CITY SHAPE, COMMUTING TIME AND WORKPLACE CHOICE: ANALYZING URBAN SPATIAL FORM IN CHINA

SHEN FU

SINGAPORE MANAGEMENT UNIVERSITY
2024

City Shape, Commuting Time and Workplace Choice: Analyzing Urban Spatial Form in China

by

Shen Fu

Submitted to School of Economics in partial fulfillment of the requirements for the Degree of Master of Philosophy in Economics

<u>Master's Thesis Committee:</u>

Lin Ma (Supervisor / Chair)
Assistant Professor of Economics
Singapore Management University

Jing Li
Associate Professor of Economics
Singapore Management University

Yuan Mei
Assistant Professor of Economics
Singapore Management University

Singapore Management University
2024

I hereby declare that this Master's thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in this thesis.

This Master's thesis has also not been submitted for any degree in any university previously.

Shen Fu

Shen Fu

18 July 2024

City Shape, Commuting Time and Workplace Choice: Analyzing Urban Spatial Form in China

Shen Fu

Abstract

This paper investigates the relationship between city shape, commuting time, and workplace choice within the context of Chinese cities. Using an urban form measure, *nCohesion*, and nighttime light data, I constructed the China City Shape Database for 263 Chinese cities. The empirical analysis reveals a counterintuitive finding: cities with less compact shapes, despite having longer within-city distances, are associated with shorter commuting times. This is because residents in poorly shaped cities tend to select workplaces closer to their homes to mitigate commuting difficulties. The findings suggest that improving city shape alone may not reduce commuting times; instead, enhancing transportation infrastructure and promoting local employment can mitigate the adverse effects of poor city shapes.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I express my deepest gratitude to my main advisor, Prof. Lin Ma, whose patient and thoughtful guidance and support were indispensable throughout my master's thesis project. His expertise and encouragement were pivotal in navigating this research and composing this thesis. Through weekly meetings, Prof. Ma guided me in conducting new research, brainstorming research ideas, applying quantitative skills, managing time effectively, and learning some coding and Geographical Information System (GIS) software from scratch. Prof. Ma's knowledge and sense of responsibility have set a lifelong example for me, inspiring my future endeavors.

My sincere thanks also go to Prof. Jing Li and Prof. Mei Yuan for their roles on my thesis committee. Their invaluable feedback and suggestions significantly enriched this work.

I am immensely grateful to my friends and family for their unwavering love and support throughout this process. Their presence made this journey not just possible, but meaningful.

Chapter 1

Introduction

The past several decades have witnessed rapid urbanization across the globe, particularly in fast-growing countries such as China and India. For instance, China's urban population surged from 26% in 1990 to 61% in 2020 (World Bank Open Data, 2023a). However, this rapid urbanization trend has raised concerns among urban planners and economists regarding urban sprawl. Urban sprawl has led to numerous market failures, including reduced access to open spaces, traffic congestion, and excessive infrastructure costs (Brueckner, 2000). A key feature of urban sprawl is the emergence of more distorted urban forms, resulting in deteriorating city shapes (Wang et al., 2023). In practice, not all cities can expand with ideal shapes. Some are hindered by geographical constraints, such as steep terrains or water bodies, while others are shaped by poor urban planning.

City shape refers to the geometric configuration and spatial distribution of urban areas, encompassing the layout, density, and physical form of a city. A compact city shape typically features a higher density of buildings and infrastructure within a smaller geographic area, fostering proximity between residential, commercial, and public spaces (Liu and Tian, 2022a). This compactness can result in shorter distances between different parts of the city, improved public transport accessibility (Burton, 2000), increased firm productivity and household consumption (Li, 2022), better access to public services and amenities (Carruthers and Ulfarsson, 2003), enhanced neighborhood satisfaction (Mouratidis, 2018), and greater social equity (Burton, 2003). However, it can also lead to overcrowding (Neuman, 2005), unaffordable housing prices (Burton, 2000), and reduced urban green spaces (Haaland and van den Bosch, 2015).

The literature on urban sprawl and urban form encompasses various strands, including monocentricity, building density, and the geometry of a city's footprint. This paper focuses on the geometric aspect, an underexplored topic for urban economists prior to Harari's work on Indian cities (Harari, 2020). Previously, much of urban economics literature treated cities as points, relying on total population data or city size information while neglecting their geometric shapes (Li, 2022).

Incorporating city shape into our analysis allows us to address questions about how a city's internal structure affects urban dwellers' intracity commuting choices. This thesis explores the relationship between city shape, commuting time, and workplace choice, providing new insights into how urban morphology influences residents' daily lives.

By constructing a measure of city shape (hereafter nCohesion), we observe that, for a given area, better-shaped cities have shorter within-city distances. Conventional wisdom might believe that shorter within-city distances would naturally lead to shorter commuting times. However, this study reveals a counterintuitive finding: less compact cities are associated with shorter commuting times because residents in poorly shaped cities tend to choose workplaces closer to their homes to mitigate commuting difficulties.

This thesis contributes to the literature by addressing several key gaps. Firstly, it constructs the China City Shape Database 2015, documenting shape metrics for 263 Chinese cities. This study is one of the few empirical investigations of city shape conducted in China, utilizing comprehensive data from multiple sources, including nightlight data. The use of nightlight data to measure urban form offers a compelling approach to accurately capture city shape for this and future research.

Moreover, this research extends the discussion by empirically analyzing city shape in the context of commuting and workplace choice, a topic previously unexplored. Our findings highlight the adaptive behaviors of residents in response to suboptimal urban forms, presenting a counterintuitive result.

The structure of this thesis is as follows: Chapter 2 reviews the relevant literature, highlighting existing research on city shape. Chapter 3 presents the measurement of city shape. Chapter 4 discusses the hypotheses and mechanisms underpinning this study. Chapter 5 details the data sources and empirical specifications used in the analysis. Chapters 6 and 7 present the empirical results, focusing on the impact of city shape on commuting time and workplace choice, respectively. Chapter 8 provides heterogeneity analysis. Chapter 9 offers robustness checks. Chapter 10 concludes the thesis with a summary of findings, policy implications, and suggestions for future research.

Through this research, we aim to contribute to a deeper understanding of how urban form affects commuting behaviors and workplace choices, offering valuable insights for urban planners and policymakers seeking to create more efficient and livable cities.

Chapter 2

Literature Review

The global urban population rate has grown from 43% in 1990 to 56% in 2020 (World Bank Open Data, 2023b). This rapid urbanization has generated a substantial body of literature on urban sprawl (Glaeser and Kahn, 2004).

Research on urban sprawl encompasses various strands, including floor area ratio (Joshi and Kono, 2009), monocentric and polycentric city design (Li et al., 2018), decentralization and mixed land use (Cervero, 1996), and the geometric layout of cities. Much of the previous economics literature has focused on city size, population, and population density, often treating cities as points, perfect circles, or entirely symmetric forms. This approach neglects the fact that real-world city shapes are rarely ideal. The work of Harari (2020) on Indian cities has brought economists' attention to the geometric shape of cities.

Subsequently, numerous studies have examined city shapes in China. Research has explored the impact of city shape on innovation (Liu and Wang, 2023), happiness (Wang et al., 2023), urban employment growth (Xiao et al., 2022), housing affordability (Sun et al., 2024), the human development index (Liu and Tian, 2022a), urban poverty (Liu and Tian, 2022b), economic development (Zou and Yang, 2024; Li, 2022), firm productivity, and household consumption (Li, 2022). Among these works, the Mandarin-written Ph.D. dissertation by Li (2022) is the most comprehensive, detailing the measurement of China's city shape, urban economic growth, firm productivity, and household consumption.

My paper measures city shape by using urban planning scholars's indicators (Angel et al., 2010). I was also inspired by Harari's idea of measuring Indian city shape (Harari, 2020), and followed Li's practice (Li, 2022). This includes adopting the same luminosity threshold of 35 and utilizing the same operational definition of a city. Details of the construction can be found in Chapter 3 and Appendix B. Additionally, the relationship between city shape and commuting has not been directly studied before.

Chapter 3

City Shape and Measurement

Before proceeding with the formal analysis, we will first introduce what is city shape and how it is computed.

3.1 City Shape

City shape refers to the geometric configuration and spatial distribution of urban areas, including the layout of streets, the density of buildings, and the arrangement of land uses. A well-shaped city typically features a compact, contiguous, and well-connected urban form, which facilitates efficient transportation and accessibility. In contrast, poorly shaped cities may exhibit fragmented, sprawling, or irregular patterns that hinder mobility and increase travel distances (Bertaud, 2004).

Angel et al. (2010) have proposed several ways of measuring city shape. One popular measure is *nCohesion*. As a normalized indicator of *Cohesion* index, it quantifies the degree of spatial cohesion and connectivity within a city. Higher *nCohesion* values indicate better city shape, characterized by more compact and connected urban areas.

The Cohesion index measures the average distance between all pairs of interior points within a shape. It is calculated using the formula:

$$\text{Cohesion} = \frac{\sum_{i,j} d_{ij}}{N}$$

where d_{ij} is the distance between points i and j, and N is the number of point pairs. The cohesion index provides an estimate of the average travel distance within the polygon, capturing the internal connectivity of the city.

