annual cross-sectional data. From the SDM model, both mandatory stay-at-home orders and mask mandate show statistically significant direct, indirect, and total effects on COVID-19 cases per 100,000 people with control variables of other socioeconomic, demographic, non-compliance, specific medical conditions, and lifestyle factors. For other NPIs, no statistically significant impacts were found on COVID-19 spread.

[11] https://www.idsociety.org/globalassets/idsa/public-health/covid-19/covid19-health-disparities-in-rural-communities_leadership-review_final_ab_clean.pdf

The indirect effects in neighboring counties are larger than the direct effects from mask mandate and mandatory stay-at-home order implemented in a particular county i. Based upon SDM results, implementing mask mandates for one-month results in 4 fewer cases per 100,000 people in mandate county and 9 fewer cases per 100,000 people in 6-nearest neighboring counties. Implementing mandatory stay-at-home order for same period results in 15 fewer cases per 100,000 people in mandate county, and 38 fewer cases per 100,000 people in the 6-nearest neighboring counties. This indicates that the contribution of stay-at-home order is larger than of mask mandate in mandate and non-mandate counties. These results corroborate the findings of Ginther and Zambrana (2021) and Huang et al. (2022) that mask mandates reduced the number of COVID-19 cases. These results further indicate that the indirect effects are larger than the direct effects of both these policies on COVID-19 spread.

These research results demonstrate the important role of spatial spillovers for policymakers to consider especially in the context of economic burden saved. From the patient perspective, individuals can avoid economic burdens from hospitalization and productivity loss by implementing NPIs. Research results also indicate that policymakers should put more emphasis on mask mandates and mandatory stay-at-home orders relative to other NPIs (such as stay-at-home advisories) to control infectious diseases like COVID-19.

Results on remote working, rurality, race, and poverty are factors indicative of less stringent NPIs policy compliance and/or less available testing resources. Compared to metro counties, policymakers should ensure equal access to testing facilities and enhance health insurance coverage in rural areas to protect rural communities from infectious diseases like COVID-19. Also, given the long-standing history of structural racism, residential segregation, and social risk in the U.S., policymakers should ensure justice in screening, and symptom presentation of infectious diseases like COVID-19 across races and ethnic communities. Federal and state governments should enhance financial benefits under the Coronavirus Aid, Relief, and Economic Security (CARES) Act to disadvantaged low-income communities particularly less educated communities of color. Finally, from the societal perspective, federal and state governments can save resource by designing cost-efficient interventions to tackle future pandemics based on these results.

Under-reporting is one of the major limitations of COVID-19 research. Since regressions are based on reported cases only, there is no accounting for unreported cases. Unreported cases lead to an undercount of actual cases. Secondly, in the early phase of 2020, a higher rate of COVID-19 infections in a period spreads the disease in the following period while lower infections in a period lessen the disease and grow immunity among the public in the following period in the second phase of 2020. Temporal dependence and the effects of NPIs on COVID-19 spread are time-dependent. This paper could neither consider differential impacts of policies across different waves[12] nor incorporate time such as time fixed effect (FE) in the model. Future researchers may want to apply a spatiotemporal autoregressive distributed lag (STADL) model to avoid misspecification of temporal dependence (Cook, et al. 2023). One challenge of using a STADL model is that the average recovery period of COVID-19 is shorter than the one-month data frequency used here. For example, if a person is infected, it takes up to fourteen days to fully recover[13] which means that the temporal overlap between months is not consistent.

Finally, one controversial NPI (school closures) is not evaluated in this research (Chernozhukoy et al. 2021b). However, no county-level school closure data could be obtained in

[12] https://www.hopkinsmedicine.org/health/conditions-and-diseases/coronavirus/first-and-second-waves-of-coronavirus
[13] https://driphydration.com/blog/covid-19-timeline/

this research. Future research should include this variable as part of NPIs into a model if data become available in order to assess the efficacy of this controversial practice.

References

Abaluck, J. et al. (2022). Impact of community masking on COVID-19: A cluster-randomized trial in Bangladesh. *Science, 375*(160).

Abedi, V. et al. (2020). Racial, Economic, and Health Inequality and COVID-19 Infection in the United States. *Journal of Racial and Ethnic Health Disparities 8*:732–42.

Arbia, G. (2014). *A Primer for Spatial Econometrics: With Applications in R*. New York: Palgrave Macmillan, pp. 33-35.

Bargain, O. and Aminjonov, U. (2020). Trust and compliance to public health policies in times of COVID-19. *Journal of Public Economics*, *192*, 104316.

Benitez, J. et al. (2020). Racial and ethnic disparities in COVID-19: Evidence from six large cities. *Journal of Economics, Race, and Policy, 3*: 243-261.

Bivand, R. et al. (2021). A Review of Software for Spatial Econometrics in R. *Mathematics,* 9: 1276. https://dx.doi.org/10.3390/math9111276

CDC (2020a). *Assessing Risk Factor for Severe COVID-19 Illness*, Centers for Disease Control and Prevention (CDC). Accessed on 19 September 2020 https://www.cdc.gov/coronavirus/2019-ncov/covid-data/investigations-discovery/assessing-risk-factors.html

CDC (2020b). *People with Certain Medical Conditions,* Centers for Disease Control and Prevention (CDC). Accessed on 19 September 2020 https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-with-medical-conditions.html

CDC (May 5, 2023). *COVID Data Tracker*. Centers for Disease Control and Prevention. Access: https://covid.cdc.gov/covid-data-tracker/#datatracker-home

Chagas, A.L., et al. (2016). A spatial difference-in-differences analysis of the impact of sugarcane production on respiratory diseases. *Reg. Sci. Urban Econ.*, *59*: 24–36.

Chen, X. et al. (2017). Impacts of air pollution and its spatial spillover effect on public health on China's big data sample. *Journal of Cleaner Production, 142* (2), 915-925.

Chernozhukov, V. et al. (2021a). Causal impact of masks, policies, behavior on early COVID-19 pandemic in the US. *Journal of Econometrics, 220* (1), 23-62.

Chernozhukov, V. et al. (2021b). The association of opening K-12 schools with the spread of COVID-19 in the United States: County-level panel data analysis. *Proceedings of the National Academy of Science, 118 (42):* e2103420118.

CNN (March 9, 2023). Opinion: Were masks useless? The deceptive interpretation of what science tells us. Accessed: https://www.cnn.com/2023/03/08/opinions/mask-mandate-cochrane-study-covid-sepkowitz-ctrp

Cook, S.J. et al. (2023). STADL Up! The spatiotemporal autoregressive distributed lag model for TSCS data analysis. *American Political Science Review, 117* (1): 59-79.

Correa-Areneda, F. et al. (2021). Environmental determinants of COVID-19 transmission across a wide climatic gradient in Chile. *Scientific Reports, 11*: 9849.

Courtemanche, C. J. et al. (2021). School reopening, mobility, and COVId-19 spread: Evidence from Texas (NBER working paper No. 28753). MA: National Bureau of Economic Research.

Crankson, S. et al. (2021). Determinants of COVID-19 outcomes: A systemic review. *MedRxiv* (pre-print version). Accessed on December 8, 2021 https://doi.org/10.1101/2021.03.21.21254068 .

Cui, Z. et al. (2020). *The Political Economy of Responses to COVID-19 in the U.S.A.* online seminar hosted by Italian Association of Environmental and Resource Economists (IAERE) on December 11, 2020. Accessed on January 9, 2021: https://www.youtube.com/watch?v=7M7jQ0PyoAY

Cui. Z. et al. (2021). Economics affects mobility, and ideology affects mask-wearing: How COVID-19 drifted to the red areas within the USA in 2020? *UCLA Anderson Review*. Accessed on September 7, 2021: https://anderson-review.ucla.edu/wp-content/uploads/2021/03/COVID-19-USA-2021-Feb-.pdf

Dobis, E.A. and McGranahan, D. (February 1, 2021). Rural Residents Appear to be More Vulnerable to Serious Infection or Death From Coronavirus COVID-19. Accessed: https://www.ers.usda.gov/amber-waves/2021/february/rural-residents-appear-to-be-more-vulnerable-to-serious-infection-or-death-from-coronavirus-covid-19/

Eikenberry, S.E. et al. (2020). To mask or not to mask: Modeling the potential for face mask use by the general public to curtail the COVID-19 pandemic. *Infectious Disease Modeling, 5*, 293-308.

Elhorst, J. P. (2014). *Spatial Econometrics: From Cross-sectional Data to Spatial Panels*. Springer Briefs in Regional Science.

Emch, M., et al. (2012). Integration of spatial and social network analysis in disease transmission studies. *Ann. Assoc. Am. Geogr. 102* (5): 1004-1015.

Espana,G. et al. (2021). Impacts of K-12 school reopening on the COVID-19 epidemic in Indiana, USA. *Epidemics 37*:100487. doi: 10.1016/j.epidem.2021.100487.

Finch, W. H. and Finch, M. E. H. (2020). Poverty and Covid-19: Rates of Incidence and Deaths in the United States During the First 10 Weeks of the Pandemic. *Frontiers in Sociology, 5*. https://doi.org/10.3389/fsoc.2020.00047.

Ginther, D.K. and Zambrana, C. (2021). Association of mask mandates and COVID-19 case rates, hospitalizations, and deaths in Kansas. *JAMA Network Open, 4(6)*: e2114514. DOI: 10.1001/jamanetworkopen.2021.14514

Guliyeb, H. (2020). Determining the spatial effects of COVID-19 using the spatial panel data model. *Spatial Statistics, 38*: 100443.

Gupta, S. et al. (2021). Tracking public and private responses to the COVID-19 epidemic: Evidence from State and Local government actions. *American Journal of Health Economics, 7* (4).

Huang, J. et al. (2022). The effectiveness of government masking mandates on COVID-19 county-level case incidence across the United States, 2020. *Health Affairs, 41 (3)*. Accessed: https://doi.org/10.1377/hlthaff.2021.01072

IRS (2023). CARES Act Coronavirus Relief Fund frequently asked questions. Internal Revenue Service. Retrieved from: https://www.irs.gov/newsroom/cares-act-coronavirus-relief-fund-frequently-asked-questions

Isphording, I.E. et al. (2021). Does re-opening schools contribute to the spread of SARS-CoV-2? Evidence from staggered summer breaks in Germany. *Journal of Public Economics, 198*: 104426.

Jackson, S.L. et al. (2021). Spatial disparities of COVID-19 cases and fatalities in United States Counties. *Int. J. Environ. Res. Public Health, 18* (16), 8259.

Kahane, L.H. (2021). Politicizing the mask: Political, economic and demographic factors affecting mask wearing behavior in the USA. *Eastern Economic Journal, 5*, 1-21.

Kim, D. et al. (2022). Support for regulation versus compliance: Face masks during COVID-19. *Public Health in Practice,* In-Press. DOI: https://doi.org/10.1016/j.puhip.2022.100324

Krisztin, T., et al. (2020). The spatial econometrics for the corona pandemic. *Letters in Spatial and Resource Sciences, 13*: 209-218.

Lee, W. et al. (2021). Urban environments and COVID-19 in three Eastern states of the United States. *Science of Total Environment*, 779: 146334.

LeSage, J.P. (2014). What Regional Scientists Need to Know about Spatial Econometrics? *The Review of Regional Studies, 44,* 13-32.

