



ESSAYS IN SPATIAL ECONOMICS AND REGIONAL DEVELOPMENT

By

Yujing Yang

A thesis submitted to

The University of Birmingham

for the degree of

DOCTOR OF PHILOSOPHY

Department of Economics

Birmingham Business School

College of Social Sciences

March 2025

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

Contents

List of Figures	v
-----------------	---

List of Tables	vii
----------------	-----

1 Which Path to Growth? Heterogeneous Impacts of Transportation Modes on Regional Industry	2
1.1 Introduction	4
1.2 Literature Review	6
1.2.1 Heterogeneous Impact of Transportation on the Regional Economy .	7
1.2.2 Geographical Spillover Effects of Infrastructure	10
1.3 Data Description and Background	12
1.4 Methodology	35
1.4.1 Preparation of the Estimation	36
1.4.2 Spatial Autocorrelation	41
1.4.3 Model Selection	49
1.4.4 Model Building	56
1.5 Results	59

1.5.1	Baseline Results	60
1.5.2	Regional Inequality and Heterogeneity	69
1.6	Conclusion	82
2	Spatial Dynamics of Freight Efficiency: Use of Synthetic Instruments to Overcome Endogeneity	85
2.1	Introduction	87
2.2	Literature Review	90
2.2.1	Freight Productivity in Regional Production	91
2.2.2	Synthetic Instrument Construction	92
2.3	Empirical Data Description	99
2.4	Methodology	102
2.4.1	Weight Matrix	104
2.4.2	Synthetic Instruments Construction	105
2.5	Econometric Analysis	111
2.5.1	Two-step GMM Estimation	112
2.5.2	Spatial Analysis in Synthetic Instruments	120
2.6	Empirical Findings: Spillover Effects of Freight Productivity	134
2.6.1	Data and Background	134
2.6.2	Empirical Result	135
2.7	Residual Analysis	137
2.8	Conclusion	140
3	All Roads Lead to Rome: The Role of Road Networks in the Economic Resilience of Cities	142

3.1	Introduction	144
3.2	Literature Review	146
3.2.1	Market Accessibility with Road Network	146
3.2.2	Urban Economic Resilience	148
3.2.3	Industry Agglomeration and Urban Cluster	150
3.3	Background and Data	153
3.3.1	Data Description	153
3.3.2	Background of Chinese Cities	155
3.4	Methodology	158
3.4.1	Industry Cluster Analysis	159
3.4.2	Regional Economic Resilience	165
3.4.3	Access to Big Markets	169
3.5	Dynamic Spatial Model	179
3.5.1	Weight Matrices and Spatial Autocorrelation	180
3.5.2	Model Setting	186
3.6	Empirical Finding	187
3.6.1	Main Result	188
3.6.2	Spatial Analysis	190
3.7	Conclusion	197
	Conclusion	200
	A Appendix to Chapter 1	214
A.1	Descriptive Statistics	214
A.2	Main Regression	215

A.3	Robust Results	217
B	Appendix to Chapter 2	223
B.1	Data Description	223
B.2	Windmeijer’s Small-Sample Correction	225
B.3	Diagnostic Tests	226
B.4	Residual Analysis for Spatial Autocorrelation	227
C	Appendix to Chapter 3	230
C.1	Data Description	230
C.2	Euclidean Distance	232
C.3	Global Moran’s I	234
D	City-level Replication	236
D.1	Abstract	236
D.2	Introduction	237
D.3	Background	238
D.4	Data Description	241
D.5	Spatial Autocorrelation	243
D.6	Spatial Estimation and Result	247
D.7	Conclusion	250
	Bibliography	252

List of Figures

1.1	Average Industry Production in Regional Groups	18
1.2	Industry Distribution Across China by Regions	20
1.3	Industry Geography Distribution Across China by Province	22
1.4	Industry Distribution of Region Division	24
1.5	Freight Volume Percentage	28
1.6	Transportation Infrastructure Density Distributions	32
2.1	Visualization of All Eigenvectors	107
2.2	Moran Coefficients for All Eigenvectors	110
2.3	Geographic Distribution on Rail Services and Instruments	121
2.4	Geographic Distribution on Road Services and Instruments	123
2.5	Spatial Autocorrelation in Synthetic Instruments	124
2.6	Residual Analysis with the Dynamic Spatial Durbin Model	138
3.1	Geographic Distribution of Chinese Cities	157
3.2	Elbow Plots in Industry Cluster	160
3.3	K-Means Cluster Analysis in Cities	162
3.4	Comparison of Resilience and Regional GDP Growth Rate	167

3.5	Geographic Distribution of Resilience	168
3.6	Intra-City Road Network	175
3.7	Inter-City Road Network	176
3.8	Road Network Analysis from 2014 to 2021	178
D.1	Average Industry Output	239
D.2	Average Road Accessibility to Big Cities	240

List of Tables

1.1	Spatial Autocorrelation test in Moran's I with W1	43
1.2	Spatial autocorrelation tests in Moran's I with W2	45
1.3	Spatial autocorrelation tests in Moran's I with W3	47
1.4	Lagrange Multiplier Test	53
1.5	Likelihood-ratio Tests and Wald Tests	54
1.6	Baseline Results of SDM with Different Matrices	61
1.7	Spatial Effects Analysis	64
1.8	Baseline Results for East, Central, and West Regions	71
1.9	Regional Analysis with Adjacency-W1	74
1.10	Regional Analysis with Distance-W2	77
1.11	Regional Analysis with Economic-W3	79
2.1	Two-Step GMM Diagnostic Tests	117
2.2	Levin-Lin-Chu Test	126
2.3	Main Results with the Dynamic Spatial Durbin Model	128
2.4	Spatial Analysis in Short-term Effects	130
2.5	Spatial Analysis in Long-term Effects	133

3.1	Global Moran's I	183
3.2	Main Results with Dynamic Spatial Autoregressive Model	188
3.3	Spatial Analysis in Short-term and Long-term Effects	192
A.1	Data Description	215
A.2	Descriptive Statistics	215
A.3	Baseline Main Results of SDM with Different Matrices	216
A.4	Robust Estimation - with GDP growth rate	219
A.5	Robust Estimation - without financial shock	220
A.6	Robust Tests - Spillover Effects1	221
A.7	Robust Tests - Spillover Effects2	222
B.1	Data Description	224
B.2	Variables Statistic Summary	224
B.3	Two-Step GMM Diagnostic Tests - with All Instruments	227
B.4	Residual Analysis on Global Moran's I statistics - with Synthetic Instruments	228
B.5	Residual Analysis on Global Moran's I statistics - with Models	229
C.1	Data Description	231
C.2	Descriptive Statistics of Variables	231
D.1	Summary Statistics of Key Variables	243
D.2	Spatial Autocorrelation Test with Moran's I for Industry Output	244
D.3	Spatial Autocorrelation Test with Moran's I for Road Accessibility	245
D.4	Regression Results	247
D.5	Spatial Analysis	249

Declaration

I certify that the thesis submitted for examination for the PhD degree at the University of Birmingham is entirely my own work, except where clearly stated otherwise. My contributions and those of the other authors to this work have been explicitly indicated within the authorship declaration forms attached at the end of the thesis. I confirm that appropriate credit has been given within this thesis where reference has been made to the work of others.

I declare that my thesis consists of approximately 65,000 words.

Abstract

Regional economic development is rarely evenly spread across space, but rather presents geographical characteristics such as agglomeration, imbalance, path dependence, and spatial spillovers. We aim to explore how infrastructure construction, economic flow patterns, and the dynamic development of cities together shape the way regional inequality evolves over time.

This thesis consists of three essays studying how transport infrastructure spatially impacts regional economic development in China. The first chapter examines the heterogeneous spillover impacts of rail, road, and inland waterway infrastructure on regional industries to explore region-specific variations and the importance of more targeted investment strategies. Then, the second chapter introduces a novel geographic approach for constructing instrumental variables with exogenous geographical weight matrices, emphasising the importance of selecting appropriate spatial filters and synthetic tools for mitigating bias in extensive observational data. The last chapter as the third one investigates how market access to major cities influences urban economic resilience and demonstrates the significant role of connectivity in enhancing economic resilience and fostering regional cooperation.

Introduction

The spatial dimension of regional economic development reflects how economic activities are distributed, interact, and evolve across geographic space. At its core, this involves tracing the path of uneven development, spatial spillovers, and structural transformation. It is not just about mapping the total volume of economic activity in space, but about understanding the underlying forces that drive it: Why can some regions take the lead in breaking through path dependence, while others are locked in inefficient or highly polluting production methods?

As urbanisation gathers pace, economic factors become more mobile, and the global industrial structure undergoes significant restructuring, spatial economic patterns are becoming increasingly complex. We are witnessing not only the polarising effects of major cities, but also the renewed prominence of peripheral areas, driven by the expansion of transport infrastructure. Infrastructure is no longer merely a tool for linking central hubs—it has become a key force in reshaping spatial structures. New motorways, rail networks, and logistics corridors are enabling previously marginalised regions to reconnect with the core of regional economic systems.

This calls for a rethink of the role of space in economic development. Space is not simply a passive backdrop for allocating resources—it plays an active and dynamic role in steering

the flow of capital and labour, shaping firms' location choices, influencing the reach and impact of policy, and ultimately contributing to patterns of social equity and environmental sustainability.

In this context, the present study responds to these theoretical challenges through a multi-scalar empirical approach. This thesis consists of three chapters, each focusing on a different spatial scale: provinces, cities, and cross-geographic industrial structures. By analysing these levels, the research fully uncovers the impact of transport infrastructure across various spatial dimensions. We provide a multi-scale analysis that provides an empirical foundation for developing a regional development strategy.

Specifically, we employ spatial econometric models with weighting matrices, to capture spatial dependencies often discounted in traditional empirical regressions. The construction of transport infrastructure not only influences the development of the local region but also generates spillover effects in adjacent or related areas. Traditional econometric regressions typically assume that data is independent and homogeneously distributed, which fails to account for spatial dependence and heterogeneity, leading to biased estimates. Using spatial econometric models, this research more accurately identifies and analyses the spatial effects of transport infrastructure. Our goal is to understand the spatial imbalance of regional development, narrow the economic disparities between regions, and promote greater social equity.

The first chapter, “*Which Path to Growth? Heterogeneous Impacts of Transportation Modes on Regional Industry*”, studies the heterogeneous spillover impacts of three types of freight transport infrastructure on regional industries in a multimodal transport network.

The multimodal mobility network improves the efficiency of connectivity, enhances regional economic integration, and facilitates the flow of resources and factors across a broader spatial area. Our research is motivated by the challenge of how to efficiently develop a multimodal transport system in developing provinces, given limited investment resources and the constraints of the existing transport network.

We want to know whether regional economic mobility is influenced more by geographical proximity or by the similarity of regional economic levels. Therefore, we measure the spatial dependencies of rail, road, and inner waterways using three spatial weight matrices at the geospatial and economic level: adjacency, distance, and economy.

Our analysis reveals different impacts of these infrastructures across regions and weighting schemes, emphasising the necessity for region-specific multimode mobility policies. Our baseline results indicate that the spatial weight matrices show significant direct and indirect effects of rail densities on industrial production. However, the magnitude and significance of these effects are contingent upon the spatial relationship, supporting our hypothesis that the spatial scale significantly influences observed impacts.

This study offers a spatial economic perspective for investment in the mobility network, suggesting that local areas should prioritize appropriate transportation infrastructure based on their unique characteristics and needs. For regions where rail infrastructure demonstrates strong positive effects on industrial production, policymakers should focus more on expanding and enhancing train networks. Conversely, in regions where road improvements may have negative impacts, alternative strategies or complementary infrastructure should be considered. Inland waterway requires careful evaluation, particularly in regions with varying impacts, to ensure beneficial improvements and avoid hindering growth. We suggest that in-

vestment in transportation infrastructure and the construction of multimode systems should be more focused and prioritized. By adopting a targeted, context-specific approach, policy-makers ensure that infrastructure investments are effective and sustainable development on multimodal transportation networks, and increase regional industrial outcomes.

In the second chapter, “*Spatial Dynamics of Freight Efficiency: Using Synthetic Instruments in Dynamic Spatial Durbin Models*”, we introduce a novel interdisciplinary approach for constructing instrumental variables using exogenous geographical weight matrices.

In empirical studies on transportation, freight productivity is often endogenously influenced by the total volume of regional transportation. This total volume directly impacts freight productivity and also indirectly affects it by altering transportation efficiency, logistics costs, and market accessibility.

In this chapter, we focus on how to address the endogenous effects of freight productivity on industrial development. Initially, we generate eigenvectors with geographic matrices, selecting suitable synthetic instrumental variables via a supervised spatial filter. Our study underscores the importance of selecting appropriate spatial filters and synthetic tools in scenarios with extensive observational data, offering valuable insights into constructing and screening synthetic tools in panel data contexts. The effectiveness of our synthetic tool is demonstrated in a two-step GMM test, showing significant bias mitigation induced by endogenous variables. While increasing the number of synthetic tools does not compromise the test, it influences the effectiveness and efficiency of estimations. Instruments with higher spatial dependencies facilitated more comprehensive spatial effects, while those with low spatial correlation failed to capture all spatial dynamics.

The empirical analysis with synthetic instruments also illustrates the spatial performance of freight productivity in regional industries. The improvements in freight productivity positively impact the industrial output in neighbouring places. We propose several points, noting that while synthetic tools effectively mitigate residual bias in time series, they do not completely eliminate spatial autocorrelation in the residuals. This issue may arise from the heightened sensitivity of spatial models to geographic dependencies and the secondary geographical weighting applied to synthetic instruments.

The third and final chapter, “*All Roads Lead to Rome: The Role of Road Networks in the Economic Resilience of Cities*”, studies how market access to big cities impacts urban economic resilience. Urban economic resilience refers to a city’s ability to regain economic vitality after external shocks, such as epidemics.

We are interested in exploring how transportation infrastructure can facilitate the recovery of cities by helping them resume production, restart industrial activities, and reconnect with the broader market network. Our research is based on a popular topic—urban shadow theory—which explores the influence of large urban areas on surrounding regions.

We use geographic information systems on road network analysis to investigate market access to major big cities. We then employ a dynamic spatial autoregressive model to examine the impact of road networks on urban resilience. The findings indicate that market access significantly enhances urban resilience in both the short and long term. Cities with better access to large markets can gain more opportunities and resources, thereby increasing economic resilience. We estimate the spatial autocorrelation index to emphasise the interconnection of urban resilience, showing that a city’s resilience is influenced by neighbouring cities. Improving market access in one city can lead to spillover effects for surrounding

regions, both in geographic distance and also within regional industrial agglomerations.

The study also identifies a negative impact of lagged resilience on current resilience, indicating that past economic challenges can persist and further decrease regional resilience. This negative impact also extends to neighbouring areas. Essentially, the critical role of market access in improving urban resilience emphasises the need for integrated regional planning and infrastructure investment. We propose that to build a stronger and more resilient urban system, strategies should prioritize regional connectivity, linking large central cities to smaller surrounding areas for mutual support.

In summary, the three studies presented here demonstrate that regional economies do not operate in isolation; rather, they are shaped by spatial spillovers and inter-regional dynamics. We argue that the development trajectories of different regions are influenced not only by local economic factors—such as capital, labour, and technology—but also by their geographical position, institutional context, infrastructure configurations, and networked relationships. Regional development is not a simple replication of a ‘core-periphery’ model, but a dynamic process of spatial adjustment and interaction.

From a theoretical perspective, Our research integrates spatial econometric modelling with new technologies. Specifically, we incorporate geographic information systems and remote data to construct more realistic and dynamic road networks and mobile systems. We also apply machine learning to uncover hidden patterns within spatial structures, and to know how spatial mechanisms influence regional development. We contributes to a deeper understanding of spatial spillovers and path dependence of regional development.

From a policy view, our findings emphasis the significance of spatial heterogeneity and regional interlinkages. More targeted spatial interventions and differentiated infrastructure

strategies are likely to be more effective in supporting balanced regional growth. Therefore, We advise development strategies that take into account each region's unique development path and also network connectivity.

Acknowledgements

The Ph.D. is a challenging journey, filled with moments of uncertainty and happiness. Throughout this journey, I was fortunate to have the support of my supervisors, colleagues, and loved ones, whose encouragement and insights were invaluable. The challenges were worth it, as they shaped me into a more capable and confident one, ready to contribute meaningfully to my field.

I am especially thankful to Professor Horsewood, who believed in me from the very beginning and encouraged me to pursue my research. He has been both a strict supervisor and a caring elder, offering me invaluable guidance, advice, and support while always looking out for my well-being. He often noticed me working late at night and reminded me to get to rest early. Work hard and play hard—this is a lesson he imparted to me that I will carry with me always.

I am equally grateful to Professor Dickinson, who continually encouraged and supported me to keep learning and become a better researcher, especially during moments of doubt and uncertainty. He is the one who motivated me to persevere and to learn from every failure as a step toward growth.

Thanks to both of them, I was able to navigate through every challenging moment of this

journey. Without their support and guidance, this achievement would not have been possible.

I would like to extend my gratitude to my viva committee members: Professor Alessandra Guariglia, Professor Oleksandr Sasha Talavera, and Dr. Yukun Shi. I am deeply appreciative of their time and effort in reviewing my dissertation and organizing my viva.

I am especially grateful to Professor Alessandra Guariglia, who not only is my annual reviewer but also chaired my final viva. Her kindness and understanding provided me with great encouragement, making it possible to successfully complete my doctoral presentation. I am also deeply thankful to Professor Oleksandr Sasha Talavera, my internal examiner, for dedicating his time to thoroughly review my dissertation and provide invaluable guidance and suggestions. His econometric suggestions greatly benefit my future research.

I also want to express my sincere thanks to Dr. Yukun Shi, my external examiner, for travelling from Glasgow to attend my viva despite his busy schedule. His feedback has enhanced my understanding of development economics and broadened my perspective on regional research.

I am also grateful to Professor Dayong Zhang from Southwestern University of Finance and Economics for the support and care provided during my visit.

Their support and expertise have been important in my doctoral journey, and I am truly grateful.

I am profoundly grateful to my parents, Hui Wang and Quanyue Yang, for their unwavering love and support. Their belief in me has been a constant source of strength throughout this journey.

I would also like to extend my heartfelt gratitude to my colleagues on the 10th floor of the Muirhead Tower: Yibei Qu, Ye Cheng, Haokun Pang, Duiyi Dai, and Yangjun Han. Your friendship, support, and companionship have been an essential part of my journey, and I am truly grateful.

To Yibei, thank you for always encouraging me and sharing so many joyful moments together. Your positivity and enthusiasm have been a constant source of inspiration. To Ye Cheng, thanks for your warm and thoughtful support, always looking after me in every possible way and making me feel cared for. To Haokun, thank you for sharing non-spicy but delicious treats with me and also for your valuable academic support. To Duiyi and Yangjun, thank you for your companionship and for sharing both laughter and challenges during this journey.

The friendship we've built, the collaboration we've shared, and the countless cups of delicious Starbucks coffee we enjoyed together have made my time here truly memorable. You have turned the Muirhead Tower into a place of warmth and happiness for me, and I am so thankful for the moments we have shared.

Lastly, I want to express my gratitude to my room mates, Yixin Li and Yiwen Wu, for their friendship and support, as well as to my friend, Xinzhu Li, for her invaluable encouragement. Their companionship has been a source of warmth, and I am deeply thankful for the countless moments of laughter and support we've shared.

Thank you all.

Chapter 1

Which Path to Growth?

Heterogeneous Impacts of

Transportation Modes on Regional

Industry

Abstract

This chapter explores the role of multimodal transportation systems, which integrate various transport modes such as road, rail, and waterways, in enhancing regional industry production. We construct a spatial Durbin model with three matrices to capture the spatial dependences of geographical proximity and economic levels between regions. The empirical results reveal that while highways may impede industrial growth in certain regions, railways and waterways contribute positively to regional industry production. Our research suggests that the construction of multimodal transport systems should consider local economic conditions and existing infrastructure when planning transportation investments to enhance the gains for regional industries.

Keyword: Multimodal transport; Spatial Durbin Model; Heterogeneity; Regional Development.

Presented at 10th International Workshop on Regional, Urban, and Spatial Economics in China, Peking University, China, July 2023.

1.1 Introduction

Multimodal transport is widely regarded as a crucial component for achieving sustainable regional development. It combines different modes of transport, such as road, rail, and inland waterways, to optimise the strengths of each mode and minimise their weaknesses. The European Commission emphasises that integrating these transport modes and ensuring interoperability at all levels are key to achieving sustainable, energy-efficient, and environmentally respectful mobility.¹ Multimodal transportation supports economic growth by improving connectivity between regions, fostering trade, and facilitating the efficient movement of goods.

Governments worldwide are realising how important strong and well-connected transport networks are for boosting the economy and environmental sustainability. Consequently, significant funds are being allocated to modernise existing infrastructure and construct new multimodal hubs. For example, one notable example is the European Union's TEN-T, also known as Trans-European Transport Network policy, which aims to build a comprehensive network of railways, roads, inland waterways, and maritime shipping routes across Europe. Recent investments include 6.2 billion euros allocated in 2023 for 107 projects, focusing on sustainable and smart mobility.² China's Belt and Road Initiative is a prime project, aiming to connect Asia with Africa and Europe through a vast network of railways, highways, and maritime routes. This is one of the largest infrastructure projects ever undertaken, with total investment estimates ranging from 1 trillion to 8 trillion US dollars.³

¹European Commission. (2023). Sustainable transport. Retrieved from: https://transport.ec.europa.eu/transport-themes/sustainable-transport_en

²Trans-European Transport Network (TEN-T) - European Commission. Retrieved from: <https://transport.ec.europa.eu/transport-themes/infrastructure-and-investment/trans-european-transport-network-ten-t>

³Centre for Strategic and International Studies: How Big Is China's Belt and Road? Retrieved from:

The World Bank is the largest global provider of development financing for transport, with a portfolio that currently includes 34.5 billion US dollars in commitments across 165 transport projects.⁴ It is not difficult to observe that enthusiasm for transportation investment remains high around the world. However, an increasing number of researchers are doubtful about the positive returns of these huge investments, such as Faber (2014), Yu et al. (2013), and Baum-Snow et al. (2017). This hesitation is rooted in the recognition that such investments are substantial and long-term, particularly in poorer areas. While improved transportation infrastructure can potentially open markets and create opportunities, it can also result in the outflow of resources and the decline of local economic strength. In this study, we explore this critical and contentious topic in transportation research, examining whether these investments truly benefit the regions or inadvertently harm them.