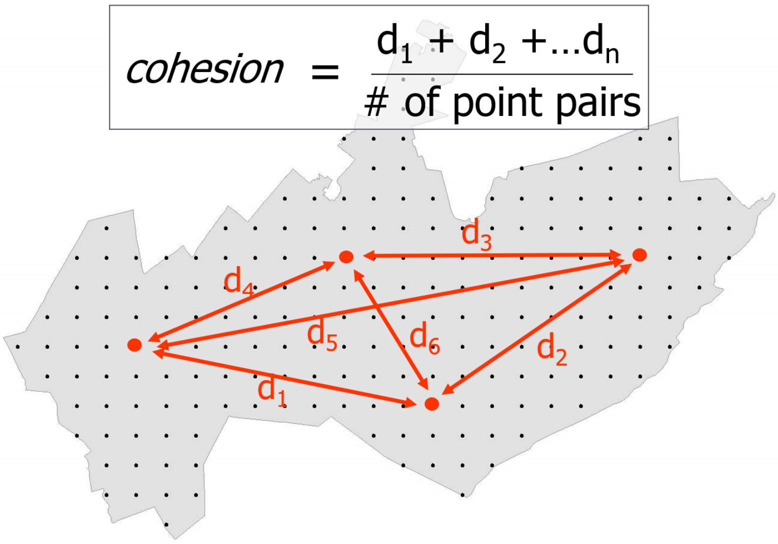


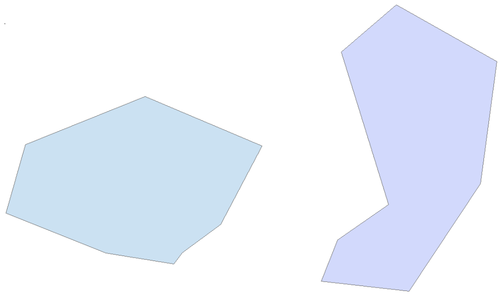
Figure 3.1: Example diagram for calculating Cohesion. This figure is from Angel, Civco, and Parent (2009)

Figure 3.1, provided by Angel, Civco, and Parent (2009), illustrates the methodology used to calculate city shape metrics. This paper employs a user-written ArcGIS plugin developed by these authors. The plugin assigns 20,000 uniformly distributed interior points within a polygon in a grid pattern. It then conducts 30 samples, with each sample computing the average distance of 1,000 randomly selected points.

The *nCohesion* index is the normalized version of the cohesion index. It is calculated by normalizing the cohesion index with respect to an Equal Area Circle (EAC), which is a circle with the same area as the polygon. The normalization formula is:

$$\text{nCohesion} = \frac{\text{Cohesion}_{\text{EAC}}}{\text{Cohesion}_{\text{Shape}}}$$

Here $\text{Cohesion}_{\text{EAC}}$ is the cohesion index of the EAC. Normalizing using the EAC removes the effect of area difference, allowing for a pure measurement of shape compactness. A higher *nCohesion* value indicates a more compact and connected



City Name	Area(km2)	Cohesion(km)	nCohesion
Neijiang	52.26	3.75	0.985
Qingyang	52.36	4.89	0.755

Figure 3.2: Example of city shape: Neijiang (Left) and Qingyang (Right)

urban form, while a lower value suggests a more sprawling and disconnected shape. *nCohesion* is between 0 to 1, where 1 means the shape is a perfect circle.

For example, Figure 3.2 are city shapes of Neijiang (left) and Qingyang (right). Although both cities are similar in size, the *Cohesion* and *nCohesion* are vastly different. Neijiang's city shape is closer to a circle, and therefore has higher *nCohesion*.

To calculate *nCohesion* index for Chinese cities, I used the global nightlight data. The detailed process are documented in Appendix B.

While *nCohesion* is a widely used measure of city shape, other indices and metrics, such as *proximity*, *spin* and *range*, also exist. However, in previous papers that have employed multiple indicators (Harari, 2020; Li, 2022), these indicators show a high relevance and similar results. Therefore this paper only adopts the most popular one, *Cohesion* and *nCohesion*.

Figure 3.3 is the density distribution of cities. Most cities' *nCohesion* are between 0.9 to 1.

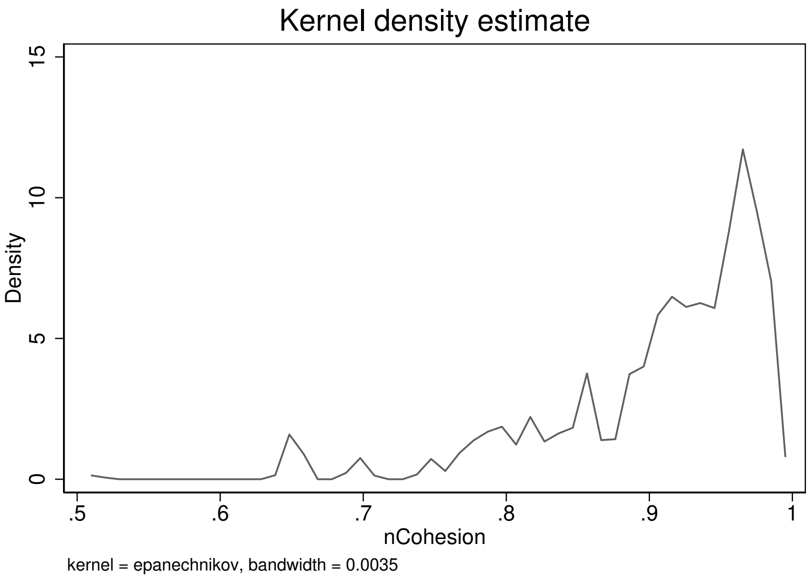


Figure 3.3: Density of ncohesion across cities

Chapter 4

How Does City Shape Affect Commuting Time? Two Hypotheses

The relationship between city shape and commuting time is an unexplored topic in urban economics. The previous literature, the work of Harari (2020), done for Indian cities, didn't investigate this problem directly due to the lack of commuting surveys. Our measurement of city shape implies that better shaped cities have shorter within city distance, which enhances transit accessibility. As a result, some may think it lead to reduced commuting times. However, our study provides an alternative perspective, hypothesizing that the relationship between city shape and commuting time is more nuanced. We propose two hypotheses to guide our empirical analysis.

4.1 Hypothesis 1: Better City Shape Improves Accessibility and Reduces Commuting Time

Urban economists have argued that a more compact and well-shaped city is associated with better transit accessibility (Harari, 2020; Wang et al., 2023). The reason is that in a well-shaped city, the distance between residential areas and workplaces is shorter, thereby reducing overall commuting time. This hypothesis aligns with the traditional compact city paradigm, which emphasizes dense, well-organized urban forms as a solution to excessive commuting times. Formally, we state:

> **Hypothesis 1:** Cities with better shapes, characterized by higher nCohesion, are associated with shorter commuting times due to improved transit accessibility (characterized by shorter within city distance).

We test this hypothesis by examining the relationship between the nCohesion metric, which quantifies the compactness of a city's shape, and the commuting time of its residents.

4.2 Hypothesis 2: Bad City Shape Influences Workplace Choice and Reduces Commuting Time

Contrary to the traditional view, our second hypothesis assumes that bad city shape may still induce a higher cost to travel, but residents will choose to adapt their behavior. In cities with poorer shapes (lower nCohesion), residents tend to work closer to home to reduce commuting costs. In contrast, better-shaped cities might encourage residents to travel farther for work due to lower marginal commuting costs.

This hypothesis suggests that the relationship between nCohesion and commuting time might be positive, reflecting the behavioral adjustments of residents to the urban form. Note that this hypothesis does not inherently contradict the concept of 'transit accessibility' discussed in the first hypothesis. Shorter within city distance can be a feature of better transit accessibility, together with other features such as longer built road mileage and better public transportation service. As Harari (2020) noted, a poorly shaped city can be perceived as offering a less favorable menu of choices. Residents in such cities may opt to work closer to home, forgo certain trips entirely, or resort to less desirable substitutes, such as shopping at more expensive local convenience stores instead of traveling to distant supermarkets. While this adaptive behavior can partially mitigate the cost of commuting, it does not imply that households can fully optimize all the negative consequences of bad city shape.

Formally, we state:

> **Hypothesis 2:** In cities with poorer shapes (lower nCohesion), residents tend to live closer to their workplaces, resulting in shorter commuting times in poorly shaped cities.

To test this hypothesis, we analyze the commuting patterns and workplace choices of residents in relation to the nCohesion metric. Specifically, we examine whether residents in poorly shaped cities (with lower nCohesion) exhibit a higher tendency to work closer to home compared to those in better-shaped cities.

By testing these two hypotheses, we aim to provide a comprehensive understanding of how city shape impacts commuting time and to challenge the conventional wisdom that better city shapes always lead to improved commuting outcomes. Our empirical analysis will employ a range of econometric models to rigorously test these hypotheses using data on city shapes and commuting patterns.

Chapter 5

Model Specification and Data

This section will first present the empirical specifications used in this study, followed by an overview of the data and summary statistics.

5.1 Empirical Specification

This section outlines the empirical specifications used in this paper. The first one examines the relationship between city shape and commuting time. The second one examines the relationship between city shape and workplace choice. We focus on a cross-sectional analysis for the year 2015, considering only employed individuals within the working age range from 15 to 65 years old.