Moreland, A. et al. (2020). Timing of state and territorial COVID-19 stay-at-home orders and changes in population movement — United States, March 1–May 31, 2020. *Morbidity and Mortality Weekly Report (MMWR), 69*:1198–1203. DOI: http://dx.doi.org/10.15585/mmwr.mm6935a2.

Nelson, P. and Cromartie, J. (2022). Migration, local mobility, and the spread of COVID-19 in rural America (COVID-19 Working Paper, AP-107). USDA Economic Research Service.

Olmo, J. and Sanso-Navarro, M. (2021). Modeling the spread of COVID-19 in New York City. *Papers in Regional Science, 100* (5), 1209-1229.

Orea, L. and Alvarez, I. C. (2021). How effective has the Spanish lockdown been to battle COVID-19? A spatial analysis of the coronavirus propagation across provinces. *Health Economics*, 1-20. Accessed: https://doi.org/10.1002/hec.4437

Rader, B. et al. (2021). Mask-wearing and control of SARS-CoV-2 transmission in the USA: a cross-sectional study. *Lancet Digital Health*. Accessed on January 31, 2021: https://doi.org/10.1016/S2589-7500(20)30293-4

Robinson, L. A. et al. (2021). Valuing COVID-19 mortality and morbidity risk reductions in U.S. Department of Health and Human Services Regulatory Impact Analyses (Report). Office of the Assistant Secretary for Planning and Evaluation. Accessed: https://aspe.hhs.gov/sites/default/files/2021-08/valuing-covid-risks-july-2021.pdf

Tori, L. et al. (2022). Lifting Universal Masking in Schools- Covid-19 Incidence among Students and Staff. *New England Journal of Medicine*. DOI: 10.1056/NEJMoa2211029

Ulimwengu, J. and Kibnonge, A. (2021). Spatial spillover and COVID-19 spread in the U.S. *BMC Public Health, 21* (1765).

Wager, E. et al. (2022). Cost of COVID-19 Hospital Admissions among People with Private Health Coverage. Retrieved from: https://www.kff.org/coronavirus-covid-19/issue-brief/cost-of-covid-19-hospital-admissions-among-people-with-private-health-coverage/

Wang, H., et al. (2015). Detecting the association between meteorological factors and hand, foot, and mouth disease using spatial panel data models. *Int. J. Infect. Dis. 34*: 66-70.

Weng, M. et al. (2017). Area deprivation and liver cancer prevalence in Shenzhen, China: A spatial approach based on social indicators. *Soc. Indic. Res., 133,* 317-332.

Yoshikawa, M. and Asaba, K. (2021). Educational Attainment Decreases the Risk of COVID-19 Severity in the European Population: A Two-Sample Mendelian Randomization Study. *Frontiers in Public Health*, 9, 673451.

Zambrana, C. and Ginther, D.K. (2020). Do masks matter in Kansas? The University of Kansas, Institute for Policy & Social Research. Accessed on September 7, 2021: https://ipsr.ku.edu/covid19/images/MaskMandateUpdate.pdf

Zhang, K. et al. (2020). The impact of mask-wearing and shelter-in-place on COVID-19 outbreaks in the United States. *International Journal of Infectious Disease*, *101,* 334-341.

Appendix:

Table 3.1A: Fitted spatial regression model coefficients for OLS and SEM models, by County, 2020.

Variables	Dep. Variable: log(COVID-19 Case per 100,000 people)	
	OLS	**SEM**
(a) Variable of Interest:		
STAYHOMEADV (% days)	-.0008***	.00119
	(.002)	(.0023)
STAYHOMEMAN (% days)	-.0096***	.00208
	(.0014)	(.00205)
MASKMANDATE (% days)	-.002***	-.00118**
	(.0003)	(.00055)
RESTLIM (% days)	-.057***	-.0017
	(.009)	(.015)
BARLIM (% days)	.051***	-.00054
	(.009)	(.015)
BARCLOSED (% days)	.049***	-.00300
	(.009)	(.0148)
(b) Non-compliance Factor:		
TRUMPVOTES (%)	.004***	.0021***
	(.0008)	(.00084)
(c) Demographic Factors		
ADULT (% aged ≥ 65 years)	-.029***	-.024***
	(.0019)	(.00147)
MALE (%)	.0012***	.01066***
	(.003)	(.0021)
EDUCATION (% with bachelor's degree or	-.005***	-.0029***
higher	(.0015)	(.0012)
URBAN (Dummy)	-.054**	-.0096
	(.017)	(.015)
NON-WHITE (%)	.0052***	.0038***
	(.00078)	(.00081)
POPDENSITY (per sq. mile)	.00001	-.000013***
	(.000006)	(.00000)
(d) Socioeconomic Factors:		
INCOME ($)	-.000006***	-.000003***
	(.000001)	(.00000)
POVERTY (% of all ages)	-.012***	.0059***
	(.003)	(.00233)
UNEMPLOYMENT (%)	-.0034	.010***
	(.0043)	(.0037)
HOUSING (% Housing Units with more	-.006	.0025
people than Rooms)	(.0044)	(.0032)
(e) Specific Medical Conditions:		
COVIDRISK (% ≥ 18 years)	-.0003	.00085***
	(.0004)	(.00032)
(f) Lifestyle Factors:		
SMOKING (% current smokers ≥ 18 years)	-.0056	-.021***
	(.003)	(.0035)
REMOTEWORK (% aged≥ 16 years)	-.01***	-.015***
	(.0025)	(.0019)
Constant	7.287***	6.6914***
	(.2114)	(.172)
Lambda		.83019***
Obs.	3045	3045

Note: *** for 0.01, ** for 0.05, * for 0.1. Numbers in the parentheses represent standard error estimates.

CHAPTER 4:

ESSAY 3: DECARBONIZATION OF CITIES: MYTH OR REALITY?

4.1 Introduction

Urbanization is the process of population migration from rural to urban areas by which social and economic characteristics change over time.[14] The economic globalization that started a century ago continuously creates the opportunity for the entire world, both developed and developing countries, to be involved in urban development. The United Nations Department of Economic and Social Affairs (UN DESA) data show that the global average urbanization rate rose from 29.6% in 1950 to 54% in 2014 (UN, 2015). It is projected that the urbanization rate in developed countries will increase to 86% by 2050, and up to 67% of the global population will live in urban areas.

Urbanization is often associated with an increasing demand for energy and rising greenhouse gas (GHG) emissions, the main contributor to global warming. GHG emissions in urban areas, particularly Carbon Dioxide (CO_2) emissions, are mostly generated from the combustion of fuels in the energy and transportation sectors which mainly rely on fossil fuels such as natural gas and crude oil (Crippa et al., 2021). Gurney et al. (2022) estimated that global urban areas were responsible for 61.8% of the total GHG emissions in 2015. O'Neill et al. (2010) report that GHG emissions from urban areas in developing countries will be responsible for more than 25% increase in the total global GHG emissions by 2100.

Reducing GHG emissions in urban areas is key to tackling global climate change. The United Nations' sustainable development goals (SDGs) state that countries should make cities more resilient and sustainable in dealing with climate change. One of the initiatives is to significantly reduce air pollution by 2030 while ensuring access to affordable, reliable, sustainable, and modern energy for all citizens. Of particular importance is accelerating the transition to

[14] See, https://www.sciencedirect.com/topics/social-sciences/urbanization

modern renewable energy, especially in the transport and energy sectors which are the main contributors to GHG emissions (IEA, 2021).

Significant regional- and country-level efforts have taken place to reduce GHG emissions from urban areas. Under the Urban Agenda Energy Transition Partnership (UAETP) action plan, the European Union (EU) aims to improve energy efficiency at the city level and increase the amount of local and renewable energy. The Association of Southeast Asian Nations (ASEAN[15]) has agreed to cooperate in developing the renewable energy sector as one of the potential solutions to reduce GHG emissions (Lidula et al., 2007). Following their efforts, the Asian Development Bank (ADB) and Japan's Ministry of Economy, Trade and Industry (METI) recently signed a cooperation agreement to enhance their endeavors to increase renewable energy consumption and reduce CO_2 emissions (ADB, 2021). In the U.S., the Biden administration announced the climate plan to reduce carbon emissions by 50% by 2035, in part by increasing energy efficiency and the share of renewable energy in the energy portfolio. China, one of the most significant contributors to GHG emissions, also made efforts to reduce carbon emissions. For example, in 2006, China proposed lowering the energy intensity by 20% and GHG emissions by 10% at the national level.[16] In June 2011, it revealed a comprehensive policy for energy-saving and emissions reduction in the country (Lin and Zhu, 2019a).

Since cities are both "causes and cures" for GHG emissions, an interesting question arises: what is the relationship between urbanization and CO_2 intensity? Further, given the dominance of energy consumption on carbon emissions, how does the development of the renewable fuel sector mitigate the impact of urbanization on CO_2 intensity? Many studies have examined how

[15] ASEAN is a regional group of countries comprises Brunei, Cambodia, Indonesia, Laos, Malaysia, Myanmar, the Philippines, Singapore, Thailand, and Vietnam, grouped for economic, political, and security cooperation.
[16] See https://www.iea.org/policies/6277-china-13th-renewable-energy-development-five-year-plan-2016-2020, accessed on 25 November 2022.

urbanization affects GHG emissions, CO_2 intensity in particular. However, the results are somewhat mixed. For example, Meng et al. (2021) and Salahuddin et al. (2019) find that CO_2 emissions is positively linked to urbanization. Meanwhile, Kim et al. (2020) and Sadorsky (2014) find that urbanization did not significantly affect GHG emissions, while Ali et al. (2017) find a significant negative relationship between urbanization and carbon emissions. Zhu et al. (2012) and Shahbaz et al. (2016), on the other hand, find a nonlinear relationship between urbanization and CO_2 emissions. While a vast majority of research efforts used CO_2 emissions as a dependent variable into their studies, only Zhang, et al. (2014) used CO_2 intensity and found long-term impact of urbanization on environmental degradation in China.

Both per capita CO_2 emissions and CO_2 intensity provide measures of emissions that primarily comes from burning fossil fuels. While CO_2 emissions per capita consider emissions with respect to population size, CO_2 intensity accounts for emissions with respect to economic growth[17]. In fact, CO_2 intensity aims to reduce emissions by maintaining normal economic growth (Zhang et al., 2014). Therefore, this study used both per capita CO_2 emissions and CO_2 intensity as dependent variables to examine how economic growth in compare to population growth influences environmental degradation. Based on global panel data, this study shows an inverted U-shaped relationship between urbanization and CO_2 intensity while U-shaped relationship between urbanization and per capita CO_2 emissions.

This lack of consensus in the literature on the urbanization-environmental degradation relationship challenges the accuracy of carbon emission forecasting models, which are vital to policy initiatives aiming to mitigate global warming. Several factors may explain the mixed conclusions, in particular the use of environmental degradation indicators, different sample

[17] https://fortune.com/2021/06/25/carbon-intensity-emissions-climate-change-paris-agreement/

periods, estimation methods, and study areas in the empirical design. For instance, most of the existing studies consider only single-country (or city) or regional-level panel data (Shah et al., 2020; Kim, 2020; Zhang et al., 2014; Yao et al., 2018; Shahbaz et al., 2016; Ali et al., 2019). Given the different stages of economic development and the vastly varying policies toward climate change and urbanization across countries and cities, conclusions regarding the urbanization-CO_2 intensity relationship are likely highly dependent on the study areas considered. A global approach accounting for countries/cities from both developed and developing economies with different environmental policies is warranted to reveal the average impact of urbanization on CO_2 intensity. A similar concern applies to the sample period used for empirical analyses. All else equal, a longer sample period encompassing various stages of economic growth is likely to yield results that are more consistent with real-world observations.