Transportation linkage mechanisms allow economic activities to transcend regional boundaries, facilitating the flow of goods, services, and labour across extensive geographical areas. These mechanisms enhance regional economic interdependence and competition while generating significant spatial interactions (Feng et al., 2009; Tong et al., 2013; Yu et al., 2013; Konno et al., 2021; Albalade and Fageda, 2016; Zhang et al., 2020).

Therefore, we build on previous research examining the heterogeneous impact of transportation on regional economies and extend the analysis to spatial spillover impacts. Prior studies have primarily focused on the causes of these heterogeneous effects, such as economic structure and geographical location. Economic factors such as financial scale, development level, and economic cycles have been identified as major determinants of negative outcomes (Qin, 2016; Baum-Snow et al., 2017; Forero et al., 2020).

<https://www.csis.org/analysis/how-big-chinas-belt-and-road>

⁴The transportation projects of the World Bank, Retrieved from: <https://www.worldbank.org/en/topic/transport/overview>

We are motivated to investigate how these impact mechanisms propagate spatially within a multimodal transportation network. Specifically, we analyse the varying spatial effects related to provincial boundaries and regional distances within geographic space, and provide further spatial evidence based on differences in economic size. We construct a spatial Durbin model to explain how highways, railways, and inland waterways shape the spatial dynamics of regional industrial production.

Our research aims to understand how transportation infrastructure influences industrial development within a region and its surrounding areas, considering multiple types of spatial dependencies. We explore the optimisation of investment in multimodal transportation systems for economies of varying sizes to maximise economic benefits.

The rest of this chapter is structured as follows. Section 1.2 provides a literature review on the theoretical framework. Next, Section 1.3 describes the empirical data and background. Section 1.4 introduces methodologies and models. Then, Section 1.5 analyses the estimation results. Finally, Section 1.6 discusses the findings and concludes this research.

1.2 Literature Review

This section engages with two broad strands of literature: (1) Heterogeneous Impacts of Transportation on Regional Economies, and (2) Geographical Spillover Effects of Transportation Infrastructure.

1.2.1 Heterogeneous Impact of Transportation on the Regional Economy

The relationship between transportation and economic development has been extensively studied within the framework of economic geography and regional economics. Classical theories, such as Marshall (1890), suggest that transportation costs significantly influence the spatial distribution of economic activities. Lower transportation costs can lead to increased market accessibility, reduced input costs, and enhanced competitive advantages for firms, thereby promoting regional economic growth (Fujita and Thisse, 2002). Regions with higher accessibility are more likely to attract businesses and investments due to lower transportation costs and shorter travel times.

Theoretically, Krugman (1991) developed a model demonstrating how regions can become economically differentiated due to economies of scale and transportation costs. This research indicates that lower transportation costs and stronger economies of scale lead to industry agglomeration and production growth. Donaldson (2018) demonstrated through a general equilibrium trade model that transportation projects reduce trade costs and price gaps between regions, promoting trade growth and enhancing real income and welfare in those regions.

Empirical studies support these theoretical findings. For instance, Michaels (2008) found that cities along highways in the United States experienced increased trade activities, which promoted employment and improved labour skills in surrounding areas. Similarly, Chankrajang and Vechbanyongratana (2020) highlighted that canal transportation in 19th century Bangkok enabled orchards to transport agricultural products more efficiently to markets, thereby boosting agricultural productivity. Rahman et al. (2020) showed that China's

provincial highways facilitated employment, increased per capita productivity, and promoted trade by enhancing opportunities for regional resource acquisition and the flow of goods and services. In summary, improved transportation infrastructure extends the scope of regional economic activities and connects regions to broader markets and resources, thereby fostering economic growth and development.

However, transportation improvements can exacerbate spatial inequalities and regional disparities. The new economic geography framework highlights that while core regions with existing economic advantages may experience significant growth due to improved transportation, peripheral or less-developed regions might see limited benefits or even negative effects (Puga, 2002). This divergence occurs because core regions can better leverage transportation infrastructure to enhance their competitive edge, attracting more investment and skilled labour, while peripheral regions may struggle to catch up.

According to Deng (2013), the controversial focus of the literature on transport infrastructure can be summarised in two categories. The first is the difference in macroeconomic fundamentals, which includes regional economic conditions and cycles. Transport infrastructure enables resource flows, which may lead to resource inequality. When studying inter-provincial railways in China, Yu et al. (2013) found that regions with poor economic foundations are more likely to experience negative impacts from an increase in track length because resources are transferred to more developed areas. Qin (2016) also concluded that the flow of investment brought by high-speed rail has uneven distributional impacts on infrastructure investments; improvements in one area may come at the expense of another.

Banerjee et al. (2020) found that per capita GDP growth did not yield the expected positive impacts, as indicated by estimates of traffic accessibility between major economic cities.

Their study demonstrated that factor mobility plays a critical role in determining the nature of economic benefits. Similarly, Li et al. (2020) reached the same conclusion using a comprehensive control method for high-speed rail networks. Faber (2014) further revealed that the benefits of infrastructure are not uniformly distributed, with marginal and remote cities struggling to achieve the same level of economic dividends as their more central counterparts.

This heterogeneity is also believed to be further exacerbated by changes in economic cycles according to Tong et al. (2013), who estimated spatial benefits for US agriculture. During an economic downturn, the positive effects of transportation infrastructure are weakened and even negative for underdeveloped areas. Moving to the post-crisis period, Sharif et al. (2019) used the Quantile-on-Quantile approach to reveal a positive correlation between transportation infrastructure and economic growth, and the relationship is particularly strong in the post-crisis period.

Overall, the varied impact of transportation on regional development is shaped by a complex interplay of economic, geographic, and social factors, which can readily be observed. Economically, transportation infrastructure influences trade efficiency, market accessibility, and regional competitiveness, driving both local growth and broader economic integration. Geographically, factors such as terrain, distance, and the spatial distribution of resources and populations determine the effectiveness and reach of transport networks. Socially, transportation affects mobility, which in turn can influence labour markets and urbanisation trends.

1.2.2 Geographical Spillover Effects of Infrastructure

When these factors flow to other regions, the impact of transportation is no longer limited to the local area. Spillover effects occur when the benefits of transportation infrastructure extend beyond the immediate area of investment, influencing neighbouring regions. External economies of scale occur when the benefits of infrastructure investment are shared by multiple regions. Improved transportation infrastructure in one area can reduce transportation costs and increase market accessibility for adjacent regions, promoting regional economic integration and growth (Marshall, 1890).

Empirical studies provide robust evidence of the spatial and spillover effects of transportation infrastructure. For example, Duranton and Turner (2012) examined the impact of highways on urban growth in the United States, finding that improved highway infrastructure not only stimulated growth in cities but also generated significant spillover effects in surrounding areas. Chandra and Thompson (2000) studied the effects of highway infrastructure on non-metropolitan counties in the United States, concluding that highways promoted regional economic growth by enhancing accessibility and connectivity, leading to spillover benefits in neighbouring regions. Gibbons et al. (2019) analysed the impact of transportation infrastructure on local and regional economies in the United Kingdom. Their findings indicate that infrastructure investments not only improved local economic performance but also produced positive spillover effects in adjacent areas.

However, as summarised in the previous subsection, underdeveloped regions face negative impacts that may further expand to surrounding areas. Ahlfeldt and Feddersen (2018) discovered that the French high-speed rail has smaller spillover effects on remote areas, indicating a limited economic impact. Li et al. (2020) employed a control method to analyse

high-speed rail networks and found that the indirect economic effects, such as improved flow efficiency and altered distribution of production factors, resulted in positive economic outcomes for some cities while causing negative effects in others.

In summary, the impact of traffic on regions is driven by the flow of economic factors. Negative effects on local areas arise from various factors, such as regional economic conditions, types of transportation infrastructure, unexpected congestion, and economic cycles. These factors further expand this influence in regionally. Our study first examines the composition of transportation networks based on previous literature. We analyse the propagation of traffic impacts across regional spatial distances at both geographical and economic levels.

Therefore, we propose the following hypotheses.

- The different modes of transportation infrastructure significantly affect the flow of economic factors, resulting in varied impacts on regional economies.
- Developed regions will benefit more from transportation improvements due to resource transfer, while economically underdeveloped regions will be more vulnerable to negative impacts.
- Transportation impacts vary between geographical and economic distances, with geographical proximity potentially yielding greater spatial effects compared to economic proximity.

1.3 Data Description and Background

Data Description

We focus on the spatial relationship between industrial production related to rail and the actual kilometres of highway transportation in China. Our analysis draws on empirical data sourced from 30 provinces, municipalities, and autonomous regions over the period 2001 to 2019. These data are primarily derived from the annual provincial sections of the National Bureau of Statistics of China and the China Statistical Yearbook.⁵ Geographic coordinates, specifically latitude and longitude, are sourced from Google Maps.

Empirical data for Hong Kong, Macau, and Taiwan are not incorporated into this analysis. The economic and administrative systems in these regions differ significantly from those of mainland China, potentially leading to inconsistencies in data standardisation and interpretation. Additionally, the research does not encompass data from Tibet. This exclusion is attributed to the lack of accessible and reliable data for the region. The challenges in data collection for Tibet are multifaceted, including geographical remoteness, infrastructural limitations, and restricted availability of comprehensive statistical information. Consequently, incorporating data from Tibet under these constraints could compromise the research’s methodological rigour and the validity of its findings. The scope of this study is confined to regions where consistent and methodologically aligned data are available.

This dataset includes both economic data and exogenous geographic data, as shown in Table A.1. Our explanatory variables on the left-hand side include the density of three types of transportation facilities.

In our research, we use the value-added industrial production as the dependent variable. This

⁵China Statistical Yearbook, from: <https://www.stats.gov.cn/english/Statisticaldata/yearbook/>

metric captures the net output of industrial enterprises at the end of the year, representing the monetary value of their production activities during the reporting year. According to the indicators explanation from the National Bureau of Statistics of China, value-added industrial production reflects the results after accounting for all costs and intermediate consumption.⁶ By using industrial added value, we aim to measure the true economic contribution of industries, excluding any external costs and inputs, thus providing a clearer picture of industrial production and growth.

For the independent variable, we use existing annual transportation coverage to represent the extent of transportation infrastructure in each region. Because transportation infrastructure, particularly roads and railways, often exhibits endogeneity, infrastructure policies are not randomly assigned but tend to prioritise regions with less developed infrastructure. For instance, policy initiatives may target underdeveloped areas to boost their economic growth.

Furthermore, regions in western provinces like Sichuan and Guizhou, which have vast land areas, typically have a greater total number of railway and highway construction projects compared to smaller, more developed regions like Shanghai and Beijing. However, simply comparing the total number of infrastructure projects across provinces can be misleading due to differences in geographic size.

To address this, our measurement of transportation infrastructure—specifically railways and highways—is adjusted by calculating their coverage relative to the actual land area of each province. We calculated the density of road and rail facilities in each province based on Feng et al. (2009)’s estimation method for the regional transportation capacity of Chinese cities. This approach provides a more accurate representation of transportation infrastruc-

⁶Indicators Explanation and Calculation notes from: <https://data.stats.gov.cn/english/easyquery.htm?cn=E0103>

ture density and allows us to assess how well different regions are connected. Additionally, these data give us insight into the farthest reachability of regional transportation networks, helping us understand the extent to which transportation development has penetrated each area.

However, inland waterway transportation differs significantly from railway and highway construction because it heavily depends on the natural river topography of a region. The complexity of river landforms means that only certain areas are naturally suitable for shipping, which significantly restricts the formation and expansion of inland waterway networks. These natural constraints make it challenging to develop a consistent and widespread waterway transportation system.

To more accurately assess the inland waterway transportation conditions across different regions, we use the shipping coverage rate, which is based on the operational data of cargo ships. Inland waterway data mainly include natural waters and artificial canals that are passable by ships. This metric was chosen because it not only reflects the geographical distribution of navigable waterways but also incorporates the actual functioning of the shipping network. By considering both the physical infrastructure and the operational dynamics, the shipping coverage rate provides a more comprehensive view of the service capacity and efficiency of the inland waterway transportation system. This metric is used to measure the specific contribution of actual navigable waters to industry.

Moreover, the farther a ship can travel, the larger the service area covered by the shipping network. Thus, the shipping coverage rate also helps to capture the extent of reachability and connectivity offered by inland waterway transportation in each region, making it a valuable measure for understanding the system's overall performance. Conversely, an increase in

the number of boats, without a corresponding increase in river mileage, indicates increased freight traffic but no substantial improvement in waterway infrastructure, thus reducing the ‘inner river per boat’ ratio. When there is no freight boat in a province, it means the inner river does not function as a freight waterway, so it accounts for no value and is equal to zero. This approach allows us to assess the efficiency and capacity of waterway transportation in relation to the physical infrastructure in place.

In our study, we also estimate the econometric models with several control variables to account for regional characteristics. Education expenditure is used as a proxy for the education level, reflecting the region’s investment in skills development, which influences workforce productivity and, consequently, industrial output. Population size is also controlled for, as it provides both a labour pool and a consumer base that are critical for industrial activity. In the appendix, Table A.1 offers a detailed description of the key data variables used in our analysis.

Additionally, nominal economic indicators are adjusted for price changes using the production price indices to obtain real values, thereby facilitating an accurate temporal comparison by controlling for inflationary effects. Foreign investment data are unified into Yuan through the annual average exchange rate. The logarithmic transformation is not applied to transportation data and the GDP growth rate to preserve the linear nature of growth measurements over time. We present a description of the data and statistics in the appendix as Table A.2.

Secondly, for the geographical component of our research, we have employed Google Maps to obtain geographic data. Specifically, we have extracted latitude and longitude coordinates for multiple locations within each province. Instead of selecting the geographical centre

or the capital cities as reference points, which might not accurately represent the entire province, we choose for a randomised approach to point selection. This strategy involves choosing random cities centre within the boundaries of each province to capture a more diverse and representative sampling of geographic locations. We aim to mitigate potential biases that might arise from using only central or administrative locations, thus allowing for a more nuanced analysis of geographic factors in relation to economic data.

Regional Diversity and Industry Development

We adopt the regional provincial classification framework proposed by Yu et al. (2013) and National Bureau of Statistics (2020), which integrates geographic spatial connectivity with China’s established regional divisions. This classification method considers both the physical proximity of provinces and the socio-economic characteristics that define regional groupings in China. By incorporating spatial connectivity, the framework aligns with the natural flow of economic activities and interactions across provinces. Also, it reflects the practical administrative and economic divisions that influence regional policy-making and development strategies.

According to National Bureau of Statistics (2020), the provinces are divided into four regions based on their geographical location and economic conditions. These are the eastern region, northeastern region, central region, and western region. Since the northeastern region only includes three provinces—Liaoning, Jilin, and Heilongjiang—we merge them into the eastern region based on their geographical location. Therefore, the 30 provinces for this study are divided as follows:

- Eastern Region: Beijing, Tianjin, Hebei, Shanghai, Jiangsu, Zhejiang, Fujian, Shandong, Guangdong, Hainan, Liaoning, Jilin, and Heilongjiang

- Central Region: Shanxi, Anhui, Jiangxi, Henan, Hubei, and Hunan
- Western Region: Inner Mongolia, Guangxi Zhuang Autonomous Region, Chongqing Municipality, Sichuan Province, Guizhou Province, Yunnan Province, Shaanxi Province, Gansu Province, Qinghai Province, Ningxia Hui Autonomous Region, and Xinjiang Uygur Autonomous Region

This regional division is not only a geographical and policy distinction, but also combines industrial production data. This division into three groups allows us to capture these distinct regional characteristics and dynamics, enabling a more nuanced analysis of how transportation and regional policies influence industrial production and economic development across different contexts.

According to the definition of China’s National Bureau of Statistics, the industrial sector comprises three main categories: mining, manufacturing, and electricity, heat, gas, and water production and supply.⁷

These categories cover a wide range of industrial activities, from resource extraction and product manufacturing to energy and water production and supply, and represent the entire range of industries included in China’s industrial value-added statistics.

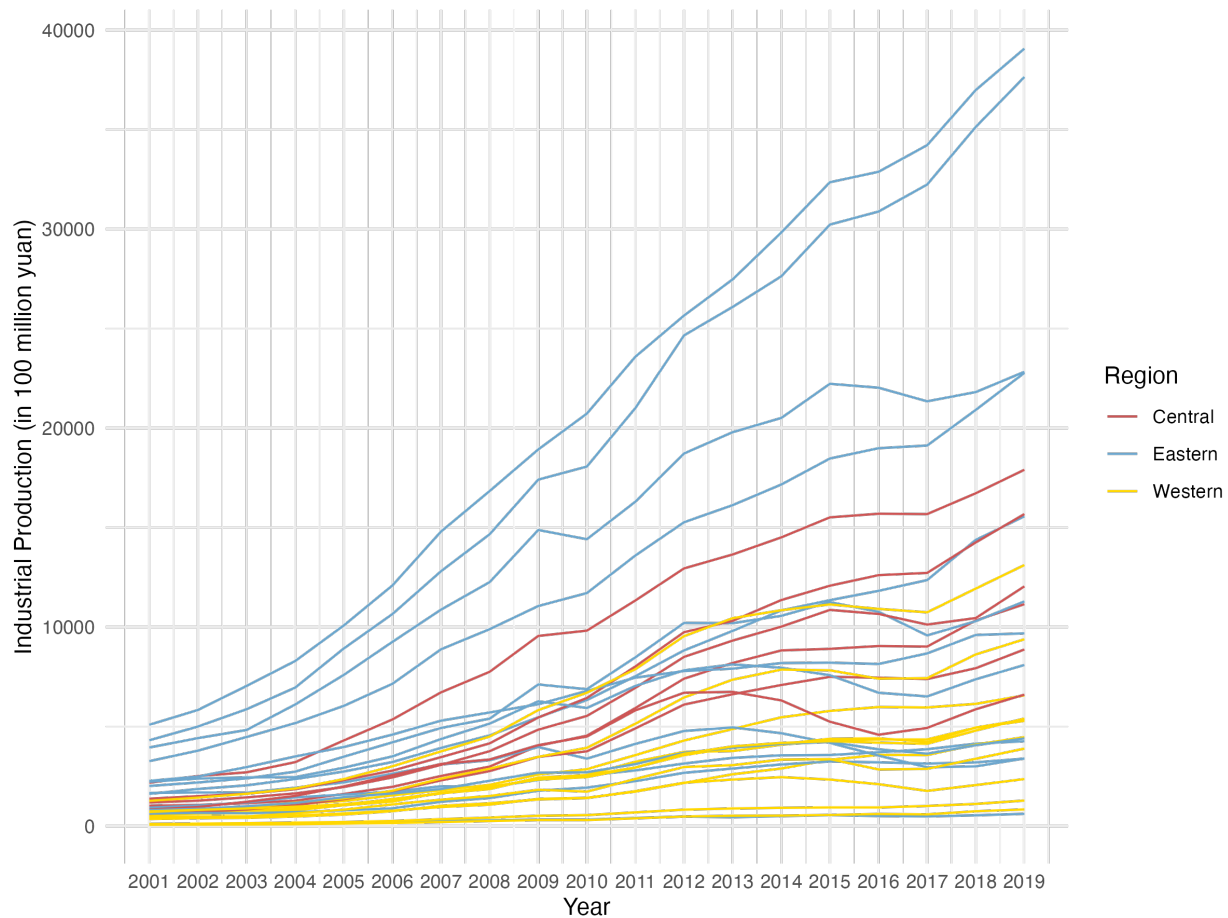
Over recent decades, especially since the economic reforms, China’s industry has made historic progress. Industrial value-added exceeded 1 trillion yuan in 1992, 10 trillion yuan in 2007, 20 trillion yuan in 2012, and 30 trillion yuan in 2018. By 2018, China’s manufacturing value accounted for more than 28% of the global total, making it a crucial driver of global industrial growth.⁸ According to data from the World Bank, the share of the manufacturing

⁷National Bureau of Statistics, Statistical systems and classifications, The definition of Industry in China, available from: https://www.stats.gov.cn/hd/cjwtjd/202302/t20230207_902279.html

⁸The statistics come from the State Council of China, available from <https://www.gov.cn/xinwen/>

industry in GDP has slightly declined from around 30% at the beginning of the 21st century but has remained above 20%, demonstrating the significant role of the manufacturing industry in the economy.⁹

Figure 1.1: Average Industry Production in Regional Groups



Note: the line graph shows the industrial production (in 100 million yuan) for each province over time.

We provide a comparative overview in Figure 1.1, which features a line graph illustrating the

2019-09/21/content_5431841.htm

⁹The data are from the World Bank, available at: <https://data.worldbank.org/indicator/NV.IND.MANF.ZS?locations=CN>

average industrial production across three regional groups in China—Central, Eastern, and Western—from 2001 to 2019. Each region is represented by a distinct colour, corresponding to the provinces included in that group.

