The Great Migration, Urban Spatial Structure, and Their Economic and Environmental Impacts in the U.S.

by

Zixing Shen

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Sustainability Management

Waterloo, Ontario, Canada, 2025

© Zixing Shen 2025

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner:	Jeff Chan Professor, Lazaridis School of Business and Economics, Wilfrid Laurier University
Supervisor(s):	Juan Moreno-Cruz Professor, School of Environment, Enterprise and Development, University of Waterloo
Internal Member:	Cameron Mccordic Professor, School of Environment, Enterprise and Development, University of Waterloo
Internal-External Member:	Jeremy Pittman Professor, School of Planning, University of Waterloo
Other Member(s):	Jevan Cherniwchan Professor, Dept. of Economics, McMaster University
	Xi Mao Professor, Dept. of Economics, University of Texas Rio Grande Valley

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

This thesis investigates the transformative impact of the Great Migration on urban spatial structure in the United States and its subsequent economic and environmental consequences. The Great Migration (1910-1970), marked by significant African American migration from the South to Northern and Western cities, reshaped the demographic and spatial landscape of receiving cities. Yet, the long-term spatial dynamics and their implications remain underexplored, a gap this study aims to address.

Using a novel dataset constructed through the City Clustering Algorithm (CCA), this research redefines historical urban boundaries, creating "synthetic cities" (Syncities) that more accurately reflect urban development from 1900 to 1970. The analysis reveals how demographic shifts influenced the size and shape of cities, with implications for their economic performance and environmental quality. By employing instrumental variable regression and mediation analysis, this thesis identifies the causal pathways through which these spatial transformations affected income levels and air quality, both immediately following the migration and decades later.

The findings highlight both economic growth and environmental challenges linked to urban expansion. Cities with larger and more dispersed urban forms benefited economically in the short term but faced greater environmental degradation over time. These results underscore the importance of urban spatial structure in shaping sustainable development trajectories.

By bridging historical demographic changes with contemporary urban outcomes, this study offers valuable insights for urban planning and policy, demonstrating how historical conditions continue to shape modern challenges. This work also provides a methodological foundation for future research on the interplay between migration, urban form, and sustainability.

This thesis is organized into seven chapters. Chapter 1 introduces the research goals and questions, explaining the importance of the study. Chapter 2 describes how the new urban spatial dataset was created and provides an overview of the key data. Chapter 3 looks at how urban areas in the U.S. changed between 1900 and 1970, providing background for the main analyses. Chapter 4 studies how the Great Migration affected the size and shape of cities, while Chapter 5 examines the economic and environmental effects of these changes. Chapter 6 explains how changes in urban areas connected the Great Migration to economic and environmental outcomes. Finally, Chapter 7 brings the findings together, discusses what they mean for policy, highlights limitations, and suggests ideas for future research. These chapters work together to show how the Great Migration reshaped American cities and what that means for sustainability today.

Acknowledgements

I want to thank everyone who made this thesis possible.

I want first to thank my supervisor, Professor Juan Moreno-Cruz. I want to thank him for his patience as well as for clear and valuable guidance and suggestions, which helped me solve research issues, improve my research skills, and help me become a better person. I am also very grateful to him for his understanding and support for my personal challenges due to a family emergency in 2023. Moreover, I want to say thank you to my committee members, Professor Cameron Mccordic, Professor Jeremy Pittman, Professor Jevan Cherniwchan, and Professor Xi Mao, as well as the external examiner, Professor Jeff Chan, for their valuable suggestions and feedback that made my research more meaningful. Also, I want to thank all the professors for my previous classes. Their classes provide the foundation for me to work on the research. Here, I want to say thank you to my master's supervisor, Professor Joel Wood, for his advice and help during this five-year research journey.

I would also like to thank the school's administrative support team, especially Anastasiya Saparaliyeva, Dragana Kostic, Jennifer Nicholson, Jennifer Doucet, Delores Lanni, and Hayley Cuthbertson. Their help on administrative matters allowed me to focus more on my research. In addition, I want to thank the school's counseling services team, especially Lihua Yang, who supported me from an emotional perspective. As an international student, I want to thank immigration consultant Louann Nhan for her help with my immigration-related applications. This ensures that I can study and do research without additional worries about my status in Canada. I also want to thank the advisors in the school's writing center who helped me with my writing revision and presentation mock, improving my English writing and speaking skills.

During the research journey, I also received support from many friends. First of all, I am grateful to my fellow group members in our lab for their care and valuable chat. I also want to thank Rundong Gan, Xi Mao, Geisha Shaina Lyn Sanchez, Naomi Sunu, Jacqueline Ouellette, Usmita Afrose, Asher, Pankaj Tewari, Kazuhito Mizutani, Julia Chow, Guofei Long, Elham, and Lida Xu for their super kindness, concern, and support to me during my research journey. In addition, I want to thank Xinyi Zhou, Xinying Zhu, and Yue Zhang, who emotionally supported me all the way from my hometown in China.

Finally, I want to thank my parents, sister Haiying Shen, and late grandmother on my father's side for their full support and encouragement of my doctoral studies. So I can face various challenges strongly. I would also like to pay tribute to my late grandfather on my father's side and my late grandparents on my mother's side. The memory of them motivated me during the doctoral journey.

Table of Contents

Examining Committee ii

Author's Declaration iii

Abstract iv

Acknowledgements v

List of Figures x

List of Tables xii

1 Introduction 1

1.1 Background & Objectives 1

1.2 Research Questions 3

1.3 Research Contributions 4

1.4 Scope of the Study 5

1.5 Thesis Organization 6

2 DATA 9

2.1 Data Sources and Dataset Construction 9

2.1.1 Population 9

2.1.2 Urban Spatial Structure . . . 10
2.1.3 Economic Performance . . . 13
2.1.4 Environmental Performance . . . 14
2.1.5 Control Variables . . . 14
2.2 Descriptive Statistics . . . 16

3 The Evolution of Urban Spatial Structure in US: 1900-1970 20
3.1 Introduction . . . 20
3.2 Evolution of Cities . . . 23
3.2.1 Descriptive Maps . . . 23
3.2.2 Trends in Urban Area and Compactness 1900-1970 . . . 26
3.2.3 The Location Distribution of Newly Formed CCA Clusters . . . 31
3.2.4 City Growth Patterns . . . 35
3.2.5 Evolution of the Urban System in the United States . . . 40
3.3 Discussion . . . 46

4 Demographic Shocks and Urban Spatial Change: Lessons from the Great Migration 48
4.1 Introduction . . . 48
4.2 Empirical Strategy . . . 52
4.2.1 Instrumental Variable Construction: Predicted Black Migration . . . 53
4.2.2 Estimating Equations . . . 54
4.3 Results . . . 55
4.4 Discussion . . . 58

5 Urban Spatial Structure and Economic and Environmental Outcomes 60
5.1 Introduction . . . 60
5.2 Empirical Strategy . . . 64

5.2.1 Instrumental Variable Construction: Predicted Urban Spatial Structures . . . 65
5.2.2 Estimating Equations . . . 67
5.3 Results . . . 68
5.3.1 Economic Outcomes . . . 69
5.3.2 Environmental Outcomes . . . 72
5.4 Discussion . . . 75

6 A Causal Mediation Analysis of the Role of Immigration Shocks on Economic and Environmental Outcomes 77
6.1 Introduction . . . 77
6.2 Empirical Strategy . . . 80
6.2.1 Estimating Equations . . . 85
6.3 Results . . . 88
6.3.1 Short-term Economic Outcomes . . . 90
6.3.2 Long-term Economic Outcomes . . . 92
6.3.3 Short-term Environmental Outcomes . . . 94
6.3.4 Long-term Environmental Outcomes . . . 97
6.4 Discussion . . . 100

7 CONCLUSION 105
7.1 Limitations of the Study . . . 109
7.1.1 Limitation of the Data Availability . . . 109
7.1.2 Limitation of Air Quality Data . . . 110
7.1.3 Limitation of Mismatch Between Data Boundaries . . . 111
7.2 Future Research . . . 111

References 114

APPENDICES 121

A Newly Formed CCA Clusters Distribution of Selected Syncities **122**

B Full Results of the Relationship Between Urban Spatial Structure and Outcomes **124**

B.1 Effects of Urban Spatial Structure on Economic Performance: Short-term . 124

B.2 Effects of Urban Spatial Structure on Economic Performance: Long-term . 125

B.3 Effects of Urban Spatial Structure on Environmental Performance: Short-term 126

B.4 Effects of Urban Spatial Structure on Environmental Performance: Long-term 129

C Full Results of the IV Mediation Analysis **133**

List of Figures

2.1 Defining CCA Clusters: An Example 11
2.2 Constructing Syncities: An Example 12

3.1 9 Representative Syncities in the US: 1900, 1940, and 1970 24
3.2 Urban Hierarchy [40]. 26
3.3 US National Level Urban Spatial Structure 1900-1970 28
3.4 US Syncity Level BUA 1900-1970: 9 Representative Examples 29
3.5 US Syncity Level Cohesion Index 1900-1970: 9 Representative Examples . 30
3.6 New CCA Clusters Distribution: kdensity 32
3.7 New CCA Clusters Distribution: Top-down View 34
3.8 Probability of Being Part of a Syncity 37
3.9 Location of the Syncity Core Center 38
3.10 Zipf's Illustration of the Rank-size Rule for US City Population: 1900-1970 41
3.11 Zipf's Illustration of the Rank-size Rule for US Urban Spatial Structure: 1900-1970 42
3.12 z Value for Rank-size Rule 45

4.1 The Effect of the Racial Composition Shock on Urban Spatial Structure . . 52

5.1 The Effect of the Urban Spatial Structure on Outcomes 64
5.2 The Construction of Predicted Urban Footprint on Map: An Example . . . 67

6.1 The Effect of the Great Migration on Outcomes Through the Urban Spatial Structure 80
6.2 General IV Mediation Model [10]. 81
6.3 Restricted IV Mediation Model [10]. 83
6.4 Mediation Analysis Causal Chains 85

A.1 Two Examples of New CCA Clusters Distribution: Kdensity 123

List of Tables

2.1 Summary Statistics of Key Variables 16

3.1 Summary Statistics of Urban Spatial Structure from 1900 to 1970 27
3.2 Coefficients: Distance to Centroid vs. Distance to Center 39

4.1 [Dependent: Ln BUA 1970] 56
4.2 [Dependent: Ln Cohesion Index 1970] 57

5.1 Effects of of Urban Size on Economic Outcomes 69
5.2 Effects of of Urban Shape on Economic Outcomes 70
5.3 Effects of of Urban Size on AQI Outcomes 72
5.4 Effects of of Urban Shape on AQI Outcomes 73

6.1 Changes in Share of Blacks (T) → Urban Size (M) → Short-term Income Outcomes (Y) 90
6.2 Changes in Share of Blacks (T) → Urban Shape (M) → Short-term Income Outcomes (Y) 91
6.3 Changes in Share of Blacks (T) → Urban Size (M) → Long-term Income Outcomes (Y) 92
6.4 Changes in Share of Blacks (T) → Urban Shape (M) → Long-term Income Outcomes (Y) 93
6.5 Changes in Share of Blacks (T) → Urban Size (M) → Short-term AQI Outcomes (Y) 94

6.6 Changes in Share of Blacks (T) → Urban Shape (M) → Short-term AQI Outcomes (Y) . . . 95
6.7 Changes in Share of Blacks (T) → Urban Size (M) → Long-term AQI Outcomes (Y) . . . 97
6.8 Changes in Share of Blacks (T) → Urban Shape (M) → Long-term AQI Outcomes (Y) . . . 98

B.1 [Dependent: Ln Total Income 1974] . . . 124
B.2 [Dependent: Ln Total Income 1974] . . . 125
B.3 [Dependent: Ln Total Income 2010] . . . 125
B.4 [Dependent: Ln Total Income 2010] . . . 126
B.5 [Dependent: Ln Median AQI 1980] . . . 126
B.6 [Dependent: Ln Median AQI 1980] . . . 127
B.7 [Dependent: Ln Max AQI 1980] . . . 127
B.8 [Dependent: Ln Max AQI 1980] . . . 128
B.9 [Dependent: Ln Ratio of unhealthy weather days 1980] . . . 128
B.10 [Dependent: Ln Ratio of unhealthy weather days 1980] . . . 129
B.11 [Dependent: Ln Median AQI 2010] . . . 129
B.12 [Dependent: Ln Median AQI 2010] . . . 130
B.13 [Dependent: Ln Max AQI 2010] . . . 130
B.14 [Dependent: Ln Max AQI 2010] . . . 131
B.15 [Dependent: Ln Ratio of unhealthy weather days 2010] . . . 131
B.16 [Dependent: Ln Ratio of unhealthy weather days 2010] . . . 132

C.1 Changes in Share of Blacks (T) → Urban Size (M) → Short-term Income Outcomes (Y) . . . 133
C.2 Changes in Share of Blacks (T) → Urban Shape (M) → Short-term Income Outcomes (Y) . . . 135
C.3 Changes in Share of Blacks (T) → Urban Size (M) → Long-term Income Outcomes (Y) . . . 136

C.4 Changes in Share of Blacks (T) → Urban Shape (M) → Long-term Income Outcomes (Y) . . . 137
C.5 Changes in Share of Blacks (T) → Urban Size (M) → Short-term AQI Outcomes (Y) . . . 138
C.6 Changes in Share of Blacks (T) → Urban Shape (M) → Short-term AQI Outcomes (Y) . . . 139
C.7 Changes in Share of Blacks (T) → Urban Size (M) → Long-term AQI Outcomes (Y) . . . 140
C.8 Changes in Share of Blacks (T) → Urban Shape (M) → Long-term AQI Outcomes (Y) . . . 141

Chapter 1

Introduction

The cities we inhabit today are the result of historical transformations that have shaped their form, function, and sustainability. Among the most significant of these transformations is the Great Migration–a massive movement of African Americans from the rural South to urban centers in the North and West during the 20th century. This migration not only redefined the demographics of American cities but also influenced their economic opportunities, spatial organization, and environmental challenges. Yet, the effects of this period on the physical structure of cities and their lasting economic and environmental consequences remain underexplored. How did this sweeping demographic shift reshape the urban landscape, and what can it teach us about the interconnected dynamics of population, economy, and sustainability? This chapter sets the stage by outlining the historical background of the Great Migration and introducing the objectives and research questions that guide this study.

1.1 Background & Objectives

Understanding how cities respond to major demographic changes is crucial for addressing today's challenges of urban sustainability, economic equity, and environmental health. One of the most significant demographic shifts in American history was the Great Migration, a period of large-scale movement of African Americans from the South to urban centers in the North and West between 1910 and 1970. This migration fundamentally reshaped the social, economic, and spatial landscape of the receiving cities. By 1950, more than 2.5 million Southern-born African Americans had moved to non-Southern areas, a number that grew to over 4 million by 1980, with additional impacts from subsequent generations

[56]. Despite the profound changes brought about by this migration, its effects on urban spatial structure and its long-term consequences remain underexplored.

This thesis addresses this gap by focusing on how the Great Migration influenced the size and shape of cities and their subsequent economic and environmental outcomes. The study is motivated by the need for better historical data on urban spatial structure and a deeper understanding of how major demographic shifts shape urban development.

1. **To construct a reasonable and useful urban spatial dataset and study the urban spatial development during the period of the Great Migration.** By redefining historical city boundaries through the City Clustering Algorithm (CCA), this study creates a dataset of synthetic cities (Syncities) from 1900 to 1970. This dataset provides a clearer picture of urban spatial development during the Great Migration, addressing the limitations of administrative boundaries and traditional datasets.

2. **To investigate the relationships between the Great Migration and the urban spatial structure by applying the novel urban spatial dataset we construct.** The thesis investigates how the influx of African Americans during the second wave of the Great Migration (1940–1970) shaped urban spatial characteristics, including size and cohesion, in receiving cities.

3. **To verify the importance of the urban spatial structure in the last century in America.** By focusing on urban size and shape, the study explores their roles in shaping income levels and air quality in the short term (1974 and 1980) and the long term (2010). It aims to demonstrate both economic benefits and environmental challenges associated with urban expansion.

4. **To examine the underlying mechanism of the impact of the Black population redistribution on economic and environmental performance.** The research examines how urban spatial structure mediated the effects of the Great Migration on economic and environmental outcomes. Using mediation analysis, it identifies direct and indirect pathways, offering a more nuanced understanding of these relationships.

Together, these objectives provide a framework for analyzing the interplay between historical demographic shifts, urban spatial evolution, and sustainability challenges.

1.2 Research Questions

To achieve the objectives of this thesis, the following research questions guide the exploration of how the Great Migration influenced urban spatial structure and its broader economic and environmental implications. These questions are designed to address gaps in existing research and connect the historical impacts of demographic shifts to contemporary urban challenges.

1. **How did cities grow spatially during the 20th century?**

   (a) ***How did urban spatial structure evolve in the United States from 1900 to 1970?***

   (b) ***What patterns characterized the growth and organization of the U.S. urban system during this period?***

   Answering these questions will help trace the trajectory of urban development in the U.S., focusing on how cities expanded horizontally, how their spatial connectivity changed, and how these changes contributed to the organization of the urban system as a whole. This includes examining newly formed urban clusters, analyzing city growth patterns, and applying Zipf's Law to quantify trends in urban hierarchy.

2. **What was the impact of the Second Wave of the Great Migration on urban spatial structure?**

   This question investigates how changes in racial composition influenced urban expansion and spatial connectivity at the end of the Great Migration in 1970. The study seeks to quantify how the arrival of African American migrants affected the built-up area and cohesion of cities, shedding light on how demographic shifts translated into changes in urban form.

3. **What are the economic and environmental consequences of urban spatial structure?**

   These questions explore the outcomes of urban spatial structure on two critical dimensions: economic performance and environmental quality. By investigating the relationships between urban size, shape, and these outcomes, the analysis aims to uncover economic benefits, such as higher income levels, and environmental costs, such as deteriorating air quality, over both short-term and long-term horizons.

4. **What is the underlying mechanism of the impact of the African American migrants' redistribution on economic and environmental performance?**

(a) ***What were the short- and long-term economic and environmental impacts of the influx of African American migrants during the second wave of the Great Migration (1940-1970) on U.S. cities?***

(b) ***How did changes in urban size and spatial structure mediate the relationship between this demographic shift and its effects?***

Addressing these questions involves identifying the pathways through which changes in racial composition shaped income levels and air quality. The study focuses on the role of urban spatial structure as a mediating factor, analyzing both the direct effects of the Great Migration on economic and environmental outcomes and the indirect effects channeled through changes in urban size and shape. This approach provides a deeper understanding of the processes that link demographic shifts to urban development.

Together, these questions form a comprehensive framework for examining how the Great Migration reshaped urban America. They connect historical demographic changes to broader themes of urban sustainability, economic equity, and environmental health, providing insights into the lasting impacts of past events on contemporary urban challenges.

1.3 Research Contributions

This thesis makes several important contributions to the study of urban spatial structure, migration, and sustainability.

First, it introduces a novel dataset of synthetic cities (Syncities) from 1900 to 1970, constructed using the City Clustering Algorithm (CCA). This dataset provides a unique and precise view of historical urban spatial structure, addressing the limitations of administrative boundaries and offering a foundation for new empirical research in urban studies. By redefining urban boundaries based on data rather than subjective criteria, the thesis ensures more accurate measurements of urban size and cohesion.

Second, this thesis advances our understanding of the Great Migration's impact on urban form. While prior research has primarily focused on the social and economic dimensions of migration, this study is the first to systematically examine its effects on urban size and shape. By quantifying changes in urban spatial structure and connecting them to racial composition shifts, the research fills a critical gap in the literature on demographic shocks and urban development.

Third, the research explores both economic and environmental outcomes, highlighting the interactions between urban expansion and sustainability. By analyzing the short- and long-term impacts of urban spatial structure on income levels and air quality, the study provides a nuanced perspective on how historical events shape contemporary challenges. This dual focus ensures a balanced consideration of urban growth's benefits and costs.

Fourth, the thesis employs a rigorous methodological framework that combines mediation analysis and instrumental variable (IV) approaches to disentangle complex causal relationships. The mediation analysis identifies the urban spatial structure as a critical intermediary, capturing how changes in racial composition indirectly influence economic and environmental outcomes through the urban form. For example, the study demonstrates how the influx of African American migrants altered city size and cohesion, which in turn affected income levels and air quality. The IV approach strengthens these findings by addressing endogeneity concerns, such as reverse causality and omitted variable bias. By using predicted Black migration and predicted urban spatial structures as instruments, the analysis isolates the causal effects with greater precision. Together, these methods provide a robust framework for studying the interplay between demographic shifts, urban form, and outcomes, advancing methodological practices in the field.

Finally, this work has significant policy implications. By linking historical demographic shifts to contemporary urban challenges, the thesis underscores the importance of considering past events in planning for sustainable cities. The findings inform debates on urban sustainability, racial equity, and environmental justice, demonstrating the long-term consequences of demographic and spatial transformations. Moreover, the research offers valuable insights for urban planning, highlighting the importance of designing cities that balance economic growth with environmental sustainability.

1.4 Scope of the Study

This study focuses on the impact of the Great Migration on urban spatial structure and its economic and environmental consequences. The scope is deliberately structured to balance historical breadth and empirical depth, ensuring a comprehensive analysis of how demographic shifts influenced urban development. Key aspects of the study's scope are outlined below:

Temporal Scope

The analysis spans the 20th century, with a primary focus on the period from 1900 to 1970. This time frame captures the historical evolution of urban spatial structure

during the Great Migration, particularly the second wave (1940–1970). To assess long-term outcomes, the study also examines economic and environmental impacts up to 2010.

Geographical Scope

The study is limited to urban areas in the United States, specifically those significantly affected by the Great Migration. By redefining urban boundaries through the City Clustering Algorithm (CCA), the research constructs synthetic cities (Syncities) that better reflect historical urban expansion than traditional administrative boundaries. This approach ensures consistency in analyzing urban forms across diverse regions.

Conceptual Scope

The research is centered on urban spatial structure, defined by two key dimensions: urban size (built-up area) and urban shape (cohesion index). These dimensions are analyzed in relation to economic outcomes (income levels) and environmental outcomes (air quality). The study explicitly focuses on causal pathways, emphasizing the mediating role of urban form in linking demographic changes to urban outcomes.

Methodological Scope

The study employs a mixed-methods approach, combining quantitative techniques like mediation analysis and instrumental variables with qualitative methods such as interpretative analysis of maps. The quantitative methods address endogeneity concerns and allow for robust causal inference, while the qualitative methods identify the trajectory of urban evolution. The study is grounded in a novel dataset constructed to overcome the limitations of existing historical urban data.

Exclusions and Limitations

While the focus on urban size and shape provides valuable insights, the study does not examine other dimensions of urban spatial structure, such as land use or transport networks. Additionally, the analysis is restricted to the U.S. context and may not fully capture cross-national variations in urban dynamics. The temporal scope ends in 2010, meaning recent trends in urban development are beyond the thesis's purview.

1.5 Thesis Organization

This thesis is organized into seven chapters, each building toward a comprehensive understanding of how the Great Migration reshaped urban spatial structure and its economic and environmental consequences. The chapters are designed to guide the reader from

the historical background to methodological innovations, empirical analyses, and broader implications, providing a cohesive narrative that addresses the research objectives.

Chapter 1 introduces the research problem, objectives, and key questions that guide the study. It provides an overview of the Great Migration's historical significance and its potential long-term impacts on urban spatial structure, economic outcomes, and environmental sustainability. The chapter also outlines the contributions of the thesis and its broader relevance to urban and environmental research.

Chapter 2 describes the construction of the novel dataset of synthetic cities (Syncities) using the City Clustering Algorithm (CCA). It explains how historical urban boundaries are redefined to better capture spatial development and provides detailed descriptive statistics for the key variables.

Chapter 3 examines the evolution of urban spatial structure in the United States from 1900 to 1970. It identifies key trends in urban growth and spatial organization, highlighting how the Great Migration fits within broader patterns of urban development. The chapter sets the stage for understanding how historical urban dynamics inform the analyses in subsequent chapters.

Chapter 4 investigates how the arrival of African American migrants influenced the size and shape of cities in receiving regions. Using an instrumental variable approach, the analysis quantifies the causal effects of demographic changes on urban form, emphasizing the interplay between migration and spatial expansion or cohesion.

Chapter 5 explores the outcomes of urban size and shape on two critical dimensions: income levels and air quality. It examines both short-term impacts (1974 and 1980) and long-term consequences (2010), highlighting the interactions between economic growth and environmental degradation. The findings underscore the dual-edged nature of urban expansion, offering insights into the sustainability challenges posed by historical urban dynamics.

Chapter 6 investigates the pathways through which the Great Migration influenced economic and environmental outcomes. Using mediation analysis, it demonstrates how changes in urban spatial structure acted as intermediaries, linking demographic shifts to broader urban outcomes. This chapter provides a deeper understanding of the mechanisms driving the observed effects.

Chapter 7, the final chapter, synthesizes the findings of the thesis, discussing their broader implications for urban sustainability, economic equity, and environmental justice. It highlights the policy relevance of the study, particularly in addressing the contemporary challenges of urban planning and demographic change. Moreover, this chapter discusses

limitations regarding data and also identifies areas for future research, building on the methodological and empirical advances of this work.

Together, these chapters offer a comprehensive analysis of the Great Migration's lasting impacts on American cities, bridging historical demographic shifts with modern urban challenges.

Chapter 2

DATA

2.1 Data Sources and Dataset Construction

This section discusses data sources and construction for key variables, including population, urban spatial structure, economic level, environmental outcomes, and necessary control variables[1].

2.1.1 Population

The city population data at the beginning (1940) and the end (1970) of the second wave of the Great Migration are derived from the County and City Data Book 1944-1977 (CCDB) [63] and from the full-count US censuses in the Integrated Public Use Microdata Series (IPUMS) [55], respectively. These city-level data are for both the overall population and the Black population. Then, we aggregate these data into the county-Syncity union level for further calculations.

Following Derenoncourt's paper [16], we define the changes in the share of Blacks between 1940 and 1970 in non-southern county-Syncity union in the United States following the equation below:

$$\Delta Bpop_{i,1940-1970} = \frac{(Bpop_{i,1970} - Bpop_{i,1940})}{pop_{i,1940}} \tag{2.1}$$

[1]Regardless of whether the original data were at the city-level, Syncity-level, or county-level, they finally aggregated to the county-Syncity union level to be comparable and consistent to each other.

where i denotes each county-Syncity union. The change in the share of Blacks is the urban Black population increase from 1940 to 1970 as a share of the urban population in 1940.

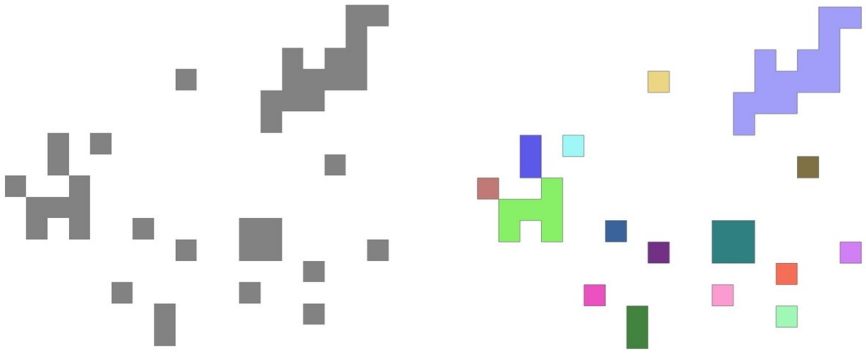
2.1.2 Urban Spatial Structure

The urban spatial structure includes two dimensions, which are urban size and urban shape. Before calculating these two urban spatial metrics, we first need to determine the historical urban boundaries. We choose to construct our own urban boundaries by using building data rather than administrative urban boundaries for three reasons. First, the administrative urban boundaries have limited data available for us to study cities from 1900 to 1970 [28]. Second, the administrative boundaries are often hard to capture the true economic and social dynamics of urban areas [24]. Third, other alternative methods to administrative boundaries, such as commuting zones, nighttime lights, and population density, are limited by the lack of historical data [5, 24]. Thus, building data is the most suitable choice for us to define the historical urban boundaries. We delineate the urban areas by following the steps below:

1. **Defining the CCA clusters.** We obtained the raster datasets of built-up area (BUA) from 1900 to 1970 derived from the Historical Settlement Data Compilation for the United States (HISDAC-US) [29]. This raw dataset divides the United States into grid cells for each half-decade from 1810 to 2015. The original data source is provided by Zillow, which is one of the most comprehensive sources of real estate information. According to this raw raster data, each grid cell is counted as 1 when there is at least one built-up property in a given year, and it is counted as 0 when there is no built-up property. In Figure 2.1(a), cells in Black are counted as 1, and white areas are counted as 0. Moreover, each grid cell represents a 250m $\times$ 250m area on the map. Following the City Clustering Algorithm (CCA), we set a 250-meter threshold for clustering cells in the raster BUA dataset [57]. For each cell that is counted as 1, they will be considered one CCA cluster whenever the distance between them is smaller than the threshold[2]. As shown in Figure 2.1(b), the cells in Figure 2.1(a) are grouped to represent continuous urban clusters rather than isolated pixels. Also, clustering at a scale of a 250-meter threshold allows for a meaningful

[2]Given that the average length of a block in America is around 200 to 900 feet (60 m - 275 m). For example, the city block in Chicago is around 330 by 660 feet (100 m $\times$ 200 m) [51]. Thus, a 250-meter threshold means whenever the distance between cells is less than the length of a city block, these cells are considered part of the same CCA cluster.

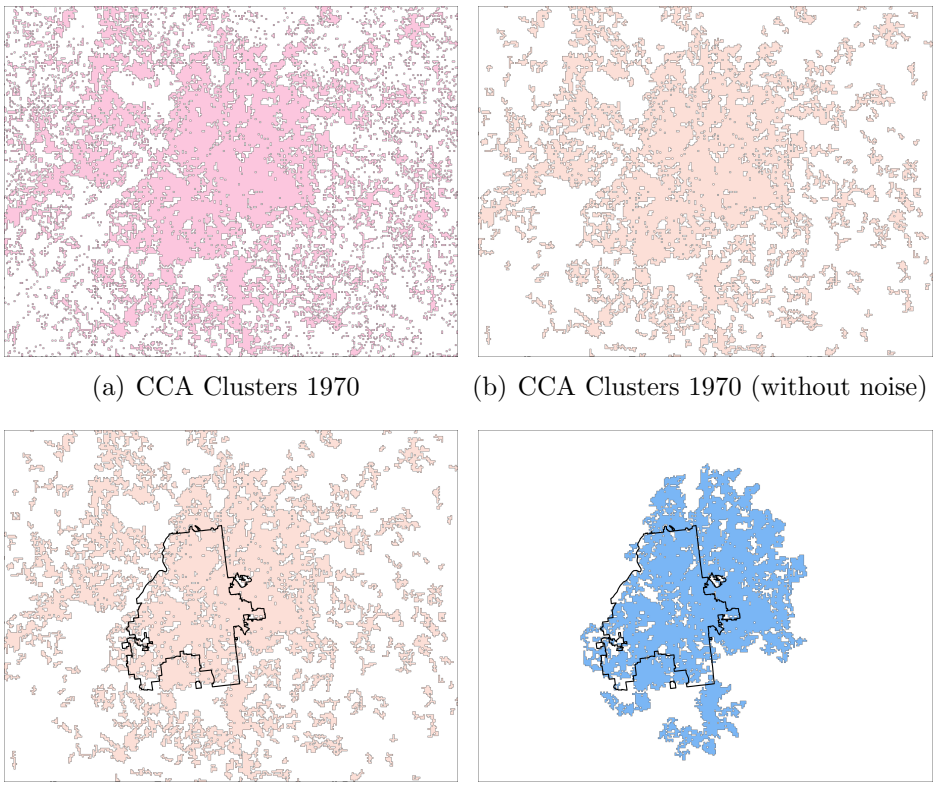
analysis of urban development patterns while avoiding the noise that might come with smaller or larger thresholds.



(a) BUA Grid Cells (b) CCA Clusters

Figure 2.1: Defining CCA Clusters: An Example
Note: For each decade in our sample, we used R to cluster the gird cells (Figure 2.1(a)) into polygons (Figure 2.1(b)) for the threshold we set. The example in Figure 2.1(b) shows the clustering in 1900 for what is now around Clark Atlanta University in Atlanta. Each color in Figure 2.1(b) identifies a cluster.

2. **Constructing the Syncities.** As shown in Figure 2.2(a), the pink polygons display the CCA clusters' appearance in 1970. However, they cannot all be considered urban clusters because some are too small. Therefore, as shown in Figure 2.2(b), we dropped the CCA clusters with seven or fewer gird cells (when the areas of the CCA cluster polygons $\leq$ 437,500 square meters). This elimination of noise allowed us to zoom in on only the significant CCA clusters, which are more likely to be urban areas rather than rural areas. Then, We intersected the 1970 CCA clusters without noise with the 2020 administrative city boundaries we obtained from the Integrated Public Use Microdata Series (IPUMS) [54], shown in Figure 2.2(c). Then, we could identify the 1970 CCA clusters that belonged to contemporary cities and then define the 1970 Syncity boundaries and shapes in Figure 2.2(d). Finally, we could identify Syncity boundaries and shapes from 1900 to 1960 by keeping the cleaned CCA clusters within and intersecting with Syncity boundaries in 1970. The final map shows how CCA clusters emerged and merged within the Syncity boundaries decade by decade from 1900 to 1970.



(a) CCA Clusters 1970 (b) CCA Clusters 1970 (without noise)

(c) CCA Clusters (d) CCA Clusters

Figure 2.2: Constructing Syncities: An Example

Note: Figure 2.2(a) shows an example of CCA clusters on the map around Atlanta in 1970. Figure 2.2(b) shows the CCA clusters when dropping the CCA cluster in Figure 2.2(a) that is $\leq$ 437,500 square meters. Figure 2.2(c) shows how the CCA clusters in Figure 2.2(b) intersect with the administrative city boundaries of Atlanta in 2020. Finally, Figure 2.2(d) shows the 1970 Atlanta Syncity urban areas in blue polygons that intersect with the 2020 Atlanta administrative boundary.