5.1.1 Econometric Specification

To test our hypotheses regarding the impact of city shape on commuting time, we employ the following econometric model:

$$\text{travel\_time}_{ijc} = \beta_0 + \beta_1 \text{nCohesion}_c + \boldsymbol{\gamma}_1 \textbf{Individual}_{ijc} + \boldsymbol{\gamma}_2 \textbf{Household}_{ijc} + \boldsymbol{\gamma}_3 \textbf{City}_c + \varepsilon_{ijc} \quad (5.1)$$

where:

- travel_time_{ijc} is the commuting time for individual i in household j in city c.

- β_0 is the intercept term.

- nCohesion_c is the key explanatory variable measuring city shape's effect for city c.

- $\textbf{Individual}_{ijc}$ is a vector of individual-level control variables, including gender, age, hukou status, education level, ethnic group, travel mode, and marital status.

- $\textbf{Household}_{ijc}$ is a vector of household-level control variables, including number of family members, floor area, number of rooms, kitchen type, restroom type, house source type, value of cars, and number of cars. While we do not have direct statistics about individual or family income, many of these variables can also serve as a proxy of income.

- $\textbf{City}_c$ is a vector of city-level control variables, including population, per capita GDP, fiscal expenditure to GDP ratio, fiscal revenue to GDP ratio, total road area, number of buses, number of college students, number of hospital beds, number of books in libraries, second industry ratio, and third industry ratio.

- ε_{ijc} is the error term clustered at the city level c.

The second model examines the relationship between workplace choice and city shape. The model is identical with the first model except that the dependent variable is *work_home*, which is a binary variable that 1 refers to the workplace and residence are in the same subdistrict within the city, and 0 refers to else:

$$\text{work\_home}_{ijc} = \beta_0 + \beta_1 \text{nCohesion}_c + \gamma_1 \textbf{Individual}_{ijc} + \gamma_2 \textbf{Household}_{ijc} + \gamma_3 \textbf{City}_c + \varepsilon_{ijc} \quad (5.2)$$

5.2 Data Sources, Variables, and Summary Statistics

Three datasets were employed to study how city shape affect commuting time and residents' workplace choice. The first one is the global nightlight database, where we extract urban polygons to calculate city shape metrics. The second one is China Intercensal Population Sample Survey of One-Percent 2015. The last one is China City Statistical Yearbook.

5.2.1 Nighttime Lights data

Two popular nighttime lights database used in China's city shape studies are DMSP/OLS database and NPP-VIIRS database (Li, 2022). The first database records global nightlight data from 1992 to 2013, with luminosity level from 0 to 63, where 35 was a commonly used nightlight threshold employed in previous studies (Li, 2022; Harari, 2020).

NPP-VIIRS night light time records nightlight data from 2012 with a new generation of satellites, offering a richer luminosity level that is over 200 . This paper employs DVNL database, which is derived from VIIRS database. It converts VIIRS readings to DMSP-like digital numbers, providing around 63 luminosity levels. This conversion addresses the incompatibility in spatial resolution and dynamic range between the two sensors, providing a continuous and comparable dataset for long-term analysis (Nechaev et al., 2021; Ghosh et al., 2021). This paper still chooses luminosity level 35 to be consistent with previous literature.

Therefore, the operational definition of a 'city' used in this paper is: the largest continuous nightlight area, above the luminosity threshold 35, within the administration boundary. City shape (measured by *nCohesion*), and urban area data are computed from this source and detailed process is documented in Appendix B.

5.2.2 Other Data Source, and Summary Statistics

All variables and the descriptive statistics are summarized in Table 5.1.

Here individual- and family-level data is from the 2015 China Intercensal Population Sample Survey, whose default time is November 1, 2015 (Xinhua News Agency, 2016). Other city-level data is from the China City Statistical Yearbook 2016, which

Table 5.1: Descriptive Statistics

	count	mean	sd	min	max
travel_time	858,841	19.02	27.77	0	300
travel_mode	858,841	2.94	2.028	1	8
ncohesion	858,841	.9066	.07781	.5132	.9916
area_km2	858,841	413.4	718.6	1.689	3,800
gender	858,841	.5783	.4938	0	1
age	858,841	40.8	12.63	15	98
hukou_loc	858,841	.7801	.4142	0	1
edu	858,841	3.492	1.454	1	8
edu_re	858,841	1.864	.4967	1	3
ethnic	858,841	.9463	.2253	0	1
employment	858,841	1	0	1	1
work_place	858,841	1.413	.686	1	3
marriage	858,841	.8143	.3889	0	1
nfam_member	858,841	4.157	2.435	1	45
house_type	837,374	1.02	.2019	1	4
floor_area	827,493	124.4	76.21	1	999
nroom	827,493	4	2.415	1	90
kitchen	827,493	1.184	.5645	1	3
restroom	827,493	2.404	1.587	1	5
house_source	827,493	4.554	1.628	1	8
value_cars	827,493	6.612	.8142	1	7
cars	827,493	.2341	.4235	0	1
popu2015_district	858,841	251.3	353.9	15.36	2,129
pcGRP_district	858,841	8.433	6.514	1.306	49.31
fiscal_exp_to_GDP	858,841	1,816	946.1	391.9	14,278
fiscal_rev_to_GDP	858,841	1,044	419.4	235.2	3,433
road	858,841	3,075	3,561	140	16,128
bus	852,067	40,983	74,216	81	406,003
students	855,151	147,700	222,753	1,030	1,043,221
beds_hospital	858,841	19,019	25,809	995	120,095
library	858,841	4,638	12,782	0	75,055
second_industry_ratio	858,841	46.2	10.03	17.07	74.45
third_industry_ratio	858,841	48.13	10.85	22.36	79.65

records the cities' statistics for 2015 (CHINAYEARBOOKS, 2016).

Key variables of interest include measures of commuting time (*travel_time*), travel mode, city shape indicator (*nCohesion*), demographic characteristics, family level conditions, and urban infrastructure. The details of variable names and labels are provided in Appendix A.

Chapter 6

City Shape and Commuting Time

This section presents the empirical results of our analysis on the relationship between city shape, measured by *nCohesion*, and commuting time. The results are displayed in two tables: Table 6.1 for the overall analysis and Table 6.2 for the analysis by travel mode.

6.0.1 Impact of City Shape on Commuting Time

To examine the relationship between city shape and commuting time, we regressed commuting time on *nCohesion*, controlling for other variables. The results are presented in Table 6.1. Column (2) shows the main result, where we regressed travel time on *nCohesion*, controlling for individual characteristics, family level controls, and city level controls. For comparison, column (1) omits all individual level, family level, and city level controls, containing only the relationship between *nCohesion* and travel time. Column (3) adds individual characteristics while still omitting family level and city level controls. Column (4) further adds city level controls, and column (5) adds family level controls.

The first hypothesis posits that better city shape (higher *nCohesion* index) should improve transitivity, resulting in shorter commuting time. Therefore, the coefficient for *nCohesion* should be negative. However, none of these columns reveal such a conlcusion, suggesting that the first hypothesis does not hold.

Table 6.1: nCohesion on Commuting Time

	(1) travel_time	(2) travel_time	(3) travel_time	(4) travel_time	(5) travel_time
nCohesion	1.890	4.520**	7.257***	4.690**	5.375***
	(2.754)	(2.116)	(1.966)	(2.103)	(1.880)
gender		1.650***	1.599***	1.568***	1.677***
		(0.109)	(0.0982)	(0.0993)	(0.106)
age		-0.0435***	-0.0470***	-0.0437***	-0.0456***
		(0.00464)	(0.00586)	(0.00562)	(0.00474)
hukou_loc		0.0360	2.178***	1.853***	0.0715
		(0.164)	(0.230)	(0.244)	(0.176)
ethnic		-1.552***	-1.932***	-1.566***	-1.914***
		(0.444)	(0.471)	(0.437)	(0.469)
marriage		-0.233	0.00340	0.0140	-0.266*
		(0.142)	(0.124)	(0.118)	(0.147)
Education:					
Secondary and Vocational		-1.018***	-1.177***	-1.060***	-1.085***
		(0.151)	(0.180)	(0.167)	(0.160)
University and Above		-0.703*	-0.879	-0.781	-0.632
		(0.402)	(0.588)	(0.510)	(0.488)
Travel mode:					
Bicycle		2.798***	2.819***	3.057***	2.640***
		(0.232)	(0.267)	(0.253)	(0.248)
Electric Bicycle		4.354***	4.275***	4.599***	4.106***
		(0.248)	(0.252)	(0.258)	(0.242)
Motorcycle		4.960***	5.173***	5.363***	4.760***
		(0.242)	(0.231)	(0.244)	(0.231)
Car		13.59***	12.91***	13.10***	13.59***
		(0.576)	(0.645)	(0.623)	(0.589)
Bus		28.91***	29.35***	29.10***	29.20***
		(1.049)	(1.002)	(1.030)	(1.038)
Rail Transit		89.55***	87.81***	88.09***	89.45***
		(11.76)	(10.96)	(11.55)	(11.09)
Other		22.26***	21.57***	21.20***	22.68***
		(1.245)	(1.169)	(1.174)	(1.249)
Family level controls	No	Yes	No	No	Yes
City level controls	No	Yes	No	Yes	No
Observations	833,881	768,231	833,881	798,007	802,741
Mean of Dep. Variable	19.20	19.43	19.20	19.09	19.54
R-squared	0.0000273	0.215	0.211	0.213	0.213

Notes: This table reports city shape's impact on commuting time. Column (2) controls for individual, family and city level controls. Column (1) omits all controls. Column (3) only has individual level controls. Column (4) omits family level controls and column (5) omits city level controls. Robust standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

In terms of individual characteristics, the coefficient on *gender* is positive and statistically significant across all models with controls, suggesting that males have longer commuting times than females. Negative *age* indicates that older individuals tend to have shorter commuting times. The coefficient on *hukou_loc* is not always significant. In terms of education level, comparing with primary school education and below, the coefficients for *Secondary and Vocational* and *University and Above* are generally negative, indicating that higher education levels are associated with shorter commuting times. *marriage*'s effect looks more trivial as well.