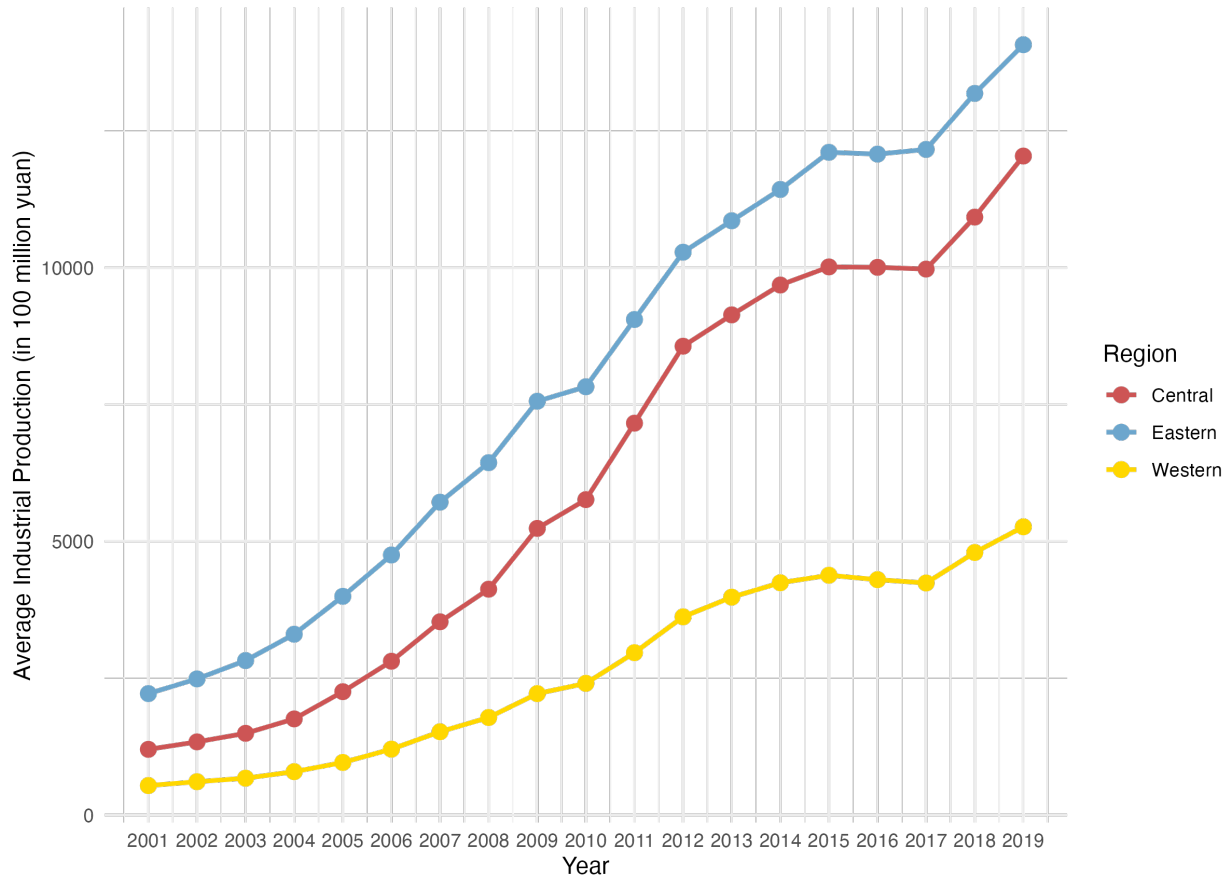
The x-axis spans the years 2001 to 2019, while the y-axis displays industrial production in units of 100 million yuan. The figure captures the industrial production trends for all provinces within each regional group over the observation period, highlighting the differences in economic performance across regions. By using distinct colours to represent each region, the figure emphasizes the rationality of this regional division and visually demonstrates the disparities among the groups. This graphical representation reinforces the validity of the chosen regional classification for our analysis.

We further illustrate the differences among three regions and the reasons for the division by taking the average of each region.

Figure 1.2 illustrates the trends in average industrial production (in 100 million yuan) across three major regions in China—Eastern, Central, and Western—between 2001 and 2019. A clear overall growth pattern is evident for all regions, reflecting the broader economic development of the country during this period. However, significant disparities in industrial production levels persist across these regions.

The Eastern Region consistently demonstrates the highest industrial production, with values rising steeply from approximately 5,000 units in 2001 to over 15,000 units in 2019. This region's dominance can be attributed to its economic advantages, such as advanced infrastructure, higher foreign investment, and its strategic coastal location. The steep and almost linear growth trend reflects the region's strong and consistent economic performance over time.

Figure 1.2: Industry Distribution Across China by Regions



Note: This graph illustrates the average data in regional group among the three regions

The Central Region shows a more moderate growth trajectory. Starting below 5,000 units in 2001, it reached approximately 10,000 units by 2019. Although the region has improved significantly, it still lags behind the Eastern Region.

The Western Region exhibits the lowest levels of industrial production, starting below 2,000 units in 2001 and rising steadily to around 5,000 units by 2019. Despite its slower growth, the Western Region's upward trajectory suggests the impact of development policies, such

as the “Western Development Strategy,”¹⁰ aimed at reducing regional disparities.

We also provide a more detailed graph by province in Figure 1.3, which illustrates the geographic distribution of industrial production across China’s provinces. We employ two key elements: the size of the circles to represent the percentage of each province’s industrial production relative to the national total, and a colour gradient to indicate the real production values.

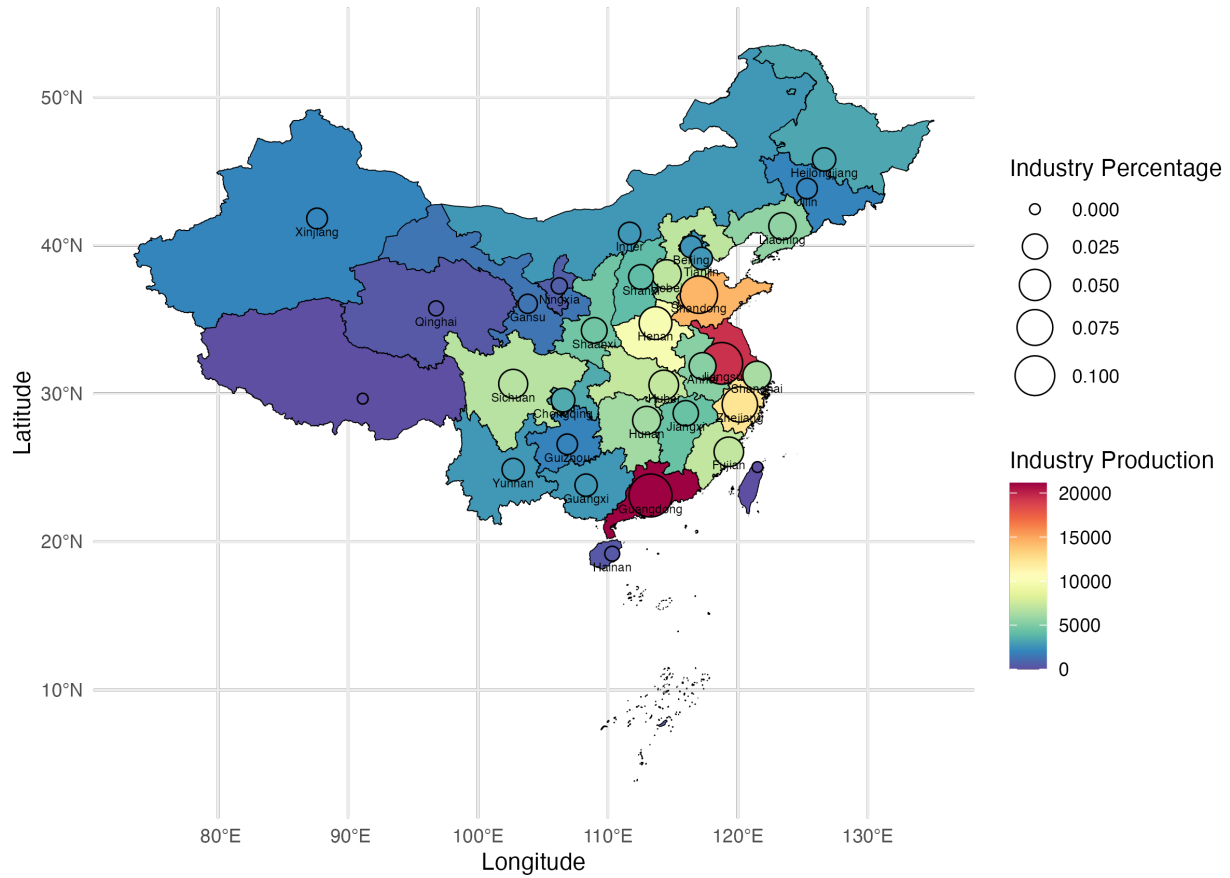
The map effectively highlights the industrial production landscape in China, showcasing pronounced regional disparities. Coastal provinces lead significantly, both in their share of total production and in absolute output levels, while inland provinces fall noticeably behind.

The colour gradient on the map ranges from blue, indicating low production, to red, indicating high production. Coastal provinces such as Guangdong, Jiangsu, and Zhejiang are depicted in warmer colours like red and orange, signifying high industrial output.

These regions are known for their advanced infrastructure, strategic location with access to international markets, and significant foreign investment, all contributing to their high production levels. In contrast, inland provinces like Qinghai, Gansu, and Ningxia are shown in cooler colours like blue and green, indicating lower industrial production.

¹⁰Western Development Strategy, from the State Council of China, available from https://www.gov.cn/gongbao/content/2001/content_60854.htm

Figure 1.3: Industry Geography Distribution Across China by Province



Note: This map illustrates the industrial distribution across various provinces in China. The size of the circles indicates the percentage share of each province's industry in the total national industrial production, while the colour gradient reflects the actual production values.

The size of the circles on the map represents the percentage share of each province's manufacturing industry in total national industrial production. Larger circles are concentrated in provinces such as Guangdong, Jiangsu, Shandong, and Zhejiang, highlighting their substantial contribution to China's industrial sector. Guangdong, in particular, stands out with

the largest circle, underscoring its role as a major industrial and economic hub. In contrast, smaller circles are seen in provinces like Xinjiang and Qinghai, indicating their relatively smaller share in national industrial production.

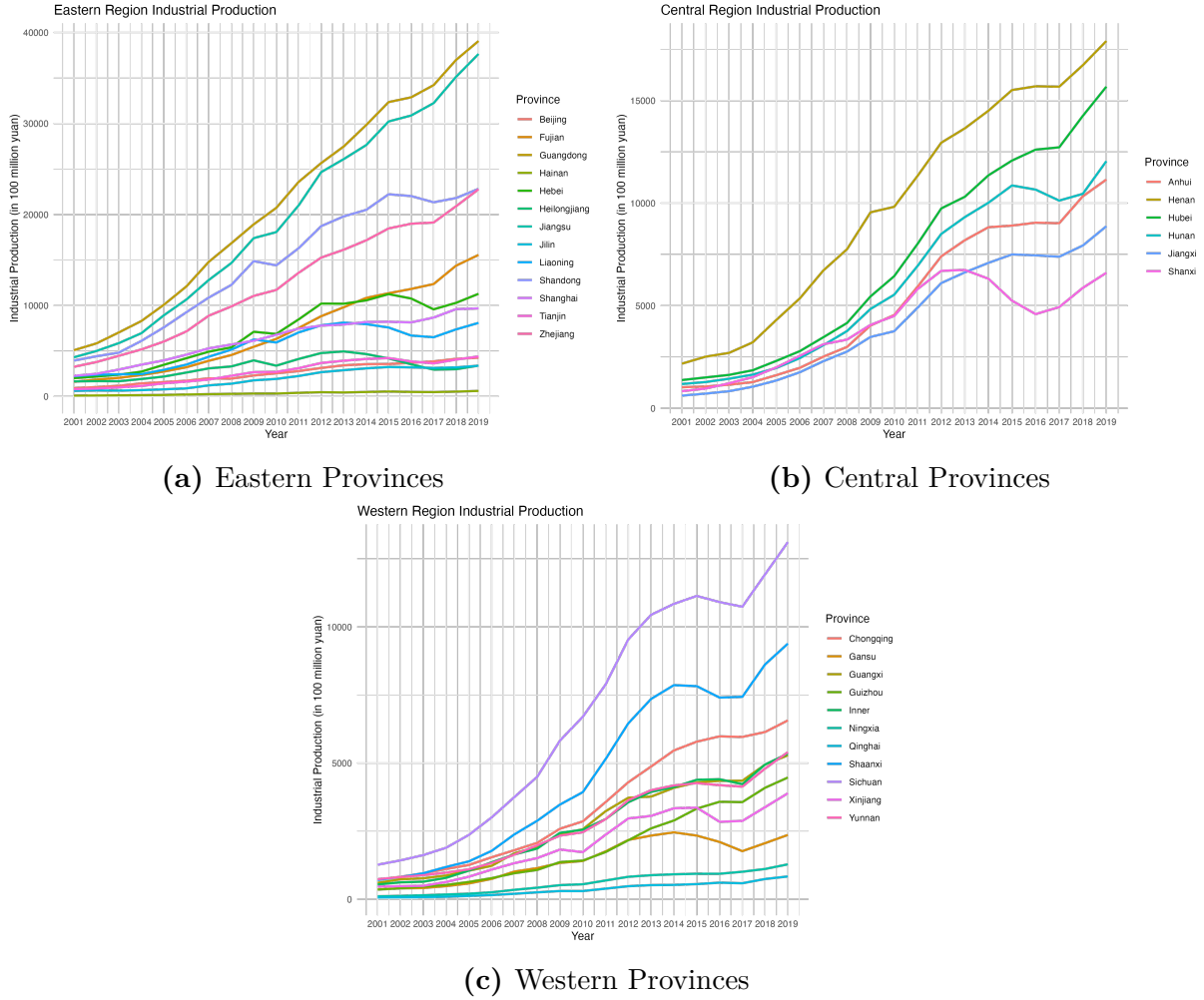
Overall, there are clear differences in average industrial production among the three regions. Industrial production generally rises from the Eastern region to the Central region and then declines in the Western region. To better understand the temporal dynamics within each region, we further analyse the industrial production trends of provinces in these regions from 2001 to 2019.

The Eastern region, as shown in Figure 1.4a, demonstrates a remarkable upward trend in industrial production, in 100 million yuan, across its provinces between 2001 and 2019 but with significant variations in growth rates and output levels.

Top-performing provinces, such as Shanghai, Jiangsu, and Zhejiang, have production outputs exceeding 40,000 units by 2019. In contrast, mid-tier provinces, which started with moderate growth, exhibit steady but slower increases, with outputs ranging from 20,000 to 30,000 units by 2019, suggesting smaller-scale industries.

Lagging provinces, with production below 10,000 units even in 2019, display minimal growth. The most significant growth occurred between 2005 and 2010, with many provinces experiencing sharp production increases. Post-2015, the leading provinces maintain high growth trajectories.

Figure 1.4: Industry Distribution of Region Division



Note: These three graphs present the time trends of industrial production from 2001 to 2019, highlighting regional details and growth dynamics across the East, Central, and Western regions.

The Central Region, as shown in Figure 1.4b, shows a diverse trajectory of industrial production, in 100 million yuan, from 2001 to 2019. The top-performing province surpasses 10,000 units by 2019, experiencing accelerated growth between 2005 and 2010. Mid-tier provinces, with outputs between 5,000 and 8,000 units, reflect steady expansion but remain behind the leader, while lagging provinces exhibit minimal growth, staying below 3,000 units

throughout the period. Growth dynamics reveal a regional surge from 2005 to 2010, followed by a slowdown for some provinces after 2015.

The Western Region, as shown in Figure 1.4c, displays significant growth in industrial production, in 100 million yuan, from 2001 to 2019. The leading province, represented by the yellow line, has 15,000 units by 2019, showing rapid growth after 2005. Mid-tier provinces, such as those represented by green and red lines, follow with outputs ranging between 10,000 and 15,000 units by 2019, reflecting steady growth but a slower pace compared to the leader. Provinces in the lower tier, including those with flatter growth curves, remain below 10,000 units. The Western Region experienced a growth surge between 2005 and 2010, similar to other regions.

While output levels remain lower overall, the average growth rates—especially from 2005 to 2010—mirror the patterns observed in other regions, reinforcing the persistence of a tiered growth structure across provinces.

We divide the provinces within each region based on their geographical location and industrial development levels. Provinces that are geographically close and exhibit similar industrial production capacities are grouped together. The grouping reflects both spatial proximity and comparable economic characteristics. By organizing the provinces in this way, we can better capture and analyse the industrial interactions and spillover effects that occur within and between geographically connected areas, providing deeper insights into the spatial dynamics of industrial development for later spatial analysis.

Transportation Analysis

Transport Modes

China's provinces exhibit significant disparities not only in industrial production but also in the state of their transportation infrastructure. These differences are reflected in the total freight volume handled in each region.

According to the annual report from the Chinese Ministry of Communications on the transportation sector's freight traffic,¹¹ China predominantly relies on railways, highways, and waterways for freight logistics.

China's railway and highway freight systems also underwent significant expansion and modernisation to meet the demands of its rapidly growing economy and logistics needs. Railway freight volume notably increased from approximately 2 billion tonnes in 2000 to over 4 billion tonnes by 2019, driven by massive infrastructure and industrialisation projects, with substantial investments in expanding and upgrading thousands of kilometres of railway lines. Meanwhile, highway freight experienced a surge with extensive upgrades to the road network. The total road network expanded to over 4 million kilometres by 2010.

The proportion of freight transport in China in 2019. When looking on industry transportation in China, we focus more on the top three freight modes—road, rail, and waterways—due to their substantial contribution to the total freight volume and their major roles in supporting industrial logistics. Together, these three modes account for nearly the entirety of freight transport, with road dominating at 74%, followed by waterways at 16%, and rail at 9%.¹² These modes form the backbone of industrial supply chains, efficiently handling bulk materials, raw resources, and manufactured goods that are essential to China's industrial sectors.

¹¹The report summarises the transport environment in Chinese. Ministry of Transport of the People's Republic of China. It is available from https://www.gov.cn/xinwen/2020-05/12/content_5510817.htm

¹²The report summarises the transport environment in Chinese. Ministry of Transport of the People's Republic of China. It is available from https://www.gov.cn/xinwen/2020-05/12/content_5510817.htm

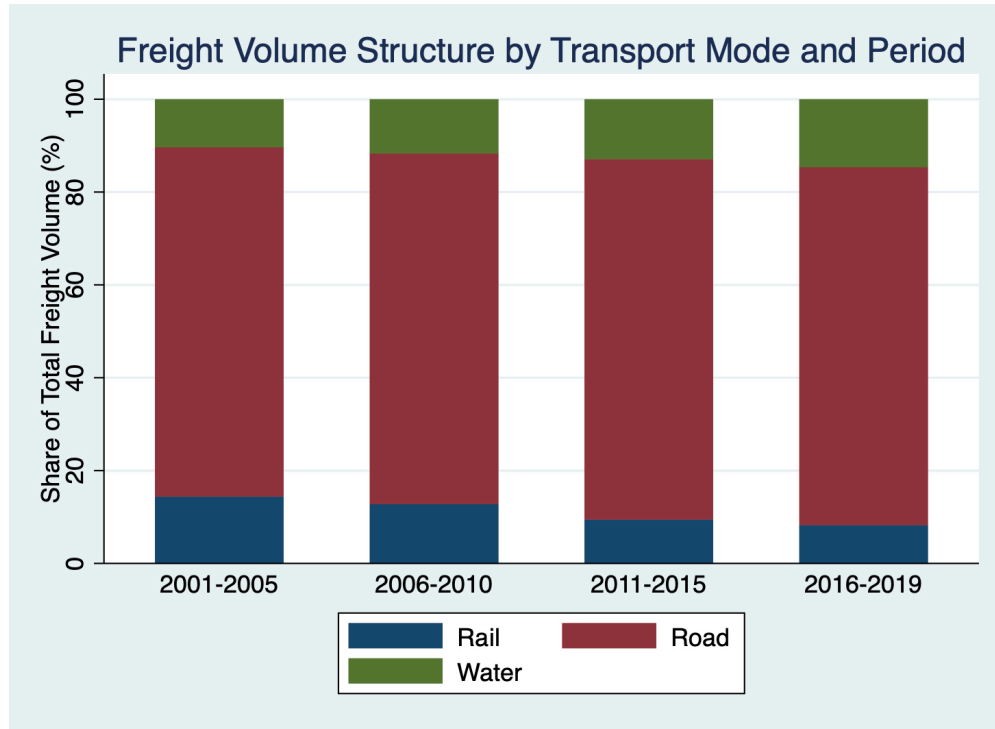
Air transport, in contrast, is excluded from the research scope as it accounts for less than 1% of the total freight volume. While air freight is indispensable for high-value, time-sensitive goods like electronics or pharmaceuticals, its application in general industrial transportation is limited due to its high costs and lower capacity compared to other modes. Industrial logistics typically require cost-effective and large-scale freight solutions, which are better addressed by road, rail, and waterways. Focusing on these three modes allows the study to provide deeper insights into the transportation systems that are most impactful for regional industrial development.

Figure 1.5 analysis examines the relative shares of freight volume transported via rail, road, and inland waterways across four time periods: 2001-2005, 2006-2010, 2011-2015, and 2016-2019. The findings reveal key structural trends in China's freight transport system and highlight the dominance of specific modes over time.

Road transport consistently holds the largest share of total freight volume across all observed periods. In 2001-2005, road freight accounted for approximately 75-80% of the total, reflecting its critical role in China's logistics network. By 2006-2010, this share had slightly increased to around 80-85%, indicating continued reliance on highways for industrial transportation. In the subsequent periods, 2011-2015 and 2016-2019, the proportion of road freight stabilized at around 85%.

Rail freight, once a key component of China's transport infrastructure, has seen a gradual decline over time. In 2001-2005, rail transport accounted for approximately 15-20% of total freight volume. However, by 2006-2010, its share had decreased to 10-15%, reflecting a shift towards road-based logistics. The decline continued through 2011-2015 and 2016-2019, with rail freight stabilizing at around 10

Figure 1.5: Freight Volume Percentage



Note: This chart illustrates the distribution of freight volume by transport mode in China. Data source: Ministry of Transport of the People's Republic of China.

Inland waterways maintain a relatively small yet stable share of total freight transport. In 2001-2005, waterborne freight accounted for less than 10%, reflecting its limited role in national logistics. A slight increase was observed in 2006-2010, reaching around 10%, and this share remained relatively stable at 10-12% in the following periods. Unlike road and rail transport, inland waterways are highly dependent on seasonal and geographical factors, which restrict their year-round utility. In northern regions, winter freezes can disrupt waterway transport, while in the south, seasonal variations in river flow affect navigability. Therefore, we focus on rail, road, and waterway transportation, as they collectively account

for the vast majority of freight volume in China. They play an important role in supporting industrial freight needs, ensuring the efficient movement of goods and underpinning regional industrial development. Air transport is excluded from this analysis due to its relatively minor role in handling large-scale industrial freight, thus making rail, road, and waterways the primary focus of our study.