Another concern with the empirical literature is the failure to account for nonlinear effects. While some researchers find that the impact of urbanization on CO_2 emissions is U-shaped, others show that the impact is inverse U-shaped (Zhu et al., 2012; Shahbaz et al., 2016). Even for studies that find the latter, the threshold at which the impact of urbanization on CO_2 emissions becomes negative is rather different. For example, Du and Xia (2018) find that urbanization increases CO_2 emissions when the urbanization rate is below 48 percent, while Zhang et al. (2017) show that the turning point is 73 percent. A more comprehensive study that accounts for nonlinear relationships is needed to determine the threshold level for urbanization.

While a few studies investigated the urbanization-CO_2 linkage using global panel data and accounted for the possible presence of a nonlinear relationship (Zhang et al., 2017; Chikaraishi et al., 2015), they do not consider the impact of the energy transition. All else equal, an economy where a larger percentage of energy consumption is derived from renewable resources is linked with lower per capita GHG emissions. For example, Cherni and Jouini (2017) find that government policies and regulations that stimulate renewable energy in their energy mix are mainly designed to reduce the reliance on fossil fuels which are the major contributor to GHG emissions. Lin and Zhu (2019b) show that renewable energy sources significantly contribute to reducing CO_2 emissions, further highlighting the importance of considering energy transition when investigating the relationship between urbanization and CO_2 intensity.

The aim of this essay is to examine the relationship between country-level urbanization and GHG emissions by applying dynamic panel data approach. There are two measures that impact GHG in the atmosphere: CO_2 intensity and CO_2 emissions. In this research, the first objective is to estimate the short-term and long-term impacts of urbanization on GHG emissions across countries. Second objective is to test the non-linear relationship between urbanization and GHG emissions in the sample countries. Final objective is to estimate the interaction effect of urbanization and renewable energy use on GHG emissions in the sample countries.

By using panel data for 105 countries from 1990 to 2018, pooled mean group (PMG) estimation technique is applied to examine the impact of urbanization on CO_2 intensity. The analysis follows the urban pollution hypothesis, which argues that the relationship between environmental degradation and urbanization is inverse U-shaped. Evidence consistent with the urban pollution hypothesis is found in the long run using CO_2 intensity as the dependent; however, no significant relationship is identified in the short run. On the other hand, this urban pollution hypothesis does not exist when per capita CO_2 emissions are used as the dependent variable. Both these models further evident that renewable energy consumption significantly reduce environmental degradation. By introducing an interaction term into these models, no significant impact of renewable energy use is found on CO_2 intensity but significant positive effect on per capita CO_2 emissions beyond a certain level (threshold) of urbanization.

This essay contributes to the literature in at least four ways. First, a large-scale, global panel data set spanning 29 years across 105 countries is used to investigate the impact of urbanization on CO2 intensity. Second, both linear and nonlinear effects are accounted for in the estimation, which allows for estimating the threshold at which the impact of urbanization changes the sign. Third, long- and short-run relationships are examined and compared between urbanization and both CO_2 intensity and emissions. Finally, the impact of energy transition is accounted for in modelling by examining how renewable energy consumption mediates the effect of urbanization on CO_2 intensity.

The rest of the paper proceeds as follows. Section two discusses the hypotheses and briefly reviews the literature. Sections three and four discuss the data and empirical strategy, respectively. Results are discussed in section five, and the last section provides conclusions and policy implications.

4.2 Testing Hypotheses and A Brief Review of Related Literature

Our theoretical underpinning draws heavily on the Environmental Kuznets Curve (EKC) literature, which argues that a nonlinear relationship exists between environmental quality and economic growth. In the early stage of development, economic development takes priority over environmental quality, resulting in an increasing marginal rate of environmental degradation as economic growth. As economic growth reaches a certain level, society becomes more aware of environmental quality, with significant efforts being paid to reduce environmental degradation. This leads to a decreasing marginal rate of environmental degradation as economic growth. Under the EKC hypothesis, the relationship between environmental degradation, such as pollution, and income should follow an inverse U shape. A vast majority of the literature investigates the relationship between income and CO_2 emissions and confirms the validity of the EKC hypothesis. For example, Apergis and Ozturk (2015) utilized data from 14 Asian countries, finding a statistically significant relationship that validates the EKC hypothesis.

In our context, Poumanyvong and Kaneko (2010) note that three other theories may help explain the relationship between urbanization and carbon emissions, namely ecological modernization, urban environmental transition, and compact city theories. Of the three, the first two theories closely resemble the EKC hypothesis. The theory of ecological modernization views urbanization as a process of social transformation toward modernization, during which environmental problems first worsen as societies move from low to middle stages of development and later lessen as societies play a higher value on environmental quality in higher stages of development (Sadorsky, 2014).

The urban transition theory places more weight on economic sector composition, arguing that the manufacturing base typically expands in the early stages of development, leading to increasing environmental problems. As cities become wealthier, residents demand higher environmental quality, putting pressure on regulations and technological innovation to reduce environmental problems. However, more affluent cities typically produce more pollution due to higher energy demand, complicating the impact of wealth on environmental quality. The compact city theory, meanwhile, argues that urbanization should lead to fewer environmental problems due to the economics of scale and more efficient use of public infrastructure.

Following Brueckner (2011), I start with the urban pollution hypothesis by applying the EKC framework and the three theories discussed above to the urbanization- GHG emissions nexus. In Figure 4.1, I apply GHG emissions as a measure of environmental degradation which shows three stages of the relationship between GHG emissions and urbanization: pre-urbanization, urbanization, and post-urbanization. In the initial stage, as urbanization increases, the amount of GHG emissions goes in the same direction at a mildly increasing rate. When cities grow into the middle stage (the urbanization stage), GHG emissions increase at a decreasing rate and reach the maximum level. Thus, the curve has a turning point when the GHG emissions are at the highest level (i.e., peak emission). During this stage, economic growth is prioritized over environmental quality. The manufacturing base expands rapidly, with little attention paid to GHG emissions. However, the growth rate in GHG emissions decreases as a higher population density generates economies of scale.

Finally, GHG emissions falls when urbanization continues to increase, the so-called post-urbanization period. As cities become wealthier, more attention is paid to environmental sustainability. Technology improvements, environmental regulations, and a shift of economic sector composition (i.e., from manufacture-based economy to service-based economy) allow carbon emissions to decrease even as urban areas continue to expand. In particular, the increasing adoption of carbon-neutral technology and consumption of renewable energy help reduce carbon intensity in advanced countries where urbanization passes the threshold.

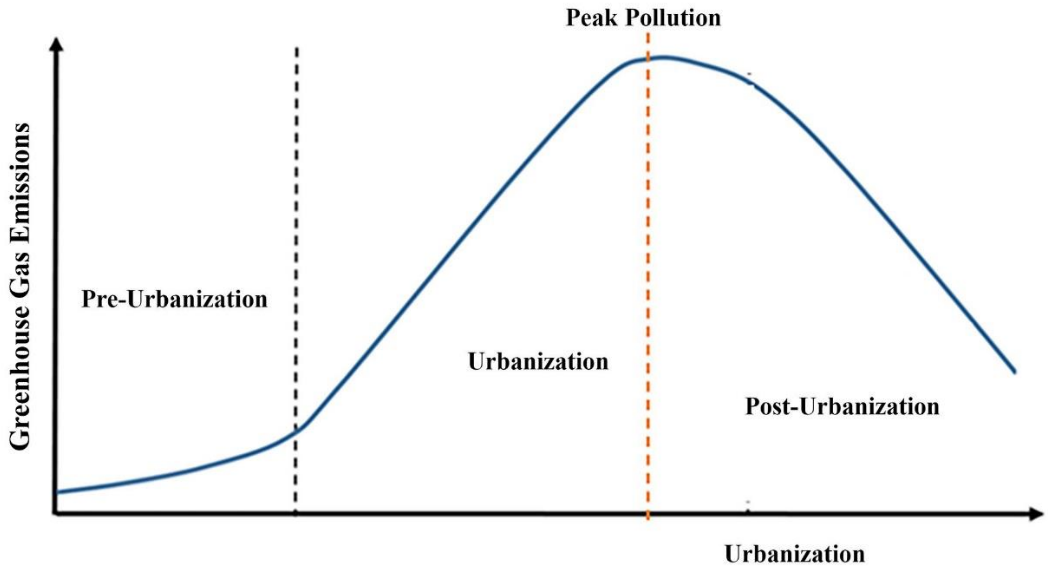


Figure 4.1: Urban Pollution Hypothesis[18].

Based on the discussion above, I propose two hypotheses linking urbanization, renewable energy use, and GHG emissions:

Hypothesis 1: The relationship between urbanization and GHG emissions is nonlinear. In particular, a threshold exists at which GHG emissions decreases even as urbanization continues to expand.

Hypothesis 2: Renewable energy use mitigates the impact of urbanization on GHG emissions. Higher rates of renewable energy use during the post-urbanization stage results in reduced GHG emissions.

A number of existing studies have tested the first hypothesis. As noted earlier, conclusions are mixed. One strand of literature found an inverted U-shaped relationship between CO_2 emissions and urbanization, particularly when using STIRPAT[19] models (Martinez-Zarzoso and

[18] This graph is generated based on Kaika and Zervas (2013).

[19] STIRPAT stands for Stochastic Impacts by Regression on Population, Affluence, and Technology.

Maruotti, 2011; Zhang et al., 2017; Chen et al., 2019). The inverted U-shaped relationship occurs due to the tradeoff between environmental sustainability and economic growth. On the other hand, a few studies rejected this hypothesis and found a U-shaped relationship between urbanization and CO_2 emissions (Shah et al., 2020; Shahbaz et al., 2016). Shah et al. (2020) claimed that the lack of policies toward energy efficiency and clean energy initiatives in Pakistan is the leading cause of this U-shaped relationship at the higher rate of urbanization. Meanwhile, Shahbaz et al. (2016) argued that Malaysia adopted several policies to protect the environment from urbanization at its early stage. However, they did not provide any explanation for the increasing emissions at the later stage of urbanization.

Some studies noted a linear impact of urbanization on carbon emissions. Poumanyvong and Kaneko (2010), for instance, find that urbanization significantly increases CO_2 emissions, with the most significant effect observed in the middle-income group of countries. By contrast, Fan et al. (2006) note that the relationship between urbanization and CO_2 emissions was negatively significant, supporting the compact city theory. In addition, Sadorsky (2014) finds a non-significant impact of urbanization on CO_2 emissions. A similar conclusion was also reached by Liddle and Lung (2010) using panel data from 17 countries for the total CO_2 emissions. Overall, the existing literature is inclusive regarding the relationship between CO_2 emissions and urbanization.