Next, we picked two representative shape metrics to show the multidimensional aspects of the Syncities we defined. Firstly, we used the built-up area (BUA) to study the size of each Syncity. Recall that the BUA on the map represents the total area of grid cells marked as one, which is 625000m^2 (250m $\times$ 250m) per such cell. We added up the total BUA for the cells included in each Syncity to serve as the horizontal size of the urban

land area[3]. A higher BUA value means a larger area of the Syncity, which means a larger horizontal expansion.

We used the cohesion index (S) to study Syncities' shape. The cohesion index (S) is used to represent the compactness characteristic of a Syncity, and the equation is shown below:

$$\mathrm{S} = \frac{d_1 + d_2 + \cdots + d_n}{\#\ of\ point\ pairs} \tag{2.2}$$

where $d_1, d_2, ..., d_n$ denotes the distance between the center point of any two CCA clusters within the same Syncity.

Similar to Harari's paper [44], we defined our cohesion index as the average distance between the center point of any two CCA clusters within a Syncity. A higher distance between CCA clusters in a Syncity indicates a more significant disconnection and a longer transportation distance between those clusters. In other words, a higher cohesion index value means a lower connectivity level and higher spatial dispersion in a Syncity.

2.1.3 Economic Performance

We are interested in past economic performance during the second wave of the Great Migration and today's economic performance. Regarding the past, we used the total income in 1974 to measure the economic performance. The total income is the product of the city population for 1975 and city-level income per capita for 1974, obtained from the County and City Data Book (CCDB) [63]. In addition, the total income in 2010, used to measure today's economic performance, is the product of the city population in 2010 and city-level income per capita in 2010. However, the CCDB only covers the period until 1977, so the city population and income data in 2010 are derived from the Integrated Public Use Microdata Series (IPUMS) [54] instead.

To analyze the effects of the changes in the share of Blacks and urban spatial structure on income levels in 1974 and 2010, we adjusted the income values using the Consumer Price Index (CPI) to account for inflation. Specifically, we converted the historical dollar values in 1974 into the dollar values in 2010 to consistently compare the income level across different periods. In other words, we can ensure that the outcomes of both causal relationships are measured on the same financial standard, making the analysis more meaningful.

[3]The BUA does not precisely capture the actual building footprint; it indicates the presence of built-up properties. Therefore, the BUA is greater than or equal to the total actual building footprint over all cells.

2.1.4 Environmental Performance

We used the Air Quality Index (AQI) data to measure environmental performance. The main index we focus on is the county-level median AQI in 1980 and 2010, derived from the United States Environmental Protection Agency (EPA) [18]. We also tested the county-level max AQI and ratio of unhealthy weather days in 1980 and 2010, derived from the EPA. In detail, AQI is an indicator that measures air quality. A higher number means worse air quality. To clarify, an AQI from 0 to 50 means good air quality, 51 to 100 means moderate air quality, 101 to 150 means unhealthy air quality for sensitive groups, 151 to 200 means unhealthy air quality, 201 to 300 means very unhealthy air quality, and values over 300 means hazardous air quality [1]. Based on these categories, the ratio of unhealthy weather days is the ratio of the number of days with AQI over 100 out of the total AQI recorded days.

2.1.5 Control Variables

Control variables are needed to eliminate the influence of confounding factors in the empirical study. Due to the similarity of the research background and some core variables, we refer to the selection of control variables in the study of Derenoncourt [16]. These control variables include shares of the manufacturing labor force, educational upward mobility, the share of the Black migrants, and region fixed effect. There may be multiple relationships between these variables and the urban Black population, urban size, shape, income, and air quality. Therefore, we need to control these potential confounding factors to improve the accuracy of the regression model in the subsequent empirical study.

Shares of the manufacturing labor force. – This variable is about the share of the labor force employed in the manufacturing sector in 1940. The county-level data on labor in the manufacturing sector in 1940 and the total labor force in 1940 were obtained from the Inter-university Consortium for Political and Social Research (ICPSR) 7736 [64]. Different cities may experience different development paths because of their different share of the manufacturing labor force at the beginning of the second wave of the Great Migration. Therefore, we control for the initial manufacturing labor share of the second wave of the Great Migration to study the economic and environmental effects of the changes in the Black population between 1940 and 1970 and the urban spatial structure in 1970 without the disturbance of the manufacturing factors.

Educational upward mobility. – The county-level education data is derived from the full-count IPUMS dataset [55]. The share of the population with Grade 9 and above education

backgrounds in 1940 is used to represent educational upward mobility. At the beginning of the second wave of the Great Migration, a higher proportion of the population with a specific education background means more education opportunities in this place. This place may attract more people who want to pursue better education, especially the Black migrants from Southern areas. They expect to find more opportunities for themselves and the next generation during the Great Migration. This may result in urban spatial structure, economy, or environment changes. Therefore, we need to control the education variable to avoid the bias brought by education's impact on the core variables' causal results.

Share of the Black migrants. – The raw data to calculate this variable is obtained from the full-count IPUMS dataset [55]. It shows the share of mean numbers of Black migration from each southern county to the non-southern city in 1940. The proportion of Black migrants at the beginning of the second wave of the Great Migration may have significantly influenced the choice of destination for subsequent Black migrants from the South. For example, they may be more likely to choose cities where they already have friends and relatives as a destination for migration. In addition, cities with a higher share of Black migrants in 1940 may have experienced different trajectories in the following decades, such as changes in urban spatial structure, economy, or environment. Therefore, we need to control for the proportion of Black migrants at the initial stage to exclude the potential impact of this initial condition on the causal analysis.

Regional fixed effects. – These are dummy variables for different census regions[4]. Different areas may differ significantly regarding history, culture, and economic and social policies. The regional characteristics may affect cities' geographical or economic development, migration patterns, and environmental changes. By controlling regional fixed effects, the study can reduce the disturbance of unobservable factors caused by regional differences in the regression results to explain the causal results more accurately.

[4]There are four census regions, including Northeast Region, Midwest Region, South Region, and West Region.

2.2 Descriptive Statistics

	obs	mean	sd	min	max
Population					
Urban Black Population Increase (1940-1970)	133	50,880.8	164,049.8	-516	1,323,171
Urban Population Increase (1940-1970)	133	46,157.8	182,434	-170,823	1,859,012
Urban Spatial Structure					
Built-up Area (million m^2) (1940)	133	156	276	1.5	1580
Built-up Area (million m^2) (1970)	133	313	522	5.7	3040
Cohesion Index (km.) (1940)	133	10.6	8.1	0	46.6
Cohesion Index (km.) (1970)	133	9.8	8.3	0	45.9
Economic Performance					
Total Income (million) (1974)	133	9,050	25,400	0	194,000
Total Income (million) (2010)	84	15,000	39,600	931	266,000
Environmental Performance					
Median AQI (1980)	103	43.1	19.4	15	134
Median AQI (2010)	118	47.3	10.7	18	87
Max AQI (1980)	103	194.4	47.4	28	336
Max AQI (2010)	118	137.6	44.9	57.5	501
Ratio of Unhealthy Weather Days (1980)	103	0.12	0.12	0	0.63
Ratio of Unhealthy Weather Days (2010)	118	0.03	0.06	0	0.40

Table 2.1: Summary Statistics of Key Variables

Table 2.1 displays descriptive statistics[5] of key variables for non-southern county-Syncity unions. First, in terms of the population, there was an average increase in the urban Black population of around 50 thousand individuals during the second wave of the Great Migration (1940-1970). Through simple calculations, we can estimate that during the second wave of the Great Migration, there was an increase of approximately 6.77 million ($50,880.8 \times 133$) in the urban Black population across non-southern urban areas. According to Tolnay's paper [56], over 4 million southern-born Blacks lived in non-southern regions by 1980. Also, based on the literature about the Great Migration, we know that a significant proportion of this, over 4 million Black migrants, moved during the second wave of the Great Migration. Therefore, the critical reason for the increase in the urban

[5]The descriptive statistics presented here are based on the observations that are analyzed in the following Chapters of empirical studies.

Black population in non-southern areas during this period was the Black migrations from the South to non-southern areas. In addition, the non-southern urban area had an approximately 45 thousand increase in urban population on average over the second wave of the Great Migration (1940-1970). Compared to around 50 thousand increase in urban Black population between 1940 and 1970, which is five thousand more than the increase in total urban population during this period, we reasonably suspect that this five thousand difference is due to the influx of African American migrants into non-southern urban areas, which push numbers of non-Black residents, mostly are White population, to leave non-southern cities. Moreover, the large standard deviations for the increase in population between 1940 and 1970 indicate a significant difference in total and Black urban population across county-syncity unions. In our samples, the urban Black population had the most significant decrease of 516 Black residents while the largest increase of 1,323,171 Black residents. Also, the urban population had the most significant reduction of 170,823 individuals, while the largest increase was 1,859,012 individuals.

Regarding the second key variable, which is the urban spatial structure[6], we focus on two typical spatial metrics, including the built-up area and cohesion index. During the second wave of the Great Migration (1940-1970), the Built-up area increased from 156 million m^2 to 313 million m^2 on average. This 157 million m^2 increase shows double urban horizontal expansion between 1940 and 1970. Moreover, during this period, the minimum built-up area increased nearly fourfold while the maximum built-up area approximately doubled. The difference in built-up areas across county-syncity unions increases over time according to the increased standard deviation between 1940 and 1970. On the other hand, the average cohesion index, which shows the average distance between clusters in Syncities, slightly decreased from 10.6 km in 1940 to 9.8 km in 1970. Similar to the built-up area, the difference in cohesion index across county-syncity unions also increased from 1940 to 1970 based on a more significant standard deviation in 1970 compared to 1940. In addition, the minimum cohesion index is always zero, meaning there was at least a Syncity with only one cluster, whether in 1940 or 1970. Like the average cohesion index, the maximum index slightly decreased from 46.6 km in 1940 to 45.9 km in 1970. These decreases show that the overall spatial connectivity of urban shape was slightly improved while the urban horizontal expansion increased during the second wave of the Great Migration. Although the average cohesion index slightly decreased between 1940 and 1970, our empirical study in Chapter 4 for studying the effect of the increased urban Black population between 1940 and 1970 on the cohesion index in 1970 shows a positive causal result. That is to say, a region with a higher increase in Black population correlates with a higher spatial dispersion

[6]Detailed descriptive statistics of urban spatial structure for all America from 1900 to 1970 are shown in Chapter 3.2.2 when discussing urban spatial development trajectory.

urban shape in that region.

Based on the summary statistics of Economic performance, the average total income in 1974 was \$9,050 million. It increased to \$15,000 million in 2010[7]. The large standard deviations for the income in 1974 and 2010 indicate a significant difference in income performance across our regional units, and the increasing standard deviation from 1974 to 2010 shows that the variation in income level today in 2010 became larger compared with the variation in income level in the last century in 1974. In 1974, the income level was as low as \$0 and as high as \$194,000 million. In addition, in 2010, the income level was as low as \$931 million and as high as \$266,000 million.

The summary statistics of Environmental performance specifically focus on median AQI, max AQI, and ratio of unhealthy weather days in 1980 and 2010. According to Table 2.1, the average median AQI was 43.1 in 1980 and 47.3 in 2010. Based on the AQI guide[8], the median AQI in 1980 and 2010 are considered satisfactory air quality with little or no risk. However, the value is increasing over time, so the air quality is generally worsening. In terms of the Max AQI, which shows the highest value of AQI for the year, the average value across regions in America was 194.4 in 1980 and decreased to 137.6 in 2010. This implies that the extremely unhealthy air quality conditions almost hit very unhealthy levels of concern in 1980, which indicates a risk of health effects for everyone. This unhealthy risk of extreme conditions is reduced over time to the level that only causes health risks to the members of sensitive groups but not the general public in 2010. Recall that the ratio of unhealthy weather days is the proportion of the days with AQI over 100 out of the total AQI recorded days. The average ratio of unhealthy weather days in 1980 was 0.12 and reduced to 0.03 in 2010. This indicates an average of 43.8 unhealthy weather days (0.12×365) annually in 1980, and the number of days reduced to 10.95 (0.03×365) annually in 2010. In addition, the standard deviations for median AQI, max AQI, and ratio of unhealthy weather days decreased over time between 1980 and 2010. That is to say, the variation in each of these three environmental performance indicators becomes smaller, meaning the air quality performance across regions in America becomes more and more similar. Based on our samples, the minimum median AQI in 1980 was as low as 15, and the maximum was as high as 134. On the other hand, 30 years later, the minimum median AQI increased to 18, and the maximum value was reduced to only 87. The max AQI had a minimum value of 28 in 1980, which increased to 57.5 in 2010; the maximum value of the max AQI in 1980 was 336 in 1980, reaching hazardous levels of concern, and this maximum value increased to 501 in 2010. Regarding the ratio of unhealthy weather

[7]We adjusted the monetary values from 1974 to their 2010 equivalents using the Consumer Price Index (CPI) to account for inflation.

[8]The AQI Basics for Ozone and Particle Pollution [1].

days, the minimum number of unhealthy weather days annually was zero in both 1980 and 2010; the maximum number of unhealthy weather days annually was as high as 229.95 (0.63×365) in 1980, and this maximum value reduced to 146 (0.4×365) days annually in 2010.

Chapter 3

The Evolution of Urban Spatial Structure in US: 1900-1970

3.1 Introduction

Urban spatial evolution is a topic that directly impacts people's lives, particularly as cities play an essential role in achieving Sustainable Development Goals (SDGs). The way cities evolve affects economic opportunities, environmental conditions, and social well-being. As of 2008, more people lived in urban areas than in rural ones, and understanding how cities develop has never been more crucial. This chapter aims to shed light on how American cities evolved between 1900 and 1970—a period of significant urban growth and change driven by technological advancements and major demographic shifts.

This chapter addresses the following questions: (i) How did urban spatial structure evolve in the United States from 1900 to 1970? (ii) What patterns characterized the growth and organization of the U.S. urban system during this period? By examining these questions, we aim to understand the physical expansion of cities and how they transformed into interconnected urban systems.

Our investigation is supported by two primary motivations. First, the United States, as a highly urbanized nation, presents a unique case, with urban areas accounting for 74% of the population in 1970 and 80% by 2008. Second, the availability of data from the Historical Settlement Data Compilation for the United States (HISDAC-US) allows us to construct detailed historical urban boundaries, helping to fill a gap in studying U.S. urban spatial development during this important period.

Many previous studies have explored the rise and functions of cities, often focusing on spatial equilibrium theories in economics. For instance, Bosker [41] provided many insights into the factors underlying the creation of cities, while Jedwab, Johnson, and Koyama [49] discussed city growth patterns during the Middle Ages, especially in Europe. However, these studies frequently overlook the initial emergence of cities and neighborhoods, assuming a simple outward expansion from a central core. Furthermore, Uhl et al. [28] highlighted that the lack of comprehensive historical urban data in the U.S. before 1970 limits our understanding of urban spatial development, which creates significant gaps in understanding the underlying mechanisms of urban evolution.

To construct an appropriate urban spatial dataset, it is essential to consider how urban boundaries are defined. Many studies have used existing geographical boundaries of official administrative units to classify and analyze geospatial data, including state, county, MSA, city, or town boundaries [28, 30, 53, 66, 70, 71]. However, relying on administrative definitions that depend on qualitative and subjective criteria can lead to inaccuracies, particularly in rapidly growing urban agglomerations [24]. Administrative boundaries often fail to capture the true economic and social dynamics of urban areas, and they are inconsistent across countries [24].

Galdo et al. [66] further showed bias in the official classification of different places, such as rural and urban, highlighting that many areas classified as rural often exhibit urban characteristics. Dingel et al. [32] also emphasized that economic interactions frequently cross administrative boundaries, which can result in misrepresentation when using traditional definitions. These observations form a core motivation for this chapter's approach: redefining urban boundaries to reflect the true extent of urban areas accurately.

Recent literature has proposed several alternative methods for defining urban boundaries without official administrative limits. Four main approaches include commuting zones, nighttime lights, population density, and building footprints. Commuting zones capture the surrounding travel-to-work regions of a city, providing a functional definition of urban areas [24]. Additionally, using nighttime lights, population density, or building density to identify new geographical boundaries through a set threshold provides a more consistent method for urban delineation. For example, Henderson et al. [24] used building coverage and population density thresholds to define the boundaries of African cities, while Daniel Arribas Bel [11] used a machine learning approach in Spain based on building and search range criteria. Ana Moreno Monroy [5] combined population and transportation data to draw new boundaries for metropolitan areas.

However, there is no reliable historical data in the United States available for us to apply these alternative methods, especially for commuting zones, nighttime lights, and population

density. In our case, building footprint is the most suitable choice, as our primary data source is built-up related data. To delineate urban areas, we applied the City Clustering Algorithm (CCA) method to cluster the building footprint data. Although Daniel Arribas Bel [11] also delineated urban areas based on building footprint, his method relied on highly accurate building data, which is very limited for historical periods. In comparison, our CCA method allows us to cluster building footprints based on relatively less precise construction data, which is more suitable for delineating historical urban boundaries. As a result, this chapter addresses the limitations of the administrative urban boundaries found in previous studies and provides a more comprehensive perspective on urban spatial evolution in the U.S. from 1900 to 1970.

Our contributions are threefold and directly address the limitations highlighted in the literature. First, we introduce a novel approach to delineate historical urban boundaries using the CCA. This methodology stands out because it overcomes the inconsistencies associated with administrative boundaries, which are often influenced by subjective criteria and fail to capture the true extent of urban areas [24, 66]. Unlike traditional administrative boundaries, the CCA uses building footprint data, allowing for a more data-driven and accurate representation of urban boundaries. This approach also builds on previous efforts by Henderson et al. [24] and others but extends their techniques by applying them to historical data in a novel context.

Second, we provide a new spatial dataset for urban areas in the United States from 1900 to 1970. This addresses a significant gap in the literature, as highlighted by Uhl et al. [28], where the lack of comprehensive historical urban data has limited our understanding of urban evolution. By using building footprint data to create this dataset, we offer a valuable tool for systematically examining the long-term evolution of U.S. urban areas. This dataset not only captures urban expansion but also allows for a deeper analysis of how cities responded to significant historical events, such as the Great Migration, which has been underexplored in the context of spatial urban studies.

Third, our study links the historical evolution of urban spatial structure with economic and environmental performance, providing new insights into the long-term implications of urban growth patterns. This extends the work of Dingel et al. [32], who emphasized the importance of understanding economic interactions beyond administrative boundaries. By examining how urban evolution impacted economic outcomes and environmental conditions, our research contributes to a more integrated understanding of urban systems, which can inform current and future urban planning and policy decisions. This linkage between spatial structure and performance is particularly important for understanding the broader implications of urbanization on social and environmental sustainability.

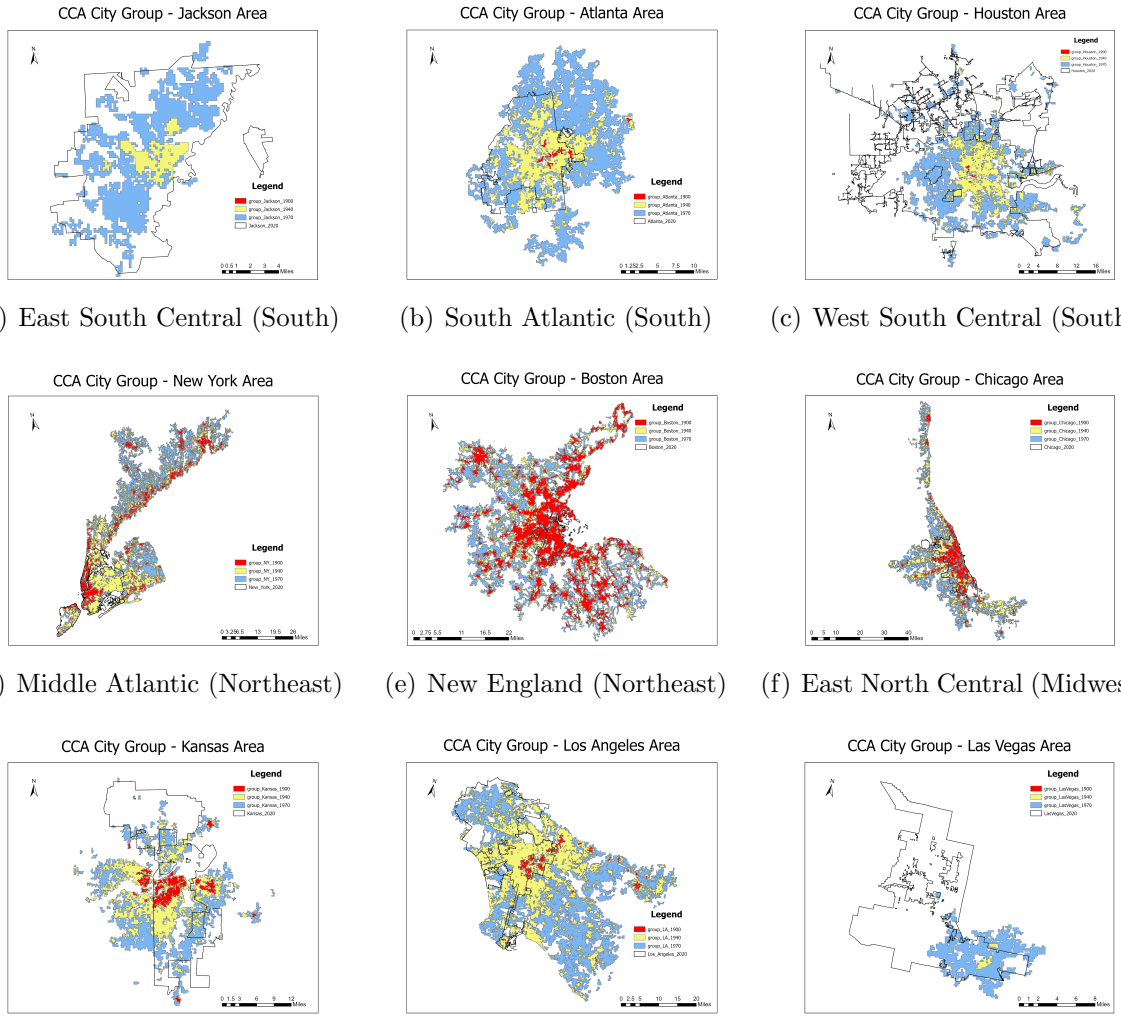
The chapter is organized as follows: Section 3.2 provides a detailed analysis of the urban growth trajectory in the United States from 1900 to 1970. Then, Section 3.3 concludes with insights on how understanding the historical evolution of urban areas can inform current urban planning and policy decisions.

3.2 Evolution of Cities

To deepen our understanding of the evolution of urban spatial structure in the United States, we begin by mapping representative Syncities to illustrate their spatial distribution. We then present development trends in key urban spatial metrics from 1900 to 1970, providing an overview of the major changes in urban form over this period. Next, we analyze the spatial distribution of newly formed clusters within Syncities and empirically examine patterns of city growth. Finally, we investigate the evolution of the broader urban system in the United States, tracing shifts in structure and scale over time.

3.2.1 Descriptive Maps

For each of the nine regions in the United States, we selected nine representative Syncities: Jackson (East South Central region), Atlanta (South Atlantic region), Houston (West South Central region), Kansas City (West North Central region), Boston (New England region), Chicago (East North Central region), New York (Middle Atlantic region), Los Angeles (Pacific region), and Las Vegas (Mountain region). These cities were chosen to observe changes in urban size and shape between 1900 and 1970. We focused on three key years—1900, 1940, and 1970—to provide a cross-sectional view of their development. The multi-polygons in Figure 3.1 show the CCA clusters for each Syncity, illustrating their horizontal size and overall spatial patterns. In the following discussion, we will analyze the changes in size and shape of these nine Syncities across the selected years.



(a) East South Central (South) (b) South Atlantic (South) (c) West South Central (South)

(d) Middle Atlantic (Northeast) (e) New England (Northeast) (f) East North Central (Midwest)

(g) West North Central (Midwest) (h) Pacific (West) (i) Mountain (West)

Figure 3.1: 9 Representative Syncities in the US: 1900, 1940, and 1970

Note: Red polygons represent the CCA city group of Syncities in 1900, yellow polygons represent the CCA city group of Syncities in 1940, and blue polygons represent the CCA city group of Syncities in 1970. In addition, the black lines indicate the administrative city boundaries in 2020.

The maps showed that cities developed at different times and rates. In 1900, the Northeast was the first region to show significant urban clustering, followed by the Midwest, then the South, and finally the West. Los Angeles was an exception in the West, having

some urban clusters in 1900 due to its coastal location, but these clusters were much smaller compared to those in Northeastern coastal cities. For example, Boston, New York, and Chicago each had a dominant urban core in 1900, surrounded by smaller sub-cores. This primary core was significantly larger and denser than the others.

In contrast, cities like Kansas City, Atlanta, and Los Angeles were slower to develop in 1900, with only an initial main core and surrounding clusters of similar size that were more randomly distributed. The slowest growing areas were Jackson, Houston, and Las Vegas, which showed minimal development in 1900, with only a few small clusters or none at all, as seen in Jackson and Las Vegas.

Over time, cities grew at different rates, with coastal cities developing more rapidly. Coastal cities had advantages like better transportation and trade opportunities, including international trade via waterways and access to marine resources, which promoted urban growth. As a result, Los Angeles, which in 1900 was at a similar stage of development as inland cities like Kansas City and Atlanta, experienced much faster growth over the next 40 years. By 1970, the difference had become even more pronounced, with Los Angeles expanding at a significantly higher rate compared to Kansas City and Atlanta.

Regardless of whether a city develops quickly or slowly, its spatial growth follows certain patterns. By 1970, with the exception of the two cities in the western regions, the other seven cities had generally evolved in line with the hierarchical urban cluster patterns observed in Boston, New York, and Chicago in 1900. These patterns are consistent with central place theory, which describes urban systems as a series of interconnected networks where cities of different sizes form a structured hierarchy [40]. This theory provides a useful framework for understanding how urban areas organize themselves spatially, as illustrated in 3.2 [40].

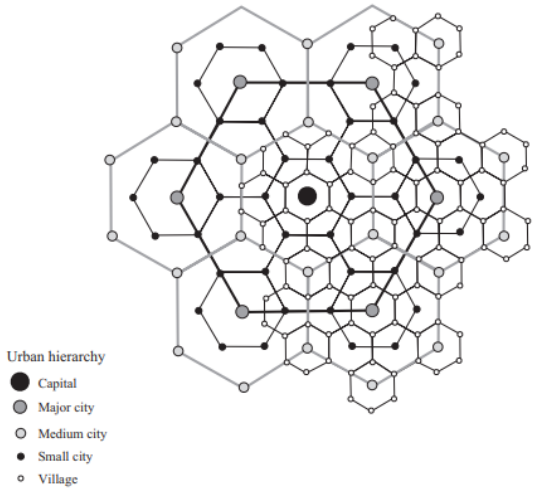


Figure 3.2: Urban Hierarchy [40].

The urban hierarchy shown in Figure 3.2 is not strictly followed in practice due to various real-world factors. Mining areas, for instance, can disrupt the typical distribution of urban clusters by attracting settlements dedicated to specific economic activities, leading to unique growth patterns. Geographic features such as mountains, rivers, and deserts can also impede the formation and growth of urban clusters. For example, Las Vegas did not develop according to the typical urban hierarchy due to the challenges posed by its desert environment. Additionally, the construction of transportation infrastructure, such as highways and railroads, significantly alters natural agglomeration trends, redirecting growth in ways that deviate from the ideal hierarchical model. These factors demonstrate that economic activities, geographic constraints, and infrastructure development can shape urban growth in ways that diverge from the idealized central city model of urban hierarchy.

3.2.2 Trends in Urban Area and Compactness 1900-1970

The statistics in Table 3.1 show urban size and shape trends from 1900 to 1970 in the United States. The urban size (BUA) increased from 23.2 million square meters to 160 million square meters. This indicates rapid urban expansions through the horizontal perspective between 1900 and 1970. Additionally, the standard deviation (SD) for BUA increased year by year, suggesting that differences in BUA among cities across the country were widening over time. In other words, cities developed at different speeds, some growing rapidly while

others lagged. Moreover, the urban spatial dispersion (cohesion index) increased from 5.77 kilometers of the average distance between clusters in 1900 to 9.70 kilometers in 1960 and then decreased to 8.89 kilometers in 1970. This shows that the urban structure became less compact year by year until 1960 and showed an increase in urban compactness after 1960.

	obs	mean	sd	min	max
1900					
BUA	357	2.32e+07	6.32e+07	484375.8	7.69e+08
Cohesion Index	357	5.771672	7.192476	0	47.56336
1910					
BUA	410	2.92e+07	7.88e+07	449629.2	8.54e+08
Cohesion Index	410	6.132561	7.255913	0	49.46152
1920					
BUA	430	4.02e+07	1.03e+08	500000	1.03e+09
Cohesion Index	430	6.595947	7.236134	0	49.53651
1930					
BUA	441	5.59e+07	1.39e+08	562503.2	1.36e+09
Cohesion Index	441	7.445908	7.172131	0	51.318
1940					
BUA	450	6.99e+07	1.64e+08	560552.4	1.58e+09
Cohesion Index	450	8.140234	7.077637	0	55.37835
1950					
BUA	456	9.38e+07	2.06e+08	500003.2	1.88e+09
Cohesion Index	456	8.917134	7.016763	0	60.63997
1960					
BUA	461	1.31e+08	2.72e+08	625000	2.59e+09
Cohesion Index	461	9.697917	7.419329	0	66.99203
1970					
BUA	462	1.60e+08	3.14e+08	812494.3	3.04e+09
Cohesion Index	462	8.888095	7.055229	0	71.72318

Table 3.1: Summary Statistics of Urban Spatial Structure from 1900 to 1970

To further study the development path of urban spatial structure, we use line charts to show the trends in the shape metrics of Syncities from 1900 to 1970 nationwide and in nine sample Syncities representing nine different regions in the United States. To answer the question about how Syncities grow over time, the national trends can help us understand overall urban spatial development. Syncity-level trends can more clearly reflect the specific locations of national trend fluctuations.



(a) BUA

(b) Cohesion Index

Figure 3.3: US National Level Urban Spatial Structure 1900-1970

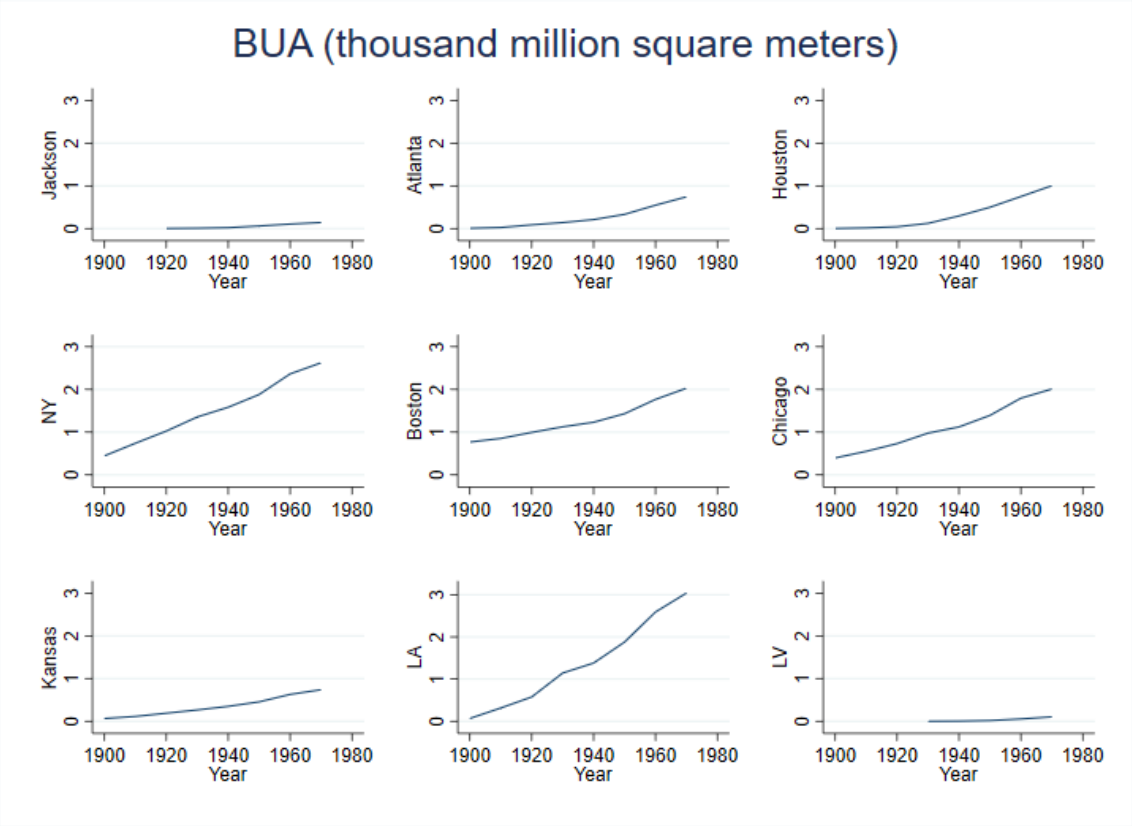


Figure 3.4: US Syncity Level BUA 1900-1970: 9 Representative Examples

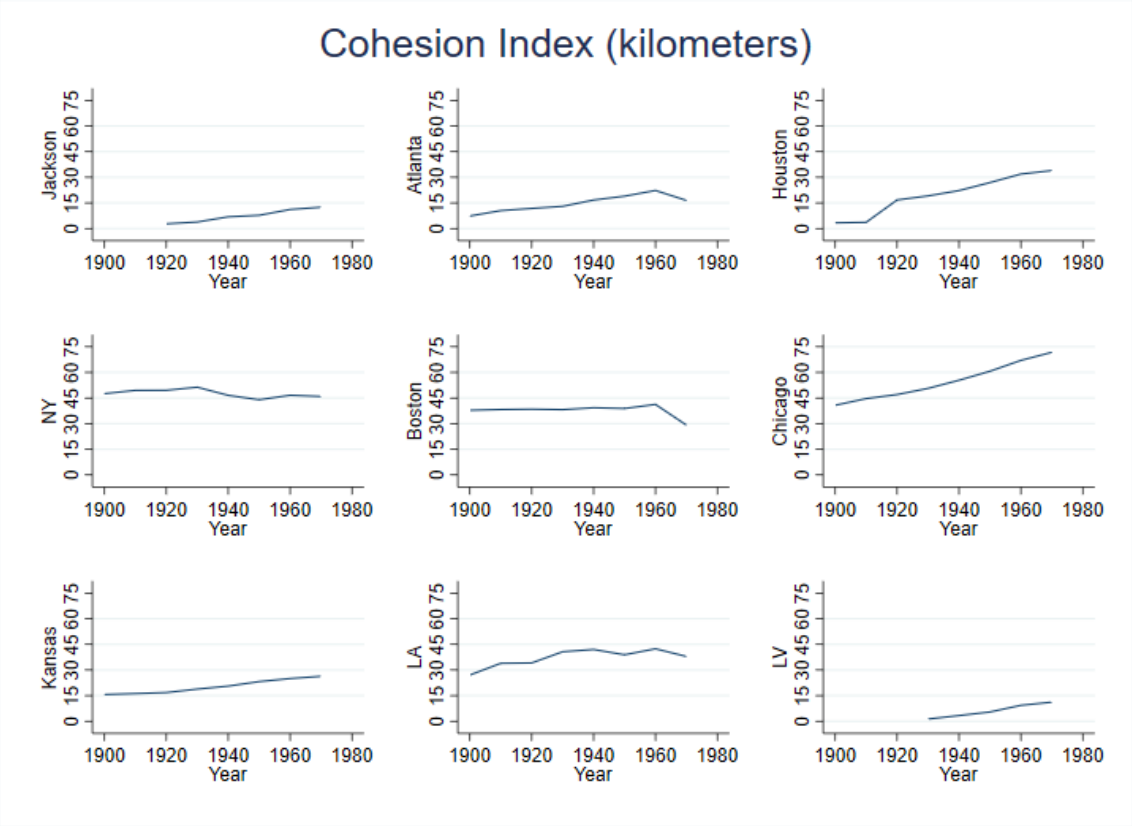


Figure 3.5: US Syncity Level Cohesion Index 1900-1970: 9 Representative Examples

From a national perspective in Figure 3.3(a), the overall trend of BUA is increasing. The magnitude in 1900 is much lower than that in 1970. Moreover, after the dramatic migration shock in 1910, the growth of the BUA in each decade following 1910 was more significant than the growth in the decade before it, and growth accelerated after the beginning of the second wave of the Great Migration (1940) compared to the earlier periods. Furthermore, according to Figure 3.4, the magnitude of the increase of the BUA is different for each region. The increase was most significant in the northern areas, followed by the southern regions, and slower in the West. Again, Los Angeles is an exception in the West due to its unique status as a coastal city.