Mobility and Regional Development

Railway transport is primarily used for moving large equipment and key industrial resources, making it essential for major industrial cities. However, railways are not well-suited for short-distance, small-volume freight, limiting their benefits for agriculture and small-scale manufacturing. Furthermore, industries that rely on rail transport are often capital-intensive, offering limited support for small and medium-sized enterprises. In terms of market connectivity, railways enhance long-distance trade and facilitate regional market integration.

The long-term costs of railway transport are relatively low, making it efficient for large-scale freight movement and sustainable long-term operations. However, high construction and maintenance costs remain a major challenge. Some rail lines struggle with low passenger flow and persistent financial losses. Additionally, once a railway line is established, its infrastructure is difficult to modify, making it less adaptable to future shifts in industrial geography.

Second, road transport is the most flexible mode, particularly for short to medium-distance transportation. Highways play a key role in supporting economic activities in small and medium-sized cities as well as rural areas. Expanding the highway network not only promotes urban growth and suburban development but also supports the expansion of manufacturing and logistics industries. However, highways require significant investment, and maintenance

costs—especially for infrastructure such as viaducts and tunnels—can impose a heavy fiscal burden.

Compared to rail and water transport, road infrastructure is cheaper to build, offers faster returns on investment, and is easier to maintain, making it particularly suitable for developing countries. However, in the long run, road transport carries high maintenance costs, especially when traffic congestion leads to road deterioration, adding further financial strain.

13

Water transport plays a crucial role in boosting port economies, supporting the growth of major global trade hubs such as Shanghai, Rotterdam, and Singapore, which serve as logistics centres for international trade. However, water transport is heavily dependent on natural geography, limiting its accessibility for inland cities and potentially widening economic disparities between coastal and inland regions.

Waterways are most suitable for transporting bulk commodities such as coal, ore, oil, and grain due to their low per-unit transport costs. However, they are not ideal for industries requiring fast deliveries, such as technology, healthcare, and consumer goods. While water transport remains the most cost-effective mode, it is constrained by port infrastructure and slow transit speeds, further limiting its suitability for time-sensitive industries.

Each mode of transportation—rail, road, and water—has its strengths and limitations. Railways support large-scale economic integration but require a long investment return period. Road transport offers flexibility and promotes regional economic activity but may contribute to uncontrolled urban sprawl. Water transport is the most cost-efficient but is geographically restricted.

¹³Baidu Congestion Data and Analysis, from: <https://jiaotong.baidu.com/congestion/city/urbanrealtime/>

To maximize economic efficiency, transportation planning should balance these different modes, ensuring they complement each other. Future infrastructure development must consider economic sustainability, coordination between government policies and market dynamics, and the strategic optimization of transportation networks. China has consistently emphasized a well-integrated transport system of various transport modes and the need to accelerate the construction of a strong transportation country in the new era in the construction of transportation infrastructure.¹⁴

Spatial Distribution of Infrastructure

We also aim to uncover the spatial patterns and regional dynamics that shape industrial transportation in China by analysing the geographical distribution of these three key modes.

The first map, Figure 1.6a, visualises the distribution of road transportation across China. The colour gradient indicates the density of road networks within each province. Regions with warm colour shades represent higher road network densities, while cool colour shades indicate lower densities.

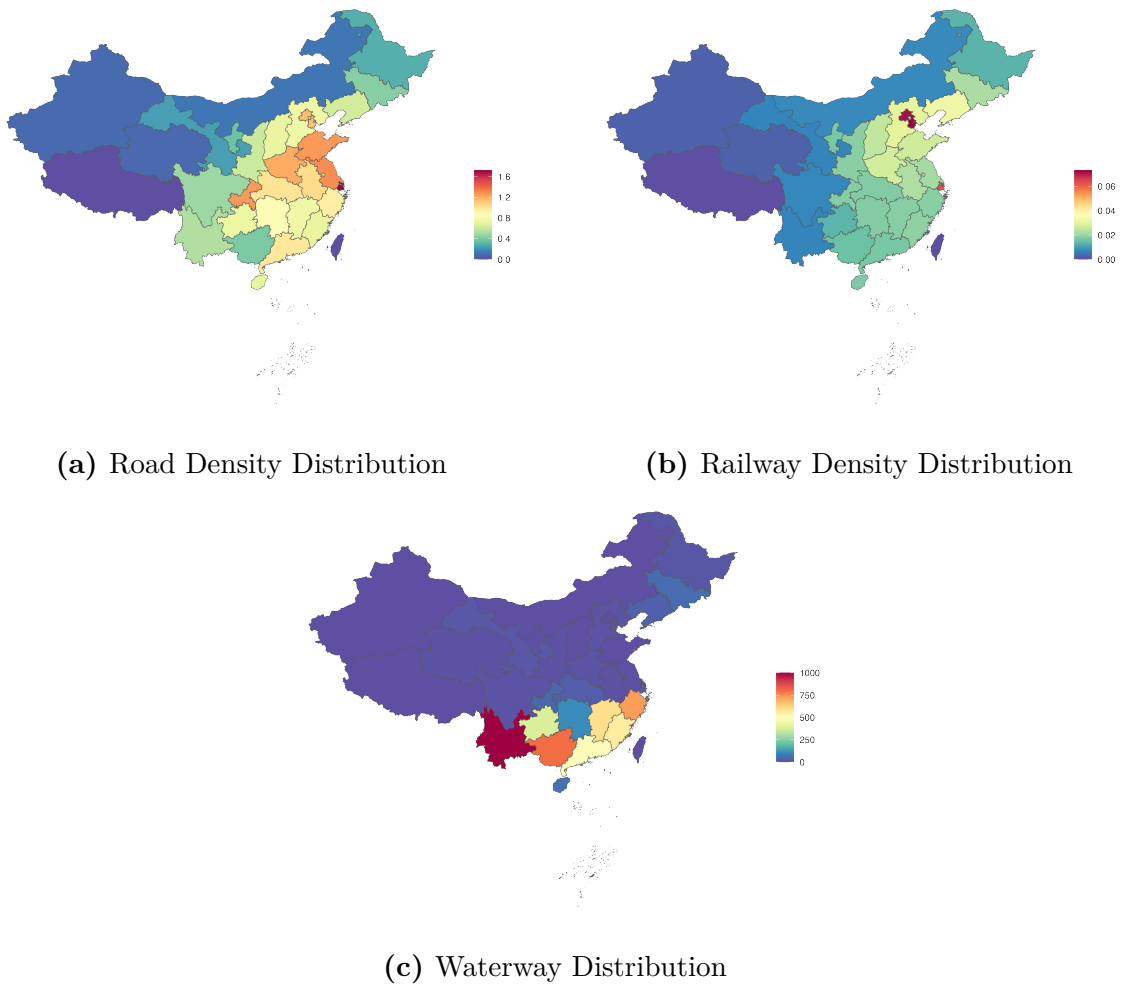
From the map, it is evident that the eastern and central parts of China, particularly areas such as Jiangsu, Shandong, and Henan, exhibit higher road network densities. Conversely, the western provinces such as Qinghai and Xinjiang display lower road densities, reflecting lower transportation capacity and possibly geographical challenges that impede road construction.

The second map, Figure 1.6b, illustrates the rail transportation infrastructure. Similar to the road distribution, warm colours signify greater rail network densities. The map highlights substantial rail infrastructure in regions such as Hebei, Henan, and Jiangsu, aligning with

¹⁴China Communication and Transportation Association, from: <https://www.cctaw.cn/jujiao/13286.html>

China's efforts to enhance connectivity and support economic growth in these areas.

The central and northeastern regions also show considerable rail densities, facilitating efficient freight and passenger movements. However, the less developed western regions have significantly lower rail infrastructure.



Note: Warm colours such as red and orange indicate higher values; cool colours such as green and blue represent smaller values.

Figure 1.6: Transportation Infrastructure Density Distributions

The third map, Figure 1.6c, presents the distribution of waterway transportation, focusing on the inland river per boats (IRB). In this case, the colour intensity reflects the extent and development of inland waterway networks. Provinces along the Yangtze River, such as Jiangsu, Anhui, and Hubei, show higher values, which is expected given the Yangtze's status as a crucial waterway for freight transportation.

Similarly, areas around the Pearl River Delta, including Guangdong, exhibit significant waterway infrastructure, supporting their roles as major economic hubs. The less coloured regions, particularly in the north and west, indicate a scarcity of navigable rivers or under-developed waterway networks.

The transportation infrastructure in China exhibits distinct regional disparities, with the eastern and central regions typically better served by road, rail, and waterway networks. This distribution aligns with the concentration of economic activities, urban centres, and population densities in these areas.

The western regions, characterised by more challenging topographies and lower population densities, display less developed transportation infrastructures. This distribution is similar to the industrial production distribution.

Standardization of Transportation data

As discussed in the previous sections and chapters, different transportation modes possess unique characteristics, measurement units, and scales. In this chapter, we focus on two primary reasons for standardizing transportation data.

First, standardization unifies transportation data across different modes and spatial scales, enhancing the comparability of results. By converting data into a consistent framework,

standardization allows us to better analyse and interpret the impacts of various transportation modes in a cohesive manner.

These variations in metrics can lead to inconsistencies when analysing or aggregating data across modes. The diversity of transportation data necessitates standardisation, as highlighted by Washington et al. (2020). Transportation modes such as roads, railways, and waterways are characterised by distinct measurement units and operational scales, including kilometres of paved surfaces, track lengths, navigable waterways, or tonnage capacities. These differences can lead to bias when integrating or comparing data across modes.

Second, standardized transportation data represent deviations from the overall average transportation level. We address more on the differences in transportation infrastructure across regions. It enables us to identify and analyse regional disparities more effectively.

Without standardisation, the disparities in how transportation data are measured and reported can introduce biases and distortions in economic analysis. If these data are not harmonised into a common framework, any comparison or aggregation may inaccurately reflect their true contributions. One example is that a region with extensive but under-utilised railways might appear more economically connected than one with highly efficient but shorter waterways, simply because the metrics are incompatible. As Washington et al. (2020) points out, using consistent frameworks and methods not only makes transportation data more reliable but also helps ensure accurate economic estimation later.

We use z-score standardisation, also known as standard score normalisation, which is widely used in dealing with variables measured on different scales or units (Gelman et al., 2021). It standardises data by centring it around a mean of zero and scaling it to have a standard deviation of one (Gelman et al., 2021). By subtracting the mean from each observation and

dividing by the standard deviation, the transformed values represent the number of standard deviations each observation is from the mean.

The mathematical basis for Z-score standardization lies in its transformation equation, which is defined as $z_i = \frac{x_i - \mu}{\sigma}$, where z_i represents the standardized score, x_i is the original value of the i -th observation, μ is the mean of the dataset, and σ is its standard deviation (Gelman et al., 2021).

To facilitate cross-variable comparison and ensure model stability, all transportation variables were standardised prior to regression estimation. This approach allows the estimated coefficients to reflect the effect of a one standard deviation increase in each transport variable, which is especially useful when the original units differ substantially.

However, for economic interpretation, we convert these standardised coefficients into elasticities, using the mean and standard deviation of the original, unstandardised data. This allows us to express the estimated effects in terms of the percentage change in output resulting from a 1% change in each transport indicator.

1.4 Methodology

In this section, we present the spatial identification and modelling of traffic networks. We begin by capturing the most fundamental spatial relationships and constructing spatial weight matrices to represent these connections. Next, we identify the spatial correlations and dependencies within traffic networks using these matrices. Our analysis reveals that the impacts of transport networks and infrastructure are spatially dependent.

We employ spatial econometric models to capture the interactions and spillover effects be-

tween regions. We expect these models to show how transport development in one region can influence economic activities and transport efficiency in neighbouring areas. According to Elhorst (2014), spatial interactions may lead to unexpected bias in traditional regression. This approach helps to avoid biases that could result from ignoring spatial factors, providing a clearer understanding of the regional effects of transport infrastructure.

1.4.1 Preparation of the Estimation

The weight matrix defines the structure of spatial dependence and is a cornerstone for analysing spatial data and modelling spatial interactions. The weight matrix W is used to specify the spatial relationships among observations. This matrix typically consists of binary values or distances that represent the proximity or connectivity between spatial units (Anselin, 1988). It determines the spillover effects, which refer to the impact that changes or events in one spatial unit have on neighbouring units.

The weight matrix as W is a positive matrix ($N * N$) that describes the dependence relationship between units, where the element W_{ij} reflects the connection between region i and region j . Researchers have proposed various methods for constructing the spatial weights. Common methods for constructing weight matrices include using contiguity, as neighbours sharing a common boundary, distance-based weights, and k-nearest neighbours (Getis and Aldstadt, 2004). In a word, they are divided into two categories: geographical spatial weights, which include geographical boundaries and actual geographical distances; and economic or social factor-related weights, such as income level and per capita GDP of regions.

We construct three types of weight matrices to explain the geographic and economic dependencies. The first type is the contiguity-based weight matrix, which assigns a weight of 1 to

spatial units that share a common boundary and 0 otherwise. This matrix is particularly useful for capturing direct spatial interactions, such as those found in urban areas where neighbouring regions are closely linked (Cliff and Ord, 1981). The contiguity matrix can be either binary, indicating simple adjacency, or row-standardised to ensure that each row sums to one, thereby facilitating the interpretation of the spatial lag and error terms in econometric models (Anselin, 1988).

In our case, an adjacency weight matrix, particularly in the context of geographical areas like provinces, is a specialised form of the adjacency matrix used to represent spatial relationships between regions. If two provinces share a common border, the corresponding cell in the matrix is set to 1. This cell is found at the intersection of the row and column corresponding to these two provinces. If two provinces do not share a common border, the corresponding cell in the matrix is set to 0.

Our use of this matrix aims to test whether policy benefits spill over within the same policy region—that is, whether a policy implemented in one province also affects neighbouring provinces. Specifically, we seek to analyse the spatial interactions between provinces to determine if policies, such as fiscal spending or infrastructure development, in one province have positive or negative spillover effects on economic activities, industrial production, or other economic indicators in neighbouring provinces through either geographical proximity or economic linkages.

$$W_{ij} = \begin{cases} 1, & \text{if } i \text{ and } j \text{ are neighbours,} \\ 0, & \text{not neighbours.} \end{cases}$$

The second type is the distance-based weight matrix, which defines weights based on the

physical distance between spatial units. Typically, weights decrease with increasing distance, reflecting the principle of spatial decay, where closer units exert a stronger influence on each other than those further apart (Tiefelsdorf, 2000). Various functional forms, such as inverse distance or exponential decay functions, are used to model this spatial relationship. Distance-based matrices are especially relevant in contexts where spatial interactions decrease with distance (LeSage and Pace, 2009).

Each cell in the matrix represents the inverse of the distance between the two locations corresponding to that cell's row and column. Commonly, the inverse distance is calculated as 1 divided by the distance, so if the distance between two locations is d , the corresponding cell value would be $1/d$ (Shepard, 1968). This assumes d is not zero; for cells representing the distance from a location to itself, which would be zero, the value is typically set to 0 to avoid division by zero.

$$W_{ij} = \begin{cases} \frac{1}{d_{ij}}, & \text{if } i \neq j \text{ and } d_{ij} \neq 0 \\ 0, & \text{if } i = j \end{cases}$$

The third type is the economic dependence weight matrix, which accounts for economic linkages between spatial units. This matrix may be constructed using economic indicators such as trade flows, employment patterns, or industry linkages (Rey and Montouri, 1999). This type of matrix captures economic spillovers and dependencies that transcend mere geographic proximity. This approach is particularly valuable for analysing regional economic policies and understanding the interconnectedness of economies across spatial scales (Fingleton, 2003).

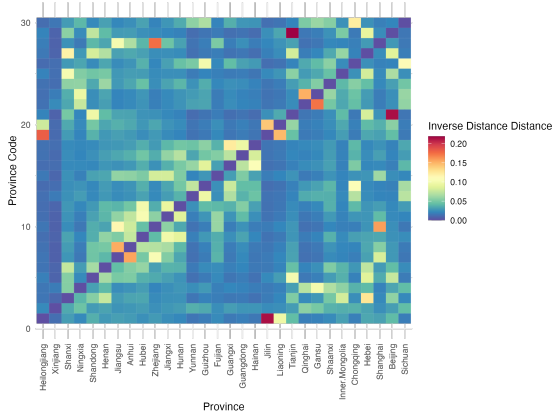
We build an economic distance weight matrix as W_3 based on GDP per capita. This matrix

captures the economic disparity and interactions between regions by considering the differences in their economic performance. Specifically, we calculate the weights using the inverse of the absolute differences in GDP per capita between regions. The rationale behind this approach is that regions with similar levels of economic development are likely to exhibit stronger economic interactions and dependencies.

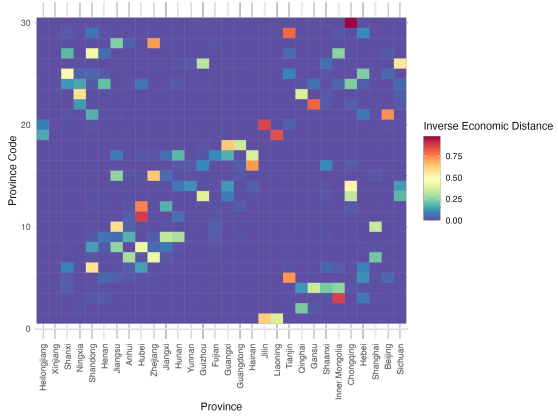
To construct W_2 , we first define the economic distance between any two regions i and j as $d_{ij} = |GDP_i - GDP_j|$, where GDP_i and GDP_j are the GDP per capita of regions i and j , respectively. We calculate the average GDP over the observation period to estimate the development level of each region to smooth out short-term fluctuations. By averaging GDP over multiple years, we reduce the impact of temporary economic shocks or anomalies, providing a clearer picture of a region's long-term economic trends and underlying development. The weights w_{ij} are then calculated as the inverse of these distances, normalized to ensure that the sum of the weights in each row of the matrix equals one. Mathematically, this can be represented as:

$$w_{ij} = \frac{1/d_{ij}}{\sum_j (1/d_{ij})}$$

The heat maps are provided for the weight matrices based on inverse geographic distance and inverse economic distance. Figures 1.7a and 1.7b reveal distinct patterns in the structure of spatial dependencies based on geographic and economic proximities.



(a) Inverse Distance - W2



(b) Economic Scale - W3

The distance heat map shows a prominent diagonal line, indicating that geographically closer regions have stronger weights. Clusters of higher weights near the diagonal reflect close geographic proximity among neighbouring regions. This structure suggests that geographic proximity significantly influences spatial dependence, as geographically close regions exhibit stronger interactions.

In contrast, the economic heat map displays a more dispersed pattern of higher weights, indicating that economic similarities are distributed less uniformly than geographic proximities. Higher weights are not concentrated solely near the diagonal, suggesting that regions with similar economic levels can be geographically distant. This structure implies that economic interactions and dependencies are not strictly confined to geographically neighbouring regions. Instead, regions with similar economic profiles, regardless of their geographic distance, exhibit stronger economic ties.

When comparing the two matrices, the first heat map shows a more concentrated pattern of weights along the diagonal, emphasising the role of physical proximity. In contrast, the second heat map presents a more dispersed pattern, highlighting economic similarities over

geographic proximity. Therefore, they provide a clear view of spatial interactions influenced by both geographic and economic factors.

1.4.2 Spatial Autocorrelation

Spatial autocorrelation refers to the statistical dependence between the values of a variable in geographic space and those of neighbouring regions or points. In other words, it describes the tendency for similar values to cluster in geographically adjacent regions (Tobler, 1970). This phenomenon reflects spatial aggregation or diffusion patterns and is often summarised by Tobler’s First Law of Geography: “Everything is related to everything else, but near things are more related than distant things” (Tobler, 1970).

Spatial autocorrelation can be categorised into two types: positive and negative. Positive spatial autocorrelation occurs when similar values cluster together in space, indicating that regions with high or low values of a variable are near each other. Negative spatial autocorrelation, on the other hand, occurs when similar values are distributed in neighbouring but different regions, suggesting a matrix pattern where high values are adjacent to low values, and vice versa.

To measure spatial autocorrelation, two commonly used statistics are Moran’s I and Geary’s C. Moran’s I measures spatial similarity, quantifying the degree to which similar values of a variable cluster together in space (Cliff and Ord, 1981; Anselin, 1988). Geary’s C, conversely, measures spatial dissimilarity, highlighting how different values of a variable are distributed across neighbouring regions.

These measures reveal underlying spatial patterns and trends that might not be apparent in traditional non-spatial analysis (Anselin, 1988; Bivand, 2008). We aim to identify these

patterns to better understand economic activities within spatial processes and improve the accuracy of our models by accounting for spatial dependence.

To assess the presence of spatial autocorrelation, we employed Global Moran's I statistic, which measures the degree of spatial clustering of similar values. A significant Moran's I indicates the presence of spatial autocorrelation, suggesting that the value of a variable in one location is influenced by the values in neighbouring locations. We also include the calculation of Global Moran's I in Appendix C.3.

We use an adjacency weight matrix as the basic weight matrix. Table 1.1 presents Moran's I-statistics for various transportation infrastructure variables across different years, including industry production, roads, railways, and inland waterways. The Moran's I values and their corresponding z-values, provided in parentheses, are calculated using a two-tailed test, indicating the presence and significance of spatial autocorrelation in these variables over time.

The Moran's I statistics for industry production consistently show positive and significant values across all years, ranging from 0.299 to 0.348. The z-values indicate strong statistical significance at the 1% level, suggesting that industry production is spatially autocorrelated. This indicates that regions with high industry production tend to cluster together, while regions with low production also form geographic clusters.