The coefficients for different travel modes show substantial variation in commuting times. For instance, using rail transit (*Rail Transit*) or a bus (*Bus*) significantly increases commuting time compared to other modes.

6.0.2 Relationship by Travel Mode

To further examine the first hypothesis, we also regressed this relationship on different travel modes. The results are listed in Table 6.2. For every odd column, we regressed commuting time on city shape while controlling for individual, family, and city-level characteristics for each specific travel mode. The goal is to detect if any travel mode is more sensitive to this relationship. Moreover, for each even column, we examined this relationship only for residents whose workplaces are in other subdistricts within the city, a sub-sample that tends to have longer commuting times than the average.

None of these results show a negative and significant relationship between *nCohesion* and commuting time, which refutes the first hypothesis.

Our findings suggest that the impact of city shape on commuting time varies by travel mode and workplace location. The positive coefficients on *nCohesion* for most travel modes indicate that better city shape is associated with longer commuting times, particularly for modes such as rail transit and bus, which typically have longer travel duration.

For individuals working in other subdistricts within the city, the results show a similar pattern, suggesting that the spatial distribution of workplaces within the city also plays a critical role in determining commuting times.

In addition, We also conduct several robustness checks to ensure the validity of our findings, which can be found in robustness checks section. We showed that the majority of cities' *nCohesion* is between 0.9 to 1. Therefore we also restrict our attention to cities whose *nCohesion* is below 0.9 and below 0.8. In addition, we also tested the non-linear relationship by adding $nCohesion^2$ term. We still have not found evidence that good city shape is associated with shorter commuting time.

Table 6.2: nCohesion on Commuting Time, by Different Travel Mode

	Walking travel_time	Walking travel_time	Bicycle travel_time	Bicycle travel_time	Electric Bicycle travel_time	Electric Bicycle travel_time	Motorcycle travel_time	Motorcycle travel_time
nCohesion	0.578	1.663	-0.774	-4.352	1.607	6.242***	3.993**	7.032***
	(1.680)	(1.922)	(1.889)	(8.369)	(1.364)	(1.689)	(1.985)	(2.336)
Individual level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Family level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Work in other subdistricts within the city	No	Yes	No	Yes	No	Yes	No	Yes
Observations	286,713	14,102	43,521	5,748	183,321	34,372	72,321	12,928
Mean of Dep. Variable	11.99	13.50	14.27	20.32	15.86	22.01	17.40	23.95
R-squared	0.0507	0.105	0.0410	0.0649	0.0260	0.0356	0.00820	0.0156

	Car travel_time	Car travel_time	Bus travel_time	Bus travel_time	Rail Transit travel_time	Rail Transit travel_time	Other travel_time	Other travel_time
nCohesion	5.981*	2.549	14.04	11.79**	126.0***	14.89	30.34***	0.317
	(3.300)	(2.666)	(9.496)	(4.865)	(47.63)	(27.77)	(11.37)	(7.399)
Individual level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Family level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Work in other subdistricts within the city	No	Yes	No	Yes	No	Yes	No	Yes
Observations	68,980	33,293	72,684	35,749	10,083	4,981	30,608	6,188
Mean of Dep. Variable	24.19	28.30	40.84	39.16	101.4	53.36	33.96	35.85
R-squared	0.0295	0.0721	0.0209	0.0601	0.268	0.121	0.0501	0.0913

Notes: This table reports city shape's impact on commuting time for different travel modes. Individual, household and city level characteristics are controlled. The odd columns does not restrict workplaces. The even columns further restrict workplaces that are in other subdistricts within the city. Robust standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Chapter 7

City Shape and Workplace Choice

The second hypothesis believes that people in worse shaped cities tend to choose workplaces that are closer to their home, resulting in a shorter commuting time. Table 6.1 indicates that, in general, worse city shape is associated with shorter travel time, which supports this hypothesis. To further validate this idea, we can test the relationship between city shape and work place choice. Table 7.1 and Table 7.2 present the results of logistic regressions (logit) between city shape (*nCohesion*) and workplace choice.

Here I categorized workplaces into two types. The variable *work_home* equals 1 if the individual's workplace is within the same subdistrict they reside in, and 0 if it is outside their subdistrict or city. This binary variable indicates proximity: 1 means the workplace is close to home, and 0 means it is far away.

7.0.1 Impact of City Shape on Workplace Choice

In Table 7.1, we showed the relationship between city shape and work place choice. Column (2) shows the main result, where we regressed *work_home* on *nCohesion*, controlling for individual characteristics, family level controls, and city level controls. For comparison, column (1) omits all individual level, family level, and city level controls, containing only the relationship between *nCohesion* and *work_home*.

In column (2), the coefficient on *nCohesion* is negative and on five percent level

Table 7.1: nCohesion on Work Place Choice

	(1) work_home	(2) work_home
nCohesion	0.457***	-0.0939**
	(0.0303)	(0.0376)
gender		-0.285***
		(0.00543)
age		0.0371***
		(0.000290)
hukou_loc		0.342***
		(0.00807)
ethnic		-0.124***
		(0.0124)
marriage		0.175***
		(0.00754)
Education:		
Secondary and Vocational		-0.538***
		(0.00909)
University and Above		-1.301***
		(0.0139)
Family level controls	No	Yes
City level controls	No	Yes
Observations	833,881	768,231
Mean of Dep. Variable	0.693	0.688
Pseudo R-squared	0.0002	0.1032

Notes: This table reports city shape's impact on workplace choice. Column (2) controls for individual, family and city level controls. Column (1) omits all controls. Robust standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

significant, indicating that worse city shape (lower *nCohesion*) is associated with a higher probability of working close to home. That means individuals in worse-shaped cities are more likely to choose workplaces within their subdistrict. This finding supports the idea that workplace choice is indeed endogenous to city shape.

The coefficient on *gender* suggests that males are less likely to work close to home compared to females. The coefficient is -0.285, indicating a substantial decrease in the likelihood of working close to home for males. *age* is positive and statistically significant, indicating that older individuals are more likely to work close to home. The coefficient is 0.0371, implying an increase in the probability of working close to home with each additional year of age. *hukou_loc* is positive as well, suggesting that those with local hukou are more likely to work close to home. The coefficients on education levels (*Secondary and Vocational* and *University and Above*) are negative and significant, indicating that higher education levels are associated with a lower probability of working close to home. *marriage* is also positive, indicating that married individuals are more likely to work close to home.

Many of these results look consistent with the story from Table 6.1. For example, males tend to work farther away and have longer commuting times, whereas older individuals live closer to their workplaces and have shorter commuting times.

These results look intuitive and can be explained as follows: under current socio-economic conditions, females, older individuals, and married people often have more household responsibilities, particularly child care, prompting them to choose jobs closer to home. In contrast, non-local hukou holders, who are typically migrant workers, may prioritize employment opportunities over proximity to home. Higher-educated individuals often possess specialized skills and qualifications that necessitate working in specific industries or firms located farther from residential areas. These individuals may also be more willing to commute longer distances for better career opportunities and higher-paying jobs.

7.0.2 Relationship by Travel Mode

Table 7.2: nCohesion on Work Place Choice, by Different Travel Mode

	Walking work_home	Bicycle work_home	Electric Bicycle work_home	Motorcycle work_home
nCohesion	0.0226	-0.288	0.150**	-0.844***
	(0.0824)	(0.198)	(0.0757)	(0.137)
Individual level controls	Yes	Yes	Yes	Yes
Family level controls	Yes	Yes	Yes	Yes
City level controls	Yes	Yes	Yes	Yes
Observations	286,713	43,521	183,321	72,321
Mean of Dep. Variable	0.844	0.798	0.736	0.764
Pseudo R-squared	0.1066	0.1089	0.0634	0.0377

	Car work_home	Bus work_home	Rail Transit work_home	Other work_home
nCohesion	-0.00731	-0.576***	-1.606**	-0.210
	(0.106)	(0.135)	(0.762)	(0.173)
Individual level controls	Yes	Yes	Yes	Yes
Family level controls	Yes	Yes	Yes	Yes
City level controls	Yes	Yes	Yes	Yes
Observations	68,980	72,684	10,081	30,608
Mean of Dep. Variable	0.427	0.196	0.0455	0.571
Pseudo R-squared	0.0556	0.0190	0.0536	0.1424

Notes: This table reports city shape's impact on workplace choice for each specific travel mode. Individual, family and city level controls are included. Robust standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

To further validate the hypothesis, we examined the relationship between city shape and workplace choice by different travel modes. The results are presented in Table 7.2. The operations are the same with column (2) from Table 7.1, but specifying for each travel mode.