Regarding the second hypothesis, only a few studies have included renewable energy as an explanatory variable in examining the EKC. The reason may be the small share of renewable energy production from the total energy production in many countries. For example, in the U.S., the renewable energy production share from the total U.S. energy production in 2021 was only 13% (EIA, 2022). Baek (2016) studied the effects of renewable and nuclear energy consumption on CO_2 emissions in the U.S., finding that nuclear energy consumption reduces CO_2 emissions in the short and long run, while renewable energy consumption does only in the short run. This conclusion contradicts Apergis et al. (2010), who find that renewable energy does not significantly affect CO_2 emissions in the short run. Kim (2020) finds that renewable energy significantly decreases CO_2 emissions in Korea, a conclusion confirmed by Hanif (2018) for Sub-Saharan African low-income countries. Meanwhile, Chen et al. (2019) found that for Chinese cities, more urbanization resulted in higher use of nonrenewable energy than renewable energy.

However, none of these studies examined how the impact of urbanization on CO_2 emissions varies with renewable energy use. To test Hypothesis 2, a renewable energy share variable is included in the urbanization-CO_2 intensity relationship along with an interaction term of urbanization and renewable energy, which is discussed in the method section.

4.3 Variables and Data

Data used for the analysis consisted of 105 countries from 1990 to 2018. A list of countries considered in the analysis, along with their urbanization levels in 2018, is provided in appendix. Table 4.1 describes the variables used in the analysis. Data on independent variables are collected from the World Development Indicators (WDI)[20] of the World Bank. Data on dependent variables on GHG impacts are collected from Our World in Data[21]. GHG impacts are measured by per capita CO_2 emissions and CO_2 intensity, which is measured by CO_2 emissions per dollar of GDP in metric tons (MT). The CO_2 intensity measure relates the growth in emissions to economic growth. The key explanatory variable is urbanization, which can be defined based on various metrics such as place, area, and population. Here, the focus is on the population dimension and the percentage of the urban population is used rather than the overall population to measure the rate of urbanization.

[20] https://data.worldbank.org/indicator
[21] https://ourworldindata.org/grapher/co2-intensity

This measure has been used in various previous studies such as Zhang et al. (2017) and Zhu et al. (2012). Zhang et al. (2017) note that the population-based measure describes both CO_2 intensity and urbanization level.

Other covariates considered are renewable energy use[22], computed as the percentage of the renewable energy consumption divided by the total final energy consumption. Renewable energy sources include, among others, solar, wind, hydro, biomass, geothermal, municipal waste, and tidal. This variable measure the degree of the energy transition for a given country. Per capita GDP at a constant price of $2010 is used to measure economic growth, which has been shown by various previous studies to significantly affect GHG emissions (e.g. Narayan et al., 2016).

Previous studies find that foreign direct investment (FDI) directly contributes to carbon emissions (Tang and Tan, 2015; Shahbaz et al., 2019; Koçak and Şarkgüneşi, 2018). Other studies find that FDI reduces CO_2 emissions through the adoption of clean energy technology in production (Zhang and Zhou, 2016; Xie et al., 2020), though Lee (2013) finds no evidence of clean energy use. FDI is measured by the net flows of all investment activities at a constant price of $2010 in the analysis. Trade openness is measured by the total exports and imports of goods and services as a percentage of the GDP. Zhang et al. (2017) find that trade openness negatively affects CO_2 emissions. Lv and Xu (2019) find no short-run effect of trade openness but a long-run positive effect on CO_2 emissions. Finally, the structure of the economy is measured by the percentage of GDP attributable to industrial activities. Previous studies note that industrial structural

[22] One criticism of using "renewable energy use" variable is that development direction in very underdeveloped countries does not support hypothesis 2 as more urbanization does not necessarily lead to more biomass use but perhaps even less biomass use and more fossil fuel use. However, Barnes and Floor (1999) claimed that as urban areas expand in developing countries, the transition to renewable energy follows EKC hypothesis of CO_2 emissions. The variable "Share of electricity that comes from low-carbon sources" can be used as a proxy for renewable energy. Since many observations are missing during 1990-1999, and adding this variable restricts the sample size to only 57 countries for 1990-2018 period, I used renewable energy consumption as a percentage of total final energy consumption to cover the longest time period (1990-2018) for the largest (105) number of countries which might not be expected to have a consistent impact on CO_2 intensity at given urbanization.

transformation reduces CO_2 emissions through technical progress (Zhao et al., 2022; Zhou et al., 2013).

Table 4.1: Variable Definition and Summary Statistics (N= 3,045)

Variable	Definition and Units	Mean	S.D.	Min	Max
CO_2 *intensity*	CO_2 emissions (metric ton) per dollar of GDP	0.29	0.25	0.00	2.79
CO_2	CO_2 emissions (metric ton) per capita	3.95	4.48	0	30.36
URBAN	Urban population from the total population, %	57.90	22.80	5.42	100.00
RENERGY	Renewable energy of total final energy consumption, %	36.53	30.89	0.00	98.30
GDP	Per capita GDP, current U.S. dollars (‘000’$)	11.17	17.61	.02	123.68
FDI	net inflows, Million Current U.S. dollars ($)	10766.30	40285.54	-344000.00	734000.00
TRADE	Imports & exports of goods & services as a % of GDP, %	74.90	50.45	0.02	437.33
INDUSTRY	Industry activities as a % of GDP, %	27.12	10.45	4.50	86.67

Notes: S.D. represents the standard deviation.

Summary statistics of the variables are also shown in Table 4.1. Figure 4.2 plots the spatial distribution of CO_2 emissions and urbanization rate in 2019. Urbanization averaged 57% across all countries during the sample period, but the variation across countries is high as evidenced by the large standard deviation (22.80). The lowest urbanization rate observed in Rwanda while the maximum of 100% occurring in Singapore. On average, CO_2 emissions are 0.29 MT per dollar of GDP, with the lowest CO2 intensity in Namibia in 1990 and highest in Mongolia in 1991. The highest per capita CO_2 emissions occurred in Luxembourg. Meanwhile, Mali had the lowest per capita CO_2 emissions, and most people in these countries live in rural areas. For Europe, CO_2 emissions are generally low relative to the urban population.

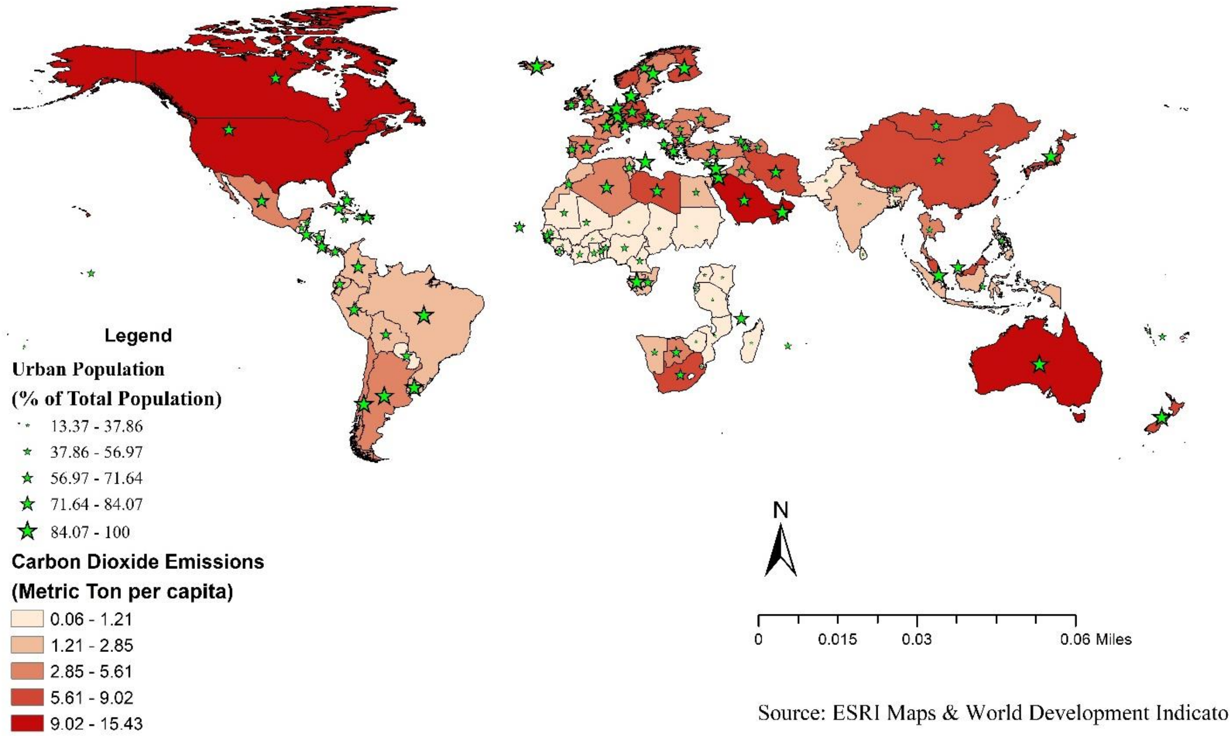


Figure 4.2: Urbanization and CO_2 Emissions in the Study Area, 2019.

For renewable energy share, the minimum percentage (zero) was observed for Oman, Niger, and Malta, while the maximum was 98.30% for Chad in 1991, which is one of the lowest-income countries. The high share of renewables in low-income countries mostly reflects the use of biomass energy for heating and cooking. On average, around 36% of global energy consumption can be attributed to renewable sources. Figure 4.3 plots the spatial distribution of renewable energy consumption and CO_2 emission where the high percentages of renewable energy are located around the world- across both very developed (e.g. some European countries) and very undeveloped parts of the world (e.g. some African countries). This reflects two opposite development directions which both reduce CO_2 emissions: (1) a higher level of development that involves replacement or non-use of fossil fuels to generate electricity via renewable energy (solar, wind, hydro, geothermal, tidal), and (2) a very low level of development that uses mainly biomass for energy, not even getting to a fossil fuel use stage of development. Consumption of renewable energy as a percentage of total energy consumption was also found high in South and Southeast Asia, and South America. Meanwhile, CO_2 emissions are lower in those areas.

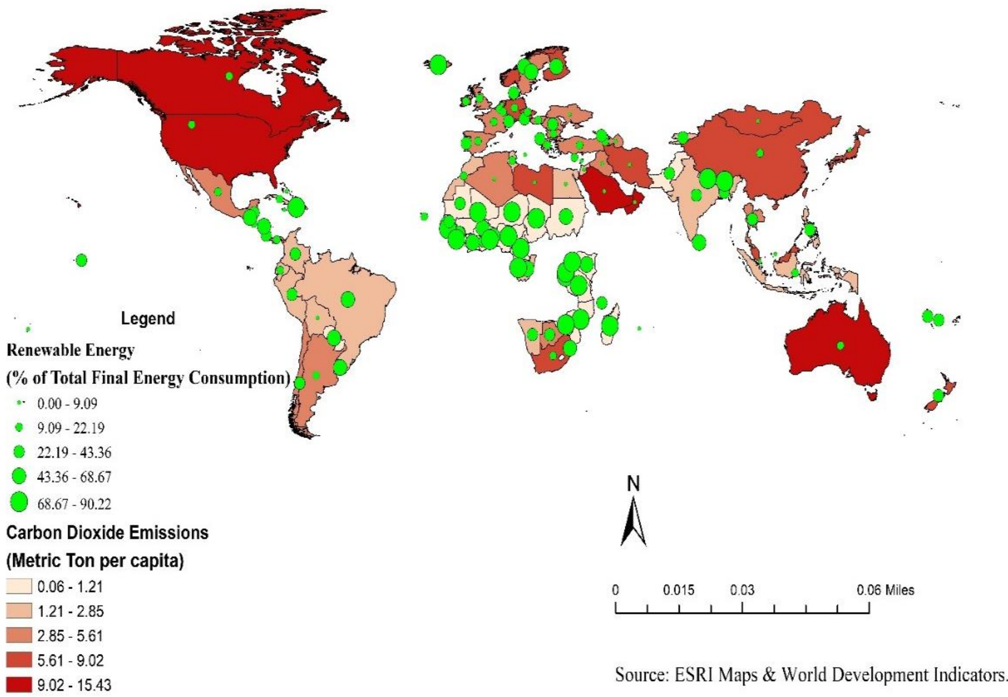


Figure 4.3: Renewable Energy Consumption and CO_2 Emissions in the Study Area, 2019.