The Moran's I statistics for road infrastructure are notably higher than those for industrial production, ranging from 0.569 to 0.622, and are statistically significant at the 1% level. This result indicates a strong degree of spatial autocorrelation, suggesting that regions with dense road networks tend to be geographically clustered. The high Moran's I values reflect the uneven distribution of road infrastructure, which is concentrated in well-developed provinces,

Table 1.1: Spatial Autocorrelation test in Moran's I with W1

Year	Industry Production	Road Density	Rail Density	Inland Waterway
2001	0.348*** (4.204)	0.569*** (6.478)	0.458*** (5.766)	0.164*** (3.782)
2002	0.343*** (4.157)	0.573*** (6.528)	0.469*** (5.815)	0.247*** (3.279)
2003	0.332*** (4.079)	0.573*** (6.533)	0.469*** (5.784)	0.275*** (3.508)
2004	0.334*** (4.113)	0.570*** (6.633)	0.471*** (5.789)	0.297*** (4.058)
2005	0.336*** (4.131)	0.570*** (5.850)	0.477*** (6.650)	0.318*** (4.283)
2006	0.331*** (4.080)	0.618*** (6.877)	0.477*** (5.864)	0.285*** (3.915)
2007	0.325*** (4.015)	0.618*** (6.903)	0.472*** (5.812)	0.211*** (4.385)
2008	0.320*** (3.970)	0.622*** (6.954)	0.467*** (5.825)	0.279*** (4.143)
2009	0.331*** (4.057)	0.617*** (6.896)	0.476*** (5.934)	0.228*** (3.949)
2010	0.322*** (3.983)	0.612*** (6.831)	0.455*** (5.576)	0.246*** (4.003)
2011	0.328*** (4.039)	0.611*** (6.817)	0.438*** (5.416)	0.251*** (3.867)
2012	0.333*** (4.063)	0.614*** (6.856)	0.451*** (5.527)	0.242*** (3.969)
2013	0.326*** (3.999)	0.617*** (6.880)	0.454*** (5.553)	0.225*** (4.517)
2014	0.319*** (3.933)	0.610*** (6.815)	0.461*** (5.604)	0.189*** (4.587)
2015	0.308*** (3.819)	0.601*** (6.714)	0.475*** (5.714)	0.165*** (4.746)
2016	0.305*** (3.793)	0.600*** (6.701)	0.457*** (5.562)	0.208*** (4.475)
2017	0.303*** (3.802)	0.599*** (6.690)	0.453*** (5.512)	0.255*** (3.734)
2018	0.301*** (3.786)	0.594*** (6.627)	0.443*** (5.449)	0.161*** (3.092)
2019	0.299*** (3.745)	0.582*** (6.513)	0.471*** (5.708)	0.131*** (3.652)

Note: ***, **, and * show the significance at of 1%, 5% and 10%, respectively. Z-values are provided in parentheses. Moran's I-statistics are based on a two-tailed test.

particularly in the eastern and central regions.

The Moran's I values for railway infrastructure are also positive and statistically significant, ranging from 0.438 to 0.477. These results confirm that railway networks, while less dense than roads, also exhibit spatial clustering. This pattern may reflect coordinated regional development strategies and the historical legacy of rail infrastructure investment, particularly in provinces like Hebei and Henan.

Inland waterway infrastructure also shows positive and significant Moran's I values, though they are generally lower than those for road and railway infrastructure, ranging from 0.131 to 0.318. The z-values again confirm statistical significance at the 1% level for most years. The lower Moran's I values suggest that while there is spatial autocorrelation, the clustering is less pronounced compared to road and railway infrastructure.

Overall, the results indicate significant spatial autocorrelation in all four variables studied. Industry production shows strong spatial dependence, suggesting that economic activities are geographically clustered. The high Moran's I values for road and railway infrastructures highlight the concentrated nature of these developments. Inner waterway infrastructure, while still showing spatial dependence is less clustered, due to geographical constraints.

The following two tables, Table 1.2 and Table 1.3, present results of Moran's I test in two tails with Inverse distance weight matrix as W2 and economic weight matrix as W3 respectively. The tables show the spatial autocorrelation statistics for industry production, road, railway, and inner waterway infrastructure from 2001 to 2019, with z-values provided in parentheses to indicate the significance levels of the results.

Table 1.2 presents Moran's I statistics calculated using the inverse distance weight matrix (W2). For industry production, Moran's I values range from 0.026 to 0.048, with most

Table 1.2: Spatial autocorrelation tests in Moran's I with W2

Year	Industry Production	Road Density	Rail Density	Inland Waterway
2001	0.039** (2.165)	0.123*** (4.525)	0.158*** (6.034)	-0.020 (0.742)
2002	0.037** (2.107)	0.124*** (4.577)	0.161*** (6.049)	0.010 (1.388)
2003	0.037** (2.119)	0.127*** (4.638)	0.165*** (6.115)	0.040** (2.273)
2004	0.035** (2.061)	0.118*** (4.489)	0.165*** (6.103)	0.078*** (3.684)
2005	0.032** (1.971)	0.115*** (4.389)	0.161*** (5.973)	0.075*** (3.555)
2006	0.028* (1.864)	0.127*** (4.546)	0.161*** (5.976)	0.057*** (2.979)
2007	0.027* (1.836)	0.124*** (4.486)	0.158*** (5.920)	0.018** (2.472)
2008	0.025* (1.793)	0.125*** (4.526)	0.168*** (6.303)	0.029** (2.253)
2009	0.026* (1.801)	0.127*** (4.568)	0.172*** (6.409)	0.020** (2.163)
2010	0.031** (1.958)	0.124*** (4.484)	0.138*** (5.250)	0.024** (2.235)
2011	0.035** (2.082)	0.123*** (4.462)	0.122*** (4.804)	0.028** (2.248)
2012	0.038** (2.160)	0.125*** (4.527)	0.129*** (4.983)	0.022** (2.148)
2013	0.038** (2.139)	0.127*** (4.573)	0.135*** (5.149)	0.011** (2.104)
2014	0.039** (2.199)	0.122*** (4.443)	0.140*** (5.271)	0.042*** (4.131)
2015	0.038** (2.152)	0.116*** (4.267)	0.145*** (5.374)	0.021*** (3.436)
2016	0.042** (2.269)	0.116*** (4.266)	0.141*** (5.321)	0.012** (2.282)
2017	0.043** (2.345)	0.114*** (4.198)	0.140*** (5.274)	0.060*** (3.248)
2018	0.047** (2.464)	0.107*** (4.005)	0.145*** (5.487)	-0.007 (1.144)
2019	0.048** (2.491)	0.103*** (3.883)	0.153*** (5.674)	-0.025 (0.545)

Note: ***, **, and * show the significance at of 1%, 5% and 10%, respectively. Z-values are provided in parentheses. Moran's I statistics are based on a two-tailed test.

years exhibiting significant positive spatial autocorrelation at the 5% or 10% level. This suggests that regions with high or low levels of industry production tend to cluster, albeit with relatively modest spatial dependence.

Road infrastructure also exhibits significant positive spatial autocorrelation, with Moran's I values ranging between 0.103 and 0.127. These consistently significant results at the 1% level suggest that road networks tend to be spatially clustered, particularly in more developed regions.

For railway infrastructure, Moran's I values are even higher, ranging from 0.122 to 0.172, and remain consistently significant at the 1% level. This indicates a stronger degree of spatial clustering in rail infrastructure development, where adjacent regions often share well-connected railway systems, reflecting both coordinated planning and historical investments.

Inland waterway infrastructure exhibits greater variability in Moran's I values, with some years showing significant positive spatial autocorrelation while others do not.

Table 1.3 presents Moran's I statistics calculated using the economic weight matrix (W_3). For industry production, Moran's I values indicate no significant spatial autocorrelation at conventional levels. This means that industry production is not as strongly clustered based on economic proximity as it is on geographic proximity.

Rail infrastructure also exhibits significant positive spatial autocorrelation, with Moran's I values ranging from 0.457 to 0.581. These values are consistently significant at the 1% level, suggesting that regions with similar economic levels tend to develop similar levels of rail infrastructure. This pattern likely reflects long-term strategic planning and infrastructure investment tied to regional industrial bases.

Table 1.3: Spatial autocorrelation tests in Moran's I with W3

Year	Industry Production	Road Density	Rail Density	Inland Waterway
2001	0.150 (1.225)	0.567 *** (3.911)	0.647 *** (4.836)	-0.015 (0.231)
2002	0.147 (1.213)	0.577*** (3.981)	0.658*** (4.845)	0.082 (0.821)
2003	0.155 (1.276)	0.581*** (4.008)	0.660*** (4.833)	0.144 (1.227)
2004	0.151 (1.252)	0.532*** (3.766)	0.662*** (4.835)	0.267 ** (2.236)
2005	0.137 (1.160)	0.515 *** (3.659)	0.645*** (4.710)	0.237** (2.005)
2006	0.125 (1.079)	0.485*** (3.314)	0.644*** (4.711)	0.163 (1.466)
2007	0.12 (1.048)	0.489*** (3.355)	0.641*** (4.698)	0.073 (1.182)
2008	0.113 (1.002)	0.507*** (3.470)	0.684*** (5.061)	0.149 (1.471)
2009	0.112 (0.983)	0.535*** (3.647)	0.707*** (5.227)	0.061 (0.875)
2010	0.132 (1.129)	0.534 *** (3.634)	0.581*** (4.247)	0.068 (0.893)
2011	0.143 (1.198)	0.538*** (3.664)	0.516*** (3.820)	0.076 (0.912)
2012	0.152 (1.253)	0.543*** (3.695)	0.529*** (3.886)	0.063 (0.848)
2013	0.15 (1.239)	0.548*** (3.722)	0.559*** (4.091)	0.091 (1.336)
2014	0.155 (1.274)	0.534*** (3.637)	0.578*** (4.199)	0.33*** (4.609)
2015	0.148 (1.234)	0.512*** (3.499)	0.589*** (4.233)	0.205*** (3.530)
2016	0.16 (1.313)	0.510*** (3.480)	0.590*** (4.284)	0.061 (1.083)
2017	0.166 (1.365)	0.502 *** (3.425)	0.583*** (4.231)	0.247 ** (2.205)
2018	0.177 (1.445)	0.480 *** (3.285)	0.602*** (4.396)	0.029 (0.610)
2019	0.18 (1.461)	0.457*** (3.138)	0.626*** (4.515)	-0.01 (0.330)

Note: ***, **, and * show the significance at of 1%, 5% and 10%, respectively. Z-values are provided in parentheses. Moran's I statistics are based on a two-tailed test.

Road infrastructure, by contrast, shows even higher Moran's I values, ranging from 0.516 to 0.707, also significant at the 1% level. This indicates a stronger degree of spatial clustering based on economic similarity, suggesting that well-developed road networks are highly concentrated in economically advanced regions, reinforcing accessibility advantages and regional development disparities.

Inland waterway infrastructure shows more mixed results. While some years show significant positive spatial autocorrelation at the 1% or 5% level, other years do not show significant clustering. This variability suggests that economic factors play a role in the spatial distribution of inland waterway infrastructure, but the influence is less consistent compared to road and railway infrastructure.

Comparing the three weight matrices, it is evident that both geographic and economic proximities influence spatial autocorrelation, but the strength and consistency of these influences vary across infrastructure types. Road and railway infrastructure exhibit strong and consistent spatial autocorrelation with both weight matrices, highlighting the importance of geographic and economic factors in their distribution. Inland waterway infrastructure shows less consistent spatial autocorrelation.

This observation suggests that regions with more developed rail infrastructure tend to be geographically close, forming distinct spatial clusters. In other words, when a region has a well-developed rail network, its neighbouring regions are also likely to have similarly advanced railway systems. This pattern is further confirmed by the higher values of Moran's I, indicating that rail infrastructure development is unevenly distributed across regions and is concentrated in areas with denser and more extensive networks.

This phenomenon of concentration and spatial aggregation highlights the strong link between

rail infrastructure and long-distance regional corridors. Railways are typically designed to support freight and intercity connectivity, reinforcing spatial development along key economic axes. As a result, rail infrastructure often reflects strategic investment patterns, favouring regions with industrial capacity and geographic centrality. The connectivity created by railway systems contributes to spatial clustering by linking economically significant regions across provinces.

In contrast, road infrastructure—although also spatially clustered—tends to follow urban agglomeration patterns more closely, connecting neighbouring cities and supporting localised mobility.

In summary, the analysis of Moran’s I-statistics using three weight matrices reveals significant spatial autocorrelation in infrastructure development, with rail and road infrastructures showing strong clustering based on both geographic and economic proximities. Inland waterway infrastructure, however, exhibits greater variability, indicating that its spatial distribution is influenced by multiple factors. More broadly, unlike roads and railways, inland waterways depend heavily on local physical geography, making them more difficult to expand or strategically plan across wide regions.

1.4.3 Model Selection

Based on the results of the previous section, two key characteristics of the spatial relationships in infrastructure construction are evident. First, the results of Moran’s I statistic exhibit spatial autocorrelation, indicating that observations closer to each other in space tend to be more similar. Second, spatial variability in statistical properties necessitates the selection of models capable of handling such changes.

Multiple spatial autocorrelations can lead to more complex spatial dependencies between variables. Spatial relationships may exist among both dependent and independent variables. Therefore, in this section, we select an appropriate spatial model based on the spatial relationships between variables.

According to Elhorst (2014), there are three primary types of spatial models: the Spatial Error Model (SEM), the Spatial Lag Model (SLM), and the Spatial Durbin Model (SDM). Each of these models addresses different kinds of spatial effects in the data.

SEM, SLM, and SDM are based on different underlying spatial processes. The appropriate model should capture the nature of spatial interactions affecting the dependent variable. SEM accounts for spatial autocorrelation in the error terms of a regression model. This model is appropriate when the correlation between observations arises from unobserved influences or shocks that are spatially correlated, rather than from the spatial diffusion or influence of the dependent variable itself.

The SLM, also known as the spatial autoregressive model, incorporates spatial dependency directly into the dependent variable. This model is suitable when the value of the dependent variable at one location is influenced by its values in neighbouring locations.

Finally, the SDM extends the SLM by adding spatial lags of the independent variables to the equation, capturing both the spillover effects of the dependent variable and the influence of neighbouring independent variables.

We begin by exploring the potential spatial mechanisms present in our data. We follow Tong et al. (2013)'s model selection process: (1) clarify the spatial dependencies, and (2) test for possible spatial interactions in regression. We then perform the Wald test and likelihood ratio test to evaluate whether spatial regression provides better results than simple linear

regression.

In time series data, observations are ordered chronologically, whereas in spatial data, relationships between data points are based on geographical features rather than linear order. This spatial arrangement introduces the concept of spatial autocorrelation, where economic activity in one region can be influenced by neighbouring regions. As a result, the traditional OLS model, which assumes independence among observations, may not be suitable for analysing spatial data, as observations in neighbouring regions are often interdependent.

Therefore, we employ the Lagrange Multiplier (LM) test to detect the presence of spatial interactions. By analysing residuals using a spatial weight matrix, the LM test identifies whether spatial dependencies are present in the data. If the LM test reveals significant spatial dependencies, it indicates the need to use an appropriate spatial econometric model, such as a Spatial Lag Model or a Spatial Error Model, to properly account for these interdependencies and produce accurate estimates.

The LM tests are performed to detect the presence of spatial dependence in the model. The LM test for spatial lag assesses whether a spatially lagged dependent variable should be included in the model. Similarly, the LM test for spatial error evaluates whether the error terms exhibit spatial autocorrelation. The results, presented in Table 1.4, evaluate spatial dependencies using three different weight matrices. These tests help determine the necessity and appropriate type of spatial model, such as the Spatial Error Model, Spatial Lag Model, or Spatial Durbin Model, by identifying spatial dependencies.

We conduct robust Lagrange Multiplier tests for spatial error and spatial lag effects using three weight matrices: W_1 , W_2 , and W_3 . These tests aim to identify spatial dependency features in the model, distinguishing between spatial autocorrelation in the error terms and

spatial lag effects of the dependent variable.

For the W_1 weight matrix, the robust LM statistic for spatial error effects is 469.429, with a p-value of 0.000, indicating highly significant spatial autocorrelation in the error terms. The robust LM statistic for spatial lag effects is 36.947, also significant with a p-value of 0.000, suggesting that the dependent variable in one region is significantly influenced by neighbouring regions. While both effects are significant, the much higher statistic for error effects indicates that spatial dependency is primarily driven by spatial autocorrelation in the error terms.

For the W_2 weight matrix, the robust LM statistic for spatial error effects is 314.326, with a p-value of 0.000, indicating strong and highly significant spatial autocorrelation in the error terms. In contrast, the robust LM statistic for spatial lag effects is 0.211, with a p-value of 0.646, suggesting that spatial lag effects are not significant under this matrix.

Under the W_3 weight matrix, the robust LM statistic for spatial error effects is 26.714, with a p-value of 0.000, confirming significant spatial autocorrelation in the error terms, though weaker than in W_1 . The robust LM statistic for spatial lag effects is 5.231, with a p-value of 0.022, indicating significant but weaker lag effects compared to error effects. While the influence of neighbouring regions on the dependent variable is more limited, it is not negligible.

In summary, the test results across the three weight matrices consistently highlight the significant presence of spatial error effects, while the significance and strength of spatial lag effects vary. We employ later tests to further the model selections between SEM and SDM.

Table 1.5 presents significant findings from both likelihood ratio tests and Wald tests applied to spatial econometric models with different weight matrices. Likelihood ratio and Wald tests

Table 1.4: Lagrange Multiplier Test

Test (W1)	Statistics	df	p-value
Robust Lagrange Multiplier (spatial error)	469.429	1	0.000
Robust Lagrange Multiplier (spatial lag)	36.947	1	0.000
Test (W2)	Statistics	df	p-value
Robust Lagrange Multiplier (spatial error)	314.326	1	0.000
Robust Lagrange Multiplier (spatial lag)	0.211	1	0.646
Test (W3)	Statistics	df	p-value
Robust Lagrange Multiplier (spatial error)	26.714	1	0.000
Robust Lagrange Multiplier (spatial lag)	5.231	1	0.022

Notes: This table present Lagrange Multiplier tests for W1, W2, and W3. LM test hold H0: no spatial error autocorrelation or no spatial lag autocorrelation. H1: spatial error autocorrelation or spatial lag autocorrelation exists in regression.

are effective tools for selecting appropriate spatial models, ensuring spatial relationships are accurately captured and modeled.

Likelihood ratio tests were conducted to compare nested models. For each weight matrix, the likelihood ratio values indicate whether including spatial lag or spatial error terms significantly improves model fit. Wald tests were employed to assess the significance of individual parameters in the spatial econometric models. These tests confirm the relevance of spatial terms in the models.

Likelihood ratio tests are used to compare the goodness of fit between nested models. They are based on the principle of maximum likelihood estimation. By comparing the likelihood values of restricted (simpler) and unrestricted (more complex) models, likelihood ratio tests evaluate whether additional parameters significantly improve model fit. The test statistic is given by $LRT = -2(\ln L_{\text{restricted}} - \ln L_{\text{unrestricted}})$, which follows a chi-square distribution under the null hypothesis that the simpler model suffices (Greene, 2012).

The chi-squared statistics are significantly high across all models and assumptions, strongly

suggesting rejection of the null hypothesis (H0) that ordinary least squares regression suffices for estimating the model. Specifically, when considering individual (ind) effects nested within both models or time effects, the likelihood ratio (LR) chi-squared values significantly exceed typical critical values, indicating that spatial components cannot be ignored and a more complex spatial regression model is required.

Table 1.5: Likelihood-ratio Tests and Wald Tests

Test	Weight	Assumption	LR chi2(6)	Prob > chi2
Likelihood Ratio	W1	ind nested in both	155.96	< 0.0000
		time nested in both	1263.92	<0.0000
	W2	ind nested in both	118.82	< 0.0000
		time nested in both	1270.12	<0.0000
	W3	ind nested in both	184.74	< 0.0000
		time nested in both	1247.89	<0.0000
Wald Test	W1	Wx	97.85	< 0.0000
		Spatial.rho	95.92	< 0.0000
	W2	Wx	31.87	< 0.0000
		Spatial.rho	23.24	<0.0007
	W3	Wx	61.69	<0.0000
		Spatial.rho	36.54	< 0.0000

Note: This table presents the results from likelihood ratio tests and Wald tests using weight matrices. These tests contain H0: OLS regression is enough for the estimation. H1: OLS is not enough to explain the estimation.

Wald tests are used to assess the significance of individual coefficients within a model. The test statistic is computed as $W = (\hat{\theta} - \theta_0)'[\text{Var}(\hat{\theta})]^{-1}(\hat{\theta} - \theta_0)$, where $\hat{\theta}$ is the estimated parameter vector, θ_0 is the null hypothesis value, and $\text{Var}(\hat{\theta})$ is the covariance matrix of the estimators (Wooldridge, 2010). Under the null hypothesis, the Wald statistic follows a chi-square distribution.

The Wald test results further support the conclusions drawn from the likelihood ratio tests. The Wald statistics for ‘Wx,’ a spatial weight matrix, are significant. ‘Wx’ represents spatial lag variables, obtained by multiplying the original variable x with the spatial weight matrix

W, indicating the influence of neighbouring regions on the current region. ‘Spatial.rho’ is a spatial autoregressive parameter. Across all three weight matrices, it is significant, indicating that spatial effects are indeed present in the data. It measures the strength of spatial dependence in a spatial autoregression model, reflecting the presence and degree of spatial autocorrelation in the model and describing the linear dependence between values in one region and those in neighbouring regions.

In essence, both types of tests strongly suggest that spatial autocorrelation violates the independence assumption of OLS regression. Since OLS cannot account for spatial autocorrelation, its estimates may be biased and inconsistent, leading to inaccurate conclusions. Thus, the use of OLS to analyse spatial datasets is inappropriate. The high chi-squared values from the likelihood ratio tests imply that models incorporating spatial dependencies provide a significantly better fit to the data compared to non-spatial models. Similarly, the significant Wald test results further confirm the importance of including spatial effects in the model.

The significant results from these tests indicate the presence of spatial dependence in our data, suggesting that models incorporating spatial lag and spatial error components are more appropriate. By comparing the log-likelihoods of models with and without spatial components, the likelihood-ratio tests demonstrate that including spatial lag and spatial error terms significantly enhances the model’s explanatory power. The Wald test results consistently show that the spatial lag, spatial error, and spatially lagged explanatory variables are highly significant.

We conclude that a model incorporating both spatial lag and spatially lagged explanatory variables is most suitable. Ultimately, we selected the SDM model.

1.4.4 Model Building

In the previous section, we finalized the selection of a spatial model to address spatial dependencies in our analysis. We assume that the control variables originate from two main sources: government spending and foreign investment. Government spending is typically directed toward infrastructure development, public services, and social programs, creating a supportive environment for economic activities and fostering long-term growth. Foreign investment, on the other hand, often involves the inflow of capital directly into firms, contributing to increased production capacity, technology transfer, and enhanced productivity.