Among all Syncities, New York experienced a higher starting point and faster growth, while Las Vegas had the slowest growth. This is to be expected because much of the northern regions in the United States, particularly the Middle Atlantic region, which includes the New York metropolitan area, is a powerful financial, business, and technological cen-

ter. Therefore, it is unsurprising that these cities have accelerated development over the decades. In the West, the construction of cities is hindered due to geographical factors such as mountains and deserts. Furthermore, due to its huge area with a small population, the western regions, especially the Mountain region, lack dense transport and infrastructure networks, which harms economic activities and development there.

Compared to the size of the city, the changes in city shape from 1900 to 1970 are more diverse. We observed the degree of urban spatial compactness using the Cohesion Index. Specifically, a large cohesion index shows a longer average distance between CCA clusters within a Syncity, which indicates a less compact Syncity. According to the nationwide Figure 3.3(b), the cohesion index before 1960 increased but peaked in 1960 and began to decline after that. This shows that the Syncities became less and less compact until 1960 and then turned to become more compact.

In addition, the trends in nine sample Syncities in Figure 3.5 provide us with further details. According to the line charts, the cohesion index for most of the Syncities shows an increasing trend, especially after 1910. However, the trend of the cohesion index in New York and Boston, two highly developed cities, remained relatively stable, suggesting that their urban structures maintained a consistent level of compactness. Notably, Boston and Atlanta even faced a decrease in the cohesion index after 1960, showing a move towards a more compact urban shape after 1960. We can reasonably predict that other cities, which experienced an increasing trend of the cohesion index from 1900-1970, may stop their urban sprawl when sufficiently developed.

3.2.3 The Location Distribution of Newly Formed CCA Clusters

The discussion above focused on all CCA clusters in each decade, including older existing and newly formed clusters. However, it is also valuable to focus only on the newly formed CCA clusters in each subsequent decade, as this provides us with new insights into how and where cities construct new buildings over time.

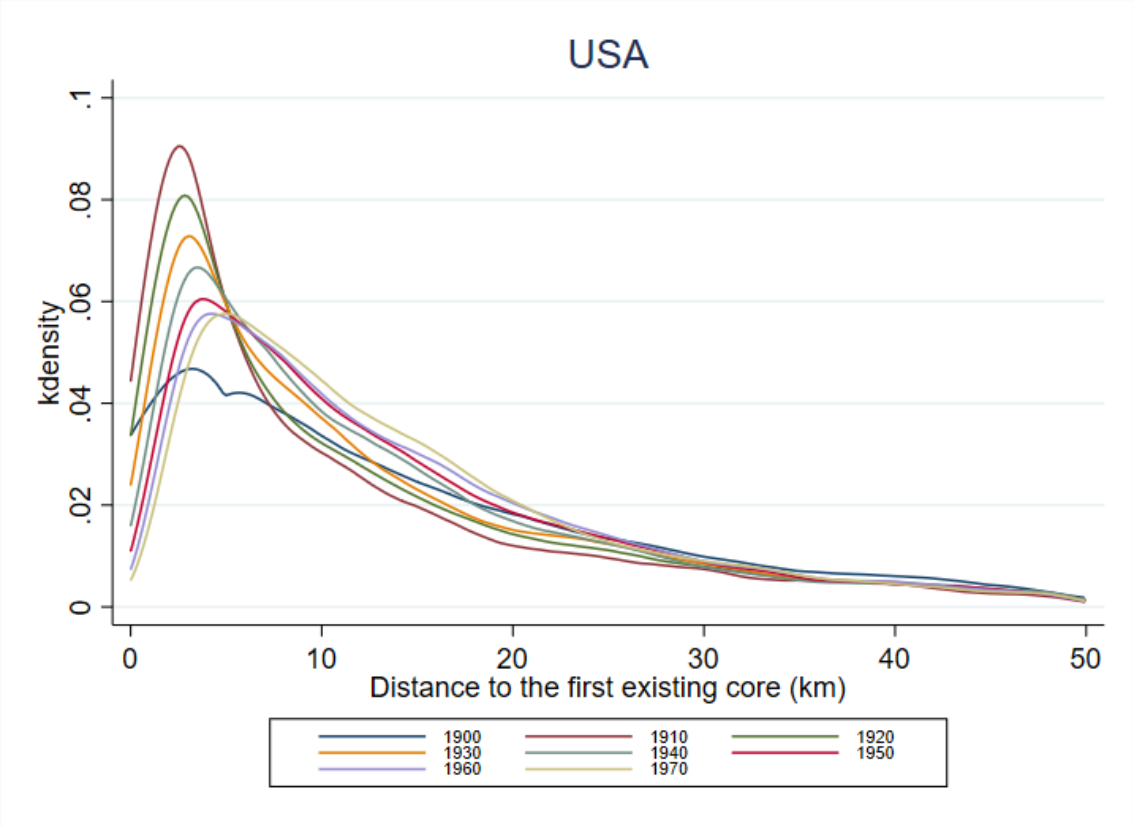


Figure 3.6: New CCA Clusters Distribution: kdensity

Figure 3.6 shows the distribution of new CCA clusters in the USA from 1900 to 1970. The x-axis represents the distance (in kilometers) from the first existing core of each Syncity to each newly formed CCA cluster; the y-axis shows the kernel density, indicating the clusters' concentration at different distances to the core.

Comparing the curve of 1900 with that of other years, we can see two apparent peaks around 1 kilometer and 6 kilometers from the central core. Both reflect that most CCA clusters stay at such distances from the center core. In 1910, a new CCA cluster was formed around 2 kilometers from the center core, so the peak moved to 2 kilometers. The peaks of the curves from 1920 to 1970 are shifting out to the right. For example, the peak remained at 2 kilometers in 1920 but moved to about 5 in 1970. This reflects that highly distributed newly formed CCA clusters spread gradually further away from the center core but within a relatively limited range.

Moreover, the shape of the curves has also changed over the decades. In 1900, the curve

was pretty broad and smoothed. 1900 is a baseline that includes all the CCA clusters that exist. Meanwhile, later years' curves are much narrower and more distinct. They include only new clusters formed in that specific year, making the data more representative and concentrated. Furthermore, the bumps of the curves from 1910 until 1970 become more expansive, which indicates that newly formed clusters are becoming increasingly dispersed.

Figure 3.7 shows an ideal visual presentation of a top-down view to visualize the spatial distribution of newly formed CCA clusters. The black dot in the center is the Syncity center core. The radius of the red and the blue concentric circles are 1 kilometer and 6 kilometer, respectively. Figure 3.7 shows that dark blue building symbols are the typical CCA clusters in 1900. A majority of these clusters are around Red and Blue circles. The intermediate black concentric circles indicate distances of 2 kilometers, 3 kilometers, and 5 kilometers from the center core. We selected three sample years (1910, 1940, and 1970) to show the distribution of newly formed clusters. The brown building symbols represent clusters in 1910 concentrated around 2 kilometers from the core, light blue building symbols are for clusters in 1940 around 3 kilometers from the core, and green building symbols are for clusters in 1970 around 5 kilometers from the core.

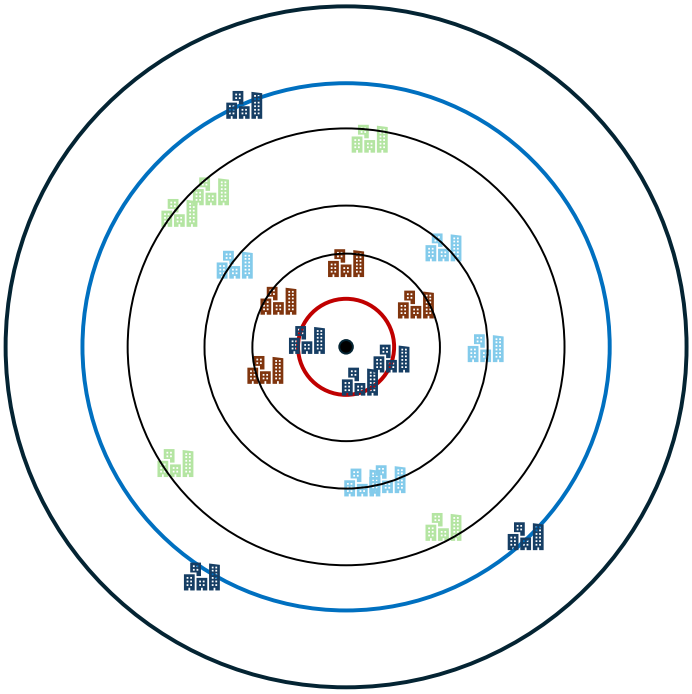


Figure 3.7: New CCA Clusters Distribution: Top-down View

Fig 3.7 shows that, whereas CCA clusters that form appear gradually farther from the center core, they also tend to fill in the spaces between the older clusters at the same time, showing a feature of urban infill rather than all the way sprawl. In other words, urban spatial development involves merging and filling areas between older clusters rather than just expanding outward[1]. Let's consider each CCA cluster as a "seed." The results based on the distribution of newly formed CCA clusters tell us that the city grows not only because each seed expands bigger or new seeds are placed radially from the center core, as the traditional economic model suggests, but also because the new seeds are placed between existing seeds.

[1]The Kdensity curve of the newly formed CCA clusters in two selected Syncities – Atlanta and Cleveland – also shows a similar result as the national-level curves. The Kdensity graphs of Atlanta and Cleveland are in Appendix A.

3.2.4 City Growth Patterns

Two Equilibrium Conditions

The first growth pattern we are interested in is how likely CCA clusters will be part of a Syncity over the years. We applied the equilibrium condition in Allen's paper [59] to test this pattern. Allen [59] proposed allocating a location to the nation whose capital is closest. Specifically, according to the definition of dynamic equilibrium, the territory of a nation at any given time t is determined by the set of locations s that is closer to its capital $Capital_{it}$ compared to other capital $Capital_{jt}$. In other words, the boundary of a nation is shaped by the locations to its nearest capital. So, based on this framework, Allen [59] regressed a dummy variable equal to one if a location s in a nation i in a given time t on the distance between this location s and the capital $Capital_{it}$.

Based on Allen's study [59] of national-level land allocation, we hypothesized that this national-level equilibrium condition can also be applied to our Syncity-level to examine Syncity-level land allocation. Specifically, we empirically tested this condition at the Syncity level by using the given location of the Syncity center, which is determined by the center point of the CCA cluster with the highest build-up intensity[2]. It is worth noting that we keep the Syncity center consistent for each decade, using the center point in the first existing year. So we can make sure that we only have the CCA clusters changing over time while the urban center is constant. In other words, the variation is not because of the changing center point but because the new cluster exists. As a result, we can successfully study if the observed Syncity affiliation of the CCA clusters is consistent with the prediction of the equilibrium condition. The equation is shown below:

$$\mathbf{1}\{a \in Syncity_{ct}\}_{act} = \beta d\left(a, x_{ct}\right) + \delta_{at} + \delta_{ct} + \delta_{ac} + \epsilon_{act} \tag{3.1}$$

The dependent variable is a dummy variable that equals one if CCA cluster a is part of Syncity c in period t. We regressed the dependent variable on the distance between a and the Syncity center x_{ct}, corresponding with three fixed effects to account for various influences. First, we included the CCA cluster – year fixed effect δ_{at}, accounting for the average distance between CCA clusters and all Syncity centers each year. Second, the Syncity – year fixed effect δ_{ct} captures the interaction between the Syncity code and year.

[2]The data of build-up intensity for each CCA cluster is derived from the Historical Settlement Data Compilation for the United States (HISDAC-US) [74]. The build-up intensity reflects the total amount of indoor building area for each CCA cluster, which shows the total built-up volume. Thus, the highest build-up intensity indicates the busiest cluster within a Syncity

Third, we included the CCA cluster – Syncity fixed effect δ_{ac}, representing the interaction between the CCA cluster and Syncity code.

The second growth pattern we are interested in is how the core center of a Syncity is located over the years. According to the definition of dynamic equilibrium in Allen's paper [59], the location of a capital $Capital_{it}$ should equal the location that minimizes the total weighted distance to all points within the territory. So, based on this framework and testing this equilibrium condition empirically, Allen regressed the distance between a location and its real capital on the distance between that location and the location of the centroid point, which can minimize the distance to all points within the territory [59].

To investigate how the urban center is located over time for our Syncity-level empirical case, we tested this equilibrium condition from Allen's paper [59], which assumes that a nation's capital is located to minimize internal transportation costs. Specifically, we tested the distance between the location and the Syncity center, and the distance between the location and the Syncity centroid point using the equation 3.2. Unlike the investigation of the first equilibrium condition, we did not make the Syncity center or the relative centroid point constant over the years but looked at the changes in these pair of points to see whether they grew further or closer to each other over the years.

$$d\left(a, x_{ct}\right)_{at} = \beta d\left(a, x_{ct}^{\text{centroid}}\right)_{at} + \delta_{ct} + \delta_{ac} + \epsilon_{at} \tag{3.2}$$

The dependent variable represents the distance of a CCA cluster to its Syncity's center. We regressed the dependent variable on the distance of the CCA cluster to its Syncity's centroid with two fixed effects for a more robust analysis. One is the Syncity – year fixed effect δ_{ct} that captures the interaction between each Syncity and year, while another is the CCA cluster – Syncity fixed effect δ_{ac} that accounts for the interaction between the CCA cluster and the corresponding Syncity.

Results

Figure 3.8 shows the results of the probability that CCA clusters will become part of the Syncity. Figure 3.9 shows whether the center core of the Syncity is located to minimize internal transportation costs. These results confirm that the two equilibrium conditions proposed by Allen [59] at the national level also apply to the Syncity-level. Additionally, these results indicate the patterns of the CCA clusters during urban spatial development.

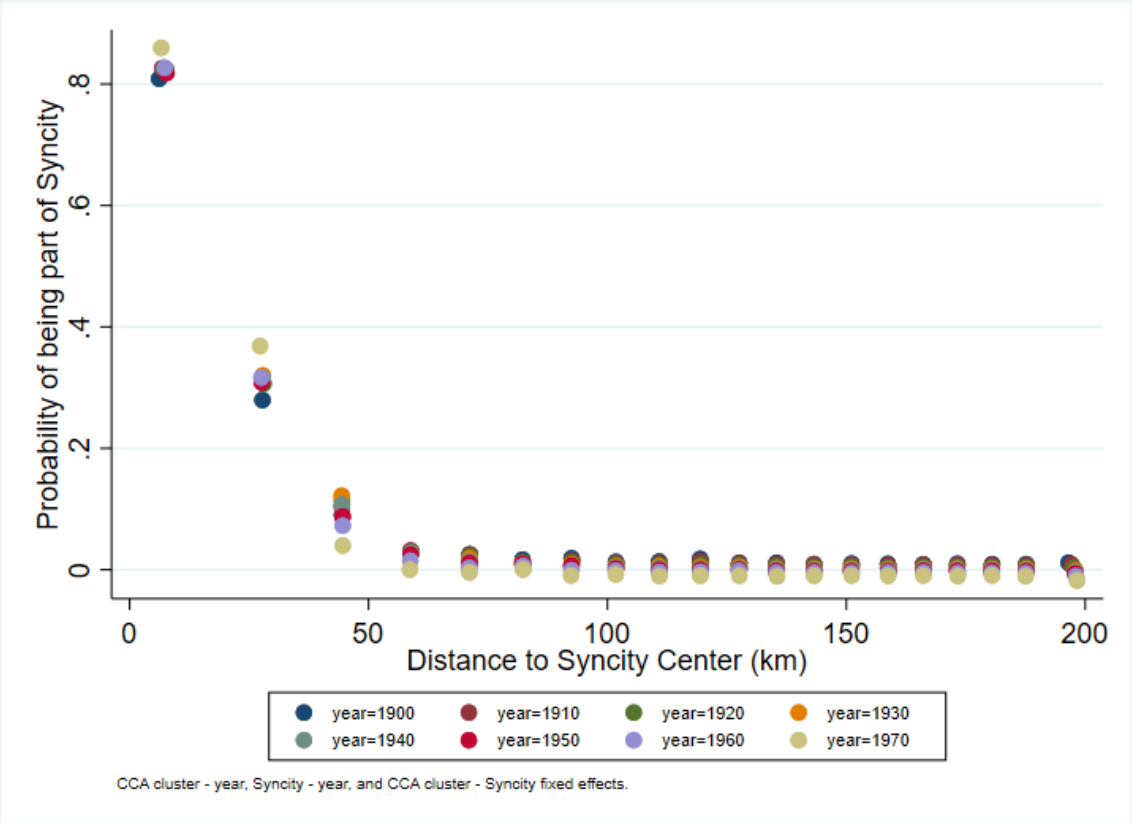


Figure 3.8: Probability of Being Part of a Syncity

Figure 3.8 shows that the outcomes from 1900 to 1970 are similar. We can see that when the distance from the CCA cluster to the center of Syncity increases, the probability of a CCA cluster taking part in the Syncity will decrease sharply. If it is nearer to the core of Syncity, the reduction will be sharper. Still, if the distance from CCA clusters to the center is far from the core of Syncity, the probability reduction will be minimal or nearly zero. The distance from CCA clusters to the center will play an essential role in Syncity for determining whether or not the CCA cluster will be compatible. Clusters close to the center will be more likely to get into Syncity, but those farther to the center will be less likely to get into Syncity.

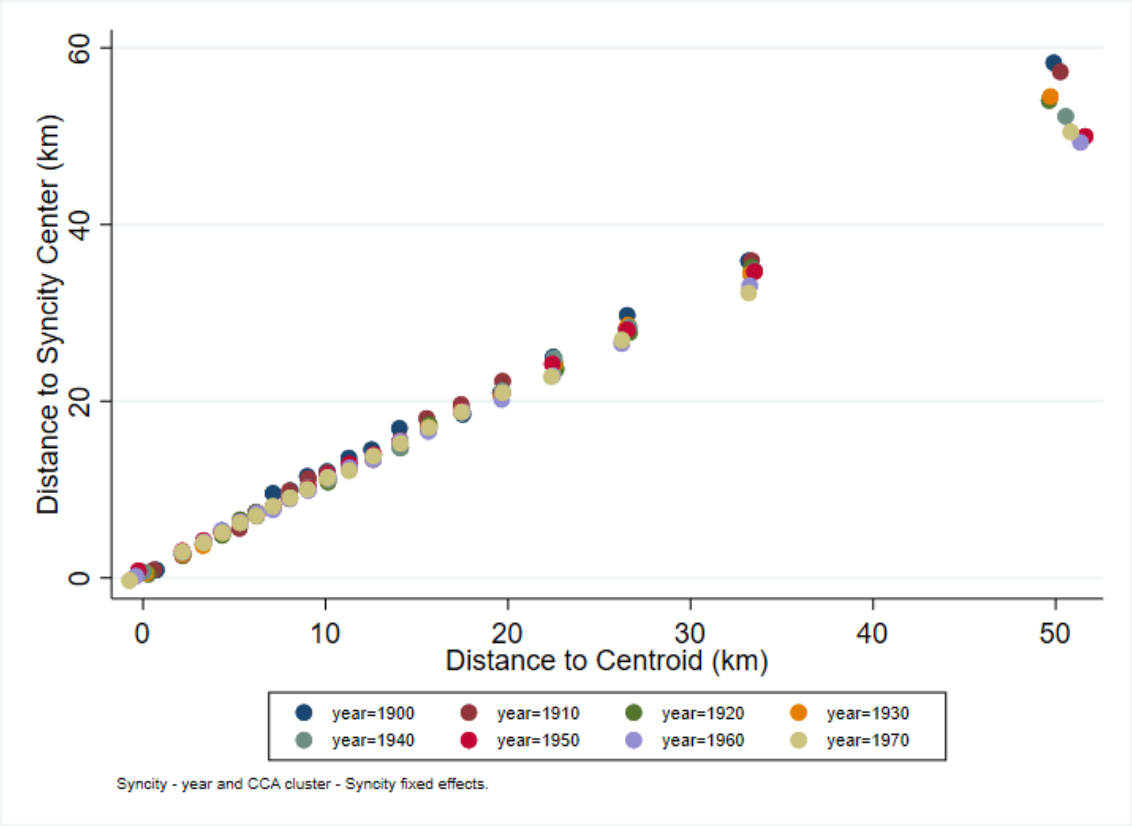


Figure 3.9: Location of the Syncity Core Center

Year	Coefficient (β)
1900	1.102[a]
	(0.034)
1910	1.115[a]
	(0.022)
1920	1.063[a]
	(0.011)
1930	1.066[a]
	(0.010)
1940	1.016[a]
	(0.006)
1950	0.965[a]
	(0.010)
1960	0.960[a]
	(0.010)
1970	0.991[a]
	(0.011)

Note: [a] denotes significance at the 1 percent level.

Table 3.2: Coefficients: Distance to Centroid vs. Distance to Center

Figure 3.9 shows the distance relationship between CCA clusters and the Syncity's center core and centroid, respectively. We can tell that CCA clusters far from the centroid of a Syncity are also far from the center core. Additionally, when the distance from CCA clusters to the Syncity centroid equals the distance to the center core (when the coefficient β equals one), there is a high probability that the centroid point overlaps with the center core point, indicating that the internal transportation costs of the Syncity are minimized at this time. The results in Table 3.2 show that, over time, the coefficient gradually reaches one, indicating that Syncities have continuously become more organized by minimizing internal transportation distance.

3.2.5 Evolution of the Urban System in the United States

Zipf's law is a well-known principle in urban geography. Scholars often use this law to show that cities are complex systems because their size distribution does not follow simple statistical patterns but follows a more complex rank-size rule [40].

$$\text{rank}(N) = \left(\frac{N_{\text{max}}}{N}\right)^{z} \tag{3.3}$$

According to the rank-size rule in Equation 3.3, N is the population size of a city, and N_{max} is the population of the biggest city in the urban system. To fully follow the rank-size rule, the z is equal to 1. In other words, in any urban system in the world, if cities are listed using the population to show the size, the second largest city will have half the population of the biggest, the third largest city will have one-third of the population, etc.

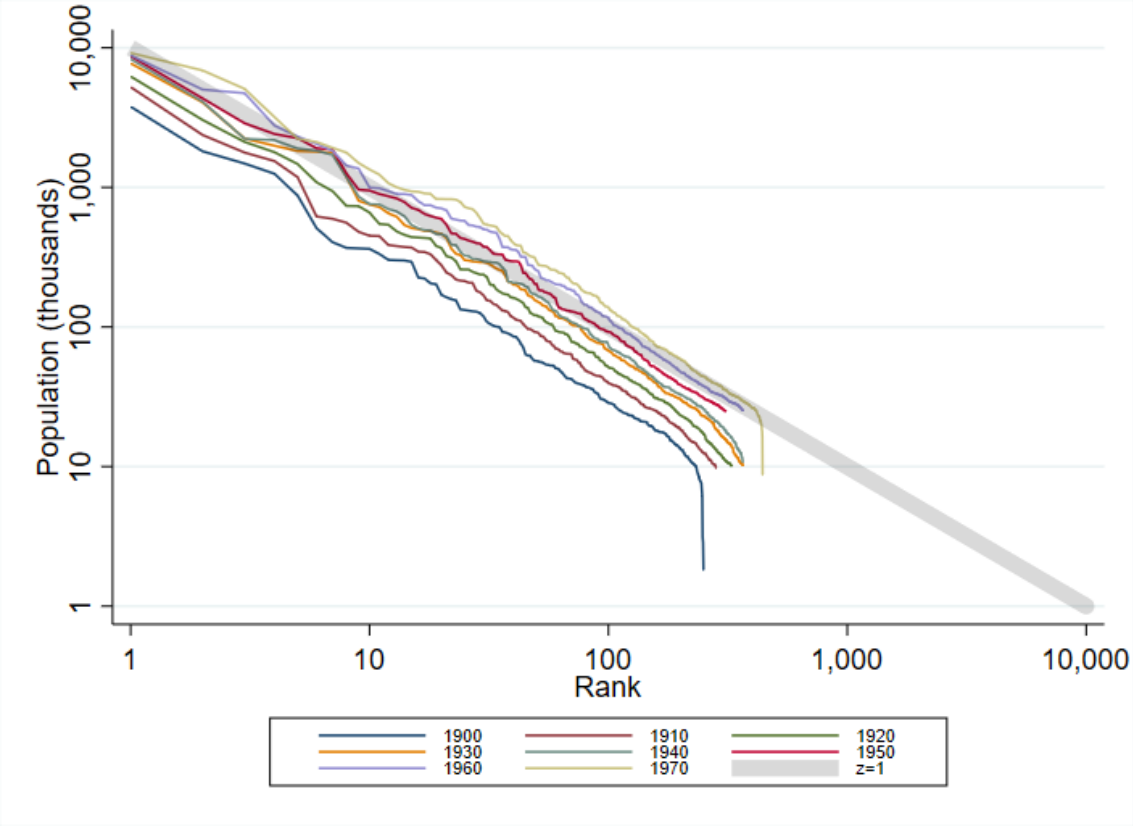


Figure 3.10: Zipf's Illustration of the Rank-size Rule for US City Population: 1900-1970

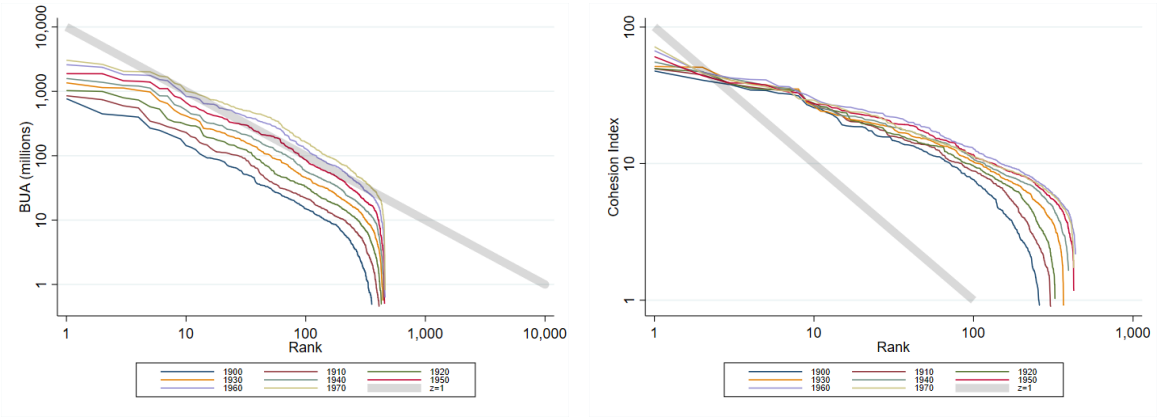
Note: The thick gray line indicates the rank-size rule with z = 1.

Inspired by the theory of Zipf's law, which suggests a relationship between a city's rank and its size, we decided to examine the relationship between a certain urban characteristic and its rank. Specifically, we tested the aggregated population in the Syncities we delineated, the built-up areas in these Syncities, and the cohesion index in these Syncities. We want to find how they behave or the patterns they have to understand these urban characteristics and also understand the urban system as a whole.

The curve for the population in Figure 3.10 shows that the rank of the population is inversely proportional to the population for each of the decades between 1900 and 1970. It means that when the rank of the population increases, the population size decreases proportionally. However, there is a decay at large ranks, which means a visible systematic deviation in terms of small cities. This result is very similar to the findings in the past literature. Bettencourt [40] tested the rank-size rule for populations in U.S. administrative

cities from 1790 to 1930 and concluded that "Zipf's law is an approximate description of the city size distribution, with some visible systematic deviations in terms of smaller city sizes and slopes." This may be because the data for very small cities is incomplete and sometimes hard to gather. Another reason may be that the tiny cities are not mature enough to follow the system patterns compared to others, which are in a relatively more mature stage.

The result above helps us understand the hierarchical structure of urban systems by studying the relationship between rank and size. Most existing research has defined size by population [40]. However, the spatial horizontal expansion of a city is also an important and direct indicator of its size. Therefore, we not only tested population but also innovatively tested for urban size, which is one indicator of urban spatial structure. Moreover, we are also interested in the relationship between rank and urban shape, which is another indicator of urban spatial structure. As a result, we can contribute to previous literature by capturing spatial dimensions beyond population, which can help us better link to the practice regarding transportation and land use.



(a) BUA (b) Cohesion Index

Figure 3.11: Zipf's Illustration of the Rank-size Rule for US Urban Spatial Structure: 1900-1970

Note: The thick gray line indicates the rank-size rule with z = 1.

According to Figure 3.11(a), the curves have a similar pattern as the curves of the population. Different from the population, the curves for built-up areas have outliers at the beginning. All in all, for cities that are not extremely large or very small, their BUA rank is inversely proportional to their BUA size. Moreover, small cities as outliers may have

similar reasons as I talked about above. Extreme large cities as outliers may be because they are more likely to build vertically but not expand horizontally. This will cause the changes in BUA ranks of larger cities to be smaller than for other cities. In addition, the main slope of the BUA curve is similar to the one for population; it is because the built-up areas can directly reflect the size of the population in a city.

Regarding the cohesion index, which represents urban shape, Figure 3.11(b) shows that the relationship between the cohesion index and its rank follows a certain proportion for most cities. However, this proportion is different from the results of BUA and population. This may be because the cohesion index deals with a different dimension from BUA. Specifically, the cohesion index is the one that reflects the spatial dispersion level in cities, but BUA reflects the land occupation in cities. They discuss different perspectives leading to different slopes for curves. This result shows motivation for future studies to investigate why the cohesion index follows a different pattern compared to the urban size.

The hierarchical structure of city sizes follows Zipf's law. According to the test above, the city's "size" can be population or geographic size, such as BUA. In other words, we can understand that an urban system emerges when the "size" follows Zipf's law. This hypothesis allows us to study the dynamic development trajectory of the urban system in the United States.

However, the Syncity-level urban size or population is not applicable when exploring the dynamic development trajectory of the urban system. In terms of the urban size, BUA for each Syncity, in our case, is derived from the CCA clusters that have already been cleaned and grouped for each Syncity. These pre-processed CCA clusters constituting each Syncity cannot fully display how randomly individual clusters gradually grow and merge to form a complete urban system. These pre-processed CCA clusters have always appeared as a "city" cluster, the Syncity. Also, there is a similar problem for urban populations. They have already been aggregated to represent the size of a complete Syncity but cannot show the dynamic process of how people gather together to form a Syncity. Therefore, the previous application of the Syncity-level rank-size rule shows a similar trend from 1900 to 1970 when following Zipf's law, which does not effectively reveal any dynamic development path of the urban system over the years.

This problem can be solved if we do not focus on the already formed Syncities but look at the random individual clusters and capture the merging path of these clusters. Although we do not have population data for the random individual clusters, we have an excellent alternative indicator, BUA, to show the size of the random individual clusters. Specifically, we do not restrict the analysis of the rank-size rule within the boundaries of Syncities but leave all the original CCA clusters within the US. Then, we applied the BUA

of these random clusters to examine the evolution of the urban system by testing how the size of these random clusters has followed Zipf's law over the years. Following the form of the rank-size rule, we applied the equation to test the relationship of rank (N) – the rank of BUA for each CCA cluster in the US – and N – the size of BUA of each CCA cluster in the US. We obtained the equation by using the logarithm of the rank-size rule form:

$$\log_{10}\left(\text{rank}\left(N\right)_a\right) = z \times \left(\log_{10}\left(N_{\max}\right) - \log_{10}\left(N\right)_a\right) \tag{3.4}$$

where a denotes each CCA cluster, and $N_{\max}$ is the BUA value of the largest CCA cluster.