4.4 Empirical Strategy

Prior studies primarily rely on fixed effects (FE), random effects (RE), ARDL, and vector autoregressive models when analyzing the impact of urbanization on carb emissions (Ali et al., 2017; Ali et al., 2019; Kim, 2020; Mahmood et al., 2020; Shah et al., 2020; Shi and Li, 2018). Those estimators usually impose homogeneity restrictions and allow only intercepts to vary across panels, which are useful when there is a large number of panels (*N*) but small time-series observations (*T*). If heterogeneity arises across panels over time and serial correlations among regressors exist, then these estimators would provide inconsistent results.

Pesaran et al. (1999) developed the PMG estimator based on a relatively large number of *T* and equal order of magnitude of *N* within the dynamic heterogeneous panels. PMG estimator is more efficient because it considers the long-run restrictions assuming coefficients are the same across panels. It also consistently and efficiently estimates long-run causality while allowing for both stationary and nonstationary variables (Pesaran and Shin, 1998). Given these advantages, the PMG estimator has become a popular approach when investigating long- and short-run relationships in panel data settings (Atasoy, 2017; Iwata et al., 2011; Martinez-Zarzoso and Bengochea-Morancho, 2004; Mert and Boluk, 2016).

I start our estimation using the following equation:

$$
\begin{aligned}
GHG_{it} = \beta_{0t} &+ \beta_{1t}URBAN_{it} + \beta_{2t}URBAN2_{it} + \beta_{3t}RENERGY_{it} + \beta_{4}URBAN * RENERGY_{it} \\
&+ \beta_{5t}GDP_{it} + \beta_{6t}GDP2_{it} + \beta_{7t}FDI_{it} + \beta_{8t}TRADE_{it} + \beta_{9t}INDUSTRY_{it} + \mu_{i} \\
&+ \varepsilon_{it} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (4.1)
\end{aligned}
$$

where GHG_{it} is the GHG emissions proxied by two variables: carbon dioxide emissions in metric tonnes per dollar of GDP, and per capita carbon dioxide emissions in metric tonnes in a country/territory *i* and year *t*, $URBAN_{it}$ and $URBAN2_{it}$ represent the percentage of urban population and its squared value, respectively, $RENERGY_{it}$ is the percentage of renewable energy consumption in total energy consumption, and $URBAN * RENERGY_{it}$ is the interaction term of urbanization and renewable energy consumption to examine how the impact of urbanization on CO_2 intensity varies with the energy transition. Furthermore, GDP_{it} and $GDP2_{it}$ represents the per capita GDP at a constant 2010 US$, and its squared value, FDI_{it} is the annual net inflows of foreign direct investment in million current U.S. dollars, $TRADE_{it}$t is the trade openness, and $INDUSTRY_{it}$ is the structure of the economy. Finally, μ_i is the group effect and ε_{it} is the error term.

Our first hypothesis holds if the coefficient of urbanization variable (β_1) is positive, and the coefficient of squared urbanization (β_2) is negative, confirming the so-called inverted U-shaped relationship. At first, GHG emissions increase with urbanization at a decreasing rate until reaching a maximum point, after which GHG emissions decrease at an increasing rate. The point at which the curve attains its maximum is called the turning point, which is the point on the curve where GHG emissions are at a maximum. The point can be calculated by setting the first derivation of equation (4.1) with respect to *URBAN* equal to zero and solving for urbanization:

$$\beta_1 + 2\beta_2 URBAN = 0 => URBAN = \frac{-\beta_1}{2\beta_2}, \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (4.2)$$

Since an inverted-U shaped relationship exists between environmental degradation and income, I included per capita GDP and its squared value into the model to avoid multicollinearity between urbanization and income (Kuznets, 1955). Further, GHG emissions are a result of the long-time impact of various causal factors, I include lagged regressors in the estimation. However, many problems could arise from using too many lags, such as the loss of degrees of freedom, multicollinearity, error term serial correlation, and misspecification errors. Here, I consider one lag for the analysis which presents the least number of problems in empirical estimation. The resulting specification takes the form of an ARDL model:

$$\begin{aligned} GHG_{it} = {} & \delta_i GHG_{i,t-1} + \varphi_{10i} URBAN_{it} + \varphi_{11i} URBAN_{i,t-1} + \varphi_{20i} URBAN2_{it} + \\ & \varphi_{21i} URBAN2_{i,t-1} + \varphi_{30i} RENERGY_{it} + \varphi_{31i} RENERGY_{i,t-1} + \varphi_{40i} URBAN * \\ & RENERGY_{it} + \varphi_{41i} URBAN * RENERGY_{i,t-1} + \varphi_{50i} GDP_{it} + \varphi_{51i} GDP_{i,t-1} + \\ & \varphi_{60i} GDP2_{it} + \varphi_{61i} GDP2_{i,t-1} + \varphi_{70i} FDI_{it} + \varphi_{71i} FDI_{i,t-1} + \varphi_{80i} TRADE_{it} + \\ & \varphi_{81i} TRADE_{i,t-1} + \varphi_{90i} INDUSTRY_{it} + \varphi_{91i} INDUSTRY_{i,t-1} + \mu_i + \vartheta_{it}, \\ & \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (4.3) \end{aligned}$$

Equation (4.3) can be further rewritten into the error correction form to compute the long- and short-run relationship between carbon intensity and urbanization:

$$\Delta GHG_{it} = \gamma_i\big(GHG_{i,t-1} - \theta_{0i} - \theta_{1i}URBAN_{it} - \theta_{2i}URBAN2_{it} - \theta_{3i}RENERGY_{it} - \theta_{4i}URBAN * RENERGY_{it} - \theta_{5i}GDP_{it} - \theta_{6i}GDP2_{it} - \theta_{7i}FDI_{it} - \theta_{8i}TRADE_{it} - \theta_{9i}INDUSTRY_{it}\big) + \varphi_{11i}\Delta URBAN_{it} + \varphi_{21i}\Delta URBAN2_{it} + \varphi_{31}\Delta RENERGY_{it} + \varphi_{41}\Delta URBAN * RENERGY_{it} + \varphi_{51i}\Delta GDP_{it} + \varphi_{61i}\Delta GDP2_{it} + \varphi_{71i}\Delta FDI_{it} + \varphi_{81i}\Delta TRADE_{it} + \varphi_{91i}\Delta INDUSTRY_{it} + \tau_{it}, \ldots \quad (4.4)$$

where, γ_i= - (1-δ_i), $\theta_{0i}=\frac{\mu_i}{1-\delta_i}$, $\theta_{1i} = \frac{\varphi_{10i}+\varphi_{11i}}{1-\delta_i}$, $\theta_{2i} = \frac{\varphi_{20i}+\varphi_{21i}}{1-\delta_i}$, $\theta_{3i} = \frac{\varphi_{30i}+\varphi_{31i}}{1-\delta_i}$, $\theta_{4i} = \frac{\varphi_{40i}+\varphi_{41i}}{1-\delta_i}$, $\theta_{5i} = \frac{\varphi_{50i}+\varphi_{51i}}{1-\delta_i}$, $\theta_{6i} = \frac{\varphi_{60i}+\varphi_{61i}}{1-\delta_i}$, $\theta_{7i} = \frac{\varphi_{70i}+\varphi_{71i}}{1-\delta_i}$, $\theta_{8i} = \frac{\varphi_{80i}+\varphi_{81i}}{1-\delta_i}$, $\theta_{9i} = \frac{\varphi_{90i}+\varphi_{91i}}{1-\delta_i}$. Here, θ represents the long-run coefficients, φ represents the short-run coefficients, and the error correction term γ represents the speed of adjustment of CO_2 intensity to the long-run equilibrium following a shock that deviates the relationship from the equilibrium. Thus, error correction term γ depends on the coefficient of lagged dependent variable, long-run coefficients θ depends on corresponding short-run coefficients and the coefficient of lagged dependent variable. The PMG approach estimates first panel-specific slope coefficients for each country using the pooled ordinary least square (OLS) approach, which is subsequently averaged to obtain the mean group estimates.

4.5 Estimation Results

I first test for panel stationarity using the Levin-Lin-Chu and the Hadri LM tests, (Hadri, 2000; Levin et al., 2002). The Levin-Lin-Chu test assumes that panels contain unit roots across cross-sections, while the Hadri LM test assumes that all panels are stationary against some panels containing unit roots. Table 4.2 shows that some series are stationary, and some contain a unit root. Mixed conclusions are reached for the two tests. One advantage of the PMG estimator is that it allows for both stationary and nonstationary variables. Therefore, I proceed to the analysis using the variables as specified in the data section.

Table 4.2. Panel Unit Root Tests

Levin, Lin and Chu tests				**Hardri LM tests**			
Series	Adjusted t	Series	Adjusted t	Series	Z	Series	Z
CO_2intensity	-3.91***	*ΔCO_2intensity*	-17.52***	*CO_2intensity*	84.53***	*ΔCO_2intensity*	5.13***
CO_2	-8.33***	*ΔCO_2*	-12.58***	*CO_2*	80.96***	*ΔCO_2*	5.69***
URBAN	-4.88***	*ΔURBAN*	-37.73***	*URBAN*	120.93***	*ΔURBAN*	71.94***
URBAN2	-3.82***	*ΔURBAN2*	-31.49***	*URBAN2*	118.89***	*ΔURBAN2*	77.79***
RENERGY	-1.67*	*ΔRENERGY*	-14.90***	*RENERGY*	79.63***	*ΔRENERGY*	0.71
GDP	-1.02	*ΔGDP*	-14.02***	*GDP*	51.33***	*ΔGDP*	7.90***
GDP2	1.10	*ΔGDP2*	-14.94***	*GDP2*	47.71***	*ΔGDP2*	5.31***
FDI	-4.34***	*ΔFDI*	-15.70***	*FDI*	19.13***	*ΔFDI*	-3.74
TRADE	-3.05***	*ΔTRADE*	-18.80***	*TRADE*	65.36***	*ΔTRADE*	-2.48
INDUSTRY	-2.30***	*ΔINDUSTRY*	-16.90***	*INDUSTRY*	87.32***	*ΔINDUSTRY*	-1.42

Note: For Levin, Lin and Chu tests, one lag is used in the ADF regression. The null hypothesis is that the panel under consideration contains a unit root, while the alternative hypothesis states the panel is stationary. Panels means and time trends are included. For the Hardri LM tests, the time trend is included. The null hypothesis is that the panel under consideration is stationary, while under the alternative hypothesis, the panel contains a unit root. One, two, and three asterisks represent statistical significance at 10%, 5%, and 1%, respectively.