Additionally, we include the GDP growth rate as a control variable to account for broader macroeconomic conditions influencing industrial output beyond the direct effects of capital, labour, and technology. By controlling for GDP growth, we aim to isolate the specific impact of our variables of interest on regional economic performance.

Numerous studies have incorporated spatial effects into the basic production function by including a spatial lag to capture the influence of neighbouring regions. For example, research by Cohen and Paul (2004), Bo and Florio (2012), Tong et al. (2013), and Konno et al. (2021) demonstrates the importance of considering spatial interdependencies in economic modelling. The SDM extends the conventional framework by including spatial lags of both dependent and independent variables. This extension allows us to account for potential spillover effects, where economic activity in one region may influence production functions in neighbouring regions.

Following the methodologies outlined by LeSage and Pace (2009) and Elhorst (2014), we employ the Spatial Durbin Model to capture spatial dependencies in both the error terms and independent variables. This approach allows us to better understand how economic activities

and policy interventions in one region affect other regions, offering a more comprehensive view of regional dynamics. This model can be represented as:

$$Y_{it} = \alpha_i + \beta X_{it} + \rho \sum_{j=1}^N w_{ij} Y_{jt} + \theta \sum_{j=1}^N w_{ij} X_{jt} + u_i + \lambda_t + \varepsilon_{it} \quad (1.1)$$

The industrial production for provinces i at time t is the dependent variable Y_{it} in above equation. The explanatory variables are X_{it} , which include road density, rail density, inner river per boat, government expenditure, foreign direct investment as enterprises, labour, and GDP growth rate. α is the constant term. The term $\sum_{j=1}^N w_{ij} Y_{jt}$ and $\sum_{j=1}^N w_{ij} X_{jt}$ are spatial lags in the dependent variable and independent variables respectively, which represents the spatially weighted average value. The weight matrices, $\sum_{j=1}^N w_{ij}$, are from last section as W_1 , W_2 , and, W_3 . To combine with our industry production equation, the final equation is following:

$$\begin{aligned} Ind_{it} = & \beta_0 + \rho \sum_{j=1}^N w_{ij} Ind_{jt} + \beta_{RoadD} RoadD_{it} + \beta_{RailD} RailD_{it} + \beta_{IRB} IRB_{it} + \beta_C C_{it} \\ & + \theta_{RoadD} \sum_{j=1}^N w_{ij} RoadD_{jt} + \theta_{IRB} \sum_{j=1}^N w_{ij} IRB_{jt} + \theta_{RailD} \sum_{j=1}^N w_{ij} RailD_{jt} + \theta_C \sum_{j=1}^N w_{ij} C_{jt} + \varepsilon_{it} \end{aligned} \quad (1.2)$$

$$\varepsilon_{it} = u_i + \nu_{it} \quad (1.3)$$

The equation represents a SDM model with fixed effects. In our case, the dependent variable Ind_{it} represents the industry production for province i at time t . The model seeks to explain

the variation in Ind_{it} , both at the individual level and through their spatial interactions.

The term β_0 is the intercept, capturing the baseline level of industry output when all other explanatory variables are held constant. The spatial dependence of the industry output is represented by the term $\rho \sum_{j=1}^N w_{ij} Ind_{jt}$, where ρ is the spatial autoregressive parameter, and w_{ij} are the elements of the spatial weight matrix W , which defines the spatial structure between different entities. This term accounts for the influence of industry output in neighbouring locations on the industry output in location i . It is important to note that the diagonal elements of the weight matrix, w_{ii} , are set to 0 to ensure that the spatial dependence only considers interactions between different entities, excluding any self-interaction effects.

The model includes a set of independent variables, $RoadD_{it}$, $RailD_{it}$, IRB_{it} , and C_{it} , which represent the factors that affect industry output. We use C_{it} for all control variables. Each variable has an associated coefficient: β_{RoadD} , β_{RailD} , β_{IRB} , and β_C , which measure the direct impact of the respective variables on the industry output in location i . To recall the data description, $RoadD_{it}$ represents the density of the road infrastructure, while $RailD_{it}$ represents the density of rail networks, IRB_{it} could represent inland waterway per boats.

In addition to the direct effects, the model also captures the spatial lag of these variables. The terms $\theta_{RoadD} \sum_{j=1}^N w_{ij} RoadD_{jt}$, $\theta_{RailD} \sum_{j=1}^N w_{ij} Rail_{jt}$, $\theta_{IRB} \sum_{j=1}^N w_{ij} IRB_{jt}$, and $\theta_C \sum_{j=1}^N w_{ij} C_{jt}$ represent the spatial lag effects, where each θ coefficient measures how the values of the corresponding variable in neighbouring entities affect the industry output in location i . This allows the model to account for spillover effects, where changes in one region can influence outcomes in nearby regions.

Finally, the equation includes an error term ε_{it} , which captures all other unobserved influ-

ences on Ind_{it} . This term accounts for the random effects or shocks that are not explained by the included variables and their spatial lags. In the spatial Durbin model with fixed effects, the error term is typically redefined as $\varepsilon_{it} = u_i + \nu_{it}$. In this equation, u_i represents the individual-specific effect for the i -th unit, often referred to as the fixed effect. This captures unobserved heterogeneity that is constant over time but varies across individuals. On the other hand, ν_{it} is the random error term, which represents random disturbances that vary across both individuals and time, accounting for factors that are not captured by the fixed effects.

1.5 Results

As we mentioned in the previous section, the regression analysis in this study is based on standardized variables, meaning the regression coefficients represent the standardized effect—that is, the change in the dependent variable for each standard deviation increase in the explanatory variable, rather than a direct economic interpretation. To provide a more intuitive understanding of the results and facilitate comparison with existing literature, we are further calculating elasticity to quantify the impact of a 1% change in the explanatory variable on the dependent variable.

During the elasticity calculation process, we are using the mean and standard deviation of the unstandardised, or original, data to convert the standardized regression coefficients into elasticity values, ensuring economic interpretability. This calculation is being conducted using both national and regional data, with separate estimations for the East, Central, and West regions to capture overall trends and regional heterogeneity. We mainly compare significant variables, so the calculation of elasticity also focuses on significant variables.

1.5.1 Baseline Results

This section starts from the estimation of Equation 1.2. We record the spatial Durbin model result in Table 1.6. It presents the production factors of industry production across provinces, considering the spatial dependencies captured by different weight matrices: Adjacency- W_1 , Distance- W_2 , and Economic- W_3 .

Due to the existence of the spatial lag structure, the coefficients of main regression and spatial lag as $W*$ are not the final causal effects. We only make a simple analysis here, and later we will analyse the specific spillover effects through the spatial analysis proposed by LeSage and Pace (2009). We also provide the main regression part in Appendix A as Table A.3 for further check.

Table 1.6 presents the coefficients estimated under each spatial weight matrix. We observe that the spatial lag of the independent variables exhibits significant variations depending on the weight matrix applied. Overall, our findings are consistent with previous literature, indicating that railway infrastructure has a positive impact on industrial outcomes at the national level. However, road transportation shows certain negative effects.

In studies of spatial Durbin models, higher goodness of fit values indicate that the model is better able to explain the variation in the dependent variable, as industrial production. The R^2 value measures the model's goodness-of-fit, reflecting the proportion of total variation explained by the model. In our study, the R^2 values are similar for three weight matrices, which capture the spatial dependence between regions and the impact of neighbouring regions on industrial production.

This table presents the baseline results of a SDM model, applied with three different weight matrices: adjacency-based (W_1), distance-based (W_2), and economic-based (W_3). To re-

Table 1.6: Baseline Results of SDM with Different Matrices

Weight Matrix Spatial Lag of X (W*)	Adjacency-W1	Distance-W2	Economic-W3
W*RoadD	0.127*** (2.620)	0.147*** (2.880)	0.220*** (7.100)
W*RailD	-0.058 (-0.880)	-0.452*** (-3.330)	-0.024 (-0.580)
W*IRB	0.030* (1.910)	0.007 (0.210)	0.012 (1.190)
W*lnGOVE	-0.681*** (-9.580)	-0.801*** (-3.710)	0.083 (0.690)
W*lnFDI	0.502*** (7.230)	0.510*** (3.050)	-0.110 (-1.240)
W*lnLabour	0.043 (1.270)	0.102*** (2.850)	-0.017 (-0.480)
Observation	570	570	570
Spatial.rho	0.743*** (29.630)	0.781*** (22.560)	0.585*** (20.650)
R-squared	0.9235	0.9157	0.9217
Region FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes

Note: Z-statistics in parentheses. All the spatial matrices are row-standardized. ***, **, and * show the significance at of 1%, 5% and 10%, respectively.

call the pervious setting, the weight matrices represent different ways of capturing spatial relationships between regions. The adjacency matrix (W_1) defines spatial proximity based on shared borders or direct connections. The distance matrix (W_2) uses the inverse of distance to measure spatial interactions, while the economic matrix (W_3) incorporates economic characteristics such as trade ties or GDP similarities to define spatial relationships.

The coefficients for the spatial lag of explanatory variables, denoted as W^* , highlight significant regional spillover effects. For road density, represented by $W \cdot \text{RoadD}$, the results are consistently positive and significant across all matrices, indicating that improvements in road density in neighbouring regions positively influence economic outcomes in a given region.

Conversely, the spatial lag for rail density, $W \cdot \ln \text{RailD}$, shows varied effects: it is insignificant under W_1 and W_2 , but becomes significantly negative under W_3 . This suggests that in economically connected regions, higher railway density might be associated with potential congestion, network misallocation, or other spatial inefficiencies that lead to negative spillovers.

We compare our results with Tong et al. (2013), who also employed multiple weight matrices. Similar to their findings, our results show that road networks produce strong positive spatial externalities under adjacency weights, whereas railway infrastructure does not consistently generate spatial benefits across different matrix structures. We discuss the implications of these spillover effects in more detail in the subsequent section.

For other control variables, such as the spatial lag of government expenditure, as $W \cdot \ln \text{GOVE}$, the results are mixed. It shows a strong negative effect under W_1 and W_2 , while it is insignificant under W_3 . This could reflect inefficiencies or potential leakages in gov-

ernment expenditure in certain regions, depending on how spatial relationships are defined. Similarly, the spatial lag of foreign direct investment FDI, exhibits positive and significant spillovers under W_1 and W_2 , emphasizing the role of economic ties in fostering regional development. The spatial lag of labour, as $W \cdot \ln \text{Labour}$, shows limited significance, becoming positive only under W_2 .

The spatial autocorrelation coefficient ρ is highly significant across all weight matrices, with values of 0.743, 0.781, and 0.585, each significant at the 1% level. This confirms the presence of spatial dependence in the data and validates the use of a spatial econometric model. However, the explanatory power of the model, as reflected by R^2 , varies significantly across matrices. While W_2 and W_3 exhibit high R^2 values, W_1 shows the highest value of 0.9235, suggesting that adjacency-based weights have better performance in spatial analysis.

Overall, these results underline the importance of carefully selecting the spatial weight matrix, as it significantly influences the estimated spillover effects and model performance. For instance, while rail density consistently demonstrates positive regional spillovers, road density exhibits negative spillovers under the distance matrix. In contrast, the two geographic weight matrices provide a more accurate reflection of geographic proximity between regions, capturing spatial interactions that extend beyond administrative borders. Additionally, the spatial dependencies derived from the economic matrix are generally lower than those from the geographic matrices.

In spatial econometrics, effects are typically categorized into direct effects, indirect effects, and total effects, each representing different dimensions of spatial influence.

Direct effects refer to the impact of a variable within a given region or unit of observation. These effects measure how changes in an independent variable directly influence the depen-

Table 1.7: Spatial Effects Analysis

Weight Matrix Spatial Lag of X (W*)		Adjacency-W1	Distance-W2	Economic-W3
Road Density	Direct	0.041 (1.440)	0.078*** (3.340)	0.102*** (3.960)
	Indirect	0.390*** (4.110)	0.836*** (4.440)	0.495*** (8.780)
	Total	0.431*** (4.540)	0.914*** (4.690)	0.597*** (8.710)
Rail Density	Direct	-0.259*** (-7.450)	-0.268*** (-6.420)	-0.118*** (-3.160)
	Indirect	-0.629*** (-3.800)	-2.734*** (-3.410)	-0.166** (-2.120)
	Total	-0.888*** (-4.990)	-3.003*** (-3.620)	-0.284*** (-2.820)
Inland Waterway	Direct	0.005 (0.640)	-0.005 (-0.600)	-0.002 (-0.200)
	Indirect	0.087* (1.760)	0.030 (0.190)	0.021 (0.920)
	Total	0.092* (1.660)	0.024 (0.150)	0.019 (0.640)
lnGOVE	Direct	-0.152*** (-3.040)	-0.165** (-2.860)	0.110* (1.940)
	Indirect	-2.002*** (-6.690)	-3.840*** (-3.100)	0.263 (1.010)
	Total	-2.154*** (-6.410)	-4.005*** (-3.120)	0.373 (1.230)
lnFDI	Direct	-0.053 (-1.090)	0.021 (0.400)	-0.227*** (-4.090)
	Indirect	1.168*** (5.220)	2.177** (2.460)	-0.431** (-2.240)
	Total	1.115*** (4.510)	2.198** (2.400)	-0.658*** (-2.850)
lnLabour	Direct	-0.072*** (-2.610)	-0.065** (-2.150)	-0.078*** (-2.610)
	Indirect	-0.030 (-0.610)	0.201* (2.070)	-0.111*** (-2.900)
	Total	-0.102** (-2.190)	0.136 (1.400)	-0.189*** (-5.410)

Note: Z-statistics in parentheses. All the spatial matrices are row-standardized. ***, **, and * show the significance at of 1%, 5% and 10%, respectively.

dent variable in the same region, without considering the influence of neighbouring regions. According to LeSage and Pace (2009), direct effects indicate the localized impact of policy interventions or changes in explanatory variables within the local unit. These effects are analogous to traditional regression coefficients but adjusted to account for spatial dependencies. In spatial models, direct effects capture how changes in a variable influence outcomes within the same spatial unit.

Indirect effects, also known as spillover effects, represent the influence of a change in an independent variable in one spatial unit on the dependent variable in neighbouring or connected spatial units. These effects explain the interconnectedness of spatial units, where changes in one location can have repercussions in adjacent areas (Anselin, 1988).

Total effects are the sum of direct and indirect effects, representing the overall impact of an independent variable on the dependent variable across all spatial units. Total effects provide a comprehensive measure of a variable's influence, accounting for both its localized impact and spillover effects (LeSage and Pace, 2009). They offer a broader perspective on how changes in one region affect not only that region but also the surrounding spatial context.

Table 1.7 presents the specific spatial impacts. Each factor's direct, indirect, and total effects reveal distinct patterns in how they influence industrial production.

Road density consistently exhibits a positive and economically meaningful impact on regional industrial output, especially under the distance and economic weight matrices. The direct effect is particularly pronounced when spatial proximity is emphasized: under the distance weight matrix, the direct coefficient reaches 0.078, translating to an elasticity of 0.031. In practical terms, a 1% increase in road density is associated with a 0.031% rise in industrial output. For a region with an industrial base of 100 billion RMB, this implies a direct gain

of 31 million RMB—a non-trivial economic benefit.

More importantly, the spillover effect—measured as the indirect impact—is substantially larger, suggesting that road investments not only benefit the local region but also stimulate production in neighbouring areas. Under the distance-based spatial matrix, the indirect coefficient reaches 0.836, corresponding to an elasticity of 0.330. This means that a 1% increase in road density in a neighbouring region is associated with a 0.33% increase in local industrial output.

We compare the results from Ahlfeldt and Feddersen (2018) whose research on the German high-speed rail (HSR) connecting Cologne and Frankfurt. Their analysis collects data from 1992 to 2009 and combines Difference-in-Differences analysis with nonlinear least squares to measure. They reveal that counties with intermediate HSR stops experienced an average GDP increase of approximately 8.5% , indicating a significant causal impact. While their context is specific to HSR and Western Europe, it underlines the broader principle: rail infrastructure can reshape regional economic hierarchies.

Assuming a 5% improvement in road density, our model predicts a direct output gain of approximately 0.16%, with even larger effects when accounting for spatial spillovers. From a policy perspective, this suggests that targeted investments in road infrastructure, especially in strategic corridors or regional hubs, can yield amplified returns. Rather than allocating infrastructure budgets evenly, governments may achieve higher economic efficiency by prioritizing projects that enhance regional linkages—particularly those connecting economically complementary areas.

In contrast, rail density exhibits a consistently negative relationship with regional industrial development across all spatial weight matrices, revealing the complex and potentially adverse

effects of excessive or uncoordinated rail expansion. This finding aligns with Banerjee et al. (2020), who highlight that certain infrastructure projects—especially those disconnected from productive agglomerations—may lead to spatial inefficiencies.

Under the distance weight matrix, the direct effect coefficient for rail density is -0.268, corresponding to an elasticity of -0.004. In practical terms, this implies that a 1% increase in local rail density is associated with a 0.004% decline in industrial output—a small but negative marginal effect.

More importantly, the indirect effect is also negative, suggesting that such inefficiencies spill over into neighbouring areas. This supports the argument that poorly coordinated or excessive railway expansion may disperse labour and investment, weakening industrial concentration. Baum-Snow et al. (2017) provide related evidence from Chinese cities, where the expansion of ring roads and radial railroads led to population decentralisation and a decline in central urban productivity.

They find that the construction of ring roads and radial infrastructure resulted in a 20% relocation of central city populations to peripheral areas, triggering industrial decentralisation and spatial fragmentation. This process reflects how excess infrastructure can facilitate outward migration, reallocating labour away from urban cores and undermining scale economies.

From this perspective, our findings align with a broader concern: increased rail density, particularly without spatial coordination, may unintentionally weaken industrial output by diffusing critical inputs such as labour, demand concentration, and capital investment.

For inland waterway, there is a spillover effects under adjacent weight matrix to surrounding regions as a elasticity as 0.431. It does not present a significant effects at other weight

matrices.

In terms of control variables, government expenditure (GOVE) shows a significant total negative effect under the distance weight matrix. The magnitude of the indirect effect underscores the broader spatial implications of inefficient fiscal policies.

Foreign Direct Investment (FDI) offers a contrasting narrative. While the direct effect under the economic weight matrix is slightly negative at -0.227, the indirect effect under the distance weight matrix is highly positive. This indicates that the spatial benefits of FDI investments primarily materialize through spillovers in geographically connected regions. The total effect underscores the importance of spatial connectivity in maximizing FDI's regional growth potential.

Overall, a well-integrated multimodal transport system leverages the strengths of each mode, ensuring comprehensive connectivity and optimizing the flow of goods and resources. The integration of road, rail, and inner river transport facilitates the efficient movement of goods and resources, enhancing industrial productivity across different regions. Road density primarily boosts local and regional connectivity, rail density supports extensive regional industrial growth, and inner river transport, though limited in indirect impact, contributes to regional connectivity and spillover benefits, likely due to the transport of heavy goods and inputs.

To further check the robustness of the model, we employ additional variables and exclude specific years. We add two more control variables to capture provincial attributes: GDP growth rate and the log of population. In addition, we control for the impact of the 2008 financial crisis by introducing dummy variables for the year.

These variables help capture key economic and demographic factors that may influence re-

gional development and industrial output. The GDP growth rate is included because it reflects the overall economic performance and momentum of a province. By controlling for GDP growth, we account for the varying levels of economic expansion across regions, ensuring that our model captures the effect of transport infrastructure beyond general economic trends. Similarly, larger populations often correlate with higher demand for infrastructure, labour supply, and market size, all of which can directly affect regional economic activities and development. Including population as a variable helps us control for these demographic differences, providing a clearer picture of how other variables impact the region.

Additionally, we exclude the shock year, 2008, from the analysis to control for the impact of the global financial crisis. The 2008 financial crisis had a significant and potentially abnormal effect on economic performance worldwide, including in China. By excluding this year, we aim to reduce any bias or distortion that the crisis might introduce into our results, allowing us to better isolate the underlying relationships between variables.

The results of these robustness tests, including the addition of new control variables and the exclusion of 2008, are recorded in Appendix A.3. Overall, our baseline estimations remain robust, confirming the stability and reliability of our findings across different model specifications.

1.5.2 Regional Inequality and Heterogeneity

In Table 1.8, we present the regional analysis divided into three groups. The provinces are categorized into three regions: east, west, and central. As mentioned earlier, this classification also reflects different economic levels, with the east representing developed regions, the central representing developing provinces, and the west as most less developed regions.

Road density remains a significant driver of spatial dynamics, but its impact varies depending on the spatial weight matrix applied.

In the East, road density demonstrates robust positive spatial spillovers under the distance weight matrix. This suggests that road infrastructure facilitates long-distance connectivity, strengthening integration between coastal urban hubs. The comparatively weaker spillover under the economic weight matrix highlights the primacy of geographic proximity over pure economic similarity in shaping spatial road-based interactions in this region.

In the Central region, road density has minimal impact. Under the distance matrix, moderate spillover effects suggest that road connectivity may support limited regional interaction. The economic matrix shows no significant spillovers, possibly due to fragmented economic structures and weaker intercity ties within this region.

In the West, road infrastructure exhibits the strongest positive spillover effects across all matrices, with the highest coefficient under the distance weight matrix. This underscores the crucial role of road infrastructure in overcoming geographical remoteness in western provinces, promoting spatial integration and reducing development disparities across vast distances.