Based on the Equation 3.4, we calculated the coefficient z for 1900, 1940, 1970, and 2010. According to Zipf's law, coefficient $z = 1$ represents an ideal urban system possessing an optimal hierarchical structure. Because some factors will lead to an actual urban system deviation from the perfect one, we expect the actual coefficient z to be as close to an ideal "1" as possible. In other words, the closer the coefficient z is to 1, the closer the hierarchical structure of an urban system is to the normativity.

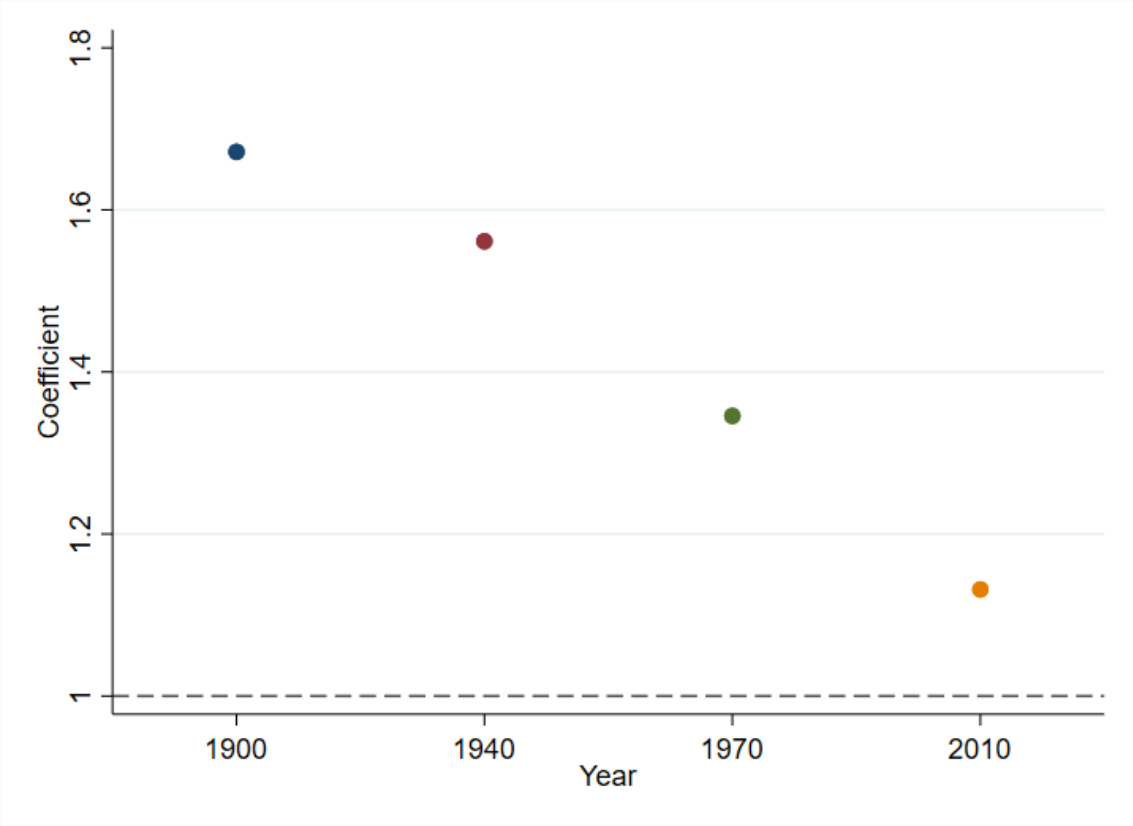


Figure 3.12: z Value for Rank-size Rule

Figure 3.12 shows the value of coefficient z for each cross-sectional year. The coefficient z gradually approaches one over time, decreasing from 1.67 in 1900 to 1.13 in 2010, indicating that the American urban system is progressively achieving a relatively ideal hierarchical structure. Additionally, we can observe that the development speed of the American urban system experienced a jump after 1940 and maintained this higher development speed until the present (2010). We reasonably suspect that the second wave of the Great Migration, which started in 1940, accelerated the development of the urban system in the United States.

3.3 Discussion

In this chapter, we studied the urban spatial development in the United States between 1900 and 1970 using reconstructed urban areas and a new spatial dataset. We applied the City Clustering Algorithm (CCA) to create historical urban areas in America, which allowed us to generate new data that helped us understand urban spatial dynamics and the evolution of the urban system over time.

This new urban spatial data not only adds to our knowledge of the historical urban structure in the U.S. before 1970 but also helps avoid the subjectivity of using administrative boundaries. As a result, we can reveal patterns and trends in urban development more accurately. Compared to analyses based on administrative boundaries, our results provide a closer representation of reality. Here are some insights from our findings:

Our analysis of cities, regardless of their location or when they emerged, showed a strong consistency in development patterns. While not all cities develop similarly or at the same pace, we saw distinct patterns across different regions, such as inland versus coastal cities. These findings highlight the orderly yet complex trajectory of American urban development.

From a broader perspective, between 1900 and 1970, American cities followed highly regular trends in expansion and shape evolution. New CCA clusters each year tended to fill gaps between existing clusters, while some expanded outward.

By applying Zipf's law, we observed that the American urban system continued to develop towards a more defined hierarchical structure. According to Zipf's law, mature urban systems have a predictable size distribution, where the rank of a city is inversely proportional to its population size. Our findings show that the urban system advanced significantly, aligning more with the expectations of Zipf's law, though it has not yet reached full maturity.

We also found that the distribution of clusters within cities follows a specific pattern. Allen's equilibrium conditions show that clusters closer to the city center are more closely tied to the city. As mentioned earlier, urban growth tends to begin at the core, where clusters are more densely packed due to higher socio-economic activity and infrastructure. This pattern suggests that city centers are highly attractive because of better access to services and economic opportunities, leading to structured growth from the center outward.

Examining equilibrium conditions also shows that cities became more organized over time by reducing internal transportation costs and improving efficiency. Cities worked to organize their spatial layouts, which led to better-connected clusters and more efficient urban structures. This process was driven by technological advancements, economic factors,

and demographic changes. Cities aimed to minimize the distance between residential, commercial, and industrial areas to improve accessibility and reduce costs. This reorganization shows cities adapt to balance growth and efficiency and minimize costs, highlighting their dynamic nature.

Our findings offer urban planners and policymakers a historical perspective to understand better how American cities have developed. This knowledge can be used to improve urban planning to support SDG 11: making cities inclusive, safe, resilient, and sustainable [62]. Our research provides valuable tools for planners and policymakers. The methods used to construct urban boundaries can be applied to historical cities as well as to study the development paths of modern cities. Urban planning departments can use similar clustering methods to analyze urban spatial dynamics. Instead of relying on administrative boundaries, planners can adjust policies based on real development needs. For example, clustering data can identify high-potential areas for urban expansion, helping planners align construction with actual needs. Planners can also use this data to adjust transportation and infrastructure, improve traffic flow, and reduce congestion, making urban areas more efficient and reducing carbon emissions.

Additionally, our findings can help predict future urban development paths. By analyzing the trends in size, compactness, and cluster evolution, urban planners can more comprehensively understand development patterns and accurately predict future trends. This will help them create better land-use plans and urban structures, leading to more sustainable cities.

Chapter 4

Demographic Shocks and Urban Spatial Change: Lessons from the Great Migration

4.1 Introduction

The second wave of the Great Migration (1940-1970) was a pivotal demographic event that reshaped the urban landscapes of many U.S. cities. African Americans, seeking better opportunities, migrated from the rural South to urban centers in the North and West, transforming these cities' racial, social, and economic dynamics. Specifically, in the white flight, which happened during the second wave of the Great Migration, a number of the white population chose to move from the city center to the suburban, directly changing the urban population structure and urban community composition [39]. To deal with the changing racial composition, policies, such as the zoning law and redlining policy, were used to enhance racial segregation [3, 76]. These social and political changes increased the spread of cities horizontally and changed the composition of who lives where. So, we expect to see how society and policy shape urban spatial structure after the dramatic influx of the Black population in urban areas.

This shift is not only a significant part of U.S. history but also a lens through which we can explore how large-scale migration impacts urban spatial structures. These changes in urban form, particularly in size and spatial distribution, have lasting effects on economic activity, housing markets, and infrastructure development. Understanding the spatial consequences of this migration wave is essential for grasping how demographic shifts reshape

cities.

This chapter addresses the question: What was the impact of the Second Wave of the Great Migration on urban spatial structure? Specifically, we examine the impact of these demographic changes on non-Southern cities, investigating whether increased racial diversity led to horizontal urban expansion and changes in the spatial cohesion of city structures. It should be noted that the Great Migration is just used as an exogenous shock to population size for us to study the effect on cities when they receive a big amount of population in a short period of time. Using an instrumental variables (IV) approach, we aim to uncover causal relationships between the changing population and urban spatial structure.

To investigate the relationship between Black migration and urban spatial changes, we employ a two-stage empirical approach using instrumental variables (IV) regression to address potential endogeneity concerns. The key challenge in estimating the causal effect of demographic shifts on urban form is that migration patterns may both influence and be influenced by pre-existing urban characteristics. To overcome this, we construct an instrumental variable based on predicted Black migration flows from 1940 to 1970, leveraging historical migration patterns prior to the second wave of the Great Migration. By relying on exogenous factors such as cultural ties and pre-established migration networks, this IV approach isolates the demographic impact of the migration shock on cities' physical expansion and spatial distribution. We then apply this instrument to estimate the causal effect of changes in the Black population on two key urban outcomes: the total built-up area (BUA) as a measure of urban size and the cohesion index, which captures the spatial dispersion of urban clusters. This methodology allows us to disentangle the independent effect of racial composition changes on urban form, providing more credible estimates than simple OLS models that may suffer from bias due to reverse causality.

A key contribution to understanding the impact of the Great Migration on urban environments is provided by Derenoncourt [16] in her study "Can You Move to Opportunity? Evidence from the Great Migration." Derenoncourt's work demonstrates that migration led to a significant reduction in the gains from growing up in northern cities for Black families, primarily due to persistent segregation, increased crime, and changes in local government policies. By interacting pre-1940 Black migrants' location choices with predicted southern county out-migration, she shows that the response of destination locations was critical in shaping the economic opportunities for migrants and their descendants. This nuanced perspective of "moving to opportunity" highlights that, despite the promise of better economic conditions in the North, the local reactions to the influx of Black migrants had enduring negative effects on upward mobility. The insights from Derenoncourt's research are highly relevant to our study, as they underscore the importance of examining

the broader consequences of demographic changes on urban spatial dynamics and socioeconomic outcomes.

Prior research on the Great Migration has largely focused on its social and economic consequences, with an emphasis on residential segregation [36, 37] and racialized public policies [3, 61]. While these studies highlight the negative consequences of discriminatory housing practices and segregation, they leave unexplored the direct effects of migration on the physical size and shape of cities. Although residential segregation is a key factor in urban spatial outcomes, the literature has yet to fully integrate an analysis of how demographic changes driven by the Great Migration influenced urban growth and the dispersal of urban clusters.

The broader social and political context of the Great Migration also played a significant role in shaping urban spaces. For instance, Sugrue [58] highlights how post-war economic restructuring and discriminatory housing policies led to stark racial divisions in urban environments, exacerbating the physical separation between Black and white communities. This separation influenced the physical growth of cities as white residents increasingly moved to suburban areas, leading to urban sprawl and a decline in urban density. Similarly, Massey and Denton [13] argue that institutionalized racism and economic barriers confined Black migrants to specific neighborhoods, affecting urban development patterns and leading to spatial segregation that persists today.

Another critical aspect of the antecedent literature involves the relationship between migration and urban infrastructure. As noted by Logan and Molotch [31], the influx of African American migrants led to shifts in urban priorities, including changes in public infrastructure and service provision. These shifts often resulted in underinvestment in predominantly Black neighborhoods, which in turn influenced urban form by restricting the growth and quality of these areas. This lack of equitable investment further entrenched spatial inequalities, with long-term implications for urban cohesion and expansion.

Moreover, Boustan [39] provides evidence that the economic incentives driving migration also had spatial implications for cities. The clustering of African American communities in specific urban areas, often driven by employment opportunities, created pressures on housing markets and influenced patterns of urban expansion. These pressures contributed to a patchwork of urban growth, where some areas expanded rapidly while others were neglected, leading to a fragmented urban landscape.

This chapter makes three main contributions to the literature. First, it extends the analysis of the Great Migration by focusing not only on social and economic impacts but also on spatial outcomes, filling a gap in our understanding of how demographic shocks reshape urban environments. While much of the existing literature has centered on resi-

dential segregation and public policy implications, this chapter provides direct empirical evidence on how migration-driven demographic changes influenced urban size and cohesion. By expanding the scope to spatial outcomes, we offer new insights into the broader ramifications of the Great Migration on urban development.

Second, this chapter uses newly defined spatial datasets that capture historical changes in urban form, allowing for a more nuanced examination of how cities grew and evolved during this period. These datasets provide a level of precision that is not available in most prior studies, which often rely on less granular data. By utilizing these newly constructed datasets, we contribute to the literature by providing a more precise and more detailed picture of urban spatial dynamics and how they were affected by demographic shifts. This data-driven approach allows us to pinpoint the specific aspects of urban growth influenced by the influx of Black migrants.

Third, the use of an IV approach to address endogeneity concerns marks a significant methodological contribution. Previous studies on the Great Migration have often relied on correlational analyses, which can suffer from biases due to reverse causality or omitted variable issues. By employing an IV approach, we provide robust causal estimates of the effect of Black migration on urban size and cohesion. This methodological rigor strengthens the validity of our findings and sets a new standard for future research on the impact of demographic shocks on urban environments. Our approach contributes to a more reliable understanding of the long-term urban effects of demographic changes and provides a robust framework for analyzing similar migration-induced transformations in other contexts.

The chapter proceeds as follows. Section 4.2 outlines the empirical strategy, detailing the construction of the instrumental variable and the regression model. Section 4.3 presents the results, showing the effects of migration on urban size and cohesion. Finally, Section 4.4 discusses the broader implications of these findings, particularly how they relate to ongoing urban planning and policy debates.

4.2 Empirical Strategy

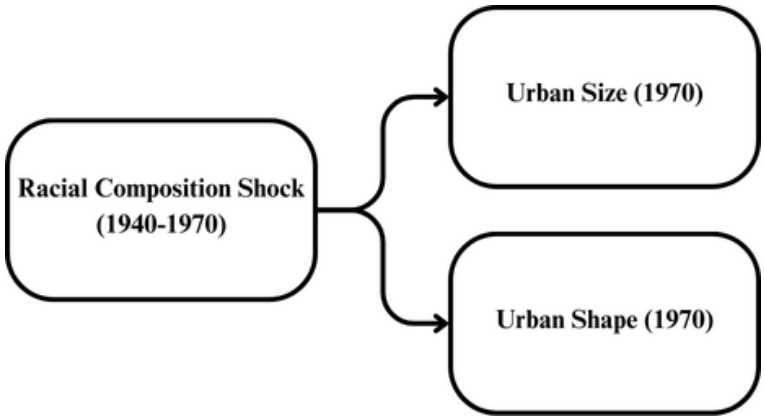


Figure 4.1: The Effect of the Racial Composition Shock on Urban Spatial Structure

In this section, we outline the empirical strategy used to identify the causal effects of the second wave of the Great Migration on urban spatial structures. As shown in Figure 4.1, we focus on the effect of the racial composition shock on urban size, which is represented by the built-up area (BUA, $area_{i,1970}$), and on urban shape, which is represented by the cohesion index ($\mathrm{S}_{i,1970}$). Moreover, the racial composition shock is measured by the changes in the share of Blacks ($\Delta Bpop_{i,1940-1970}$). The equations for Ordinary Least Squares (OLS) regression to examine the simple linear relationships between the racial composition shock and urban spatial structures are given below:

$$ln(area_{i,1970}) = a \times ln(\Delta Bpop_{i,1940-1970}) + \epsilon_i \tag{4.1}$$

$$ln(\mathrm{S}_{i,1970}) = b \times ln(\Delta Bpop_{i,1940-1970}) + \epsilon_i \tag{4.2}$$

where i represents each county-Syncity union.

However, the key challenge in this analysis is to address potential endogeneity concerns. Specifically, migration patterns are likely influenced by characteristics of the destination cities while simultaneously shaping these cities, thus making it difficult to disentangle cause and effect using ordinary least squares (OLS) regression alone. To mitigate these endogeneity issues, we employ an instrumental variables (IV) approach, leveraging historical

migration patterns to create an exogenous variation in Black migration flows. This allows us to isolate the impact of demographic changes on urban growth while reducing biases from reverse causality or omitted variables.

Our empirical approach has two main components: the construction of an instrumental variable (IV) and the estimation of effects on urban outcomes.

4.2.1 Instrumental Variable Construction: Predicted Black Migration

The instrumental variable used in this study is based on predicted Black migration flows from southern states to non-Southern urban centers during the period 1940-1970. The construction of this IV follows the method introduced in Derenoncourt's paper [16]. The IV is designed to capture exogenous variations in migration flows that are not correlated with contemporary urban dynamics. This is done by predicting migration using pre-1940 settlement patterns driven by cultural ties, community networks, and the historical distribution of Black populations.

The validity of this IV rests on two main assumptions:

1. ***Relevance.*** The IV must be correlated with the endogenous variable, which in this case is the actual change in the share of Black residents in non-southern urban areas during the migration period. Based on historical trends, this assumption is justified as the predicted Black migration rate is highly correlated with actual migration patterns due to persistent cultural and community factors.

2. ***Exclusion Restriction.*** The IV must not be correlated with the error term of the outcome equations. This implies that the predicted Black migration should influence urban spatial structures only through its impact on the share of Black residents, not directly through other unobserved factors. Although this requirement is difficult to test directly, it is reasonable to assume that the predicted migration flows—based on historical trends before the second wave—are exogenous to contemporaneous shocks affecting urban outcomes.

Following Derenoncourt [16], the predicted Black migration rate for each southern county was calculated by dividing the number of out-migrants from each southern county between 1935 and 1940 by the total population of that county in 1935. Subsequently, the predicted Black out-migration for each southern county from 1940 to 1970 was estimated

using economic predictors identified by Boustan [39], including factors like agricultural mechanization and changes in employment opportunities. By interacting these predicted migration rates with pre-1940 settlement patterns, we constructed an instrumental variable that measures the expected influx of migrants to each urban area during the second wave of the Great Migration.

Different from Derenoncourt [16], we used city-level observations instead of the commuting zone level that she used to further aggregate up to the county-Syncity union level.

The instrumental variable is thus constructed as follows:

$$\hat{M}_{i,1940-1970} = \sum_{j \in S} \sum_{c \in i} \omega_{j,c}^{1935-1940} \cdot \hat{m}_j^{1940-1970} \tag{4.3}$$

where i denotes each county-Syncity union, j denotes southern counties, and $(\omega_{jc}^{1935-1940})$ is the pre-1940 migration rate from county j to city c. The term $\hat{m}_j^{1940-1970}$ represents the predicted out-migration from southern county j during the study period.

4.2.2 Estimating Equations

Using the predicted Black migration ($\hat{M}_{i,1940-1970}$) as an instrumental variable, we estimate the impact of demographic changes on two key urban outcomes: the total built-up area (BUA) as a measure of urban size ($area_{i,1970}$), and a cohesion index that captures the spatial dispersion and connectivity of urban clusters ($\mathrm{S}_{i,1970}$).

The estimation involves a two-stage least squares (2SLS) approach. In the first stage, we estimate the relationship between the instrumental variable and the endogenous regressor—the change in the share of Black residents between 1940 and 1970:

$$ln(\Delta Bpop_{i,1940-1970}) = \alpha_T^Z \times ln(\hat{M}_{i,1940-1970}) + \mathbb{X}_i + \epsilon_i \tag{4.4}$$

where $\mathbb{X}_i$ is a set of control variables, including the share of the manufacturing labor force, educational upward mobility, the share of 1940 Black migrants, and regional fixed effects.

In the second stage, we use the predicted values from the first stage to estimate the effect of changes in the share of Black residents on urban outcomes:

$$ln(area_{i,1970}) = \alpha_M^T \times ln(\widehat{\Delta Bpop}_{i,1940-1970}) + \mathbb{X}_i + \epsilon_i \tag{4.5}$$

$$ln(\mathrm{S}_{i,1970}) = \beta_M^T \times ln(\widehat{\Delta Bpop}_{i,1940-1970}) + \mathbb{X}_i + \epsilon_i \tag{4.6}$$

In Equation 4.5 and 4.6, α_M^T and β_M^T represent the effect of the changes in the share of Blacks on urban size and urban shape, respectively. Using the IV approach, we aim to provide causal estimates of the demographic impacts on urban spatial changes, accounting for potential biases that could arise if migration patterns were endogenously determined by city characteristics.

This empirical strategy allows us to answer our central research question: What was the impact of the Second Wave of the Great Migration on urban spatial structure? By using an instrumental variable approach grounded in historical migration patterns, we provide credible causal estimates that contribute to our understanding of how demographic shifts reshape urban environments.

4.3 Results

In this section, we present the results of our analysis of the effects of the second wave of the Great Migration on urban spatial structures. Using the methods discussed in the previous sections, we estimated the impact of changes in the share of Black residents on both urban size and urban cohesion.

Specifically, we obtained regression results that are reported in Tables 4.1 and 4.2, illustrating the effects of the migration-induced changes in racial composition on the built-up urban area (BUA) and urban shape, respectively.

Column (1) displays the results without control variables. Column (2) reports the results of controlling the characteristics and region fixed effects. Column (3) shows the OLS results as a baseline reference. The ordinary least squares (OLS) results also indicate a positive relationship, though the coefficient is smaller than that obtained using the IV approach, highlighting the endogeneity issues present in the OLS estimates. Moreover, we cluster at the County-Syncity Union level, which is our merged spatial unit, to make sure the standard errors will not be underestimated.

According to Table 4.1, the second-stage instrumental variables (IV) regression results show a significant positive relationship between the change in the share of Black residents during the Great Migration and urban size. Our preferred estimate, which includes all control variables (column 2), suggests that a 1% increase in the share of Black residents between 1940 and 1970 led to a 1.644% increase in the built-up urban area by 1970. Given

that cities in 1970 were, on average, 354 square kilometers in size, this implies an expansion of approximately 5.82 square kilometers in response to the migration shock.

During this period, the share of the Black population increased by an average of 11%, which corresponds to roughly 58,800 additional Black residents in an average city. This implies that for every 100 new Black residents, there was an expansion of around 1 square kilometer. While this figure suggests a relatively high population density, it is important to note that not all new migrants settled in newly expanded areas. Many likely moved into existing neighborhoods, filling vacant properties or contributing to vertical growth through higher-density developments.

Table 4.1: [Dependent: Ln BUA 1970]

	(1)	(2)	(3)
	(IV)Second Stage		**OLS**
Ln Changes in Share of Blacks 1940-1970 (α_M^T)	1.289[a]	1.644[a]	.638[a]
	(.164)	(.234)	(.091)
	(IV)First Stage		
Ln Predicted Black Migration 1940-1970	.300[a]	.240[a]	
	(.030)	(.032)	
Observations	115	115	115
Cragg-Donald Wald F statistic	75.048	33.640	
Control		X	

Note: [a], [b], and [c] denote significance at the 1 percent, 5 percent, and 10 percent levels, respectively.

Table 4.2 presents the results for urban shape, as measured by the cohesion index, which captures the spatial dispersion and connectivity of urban areas. The instrumental variables (IV) regression results indicate a significant positive relationship, showing that each 1% increase in the share of Black residents led to a 0.44% increase in the cohesion index. Given that the average cohesion index was approximately 10 kilometers, this translates to a rise of around 45 meters in the average distance between built clusters within a city. Using the same baseline population numbers as above implies an increase of about 8 meters in spatial dispersion per 100 new migrants.

This finding suggests that migration-driven population changes contributed to a more dispersed urban form, as evidenced by the growing distance between urban clusters. Such

increased dispersion reflects the pressures placed on existing infrastructure and the need for expanding spatial footprints to accommodate the incoming population, resulting in a less cohesive urban landscape. This pattern also aligns with historical accounts of the Great Migration, where exclusionary zoning practices and residential segregation often pushed new arrivals to the urban periphery, thereby reshaping the spatial structure of cities.

Table 4.2: [Dependent: Ln Cohesion Index 1970]

	(1)	(2)	(3)
	(IV)Second Stage		**OLS**
Ln Changes in Share of Blacks 1940-1970 (β_M^T)	.425[a]	.440[b]	.120[b]
	(.130)	(.219)	(.060)
	(IV)First Stage		
Ln Predicted Black Migration 1940-1970	.300[a]	.240[a]	
	(.030)	(.032)	
Observations	115	115	115
Cragg-Donald Wald F statistic	75.048	33.640	
Control		X	

Note: [a], [b], and [c] denote significance at the 1 percent, 5 percent, and 10 percent levels, respectively.

The findings indicate that migration-induced increases in the share of Black residents had a substantial impact on urban form. The horizontal expansion of cities (increased urban size) likely responded to the increased demand for housing, infrastructure, and services driven by a growing population. Meanwhile, the increase in the cohesion index suggests that urban areas became more spatially fragmented as new residential or commercial clusters emerged at the city periphery or existing clusters expanded.

These patterns align with the historical context of the Great Migration, during which many African Americans moved to urban centers in search of economic opportunities. As a result, non-Southern cities underwent a significant transformation, both in terms of expanding their physical footprint and in the evolving spatial distribution of their neighborhoods and commercial centers.

4.4 Discussion

We examined how migration-induced changes in the share of Black residents during the second wave of the Great Migration influenced urban spatial structures in non-Southern cities. Specifically, we employed an instrumental variables (IV) approach to address endogeneity concerns, using predicted Black migration from 1940 to 1970 as an instrument to isolate exogenous variation in demographic changes. This methodology provides a way to capture the causal impact of these demographic shocks on urban size and spatial dispersion, overcoming biases that arise from simultaneous causation between population growth and urban development.

The use of an instrumental variable is crucial in addressing potential endogeneity issues. The changes in the share of Black residents may not only influence urban form but also be influenced by it. For example, as cities grow and economic opportunities expand, more individuals might be attracted to these areas, further complicating the relationship between migration and urban expansion. By leveraging historical migration patterns to create an instrument, we ensure that our estimated effects reflect the causal impact of the Great Migration on urban outcomes rather than correlations driven by other factors. This approach builds upon and enhances the robustness of earlier empirical strategies, providing a clearer picture of the urban transformations induced by this significant migration wave.

Our findings contribute to several ongoing debates in the literature on the Great Migration and its implications for urban development. Prior work, such as Derenoncourt [16], highlights that the influx of Black migrants to Northern cities during the Great Migration led to significant shifts in urban environments, transforming previously high-opportunity locations into "opportunity deserts" for Black families due to rising segregation and declining economic opportunities. Our results are consistent with this narrative. We find that the migration-induced growth in the share of Black residents contributed to the horizontal expansion of urban areas and increased spatial fragmentation, as evidenced by a rise in the cohesion index.

This expansion and increased dispersion can be understood in light of the segregation dynamics emphasized by Gotham [36] and others, where discriminatory practices such as racially restrictive covenants and exclusionary zoning laws drove racial segregation in Northern cities. Our findings show that the increased population pressure did not necessarily lead to greater density but led to a more dispersed urban form, which may have responded to the exclusionary pressures faced by African American migrants. This aligns with the work of Sahn [3], which suggests that increased racial diversity often resulted in restrictive zoning that limited multifamily housing, further contributing to the sprawling nature of cities during this period.

Moreover, our findings also relate to the work of Reny and Newman [61], who argue that the increased presence of African American migrants led to heightened racial threats that influenced public policy, such as support for Proposition 14 to protect racial discrimination in housing. The resulting fragmented urban form, reflected in our cohesion index, may, therefore, be partially attributed to such racially motivated policy responses, which hindered integrated urban growth and fostered spatial separation.

The evidence suggests that the racial composition shock during the Great Migration had significant effects on both the physical expansion and spatial arrangement of Northern cities. These results imply that urban development during periods of large demographic change can be highly path-dependent, shaped not only by economic opportunities but also by prevailing social attitudes and policies that influence who has access to different parts of a city. The fragmented urban growth observed in our study illustrates how discriminatory practices and policies can shape the physical structure of cities in ways that have long-lasting effects on social and economic outcomes.

Chapter 5

Urban Spatial Structure and Economic and Environmental Outcomes

5.1 Introduction

Urbanization is a defining feature of modern society, transforming how people live, work, and interact. However, as cities grow and evolve, their spatial structure often brings unintended consequences for economic outcomes and environmental quality. This dual impact—where expanding urban spaces can lead to both economic growth and environmental degradation—is at the heart of policy debates today. Why do larger, more dispersed cities often experience worse air quality? And how can policymakers balance economic benefits with the environmental costs of urban sprawl? These questions make understanding urban spatial structure both an intriguing and critical issue for society.

This paper addresses the question: What are the economic and environmental consequences of urban spatial structure? Specifically, it examines whether urban sprawl, characterized by horizontal expansion and spatial dispersion, correlates with improved economic metrics such as total income while simultaneously contributing to worse environmental indicators, such as air quality. The study investigates this relationship across short- and long-term horizons, providing insights into the economic-environmental interactions that emerge from urban growth patterns.

To answer this question, we employ an econometric approach that utilizes an instrumental variable (IV) methodology to address potential endogeneity issues. This ensures that

the estimated effects of urban spatial structure on economic and environmental outcomes are causal rather than driven by reverse causality or omitted variable bias. The analysis draws on data from multiple periods to distinguish between short-term and long-term impacts, allowing for a dynamic understanding of how urban growth patterns influence city-level outcomes. By employing rigorous statistical techniques, we aim to uncover the causal mechanisms linking urban spatial characteristics to both economic and environmental performance.

Prior studies have explored the relationship between urban form and economic or environmental outcomes. Research by Harari [44] and Liu et al. [73] demonstrates the economic advantages of larger urban areas, while Sajjad Abdollahpour et al. [52] show how sprawling cities tend to experience higher levels of pollution. However, the existing literature often treats economic and environmental impacts as separate dimensions, failing to analyze how these forces interact over different temporal scales comprehensively. This gap in understanding—particularly concerning the balance between short-term and long-term effects—is where this paper aims to make a significant contribution.

Glaeser and Kahn [14] provide a foundational analysis of the economic benefits of urban density, arguing that compact urban forms can reduce commuting times and improve productivity. However, their work does not adequately address the environmental drawbacks that come with such density, particularly in terms of air quality and public health. In contrast, studies like those by Stone [8] emphasize the environmental costs of urban sprawl, particularly increased vehicle emissions, but often do not integrate these findings with economic outcomes, leading to an incomplete picture of both dimensions involved.

More recent work by González-Navarro and Turner [42] highlights the role of infrastructure investment in shaping urban spatial structures, suggesting that transportation networks can influence both the economic efficiency and environmental sustainability of cities. Their findings imply that infrastructure policies can exacerbate or mitigate the interactions between economic growth and environmental degradation, depending on how they shape urban form. This perspective underscores the importance of considering policy interventions when analyzing urban spatial structure, an aspect this paper incorporates by examining how different urban growth patterns respond to policy changes over time.

Furthermore, Duranton and Puga [21] explore the concept of urban agglomeration economies, where the benefits of economic concentration can lead to increased innovation and productivity. However, these benefits are often accompanied by negative externalities such as congestion and pollution. The current paper builds on this understanding by explicitly modeling how these agglomeration forces interact with environmental outcomes, providing a more nuanced view of the costs and benefits associated with urban expansion

and dispersion.

Moreover, researchers have focused on different spatial perspectives when discussing the impact of urban spatial structure on economic and environmental outcomes. One prominent perspective is the distinction between polycentric and monocentric urban structures. Zhang et al. [33] find that the presence of multiple city centers contributes significantly to economic development. Moreover, polycentric structures not only promote economic growth but also help narrow the urban-rural income gap and reduce regional inequalities [12]. Conversely, improvements in monocentric urban structures can enhance employment opportunities, as shown by Xiao et al. [67]. Additionally, Zeng et al. [17] demonstrate that polycentric urban structures are associated with higher water use efficiency, supporting sustainable ecological development.

City size is another dimension of urban spatial structure that has received considerable attention. Liu et al. [73] suggest that larger urban areas are associated with higher short-term GDP growth, pointing to urban expansion's economic benefits. However, these benefits come at an environmental cost, as larger urban extensions also tend to increase air pollution concentrations [52].

Compactness is the third key aspect of urban spatial structure. While higher urban compactness is linked to employment growth [67], it also brings certain drawbacks, such as lower wages, higher rents, and reduced GDP per capita [44, 73]. In terms of environmental performance, compact urban forms have been found to reduce air pollution concentrations, whereas dispersed urban forms exacerbate them [52].

Green space is another critical component of urban spatial structure with significant environmental implications. Studies by Diem et al. [45] and Y. Zhang et al. [75] show that green space plays a key role in mitigating the urban heat island (UHI) effect and reducing air pollution concentrations. These findings highlight the importance of incorporating green space into urban planning to improve environmental quality.

Regarding the contributions, first, this paper integrates economic and environmental perspectives to analyze the simultaneous impact of urban spatial structure on both dimensions. Our findings show that larger or more dispersed cities tend to be associated with higher total incomes but also with worse air quality, extending the findings identified by previous studies like Glaeser and Kahn [14] and Stone [8]. By highlighting both the economic benefits and environmental costs, we contribute to a more nuanced understanding of the interactions between economic benefits and environmental costs, building on and extending their work.