Panel cointegration is tested using the Kao test to determine whether a long-run relationship exists between dependent variable and its regressors (Kao, 1999). Table 4.3 provides Kao test statistic results using CO_2 intensity as the dependent variable, and table 4.4 gives testing results of this statistic using per capita CO_2 emissions as the dependent variable. In both tables, the null hypothesis of this test is that there exists no cointegration relationship, while under the alternative hypothesis, all panels are cointegrated. As can be seen in the table 4.3, the Dickey-Fuller t-statistic in the Kao test is -7.00, which is significant at 1%. Table 4.4 also provides Dickey-Fuller t-statistic at the same level of significance. Therefore, the Kao test suggests that the series are cointegrated. Further, the Augmented Dicky-Fuller (ADF) t-statistic is also highly significant, again suggesting a long-run relationship between the series.

Table 4.3. Panel Cointegration Test

Kao Test		
Ho: No cointegration Ha: All panels are cointegrated		
	Statistic	p-value
Modified Dickey-Fuller t	-0.88	0.1878
Dickey-Fuller t	-7.00	0.0000
Augmented Dickey-Fuller t	-7.43	0.0000
Unadjusted modified Dickey-Fuller t	-1.70	0.0450
Unadjusted Dickey-Fuller t	-7.48	0.0000
Cointegrating vector: Same Panel means: Included	Time trend: Not included A.R. parameter: Same	

Note: Here, CO_2 intensity is the dependent variable.

Table 4.4. Panel Cointegration Test

Kao Test		
Ho: No cointegration Ha: All panels are cointegrated		
	Statistic	p-value
Modified Dickey-Fuller t	-0.6572	0.2555
Dickey-Fuller t	-2.5873	0.0048
Augmented Dickey-Fuller t	-2.6196	0.0044
Unadjusted modified Dickey-Fuller t	-3.0497	0.0011
Unadjusted Dickey-Fuller t	-5.1542	0.0000
Cointegrating vector: Same Panel means: Included	Time trend: Not included A.R. parameter: Same	

Note: Here, CO_2 is the dependent variable.

Given the presence of a long-run relationship, I now proceed to estimate the relationship using the PMG estimator. Estimation results are presented in Table 4.5. As can be seen in model 1, all long-run coefficients are consistently statistically significant except for interaction term of urbanization and renewable energy, and foreign direct investment, suggesting that urbanization, renewable energy consumption, economic growth, and industry composition all impact CO_2

intensity in the long term. Also, the positive sign of urbanization and the negative sign of squared urbanization supports the inverted U-shaped relationship noted in Hypothesis 1. In other words, CO_2 intensity initially increase at a decreasing rate with urbanization until reaching the peak emission level, after which CO_2 intensity decreases with urbanization. This result is consistent with previous studies such as Zhu et al. (2012) and Shahbaz et al. (2016). On the contrary, Model 2 does not support hypothesis 1 which indicates that per capita CO_2 emissions initially decreases as more people lives in urban areas, however, per capita CO_2 emissions increases as urban areas grows further beyond the minimum emission level. One possible explanation behind the difference between models is that population grows disproportionately than the economic growth at the given level of CO_2 emissions.

Table 4.5. Pooled Mean Group (PMG) Estimates (1990 - 2018)

	Dependent Variable	
Independent Variables	***Model 1 (CO_2 intensity)***	***Model 2 (CO_2)***
Long Run:		
URBAN	.0228723*** (.0036887)	-.024771*** (.007282)
URBAN²	-.0001933*** (.000027)	.0001254*** (.0000492)
RENERGY	-.0009548* (.0005455)	-.0186426*** (.0019854)
*URBAN*RENERGY*	.00000871 (.00000947)	.0002161*** (.000054)
GDP	-.0047433*** (.0007334)	-.0392729*** (.0094004)
GDP²	.0000550*** (.0000104)	.0040182*** (.0004853)
FDI	-.0000000499 (.0000000460)	.00000792*** (.00000114)
TRADE	-.0001827** (.0000868)	-.000121 (.0001097)
INDUSTRY	.0021054*** (.0003454)	-.0014017*** (.0003511)
Short Run:		
ECT	.2710863*** (.0275671)	.189295*** (.0241367)
URBAN	-.4197192 (.8927195)	-4.557598 (8.306878)
URBAN²	-.0012388 (.0075403)	-.0059886 (.0529963)
RENERGY	1.171687 (1.292493)	.7282914 (3.085353)
*URBAN*RENERGY*	-.0148018 (.0161191)	-.0194772 (.0410447)
GDP	-.0150162 (.0129915)	.1149479** (.0394451)
GDP²	.0024431 (.0106957)	-.0195317** (.0087133)
FDI	-.0000139 (.0000122)	.0000184 (.0000156)
TRADE	.0000586 (.0001187)	.0059587*** (.0012663)
INDUSTRY	-.0000207 (.0004732)	.0229923* (.0128885)
Constant	.0610591*** (.0165196)	-.6672085*** (.1044824)
Number of observations	2940	2940

Note: Standard errors in parentheses. One, two, and three asterisks represent statistical significance at 10%, 5%, and 1%, respectively.

Based on equation (4.2), the turning point of urbanization, in the long run, is estimated by using average coefficient values of URBAN and $URBAN^2$ from model 1 (Table 4.5). The turning point is computed to 59.16 percent. Countries with urbanization levels below 59.16 percent will show increases in in CO_2 intensity as more people migrate into cities. Meanwhile, countries with greater than 59.16 percent of urbanization are projected to lower CO_2 intensity even as urban areas expand in population. This turning point is consistent with the results of previous studies. For instance, Du and Xia (2018) find that urbanization increases CO_2 emissions when the urbanization rate is below 48 percent, while Zhang et al. (2017) show that the turning point is 73 percent. From Table 4.5, PMG estimates thus shows inverted U-shaped relationship between urbanization and CO_2 intensity.

Table 1A in the appendix shows the list of countries with level of urbanization in 2018, where 44 countries considered in the present study had urbanization below this threshold, and 61 countries had more than 59.16 percent of urbanization. Most of the developed economies (e.g. Belgium, Iceland, Uruguay, Malta, and Singapore) fall within the latter category, i.e. their urbanization exceeds this threshold. On the other hand, countries below the threshold are mostly located in Africa or Asia, suggesting that further efforts be devoted in these regions to mitigate the positive impact of urbanization on CO_2 intensity. The turning point identified in this research suggests modeling efforts that expect CO_2 intensity to decrease simply with a higher urbanization rate without considering other factors may be mis specified. Model 1 further show that the share of renewable energy in the total energy portfolio significantly lowers CO_2 intensity which supports the findings of Baek (2016), Apergis et al. (2010), and Kim (2020).

From the estimation results in models 1 and 2, an interaction term between urbanization and renewable energy consumption is included to show how the impact of the former is mitigated by the latter (Hypothesis 2). All else equal, the interaction coefficient evident that additional urbanization does not significantly reduce CO_2 intensity but does significantly increase per capita CO_2 emissions at the given level of renewable energy share. Thus, these results do not provide significant evidence for Hypothesis 2.

Based on long-run coefficients of GDP and GDP^2, models 1 and 2 show U-shaped relation between GDP and GHG emissions, i.e., past the turning point, countries with lower GDP emits less CO_2 per dollar of GDP and less per capita CO_2 emissions while countries with higher GDP emits more CO_2 per dollar of GDP and more per capita CO_2 emissions. From Table 4.5, I computed GDP turning point (43.12 thousand dollars) which is higher than the average GDP per capita. Based on this turning point, countries with annual per capita GDP below 43.12 thousand dollars will continue to increase CO_2 intensity and per capita CO_2 emissions as higher growth leads to lower CO_2 intensity and per capita CO_2 emissions in those countries. On other hand, countries with greater than 43.12 thousand dollars of per capita GDP are projected to increase CO_2 intensity and per capita CO_2 emissions as GDP increases. Thus, this study does not support EKC hypothesis which is consistent with Begum et al. (2015), Al-Mulali et al. (2016), Dogan and Turkekul (2016), Jebli and Youssef (2015), Liu et al. (2017) and Pao et al. (2011).

The short-run coefficients in model 1 are non-significant except for the adjustment coefficient. Models 1 and 2 also suggest that neither urbanization nor renewable energy use impacts CO_2 emissions or intensity in the short run. Edenhofer et al. (2014) note that since urbanized areas are already set up with energy, transport, and other infrastructure, a change in CO_2 intensity due to increased urban population may be negligible in the short run. Further, even with an increasing share of renewable energy consumption, the impact may be too small to significantly affect CO_2 intensity in the short run. The adjustment coefficients known as error correction term

(ECT) in both models suggest that when deviation from the long-run relationship occurs, between 20 and 30% of this disequilibrium will be adjusted in the first year. Finally, both models estimate ECT based on coefficients of lagged dependent variables which is why 105 observations are lost in 1990.

4.6 Conclusion and Policy Implications

Global climate change continues to challenge the international community. This study aims to investigate whether the urbanization process influences GHG emissions as represented by CO_2 emissions per capita and carbon intensity. An accurate understanding of the urbanization-CO_2 relationships is essential for forecasting models aiming to project future CO_2 emissions and policy initiatives to mitigate global warming. Using a panel of 105 countries from 1990 to 2018, the PMG estimation technique is applied to estimate the relationship between urbanization and CO_2 emissions per capita and intensity. Results show that impacts differ between long and short runs. In the long run, urbanization leads to more CO_2 intensity in the early stages of urbanization, while in later stages, CO_2 intensity reduce. Thus, the study supports the inverted U-shaped urban pollution hypothesis, consistent with the findings of Martinez-Zarzoso and Maruotti (2011), Zhang et al. (2017), and Chen et al. (2019). However, consistent with Kim (2020), I fail to identify a statistically significant relationship in the short run. Our results show that the urban pollution hypothesis is perhaps a long-term issue.

The estimated turning point is at 59.16 percent of urbanization based on sample countries. In other words, when the urbanization rate exceeds 59.16 percent, additional urbanization in the country decreases the CO_2 intensity. The turning point appears to be consistent with those identified in previous studies. Since renewable energy consumption significantly reduce CO_2 intensity in the long run, increasing renewables in a country's energy portfolio, along with other sustainable city initiatives, can reduce CO_2 emissions. Those sustainable initiatives are more important for countries where cities are growing very fast. By introducing an interaction term into the model, I find that higher renewable energy use does not significantly influence the marginal impact of urbanization on CO_2 intensity. Thus, more research efforts are needed to reinvestigate the interaction effect of urbanization and renewable energy consumption on CO_2 intensity.

Over sixty percent of carbon emissions within cities are released from buildings (JLL Research, May 2022). Each city should draw building codes, reporting and disclosure frameworks, and energy audits, minimum building standards, incentives, and accelerators to reduce GHG emissions (World Economic Forum, 2022). Based on this research findings, cities should enhance renewable energy portfolio in their building structures as part of best management practice. Secondly, urban transport significantly contributes to GHG emissions. Cities in low-and-middle income countries are facing challenges to low carbon and efficient urban transport to minimize GHG emissions (Bianchi Alves, et al. 2023). Findings in this study indicate that cities in developing countries should prioritize investment on low-carbon, efficient and inclusive transport services. In this regard, developed countries along with local governments should cooperate with those developing countries to take necessary policy actions, providing technical and financial assistance to attain global long-term net zero carbon (NZC) goal. The global facility to decarbonize transport (GFDT) is an example of a global initiative funded by developed countries to help developing countries achieving carbon-neutral transportation by 2050 (World Bank, n.d).