Table 1.8: Baseline Results for East, Central, and West Regions

Weight Matrix	East			Central			West		
	Adjacency-W1	Distance-W2	Economic-W3	Adjacency-W1	Distance-W2	Economic-W3	Adjacency-W1	Distance-W2	Economic-W3
Spatial Lag of X (W*)									
W*RoadD	0.259*** (6.850)	0.179*** (3.130)	0.072 (1.640)	0.038 (0.810)	0.031 (0.640)	0.044 (1.170)	0.250*** (2.960)	0.216*** (3.350)	0.147*** (3.870)
W*RailD	0.231 *** (6.850)	0.117 (0.980)	0.084 (1.430)	-0.054 (-0.720)	-0.014 (-0.170)	0.055 (1.030)	0.003 (0.040)	0.037 (0.340)	0.028 (0.590)
W*IRB	-0.015 (-1.600)	-0.015 (-0.700)	0.005 (0.480)	-0.018 (-0.470)	-0.033 (-0.730)	-0.003 (-0.100)	0.058* (1.690)	0.034 (0.880)	0.076** (2.540)
W*lnGOVE	1.819** (9.590)	2.189*** (6.940)	1.905*** (7.520)	1.477*** (3.490)	0.686 (1.450)	-0.273 (-0.570)	0.124 (1.290)	0.294 (1.330)	0.468*** (2.390)
W*lnFDI	-0.701*** (-5.280)	-1.188*** (-5.100)	-1.045*** (-6.340)	-0.936*** (-3.350)	-0.373 (-1.100)	0.101 (0.360)	-0.199* (-2.240)	-0.274* (-1.980)	-0.388*** (-2.930)
W*lnLabour	0.107*** (3.670)	0.121*** (3.000)	0.025 (0.620)	-0.149* (-2.060)	-0.198*** (-2.880)	-0.087* (-1.700)	-0.185*** (-2.750)	-0.201*** (-3.820)	-0.264*** (-3.450)
Observation	247	247	247	114	114	114	209	209	209
Spatial.rho	0.764*** (26.570)	0.647*** (12.340)	0.486*** (11.380)	0.785*** (21.130)	0.724*** (15.220)	0.759*** (19.100)	0.766*** (22.750)	0.777*** (17.610)	0.690*** (17.870)
R-squared	0.0407	0.9518	0.9247	0.9471	0.9530	0.9537	0.9535	0.9466	0.9263
Region FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: Z-statistics in parentheses. All the spatial matrices are row-standardized. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Rail density displays region-specific effects that shift under the revised matrix structures. In the East, rail density has significant positive spillovers under the distance weight matrix, indicating that rail infrastructure supports regional economic activity by reinforcing long-distance intercity linkages. However, the effects diminish under the economic weight matrix, suggesting that strong economic linkages alone are insufficient to sustain rail-induced spillovers in already well-connected eastern regions—possibly due to marginal returns on added capacity.

In the Central and Western regions, rail density does not exhibit significant spillovers under any matrix. These results imply that rail infrastructure in these areas may face barriers such as weaker economic clustering, insufficient complementary industrial capacity, or underutilised transport corridors. The absence of spillovers highlights the importance of spatial coordination and economic demand alignment in generating broader benefits from rail investments.

Waterway transportation shows minor but region-specific effects under the matrices. In the West, IRB exhibits small positive spillovers under the economic weight matrix and the border weight matrix. These findings suggest that waterways contribute to regional integration in western provinces, particularly for geographically dispersed areas with navigable rivers. In contrast, the East and Central regions show no significant impacts, indicating that waterways face competition from other transport modes or limited geographic coverage.

Government expenditure demonstrates substantial positive spillovers in the East under all matrix, reflecting the benefits of fiscal policies in regions with greater geographic connectivity. In the Central and West regions, GOVE effects are less pronounced, with significant impacts observed only under specific matrices.

Labour continues to display mixed effects. In the East, positive spillovers emerge under the geographic weight matrix, reinforcing the role of human capital in regional growth.

Foreign direct investment consistently shows negative spatial spillovers across all regions, particularly in the East under all weight matrix. This suggests that FDI may crowd out local investments or exacerbate regional disparities.

Road density exhibits strong spillover effects across all regions, emphasizing its role in regional accessibility and economic integration. In the East, direct effects are insignificant, but indirect effects are large and highly significant, indicating substantial benefits to neighbouring provinces through improved road connectivity. The total effect underlines the importance of road infrastructure in promoting regional development beyond local impacts, particularly by linking economically dynamic urban areas.

In the Central region, direct effects are significant and positive, highlighting the local importance of road density. Indirect effects are also significant, indicating moderate inter-regional accessibility benefits. The total effect suggests that road networks in the Central region contribute to both local industrial activity and broader regional integration.

In the West, indirect effects dominate, with total effects being the highest among all regions. This reflects the critical role of road infrastructure in reducing geographic isolation, supporting interprovincial economic ties, and enabling wider spatial spillovers in historically disconnected western provinces.

Therefore, we can conclude that road density supports industrial production and regional development, particularly through spillover effects that enhance economic connectivity between provinces.

Table 1.9: Regional Analysis with Adjacency-W1

Weight Matrix		East	Central	West
Road Density	Direct	0.025	0.107***	0.056
		(0.830)	(2.810)	(1.630)
	Indirect	0.485***	0.259**	0.716***
		(7.780)	(2.330)	(3.890)
	Total	0.511***	0.365***	0.772***
		(6.670)	(2.730)	(4.020)
Rail Density	Direct	0.081	-0.104*	-0.047
		(1.480)	(-1.710)	(-1.250)
	Indirect	0.514***	-0.296	-0.073
		(3.060)	(-1.300)	(-0.340)
	Total	0.595***	-0.400	-0.120
		(2.760)	(-1.450)	(-0.520)
Inland Waterway	Direct	-0.005	-0.092***	0.021
		(-0.440)	(-2.700)	(1.190)
	Indirect	-0.033	-0.194	0.185*
		(-1.180)	(-1.390)	(1.650)
	Total	-0.037	-0.286*	0.206
		(-1.000)	(-1.660)	(1.620)
lnGOVE	Direct	-1.856***	-1.032***	-0.032
		(-10.530)	(-2.700)	(-0.560)
	Indirect	0.688*	1.443	0.250
		(1.690)	(1.280)	(0.750)
	Total	-1.167**	0.411	0.217
		(-2.230)	(0.300)	(0.570)
lnFDI	Direct	0.915***	0.387	-0.023
		(7.240)	(1.430)	(-0.360)
	Indirect	0.006	-1.373	-0.541**
		(0.020)	(-1.570)	(-2.140)
	Total	0.921**	-0.986	-0.564*
		(2.270)	(-0.910)	(-1.960)
lnLabour	Direct	-0.146***	0.119**	0.084
		(-5.560)	(2.120)	(1.630)
	Indirect	-0.010	-0.130	-0.307***
		(-0.240)	(-1.260)	(-3.770)
	Total	-0.157***	-0.010	-0.223***
		(-2.970)	(-0.100)	(-3.710)

Note: Z-statistics in parentheses. All the spatial matrices are row-standardized. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Rail density, in contrast, exhibits mixed and region-specific effects. In the East, indirect effects are significant and positive, suggesting that rail infrastructure contributes to regional production spillovers, although direct effects remain insignificant, possibly due to already saturated transport corridors. The total effect reinforces the broader benefits of rail investment in facilitating interregional flows.

In the Central region, rail density shows a negative direct effect, indicating potential local inefficiencies, such as underutilized capacity or misaligned infrastructure. Indirect and total effects are insignificant, suggesting limited regional benefits from current railway configurations.

In the West, all effects are insignificant, implying that rail density has minimal influence on regional economic dynamics in this region, possibly due to weak industrial demand, sparse population, or lack of complementary infrastructure and economic clustering.

Waterway transportation has region-specific effects, with significant spillovers only in the West. In the East and Central regions, both direct and indirect effects are insignificant, highlighting the limited role of waterways in shaping regional dynamics in these areas. However, in the West, indirect effects are positive and marginally significant suggesting that waterways provide regional spillover benefits, particularly for provinces with navigable rivers.

In terms of control variables, they provide additional context for regional dynamics. Government Expenditure shows significant negative direct effects in the East and Central, suggesting inefficiencies or misallocation of fiscal resources. However, the indirect effects are positive but only marginally significant in the East, indicating limited regional spillover benefits. Foreign Direct Investment demonstrates significant positive direct effects in the East, reflecting the local benefits of FDI. However, the indirect effects are negative in the West

and marginally significant in the Central region, suggesting uneven spillover impacts.

In summary, transportation infrastructure exhibits varied impacts on industrial production and regional dynamics across the East, Central, and West provinces, with rail density emerging as the most significant driver of regional development through its robust spillover effects. Road density, however, has mixed results, with significant positive spillovers only in the East, indicating its limited role in broader regional development elsewhere. Waterway transportation demonstrates modest positive spillovers exclusively in the West, emphasizing its niche contribution to regional integration in areas with navigable rivers.

We also present the spillover result with the distance weight matrix in Table 1.10. It provides additional insights into the spatial spillover effects of transportation infrastructure on industrial production across East, Central, and West provinces.

Road density continues to exhibit significant direct, indirect, and total effects across all regions, with spillovers most pronounced under W_2 . Direct effects are positive and significant, smaller than under W_3 but larger than W_1 . Indirect effects are substantial and highly significant, surpassing those under W_3 and W_1 .

In the Central region, direct effects are slightly lower than under W_3 but higher than W_1 , showing consistent local benefits. Indirect effects are lower than under W_3 but still exceed those under W_1 .

For the Western region, direct effects are significant and consistent with W_3 . Indirect effects are significantly higher under W_2 , almost doubling the spillovers seen under W_3 and W_1 . These results reinforce the strong spatial spillover role of road infrastructure, especially in under-connected areas.

Table 1.10: Regional Analysis with Distance-W2

Weight Matrix		East	Central	West
Road Density	Direct	0.087*** (3.000)	0.138*** (3.440)	0.138*** (3.610)
	Indirect	0.561*** (4.730)	0.307** (2.350)	1.072*** (3.700)
	Total	0.648*** (5.040)	0.445*** (2.880)	1.211*** (3.790)
Rail Density	Direct	-0.019 (-0.350)	-0.176*** (-2.660)	0.007 (0.120)
	Indirect	0.227 (0.680)	-0.346 (-1.280)	0.164 (0.340)
	Total	0.208 (0.560)	-0.523 (-1.620)	0.171 (0.320)
Inland Waterway	Direct	-0.008 (-0.760)	-0.093*** (-2.530)	0.026 (1.320)
	Indirect	-0.039 (-0.660)	-0.223 (-1.370)	0.192 (1.100)
	Total	-0.047 (-0.700)	-0.316 (-1.590)	0.218 (1.140)
lnGOVE	Direct	-2.138*** (-11.620)	-1.221*** (-3.240)	0.087 (0.780)
	Indirect	1.870** (2.230)	-0.517 (-0.340)	1.242 (1.140)
	Total	-0.268 (-0.290)	-1.738 (-0.990)	1.328 (1.110)
lnFDI	Direct	1.140*** (9.570)	0.614** (2.140)	-0.053 (-0.600)
	Indirect	-1.041 (-1.600)	0.200 (0.170)	-1.075 (-1.590)
	Total	0.098 (0.140)	0.814 (0.590)	-1.128 (-1.500)
lnLabour	Direct	-0.131*** (-4.230)	0.188*** (3.070)	0.148*** (2.950)
	Indirect	0.086 (1.100)	-0.151 (-1.420)	-0.287*** (-3.400)
	Total	-0.046 (-0.570)	0.037 (0.330)	-0.138 (-1.590)

Note: Z-statistics in parentheses. All the spatial matrices are row-standardized. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Rail density, by contrast, shows weaker and more variable effects under W_2 compared to W_3 and W_1 . In the East, direct effects are insignificant, similar to W_3 but contrasting with W_1 , where indirect spillovers (0.514) were significant.

In the Central region, direct effects are significantly negative, consistent across W_2 and W_1 . Indirect and total effects remain negative and insignificant, reinforcing the view that rail infrastructure faces inefficiencies in this region.

In the Western region, all effects remain insignificant, consistent with findings under W_1 and W_2 , suggesting minimal contributions of rail density to economic spillovers in the West.

Waterway transportation has limited but region-specific effects under W_2 , emphasizing its role in the West. All effects are insignificant, consistent with W_1 and W_3 , confirming that waterways are not a critical factor in shaping economic outcomes in the East. The other results are consistent with W_1 and W_3 .

Across all regions and matrices, road density consistently shows a positive impact on industrial production, indicating that improvements in road infrastructure tend to increase output levels. Notably, the direct effects are significant in all cases, underscoring the local importance of road connectivity in facilitating industrial activity. Indirect effects are also significant in most scenarios, suggesting robust spatial spillover benefits to neighbouring regions. For example, under the distance matrix W_2 , the total effect is highest in the Western region, reflecting a strong combined local and spillover impact of road density on provinces with similar economic levels. This pattern highlights the crucial role of road networks in bridging spatial and economic gaps, particularly in less-developed areas.

The economic spatial dependency as Table 1.11, which is based on economic linkages rather than purely geographic factors. Compared to the adjacency W_1 and distance W_2 weight

Table 1.11: Regional Analysis with Economic-W3

Weight Matrix		East	Central	West
Road Density	Direct	0.179*** (5.160)	0.157*** (4.060)	0.129*** (3.460)
	Indirect	0.219*** (3.770)	0.359*** (3.310)	0.511*** (4.880)
	Total	0.398*** (5.260)	0.517*** (3.840)	0.640*** (4.940)
Rail Density	Direct	0.003 (0.040)	-0.171*** (-2.720)	-0.048 (-1.060)
	Indirect	0.113 (1.100)	-0.175 (-0.880)	-0.007 (-0.060)
	Total	0.115 (0.740)	-0.347 (-1.370)	-0.055 (-0.350)
Inland Waterway	Direct	-0.009 (-0.740)	-0.083*** (-2.800)	0.051** (2.490)
	Indirect	0.002 (0.120)	-0.138 (-1.370)	0.246*** (2.780)
	Total	-0.006 (-0.230)	-0.222* (-1.710)	0.297*** (2.790)
lnGOVE	Direct	-1.786*** (-8.330)	-0.673 (-1.630)	0.192 (1.660)
	Indirect	1.413*** (3.840)	-1.790 (-1.460)	1.304** (2.040)
	Total	-0.373 (-0.770)	-2.463 (-1.690)	1.496** (2.000)
lnFDI	Direct	0.882*** (5.930)	0.485 (1.640)	-0.260** (-2.400)
	Indirect	-0.839** (-3.140)	1.074 (1.160)	-1.233*** (-2.700)
	Total	0.043 (0.120)	1.559 (1.350)	-1.493*** (-2.700)
lnLabour	Direct	-0.138*** (-4.220)	0.091* (1.950)	0.108 (1.530)
	Indirect	-0.058 (-1.300)	-0.045 (-0.520)	-0.416*** (-5.920)
	Total	-0.196*** (-3.890)	0.046 (0.460)	-0.307*** (-5.370)

Note: Z-statistics in parentheses. All the spatial matrices are row-standardized. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

matrices, W_3 emphasizes inter-regional economic ties, revealing differences in the magnitude and direction of effects.

Road density maintains its importance across all regions under the economic distance weight matrix W_3 , with both local and spillover effects contributing to industrial production. In the East, direct effects are significant and positive (0.179), with an elasticity of 0.070, unlike the adjacency matrix W_1 , where direct effects were insignificant. This indicates that road infrastructure directly supports local industrial production when economic similarity is considered rather than pure geographic adjacency. Indirect effects, at 0.219 (elasticity 0.086), remain significant but are smaller than under W_1 , suggesting that while spillovers to economically similar neighbouring provinces persist, they may weaken with geographic distance. This result aligns with Donaldson (2018), who found that improvements in road infrastructure can generate positive impacts on regional industrial output.

In the Central region, direct effects are significant and positive, showing a stronger local role for road density under W_3 compared to W_1 . Indirect effects are also larger than in W_1 , indicating that economic proximity enhances spatial spillovers in this context. In the West, road density continues to exhibit the strongest spillover effects, with indirect effects dominating, highlighting its importance in facilitating economic integration across distant but economically similar provinces.

Rail density, in contrast, exhibits weaker and more negative impacts under W_3 . In the East, both direct and indirect effects are insignificant, contrasting with the positive spillovers observed under W_1 . In the Central region, direct effects are significantly negative, consistent with W_1 , and indirect and total effects remain insignificant, suggesting limited regional benefits and persistent inefficiencies in the spatial deployment of rail infrastructure. In the

West, all effects remain insignificant, reinforcing the view that rail density currently plays a minimal role in shaping economic outcomes in this region.

Waterway transportation exhibits more pronounced and region-specific effects under W_3 , particularly in the West. In the East, all effects remain insignificant, consistent with W_1 , indicating the limited role of waterways in this region. In the Central region, direct effects are significantly negative, highlighting local inefficiencies. Indirect and total effects are insignificant, reflecting limited regional spillovers. In the West, waterway transportation demonstrates significant positive spillovers, with indirect effects and total effects contributing to regional integration and industrial production.

Overall, the results for road density consistently demonstrate the strongest and most widespread benefits across all regions, particularly under the economic weight matrix W_3 . Under W_3 , road infrastructure excels in fostering regional integration through strong economic spillovers, especially in economically linked regions such as the West, where its effects are most pronounced. In contrast, adjacency-based spillovers under W_1 emphasize the importance of local accessibility, while geographic distance under W_2 highlights road density's dual role in supporting both local and regional industrial production.

Rail density, by contrast, exhibits more localized and limited effects, with benefits diminishing significantly as spatial distance or economic linkages are considered. Under W_1 , rail networks show positive effects in geographically proximate regions like the East, where they enhance short-range industrial connectivity. However, these benefits are largely absent under W_2 and W_3 , where negative or insignificant effects dominate, particularly in the Central region, reflecting possible inefficiencies or misalignment between rail infrastructure and industrial demand.

Waterway transportation, while less impactful than rail or road infrastructure, shows a niche role in addressing geographic isolation, particularly in the West. Spillover effects are most evident under W_1 and W_3 , where waterways support regional integration in provinces with navigable rivers.

Our study demonstrates that different economic regions experience varying levels of spillover effects, similar with the findings of Yu et al. (2013), which indicate that positive spatial spillover benefits are weaker in economically less developed regions, particularly in the western provinces.

Our research shows that when building a multi-mode transportation system, priority construction options need to be considered based on local conditions. The differences in effects across matrices and regions suggest that policymakers should tailor infrastructure investments to the specific spatial dynamics and economic contexts of the regions they aim to develop.

1.6 Conclusion

Our study makes several key contributions to the literature on spatial economics and transportation infrastructure.

First, we contribute to the understanding of the heterogeneous regional impacts of transportation infrastructure by providing further empirical evidence supporting prior findings, such as those in Yu et al. (2013), Faber (2014), Qin (2016), and Banerjee et al. (2020). On one hand, our results confirm that certain types of infrastructure, particularly rail networks, can produce negative or negligible effects in some regions, consistent with the concerns

raised by Qin (2016) and Baum-Snow et al. (2017) . On the other hand, we reaffirm findings from Yu et al. (2013), emphasizing that local economic development levels are critical in determining the direction and strength of these effects.

Our analysis reveals substantial regional variation in how different transportation modes—roads, railways, and inland waterways—influence industrial production. These effects are not uniform, and they vary significantly depending on both transport mode and the spatial weight matrix employed, highlighting the need for region-specific multimodal transportation strategies.

Second, we explore the spatial spillover effects of transportation infrastructure using three distinct spatial weight matrices: adjacency W_1 , geographic distance W_2 , and economic connectivity W_3 .

Our baseline results confirm that both road and rail density exert statistically significant direct and indirect effects on regional industrial output. However, the magnitude and significance of these effects are highly sensitive to the spatial matrix applied. Compared to the 0–1 adjacency matrix, the distance and economic matrices yield a more accurate reflection of spatial dynamics, as they better capture functional and economic relationships beyond administrative boundaries.

The distance matrix W_2 accounts for true geographic proximity and enables the detection of spillover effects across physically close regions, regardless of jurisdiction. The economic matrix W_3 , by contrast, identifies linkages through industrial interdependence, trade flows, and economic similarity, making it particularly useful in capturing non-contiguous but economically tied regions. Together, these spatial matrices allow us to reveal stronger, more meaningful spatial interactions than those captured by adjacency alone.

Finally, this study bridges spatial econometric modelling with policy implementation by demonstrating how transportation infrastructure effects vary across space and by infrastructure type. Policymakers must consider not just whether to invest in infrastructure, but what type of infrastructure, where, and for whom.

For instance, where road networks show strong spillover effects—especially in less-developed but geographically central regions—investments should prioritize enhancing road connectivity. In areas where rail infrastructure yields limited or even negative returns, complementary policies such as logistics integration or demand-side restructuring may be needed. Meanwhile, inland waterway infrastructure, due to its context-dependent impacts, requires careful evaluation tailored to regional geography, industrial mix, and trade accessibility.

In conclusion, infrastructure policy should shift from “one-size-fits-all” to “region-specific and mode-appropriate.” By aligning transport investments with spatial dynamics, regional development goals, and infrastructure spillover patterns, governments can foster more efficient, equitable, and sustainable economic growth. This spatially targeted approach not only improves local industrial capacity but also strengthens regional linkages, laying the foundation for balanced and resilient development.

Chapter 2

Spatial Dynamics of Freight Efficiency: Use of Synthetic Instruments to Overcome Endogeneity

Abstract

This paper explores the application of eigenvector-based spatial filters to construct synthetic instruments within a dynamic spatial Durbin model. By employing an exogenous geographic matrix, the technique strategically avoids weak instruments through a controlled selection process. The validity of these instruments is then confirmed using the generalised method of moments (GMM). Additionally, we conduct empirical regional analyses to gauge the spatial effects of these instruments. The results indicate that synthetic instruments significantly capture spatial effects in the model, although their performance is contingent upon the rigour of the selection process and the appropriate threshold of instrument quantity. The empirical results indicate that freight efficiency clearly impacts spatial dynamics. In other words, when rail and road freight productivity improves in one region, it positively influences industrial production in neighbouring areas.

Keyword: Spatial Econometrics, Heterogeneity, Synthetic Instruments, Two-step GMM, Dynamic Spatial Durbin Model.

Presented at 18th Spatial Econometrics World Conference, University of Groningen, the Netherlands, July 2024.

2.1 Introduction

Road freight and rail transport are indispensable to the functioning of modern industry and regional economies, playing a vital role in linking supply and demand. These transportation modes occupy a unique socioeconomic position, as they facilitate the movement of goods essential for the continuity of industrial operations and economic activities. The core purpose of transportation is to connect markets and production hubs.