The coefficients on *nCohesion* for walking, bicycle and car are not statistically significant. While the coefficients for bus and rail transit are more significant. Motorcycle's coefficient is significant too. Electric bicycle's p-value is 0.048, at the margin of 5 percent significance.

The result is pretty interesting, because we can categorize travel modes into two types in terms of routes and schedules: walking, bicycle, electric bicycle, and car as 'flexible' one, and bus and rail transit as 'inflexible' one. Individuals using these flexible travel modes are less sensitive to city shape when choosing their workplaces, they can adapt to varying city shapes without significantly altering their workplace choices.

In contrast, the coefficients on *nCohesion* for inflexible travel modes such as bus and rail transport are negative and statistically significant. For bus users, the coefficient is -0.576, and for rail transit users, the coefficient is -1.606. This indicates that worse city shape is associated with a higher probability of working close to home for individuals using these inflexible travel modes. The inflexibility of these modes, due to fixed routes and schedules, makes workplace choice more sensitive to city shape.

In addition, the coefficient on *nCohesion* for motorcycle users is also negative and statistically significant. This suggests that motorcycle users are also sensitive to workplace choice in worse-shaped cities. This might be due to motorcycle restrictions in many urban areas, making them less flexible in practice. According to Zuo (2009, as cited in Guo et al. (2020)), by 2009, more than 30% of the 363 prefecture-level cities had enacted at least one form of motorcycle restrictions, including prohibiting motorcycles on main streets, banning them from the central business district (CBD), restricting non-local licensed motorcycles, and halting the issuance of new motorcycle licenses.

Overall, the results from Table 7.1 and Table 7.2 support the second hypothesis that people in worse-shaped cities tend to choose workplaces closer to their homes, resulting in shorter commuting times. This effect is more pronounced for individuals using inflexible travel modes such as bus and rail transport, as well as motorcycles, due to their limited flexibility.

The findings highlight the importance of city shape in influencing workplace choice and commuting behavior, with significant implications for urban planning and transportation policy. The distinction between flexible and inflexible travel modes further underscores the need for tailored approaches in addressing commuting challenges in cities with varying shapes.

Chapter 8

Heterogeneity Analysis

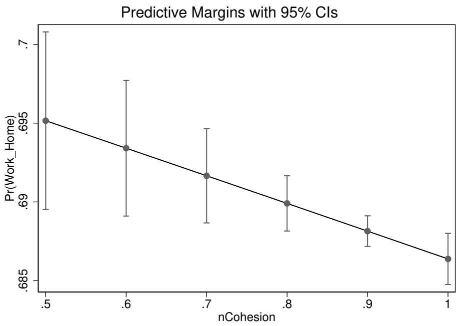
8.1 Heterogeneity Analysis 1

This section tests the heterogeneity in the relationship between city shape (*nCohesion*) and the likelihood of working close to home (*work_home*), to check if the relationship is linear. The analysis begins with the main regression and examines the marginal effects for various subgroups, including *nCohesion*, age, hukou status, education, ethnicity, and marital status. It presents the analysis following Table 7.1 column (2)'s work.

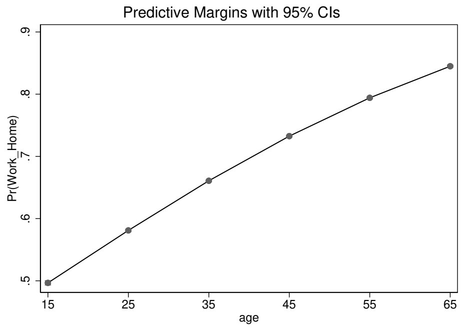
The marginal effects of *nCohesion* on the probability of working close to home were examined at values ranging from 0.5 to 1, in increments of 0.1. The results indicate that as *nCohesion* increases, the likelihood of working close to home decreases, as shown in Figure 8.1a. This suggests that higher *nCohesion* decreases the likelihood of working close to home.

The analysis of age shows that the probability of working close to home increases with age, particularly between the ages of 15 and 35, before stabilizing at older ages. This implies that younger individuals are less likely to work close to home compared to older individuals.

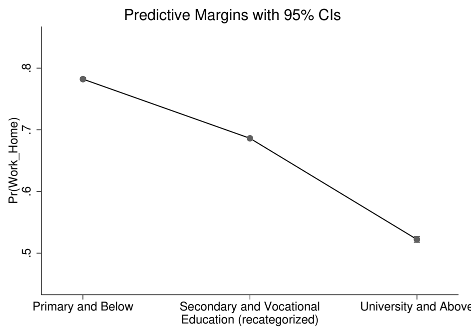
Education was categorized into three levels: primary or below, secondary and vocational, and university and above. The marginal effects analysis shows that higher



(a) Work place choice on different city shape level



(b) Work place choice on different age level



(c) Work place choice on different education level

Figure 8.1: Comparison of work place choice across different variables

education levels are associated with a lower probability of working close to home.

For hukou status, individuals with a local hukou are 70.19% likely to work close to home, whereas those with a non-local hukou are 63.53% likely to work close to home. This significant difference indicates that local hukou status positively influences the likelihood of working close to home.

Marital status also shows a notable difference, with unmarried individuals being 66.17% likely to work close to home, while married individuals are 69.47% likely to do so. This indicates that married people are more likely to work close to home, maybe because of familial reasons like the need to take care of children.

These findings suggest that while city shape (*nCohesion*) plays a role in workplace location choices, individual and household characteristics, as well as local residency status, are significant determinants of the likelihood of working close to home.

8.2 Heterogeneity Analysis 2

This section examines the heterogeneity in the relationship between city shape (*nCohesion*) and the likelihood of working close to home (*work_home*) by including interaction terms of *nCohesion* with various individual characteristics. The logit regression includes interactions with gender, age, hukou status, education, ethnicity, and marital status, along with the main effects and controls for individual, household, and city-level characteristics. I also draw the margins plot for each interaction term. The result is listed in Table 8.1

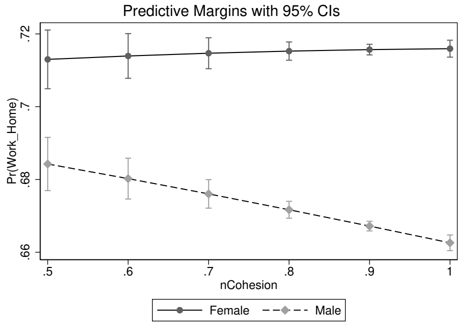
The interaction between *nCohesion* and gender is negative and significant, suggesting that comparing with females, higher *nCohesion* decreases male's choice of work close to home more. From Figure 8.2 (a), female's work place choice does not vary much with city shape. However, male's probability of work close to home significantly decreases with the increase of *nCohesion*. That means comparing with female, which is the baseline, males are much more sensitive to city shape variation.

Table 8.1: nCohesion on work place choice, with all interaction terms

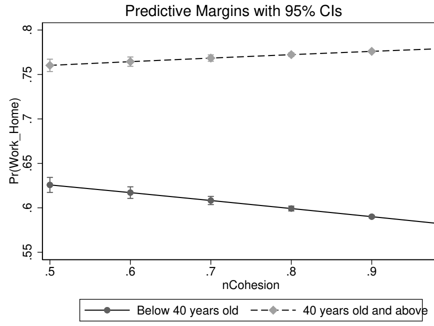
	work_home
nCohesion	-1.752***
	(0.217)
Male × nCohesion	-0.287***
	(0.0685)
Age 40 and above × nCohesion	0.593***
	(0.0734)
Local hukou × nCohesion	1.123***
	(0.0822)
Han × nCohesion	0.0429
	(0.165)
Married × nCohesion	0.372***
	(0.0886)
Education×interaction terms:	
Secondary and Vocational × nCohesion	0.271**
	(0.114)
University and Above × nCohesion	1.317***
	(0.165)
Male	0.0108
	(0.0623)
Age 40 and above	0.0680
	(0.0668)
Local hukou	-0.639***
	(0.0743)
Han	-0.156
	(0.152)
Married	0.0458
	(0.0807)
Education:	
Secondary and Vocational	-0.944***
	(0.104)
University and Above	-2.645***
	(0.149)
Family level controls	Yes
City level controls	Yes
Observations	768231
Mean of Dep. Variable	0.688

Notes: This table reports city shape's impact on workplace choice, with all interaction terms between nCohesion and individual level characteristics. Robust standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

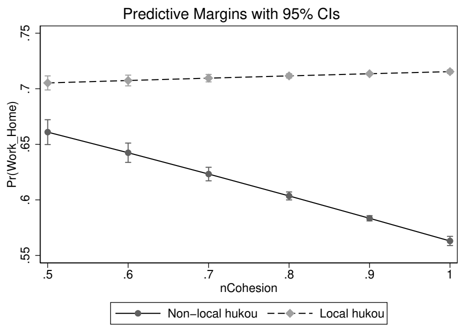
In other words, males drives the main negative overall trend of city shape and work close to home.