Second, it differentiates between short-term and long-term impacts, providing evidence that urban structure changes lead to varying outcomes over different time horizons. Specif-

ically, we find that the negative environmental impacts of urban expansion are more pronounced in the short term, whereas cities have some capacity to adapt and mitigate these effects over the long term, consistent with the dynamic changes discussed by Li [72]. This distinction between temporal effects adds depth to the existing literature by revealing how urban growth patterns can evolve to offset initial environmental detriments partially.

Third, the paper utilizes an instrumental variable (IV) approach to address potential endogeneity issues, ensuring that the observed relationships between urban structure, economic outcomes, and air quality are robust and not driven by reverse causality or omitted variable bias. The robustness of our results lends credibility to the findings of González-Navarro and Turner [42] regarding the role of infrastructure investment and policy interventions in shaping urban form. By using historical instruments and rigorous econometric techniques, we provide stronger causal evidence that supports and extends the arguments made by previous scholars regarding urban spatial policies and their effects.

The paper proceeds as follows. Section 5.2 outlines the empirical strategy, detailing the instrumental variable methodology used to identify causal relationships. Section 5.3 presents the results, distinguishing between short-term and long-term impacts. Section 5.4 concludes the findings and implications, focusing on how cities can balance economic growth with environmental sustainability.

5.2 Empirical Strategy

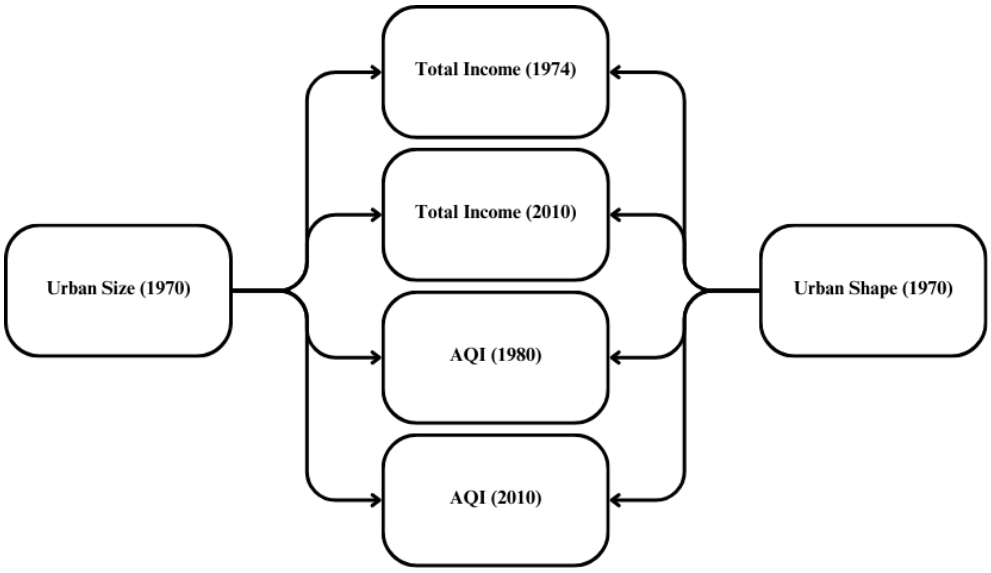


Figure 5.1: The Effect of the Urban Spatial Structure on Outcomes

In this section, we present our empirical strategy for assessing the impact of urban spatial structure on both economic and environmental performance. As shown in Figure 5.1, we focus on two key dimensions of urban spatial structure: size and shape. The urban shape is represented by the cohesion index ($S_{i,1970}$), and urban size is represented by the built-up area (BUA, $area_{i,1970}$). We explore how these dimensions influenced economic outcomes, measured by total income in 1974 and 2010, and environmental outcomes, measured by Air Quality Index (AQI) indicators in 1980 and 2010, capturing both short- and long-term effects.

To start, we employ Ordinary Least Squares (OLS) regression to examine the simple linear relationships between urban size, shape, and these outcomes. The equations for the OLS regressions are given below:

$$ln(Y_{i,t}) = c \times ln(area_{i,1970}) + \epsilon_i \tag{5.1}$$

$$ln(Y_{i,t}) = d \times ln(S_{i,1970}) + \epsilon_i \tag{5.2}$$

where i represents each county-Syncity union and $Y_{i,t}$ denotes the outcome variable for county-Syncity union i at time t.

However, using OLS alone may lead to biased estimates due to endogeneity concerns. Urban spatial structure can be both a driver and a result of economic and environmental outcomes, leading to issues such as reverse causality and omitted variable bias. To address these concerns, we employ an instrumental variable (IV) approach to better isolate the causal effects of urban spatial structure on economic and environmental outcomes.

Our empirical approach has two main components: First, we construct instrumental variables to address the endogeneity issues inherent in OLS regressions. Second, we use these instruments to estimate the causal impact of urban spatial structure.

5.2.1 Instrumental Variable Construction: Predicted Urban Spatial Structures

To address endogeneity, we use exogenous instrumental variables derived from the predicted urban footprint of Syncities in the United States in 1970. Building on Harari's methodology [44], we calculate predicted urban spatial structures based on historical population size and density, providing instruments that capture variation in urban spatial characteristics independent of economic and environmental conditions.

The validity of these instrumental variables relies on two key assumptions:

1. ***Relevance.*** The IV must be correlated with the endogenous variable. We test this assumption later in the first stage of our estimating equations. The predicted urban footprint is constructed using predicted population data from 1970 and population density from 1940, which captures the urban expansion trajectory that would have occurred if population growth continued as it had prior to the second wave of the Great Migration.

2. ***Exclusion Restriction.*** The IV must affect the dependent variable only through the endogenous explanatory variable. This means the predicted urban spatial structure should influence economic and environmental outcomes only through its impact on the actual urban spatial structure. Although this assumption is difficult to test directly, it is reasonable to assume that the predicted urban structure, rooted in historical demographic trends, would not have a direct effect on outcomes beyond its influence on urban form.

To construct our exogenous instruments, we start by predicting the urban footprint, closely linked to population, policy, and natural features [44]. We predict the urban footprint based on expected urban population growth, accounting for policy and natural constraints. We mechanically determine Syncity population growth to avoid endogeneity issues. We also consider distinct exogenous constraints for each Syncity during expansion.

Precisely, we predict urban population growth to determine a radius for each Syncity. Using this predicted radius, we identify all developable land within the radius, excluding areas restricted by geographic or political factors, to define the predicted urban footprint. From this footprint, we calculate the predicted urban areas and cohesion index, which serve as instrumental variables for actual urban size and shape.

The construction of these IVs involves several steps. First, we predict the urban population for each Syncity in 1970 by aggregating administrative-level population data to the Syncity level. We then project population growth for each Syncity log-linearly over the decades 1950, 1960, and 1970, using data from the period before the second wave of the Great Migration (1900-1940). This is represented by the following regression:

$$\log(pop_{c,t}) = \tau \times year + \epsilon_{c,t} \tag{5.3}$$

The resulting predicted populations for 1950, 1960, and 1970 reflect historical trends without the influence of later factors, such as migration waves or economic shocks.

Next, we construct the predicted Minimum Bounding Circle (MBC), which is the smallest possible circle that can encompass all the points within a given urban area for each Syncity in 1970, based on the predicted urban population and historical population density from 1940, using the equation:

$$\log(\text{MBC area}_{c,1970}) = e \times \log(\widehat{pop}_{c,1970}) + f \times \log\left(\frac{pop_{c,1940}}{\text{MBC area}_{c,1940}}\right) + \epsilon_c \tag{5.4}$$

The predicted MBC area provides an estimate of the potential urban footprint in 1970, considering historical population growth patterns.

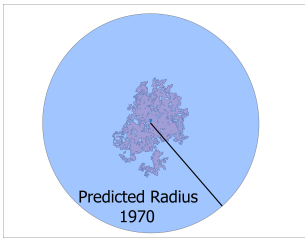
We then calculate the radius of the predicted MBC area for each Syncity in 1970 using:

$$\widehat{r}_{c,1970} = \sqrt{\frac{\widehat{\text{MBC area}}_{c,1970}}{\pi}} \tag{5.5}$$

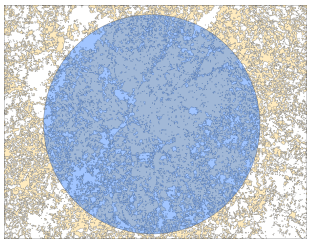
Using the predicted radius $\widehat{r}_{c,1970}$, we construct the predicted MBC for each Syncity, centered at the actual MBC from 1940 (Figure 5.2(a)).

Next, we apply the City Clustering Algorithm (CCA) to cluster built-up areas from 2010, which allows us to determine developable land within each predicted MBC. We assume that the built-up area (BUA) within Syncity boundaries had reached its maximum potential by 2010. Areas without buildings, or "white holes" as shown in Figure 5.2(b), were likely undevelopable due to geographic or policy constraints. Using the 2010 BUA, we ensure that our determination of 1970 developable land is exogenous.

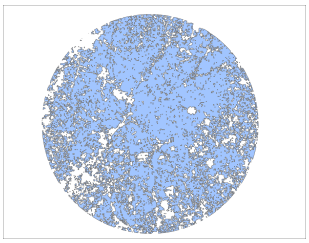
Finally, we intersect the predicted MBC with the 2010 CCA clusters to obtain the final predicted urban footprint without undevelopable land for each Syncity (Figure 5.2(c)). These predicted footprints allow us to calculate the predicted urban areas (BUA) and cohesion index, following the methods described in Chapter 2 for constructing the actual BUA and cohesion index.



(a) The pink shapes underneath represent Atlanta Syncity in 1970, and the light blue circle on top is the predicted MBC.



(b) The graph shows the partial areas around Atlanta. The yellow patches are the CCA clusters in 2010.



(c) The blue multi-polygons in the graph indicate the predicted urban footprint of Atlanta Syncity in 1970.

Figure 5.2: The Construction of Predicted Urban Footprint on Map: An Example

5.2.2 Estimating Equations

We assess how urban size ($area_{i,1970}$) and shape ($\mathrm{S}_{i,1970}$) affect economic and environmental outcomes by employing the corresponding IVs: predicted size ($\tilde{\mathrm{area}}_{i,1970}$) and predicted shape ($\tilde{\mathrm{S}}_{i,1970}$). Recall that we use the built-up area (BUA) to measure urban size and the cohesion index to measure urban shape, capturing spatial dispersion and connectivity of urban clusters. The economic outcomes refer to total income in 1974 and 2010, while environmental outcomes are measured using AQI indicators in 1980 and 2010.

We apply the standard two-stage least squares (2SLS) estimation. In the first stage, we model the relationship between the predicted urban size and shape and their actual counterparts. The first-stage estimating equations are as follows:

$$\text{First Stage:} \quad \ln(area_{i,1970}) = \gamma_T^Z \times \ln(\tilde{\text{area}}_{i,1970}) + \mathbb{X}_i + \epsilon_i \tag{5.6}$$

$$\text{First Stage:} \quad \ln(\text{S}_{i,1970}) = \theta_T^Z \times \ln(\tilde{\text{S}}_{i,1970}) + \mathbb{X}_i + \epsilon_i \tag{5.7}$$

The control variables ($\mathbb{X}_i$) include factors such as the share of the manufacturing labor force, educational mobility, the share of the 1940 Black migrant population, and regional fixed effects.

In the second stage, we use the predicted values from the first stage to estimate the impact of urban size and shape on economic and environmental outcomes:

$$\text{Second Stage:} \quad \ln(Y_{i,t}) = \gamma_Y^T \times \ln(\widehat{area}_{i,1970}) + \mathbb{X}_i + \epsilon_i \tag{5.8}$$

$$\text{Second Stage:} \quad \ln(Y_{i,t}) = \theta_Y^T \times \ln(\widehat{\text{S}}_{i,1970}) + \mathbb{X}_i + \epsilon_i \tag{5.9}$$

In Equations 5.8 and 5.9, the coefficients γ_Y^T and θ_Y^T indicate the effects of urban size and shape, respectively, on economic and environmental outcomes. By employing the IV, we aim to provide unidirectional causal estimates of the effects of urban spatial structures on economic and environmental outcomes. This method helps account for potential biases if existing economic or environmental conditions endogenously influence urban spatial changes.

This empirical strategy allows us to answer our research questions: What are the economic and environmental consequences of urban spatial structure? The use of the IV approach can ensure that our findings accurately and credibly reflect the true impact of urban spatial structure on outcomes.

5.3 Results

This section shows how urban spatial structure influences economic performance in Tables 5.1 and 5.2 and environmental performance in Tables 5.3 and 5.4.

In Tables 5.1 - 5.4, we report the results of our preferred coefficient estimate, including all control variables.[1] The complete regression results, including IV regression results with and without control variables and ordinary least squares (OLS) results, are shown in Appendix B. In addition, we cluster at the County-Syncity Union level, which is our merged spatial unit, to make sure the standard errors will not be underestimated.

5.3.1 Economic Outcomes

Table 5.1: Effects of of Urban Size on Economic Outcomes

	Total Income	
	1974	**2010**
	(IV)Second Stage	
Ln BUA 1970 (γ_Y^T)	1.156[a]	.643[a]
	(.329)	(.100)
	(IV)First Stage	
Ln Predicted BUA 1970	.845[a]	.862[a]
	(.031)	(.041)
Observations	114	74
Cragg-Donald Wald F statistic	887.943	512.139
Control	X	X

Note: [a], [b], and [c] denote significance at the 1 percent, 5 percent, and 10 percent levels, respectively. The observations are different for different years because we use different outcomes.

[1]The control variables include shares of the manufacturing labor force, educational upward mobility, the share of the 1940 Black migrants, and region fixed effects.

Table 5.2: Effects of of Urban Shape on Economic Outcomes

	Total Income	
	1974	**2010**
	(IV)Second Stage	
Ln Cohesion Index 1970 (θ_Y^T)	4.670[b]	2.465[b]
	(1.985)	(1.042)
	(IV)First Stage	
Ln Predicted Cohesion Index 1970	.499[a]	.483[b]
	(.196)	(.235)
Observations	114	74
Cragg-Donald Wald F statistic	12.989	8.898
Control	X	X

Note: [a], [b], and [c] denote significance at the 1 percent, 5 percent, and 10 percent levels, respectively. Compared to the relationship between the cohesion index in 1970 and the income level in 1974, the relationship between the cohesion index in 1970 and the income level in 2010 shows a weak IV with a low F-statistic. One reason for this could be the smaller sample size. The number of observations in 2010 might be too limited, which could lead to a weaker correlation between the IV and the cohesion index. Therefore, it cannot provide reliable statistical support to identify the causal relationship. In other words, it needs to be cautious with the reported results due to the limitation of the statistical results.

In Table 5.1, the second-stage results estimate the effect of urban size in 1970 on total income. In the short term, every additional 1% in urban size in 1970 led to an additional 1.156% in total income in 1974. Given that the average total income in 1974 was approximately $10,400 million[2], this 1.156% increase represents roughly a $120 million rise in total income due to urban expansion. Furthermore, with an average urban size of 357 square kilometers in 1970, each additional square kilometer of urban expansion resulted in approximately $34 million of additional total income in 1974, or about $90 per person annually, given the average population size of around 380,000 in 1975.

Urban shape, represented by the cohesion index, also positively impacts short-term income performance. In Table 5.1, the second-stage results present the effect of urban shape in 1970 on total income. Specifically, every 1% increase in the cohesion index in

[2]We adjusted the monetary values from 1974 to their 2010 equivalents using the Consumer Price Index (CPI) to account for inflation.

1970 resulted in a 4.67% increase in total income in 1974, equivalent to an increase of about $485.68 million, based on the average income level mentioned previously. Given that the average cohesion index was 10 kilometers in 1970, each 100-meter increase in the distance between urban clusters was associated with approximately $485.68 million in additional income in 1974, or around $1,300 per person annually.

Long-term income effects are also significant, although smaller in magnitude. A 1% increase in urban size in 1970 resulted in a 0.643% increase in total income in 2010. Given the average total income of $16,600 million in 2010,[3] this represents an increase of about $106.74 million, or $21.26 million per square kilometer of urban expansion, which is equivalent to about $33 per person annually, given the average population of 645,000. Additionally, a 1% change in the cohesion index in 1970 led to a 2.465% change in total income in 2010, or an increase of around $409.19 million. This implies a change of about $314.7 million per kilometer of urban shape dispersion, equivalent to approximately $500 per person annually for every additional 100 meters of distance between clusters.

[3]The average urban size or shape in 1970 may be different across different causal chains, depending on the sample size available for each chain. Specifically, because different outcomes have differences in data availability, the useful number of observations in each causal chain may differ when the independent variable remains the same. Due to data limitations, we did not balance variables across all causal chains. We expect to maximize the sample size for each causal chain individually to keep as much data as possible for each individual relationship. Thus, the average urban size in 1970 here differs from the average urban size mentioned above. Similarly, the average urban size or shape in the same year may differ in subsequent contexts.

5.3.2 Environmental Outcomes

Table 5.3: Effects of of Urban Size on AQI Outcomes

	Median AQI		**Max AQI**		**Unhealthy Weather**	
	1980	**2010**	**1980**	**2010**	**1980**	**2010**
			(IV)Second Stage			
Ln BUA 1970 (γ_Y^T)	.165[a]	.100[a]	.140[a]	.081[a]	.053[a]	.019[a]
	(.036)	(.025)	(.030)	(.023)	(.011)	(.006)
			(IV)First Stage			
Ln Predicted BUA 1970	.843[a]	.841[a]	.843[a]	.841[a]	.843[a]	.841[a]
	(.043)	(.034)	(.043)	(.034)	(.043)	(.034)
Observations	93	103	93	103	93	103
Cragg-Donald Wald F statistic	505.980	731.096	505.980	731.096	505.980	731.096
Control	X	X	X	X	X	X

Note: [a], [b], and [c] denote significance at the 1 percent, 5 percent and 10 percent levels, respectively.

Table 5.4: Effects of of Urban Shape on AQI Outcomes

	Median AQI		Max AQI		Unhealthy Weather	
	1980	**2010**	**1980**	**2010**	**1980**	**2010**
			(IV)Second Stage			
Ln Cohesion Index 1970 (θ_Y^T)	.475[c]	.347[b]	.400[b]	.251[b]	.164[c]	.063[b]
	(.267)	(.149)	(.194)	(.128)	(.085)	(.026)
			(IV)First Stage			
Ln Predicted Cohesion Index 1970	.583[b]	.498[b]	.583[b]	.498[b]	.583[b]	.498[b]
	(.263)	(.212)	(.263)	(.212)	(.263)	(.212)
Observations	93	103	93	103	93	103
Cragg-Donald Wald F statistic	10.110	11.580	10.110	11.580	10.110	11.580
Control	X	X	X	X	X	X

Note: [a], [b], and [c] denote significance at the 1 percent, 5 percent, and 10 percent levels, respectively.

Tables 5.3 and 5.4 summarize the effects of urban spatial structure on air quality, represented by different AQI metrics: median AQI, max AQI, and the ratio of unhealthy weather days. Median AQI reflects general air pollution levels, max AQI measures the worst levels of air pollution, and the ratio of unhealthy weather days shows the proportion of days classified as having poor air quality.

According to Table 5.3, each additional 1% in urban size in 1970 led to a 0.165% change in the median AQI in 1980. With an average median AQI of 44.14 in 1980, this 0.165% translates to a 0.073 change in median AQI. Every additional 10-square-kilometer increase in urban size resulted in a roughly 0.171 increase in median AQI in 1980.[4] For the long-term effect, each additional 1% in urban size in 1970 led to a 0.1% increase in the median AQI in 2010, equivalent to a 0.0485 increase based on an average AQI of 48.50 in 2010. The results suggest that every 10-square-kilometer increase in urban size in 1970 led to an increase of approximately 0.125 in median AQI in 2010.

The results for max AQI and the ratio of unhealthy weather days align with those for median AQI. Each 1% increase in urban size in 1970 led to a 0.14% increase in max AQI in 1980, which dropped to 0.081% in 2010. Given that the average max AQI in 1980 was 196.21 and went down to 141.23 in 2010, every 10-square-kilometer increase in urban

[4] A 10 square kilometer similar to the size of a large-sized community.

size[5] resulted in an approximately 0.6448 change in max AQI in 1980 and a 0.295 change in 2010. Similarly, each 1% increase in urban size led to a 0.053% increase in unhealthy weather days in 1980, which decreased to 0.019% in 2010. The findings imply that every 10-square-kilometer increase in urban size led to an additional 0.06 unhealthy weather days annually in 1980, compared to 0.01 in 2010.

The AQI is an index where values above 100 are considered unhealthy. Although average AQI values were generally acceptable in both 1980 and 2010, the findings suggest that urban size expansion had a negative impact on air quality over time. The short-term impact of urban expansion on air quality is more pronounced, with effects persisting over the long term. These results indicate the potential environmental cost of continued urban sprawl and its cumulative effects on air quality.

The cohesion index also has a positive relationship with AQI indicators. Specifically, each additional 1% in urban shape in 1970 led to a 0.475% change in median AQI in 1980. Based on an average cohesion index of 11 kilometers and an average median AQI of 44.14 in 1980, this translates to a 0.21 change in median AQI, implying that each kilometer increase in cohesion led to a 4.5% change in AQI. In the long term, a 1% increase in urban shape in 1970 led to a 0.347% change in median AQI in 2010, representing a 0.17 change in median AQI. This indicates that each kilometer of cohesion expansion led to a roughly 3.1% increase in air pollution levels.

The impact of urban shape on max AQI was also positive, with a 1% increase in urban shape in 1970 leading to a 0.4% change in max AQI in 1980, and 0.251% in 2010. Based on the average cohesion index of 11 kilometers in 1970, average max AQI of 196.21 in 1980, and 141.23 in 2010, the results imply that every 1-kilometer change in the average distance between clusters in 1970 resulted in the change of a max AQI of 7 in 1980 and 3.2 in 2010.

Similarly, for unhealthy weather days, each 1% increase in urban shape resulted in a 0.164% increase in 1980 and 0.063% in 2010. Based on the average cohesion index of 11 kilometers in 1970, the average ratio of unhealthy weather days of 12.43% in 1980 and 3.57% in 2010, the results indicate that every 1-kilometer change in the average distance between clusters in 1970 caused around one unhealthy weather day changed out of a year in 1980. In comparison, every 1-kilometer change in the average distance between clusters in 1970 caused only 0.075 unhealthy weather days to change out of a year in 2010. The long-term effects are less severe than the short-term, yet the cumulative pressure on air quality persists.

[5]The average urban size in the dataset for the environmental outcomes in 1980 is 426 square kilometers, and it is 388 square kilometers when in the dataset for the outcomes in 2010.

Overall, the short-term impacts of urban shape and size on air quality are more dramatic than the long-term effects. However, the continued expansion of cities and the resulting changes in urban spatial structure have enduring environmental consequences, which must be considered in urban planning and policy development.

5.4 Discussion

In this chapter, we examined how urban spatial structure in 1970 influenced total income and air quality in non-southern urban areas, both in the short term and long term. We employed an instrumental variable (IV) approach to address endogeneity issues. Specifically, we used the predicted urban spatial structure as the IV for the actual urban spatial structure, thereby mitigating biases resulting from reverse causality between urban spatial structure and outcomes, including income and air quality performance.

Reverse causality is a significant challenge that leads to endogeneity in causal relationships. Urban size or shape can determine, or be determined by, economic or environmental conditions. For instance, a geographically larger city might foster economic growth by providing richer resources, such as a larger population, labor force, and more businesses. Conversely, a more compact city might improve resource efficiency and reduce transportation costs, enhancing economic performance. On the other hand, faster economic growth can influence urban shape and form. In rapidly growing areas, higher land values can lead to more efficient land use, with denser building areas making cities more compact. Such reciprocal causation complicates the one-way impact analysis of urban size and shape on economic and environmental outcomes if only ordinary least squares (OLS) regressions are used. Using an instrument based on historical population size and density, we ensured that the causal impact of urban spatial structure on outcomes was not confounded by other factors, thus yielding more credible results regarding the effects of urban spatial changes on economic growth and environmental resilience.

Our results demonstrate that larger urban size in 1970 led to higher total income levels in 1974. This suggests that the immediate economic benefits were driven by the expansion of a city's horizontal size. Similarly, cities with more dispersed urban shapes in 1970, characterized by longer average distances between clusters, also experienced higher total income levels in 1974.

In the long term, the effects were consistent: both larger urban size and more dispersed urban shape in 1970 were associated with higher total income levels in 2010. This finding emphasizes the lasting importance of urban spatial structure; the size and shape of cities in

1970 not only influenced immediate economic performance but also established a foundation for economic development that persisted over several decades. However, the long-term influence of urban size and shape on total income in 2010 was less pronounced compared to the short-term influence in 1974. This diminishing impact over time might be attributed to various intervening factors, such as technological advancements and policy changes, which affected the initial influence of urban spatial structure on income performance.

Regarding environmental outcomes, the results indicate that larger urban size in 1970 led to poorer air quality outcomes in 1980, as measured by the AQI. Similarly, cities with more dispersed shapes in 1970 also experienced worse air quality in 1980. This could be due to increased pressure on traffic emissions and industrial pollution in larger or more dispersed cities, translating into unhealthy air quality in the short term. In the long term, both larger urban size and more dispersed shape in 1970 contributed to poor air quality in 2010, suggesting that urban spatial structure played an essential role in shaping environmental outcomes. The physical structure of cities in 1970 influenced the industrial framework, which persisted and contributed to air pollution issues until 2010. As with income outcomes, the impact of urban spatial structure on air quality was more significant in the short term compared to the long term. In the short term, physical urban structures remained largely unchanged, leading to a direct and immediate impact on air quality. However, over the long term, cities could upgrade infrastructure or implement policies to mitigate air pollution, thereby lessening the initial impact of urban spatial structure.

Our findings suggest that larger or more dispersed cities tend to experience higher total incomes and worse air quality in the short and long term. These results align with existing research on the economic and environmental impacts of urban spatial structure across different periods and locations [44, 52, 73]. This consistency suggests that the effects of urban spatial structure—particularly in terms of size and spatial dispersion—on economic and environmental outcomes may be universal across different contexts.

A key observation from our results is the evident outcomes of both economic benefits and environmental costs in response to changes in urban spatial structure. These outcomes imply that adjustments to urban spatial structure can spur economic growth while degrading environmental performance.

Chapter 6

A Causal Mediation Analysis of the Role of Immigration Shocks on Economic and Environmental Outcomes

6.1 Introduction

The Great Migration shock, especially the second wave (1940-1970), represents one of the most significant demographic shifts in U.S. history, reshaping the demographic landscape while playing an important role in influencing both economic development and environmental performance in cities. The influx of African American migrants impacted key outcomes such as income levels and air quality, but these changes may not have been driven solely by the racial composition shock itself. Instead, they were potentially mediated through other factors, such as urban spatial structure. Understanding how urban spatial structure—including urban size and spatial dispersion—served as a bridge for the impacts of this demographic shift allows us to gain a deeper insight into the relationship between population movement, economic outcomes, and environmental quality. This chapter explores this underexamined yet crucial aspect, shedding light on both direct and indirect mechanisms at play.

This chapter addresses two main questions: (i) What were the short- and long-term economic and environmental impacts of the influx of African American migrants during the second wave of the Great Migration (1940-1970) on U.S. cities? (ii) How did changes

in urban size and spatial structure mediate the relationship between this demographic shift and its effects? To answer these questions, we apply an instrumental variable (IV) mediation analysis to estimate the total, direct, and indirect effects of the Great Migration on economic and environmental outcomes, with urban spatial structure serving as the mediator. This approach allows us to capture the dynamic influence of migration flows by examining both short-term and long-term outcomes.

During the Great Migration, many African American migrants left the South for better economic opportunities in non-Southern cities, and they saw significant economic gains that narrowed the economic status gap between Blacks and Whites [68, 69].

The Great Migration also significantly impacted the financial situation in northern cities. For example, African American migrants decreased property values in northern cities, negatively impacting public spending and tax revenues [43].

The influence of historical conditions on present-day economic development is well-documented in the literature. Research demonstrates that the Great Migration not only transformed the lives of African American migrants but also shaped the economic prospects of subsequent generations. Specifically, the migration had substantial long-term benefits for educational attainment, income levels, and poverty reduction among the descendants of African Americans who moved from the South, compared to those who remained in the region [23]. However, the Great Migration also had unintended negative consequences for economic mobility in non-southern cities, largely due to local government responses. Cities receiving large numbers of Black migrants increased spending on policing and incarceration, which hindered economic opportunities for African American families [16].

Beyond economic impacts, the Great Migration also significantly influenced the urban environment. One important perspective is how African American migrants reshaped the physical landscape to reflect their collective identity and resist urban racism. In Baltimore, for instance, the interplay between rural and urban environments demonstrated how migrants from the rural South transformed urban spaces, creating a "rural" presence within the city [6]. Migrants drew on their rural heritage to redefine their identity and their surroundings in non-southern cities. Similarly, in Chicago, African Americans utilized green spaces to mitigate the negative effects of urban racism, reshaping these areas as symbols of resistance and community resilience [7].

Another crucial aspect of environmental performance to consider is emissions. Although there is limited literature specifically examining the link between the Great Migration and emissions, existing research on internal migration and environmental quality in the United States provides valuable insights. Studies suggest that migration can influence energy consumption patterns, thereby affecting environmental quality [2]. Given these findings, it

is plausible to explore how a significant migration event like the Great Migration impacted environmental conditions, particularly air quality, which will be the focus of this chapter.

This chapter makes four significant contributions to the existing body of knowledge. First, it addresses a gap in the literature by exploring the relationship between demographic shifts and environmental quality. While existing studies rarely examine the effects of racial composition shocks or migration flows on air quality, we investigate how changes in the share of African Americans during the second wave of the Great Migration influenced both short- and long-term air quality. This exploration highlights the role of demographic and racial diversity in shaping environmental outcomes, offering a fresh perspective on the existing literature.

Second, this chapter examines the impact of racial composition changes on both economic and environmental dimensions. Unlike most studies that focus exclusively on either economic or environmental outcomes, we consider both economic growth and environmental costs. Our findings indicate that an increase in the share of African Americans was associated with higher total income but poorer air quality, underscoring the complex relationship between demographic shifts and sustainable development. By adopting a holistic approach, we contribute to a more nuanced understanding of these interconnected dimensions.

Third, this chapter examines both the short-term and long-term effects of demographic changes. We treat the demographic shift caused by the Great Migration as a historical condition and analyze how it continues to shape contemporary economic and environmental conditions. This contribution broadens our understanding of the long-lasting effects of demographic shifts and offers empirical support for the concept of path dependency.

Fourth, this chapter employs instrumental variable (IV) mediation analysis to uncover the underlying mechanisms linking demographic shocks to outcomes. While most research on the Great Migration focuses on its overall effects on economic or environmental performance, we add depth by introducing urban spatial structure as a mediator. This allows us to reveal the direct and indirect pathways through which the migration shock influenced economic and environmental outcomes. By doing so, we provide a more comprehensive understanding beyond examining only the superficial impacts of migration, contributing valuable insights to the existing literature.

This chapter proceeds as follows: After reviewing the literature, highlighting the gap, and discussing the motivation in this introduction section, the next section will outline the empirical strategy of the IV mediation analysis. Section 6.3 will report the results in the short and long term. The Section 6.4 will discuss the implications of these results.

6.2 Empirical Strategy

In this section, we discuss the empirical strategy for studying the total effect of the racial composition shock on economic and environmental performance, as well as the underlying mechanism through the impact of the racial composition shock on urban spatial structure. The causal relationships in Figure 6.1 show how the racial composition shock, which is represented by changes in the share of Blacks, during the second wave of the Great Migration influenced the urban spatial structures at the end of the Great Migration and then influenced both short and long term economic and environmental performance. Previous findings, such as the economic gains caused by the Great Migration in Collins & Wanamaker's [69] and Collins's [68] papers, have shown some evidence of the relationships between the Great Migration and the outcomes. Thus, building on the findings from previous studies, we expect to study the effect of the racial composition shock on the economy and the environment through a more complex mechanism by using the urban spatial structure as a bridge in between.

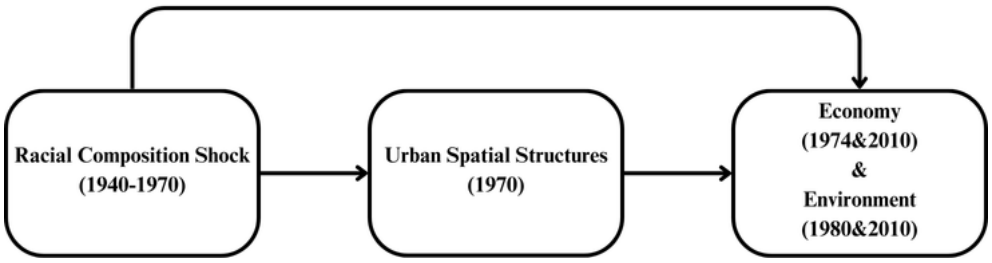


Figure 6.1: The Effect of the Great Migration on Outcomes Through the Urban Spatial Structure

To understand the effect of our interested independent variable on the interested outcomes through the underlying mechanism, we apply the IV mediation analysis method following Dippel et al.'s paper [9]. According to their paper, a mediator (M) is a variable between an independent variable - treatment (T) - and a dependent variable - outcome (Y) - to explain how the treatment affects the outcome. Based on the framework in Figure 6.1, the urban spatial structure plays such a mediating role. We expect to investigate how the urban spatial structure (M), as a mediating factor, can explain how the racial composition shock (T) influences economic and environmental performance (Y).

Specifically, to identify the total effect of T on Y and how M explains this total effect, Dippel et al. [9] first decompose the complex relationships into three parts: the impact of

T on M, the effect of T on Y, and the effect of M on Y. The simple IV model can examine the first two relationships without bias from reverse causality or omitted variable issues. However, it is hard to use a simple IV model to test the causal relation between M and Y. Therefore, Dippel et al. [9] introduced a mediation model that allows the identification of the relationship between M and Y.