References

ADB - Asian Development Bank. (2021). Japan to strengthen cooperation on clean energy in ASEAN region. Available at: https://www.adb.org/news/adb-japan-strengthen-cooperation-clean-energy-asean-region. (Accessed on November 3, 2021).

Ali, H. S., Abdul-Rahim, A. S., and Ribadu, M. B. (2017). Urbanization and carbon dioxide emissions in Singapore: evidence from the ARDL approach. *Environmental Science and Pollution Research*, 24(2), 1967-1974.

Ali, R., Khuda, B., and Yasin, M. A. (2019). Impact of urbanization on CO_2 emissions in emerging economy: Evidence from Pakistan. *Sustainable Cities and Society*, *48*: 101553.

Al-Mulali, U., Ozturk, I., Solarin, S.A., (2016). Investigating the environmental Kuznets curve hypothesis in seven regions: the role of renewable energy. *Ecol. Indic. 67*, 267–282.

Apergis, N., and Ozturk, I. (2015). Testing environmental Kuznets curve hypothesis in Asian countries. *Ecological Indicators,* 52, 16-22.

Apergis, N., Payne, J. E., Menyah, K., and Wolde-Rufael, Y. (2010). On the causal dynamics between emissions, nuclear energy, renewable energy, and economic growth. *Ecological Economics*, 69(11), 2255–2260.

Atasoy, B.S. (2017). Testing the environmental Kuznets curve hypothesis across the U.S.: Evidence from panel mean group estimators. *Renewable and Sustainable Energy Reviews*, 77: 731-747.

Baek, J. (2016). Do nuclear and renewable energy improve the environment? empirical evidence from the United States. *Ecological Indicators*, 66, 352–356.

Barnes, D.F. and Floor, W. (1999). Biomass energy and the poor in the development world. *Journal of International Affairs, 53* (1), 237-259.

Begum, R.A., Sohag, K., Abdullah, S.M.S., Jaafar, M., (2015). CO_2 emissions, energy consumption, economic and population growth in Malaysia. *Renew. Sust. Energ. Rev. 41*, 594–601.

Bianchi Alves, B., Bou Mjahed, L., and Moody, J. (2023). *Decarbonizing Urban Transport for Development* (Mobility and Transport Connectivity Series). Washington, DC: World Bank. Accessed on 29 October 2023: http://hdl.handle.net/10986/40373

Brueckner, J.K. (2011). *Lectures on Urban Economics*. Cambridge, MA: The MIT Press.

Chen, S., Jin, H., and Lu, Y. (2019). Impact of urbanization on CO_2 emissions and energy consumption structure: A panel data analysis for Chinese prefecture-level cities. *Structural Change and Economic Dynamics, 49*: 107-119.

Cherni, A., and Jouini, S. E. (2017). An ARDL approach to the CO_2 emissions, renewable energy and economic growth nexus: Tunisian evidence. *International Journal of Hydrogen Energy*, 42(48), 29056-29066.

Chikaraishi, M., Fujiwara, A., Kaneko, S., Poumanyvong, P., Komatsu, S., and Kaluain, A. (2015). The moderating effects of urbanization on carbon dioxide emissions: A latent class modeling approach. *Technological. Forecasting and Social Change, 90* (Part A): 302-317.

Crippa, M., Guizzardi, D., Pisoni, E., Solazzo, E., Guion, A., Muntean, M., ... and Hutfilter, A. F. (2021). Global anthropogenic emissions in urban areas: patterns, trends, and challenges. *Environmental Research Letters*, 16(7), 074033.

Dogan, E., Turkekul, B., (2016). CO_2 emissions, real output, energy consumption, trade, urbanization and financial development: testing the EKC hypothesis for the USA. *Environ. Sci. Pollut. Res. 23* (2), 1203–1213.

Du, W. C., and Xia, X. H. (2018). How does urbanization affect GHG emissions? A cross-country panel threshold data analysis. *Applied energy*, 229, 872-883.

Edenhofer O., et al. (2014). Technical Summary. In: Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Edenhofer, O., et al. (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

EIA – U.S. Energy Information Association. (2022). U.S. energy facts explained. Available at: https://www.eia.gov/energyexplained/us-energy-facts/

Fan, Y., Liu, L. C., Wu, G., and Wei, Y. M. (2006). Analyzing impact factors of CO_2 emissions using the STIRPAT model. *Environmental Impact Assessment Review*, 26(4), 377-395.

Gurney, K. R., Kılkış, Ş., Seto, K. C., Lwasa, S., Moran, D., Riahi, K., ... and Luqman, M. (2022). Greenhouse gas emissions from global cities under SSP/RCP scenarios, 1990 to 2100. *Global Environmental Change*, 73, 102478.

Hadri, K. (2000). Testing for stationarity in heterogenous panel data. *The Econometrics Journal*, 3(2): 148-161.

Hanif, I. (2018). Impact of economic growth, nonrenewable and renewable energy consumption, and urbanization on carbon emissions in Sub-Saharan Africa. *Environmental Science and Pollution Research, 25* (15): 15057–15067.

IEA (2021). Greenhouse Gas Emissions from Energy: Overview. Paris: International Energy Agency. Accessed on October 16, 2021 from https://www.iea.org/reports/greenhouse-gas-emissions-from-energy-overview

Iwata, H., Okada, K., and Samreth, S. (2011). A note on the environmental Kuznets curve for CO_2: A pooled mean group approach. *Applied Energy*, 88(5): 1986-1996.

Jebli, M.B., and Youssef, S.B. (2015). The environmental Kuznets curve, economic growth, renewable and non-renewable energy, and trade in Tunisia. Renew. *Sust. Energ. Rev. 47*, 173–185.

JLL Research, (May 2022). *Responding to the climate emergency*. Access on 29 October 2023: https://www.us.jll.com/en/trends-and-insights/research/decarbonizing-cities-and-real-estate

Kao, C. (1999). Spurious regression and residual-based tests for cointegration in panel data. *J. Econometrics*, 90(1): 1-44.

Kaika, D. and Zervas, E. (2013). The Environmental Kuznets Curve (EKC) theory- Part A: Concept, causes, and the CO_2 emissions case. *Energy Policy*, 62: 1392-1402.

Koçak, E., and Şarkgüneşi, A. (2018). The impact of foreign direct investment on CO_2 emissions in Turkey: new evidence from cointegration and bootstrap causality analysis. *Environmental Science and Pollution Research*, 25(1), 790-804.

Kim, S. (2020). The effects of foreign direct investment, economic growth, industrial structure, renewable and nuclear energy, and urbanization on Korean greenhouse gas emissions. *Sustainability*, *12*, 1625.

Kuznets, S. (1955). Economic growth and income inequality. *American Economic Review, 45,* 1-28.

Lee, J. W. (2013). The contribution of foreign direct investment to clean energy use, carbon emissions and economic growth. *Energy policy*, 55, 483-489.

Levin, A. et. al. (2002). Unit root tests in panel data: Asymptotic and finite-sample properties. *Journal of Econometrics, 108* (1): 1-24.

Liddle, B., and Lung, S. (2010). Age-structure, urbanization, and climate change in developed countries: revisiting STIRPAT for disaggregated population and consumption-related environmental impacts. *Population and Environment*, 31(5), 317-343.

Lidula, N. W. A., Mithulananthan, N. Ongsakul, W. Widjaya, C. and Henson, R. (2007). ASEAN towards clean and sustainable energy: potentials, utilization and barriers. *Renewable Energy*, *32* (9): 1441–1452. https://doi.org/10.1016/j.renene.2006.07.007

Lin, B., and Zhu, J. (2019a). Impact of energy saving and emission reduction policy on urban sustainable development: empirical evidence from China. *Applied Energy*, 239, 12–22.

Lin, B., and Zhu, J. (2019b). Determinants of renewable energy technological innovation in China under CO_2 emissions constraint. *Journal of Environmental Management*, 247, 662-671.

Liu, X., Zhang, S., Bae, J., (2017). The impact of renewable energy and agriculture on carbon dioxide emissions: investigating the environmental Kuznets curve in four selected ASEAN countries. *J. Clean. Prod. 164*, 1239–1247.

Lv, Z., and Xu, T. (2019). Trade openness, urbanization and CO_2 emissions: dynamic panel data analysis of middle-income countries. *The Journal of International Trade & Economic Development*, 28(3), 317-330.

Mahmood, H., Alkhateeb, T. T. Y., and Furqan, M. (2020). Industrialization, urbanization and CO_2 emissions in Saudi Arabia: Asymmetry Analysis. *Energy Reports*, *6*: 1553-1560.

Martinez-Zarzoso I. and Bengochea-Morancho, A. (2004). Pooled mean group estimation of an environmental Kuznets curve for CO_2. *Economics Letters*, 82(1): 121-126.

Martinez-Zarzoso, I. and Maruotti, A. (2011). The impact of urbanization on CO_2 emissions: Evidence from developing countries. *Ecological Economics*, *70*(7): 1344-1353.

Meng, G., Guo, Z., and Li, J. (2021). The dynamic linkage among urbanization, industrialization, and carbon emissions in China: Insights from spatiotemporal effect. *Science of The Total Environment*, 760, 144042.

Mert, M. and Boluk, G. (2016). Do foreign direct investment and renewable energy consumption affect the CO_2 emissions? New evidence from a panel ARDL approach to Kyoto Annex countries. *Environmental Science and Pollution Research*, 23: 21669-21681.

Narayan, P. K., Saboori, B., and Soleymani, A. (2016). Economic growth and carbon emissions. *Economic Modelling*, 53, 388-397.

O'neill, B. C., Dalton, M., Fuchs, R., Jiang, L., Pachauri, S., and Zigova, K. (2010). Global demographic trends and future carbon emissions. *Proceedings of the National Academy of Sciences*, 107(41), 17521-17526.

Pao, H.T., Yu, H.C., Yang, and Y.H. (2011). Modeling the CO_2 emissions, energy use, and economic growth in Russia. *Energy 36* (8), 5094–5100.

Pesaran, M.H. and Shin, Y. (1998). Generalized impulse response analysis in linear multivariate models. *Economics Letters*, 58(1):17-29.

Pesaran, M.H, Shin, Y., and Smith, R.P. (1999). Pooled mean group estimation of dynamic heterogeneous panel. *Journal of the American Statistical Association*, 94:621–634.

Poumanyvong, P., and Kaneko, S. (2010). Does urbanization lead to less energy use and lower CO_2 emissions? A cross-country analysis. *Ecological Economics*, 70(2), 434-444.

Sadorsky, P. (2014). The effect of urbanization on CO_2 emissions in emerging economies. *Energy Economics, 41*: 147-153.

Salahuddin, M., Ali, M., Vink, N., and Gow, J. (2019). The effects of urbanization and globalization on CO_2 emissions: evidence from the Sub-Saharan Africa (SSA) countries. *Environmental Science and Pollution Research*, 26(3), 2699-2709.