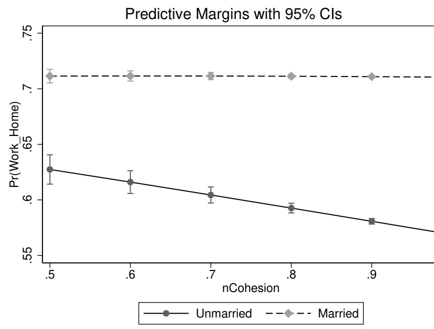
(a) Work place choice on different city shape level, by gender



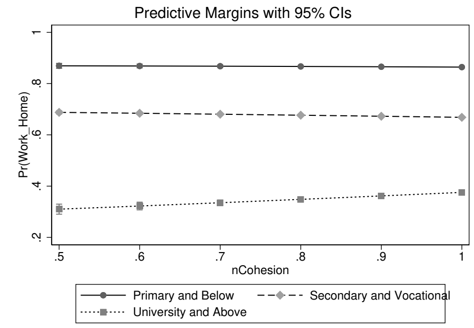
(b) Work place choice on different city shape level, by age group



(c) Work place choice on different city shape level, by hukou



(d) Work place choice on different city shape level, by marital status



(e) Work place choice on different city shape level, by education

Figure 8.2: Comparison of work place choice on different city shape levels, segmented by various demographics

For age, I re-categorized age into two groups: 'below 40 years old' and '40 years old and above'. The interaction between *nCohesion* and age is positive and significant, indicating that compared with people below 40 year old, higher *nCohesion* increases older people's choice of work close to home more. This is shown in Figure 8.2 (b).

Hukou status is shown on Figure 8.2 (c). People with local hukou's work place choice is basically invariant with regard to city shape differences. However, for non-local hukou people, their work place choice is heavily affected by city shape. This underscores the importance of local residency status in job location decisions. The interaction term is positive and highly significant, because non-local hukou are selected as the base.

The interaction term between *nCohesion* and marital status indicates that the relationship between city shape and workplace choice is different for married and unmarried individuals. The first observation is that married individuals are more likely to work close to home compared to unmarried individuals. The second observation is that married individuals' workplace choice are less sensitive to city shape variation. While the unmarried people's choice drive the overall negative trend.

In terms of education, there are two clear patterns. Firstly, highly educated people are less likely to work close to home. In addition, Secondary, vocational school and below educated people's work place choice are not very sensitive to city shape change. People with university and above level education are more sensitive. For better shaped cities, more of them are likely to work close to home.

To summarize, male, young, non-local hukou and unmarried subgroups are more sensitive to city shape variation. Good city shape often encourages them to work farther from home. At the same time, highly educated people are encouraged to work closer to home. This validates the result in Table 6.1, where highly educated people's commuting time is lower than others.

These interaction effects highlight the complex nature of the relationship between city shape and work-home proximity, demonstrating that individual characteristics such as gender, age, hukou status, education, and marital status play crucial roles in moderating this relationship. The significant interaction terms suggest that policies aimed at improving city shape should consider these demographic factors to effectively enhance work-home proximity.

Chapter 9

Robustness Checks

9.1 City Shape and Commuting Time Part

As shown in Figure 3.3, most cities nCohesion metrics are between 0.9 to 1. Therefore we restrict nCohesion range for less and greater than 0.9 and 0.8 as robustness checks. The result is shown in Table 9.1.

We examined the relationship between nCohesion and commuting time, and found that none of the results are negative and significant, which is consistent with our main finding that refutes the idea that better city shape results in less commuting time.

Table 9.1: nCohesion on Commuting Time, Restricting nCohesion Range

	nCohesion <0.9	nCohesion <0.8	nCohesion >0.9	nCohesion >0.8
	travel_time	travel_time	travel_time	travel_time
nCohesion	6.408**	-4.491	13.13*	8.869**
	(2.653)	(4.054)	(7.408)	(3.781)
Individual level controls	Yes	Yes	Yes	Yes
Family level controls	Yes	Yes	Yes	Yes
City level controls	Yes	Yes	Yes	Yes
Observations	241,512	85,305	526,719	682,926
Mean of Dep. Variable	19.34	18.84	19.47	19.50
R-squared	0.190	0.198	0.229	0.220

Notes: This table reports city shape's impact on commuting time for cities' nCohesion smaller or larger than 0.9 and 0.8. Individual, household and city level controls are used. Robust standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

In addition, we also introduced a squared term for *nCohesion* in our regression model. This approach allows us to capture potential non-linear relationships between city shape and commuting time that a linear model might miss. The result is shown below, suggesting a U-shaped relationship between city shape and commuting time. The turning point is 0.747. To further investigate this, I also restricted the *nCohesion* range to values smaller or greater than 0.747, and still could not find a negative and significant *nCohesion* coefficient. Overall, this result does not support the first hypothesis.

Table 9.2: nCohesion on Commuting Time, adding $nCohesion^2$ Term

	add $nCohesion^2$
	travel_time
nCohesion	-37.29
	(24.06)
nCohesion × nCohesion	24.96*
	(14.58)
Individual level controls	Yes
Family level controls	Yes
City level controls	Yes
Observations	768,231
Mean of Dep. Variable	19.43
R-squared	0.215

Notes: This table reports city shape's impact on commuting time while adding squared nCohesion term. Individual, household and city level controls are used. Robust standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

9.2 City Shape and Workplace Choice Part

For nCohesion and workplace choice, we also tested it with nCohesion range restriction. The result, Table 9.3, is consistent with our main results, which state that in worse shaped cities, people are more inclined to work close to home.

Since we used logit model in our main results part, here we employ the probit model and the linear probability model (LPM) to validate the consistency of our results. The consistent results, shown in Table 9.4, show that our findings are robust to different model specifications.

Table 9.3: nCohesion on Workplace Choice, Restricting nCohesion Range

	nCohesion <0.9 work_home	nCohesion <0.8 work_home	nCohesion >0.9 work_home	nCohesion >0.8 work_home
nCohesion	-0.492***	-1.333***	-0.700***	-0.983***
	(0.0774)	(0.260)	(0.138)	(0.0698)
Individual level controls	Yes	Yes	Yes	Yes
Family level controls	Yes	Yes	Yes	Yes
City level controls	Yes	Yes	Yes	Yes
Observations	241,512	85,305	526,719	682,926
Mean of Dep. Variable	0.674	0.670	0.694	0.690
Pseudo R-squared	0.1028	0.1104	0.1074	0.1037

Notes: This table reports city shape's impact on workplace choice for cities' nCohesion less or greater than 0.9 and 0.8. Individual, household and city level controls are used. Robust standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 9.4: nCohesion on Workplace Choice, using alternative models

	Probit work_home	Linear Probability Model work_home
nCohesion	-0.0514**	-0.0219***
	(0.0223)	(0.00704)
Individual level controls	Yes	Yes
Family level controls	Yes	Yes
City level controls	Yes	Yes
Observations	768,231	768,231
Mean of Dep. Variable	0.688	0.688
Pseudo R-squared/R-squared	0.1041	0.1227

Notes: This table reports city shape's impact on workplace choice by using probit and linear probability model. Individual, household and city level controls are used. Robust standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Chapter 10

Conclusion

This thesis provides a comprehensive analysis of the relationship between city shape, commuting time, and workplace choice, revealing counterintuitive insights that challenge the conventional notion that better city shape is associated with shorter commuting times. The findings, based on an empirical analysis of a rich dataset encompassing city shape metrics, individual characteristics, and commuting patterns, are summarized as follows:

The analysis reveals that less compact cities are associated with shorter commuting times, contrary to the intuitive belief that compact urban forms enhance accessibility and reduce travel times. This is because residents in poorly shaped cities tend to choose workplaces closer to their homes to mitigate the difficulties associated with commuting in such environments. This finding highlights the adaptive behavior of urban dwellers in response to suboptimal city shapes, prioritizing convenience in their workplace choices.

The heterogeneity analysis further enriches these findings by uncovering variations in the impact of city shape across different demographic groups and urban settings. Notably, the study finds that the workplace choices of males, younger individuals, unmarried people, and non-local hukou holders are more sensitive to city shape changes. They are more likely to be encouraged to work farther from home due to good city shape. Meanwhile, highly educated individuals exhibit contrary commuting patterns in response to urban form. These results emphasize the need for demographic considerations in urban planning and policy formulation.

The implications of this study are significant for urban planners and policymakers. The first takeaway is that efforts to improve city shape and compactness should be carefully evaluated because it may not reduce residents' commuting times due to their adaptive behaviors.

The second takeaway is that city shape is often determined by historical trends and geographical barriers, making it difficult to change. Despite of this, urban planners can still improve transportation infrastructure to enhance transit accessibility, such as building roads or tunnels when parts of a city are segregated. Although this may not reduce commuting times to work, it could provide residents with a better menu of commuting choices.

The third takeaway is that when city shape is hard to change, offering decentralized services and promoting local employment opportunities can mitigate the negative effects of poorly shaped cities, leading to more efficient and livable urban environments.