This mediation model allows us to solve the potential bias due to endogeneity and identify three causal relations between treatment (T) & outcome (Y), treatment (T) & mediator (M), and mediator (M) & outcome (Y) using a single instrumental variable (IV). In this chapter, the T is the racial composition shock, which is an endogenous variable in the relation between the racial composition shock (T) & urban spatial structure (M) and the racial composition shock (T) & economic and environmental outcomes (Y). As shown in Figure 6.2, the demographic shift (T) causes economic and environmental outcomes (Y) through observed urban spatial structure (M) and also through unobserved mediators U. Moreover, there are unobserved variables (V) that cause the demographic shift (T) and urban spatial structure (M) simultaneously, which will cause endogeneity issues in the causal relation between T and M. For example, social-cultural preference might influence demographic distribution and how people organize their communities' spatial structure. In addition, the reverse causality that the demographic shift and urban spatial structure might be both cause and effect to each other may also cause endogeneity problems. Similar endogeneity issues may also occur in the causal relation between T and Y. The unobserved variable (V) in Figure 6.2 causes the demographic shift (T) and outcomes (Y), including economic and environmental conditions, simultaneously. For example, industrial development can lead to economic growth and environmental pressure and, at the same time, attract immigrations, especially the African American migrants who look for economic opportunities in cities during the racial composition shock. Also, there might be an issue of reverse causality between the demographic shift and outcomes.

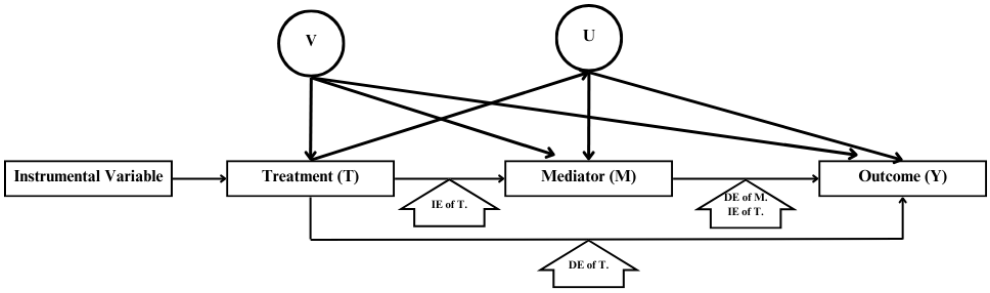


Figure 6.2: General IV Mediation Model [10].

To solve these endogeneity issues, we pick predicted Black migrations[1] as the IV for the treatment. The predicted Black migration could satisfy instrument validity conditions, including IV relevance and exclusion restrictions. In detail, this IV is a Black migration data that is projected relies on cultural ties, community networks, and distribution of the Black population before the second wave of the Great Migration. Thus, this IV is correlated with the actual change in the share of Blacks during the second wave of the Great Migration (T) and independent of the unobserved confounder V.

In terms of the causal relation between M and Y, as the only IV to be used in the mediation model, the predicted Black migration should be valid for the endogenous variable M. This means the predicted Black migration should correlate with urban spatial structure (M) and independent of the unobserved confounder V. However, as shown in Figure 6.2, the IV does not directly cause M. The joint unobserved confounder V, which causes T, M, and Y, is too general to identify the relation between M and Y. Thus, Dippel et al. [10] improved the mediation model from the general one in Figure 6.2 to a restricted one in Figure 6.3. According to Figure 6.3, the joint unobserved variable V is restricted to V_T, which directly causes T and M, and V_Y directly causes M and Y. The distinction between V_T and V_Y is based on different economic agents. In our empirical context, demographic shift (T) is endogenous. The endogenous issue is due to the heterogeneity of the demographic characteristics. The social-cultural preference discussed above is an example of the demographic unobserved factor (V_T), which influences demographic shift (T) and also influences the urban spatial structure (M). Moreover, urban spatial structure (M) is endogenous in a regression of M on economic and environmental outcomes (Y). The endogenous issue is due to the heterogeneity of the urban spatial characteristics. The urban spatial unobserved factors (V_Y), such as land use patterns, that influence urban spatial structure (M) would also influence the economic and environmental outcomes (Y). Specifically, the land use patterns directly influence the expansion or dispersion of the spatial structure. The distribution patterns of commercial or industrial areas directly influence the economy's growth pattern and pollution distribution.

[1]The construction of this IV is discussed in Chapter 4.2.1.

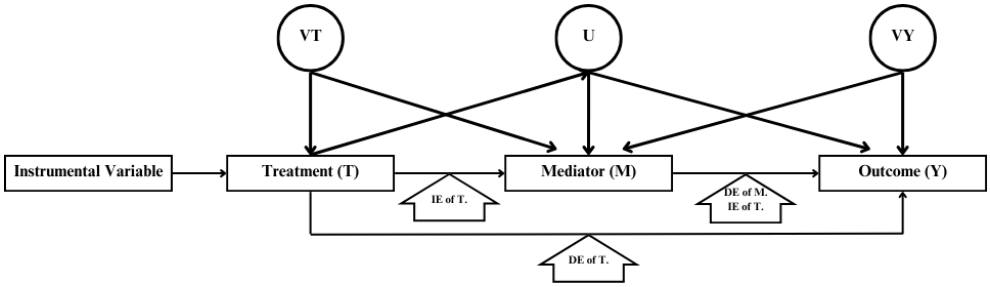


Figure 6.3: Restricted IV Mediation Model [10].

In the improved restriction mediation model, Dippel et al. [10] innovatively state that IV is a valid instrument to identify the mediator's causal effect on outcome when conditioned on treatment. In other words, Dippel et al. [10] solve the issue that the IV is not directly correlated with the mediator, or it is hard to identify exclusion restrictions. In our empirical context, the predicted Black migration (IV) is unconditionally independent of V_T (ex. social-cultural preference), which satisfies the exclusion restriction for the causal relation between demographic shift (T) and urban spatial structure (M). Nevertheless, the predicted Black migration (IV) is correlated with the social-cultural preference (V_T) when conditioned on the actual demographic shift because both predicted Black migration and socio-cultural preferences are associated with the actual immigration pattern, which reflects the actual demographic shift. Then, we can tell that, conditioned on the demographic shift (T), the predicted Black migration (IV) influences urban spatial structure (M) through social-cultural preference (V_T) because V_T causes M according to the Figure 6.3. In other words, the IV conditioned on T, becoming representative for V_T, then influencing M and instrumenting for M. In addition, as shown in Figure 6.3, the IV is independent of V_Y when conditioned on the treatment. Therefore, the predicted migration (IV) can be a valid IV for urban spatial structure (M), conditioned on the changes in the share of Blacks.

So far, we have explained the novel IV mediation model from Dippel et al.'s papers [9, 10] and how we identify the role of the IV in the model. Before applying this IV mediation analysis to our interested causal chains, we need to know how we use the three causal parameters, including the effect of T on M, the impact of T on Y, and the effect of M on Y, to understand the total impact of T on Y and the underlying mechanism of this total effect.

The corresponding equations of the IV mediation analysis are shown in Equations 6.1 - 6.5. The two components for the total effect in the IV mediation model include an indirect effect of treatment on the outcome that runs through the mediator and a direct residual impact that does not run through the mediator. In detail, in Model 1, for the relationship

between treatment and mediator in Figure 6.3, the treatment (T) is determined by the IV (Z) in Equation 6.1, and the estimated values of the treatment (T) are used in the Equation 6.2 for testing the effect of the treatment on the mediator, which is represented by the arrow directly from treatment to mediator in Figure 6.3. The coefficient η_T^M shows the effect of the treatment on the mediator, which indicates the indirect effect of treatment on the outcome that operates through the mediator.

Model 1

First Stage

$$T = \eta_Z^T \times Z + \epsilon_T \tag{6.1}$$

Second Stage

$$M = \eta_T^M \times \hat{T} + \epsilon_M \tag{6.2}$$

Moreover, in Model 2, the same IV (Z) is used as an instrument for the mediator (M) when conditioned on treatment (T) in Equation 6.3. Then, the estimated mediator is used in Equation 6.4 for testing the effect of the mediator on the outcome, which is represented by the arrow directly from mediator to outcome in Figure 6.3. In Equation 6.4, the mediator is an endogenous variable conditioned on the treatment. The coefficient η_M^Y shows the effect of the mediator on the outcome, which reveals another indirect effect of treatment on the outcome through the mediator, and the coefficient η_T^Y indicates the residual direct impact of the treatment on the outcome.

Model 2

First Stage

$$M = \theta_Z^M \times Z + \theta_T^M \times T + \epsilon_T \tag{6.3}$$

Second Stage

$$Y = \eta_M^Y \times \hat{M} + \eta_T^Y \times T + \epsilon_Y \tag{6.4}$$

Finally, by substituting the expression for M from Equation 6.2 into Equation 6.4, we obtain Equation 6.5, which shows an estimated total effect of T on Y. According to Equation 6.5, the total effect is the product of two indirect effects ($\eta_M^Y \times \eta_T^M$) plus the direct effect (η_T^Y).

Total Effect

$$Y = \eta_M^Y \times (\eta_T^M \times T + \epsilon_M) + \eta_T^Y \times T + \epsilon_Y = (\eta_M^Y \times \eta_T^M + \eta_T^Y) \times T + \eta_M^Y \epsilon_M + \epsilon_Y \quad (6.5)$$

6.2.1 Estimating Equations

In this section, we implement the IV mediation approach discussed in the previous section to estimate the total and direct effect of the racial composition shock, which is represented by the change in the share of Blacks ($\Delta Bpop_{i,1940-1970}$), on outcomes ($Y_i$), as well as the indirect effect through the effect on urban spatial structure, including total built-up area (BUA) as a measure of urban size ($area_{i,1970}$), and a cohesion index that captures the spatial dispersion and connectivity of urban clusters ($S_{i,1970}$).

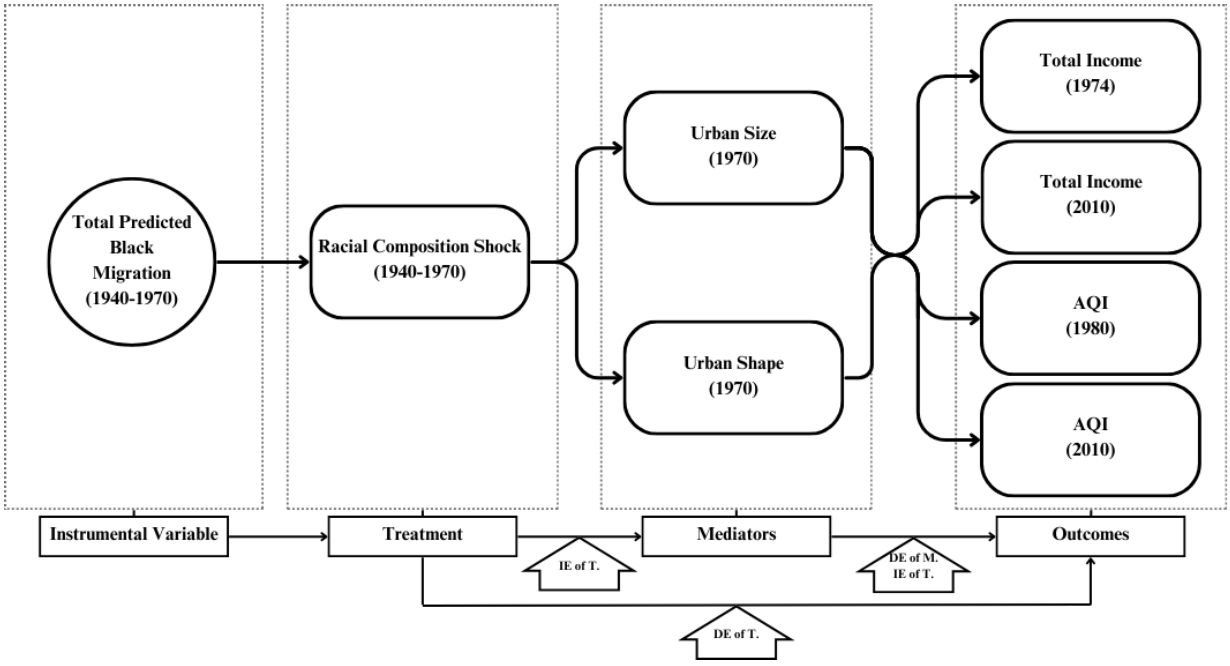


Figure 6.4: Mediation Analysis Causal Chains

Applying the IV mediation approach, the detailed causal chains for studying the underlying mechanism of the relationship between the racial composition shock and outcomes are shown in Figure 6.4. Recall that the arrow from the treatment to mediators represents the conditional indirect effect of the racial composition shock on outcomes through urban

spatial structures. The arrow from the treatment to outcomes shows the direct impact of the racial composition shock. In addition, the arrow from the mediators to the outcomes shows the direct effect of the urban spatial structure, which also represents the indirect effect of the racial composition shock on the outcomes through the urban spatial structure.

We then transform the causal chains into estimation equations. These equations follow the IV mediation method in Dippel et al.'s paper [9]. Recall that the estimation procedure includes two models. The estimating equations in Model 1 follow the standard two-stage least squares (2SLS) estimation. They are the same equations as those in Chapter 4. To ensure clarity, these equations are presented again in this chapter as equations 6.6 - 6.8. The first stage of Model 1 estimates the relationship between the instrumental variable (IV), which is the predicted Black migration ($\hat{M}_{i,1940-1970}$), and the treatment, which is the changes in the share of Blacks ($\Delta Bpop_{i,1940-1970}$).

Model 1

First Stage

$$ln(\Delta Bpop_{i,1940-1970}) = \alpha_T^Z \times ln(\hat{M}_{i,1940-1970}) + \mathbb{X}_i + \epsilon_i \tag{6.6}$$

where $\mathbb{X}_i$ is a set of control variables, including the share of the manufacturing labor force, educational upward mobility, the share of 1940 Black migrants, and regional fixed effects.

Then, the predicted value of the changes in the share of Blacks from the first stage is added to the equations in the second stage to estimate the effect of the changes in the share of Blacks on urban spatial structures. The parameters α_M^T and β_M^T in the second stage of the first model are to identify the effects of the changes in the share of Blacks on the urban size and urban shape, respectively. Moreover, within the context of IV mediation analysis, the meaning of the parameters α_M^T and β_M^T in the second stage is beyond a simple relationship between the racial composition shock and urban spatial structure. Still, it reveals how the racial composition shock influences the final outcomes through their impact on the urban spatial structure, which shows an indirect effect of the racial composition shock on the final outcomes. Then, the role of urban spatial structure becomes a mediating factor that can be used to study the underlying pathway for the relationships.

Second Stage

$$ln(area_{i,1970}) = \alpha_M^T \times ln(\widehat{\Delta Bpop}_{i,1940-1970}) + \mathbb{X}_i + \epsilon_i \tag{6.7}$$

$$ln(\mathrm{S}_{i,1970}) = \beta_M^T \times ln(\widehat{\Delta Bpop}_{i,1940-1970}) + \mathbb{X}_i + \epsilon_i \tag{6.8}$$

Model 2 is to identify the effects of the urban spatial structure on outcomes, conditional on the changes in the share of Blacks, by using predicted Black migration as the IV. Specifically, in this model, we control the changes in the share of Blacks to make sure the model has captured the direct effect of the changes in the share of Blacks on the outcomes to assume that the remaining unobserved disturbance has nothing to do with the changes in the share of Blacks. In other words, by controlling the changes in the share of Blacks, any effect of the predicted Black migration as an instrument can only be transmitted to outcomes through the urban size and shape, which are the mediators. This ensures that the predicted Black migration can be a valid IV while the changes in the share of Blacks is a conditioning variable, urban spatial structure is the endogenous variable, and the outcome is the dependent variable.

The estimating equations in the second model, when considering different mediators – urban size or shape – are shown in equations 6.9 - 6.12.

- **When using urban size (BUA) as the mediator:**

  First Stage

$$ln(area_{i,1970}) = \lambda_M^Z \times ln(\hat{M}_{i,1940-1970}) + \lambda_M^T \times ln(\Delta Bpop_{i,1940-1970}) + \mathbb{X}_i + \epsilon_X \quad (6.9)$$

  Second Stage

$$ln(Y_i) = \alpha_Y^M \times ln(\widehat{area}_{i,1970}) + \alpha_Y^T \times ln(\Delta Bpop_{i,1940-1970}) + \mathbb{X}_i + \epsilon_Y \quad (6.10)$$

- **When using urban shape (cohesion index) as the mediator:**

  First Stage

$$ln(\mathrm{S}_{i,1970}) = \delta_M^Z \times ln(\hat{M}_{i,1940-1970}) + \delta_M^T \times ln(\Delta Bpop_{i,1940-1970}) + \mathbb{X}_i + \epsilon_X \quad (6.11)$$

  Second Stage

$$ln(Y_i) = \beta_Y^M \times ln(\widehat{\mathrm{S}}_{i,1970}) + \beta_Y^T \times ln(\Delta Bpop_{i,1940-1970}) + \mathbb{X}_i + \epsilon_Y \quad (6.12)$$

The Equation 6.9 and 6.11 shows the first stage. The $area_{i,1970}$ and $\mathrm{S}_{i,1970}$ are the BUA and cohesion index in each county-Syncity union i in 1970, respectively. $\hat{M}_{i,1940-1970}$ is the predicted Black migration during the second wave of the Great Migration, which is used as the IV for BUA ($area_{i,1970}$) in 6.9 and for cohesion index ($\mathrm{S}_{i,1970}$) in 6.11. $\Delta Bpop_{i,1940-1970}$ is the changes in the share of Blacks, which is used as a conditioning variable.

Then the fitted BUA $\widehat{area}_{i,1970}$ and fitted cohesion index $\widehat{\mathrm{S}}_{i,1970}$ from equations 6.9 and 6.11 are added to equations 6.10 and 6.12, respectively. In the equations 6.10 and 6.12, which are second stages, outcomes Y_i include total income in 1974 and 2010 and air quality performance in 1980 and 2010. In Equation 6.10, the parameter α_Y^M shows the indirect effects of the changes in the share of Blacks on the outcomes that run through the BUA and the parameter α_Y^T shows the direct effects of changes in the share of Blacks on the outcomes. Then, according to Dippel et al.'s paper [9], the final indirect effect when using urban size as the mediator equals α_M^T in Equation 4.5 multiply α_Y^M in Equation 6.10. Similarly, when using urban shape as the mediator, the final indirect effect equals to β_M^T in Equation 4.6 multiply β_Y^M in Equation 6.12. Also, the β_Y^T in Equation 6.12 is the direct effect of the changes in the share of Blacks on the outcomes. Finally, the estimate of the total effect equals indirect effect plus direct effect:

- **When using urban size (BUA) as the mediator:**

$$Total\ Effect\ = \alpha_Y^T + \alpha_M^T \times \alpha_Y^M \tag{6.13}$$

- **When using urban shape (cohesion index) as the mediator:**

$$Total\ Effect\ = \beta_Y^T + \beta_M^T \times \beta_Y^M \tag{6.14}$$

Using the IV mediation approach, we can provide credible causal estimates to explain how the influx of African American migrants during the second wave of the Great Migration (1940-1970) affected the short— and long-term economic and environmental outcomes. We can also explain how urban size and shape play a role as a bridge in the relationship between the racial composition shock and the outcomes.

6.3 Results

This section presents the findings of our study on the impact of Black population growth on economic and environmental outcomes. Specifically, the mediation results[2] for different causal chains in Figure 6.4 are shown in tables 6.1 – 6.8[3]. In each table for economic

[2]Detailed results of the IV mediation analysis can be found in Appendix C.

[3]The quick information of each causal chain are provided in the captions for each table, in which arrows indicate the direction of the relationships. In the captions, 'T' stands for treatment, 'M' represents the mediator, and 'Y' stands for the outcomes.

outcomes, column (1) displays the results without control variables. Column (2) reports the results of controlling the characteristics[4] and region fixed effects. Moreover, in each table for environmental outcomes, columns (1), (3), and (5) report the results without control variables, and columns (2), (4), and (6) contain results when applying control variables. In the interpretation, we will focus on reporting the results with all controls included, which are our preferred results. In addition, we cluster at the County-Syncity Union level, which is our merged spatial unit, to make sure the standard errors will not be underestimated.

Following the method we discussed in the previous section, the final indirect effect in the IV mediation analysis is calculated by multiplying the coefficient of change in the share of Blacks from the first model with the coefficient of urban spatial metrics from the second model. Specifically, according to the detailed regression results for Model 1 and 2 in tables C.1 - C.8 in Appendix C, the coefficient in the second stage of Model 1 reveals one component of the final indirect effect. Recall that this indirect effect not only presents how the change in the share of Blacks impacts the urban spatial structure, including urban size and shape, but also reveals a deeper meaning for the mediation analysis: it shows how the change in the share of Blacks indirectly affects the outcomes through its influence on urban spatial structure. Moreover, the second component of the final indirect effect is shown in Model 2. The coefficient of the effect of urban spatial structure on outcomes in the second stage in Model 2 reports the second component of the indirect effect. Finally, the final indirect results shown in summary tables 6.1 – 6.8 are the product of these two components of the indirect effect. In addition, the direct effect is shown by the coefficient of change in the share of Blacks in the second stage in the second model in Appendix C, and the value of the total effect, shown in summary tables in this section, is the sum of the final indirect effect and direct effect. As a result, the proportion of indirect effect, which shows how important the mediator is in the relationship between the treatments and outcomes, refers to the share of the total effect represented by the final indirect effect. The following are the summary results of indirect, direct, and total effects for different causal chains with different final outcomes, including income performance in 1974 and 2010 and air quality performance in 1980 and 2010.

[4]The conditional characteristics including shares of the manufacturing labor force, educational upward mobility, and share of the 1940 Black migrants.

6.3.1 Short-term Economic Outcomes

Table 6.1: Changes in Share of Blacks (T) → Urban Size (M) → Short-term Income Outcomes (Y)

	Total Income 1974	
	(1)	(2)
Indirect effect (α_M^T * α_Y^M)	1.486[a]	1.581[a]
	(.330)	(.445)
Direct effect (α_Y^T)	.305	.456
	(.450)	(.561)
Total effect	1.790[a]	2.037[a]
	(.388)	(.390)
Proportion of Indirect Effect	83.02%	77.61%
Control		X

Note: [a], [b], and [c] denote significance at the 1 percent, 5 percent, and 10 percent levels, respectively.

Table 6.2: Changes in Share of Blacks (T) → Urban Shape (M) → Short-term Income Outcomes (Y)

	Total Income 1974	
	(1)	(2)
Indirect effect (β_M^T * β_Y^M)	1.047[b]	.992
	(.429)	(.605)
Direct effect (β_Y^T)	.743[c]	1.045[b]
	(.418)	(.456)
Total effect	1.790[a]	2.037[a]
	(.388)	(.390)
Proportion of Indirect Effect	58.47%	48.70%
Control		X

Note: [a], [b], and [c] denote significance at the 1 percent, 5 percent, and 10 percent levels, respectively.

According to Table 6.1, the total effect estimates that every 1% increase in the changes in the share of Blacks from 1940 to 1970 increases total income in 1974 by 2.037% when having full controls. Given that the average total income in 1974 was approximately $10,400 million [5], this 2.037% increase represents roughly a $212 million rise in total income due to the growth of urban Black population during the second wave of the Great Migration. Given that the average increase in the share of Blacks between 1940 and 1970 is around 11%, which corresponds to roughly 59,000 additional Black populations on average. Then, the result implies that for every 100 new Black residents, there was approximately $36 million of additional total income on average in 1974, or equivalent to $95 per person annually, given the average population size of around 380,000 in 1975.

However, the direct effects of the change in the share of Blacks itself are small and not statistically significant. Furthermore, the final indirect effect of the changes in the share of Blacks from 1940 to 1970 on the total income in 1974 shows that each 1% increase in the changes of the share of Blacks indirectly corresponds to a 1.581% rise in total income in 1974. The indirect effect explains around 78% of the total effect. In other words, the urban size in 1970 explains the majority of the total effects of the changes in the share of Blacks from 1940 to 1970 on the total income in 1974 that we discussed above.

[5]We adjusted the monetary values from 1974 to their 2010 equivalents using the Consumer Price Index (CPI) to account for inflation.

In Table 6.2, the results of the total effect of the increase in the share of Blacks on total income in 1974 are explained previously. The only difference in this causal chain is that we change the mediator from urban size to urban shape compared with the previous chain. When applying the control variables, the direct effect estimates that the 1.045% increase in the total income in 1974 is because of the changes in the share of Blacks from 1940 to 1970 itself. However, the final indirect effect is statistically insignificant. That is to say, different from urban size, urban spatial connectivity does not play an important role in transferring the effect of an increased Black population on economic growth in the short term.

6.3.2 Long-term Economic Outcomes

Table 6.3: Changes in Share of Blacks (T) → Urban Size (M) → Long-term Income Outcomes (Y)

	Total Income 2010	
	(1)	(2)
Indirect effect (α_M^T * α_Y^M)	1.109[a]	1.369[a]
	(.331)	(.358)
Direct effect (α_Y^T)	.010	.041
	(.146)	(.154)
Total effect	1.120[a]	1.410[a]
	(.268)	(.340)
Proportion of Indirect Effect	99.02%	97.09%
Control		X

Note: [a], [b], and [c] denote significance at the 1 percent, 5 percent, and 10 percent levels, respectively.

Table 6.4: Changes in Share of Blacks (T) → Urban Shape (M) → Long-term Income Outcomes (Y)

	Total Income 2010	
	(1)	(2)
Indirect effect (β_M^T * β_Y^M)	.965[b]	1.107
	(.490)	(.791)
Direct effect (β_Y^T)	.155	.303[c]
	(.157)	(.174)
Total effect	1.120[a]	1.410[a]
	(.268)	(.340)
Proportion of Indirect Effect	86.16%	78.51%
Control		X

Note: [a], [b], and [c] denote significance at the 1 percent, 5 percent, and 10 percent levels, respectively.

Regarding the long-term effect in Table 6.3, the total impact estimates that every 1% increase in the changes in the share of Blacks from 1940 to 1970 increases total income in 2010 by 1.410% when having controls. Given that the average total income in 2010 was around $16600 million, this 1.410% increase indicates an approximately $234 million increase in total income in 2010 due to the increasing Black population between 1940 and 1970. During this period, the average change in the share of Blacks is around 13%, which corresponds to roughly 89,000 additional Black populations on average[6]. The results imply that every 100 more Black residents in cities caused approximately $26 million more total income on average in 2010, which is equivalent to $40 per person more annual income on average, given the average population size of around 645,000 in 2010.

Regarding the underlying direct effect, the effects on total income in 2010 due to the change in Blacks' share are small and not statistically significant. Moreover, the final

[6]The average Black population change between 1940 and 1970 may be different across different causal chains due to different sample sizes available for each chain. Specifically, because different outcomes have differences in data availability, the final useful number of observations in each causal chain may differ. Due to data limitations, we did not balance variables across all causal chains. We expect to maximize the sample size for each causal chain individually to keep as much data as possible for each individual relationship. Thus, the average Black population change between 1940 and 1970 discussed here differs from the one we mentioned for another chain. Similarly, the average Black population change between 1940 and 1970 may vary in subsequent contexts.

indirect effects show that every 1% increase of the changes in the share of Blacks indirectly caused a 1.369% rise in the total income in 2010 through the impact on urban size. Thus, there is a full mediation effect that the urban horizontal expansion in 1970 explains almost all of the total impacts of the changes in the share of Blacks from 1940 to 1970 on the total income in 2010 that we discussed previously. In comparison, Table 6.4 shows a small but statistically significant direct effect, while the indirect effect shows economic significance with a high fraction but statistically insignificant. This implies that the effect of the changes in the share of Blacks on total income in 2010 did not run through the urban shape, represented by the cohesion index.

6.3.3 Short-term Environmental Outcomes

Table 6.5: Changes in Share of Blacks (T) → Urban Size (M) → Short-term AQI Outcomes (Y)

	Median AQI 1980		**Max AQI 1980**		**Unhealthy Weather 1980**	
	(1)	(2)	(3)	(4)	(5)	(6)
Indirect Effect (α_M^T * α_Y^M)	-.002	.130	.101	.138[c]	-.001	.040[c]
	(.103)	(.087)	(.065)	(.074)	(.032)	(.024)
Direct Effect (α_Y^T)	.075	.030	.040	.039	.033	.020
	(.074)	(.060)	(.051)	(.059)	(.022)	(.016)
Total Effect	.073	.160[b]	.141[a]	.177[a]	.033[b]	.060[a]
	(.055)	(.067)	(.039)	(.046)	(.015)	(.017)
Proportion of Indirect Effect	2.74%	81.25%	71.63%	77.97%	3.03%	66.67%
Control		X		X		X

Note: [a], [b], and [c] denote significance at the 1 percent, 5 percent, and 10 percent levels, respectively.

Table 6.6: Changes in Share of Blacks (T) → Urban Shape (M) → Short-term AQI Outcomes (Y)

	Median AQI 1980		**Max AQI 1980**		**Unhealthy Weather 1980**	
	(1)	(2)	(3)	(4)	(5)	(6)
Indirect effect (β_M^T * β_Y^M)	-.001	.072	.072	.076	-.001	.022
	(.074)	(.075)	(.057)	(.065)	(.023)	(.021)
Direct effect (β_Y^T)	.074	.088[c]	.068[c]	.100[b]	.033[b]	.038[a]
	(.052)	(.046)	(.041)	(.044)	(.015)	(.012)
Total effect	.073	.160[b]	.141[a]	.177[a]	.033[b]	.060[a]
	(.055)	(.067)	(.039)	(.046)	(.015)	(.017)
Proportion of Indirect Effect	1.37%	45%	51.06%	42.94%	3.03%	36.67%
Control		X		X		X

Note: [a], [b], and [c] denote significance at the 1 percent, 5 percent, and 10 percent levels, respectively.

According to our preferred estimation results, which control all characteristics and fixed effects, in Table 6.5, the total impact estimates that every 1% increase in the changes in the share of Blacks increases the median AQI in 1980 by 0.16%. Given that the average median AQI in 1980 was approximately 44, the 0.16% increase means a 0.0704 increase in the average median AQI. During the second wave of the Great Migration, there were roughly 72,000 additional Black residents in cities on average, which implies that every 10,000 additional Black residents caused the median AQI to increase by around 1.

Similarly, the results indicate that every 1% increase in the changes in the share of Blacks increases the max AQI in 1980 by 0.177%. Given that the average max AQI in 1980 was around 196 and the average Black population rise between 1940 and 1970 was around 72,000, the results show an increase of max AQI by around 5 for every 10,000 additional Black residents. In comparison, the increasing Black population has a five times greater impact on the max AQI than the median AQI because the max AQI amplifies the effect on high pollution areas such as areas with high industrial or traffic density. In other words, the same increase in the Black population leads to a higher risk of having extremely high AQI on some days in high-pollution areas, highlighting the need for greater attention to these regions when handling the demographic shift.

The last AQI-related outcome is the ratio of unhealthy weather days. The results show

that every 1% increase in the change in the share of Blacks increases the ratio of unhealthy weather days in 1980 by 0.06%. Given that the average ratio of unhealthy weather days is around 12% and the average Black population rise between 1940 and 1970 was around 72,000, the results indicate that every 10,000 additional Black residents leads to around half more unhealthy weather days out of a year in 1980.

The underlying indirect impact of the demographic shift on the ratio of unhealthy weather days and max AQI have statistically significant results, while the direct effects are insignificant. They show a mediation effect that the indirect effect explains more than half of the total effect. In detail, the changes in the share of Blacks from 1940 to 1970 indirectly influenced the air quality outcomes in 1980, and this indirect influence ran through urban size in 1970. However, there is no statistically significant result for direct or indirect effects regarding median AQI.

In Table 6.6, the total effect of the changes in the share of Blacks on AQI outcomes in 1980 has been discussed previously in Table 6.5. Unlike the urban size, when the urban shape is a mediator, the results show that the demographic shift only directly affected AQI outcomes in 1980 but not indirectly through the mediator. That is to say, the effect of the demographic shift on the AQI outcomes in 1980 is not through the changes in urban spatial dispersion.

6.3.4 Long-term Environmental Outcomes

Table 6.7: Changes in Share of Blacks (T) → Urban Size (M) → Long-term AQI Outcomes (Y)

	Median AQI 2010		**Max AQI 2010**		**Unhealthy Weather 2010**	
	(1)	(2)	(3)	(4)	(5)	(6)
Indirect effect (α_M^T * α_Y^M)	.025	.077	.025	.100[b]	-.020	.002
	(.053)	(.053)	(.049)	(.047)	(.021)	(.011)
Direct effect (α_Y^T)	.093[b]	.087[b]	.051	.030	.027[c]	.020[b]
	(.042)	(.038)	(.048)	(.052)	(.015)	(.010)
Total effect	.118[a]	.164[a]	.076[a]	.131[a]	.007	.022[a]
	(.030)	(.040)	(.030)	(.046)	(.007)	(.006)
Proportion of Indirect Effect	21.19%	46.95%	32.89%	76.34%	285.71%	9.09%
Control		X		X		X

Note: [a], [b], and [c] denote significance at the 1 percent, 5 percent, and 10 percent levels, respectively.

Table 6.8: Changes in Share of Blacks (T) → Urban Shape (M) → Long-term AQI Outcomes (Y)

	Median AQI 2010		**Max AQI 2010**		**Unhealthy Weather 2010**	
	(1)	(2)	(3)	(4)	(5)	(6)
Indirect effect (β_M^T * β_Y^M)	.017	.050	.018	.065	-.014	.001
	(.038)	(.044)	(.036)	(.051)	(.017)	(.007)
Direct effect (β_Y^T)	.100[a]	.114[a]	.059	.065	.021[b]	.021[a]
	(.030)	(.029)	(.037)	(.047)	(.010)	(.007)
Total effect	.118[a]	.164[a]	.076[a]	.131[a]	.007	.022[a]
	(.030)	(.040)	(.030)	(.046)	(.007)	(.006)
Proportion of Indirect Effect	14.41%	30.49%	23.68%	49.62%	200%	4.55%
Control		X		X		X

Note: [a], [b], and [c] denote significance at the 1 percent, 5 percent, and 10 percent levels, respectively.