Shah, S. A. R., Naqvi, S. A. A., and Anwar, S. (2020). Exploring the linkage among energy intensity, carbon emission and urbanization in Pakistan: Fresh evidence from ecological modernization and environment transition theories. *Env. Sc. and Poll. Res., 27*, 40907-40929.

Shahbaz, M., Loganathan, N., Muzaffar, A.T., Ahmed, K., and Ali, M. J. (2016). How urbanization affects CO_2 emissions in Malaysia? The application of STIRPAT model. *Renewable and Sustainable Energy Reviews, 57*: 83-93.

Shahbaz, M., Balsalobre-Lorente, D., & Sinha, A. (2019). Foreign direct investment–CO_2 emissions nexus in Middle East and North African countries: Importance of biomass energy consumption. *Journal of Cleaner Production*, 217, 603-614.

Shi and Li (2018). Research on three-stage dynamic relationship between carbon emission and urbanization rate in different city groups. *Ecological Indicators, 91*: 195-202.

Tang, C. F., and Tan, B. W. (2015). The impact of energy consumption, income, and foreign direct investment on carbon dioxide emissions in Vietnam. *Energy*, 79, 447-454.

UN (2015). World Urbanization Prospects: The 2014 Revision (Ref: ST/ESA/SER.A/366). Department of Economic and Social Affairs, Population Division. New York: United Nations. Accessed at: https://population.un.org/wup/publications/files/wup2014-report.pdf

World Bank (n.d). *The Global Facility to Decarbonize Transport: Changing the Way We Move*. Accessed on November 6, 2023: https://www.worldbank.org/en/programs/global-facility-to-decarbonize-transport.

World Economic Forum (2022). *Why we need global cooperation on decarbonizing cities and real estate*. Accessed on 29 October 2023: https://www.weforum.org/agenda/2022/05/global-cooperation-decarbonizing-cities-and-real-estate/

Xie, Q., Wang, X., and Cong, X. (2020). How does foreign direct investment affect CO_2 emissions in emerging countries? New findings from a nonlinear panel analysis. *Journal of Cleaner Production*, 249, 119422.

Yao, X., Kou, D., Shao, S., Li, X., Wang, W., and Zhang, C. (2018). Can urbanization process and carbon emission abatement be harmonious? New evidence from China. *Env. Impact Assessment Rev., 71*: 70-83.

Zhang, Y., et al. (2014). The impact of economic growth, industrial structure and urbanization on carbon emission intensity in China. *Natural Hazards, 73*:579–595.

Zhang, C., and Zhou, X. (2016). Does foreign direct investment lead to lower CO_2 emissions? Evidence from a regional analysis in China. *Renewable and Sustainable Energy Reviews*, 58, 943-951.

Zhang, N., Yu, K., and Chen, Z. (2017). How does urbanization affect carbon dioxide emissions? A cross-country panel data analysis. *Energy Policy, 107*: 678-687.

Zhang, S., Liu, X., and Bae, J. (2017). Does trade openness affect CO_2 emissions: evidence from ten newly industrialized countries? *Environmental Science and Pollution Research*, 24(21), 17616-17625.

Zhao, J., Jiang, Q., Dong, X., Dong, K., and Jiang, H. (2022). How does industrial structure adjustment reduce CO_2 emissions? Spatial and mediation effects analysis for China. *Energy Economics*, 105, 105704.

Zhou, X., Zhang, J., and Li, J. (2013). Industrial structural transformation and carbon dioxide emissions in China. *Energy Policy*, 57, 43-51.

Zhu, H., You, W., and Zeng, Z. (2012). Urbanization and CO_2 emissions: A semi-parametric panel data analysis. *Economics Letters, 117* (3): 848-850.

APPENDIX:

Table 4.1A: List of Countries and Level of Urbanization (%) in 2018.

Country	URBAN	Country	URBAN	Country	URBAN	Country	URBAN	Country	URBAN
Burundi	13.03	Mali	42.36	Georgia	58.63	Algeria	72.63	Saudi Arabia	83.84
Niger	16.43	Egypt	42.70	China	59.15	Switzerland	73.80	Oman	84.54
Rwanda	17.21	Guinea-Bissau	43.36	Albania	60.32	Iran	74.90	Finland	85.38
Sri Lanka	18.48	Philippines	46.91	Gambia	61.27	Bulgaria	75.01	Australia	86.01
Nepal	19.74	Senegal	47.19	Paraguay	61.59	Malaysia	76.04	New Zealand	86.54
Chad	23.06	Benin	47.31	Morocco	62.45	Cuba	77.04	Brazil	86.57
Uganda	23.77	Thailand	49.95	Armenia	63.15	Germany	77.31	Sweden	87.43
Eswatini	23.80	Namibia	50.03	Ireland	63.17	Peru	77.91	Chile	87.56
Kenya	27.03	Nigeria	50.34	Ecuador	63.82	Greece	79.06	Denmark	87.87
Burkina Faso	29.36	Cote d'Ivoire	50.78	Portugal	65.21	Costa Rica	79.34	Lebanon	88.59
Zimbabwe	32.21	Guatemala	51.05	South Africa	66.36	Libya	80.10	Gabon	89.37
Tanzania	33.78	Mauritania	53.67	Cyprus	66.81	Mexico	80.16	Jordan	90.98
India	34.03	Romania	54.00	Panama	67.71	Spain	80.32	Luxembourg	90.98
Mozambique	35.99	Haiti	55.28	Mongolia	68.45	France	80.44	Netherlands	91.49
Guinea	36.14	Indonesia	55.33	Tunisia	68.95	Colombia	80.78	Japan	91.62
Bangladesh	36.63	Jamaica	55.67	Ukraine	69.35	Comoros	80.78	Argentina	91.87
Pakistan	36.67	Azerbaijan	55.68	Bolivia	69.43	Dominican Republic	81.07	Iceland	93.81
Madagascar	37.19	Ghana	56.06	Botswana	69.45	Canada	81.41	Malta	94.61
Mauritius	40.79	Cameroon	56.37	Iraq	70.47	Norway	82.25	Uruguay	95.33
Togo	41.70	Austria	58.30	Hungary	71.35	United States	82.26	Belgium	98.00
Sierra Leone	42.06	Nicaragua	58.52	El Salvador	72.02	United Kingdom	83.40	Singapore	100.00

CHAPTER 5:
SUMMARY AND CONCLUSIONS

5.1 Study Summary

Spatiotemporal models play an important role in economic research. Spatial econometrics captures the influence of geographic locations of observations (see Partridge et al. (2012) for a general discussion of the importance of spatial econometrics). In presence of spatial autocorrelation, I apply spatial econometric models in the first and second essays and demonstrate that the conventional non-spatial model (such as OLS) provides biased estimate. Since the sample countries in essay three is not contiguous, and pollution is a long-term problem, I apply dynamic panel data econometric models (such as DFE, PMG) to estimate short-run and long-run values of coefficients.

In the first essay, I found that air quality standard non-compliance impacts lung and bronchus cancer incidence regionally while water quality standards non-compliance impacts the cancer incidence locally. Based on state-level calculations, a 10% reduction in population exposure to air and water quality violations saves six people in Utah plus three-nearest neighboring states, and five people in Oklahoma from cancer incidence annually, respectively. The corresponding annual monetary benefits from improved compliance are $20.6 and $24.7 million, respectively for water and air.

Second essay investigates the spatial spillover effects of NPIs policies on the reduction of COVID-19 cases in the contiguous U.S. counties. Based upon annual cross-sectional data, I found statistically significant spillover effects of stay-at-home orders and mask mandates on reducing COVID-19 cases in the contiguous U.S. No significant spillover effects were found for restrictions on restaurants and bars. Using annual cross-sectional data, on an average, 4 cases and 9 cases per 100,000 people can be reduced from COVID-19 infections in the mandate county and six-nearest neighboring counties, respectively by implementing one additional month of mask mandate policies. By implementing mandatory stay-at-home order for one additional month, on an average, 15 and 38 fewer cases per 100,000 people can be reduced in mandate county, and three-nearest neighboring counties, respectively. From the annual cross-sectional model, I found $0.70 million, $2.95 million, and $3.65 million of monetary benefits from reduced COVID-19 cases in three neighboring counties of Wyoming (non-mandate state) by implementing a mask mandate, mandatory stay-at-home order, and both these policies, respectively in Powder River County in the state of Montana (mandate state).

In the third essay, I found a significant nonlinear relationship between urbanization and GHG emissions. That is, a higher urbanization rate increases CO_2 intensity when the urbanization rate is below 59.16%, after which the marginal effect of urbanization turns negative. From the interaction term coefficient, this study finds that renewable energy use does not significantly mitigates the impact of urbanization on CO_2 intensity.

Based on the results, I will explain the policy implication for each essay in the following section.

5.2 Implications

The findings from the essays one and two have some policy implications for state and local governments while essay three has implications for regional and international communities. Environmental injustice issues arise from the essay one because non-white communities suffer disproportionately from than white counterpart given that water quality violations occur more often among the communities of color (Switzer and Teodoro, 2017). Racial disparity also exists in case of air quality violations, thus environmental justice across race and class to CAA is equally important for better compliance.

From the results in second essay, it is evident that the spillover effects of stay-at-home orders and mask mandates are larger than the effects within the mandate county. Also, the monetary benefit from this spillover effects is sufficiently high. Thus, individuals (from the patient perspective) as well as governments (from the societal perspectives) can save resources by reducing infections. Since stay-at-home orders and mask mandate show statistically significant impacts on the COVID-19 spread, and restrictions on restaurants and bars show no significant impacts, a cost-efficient intervention can be designed to tackle future pandemics based on these results.

Based upon dynamic panel data models, third essay claims that the urban pollution as a long-term issue. Thus, countries in pre-urbanization period should increase the share of renewable energy consumption to lower the marginal effects of urbanization on GHG emissions. In this regard, developed countries should cooperate developing countries to decarbonize cities.

5.3 Limitations

I acknowledge several limitations in my research. In the essay one, I employ a three-nearest neighbors weight matrix to estimate the effect of drinking water and air quality violations on the lung and bronchus cancer incidence. By utilizing a consistent and heterogenous spatial weight matrix based upon annual prevailing wind direction and hydrological pattern, future researchers may predict the effects of air and water quality violations on the cancer incidence. Also, I could not capture the influence of family genetic background on the incidence of lung and bronchus cancers.

In the second essay, I used only reported COVID-19 cases as a dependent variable. By incorporating unreported cases, future researchers could estimate spillover effects of NPIs on

COVID-19 spread. Future researchers can also include school closure data into the model and apply STADL model to address temporal lag of NPIs on the COVID-19 spread.

In the third essay, future studies may wish to extend the analyses in several ways. First, I use the share of the population living in urban areas as a proxy for urbanization. Future research may use alternative indicators, such as population density in urban areas (people living per square mile), to examine whether the relationship between urbanization and GHG emissions is robust to the use of different measurements. Second, although I include various control variables in the analysis, such as FDI, trade openness, and sector composition, they are by no means inclusive of all possible variables that affect carbon intensity. Depending on the data availability, future studies should consider more right-hand-side variables. However, estimation issues may occur with the inclusion of more control variables, such as reverse causality and multicollinearity. Future studies may also wish to consider alternative estimation strategies that can account for these issues.