There are several directions for future research. Most current city shape studies treat cities as segregated entities on the map. However, some cities grow and integrate, forming urban agglomerations such as the Yangtze River Delta Urban Agglomeration, the Pearl River Delta Urban Agglomeration, and the Tokyo Metropolitan Area. Transportation in these urban agglomerations warrants further exploration.

Additionally, further work could employ instrumental variable approaches to strengthen causal identification.

In conclusion, this thesis challenges conventional urban planning wisdom by presenting evidence that less compact cities are associated with shorter commuting times and a higher likelihood of local employment. These findings underscore the adaptive strategies employed by residents in response to suboptimal urban forms and provide insights for future research and policy interventions aimed at creating more sustainable and livable cities.

Bibliography

S. Angel, D. L. Civco, and J. Parent. Shape metrics, 2009. URL https://drive.google.com/drive/folders/1U2lmfuFE7_0ucfim8ziyrpjAmoPcujN7. Accessed: 2024-06-20.

S. Angel, J. Parent, and D. L. Civco. Ten compactness properties of circles: measuring shape in geography. *Canadian Geographies / Géographies canadiennes*, 54(4):441–461, 2010. ISSN 1541-0064. doi: 10.1111/j.1541-0064.2009.00304.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1541-0064.2009.00304.x.

A. Bertaud. The Spatial Organization of Cities: Deliberate Outcome or Unforeseen Consequence? July 2004. URL https://escholarship.org/uc/item/5vb4w9wb.

J. K. Brueckner. Urban Sprawl: Diagnosis and Remedies. *International Regional Science Review*, 23(2):160–171, Apr. 2000. ISSN 0160-0176. doi: 10.1177/016001700761012710. URL https://doi.org/10.1177/016001700761012710. Publisher: SAGE Publications Inc.

E. Burton. The Compact City: Just or Just Compact? A Preliminary Analysis. *Urban Studies*, 37(11):1969–2006, Oct. 2000. ISSN 0042-0980. doi: 10.1080/00420980050162184. URL https://doi.org/10.1080/00420980050162184. Publisher: SAGE Publications Ltd.

E. Burton. Housing for an Urban Renaissance: Implications for Social Equity. *Housing Studies*, 18(4):537–562, July 2003. ISSN 0267-3037. doi: 10.1080/02673030304249. URL https://doi.org/10.1080/02673030304249. Publisher: Routledge _eprint: https://doi.org/10.1080/02673030304249.

J. I. Carruthers and G. F. Ulfarsson. Urban Sprawl and the Cost of Public Services. *Environment and Planning B: Planning and Design*, 30(4):503–522, Aug. 2003. ISSN 0265-8135. doi: 10.1068/b12847. URL https://doi.org/10.1068/b12847. Publisher: SAGE Publications Ltd STM.

R. Cervero. Mixed land-uses and commuting: Evidence from the American Housing Survey. *Transportation Research Part A: Policy and Practice*, 30(5):361–377, Sept. 1996. ISSN 0965-8564. doi: 10.1016/0965-8564(95)00033-X. URL https://www.sciencedirect.com/science/article/pii/096585649500033X.

CHINAYEARBOOKS. China city statistical yearbook 2016, 2016. URL https://www.chinayearbooks.com/china-city-statistical-yearbook-2016.html. Accessed: 2024-07-11.

E. L. Glaeser and M. E. Kahn. Chapter 56 - Sprawl and Urban Growth. In J. V. Henderson and J.-F. Thisse, editors, *Handbook of Regional and Urban Economics*, volume 4 of *Cities and*

Geography, pages 2481–2527. Elsevier, Jan. 2004. doi: 10.1016/S1574-0080(04)80013-0. URL https://www.sciencedirect.com/science/article/pii/S1574008004800130.

Y. Guo, J. Wang, S. Peeta, and P. Ch. Anastasopoulos. Personal and societal impacts of motorcycle ban policy on motorcyclists' home-to-work morning commute in China. *Travel Behaviour and Society*, 19:137–150, Apr. 2020. ISSN 2214-367X. doi: 10.1016/j.tbs.2020.01.002. URL https://www.sciencedirect.com/science/article/pii/S2214367X19303151.

C. Haaland and C. K. van den Bosch. Challenges and strategies for urban green-space planning in cities undergoing densification: A review. *Urban Forestry & Urban Greening*, 14(4):760–771, Jan. 2015. ISSN 1618-8667. doi: 10.1016/j.ufug.2015.07.009. URL https://www.sciencedirect.com/science/article/pii/S161886671500103X.

M. Harari. Cities in Bad Shape: Urban Geometry in India. *American Economic Review*, 110(8):2377–2421, Aug. 2020. ISSN 0002-8282. doi: 10.1257/aer.20171673. URL https://www.aeaweb.org/articles?id=10.1257/aer.20171673.

K. K. Joshi and T. Kono. Optimization of floor area ratio regulation in a growing city. *Regional Science and Urban Economics*, 39(4):502–511, July 2009. ISSN 0166-0462. doi: 10.1016/j.regsciurbeco.2009.02.001. URL https://www.sciencedirect.com/science/article/pii/S0166046209000155.

C. Li. Chengshi kongjian xingtai de jingji jixiao pinggu ji yingxiang jizhi yanjiu (economic performance evaluation and impact mechanism of urban spatial form), 2022. URL https://wf.pub/thesis/article:D02671501.

X. Li, Y. Mou, H. Wang, C. Yin, and Q. He. How Does Polycentric Urban Form Affect Urban Commuting? Quantitative Measurement Using Geographical Big Data of 100 Cities in China. *Sustainability*, 10(12):4566, Dec. 2018. ISSN 2071-1050. doi: 10.3390/su10124566. URL https://www.mdpi.com/2071-1050/10/12/4566. Number: 12 Publisher: Multidisciplinary Digital Publishing Institute.

L. Liu and Y. Tian. Compact Urban Form and Human Development: Retest Based on Heterogeneous Effects. *International Journal of Environmental Research and Public Health*, 19(4):2198, Jan. 2022a. ISSN 1660-4601. doi: 10.3390/ijerph19042198. URL https://www.mdpi.com/1660-4601/19/4/2198. Number: 4 Publisher: Multidisciplinary Digital Publishing Institute.

L. Liu and Y. Tian. Does the Compact City Paradigm Help Reduce Poverty? Evidence from China. *International Journal of Environmental Research and Public Health*, 19(10):6184, Jan. 2022b. ISSN 1660-4601. doi: 10.3390/ijerph19106184. URL https://www.mdpi.com/1660-4601/19/10/6184. Number: 10 Publisher: Multidisciplinary Digital Publishing Institute.

X. Liu and Q. Wang. Kongjian fazhan moshi, zhishi yishuo yu chengshi chuangxin jixiao (spatial development patterns, knowledge spillovers, and urban innovation performance), 2023. URL https://cmjj.ajcass.com/Magazine/Show?ID=89558.

K. Mouratidis. Is compact city livable? The impact of compact versus sprawled neighbourhoods on neighbourhood satisfaction. *Urban Studies*, 55(11):2408–2430, Aug.

2018. ISSN 0042-0980. doi: 10.1177/0042098017729109. URL https://doi.org/10.1177/0042098017729109. Publisher: SAGE Publications Ltd.

M. Neuman. The Compact City Fallacy. *Journal of Planning Education and Research*, 25(1): 11–26, Sept. 2005. ISSN 0739-456X. doi: 10.1177/0739456X04270466. URL https://doi.org/10.1177/0739456X04270466. Publisher: SAGE Publications Inc.

J. Sun, X. Xing, Q. Xi, and W. Shi. Impact of urban form on housing affordability stress in Chinese cities: Does public service efficiency matter? *Cities*, 145:104682, Feb. 2024. ISSN 0264-2751. doi: 10.1016/j.cities.2023.104682. URL https://www.sciencedirect.com/science/article/pii/S0264275123004948.

Q. Wang, X. Liu, F. Zhang, Y. Gu, and X. Zhou. Does compact city shape matter to residents' happiness? *Cities*, 141:104524, Oct. 2023. ISSN 0264-2751. doi: 10.1016/j.cities.2023.104524. URL https://www.sciencedirect.com/science/article/pii/S0264275123003360.

World Bank Open Data. Urban population (% of total population) - china, 2023a. URL https://data.worldbank.org/indicator/SP.URB.TOTL.IN.ZS?end=2023&locations=CN&start=1960&view=chart. Accessed: 2024-07-11.

World Bank Open Data. Urban population (% of total population), 2023b. URL https://data.worldbank.org/indicator/SP.URB.TOTL.IN.ZS?skipRedirection=true. Accessed: 2024-07-11.

W. Xiao, W. Liu, and C. Li. Can the urban spatial structure accelerate urban employment growth? Evidence from China. *Growth and Change*, 53(4):1668–1693, 2022. ISSN 1468-2257. doi: 10.1111/grow.12594. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/grow.12594.

Xinhua News Agency. Tongji ju 1% renkou chouyang diaocha: quanguo dalu renkou wei 13.7349 yi ren (national bureau of statistics 1% population sampling survey: The population of mainland china is 1.37349 billion), Apr. 2016. URL https://www.gov.cn/xinwen/2016-04/20/content_5066345.htm.