Regarding the long-term effect, the total impact on Tables 6.7 and 6.8 shows that every 1% increase in the changes in the share of Blacks increases the median AQI by 0.164%. This 0.164% increase can be translated to a 0.07954 increase in median AQI in 2010, with the average median AQI in 2010 being approximately 48.5. Given that the average increase of the Black population during the second wave of the Great Migration was around 65,000, the result implies that every 10,000 increase in Black residents caused the median AQI to increase by around 1 in 2010. Compared with the median AQI performance in 1974, the impact of past demographic shifts between 1940 and 1970 on current air quality conditions in 2010 has not weakened as time passed; instead, it has maintained the same intensity as the short-term effects we discussed above. Also, the average median AQI has increased from 1980 to 2010.

Moreover, the results show that every 1% increase in the changes in the share of Blacks increases the max AQI by 0.131%. Given that the average max AQI was around 141 in 2010, and the average increase of the Black population during the second wave of the Great Migration was around 65,000, this result shows every 10,000 more Black residences caused the max AQI increase by around 3 in 2010. The impact of Black population growth on extreme AQI in 2010 is three times greater than the impact on median AQI in 2010. However, compared to the short-term effects discussed previously, the difference between

the long-term impacts on max AQI and the long-term impacts on median AQI is narrowed. This suggests that environmental strategies have improved over time to control air quality better in high-pollution areas, or relevant policies have become more effective in managing high-pollution sources. As a result, the gap between the effects of the demographic shift on max AQI and median AQI is reducing, and the average max AQI is decreasing from 1980 to 2010 as well.

Regarding the ratio of the unhealthy weather days in 2010, the result shows that every 1% increase in the changes in the share of Blacks increases this outcome by 0.022%. Given that the average ratio of unhealthy weather days in 2010 was approximately 3.57%, and the average increase of the Black population during the second wave of the Great Migration was around 65,000, the result displays that every 10,000 Black population growth causes around 0.044 more unhealthy weather days out of a year in 2010.

The underlying mechanisms for the total effect of the demographic shift on air quality performance are similar across different AQI outcomes, except for the max AQI. Specifically, the statistically significant direct effect and the insignificant indirect effect shown in tables 6.7 and 6.8 indicate that the changes in the share of Blacks directly impact the long-term air quality outcomes but not indirectly through the impact on urban spatial structures, including urban size and spatial dispersion. In other words, the demographic shift in the past has a sustained direct effect on air quality outcomes today, independent of changes in urban spatial structure caused by the population difference in the past.

Moreover, different from the median AQI and ratio of the unhealthy weather days, the max AQI represents an extreme air quality condition. The long-term results show a statistically significant indirect effect of the increasing Black population on max AQI, and the direct effect is statistically insignificant. This implies that the effect of the changes in the share of Blacks in the past on max AQI today is majority through the effect of the demographic shift on urban size[7] in the past. Compared with the result for average air quality conditions (ex., median AQI and the ratio of the unhealthy weather days), which shows a close and direct correlation to the change in population size, extreme air quality conditions in the long term, such as a day with the highest air pollution, are more depend on how the increasing Black population influence the urban horizontal expansion. In other words, the increase in the Black population in the past caused the urban horizontal expansion, which might lead to long-term spatial structure issues such as improper numbers and location of the industrial infrastructure. These issues persist today and quickly worsen

[7]When the mediator is the urban shape and outcome is max AQI in Table 6.8, there is no statistically significant direct or indirect effect. This implies no noticeable mediation effect of the changes in the share of Blacks on long-term max AQI when the observed mediator is urban shape.

the extreme air pollution conditions when occasionally extreme events happen, such as extremely high temperatures, which cause a very high air pollution level in a day or couple of days. On the other hand, unlike the extreme air quality conditions, the average air quality conditions reflect the general environmental performance, which is influenced by continuous everyday human activities but not the extreme conditions in one or couple of days. Thus, in terms of the long-term effects of average air quality conditions, the urban spatial structure is no longer a significant mediator, and the effect of the increasing Black population in the past directly influences the average air quality performance.

6.4 Discussion

In this chapter, we examined the effect of the racial composition shock between 1940 and 1970 on economic and environmental outcomes in the short and long term. We also studied the role of the urban spatial structure in 1970 as a mediator in transferring the indirect effect of the demographic shift during the second wave of the Great Migration on outcomes. Specifically, we employed an IV mediation approach to study the total impact of the racial composition shock on outcomes and the underlying indirect and direct effects when considering urban spatial structure as the mediator. Also, this IV mediation approach allows us to use the predicted Black migration as the IV to avoid biases caused by the issues of reverse causality and unobserved factors. Finally, we can capture credible results estimates to understand the causal relationship between the demographic shift and outcomes and how the racial composition shock influences the outcomes through urban size or spatial dispersion.

The total effect of the racial composition shock on economic and environmental performance answers the first main question in this chapter about how the influx of African American migrants during the second wave of the Great Migration (1940-1970) affected the short—and long-term economic and environmental outcomes. Specifically, an increase in the Black population would drive an increase in income and AQI outcomes both short-term and long-term. Recall that a higher value of AQI outcomes means a worse environmental performance. Thus, the growth of the Blacks causes a worse environmental outcome. Although the existing literature did not examine the same causal relationships as we did, they have similar results that are consistent with our results [68, 69].

Moreover, this chapter expects to further explore the complex mechanisms behind these general causal relationships by using a mediation analysis approach to reveal the truth. By adding a mediator, we can separate the influence not directly caused by independent variables on outcome variables to more accurately understand how independent variables

affect outcome variables through different paths. In detail, for the racial composition shock during the second wave of the Great Migration, we consider the importance of the urban spatial structure that was influenced by the changes in the share of Blacks during the immigration shock. Therefore, we choose the urban spatial structure as a mediator to more comprehensively reveal the complex influence mechanism of the racial composition shock on the economy and environment, answering the second main question in this chapter: How did urban size and shape play a role as a bridge in the relationship between the racial composition shock and the outcomes?

When we consider the total income in 1974 to represent the short-term economic performance, the results suggest that an increase in the urban Black population leads to an expansion of urban size, leading to a rise in income. The expansion of urban size may represent more space for economic activity, a larger housing market, and more employment opportunities, which cause an increase in income. By applying the mediator, we find that the direct effect of the changes in the share of Blacks on short-term income outcomes is not significant. This indicates that urban size as a mediator can explain most of the impact of the racial composition shock on income in the short term. In other words, the short-term income effect of the racial composition shock between 1940 and 1970 entirely depends on how the changing Black population changes the city's physical size at the end of the Great Migration. However, urban shape, which represents another dimension of urban spatial structure, is not an important mediator in transferring the effect of the changes in the share of Blacks on short-term income performance.

When the outcome variable becomes long-term income performance in 2010, and the mediator is urban size, the positive impact of the increase in the Black population between 1940 and 1970 on income in 2010 is entirely run through urban expansion. In other words, the increase in the Black population during the Great Migration drove the expansion of cities at that time, and this change in past conditions led to long-term income growth. Similar to the short-term result when the mediator is urban shape, the effect of the increasing Black population in the past on income persists until today, but this influence does not run through urban shape.

Regarding the specific value of coefficients, the long-term impact of the change in the share of Blacks during the Great Migration on the economy in 2010 is weaker than that in 1974. From this comparison, we can see that the growth in the Black population in the past rapidly affected the horizontal expansion of cities in the past, which in turn translated into short-term income growth. On the other hand, although the past racial composition shock still affects future income through the past urban size, the direct, indirect, and total effects are diminished by the complexity of the economic structure and other factors in the long term.

In urban development discussions, the economy is one dimension, while the environment represents another critical aspect. Specifically, we use the median and max AQI and the ratio of unhealthy weather days as environmental performance measures. Recall that a higher value of these variables means worse air quality performance. The results show that the negative impact of Blacks' growth on the environment in the short term highly depends on the expansion of the urban size, as evidenced by the mediation effect observed when analyzing the max AQI and the ratio of unhealthy weather days. In other words, the increase in the Black population leads to the expansion of the urban size, and this expansion leads to increased construction activity and traffic flow, and then causes deterioration of air quality, showing a significant impact on extreme air quality conditions and the ratio of unhealthy weather days. This may be because urban size expansion is usually accompanied by the development of more land for residential, commercial, and industrial uses, which increases pollution sources and, in turn, worsens air pollution immediately. However, when examining median AQI as outcomes, the mediation analysis reveals only a total effect of the racial composition shock on AQI, implying other unconsidered variables or pathways affecting the median AQI.

When we use urban shape as the mediator, the results reveal a total and direct negative effect of racial composition shock on the environment in the short term. In detail, the growth of urban Blacks caused by significant Black immigration from the South to non-southern cities might lead to increased economic activities, accompanied by more industrial pollution and car usage, negatively influencing air quality. Moreover, unlike urban size, urban shape does not play a role as the mediator in the relationship between racial composition shock and air quality outcomes in the short term. In other words, the effect of the racial composition shock on air quality outcomes does not need to proceed through urban shape. Building on our previous discussion, where urban size is used as the mediator, it becomes evident that the negative environmental impacts of the racial composition shock are primarily mediated by the expansion of urban areas compared with the urban shape.

For the long-term outcomes in terms of median AQI and the ratio of unhealthy weather days, regardless of whether the mediator is urban size or urban shape, the results show that the growth in the Black population from 1940 to 1970 had a negative total and direct impact on the environment by 2010. This long-term negative effect indicates sustained environmental degradation over time because of the racial composition shock that happened in the past, which led to a dramatic increase in the Black population in the past. However, according to the results, we did not observe any indirect effect when urban size and shape were used as mediators. This means that the mediating effect of urban size and shape on long-term environmental outcomes, particularly for median AQI and the ratio of unhealthy

weather days, are not noticeable. From the perspective of time, according to our previous discussion, urban size in 1970, caused by the growth of the Black population from 1940 to 1970, is an effective mediator for the relationship between the racial composition shock and the environmental outcome in 1980. Because this impact is short, the environmental performance is more reflected by the physical expansion of urban size. Therefore, as a mediator, urban spatial structure, especially urban size, can better transfer the influence of racial composition shock on the environmental outcome in the short term. However, by 2010, when the urban spatial structure in 1970 was used as a mediator, there was a relatively ineffective mediating effect. At this time, the impact of past racial composition shock on the environment today may not run through the changes in urban spatial structure in the past but through the increasing human activities due to the increasing Black population in the past or through other long-term factors, such as policy changes or technology development over time. In addition, the long-term max AQI performance tells a different story. Although the effect of the growth of the Black population in the past on the general air quality level today is not through the effect on urban size in the past, the potential spatial structure issues based on different urban sizes, which are caused by different Black population levels, would persist until today and affect air quality under extreme conditions.

According to the findings, we can see the dilemma between economic growth and environmental protection because of the increasing number of urban Black residents during the second wave of the Great Migration. We can tell that an increased environmental burden often accompanies economic development. Therefore, how to reduce environmental pressure while promoting economic growth is a key challenge to achieving sustainable urban development. Although this chapter focuses on the racial composition shock between 1940 and 1970 in the last century, the findings still have important implications for modern urban planners and policymakers that can help optimize urban planning to achieve SDG 11: Make cities and human settlements inclusive, safe, resilient, and sustainable [62].

As mediators, urban size plays an important role in transferring the effect of racial composition shock to economic outcomes. Also, urban size plays an important role as the mediator in transferring the impact of the racial composition shock on environmental outcomes, especially in the short term. Therefore, when policymakers plan for population mobility, the close connection between urban spatial structure and population change, especially the Black population change, and the important role of the urban size as a mediator should be recognized. We suggest that policymakers analyze the population prediction data to make corresponding plans for infrastructure construction in advance to adapt to population growth and unavoidable migration in the future. For example, policymakers can predict the migration status to recognize the place more likely to become a migration

hotspot and set the urbanization construction and regional planning according to the needs to prevent the chaotic expansion and urban sprawl caused by population increase. In this way, they can ensure short-term environmental protection while the economy develops to achieve urban sustainability. Additionally, due to the direct effect of Black population growth on outcomes, policymakers can focus on Black population redistribution and take reasonable intervention and guidance. For example, they can rationally guide existing or potential residents to low-pollution areas by providing better employment opportunities and living conditions there. As a result, this can effectively reduce the intense negative air quality impact caused by population concentration in high-pollution areas and, at the same time, keep the potential labor force and economic activities to motivate economic growth.

Chapter 7

CONCLUSION

This thesis innovatively constructed an urban spatial dataset and applied this dataset in empirical studies to reveal the relationship between the Great Migration and urban spatial structure, as well as further links to economic and environmental performance. First, we showed the historical development trajectory of urban spatial structure in the United States between 1900 and 1970. We also quantitatively examine the evolution of the urban system in America over time. Secondly, we focused on the role of the Great Migration in altering urban spatial structure in American non-southern urban areas. Thirdly, we focused on the importance of the urban spatial structure in shaping economic and environmental performance over time. Lastly, we investigated the effect of the Great Migration on economic and environmental outcomes. Also, we studied the underlying direct and indirect effect of the Great Migration on the outcomes through the effect on urban spatial structure. The following are the key findings in this Thesis for answering research questions:

- How did urban spatial structure evolve in the United States from 1900 to 1970, and what patterns characterized the growth and organization of the U.S. urban system during this period?
  - At the Syncity level, the analyses of multiple Syncities — regardless of their location in the United States or when they emerged — revealed a strong consistency in spatial development patterns. This is not to say that all Syncities develop like each other or grow at the same pace; however, there are different patterns across different regions and certain types of geography, such as being located inland versus on a coast. These patterns demonstrate the American urban spatial structure's complex but orderly development trajectory.

  - From a macro perspective, from 1900 to 1970, American Syncities showed highly regular trends in horizontal size development and shape evolution. Moreover, each decade's new clusters also show a specific pattern – they fill in between existing clusters while some expand outward.
  - The first examination of equilibrium conditions shows that CCA clusters have a strong distribution pattern in our re-defined cities – clusters near the center are more affiliated with the city and vice versa. This distribution pattern represents that city centers have relatively strong attraction and shows the trend of urban expansion from the center to the periphery.
  - The second examination of the equilibrium conditions indicates that cities generally continuously organized their overall structures over the years to minimize internal transportation costs during development.
  - The population and size of Syncities fit relatively well with Zipf's law, which indicates a hierarchical structure of Syncities. In addition, the American urban system has shown continuous improvement and is now near enough to an absolute hierarchical structure. However, it is still not perfect.

- What was the impact of the Second Wave of the Great Migration on urban spatial structure?
  - The more the Black population increased during the second wave of the Great Migration (1940-1970), the more horizontal expansion in cities would be at the end of the Great Migration (1970).
  - The more the Black population increased between 1940 and 1970, the more urban spatial dispersion would be in 1970.

- What are the economic and environmental consequences of urban spatial structure?
  - The more urban horizontal expansion was in 1970, the higher the income level would be in 1974 and 2010. Moreover, the influence on income in 2010 was smaller than in 1974.
  - The more urban spatial dispersion was in 1970, the higher the income level would be in 1974 and 2010. Similar to urban size, the magnitude of the effect on income level reduced over time.
  - The larger the urban size was in 1970, the worse the air quality would be in 1980 and 2010. Moreover, the effect of urban size on air quality performance in 2010 was smaller than in 1980.

  - The more urban spatial dispersion was in 1970, the worse the air quality would be in 1980 and 2010. Also, the magnitude of the effect reduced from 1980 to 2010.

- What were the short- and long-term economic and environmental impacts of the influx of African American migrants during the second wave of the Great Migration (1940-1970) on U.S. cities, and how did changes in urban size and spatial structure mediate the relationship between this demographic shift and its effects?

  - The more the Black population increased during the second wave of the Great Migration (1940-1970), the higher the income level would be in 1974 and 2010. In addition, the effect of the changes in the share of Blacks on income level in 2010 was smaller than that in 1974.
  - The more the Black population increased between 1940 and 1970, the worse the air quality would be in 1980 and 2010. Moreover, the effect of the changes in the share of Blacks on median AQI in 1980 has a similar magnitude compared with the effect in 2010. On the other hand, the influence on max AQI and ratio of unhealthy weather days shows a decreasing magnitude over time.
  - Black population growth between 1940 and 1970 has no direct effect on income level in 1974 and 2010 when considering urban size as the mediator. That is to say, the urban size in 1970, as a mediator, provides an important mediation effect for the effect of the Black population growth on income. In comparison, urban spatial dispersion in 1970 is not a mediator for the effect of the changes in the share of Blacks on income.
  - Regarding the short-term effect on air quality performance, the effect of Black population growth between 1940 and 1970 on air quality performance in 1980 was mainly through the urban horizontal expansion in 1970, especially when max AQI and the ratio of unhealthy weather days represent the air quality performance. In comparison, urban spatial dispersion in 1970 is not a proper mediator to transfer the influence of the change in the share of Blacks between 1940 and 1970 on air quality performance in 1980.
  - Regarding the long-term effect on air quality performance, urban size or urban spatial dispersion in 1970 does not mediate the effect of the changes in the share of Blacks between 1940 and 1970 on air quality performance in 2010. In other words, the demographic shift is more likely to influence air quality performance in the long term directly due to the changes in the share of Blacks in urban areas. However, the effect on the max AQI is an exception, especially when the urban

size is the mediator. Our finding shows that the urban size in 1970 mediates the majority effect of the Black population growth between 1940 and 1970 on max AQI in 2010. This indicates that the issues of urban spatial structure caused by the horizontal expansion in the past, which was driven by the increase in the Black population in the past, persist until today and transfer the effect of the changes in the share of Blacks in the past on air quality in extreme cases today.

These findings have important implications for current and future urban planning. Understanding the rules of urban spatial structure and historical migration patterns can help planners formulate more rational urban development strategies to achieve Sustainable Development Goal (SDG) 11: "Make cities and human settlements inclusive, safe, resilient and sustainable"[1] [62]. Specifically, this thesis provides a new historical urban spatial dataset in America and provides urban planners or policymakers with a novel lens for defining urban boundaries using the City Clustering Algorithm (CCA). This novel method can help planners detect real development needs for sustainable cities without bias from misleading subjective administrative boundaries.

In addition, empirical studies show a correlation between the changes in the share of Blacks and urban spatial structure in America, which further correlates to economic and environmental performance. The lessons from the empirical studies first show urban planners the importance of population redistribution in formulating urban development plans, especially regarding migrants of different races with different cultures and life backgrounds. Moreover, when planners make strategies and policies, they should pay attention to the significant role of urban spatial structure and the changes in population on economic and environmental performance, thinking about both economic growth and environmental degradation. Also, our findings highlight the importance of urban horizontal expansion as a mediator, and planners should especially consider the impact of population flow on urban size when they regulate the impact of urban population flow on the economy and the environment. Planners can rationally use this mediating effect to optimize policies to better control the balance between economic and environmental outcomes.

Building on all insights in this thesis, the following are two sections, including a discussion of the main limitation of this thesis and potential future research that could further expand upon this thesis.

[1] This thesis relates directly to SDG 11.3: "By 2030, enhance inclusive and sustainable urbanization and capacity for participatory, integrated and sustainable human settlement planning and management in all countries.", and SDG 11.6: "By 2030, reduce the adverse per capita environmental impact of cities, including by paying special attention to air quality and municipal and other waste management." [62].

7.1 Limitations of the Study

7.1.1 Limitation of the Data Availability

The main limitation of this thesis is that there are relatively few observations in the final dataset. The main reason for this limitation is the incompleteness of historical data. To study the correlation between the Black population redistribution and urban spatial structure during the second wave of the Great Migration, we need to make sure the urban Black population data in 1940 matches the data in 1970 to calculate the growth of the Black population between 1940 and 1970. Currently, 711 non-Southern cities have data on the urban Black population in 1940, and 891 non-Southern cities have data in 1970. However, these data do not cover all non-southern cities, and the number of cities with available data on the Black population in both 1940 and 1970 is only 296. In other words, there are only 296 observations that we can use to calculate the changes in the share of Blacks between 1940 and 1970.

The second reason for fewer observations is that different variables have different geographical levels, and the boundaries at various geographical levels intersect or contain with each other. For example, the urban black population data has a different geographic level from urban spatial data: Black population data are at the official administrative city level, while urban spatial data are at the redefined Syncity level. Some boundaries of urban areas at different geographical levels may intersect, or some syncities may contain multiple small cities. Therefore, we constructed a city-syncity crosswalk. This crosswalk can satisfy the requirement that each Syncity can match one or a couple of cities, and the cities are not repeated in the dataset. Using this crosswalk, we can match the Black population at the administrative city level with the spatial data at the Syncity level. Then, we can collapse the Black population data at the city level to the Syncity level. At this point, the 296 observations at the city level mentioned earlier collapsed to 168 at the syncity level. According to this method, other city-level data, such as income level, is also collapsed to the Syncity level.

Some key variables lack city-level data, including air quality data in 1980 and 2010, manufacturing labor share, and education data. For these variables, we can only obtain county-level data. Therefore, we constructed a county-Syncity crosswalk to match the data at different geographic levels. This crosswalk is constructed based on the same principle as the city-Syncity crosswalk. When matching city and Syncity, a Syncity contains one or several administrative cities, but cities are too small to contain any Syncity. However, when matching county and Syncity, some counties may include multiple syncities while other syncities may consist of several counties. This means we cannot collapse the data

to a specific geographic level. To address this issue, we constructed a new set of codes representing geographic locations for county-Syncity unions. This union could be one Syncity that overlaps with one county, several syncities overlap with the same county, or several counties overlap with the same Syncity. Then, we can collapse the county-level data to this new county-Syncity union level. At the same time, the previously unified Syncity level data can be collapsed to the same county-Syncity union level. So, we can keep all data in the final dataset at the same geographical level to ensure the consistency of the study. As a result, the final county-Syncity union level dataset has 160 observations[2].

Lastly, among these 160 observations, we excluded 27 with a negative total predicted Black migration. These negative values indicate that 27 locations are predicted to have a certain number of Black people leaving the North or West to the South between 1940 and 1970. However, our research primarily focuses on the impact on Northern and Western cities when Southern Blacks migrate to the North and West but not when Northern or Western Blacks leave. This means these 27 observations do not apply to our study. Moreover, When the total predicted migration includes both positive and negative values, considering both the predicted inflows to some non-southern cities and the outflows from other non-southern cities. The mechanisms by which outflows impact non-southern cities may differ significantly from those of inflows. As a result, as the IV, the data of migration outflows may introduce noise, weakening the explanatory power of the IV when it is used for a complex IV mediation analysis. Thus, it is reasonable to exclude these 27 observations.

7.1.2 Limitation of Air Quality Data

In this Thesis, as an essential variable representing environmental performance, AQI has limitations. Many objective factors, based on the geographical location of air quality monitoring stations, may impact the accuracy of air quality data. For example, the air quality data from the EPA may not accurately reflect the local air quality situation because meteorological conditions have a particular impact on the trajectory of air pollutants. In

[2]There are 160 available observations for the changes in Black population, predicted Black migration, and urban spatial data. This means that 160 available observations can be used to study the relationship between the change in the Black population and urban spatial structure. However, there may be fewer available observations for studying other relationships. For example, the available historical environment data is limited, which means only limited urban spatial data matched with environmental data will be kept when studying the relationship between urban spatial structure and environmental performance. In other words, the number of observations for studying this relationship will be less than 160, depending on how many observations have available environmental data.

detail, before the monitoring station detects air pollutants, the wind may carry them away from the source of the pollution, making the monitoring station not accurately reflect the local air pollutant emission level. EPA has made an effort to consider the weather, such as wind speed and direction, as an essential evaluation factor; they also pay attention to the impact of physical features of the land on pollution distribution and the contribution of potential emissions [65]. However, we should still be cautious and understand that potential biases may exist, and the limitation may lead to inaccuracy in our final results.

7.1.3 Limitation of Mismatch Between Data Boundaries

The mismatch of data boundaries will bring limitations to this thesis. In the study, our urban spatial dataset, calculated according to a redefined urban area, is mainly based on the agglomeration of the building footprint. Other variables include the Black population, income, and air quality; their data are obtained based on administrative boundaries, which rely more on political or historical factors. Therefore, although the new urban spatial dataset can avoid research biases caused by the official administrative boundaries, the inconsistency in the boundary definition between urban spatial variables and other variables may lead to a spatial mismatch among datasets and then affect the accuracy of the analysis of the relationship between different variables.

One solution to avoid economic, environmental, or population data being inaccurately applied to new urban boundaries is to use data in a smaller level of the spatial units. For example, population data that was collected per grid cell can be used to match the newly defined urban areas better. However, the available historical data do not provide such small spatial units. In this thesis, we have used the data from the smallest geographical unit that can be found, but the boundary mismatch cannot be avoided entirely now. In the future, we expect to have the opportunity to see more refined raw data to reduce the potential overestimation or underestimation of economic or environmental impacts in this Thesis.

7.2 Future Research

This thesis focuses on the second wave of the Great Migration. However, the first wave is also interesting because a significant agriculture shock occurred in Southern America that may have significantly influenced the patterns of Black migration at that time. This agricultural shock is the Boll Weevil infestation shock. The Boll Weevil is a tiny insect

that feeds on cotton. Due to its rapid and long-distance movement, it caused widespread and severe damage to cotton crops. 1892, the Boll Weevil entered the United States in Brownsville, Texas. By 1920, all cotton-growing areas in the southern U.S. were infested. This agricultural shock severely destroyed the Southern cotton economy, leading to a wave of migration from the Southern to non-Southern urban areas, especially among African Americans. In the late 19th and early 20th centuries, even though slavery had been abolished, the Southern cotton industry still heavily relied on cheap Black labor. The agricultural impact of the Boll Weevil on Southern cotton farming directly affected these cotton workers' livelihoods. As the cotton industry suffered, many African American cotton workers faced unemployment. Although no longer slaves, these workers had no land or property and lacked the funds to return to their distant hometowns. Thus, migrating from farms and small towns in Southern areas to non-southern urban areas to search for new opportunities was a relatively better option for them. Therefore, in future research, exploring the first wave of the Great Migration related to such agriculture shock as the push factor for the Black migration from South to North could provide valuable insights.

The development path of the urban spatial structure indicates a transformation in the trend of the cohesion index, which declined rapidly after 1960 at the national level. This change suggests that starting from 1960, the average distance between CCA clusters began to decrease. There may be some reasons for this historical change in the urban spatial patterns. Although this thesis did not research these reasons, our findings provide a new direction for future research to understand historical changes in urban spatial patterns. We suggest that by taking key policies and projects in the 1950s as the cases[3], future research could explain the shift of urban spatial patterns from less compact to more compact. By analyzing past experiences, urban planners can understand how historical policies or projects were reflected in urban structures while avoiding the same mistakes or replicating the successful strategies for current urban planning.

This thesis only focuses on internal migration in the United States, but other types of migration might also have relationships with the urban spatial structure. For example, future work could look at the relationships between international immigration and the spatial structure of American cities. Therefore, the work can study the contributions of migrants from different cultures and backgrounds to urban spatial development and how different urban spatial structures attract populations with different cultural backgrounds.

Another area for future work could examine how migration influences cities of different sizes - large, medium, small, and newly emerging cities - and whether immigration impacts

[3]The highway project was implemented in the 1950s and 1960s, accelerating American urban planning and construction.

cities at various stages of development differently. On the other hand, the opposite direction of the relationship - how cities in different development stages influence immigration differently - is also interesting.

Additionally, historical immigration affects cities' initial development orientation, and we suspect that this orientation sets the conditions for subsequent urban spatial development. Thus, how past immigration reshaped the current urban spatial structure is also an interesting and meaningful direction for future research.

Lastly, different urban administrations may have different policies toward migrants. When migrants aim at a particular city, such a policy affects immigration flow. Thus, for future research, it is interesting to see how different policies in different cities affect the impact of immigration on urban spatial structure.

References

[1] AirNow. Air Quality Index (AQI) Basics.

[2] Alexander Hill. The great U.S. emissions migration. *Journal of Environmental Economics and Management*, 127:103018, September 2024.

[3] Alexander Sahn. Racial Diversity and Exclusionary Zoning: Evidence from the Great Migration. January 2021.

[4] Alferdteen Harrison. *Black Exodus: The Great Migration from the American South*. Univ. Press of Mississippi, 1991.

[5] Ana I. Moreno-Monroy, Marcello Schiavina, and Paolo Veneri. Metropolitan areas in the world. Delineation and population trends. *Journal of Urban Economics*, 125:103242, September 2021.

[6] Austin Zeiderman. RURALIZING THE CITY: THE GREAT MIGRATION AND ENVIRONMENTAL REHABILITATION IN BALTIMORE, MARYLAND. *Identities*, 13(2):209–235, 2006. Publisher: Taylor & Francis Group.

[7] Brian McCammack. *Landscapes of hope: nature and the Great Migration in Chicago*. Harvard University Press, Cambridge, Massachusetts, 2017.

[8] Brian Stone. Urban sprawl and air quality in large US cities. *Journal of Environmental Management*, 86(4):688–698, March 2008.

[9] Christian Dippel, Andreas Ferrara, and Stephan Heblich. Causal mediation analysis in instrumental-variables regressions. *The Stata Journal*, 20(3):613–626, September 2020. Publisher: SAGE Publications.

[10] Christian Dippel, Robert Gold, Stephan Heblich, and Rodrigo Pinto. Instrumental Variables and Causal Mechanisms: Unpacking The Effect of Trade on Workers and Voters. *NBER Working Paper Series*, pages 23209–, 2017. Place: Cambridge Publisher: National Bureau of Economic Research, Inc.

[11] Daniel Arribas-Bel, M. -À. Garcia-López, and Elisabet Viladecans-Marsal. Building(s and) cities: Delineating urban areas with a machine learning algorithm. *Journal of Urban Economics*, 125:103217, September 2021.

[12] Dongyang Zhang, Qunxi Kong, and Meng Shen. Does polycentric spatial structure narrow the urban-rural income gap? – Evidence from six urban clusters in China. *China economic review*, 80:101999–, 2023. Publisher: Elsevier Inc.

[13] Douglas S. Massey. *American apartheid segregation and the making of the underclass*. ACLS Humanities E-Book (Series). Harvard University Press, Cambridge, Mass, 1993.

[14] Edward L. Glaeser and Matthew E. Kahn. The greenness of cities: Carbon dioxide emissions and urban development. *Journal of Urban Economics*, 67(3):404–418, May 2010.

[15] Elizabeth Anne Martin. *Detroit and the Great Migration, 1916-1929*. Michigan Historical Collections/Bentley Historical Library, University of Michigan, 1993. Google-Books-ID: qbThAAAAMAAJ.

[16] Ellora Derenoncourt. Can You Move to Opportunity? Evidence from the Great Migration. *American Economic Review*, 112(2):369–408, February 2022.

[17] Enyu Zeng, Pengfei Li, and Jing Zhou. Water environmental efficiency in different urban spatial structure patterns: Evidence from a panel of Chinese urban districts. *Journal of cleaner production*, 379:134834–, 2022. Publisher: Elsevier Ltd.

[18] EPA. AirData.

[19] Eric Arnesen. *Black protest and the Great Migration: A brief history with documents*. 2003.

[20] Erin Royston Battat. *Ain't got no home: America's great migrations and the making of an interracial left*. The University of North Carolina Press, Chapel Hill, 2014.

[21] Gilles Duranton and Diego Puga. Chapter 8 - Urban Land Use. In Gilles Duranton, J. Vernon Henderson, and William C. Strange, editors, *Handbook of Regional and Urban*

Economics, volume 5 of *Handbook of Regional and Urban Economics*, pages 467–560. Elsevier, January 2015.

[22] Isabel Wilkerson. *The warmth of other suns: the epic story of America's great migration*. Vintage Books, a division of Random House, Inc., New York, first vintage books edition. edition, 2011.

[23] J. Trent Alexander, Christine Leibbrand, Catherine Massey, and Stewart Tolnay. Second-Generation Outcomes of the Great Migration. *Demography*, 54(6):2249–2271, November 2017.

[24] J. Vernon Henderson, Dzhamilya Nigmatulina, and Sebastian Kriticos. Measuring urban economic density. *Journal of Urban Economics*, 125:103188, September 2021.

[25] James N. Gregory. *The Southern Diaspora: How the Great Migrations of Black and White Southerners Transformed America*. Univ of North Carolina Press, May 2006. Google-Books-ID: 6lLqCQAAQBAJ.

[26] Jeffrey Lin and Ferdinand Rauch. What future for history dependence in spatial economics? *Regional Science and Urban Economics*, 94:103628, May 2022.

[27] Joe William Trotter. *The Great Migration in Historical Perspective: New Dimensions of Race, Class, and Gender*. Indiana University Press, November 1991. Google-Books-ID: 1WdRAqTNSRcC.

[28] Johannes H. Uhl, Dylan S. Connor, Stefan Leyk, and Anna E. Braswell. Urban Spatial Development in the United States from 1910 to 2010: A Novel Data-Driven Perspective, February 2020.

[29] Johannes H. Uhl and Stefan Leyk. Historical built-up areas (BUA) - gridded surfaces for the U.S. from 1810 to 2015, January 2021.

[30] John B. Vogler and Jelena Vukomanovic. Trends in United States Human Footprint Revealed by New Spatial Metrics of Urbanization and Per Capita Land Change. *Sustainability*, 13(22):12852, January 2021. Number: 22 Publisher: Multidisciplinary Digital Publishing Institute.

[31] John R. Logan. *Urban fortunes: the political economy of place*. University of California Press, Berkeley, CA, 1987.

[32] Jonathan I. Dingel, Antonio Miscio, and Donald R. Davis. Cities, lights, and skills in developing economies. *Journal of Urban Economics*, 125:103174, September 2021.

[33] Jun Zhang, Xiong He, Xiao-Die Yuan, and Bing Xue. Research on the relationship between Urban economic development level and urban spatial structure—A case study of two Chinese cities. *PloS one*, 15(7):e0235858–e0235858, 2020. Place: San Francisco Publisher: Public Library of Science.

[34] Katherine J. Curtis White. Women in the Great Migration: Economic Activity of Black and White Southern-Born Female Migrants in 1920, 1940, and 1970. *Social Science History*, 29(3):413–455, October 2005.

[35] Kenneth L. Sokoloff and Stanley L. Engerman. Institutions, Factor Endowments, and Paths of Development in the New World. *Journal of Economic Perspectives*, 14(3):217–232, September 2000.

[36] Kevin Fox Gotham. Urban Space, Restrictive Covenants and the Origins of Racial Residential Segregation in a US City, 1900–50. *International Journal of Urban and Regional Research*, 24(3):616–633, 2000. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1468-2427.00268.

[37] Kurt Schlichting, Peter Tuckel, and Richard Maisel. Residential Segregation and the Beginning of the Great Migration of African Americans to Hartford, Connecticut A GIS-Based Analysis. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 39(3):132–144, 2006. Publisher: Heldref.

[38] Lawrence R. Rodgers. *Canaan bound: the African-American great migration novel*. University of Illinois Press, Urbana ;, 1997.

[39] Leah Platt Boustan. Was Postwar Suburbanization "White Flight"? Evidence from the Black Migration. *The Quarterly Journal of Economics*, 125(1):417–443, 2010. Publisher: Oxford University Press.

[40] Luís M. A. Bettencourt. *Introduction to Urban Science: Evidence and Theory of Cities as Complex Systems*. The MIT Press, August 2021.

[41] Maarten Bosker. City origins. *Regional Science and Urban Economics*, 94(C), 2022. Publisher: Elsevier.

[42] Marco Gonzalez-Navarro and Matthew A. Turner. Subways and urban growth: Evidence from earth. *Journal of Urban Economics*, 108:85–106, November 2018.

[43] Marco Tabellini. Racial Heterogeneity and Local Government Finances: Evidence from the Great Migration, January 2020.

[44] Mariaflavia Harari. Cities in Bad Shape: Urban Geometry in India. *The American economic review*, 110(8):2377–2421, 2020. Publisher: American Economic Association.

[45] N. K. Diem, P. K. Diem, P. M. Thien, and N. K. Downes. Monitoring Urban Heat Island Spatial Variability over Urban Structure Types – A Case Study from a Fast-growing City in the Vietnamese Mekong Delta. *IOP conference series. Earth and environmental science*, 1345(1):12001–, 2024. Place: Bristol Publisher: IOP Publishing.

[46] Nathan Nunn and Diego Puga. Ruggedness: The Blessing of Bad Geography in Africa. 2012.

[47] Nicholas Lemann. *The Promised Land: The Great Black Migration and how it Changed America*. Vintage Books, 1991. Google-Books-ID: i6TS31JqdIoC.

[48] Paul A. David. Clio and the Economics of QWERTY. *The American Economic Review*, 75(2):332–337, 1985. Publisher: American Economic Association.

[49] Remi Jedwab, Noel D. Johnson, and Mark Koyama. The Economic Impact of the Black Death. *Journal of Economic Literature*, 60(1):132–178, March 2022.

[50] Richard Hornbeck and Suresh Naidu. When the Levee Breaks: Black Migration and Economic Development in the American South. *The American Economic Review*, 104(3):963–990, 2014. Publisher: American Economic Association.

[51] Richard M. Daley. Street and Site Plan Design Standards, April 2007.

[52] Seyed Sajjad Abdollahpour, Meng Qi, Huyen TK Le, and Steve Hankey. Urban spatial structure and air quality in the United States: Evidence from a longitudinal approach. *Environment international*, 190:108871–, 2024. Publisher: Elsevier Ltd.

[53] Stefan Leyk, Johannes H. Uhl, Dylan S. Connor, Anna E. Braswell, Nathan Mietkiewicz, Jennifer K. Balch, and Myron Gutmann. Two centuries of settlement and urban development in the United States. *Science Advances*, 6(23):eaba2937, June 2020. Publisher: American Association for the Advancement of Science.

[54] Steven Manson, Jonathan Schroeder, David Van Riper, Katherine Knowles, Tracy Kugler, Finn Roberts, and Steven Ruggles. IPUMS National Historical Geographic Information System: Version 19.0, 2024.

[55] Steven Ruggles, Matt A. Nelson, Matthew Sobek, Catherine A. Fitch, Ronald Goeken, J. David Hacker, Evan Roberts, and J. Robert Warren. IPUMS Ancestry Full Count Data: Version 4.0, 2024.

[56] Stewart E. Tolnay. The African American "Great Migration" and Beyond. *Annual Review of Sociology*, 29(Volume 29, 2003):209–232, August 2003. Publisher: Annual Reviews.

[57] Steffen Kriewald. Introduction to the City Clustering Algorithm. December 2019.

[58] Thomas J. Sugrue. *The origins of the urban crisis: race and inequality in postwar Detroit*. Princeton studies in American politics. Princeton University Press, Princeton, N.J, 1st paperback ed. edition, 1998.

[59] Treb Allen. The Topography of Nations, October 2023.

[60] Treb Allen and Dave Donaldson. Persistence and path dependence: A primer. *Regional Science and Urban Economics*, 94:103724, May 2022.

[61] Tyler T. Reny and Benjamin J. Newman. Protecting the Right to Discriminate: The Second Great Migration and Racial Threat in the American West. *The American Political Science Review*, 112(4):1104–1110, November 2018. Num Pages: 7 Place: Washington, United Kingdom Publisher: Cambridge University Press.

[62] United Nations. Transforming our World: The 2030 Agenda for Sustainable Development, 2016.

[63] United States Bureau of the Census. County and City Data Book [United States] Consolidated File: City Data 1944-1977, May 2000.

[64] United States Bureau of the Census. County and City Data Book [United States] Consolidated File: County Data, 1947-1977, September 2012.

[65] U.S. Environmental Protection Agency. Initial Area Designations for the National Ambient Air Quality Standards, July 2017.

[66] Virgilio Galdo, Yue Li, and Martin Rama. Identifying urban areas by combining human judgment and machine learning: An application to India. *Journal of Urban Economics*, 125:103229, September 2021.

[67] Wei Xiao, Wenhua Liu, and Chunzhi Li. Can the urban spatial structure accelerate urban employment growth? Evidence from China. *Growth and change*, 53(4):1668–1693, 2022. Place: Lexington Publisher: Blackwell Publishing Ltd.

[68] William J. Collins. The Great Migration of Black Americans from the US South: A guide and interpretation. *Explorations in Economic History*, 80:101382, April 2021.

[69] William J. Collins and Marianne H. Wanamaker. Selection and Economic Gains in the Great Migration of African Americans: New Evidence from Linked Census Data. *American Economic Journal: Applied Economics*, 6(1):220–252, 2014. Publisher: American Economic Association.

[70] Xuecao Li, Yuyu Zhou, Mohamad Hejazi, Marshall Wise, Chris Vernon, Gokul Iyer, and Wei Chen. Global urban growth between 1870 and 2100 from integrated high resolution mapped data and urban dynamic modeling. *Communications earth & environment*, 2(1):1–10, 2021. Place: London Publisher: Nature Publishing Group.

[71] Yali Wei, Ying Li, Siying Wang, Junyi Wang, and Yu Zhu. Research on the Spatial Expansion Characteristics and Industrial and Policy Driving Forces of Chengdu–Chongqing Urban Agglomeration Based on NPP-VIIRS Night Light Remote Sensing Data. *Sustainability (Basel, Switzerland)*, 15(3):2188–, 2023. Place: Basel Publisher: MDPI AG.

[72] Yingcheng Li. Towards concentration and decentralization: The evolution of urban spatial structure of Chinese cities, 2001–2016. *Computers, environment and urban systems*, 80:101425–, 2020. Place: Oxford Publisher: Elsevier Ltd.

[73] Yong Liu, Xiaolan Chen, and Dayong Liu. How Does Urban Spatial Structure Affect Economic Growth? Evidence from Landsat Data in China. *Journal of economic issues*, 54(3):798–812, 2020. Publisher: Routledge.

[74] Yoonjung Ahn, Stefan Leyk, and Johannes H. Uhl. Historical built-up intensity layer (BUI) Version II - gridded surfaces for the U.S. from 1810 to 2020, February 2024.

[75] Yu Zhang, Yuchen Wang, Nan Ding, and Xiaoyan Yang. Assessing the Contributions of Urban Green Space Indices and Spatial Structure in Mitigating Urban Thermal Environment. *Remote sensing (Basel, Switzerland)*, 15(9):2414–, 2023. Place: Basel Publisher: MDPI AG.

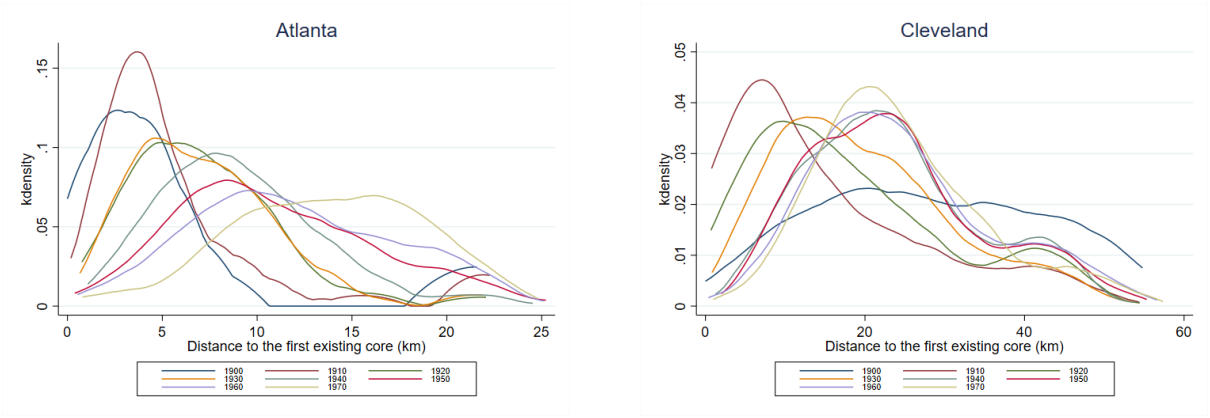
[76] Yves Zenou and Nicolas Boccard. Racial Discrimination and Redlining in Cities. *Journal of Urban Economics*, 48(2):260–285, September 2000.

APPENDICES

Appendix A

Newly Formed CCA Clusters Distribution of Selected Syncities

Following are the results of the Kdensity curve of the newly formed CCA clusters in Atlanta and Cleveland, respectively. They show a similar idea compared to the national-level result. The curve for 1900 in Atlanta has a big bump and a small bump at the distance of 5 kilometers and 20 kilometers, respectively. The peaks of the curves for the following decades are shifting out to the right and filling in the space between the two obvious peaks in 1900, showing that the newly formed clusters in Atlanta kept filling in the spaces between the older clusters. Additionally, Cleveland, as a city that borders Lake Erie, only had a big flat bump in 1900, showing a relatively even distribution of the CCA clusters in 1900. This is slightly different from the distribution in 1900 at the national-wide level. However, the curves starting from 1910 all show two peaks (a high one and a low one), and the higher peaks of the curves of the later decades are trying to fill in between the two peaks in the previous decades. This reflects that Cleveland also follows the same rules as the rules we found in the national-level results – the newly formed clusters are filling in the spaces between the older ones.



(a) (b)

Figure A.1: Two Examples of New CCA Clusters Distribution: Kdensity

Appendix B

Full Results of the Relationship Between Urban Spatial Structure and Outcomes

B.1 Effects of Urban Spatial Structure on Economic Performance: Short-term

Table B.1: [Dependent: Ln Total Income 1974]

	(1)	(2)	(3)
	(IV)Second Stage		**OLS**
Ln BUA 1970 (γ_Y^T)	1.082[a]	1.156[a]	1.057[a]
	(.281)	(.329)	(.208)
	(IV)First Stage		
Ln Predicted BUA 1970	.874[a]	.845[a]	
	(.029)	(.031)	
Observations	114	114	114
Cragg-Donald Wald F statistic	1079.129	887.943	
Control		X	

Note: [a], [b], and [c] denote significance at the 1 percent, 5 percent, and 10 percent levels, respectively.

Table B.2: [Dependent: Ln Total Income 1974]

	(1)	(2)	(3)
	(IV)Second Stage		**OLS**
Ln Cohesion Index 1970 (θ_Y^T)	3.864[a]	4.670[b]	1.002[a]
	(1.104)	(1.985)	(.203)
	(IV)First Stage		
Ln Predicted Cohesion Index 1970	.616[a]	.499[a]	
	(.147)	(.196)	
Observations	114	114	114
Cragg-Donald Wald F statistic	23.491	12.989	
Control		X	

Note: [a], [b], and [c] denote significance at the 1 percent, 5 percent, and 10 percent levels, respectively.

B.2 Effects of Urban Spatial Structure on Economic Performance: Long-term

Table B.3: [Dependent: Ln Total Income 2010]

	(1)	(2)	(3)
	(IV)Second Stage		**OLS**
Ln BUA 1970 (γ_Y^T)	.579[a]	.643[a]	.652[a]
	(.101)	(.100)	(.103)
	(IV)First Stage		
Ln Predicted BUA 1970	.872[a]	.862[a]	
	(.035)	(.041)	
Observations	74	74	74
Cragg-Donald Wald F statistic	699.206	512.139	
Control		X	

Note: [a], [b], and [c] denote significance at the 1 percent, 5 percent, and 10 percent levels, respectively.

Table B.4: [Dependent: Ln Total Income 2010]

	(1)	(2)	(3)
	(IV)Second Stage		**OLS**
Ln Cohesion Index 1970 (θ_Y^T)	2.468[a]	2.465[b]	.916[a]
	(.596)	(1.042)	(.192)
	(IV)First Stage		
Ln Predicted Cohesion Index 1970	.546[a]	.483[b]	
	(.168)	(.235)	
Observations	74	74	74
Cragg-Donald Wald F statistic	12.554	8.898	
Control		X	

Note: [a], [b], and [c] denote significance at the 1 percent, 5 percent,t, and 10 percent levels, respectively.

B.3 Effects of Urban Spatial Structure on Environmental Performance: Short-term

Table B.5: [Dependent: Ln Median AQI 1980]

	(1)	(2)	(3)
	(IV)Second Stage		**OLS**
Ln BUA 1970 (γ_Y^T)	.121[a]	.165[a]	.125[a]
	(.029)	(.036)	(.029)
	(IV)First Stage		
Ln Predicted BUA 1970	.873[a]	.843[a]	
	(.040)	(.043)	
Observations	93	93	93
Cragg-Donald Wald F statistic	612.345	505.980	
Control		X	

Note: [a], [b], and [c] denote significance at the 1 percent, 5 percent, and 10 percent levels, respectively.

Table B.6: [Dependent: Ln Median AQI 1980]

	(1)	(2)	(3)
	(IV)Second Stage		**OLS**
Ln Cohesion Index 1970 (θ_Y^T)	.340[a]	.475[c]	.069[c]
	(.132)	(.267)	(.039)
	(IV)First Stage		
Ln Predicted Cohesion Index 1970	.776[a]	.583[b]	
	(.188)	(.263)	
Observations	93	93	93
Cragg-Donald Wald F statistic	21.632	10.110	
Control		X	

Note: [a], [b], and [c] denote significance at the 1 percent, 5 percent, and 10 percent levels, respectively.

Table B.7: [Dependent: Ln Max AQI 1980]

	(1)	(2)	(3)
	(IV)Second Stage		**OLS**
Ln BUA 1970 (γ_Y^T)	.132[a]	.140[a]	.118[a]
	(.026)	(.030)	(.026)
	(IV)First Stage		
Ln Predicted BUA 1970	.873[a]	.843[a]	
	(.040)	(.043)	
Observations	93	93	93
Cragg-Donald Wald F statistic	612.345	505.980	
Control		X	

Note: [a], [b], and [c] denote significance at the 1 percent, 5 percent, and 10 percent levels, respectively.

Table B.8: [Dependent: Ln Max AQI 1980]

	(1)	(2)	(3)
	(IV)Second Stage		**OLS**
Ln Cohesion Index 1970 (θ_Y^T)	.298[a]	.400[b]	.024
	(.096)	(.194)	(.027)
	(IV)First Stage		
Ln Predicted Cohesion Index 1970	.776[a]	.583[b]	
	(.188)	(.263)	
Observations	93	93	93
Cragg-Donald Wald F statistic	21.632	10.110	
Control		X	

Note: [a], [b], and [c] denote significance at the 1 percent, 5 percent, and 10 percent levels, respectively.

Table B.9: [Dependent: Ln Ratio of unhealthy weather days 1980]

	(1)	(2)	(3)
	(IV)Second Stage		**OLS**
Ln BUA 1970 (γ_Y^T)	.039[a]	.053[a]	.034[a]
	(.008)	(.011)	(.008)
	(IV)First Stage		
Ln Predicted BUA 1970	.873[a]	.843[a]	
	(.040)	(.043)	
Observations	93	93	93
Cragg-Donald Wald F statistic	612.345	505.980	
Control		X	

Note: [a], [b], and [c] denote significance at the 1 percent, 5 percent, and 10 percent levels, respectively.

Table B.10: [Dependent: Ln Ratio of unhealthy weather days 1980]

	(1)	(2)	(3)
	(IV)Second Stage		**OLS**
Ln Cohesion Index 1970 (θ_Y^T)	.105[a]	.164[c]	.017
	(.039)	(.085)	(.011)
	(IV)First Stage		
Ln Predicted Cohesion Index 1970	.776[a]	.583[b]	
	(.188)	(.263)	
Observations	93	93	93
Cragg-Donald Wald F statistic	21.632	10.110	
Control		X	

Note: [a], [b], and [c] denote significance at the 1 percent, 5 percent, and 10 percent levels, respectively.

B.4 Effects of Urban Spatial Structure on Environmental Performance: Long-term

Table B.11: [Dependent: Ln Median AQI 2010]

	(1)	(2)	(3)
	(IV)Second Stage		**OLS**
Ln BUA 1970 (γ_Y^T)	.070[a]	.100[a]	.058[a]
	(.022)	(.025)	(.019)
	(IV)First Stage		
Ln Predicted BUA 1970	.872[a]	.841[a]	
	(.031)	(.034)	
Observations	103	103	103
Cragg-Donald Wald F statistic	897.567	731.096	
Control		X	

Note: [a], [b], and [c] denote significance at the 1 percent, 5 percent, and 10 percent levels, respectively.

Table B.12: [Dependent: Ln Median AQI 2010]

	(1)	(2)	(3)
	(IV)Second Stage		**OLS**
Ln Cohesion Index 1970 (θ_Y^T)	.239[a]	.347[b]	.049[b]
	(.085)	(.149)	(.024)
	(IV)First Stage		
Ln Predicted Cohesion Index 1970	.608[a]	.498[b]	
	(.155)	(.212)	
Observations	103	103	103
Cragg-Donald Wald F statistic	20.530	11.580	
Control		X	

Note: [a], [b], and [c] denote significance at the 1 percent, 5 percent, and 10 percent levels, respectively.

Table B.13: [Dependent: Ln Max AQI 2010]

	(1)	(2)	(3)
	(IV)Second Stage		**OLS**
Ln BUA 1970 (γ_Y^T)	.081[a]	.081[a]	.074[a]
	(.017)	(.023)	(.015)
	(IV)First Stage		
Ln Predicted BUA 1970	.872[a]	.841[a]	
	(.031)	(.034)	
Observations	103	103	103
Cragg-Donald Wald F statistic	897.567	731.096	
Control		X	

Note: [a], [b], and [c] denote significance at the 1 percent, 5 percent, and 10 percent levels, respectively.

Table B.14: [Dependent: Ln Max AQI 2010]

	(1)	(2)	(3)
	(IV)Second Stage		**OLS**
Ln Cohesion Index 1970 (θ_Y^T)	.229[a]	.251[b]	.021
	(.074)	(.128)	(.021)
	(IV)First Stage		
Ln Predicted Cohesion Index 1970	.608[a]	.498[b]	
	(.155)	(.212)	
Observations	103	103	103
Cragg-Donald Wald F statistic	20.530	11.580	
Control		X	

Note: [a], [b], and [c] denote significance at the 1 percent, 5 percent, and 10 percent levels, respectively.

Table B.15: [Dependent: Ln Ratio of unhealthy weather days 2010]

	(1)	(2)	(3)
	(IV)Second Stage		**OLS**
Ln BUA 1970 (γ_Y^T)	.013[a]	.019[a]	.010[a]
	(.003)	(.006)	(.002)
	(IV)First Stage		
Ln Predicted BUA 1970	.872[a]	.841[a]	
	(.031)	(.034)	
Observations	103	103	103
Cragg-Donald Wald F statistic	897.567	731.096	
Control		X	

Note: [a], [b], and [c] denote significance at the 1 percent, 5 percent, and 10 percent levels, respectively.

Table B.16: [Dependent: Ln Ratio of unhealthy weather days 2010]

	(1)	(2)	(3)
	(IV)Second Stage		**OLS**
Ln Cohesion Index 1970 (θ_Y^T)	.042[a]	.063[b]	.008[c]
	(.014)	(.026)	(.004)
	(IV)First Stage		
Ln Predicted Cohesion Index 1970	.608[a]	.498[b]	
	(.155)	(.212)	
Observations	103	103	103
Cragg-Donald Wald F statistic	20.530	11.580	
Control		X	

Note: [a], [b], and [c] denote significance at the 1 percent, 5 percent, and 10 percent levels, respectively.

Appendix C

Full Results of the IV Mediation Analysis

Table C.1: Changes in Share of Blacks (T) → Urban Size (M) → Short-term Income Outcomes (Y)

	(1)	(2)
Model 1 [Dependent: Ln BUA 1970]		
	Second Stage	
Ln Changes in Share of Blacks 1940-1970 (α_M^T)	1.322[a]	1.627[a]
	(.169)	(.229)
	First Stage	
Ln Predicted Total Black Migration 1940-1970	.292[a]	.243[a]
	(.029)	(.031)
Observations	114	114
Cragg-Donald Wald F statistic	77.782	36.733
Control		X
Model 2 [Dependent: Ln Total Income 1974]		
	Second Stage	
Ln BUA 1970 (α_Y^M)	1.124[a]	.971[a]
	(.204)	(.237)

Ln Changes in Share of Blacks 1940-1970 (α_Y^T)	.305	.456
	(.450)	(.561)
	First Stage	
Ln Predicted Total Black Migration 1940-1970	.321[a]	.349[a]
	(.058)	(.052)
Observations	114	114
Cragg-Donald Wald F statistic	37.652	50.711
Control		X

Note: [a], [b], and [c] denote significance at the 1 percent, 5 percent, and 10 percent levels, respectively.

Table C.2: Changes in Share of Blacks (T) → Urban Shape (M) → Short-term Income Outcomes (Y)

	(1)	(2)
Model 1 [Dependent: Ln Cohesion Index 1970]		
	Second Stage	
Ln Changes in Share of Blacks 1940-1970 (β_M^T)	.440[a]	.435[b]
	(.134)	(.214)
	First Stage	
Ln Predicted Total Black Migration 1940-1970	.292[a]	.243[a]
	(.029)	(.031)
Observations	114	114
Cragg-Donald Wald F statistic	77.782	36.733
Control		X
Model 2 [Dependent: Ln Total Income 1974]		
	Second Stage	
Ln Cohesion Index 1970 (β_Y^M)	2.381[a]	2.281[a]
	(.652)	(.818)
Ln Changes in Share of Blacks 1940-1970 (β_Y^T)	.743[c]	1.045[b]
	(.418)	(.456)
	First Stage	
Ln Predicted Total Black Migration 1940-1970	.152[a]	.149[a]
	(.048)	(.052)
Observations	114	114
Cragg-Donald Wald F statistic	11.579	10.554
Control		X

Note: [a], [b], and [c] denote significance at the 1 percent, 5 percent, and 10 percent levels, respectively.

Table C.3: Changes in Share of Blacks (T) → Urban Size (M) → Long-term Income Outcomes (Y)

	(1)	(2)
Model 1 [Dependent: Ln BUA 1970]		
	Second Stage	
Ln Changes in Share of Blacks 1940-1970 (α_M^T)	1.462[a]	1.671[a]
	(.241)	(.307)
	First Stage	
Ln Predicted Total Black Migration 1940-1970	.257[a]	.237[a]
	(.033)	(.037)
Observations	74	74
Cragg-Donald Wald F statistic	56.164	33.186
Control		X
Model 2 [Dependent: Ln Total Income 2010]		
	Second Stage	
Ln BUA 1970 (α_Y^M)	.759[a]	.819[a]
	(.188)	(.153)
Ln Changes in Share of Blacks 1940-1970 (α_Y^T)	.010	.041
	(.146)	(.154)
	First Stage	
Ln Predicted Total Black Migration 1940-1970	.391[a]	.403[a]
	(.067)	(.067)
Observations	74	74
Cragg-Donald Wald F statistic	41.194	46.599
Control		X

Note: [a], [b], and [c] denote significance at the 1 percent, 5 percent, and 10 percent levels, respectively.

Table C.4: Changes in Share of Blacks (T) → Urban Shape (M) → Long-term Income Outcomes (Y)

	(1)	(2)
Model 1 [Dependent: Ln Cohesion Index 1970]		
	Second Stage	
Ln Changes in Share of Blacks 1940-1970 (β_M^T)	.477[a]	.486[c]
	(.183)	(.270)
	First Stage	
Ln Predicted Total Black Migration 1940-1970	.257[a]	.237[a]
	(.033)	(.037)
Observations	74	74
Cragg-Donald Wald F statistic	56.164	33.186
Control		X
Model 2 [Dependent: Ln Total Income 2010]		
	Second Stage	
Ln Cohesion Index 1970 (β_Y^M)	2.023[a]	2.277[b]
	(.673)	(1.022)
Ln Changes in Share of Blacks 1940-1970 (β_Y^T)	.155	.303[c]
	(.157)	(.174)
	First Stage	
Ln Predicted Total Black Migration 1940-1970	.147[b]	.145[c]
	(.063)	(.076)
Observations	74	74
Cragg-Donald Wald F statistic	8.173	6.886
Control		X

Note: [a], [b], and [c] denote significance at the 1 percent, 5 percent, and 10 percent levels, respectively.

Table C.5: Changes in Share of Blacks (T) → Urban Size (M) → Short-term AQI Outcomes (Y)

	(1)	(2)	(3)	(4)	(5)	(6)
Model 1 [Dependent: Ln BUA 1970]						
			Second Stage			
Ln Changes in Share of Blacks 1940-1970 (α_M^T)	1.217[a]	1.385[a]	1.217[a]	1.385[a]	1.217[a]	1.385[a]
	(.174)	(.200)	(.174)	(.200)	(.174)	(.200)
			First Stage			
Ln Predicted Total Black Migration 1940-1970	.286[a]	.272[a]	.286[a]	.272[a]	.286[a]	.272[a]
	(.033)	(.036)	(.033)	(.036)	(.033)	(.036)
Observations	93	93	93	93	93	93
Cragg-Donald Wald F statistic	72.865	42.365	72.865	42.365	72.865	42.365
Control		X		X		X
Model 2 [Dependent: Ln AQI Outcomes 1980]						
			Second Stage			
Ln BUA 1970 (α_Y^M)	-.001	.094	.083	.099[c]	-.001	.029[c]
	(.084)	(.061)	(.052)	(.051)	(.026)	(.017)
Ln Changes in Share of Blacks 1940-1970 (α_Y^T)	.075	.030	.040	.039	.033	.020
	(.074)	(.060)	(.051)	(.059)	(.022)	(.016)
			First Stage			
Ln Predicted Total Black Migration 1940-1970	.316[a]	.330[a]	.316[a]	.330[a]	.316[a]	.330[a]
	(.054)	(.050)	(.054)	(.050)	(.054)	(.050)
Observations	93	93	93	93	93	93
Cragg-Donald Wald F statistic	33.981	42.279	33.981	42.279	33.981	42.279
Control		X		X		X

Note: [a], [b], and [c] denote significance at the 1 percent, 5 percent, and 10 percent levels, respectively. Columns (1) and (2) display the results when the outcome is median AQI, Columns (3) and (4) are for max AQI, and Columns (5) and (6) are for the ratio of the unhealthy weather days.

Table C.6: Changes in Share of Blacks (T) → Urban Shape (M) → Short-term AQI Outcomes (Y)

	(1)	(2)	(3)	(4)	(5)	(6)
Model 1 [Dependent: Ln Cohesion Index 1970]						
			Second Stage			
Ln Changes in Share of Blacks 1940-1970 (β_M^T)	.476[a]	.383[c]	.476[a]	.383[c]	.476[a]	.383[c]
	(.168)	(.210)	(.168)	(.210)	(.168)	(.210)
			First Stage			
Ln Predicted Total Black Migration 1940-1970	.286[a]	.272[a]	.286[a]	.272[a]	.286[a]	.272[a]
	(.033)	(.036)	(.033)	(.036)	(.033)	(.036)
Observations	93	93	93	93	93	93
Cragg-Donald Wald F statistic	72.865	42.365	72.865	42.365	72.865	42.365
Control		X		X		X
Model 2 [Dependent: Ln AQI Outcomes 1980]						
			Second Stage			
Ln Cohesion Index 1970 (β_Y^M)	-.003	.189	.152	.200	-.001	.059
	(.155)	(.166)	(.107)	(.130)	(.048)	(.046)
Ln Changes in Share of Blacks 1940-1970 (β_Y^T)	.074	.088[c]	.068[c]	.100[b]	.033[b]	.038[a]
	(.052)	(.046)	(.041)	(.044)	(.015)	(.012)
			First Stage			
Ln Predicted Total Black Migration 1940-1970	.172[a]	.164[b]	.172[a]	.164[b]	.172[a]	.164[b]
	(.062)	(.067)	(.062)	(.067)	(.062)	(.067)
Observations	93	93	93	93	93	93
Cragg-Donald Wald F statistic	9.332	7.621	9.332	7.621	9.332	7.621
Control		X		X		X

Note: [a], [b], and [c] denote significance at the 1 percent, 5 percent, and 10 percent levels, respectively. Columns (1) and (2) display the results when the outcome is median AQI, Columns (3) and (4) are for max AQI, and Columns (5) and (6) are for the ratio of the unhealthy weather days.

Table C.7: Changes in Share of Blacks (T) → Urban Size (M) → Long-term AQI Outcomes (Y)

	(1)	(2)	(3)	(4)	(5)	(6)
Model 1 [Dependent: Ln BUA 1970]						
			Second Stage			
Ln Changes in Share of Blacks 1940-1970 (α_M^T)	1.331[a]	1.642[a]	1.331[a]	1.642[a]	1.331[a]	1.642[a]
	(.183)	(.243)	(.183)	(.243)	(.183)	(.243)
			First Stage			
Ln Predicted Total Black Migration 1940-1970	.288[a]	.241[a]	.288[a]	.241[a]	.288[a]	.241[a]
	(.032)	(.034)	(.032)	(.034)	(.032)	(.034)
Observations	103	103	103	103	103	103
Cragg-Donald Wald F statistic	68.073	31.537	68.073	31.537	68.073	31.537
Control		X		X		X
Model 2 [Dependent: Ln AQI Outcomes 2010]						
			Second Stage			
Ln BUA 1970 (α_Y^M)	.019	.047	.019	.061[b]	-.015	.001
	(.040)	(.031)	(.037)	(.027)	(.016)	(.006)
Ln Changes in Share of Blacks 1940-1970 (α_Y^T)	.093[b]	.087[b]	.051	.030	.027[c]	.020[b]
	(.042)	(.038)	(.048)	(.052)	(.015)	(.001)
			First Stage			
Ln Predicted Total Black Migration 1940-1970	.305[a]	.343[a]	.305[a]	.343[a]	.305[a]	.343[a]
	(.061)	(.056)	(.061)	(.056)	(.061)	(.056)
Observations	103	103	103	103	103	103
Cragg-Donald Wald F statistic	30.301	44.038	30.301	44.038	30.301	44.038
Control		X		X		X

Note: [a], [b], and [c] denote significance at the 1 percent, 5 percent, and 10 percent levels, respectively. Columns (1) and (2) display the results when the outcome is median AQI, Columns (3) and (4) are for max AQI, and Columns (5) and (6) are for the ratio of the unhealthy weather days.

Table C.8: Changes in Share of Blacks (T) → Urban Shape (M) → Long-term AQI Outcomes (Y)

	(1)	(2)	(3)	(4)	(5)	(6)
Model 1 [Dependent: Ln Cohesion Index 1970]						
			Second Stage			
Ln Changes in Share of Blacks 1940-1970 (β_M^T)	.461[a]	.460[c]	.461[a]	.460[c]	.461[a]	.460[c]
	(.146)	(.237)	(.146)	(.237)	(.146)	(.237)
			First Stage			
Ln Predicted Total Black Migration 1940-1970	.288[a]	.241[a]	.288[a]	.241[a]	.288[a]	.241[a]
	(.032)	(.034)	(.032)	(.034)	(.032)	(.034)
Observations	103	103	103	103	103	103
Cragg-Donald Wald F statistic	68.073	31.537	68.073	31.537	68.073	31.537
Control		X		X		X
Model 2 [Dependent: Ln AQI Outcomes 2010]						
			Second Stage			
Ln Cohesion Index 1970 (β_Y^M)	.038	.110	.038	.142[c]	-.031	.003
	(.081)	(.078)	(.078)	(.083)	(.035)	(.015)
Ln Changes in Share of Blacks 1940-1970 (β_Y^T)	.100[a]	.114[a]	.059	.065	.021[b]	.021[a]
	(.030)	(.029)	(.037)	(.047)	(.010)	(.007)
			First Stage			
Ln Predicted Total Black Migration 1940-1970	.150[a]	.147[b]	.150[a]	.147[b]	.150[a]	.147[b]
	(.051)	(.057)	(.051)	(.057)	(.051)	(.057)
Observations	103	103	103	103	103	103
Cragg-Donald Wald F statistic	10.187	9.039	10.187	9.039	10.187	9.039
Control		X		X		X

Note: [a], [b], and [c] denote significance at the 1 percent, 5 percent, and 10 percent levels, respectively. Columns (1) and (2) display the results when the outcome is median AQI, Columns (3) and (4) are for max AQI, and Columns (5) and (6) are for the ratio of the unhealthy weather days.