W. Zou and F. Yang. Does City Shape Affect China's Economic Development? *China & World Economy*, 32(1):21–56, 2024. ISSN 1749-124X. doi: 10.1111/cwe.12515. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/cwe.12515.

Appendix A

Appendix A: Variable Definitions and Labels

Variables are labeled this way:

A.0.1 Individual-Level Data from the China Intercensal Population Sample Survey

- *travel_mode*: 1: 'Walking', 2: 'Bicycle', 3: 'Electric bike', 4: 'Motorcycle', 5: 'Private car', 6: 'Public bus', 7: 'Rail transit', 8: 'Other'.
- *gender*: 1: 'Male', 0: 'Female'.
- *hukou_loc*: 1: 'Local', 0: 'Non-local'.
- *edu*: 1: 'None', 2: 'Primary', 3: 'Middle School', 4: 'High School', 5: 'Vocational School', 6: 'Associate Degree', 7: 'University', 8: 'Graduate and Above'.
- *ethnic*: 1: 'Han', 0: 'Non-Han'.
- *employment*: 1: 'Employed', 0: 'Unemployed'.
- *work_place*: 1: The workplace is with the same subdistrict with the resident's home location, 2: The workplace is in other subdistrict within the city, 3: The workplace is outside the city.
- *marriage*: 1: 'Married', 0: 'Unmarried'.

A.0.2 Family-Level Data from the China Intercensal Population Sample Survey

- *house_type*: 1: 'Ordinary residence', 2: 'Collective dormitory', 3: 'Accommodation at work', 4: 'No housing'.
- *kitchen*: 1: 'Exclusive use', 2: 'Shared with other households', 3: 'None'.
- *restrooms*: 1: 'Exclusive use of pump/flush toilet', 2: 'Shared use of pump/flush toilet', 3: 'Exclusive use of other types', 4: 'Shared use of other types', 5: 'None'.

- *house_source*: 1: 'Purchase of newly built commercial housing', 2: 'Purchase of second-hand housing', 3: 'Purchase of former public housing', 4: 'Purchase of affordable housing', 5: 'Self-built housing', 6: 'Rental of low-rent housing', 7: 'Rental of other housing', 8: 'Other'.

- *value_of_cars*: 1: 'Owns a car worth more than 1 million yuan', 2: 'Owns a car worth 500,000 to 1 million yuan', 3: 'Owns a car worth 300,000 to 500,000 yuan', 4: 'Owns a car worth 200,000 to 300,000 yuan', 5: 'Owns a car worth 100,000 to 200,000 yuan', 6: 'Owns a car worth less than 100,000 yuan', 7: 'Does not own a car'.

A.0.3 City-Level Data from the China City Statistical Yearbook

- *popu2015_district*: Population of the district in 2015.

- *pcGRP_district*: Per capita local GDP.

- *fiscal_exp_to_GDP*: Fiscal expenditure relative to local GDP. Not normalized to the same unit here.

- *fiscal_rev_to_GDP*: Fiscal revenue relative to GDP. Not normalized to the same unit here.

- *road*: Year-end actual urban road area (in ten thousand square meters).

- *bus*: Annual total passenger traffic of buses and trolleybuses (in ten thousand person-trips).

- *students*: Number of students enrolled in regular institutions of higher education.

- *beds_hospital*: Number of hospital beds.

- *library*: Number of books in public libraries (in thousands).

- *second_industry_ratio*: Ratio of the second industry.

- *third_industry_ratio*: Ratio of the third industry.

Table A.1: Variable Definitions

Variable	**Description**
City-Level Data from Nightlight Data	
ncohesion	Normalized measure of city shape
area_km2	Area of the city in square kilometers
Individual-Level Data from the China Intercensal Population Sample Survey	
travel_time	Duration (in minutes) from residence to workplace
travel_mode	Mode of travel to workplace
gender	Gender of the individual
age	Age of the individual
hukou_loc	Hukou status
edu	Educational attainment
ethnic	Ethnicity
employment	Employment status
work_place	Workplace location
marriage	Marital status
Family-Level Data from the China Intercensal Population Sample Survey	
nfam_member	Number of family members
house_type	Type of housing
floor_area	Size of the house in square meters
nroom	Number of rooms
kitchen	Type of kitchen
restrooms	Type of restrooms
house_source	Source of housing acquisition or occupancy
value_of_cars	Price range of personal cars
City-Level Data from the China City Statistical Yearbook 2016	
popu2015_district	Population of the district in 2015
pcGRP_district	Per capita local GDP
fiscal_exp_to_GDP	Fiscal expenditure relative to local GDP
fiscal_rev_to_GDP	Fiscal revenue relative to GDP
road	Year-end actual urban road area (in ten thousand square meters)
bus	Annual total passenger traffic of buses and trolleybuses (in ten thousand person-trips)
students	Number of students enrolled in regular institutions of higher education
beds_hospital	Number of hospital beds
library	Number of books in public libraries (in thousands)
second_industry_ratio	Ratio of the second industry
third_industry_ratio	Ratio of the third industry

Appendix B

Appendix B: City Shape Metrics Computation

This section provides a step-by-step guide to calculate the city shape index using ArcGIS. The process involves downloading relevant data, processing raster images, applying luminosity thresholds, performing spatial joins, and conducting various calculations. The detailed steps are as follows:

1. Firstly, the 2015 data was downloaded from the DMSP-like Nighttime Lights Derived from VNL (DVNL) [1].

2. A new project was created in ArcGIS, and the downloaded raster data was added to a new geodatabase. Pyramids were created using cubic convolution to enhance visualization and processing.

3. A luminosity threshold of 35 was applied using the Raster Calculator in the Geoprocessing tools.

4. For polygonization, the Raster to Polygon conversion tool in the Geoprocessing tools was utilized, with the output being simplified.

5. Next, two additional layers were added: *city points.shp* and an administrative boundary shapefile. The first file provided each city government's latitude and longitude information, while the second file offered the prefecture-level cities' administrative boundaries.

6. The nightlight polygons were then integrated with the administrative boundaries using the Intersect tool under Analysis Tools in ArcToolbox. The polygons were split using the Multipart to Singlepart tool, resulting in a layer with separated polygons.

7. City points were joined to these split polygons using the Spatial Join tool in ArcToolbox. The polygons were set as the target features, and the city points as the join features, using the JOIN_ONE_TO_ONE option.

8. Polygons that were not matched to any city points were dropped. For the layer obtained from the previous step, the attribute table was opened, and polygons with a "Join Count" of 0 were selected using the "Select By Attributes" tool and deleted. This ensured that every administrative boundary only had polygons with corresponding city points. However, some city points were not correctly matched to the city polygon, which required manual adjustments. The city points were manually adjusted, and after saving these edits, the join process was repeated until all city points were matched.

9. To project the data to square kilometers, the Project tool under Data Management Tools in ArcToolbox was used. The "Asia North Albers Equal Area Conic" projection was selected.

[1] Available at: `https://eogdata.mines.edu/products/dmsp/#v4_dmsp_download_intercal`

10. A new field for the area calculation was added. In the attribute table, a field named *Area_km2* with a double data type was added. Using the field calculator with the Python expression *!shape.area@SQUAREKILOMETERS!*, the area in square kilometers was calculated.

11. Next, shape metrics were computed using the Shape Metrics tool in ArcToolbox. It is a user written tool provided by Angel, Civco, and Parent (2009). For the layer obtained from the previous step, metrics such as cohesion were calculated with an edge width of 0, and the output was saved.

12. After completing the calculations in ArcGIS, the resulting data was added and the attribute table was opened. The table was then exported to a CSV file, named *2015_ShapeMetrics*. The results were verified by checking the most and least cohesive cities.

13. Upon reviewing the CSV file, it was noticed that some cities' cohesion and nCohesion values were misaligned due to null columns on the left side. These columns were manually adjusted to ensure the correct alignment of the data.

The final observations were as follows: Cities with small cohesion indices were typically smaller cities, such as Zhangjiajie, while cities with large cohesion indices included major cities like Beijing, Tianjin, Nanjing, and Chengdu. This observation is consistent with the concept of Cohesion, as it measures average mutual distance, naturally growing with urban area. Cities with high nCohesion indices, such as Liaocheng and Xuzhou, typically had a circular shape. Conversely, cities with low nCohesion indices, such as Yantai and Baoji, had more elongated shapes. These observations align with the definitions of *Cohesion* and *nCohesion*.

One concern during the operational phase is the possibility of having multiple nightlight polygons within a single administrative boundary, indicating several cities within that boundary. Our operational definition of a city is the largest continuous nightlight area above the luminosity threshold of 35 within the administrative boundary. This raises the question: is the largest continuous nightlight area necessarily the capital city within the boundary? For example, if cities A, B, and C are all within the same administrative boundary and A is the capital city, but B has a larger nightlight area, could our definition mistakenly identify city B as the capital?

The result is no. This was verified using the city points mapping (*city points.shp*). On the one hand, I use city points mapping to discard all irrelevant urban polygons. On the other hand, I solely kept the largest polygons within each boundary. After comparing these two approaches, I found that for 2015, all largest nightlight urban polygons are exactly the capital cities within their respective administrative boundaries.

Eventually, the China City Shape Database 2015 was successfully computed, containing urban area and city shape information for 263 cities.