Worldwide, logistics expenditures represent approximately 10%–15% of global GDP (Rodrigue, 2024). For the transport of commodities, logistics costs typically range between 20% and 50% of their total costs (Rodrigue, 2024). Efficient freight connectivity serves as an essential driver of regional GDP growth and employment, as it reduces transportation costs, enhances supply chain efficiency, and supports economic activities across sectors.

Authorities also recognise that efficient transportation infrastructure is a key to ensure fast and reliable cargo movement. A dependable transportation system minimises delays and reduces transportation costs, both of which are crucial for enhancing industrial productivity and maintaining competitiveness in the market.

However, the health of transportation systems does not solely depend on the construction and investment in large-scale transportation infrastructure. While these are undoubtedly important, broader factors such as freight services contribute significantly to the efficiency of delivering resources and goods.

A crucial yet often overlooked aspect of transportation efficiency is the range of supporting services that underpin transportation infrastructure. These services include logistics management, storage solutions, and the dedicated personnel involved in freight operations.

Service personnel—such as drivers, logistics planners, warehouse workers, and maintenance staff—form the backbone of the transportation sector. For instance, the express delivery business in China experienced a significant revenue increase from 2018 to 2019, reaching 749.782 billion yuan, an increase of 24.2%.¹ This growth addresses the critical role of these supporting services and personnel in driving economic development.

Nevertheless, analysing the economic impact of these transport factors presents a core challenge in transport research: the endogeneity of transport. Freight transport is considered endogenous because it is influenced by interactions within the economic system. Logistics efficiency is shaped not only by market structure and demand dynamics but also by the level of economic activity. For instance, as manufacturing output increases, firms' demand for freight services rises. Efficient freight services can reduce transportation costs and shorten delivery times, thereby improving corporate productivity. Additionally, logistics efficiency is often shaped by regional policies, such as infrastructure development, traffic regulations, and government investments, further complicating the issue of endogeneity in transport factors.

To address these issues, this study focuses on two main themes: first, how to effectively address the endogeneity of freight transport efficiency; and second, assessing the direct and indirect impacts of freight efficiency on regional industrial production. By quantifying the contribution of freight efficiency to industrial development, we aim to better understand the role of the logistic system in driving regional industries.

In pursuit of the first goal, we build on geographical research employing the eigenvector-based spatial filtering approach, introduced by Griffith (2003). The spatial filtering approach uses eigenvectors derived from a transformed connectivity matrix to construct filtered synthetic

¹Statistical Bulletin on Transport Sector Development 2019, Ministry of Transport, Retrieved from: https://www.gov.cn/xinwen/2020-05/12/content_5510817.htm

instruments that remove spatial patterns from the model residuals. This method allows for the use of exogenous instruments from the geographic weight matrix while relying on the endogenous variables to select the relevant eigenvectors.

We also pose several research questions to guide our inquiry. First, how effective are synthetic instruments—specifically those derived from geographic weight matrices—in mitigating the endogeneity of transportation variables within spatial models? This question leads to a deeper examination of whether the eigenvector-based spatial filtering approach can effectively isolate and neutralise endogenous correlations, thereby eliminating biases from the results. Another critical area of inquiry concerns the robustness of these methodologies: considering the interaction between geographic impacts and temporal changes in regional economic outcomes, how do variations in spatial thresholds affect the predictive power and accuracy of dynamic models?

Regarding the second objective, we test the instrumental variables we constructed using empirical data from 30 provinces in China from 2001 to 2019. Our analysis includes an evaluation of the robustness of these instruments in reflecting the interconnectedness of regional economies. Furthermore, we discuss the ability of the chosen instrumental variables to capture spatial relationships.

Our results show that the selection of synthetic instrumental variables must carefully account for their spatial characteristics. Instruments with higher spatial coefficients are more effective because they not only align with the spatial tendencies of the endogenous variables but also capture a broader range of spillover effects within a dynamic spatial Durbin model. These variables maintain consistency in terms of spatial dependence, ensuring that the spatial relationships observed in the endogenous variables are accurately reflected in the

instruments.

The empirical findings also reveal that selecting the most spatially efficient synthetic instrumental variable enables us to better capture the dynamic impact of freight efficiency on regional industrial production. In the short term, increased freight efficiency can lead to resources and economic activity flowing to surrounding regions, potentially resulting in a temporary decline in local production value as firms and industries benefit from improved access to neighbouring markets. However, over the long term, this redistribution of resources and enhanced connectivity foster greater regional integration and efficiency, ultimately exerting a positive impact on local industrial production.

As this chapter addresses both spatial econometrics and empirical research findings, it is extensive in length. The remainder of the paper is organised as follows: Section 2.2 provides a literature review. Section 2.3 presents a description of the empirical data. Section 2.4 outlines our methodology, including an analysis of endogeneity. Section 2.5 focuses on the construction and selection of eigenvectors within the econometric analysis. Section 2.6 illustrates our empirical findings. Section 2.7 offers a residual analysis. Finally, Section 2.8 concludes the paper with a summary of the findings and their implications.

2.2 Literature Review

This section addresses two key aspects of the literature: (1) Freight Productivity in Regional Production, and (2) Synthetic Instrument Construction.

2.2.1 Freight Productivity in Regional Production

In many major economic studies, freight efficiency is typically considered part of transport costs (Behrens and Picard, 2011; Donaldson and Hornbeck, 2016; Donaldson, 2018; Braccaccio et al., 2020). However, there remains a significant gap in research directly examining the impact of freight productivity on regional economies. This paper contributes to the literature by exploring the role of freight service productivity in regional economic development. Specifically, we investigate the mechanisms through which freight services influence regional growth and provide an overview of existing studies addressing the endogeneity of freight services.

Freight distribution is receiving increased attention in recent research. Early studies emphasised that productivity gains in manufacturing increasingly originate from improvements at terminals rather than from the efficiency of transportation modes themselves (Rodrigue, 1999). The efficiency and effectiveness of transportation systems are intrinsically tied to freight labour productivity, as transportation forms the backbone of logistics operations. When transportation systems are well supported, freight labour can operate at optimal levels, ensuring timely deliveries while reducing the time and resources required for each shipment.

Transport cost theory posits that lower transportation costs—achieved through high freight labour productivity—lead to reduced production costs for industries. This cost reduction enhances the competitiveness of industries both domestically and internationally (Blair and Carroll, 2009). Furthermore, effective supply chain management depends heavily on the productivity of the freight labour sector. High productivity facilitates seamless coordination between various stages of production, from raw material procurement to final product

delivery, thereby improving overall supply chain efficiency (Hesse and Rodrigue, 2004).

However, freight services, like other transport factors, present endogenous challenges that conflict with the standard regression assumption of exogeneity. One primary challenge is that freight services are intricately linked to infrastructure development. Regions with superior transport networks tend to exhibit higher freight labour productivity, which enhances the competitiveness and productivity of local industries, contributing to regional economic growth and development (Rodrigue, 2024). Similarly, the structure of the regional economy, including market and industrial composition, also has a significant impact on freight services. For instance, market and industrial structures directly influence the product range and timeliness of freight transport, with regions favouring manufacturing industries relying more heavily on freight transport efficiencies (Kent, 2010).

To prevent endogeneity from compromising the validity of our model, we now turn to the approaches discussed in the existing literature. Various studies have employed a range of techniques to address endogeneity, such as the use of instrumental variables to isolate the exogenous variation in independent variables. Additionally, methods such as the two-stage least squares (2SLS) approach, the generalised method of moments (GMM), and control function approaches are commonly used to correct for endogeneity. These methods introduce external instruments or controls that are correlated with the endogenous regressors but uncorrelated with the error term, thereby improving the robustness of the estimation process.

2.2.2 Synthetic Instrument Construction

To address the process of constructing synthetic instruments, we have organised this section into three subsections. First, we examine endogeneity in spatial econometrics, focusing on

the inherent endogeneity issues that arise in this field. Next, we explain the methodology for constructing instruments using a supervised spatial filter threshold. This technique involves filtering out spatial noise and isolating significant spatial patterns that influence the dependent variable, thereby improving the precision of the instruments. Finally, we evaluate the effectiveness of the constructed instruments, including an analysis of their performance in addressing endogeneity and capturing the true spatial relationships.

Endogeneity in Spatial Econometrics

We now discuss how spatial dependencies and interactions within the data can bias traditional econometric models. Spatial models examine the relationships between locations in close proximity to one another. To capture spatial interaction, spatial regressions incorporate three types of effects: spatial endogenous interaction effects among the dependent variable, exogenous interaction effects among the independent variable, and interaction effects on the error term (Elhorst, 2014). These effects facilitate the analysis of spatial linkages among regions but also introduce endogeneity issues in spatial models.

Endogeneity in spatial models becomes particularly problematic due to spatial spillovers, where behaviour in one location influences and is influenced by neighbouring locations. This interdependence means that economic or social changes in one area can propagate to adjacent areas, creating complex relationships that are challenging to untangle. For example, an increase in property values in one neighbourhood can lead to rising property values in surrounding neighbourhoods. Similarly, a policy change in one city, such as the introduction of a new tax or subsidy, can affect economic outcomes in neighbouring cities through mechanisms like increased trade, migration, or commuting.

These spatial spillovers introduce endogeneity because the observed outcomes in one loca-

tion are not independent of outcomes in neighbouring locations (Elhorst, 2014). This interconnectedness violates the standard econometric assumption that explanatory variables are uncorrelated with the error term, leading to biased and inconsistent estimates if not properly addressed.

Omitted variable bias is another critical issue in spatial models. If significant spatially varying factors—such as environmental conditions, infrastructure quality, or local policies—are excluded from the model, their effects may spill over into neighbouring regions. This omission can result in overestimation or underestimation of the true spatial effects, further compromising the validity of the estimates.

Measurement errors also present challenges in spatial models. Inaccuracies in data collection, such as imprecise geographic boundaries or incorrect population counts, can introduce spatially correlated errors. These errors can bias the estimation of spatial effects, leading to incorrect conclusions about the relationships between variables.

Simultaneity in spatial models arises when the dependent variable and one or more explanatory variables influence each other simultaneously. For instance, economic growth in a region may attract new businesses, which in turn stimulates further economic growth, creating a bidirectional relationship. This simultaneity complicates the establishment of causality and can result in biased estimates if not adequately accounted for.

Addressing endogeneity in spatial models is essential for obtaining unbiased and consistent estimates of spatial effects and parameters. Techniques such as instrumental variable (IV) methods, spatial lag models, and spatial error models are commonly employed to mitigate endogeneity. By correctly specifying the model and adopting appropriate estimation techniques, researchers can draw valid and reliable conclusions about the spatial relationships

and effects in their study.

Numerous spatial econometricians have developed methods to address endogeneity in spatial models, aiming to ensure that spatial relationships and effects are accurately estimated while avoiding biases introduced by endogeneity.

One notable contribution is the development of methods for spatial models that include endogenous variables among the regressors. For instance, Fingleton and Gallo (2008) introduced a generalised spatial two-stage least squares estimator for spatial autoregressive models. This approach employs spatially lagged exogenous variables as instruments for the endogenous variables, thereby mitigating the endogeneity problem and producing more reliable estimates.

Building on this work, Gallo and Páez (2013) integrated the geographic spatial filter developed by Griffith (2003) to construct eigenvectors for endogenous variables in cross-sectional spatial models. This approach helps to extract the spatially correlated component of the endogenous variable, thereby reducing measurement error bias. The geographic spatial filter effectively isolates the spatial structure in the data, facilitating the resolution of endogeneity issues.

More recently, Fingleton (2022) advanced the field further by developing a dynamic spatial panel model with synthetic instruments and a difference generalised method of moments estimation. This method is particularly effective in overcoming the challenges of weak or invalid instruments, which can undermine the reliability of IV methods. By incorporating synthetic instruments and addressing simultaneity bias, Fingleton's approach provides a robust framework for estimating dynamic spatial models.

Eigenvectors, Spatial Filter, and Synthetic instruments

Addressing endogeneity in spatial econometric models presents a significant challenge: maintaining correct spatial dependencies while ensuring the exogeneity of the instruments used. Directly employing other exogenous instruments often fails to replicate the same spatial dependence as the endogenous variables, resulting in inaccurate estimations and biased results. To overcome this, constructing synthetic instruments tailored to the spatial characteristics of the data is essential. These synthetic instruments preserve spatial dependencies and capture the true relationships within the model.

The eigenvector-based spatial filtering method developed by Griffith (2003) has been widely applied across various disciplines to address this issue. This method generates proxy variables, known as spatial filters, by decomposing a transformed and exogenously defined connectivity matrix that represents the dependence structure among the units of analysis (Juhl, 2022). Spatial filters constructed from eigenvectors can then be used to create synthetic instruments.

Eigenvectors represent distinct orthogonal spatial patterns that capture the spatial structure of the data. They are derived through a principal component analysis of the spatial autocorrelation matrix, which quantifies the degree of similarity between neighbouring observations (Juhl, 2022). The spatial autocorrelation matrix is itself derived from a spatial connectivity matrix, which defines geographic relationships among observations based on criteria such as distance, adjacency, or contiguity.

Each eigenvector corresponds to an eigenvalue that indicates the amount of spatial variation explained by that eigenvector. These eigenvectors can be ranked according to their eigenvalues and used to decompose the spatial variation of a variable into different scales or

patterns (Griffith, 2003; Tiefelsdorf and Griffith, 2007).

Spatial filters, which are weighted linear combinations of eigenvectors, extract the spatially correlated component of a variable. They account for spatial heterogeneity or spillover effects, which might otherwise bias the coefficient estimates of interest. The selection of eigenvectors for spatial filters depends on research priorities. For example, Pace et al. (2012) recommend including only a subset of eigenfunctions with the largest eigenvalues. This careful selection ensures that spatial autocorrelation effects and their directions are appropriately captured.

Synthetic instruments are artificial variables constructed from the spatial configuration of the system, satisfying the requirements of instrumental variables. According to Fingleton (2022), synthetic instruments are the predicted values obtained by regressing variables influenced by spatial terms and their neighbouring values on weighted sums of some orthogonal map patterns from a symmetrical N by N matrix of geographic connections. These synthetic instruments can overcome the problem of weak or invalid instruments in spatial panel data models, particularly when the spatial lag or error term is endogenous.

By constructing synthetic instruments tailored to the spatial characteristics of the data, researchers can preserve spatial dependencies and achieve more accurate estimations. This approach mitigates the risk of biased results and enhances the reliability of spatial econometric models.

Estimation of Instruments

Finally, our research contributes to the literature on synthetic instrument estimation. The choice and validity of instruments are critical for obtaining reliable and consistent estimates

of model parameters. However, the use of instruments presents numerous challenges, including over-identification, instrument proliferation, weak instrument bias, multicollinearity, and endogeneity.

Over-identification occurs when the number of instruments exceeds the number of endogenous variables in an instrumental variable (IV) model. This imbalance can lead to issues such as instrument proliferation and weak instrument bias, undermining the validity of IV estimates. For example, Fingleton (2022) conducted an over-identification test using the number of eigenvectors as instruments and found evidence of instrument proliferation. To address this, Baltagi et al. (2019) employed a spatial version of the difference generalised method of moments estimator, which is well-suited for handling multiple synthetic instruments and spatial dependence. This approach mitigates the errors associated with the standard two-stage least squares estimator in spatial regression models.

Multicollinearity often arises when the independent variables in a model are highly correlated, which can result from the data structure or a lack of variability among predictors. This issue complicates the estimation of precise and consistent model coefficients. Griffith (2003) highlighted that spatial lag in spatial regression models can exacerbate multicollinearity, particularly in models such as the spatial Durbin model and the spatial autocorrelation lag model, where spatial lag and synthetic instruments coexist. To address multicollinearity, the method proposed by Juhl (2022), based on the variance inflation factor, can be utilised in the final residual estimation. This method helps identify and reduce the impact of multicollinearity, ensuring more reliable estimates of the model coefficients.

Based on these reviews of the literature, we developed the following hypotheses to guide our estimation:

- Synthetic instruments derived from the eigenvector-based spatial filter are valid and exogenous, and can be used to measure the impact of endogenous transport services on regional economic outcomes.
- Synthetic instruments derived from the eigenvector-based spatial filter are robust across different spatial model specifications and estimation methods.
- The spatial lag and synthetic instruments are sufficiently distinct, avoiding issues of collinearity.

In summary, our research makes several key contributions to the field of spatial econometrics. Firstly, we address the selection process for eigenvectors by implementing a refined methodology for their identification and utilisation in our analysis. This approach improves the understanding of how these vectors are chosen and applied in empirical research, offering a clearer framework for other researchers.

Secondly, our study aims to validate the effectiveness and robustness of synthetic instruments in spatial models. By conducting rigorous validity tests within our analytical framework, we contribute to the empirical validation and understanding of their utility in capturing the true impacts of transportation on regional economic outcomes.

In the next section, we describe the methodology of this chapter in greater detail, including the data sources, model specification, and estimation techniques.

2.3 Empirical Data Description

Our analysis draws on annual data from 30 provinces, municipalities, and autonomous regions in China, spanning the period from 2001 to 2019. These data, primarily sourced

from the annual provincial sections of the National Bureau of Statistics of China and the China Statistical Yearbook, are complemented by geographic coordinates—latitude, longitude, and provincial borders—obtained from Google Maps.² This combination provides a comprehensive spatial and temporal context for our study.

We exclude Hong Kong, Macau, Taiwan, and Tibet from the analysis. Taiwan, Hong Kong, and Macau have distinct economic and administrative systems, making data standardisation and interpretation inconsistent with those of mainland China. Tibet is excluded due to limited reliable data, primarily resulting from its remote location and infrastructural challenges. Including these regions could compromise the consistency and validity of our findings.

The data description includes both economic and geographic data, as detailed in Table B.1 in the appendices. Our research focuses on the annual value-added of industrial production as the dependent variable. This metric captures the net output of industrial enterprises at year-end, representing the monetary value of their production activities after accounting for all costs and intermediate consumption. According to the National Bureau of Statistics of China, value-added industrial production reflects the net results of industrial production activities, excluding external costs and inputs, thereby offering a clearer picture of industrial production and growth.³

For the independent variables, we assess freight efficiency as a measure of freight services within the transportation system. Freight efficiency is defined as the ratio of total freight volume to the number of employees engaged in freight transportation activities within a given region. We adopt the measurement proposed by Rodrigue et al. (2006), which evaluates transportation systems in terms of productivity. Productivity metrics, such as the output-

²China Statistical Yearbook, from: <https://www.stats.gov.cn/english/Statisticaldata/yearbook/>

³Indicators Explanation and Calculation notes from: <https://data.stats.gov.cn/english/easyquery.htm?cn=E0103>

to-labour ratio, are central to assessing transport services, as they reflect the efficiency of labour in handling freight volumes. This ratio is particularly effective for evaluating transport service productivity, as it directly relates freight volume to the number of workers, providing clear insights into labour efficiency and overall service performance.

Specifically, we consider all labour involved in freight services as carriers supplying transport services for shipping industrial goods. Freight efficiency is calculated by dividing the combined freight volume of rail and road by the local employment figures in the freight transport sector. Transport freight productivity reflects how effectively the sector contributes to economic activity by examining the industrial output generated per worker. Higher productivity in the transport sector indicates that each transport worker facilitates greater industrial output, signifying increased economic value generated with fewer labour inputs.

As noted in the previous section, freight services are not expected to be exogenous. Shipping factors are often subject to endogeneity due to omitted variables, such as regional policy interventions or broader economic influences that significantly affect freight services. The provision of transportation services is inherently linked to strategic decisions regarding investment in specific transport technologies (Hafner et al., 2022). Additionally, Behrens and Picard (2011) emphasise that freight services are shaped by the spatial distribution of demand within regional markets. The spatial concentration of industries can result in inelastic transportation services due to high localised demand, further contributing to the endogenous nature of freight services in regional studies.

Our empirical regression incorporates several control variables to account for regional characteristics. These include population size, education level, labour force, government support, and regional economic level. Population size is included as it provides both a labour pool

and a consumer base critical for industrial activity. Government support, measured through general government expenditure, evaluates the role of public policies and investments in facilitating or obstructing industrial activities.

Education expenditure is used as a proxy for education level, reflecting the region's investment in skills development and its influence on workforce productivity and industrial output. We select total education expenditure as the control variable instead of per capita education expenditure to avoid multicollinearity, as per capita expenditure is derived from total expenditure and population, the latter of which is already included in the model. Total education expenditure provides a direct measure of educational investment and aligns with policy perspectives, which typically focus on the overall allocation of educational resources and their contributions to regional development rather than per capita levels.

Furthermore, nominal economic indicators are adjusted for price changes using production price indices to obtain real values. This adjustment facilitates accurate temporal comparisons by accounting for inflationary effects. For statistical estimation, all variables are transformed into their logarithmic forms to stabilise variance and normalise data distribution. However, the logarithmic transformation is not applied to the GDP growth rate, preserving the linear nature of growth measurements over time. A statistical description is presented in Table B.2 in the Appendix.

2.4 Methodology

This section builds on earlier results by integrating a dynamic spatial Durbin model (SDM) with synthetic instruments to refine our estimation approach. We begin by developing a dynamic SDM model that incorporates the empirical data discussed in the preceding section.

First, we define our model using a binary weight matrix to represent inter-provincial connectivity in China. This weight matrix captures spatial relationships between provinces, reflecting how changes in one region may influence neighbouring regions. Next, we employ supervised spatial filters to generate feature eigenvectors. These eigenvectors are analysed using the coefficient of spatial autocorrelation to ensure they accurately represent the spatial structure of the data. This step is critical for preserving spatial dependencies in subsequent estimations.

To evaluate the efficacy and potential over-identification of the synthetic instruments, we apply the two-step GMM. This method enables the use of instrumental variables that are correlated with endogenous regressors but uncorrelated with the error term. By employing GMM, we test the validity of our instruments through over-identification tests, such as the Hansen J-test, which assesses whether the instruments are appropriate and free from bias. The two-step GMM approach further improves efficiency by utilising a robust weighting matrix in the second step.

After establishing the validity of the instruments, we analyse their effectiveness in capturing spatial spillover effects over both short-term and long-term periods. In the short term, spillover effects may be limited or localised, but over time, these impacts can extend across a broader geographic area as economic linkages and infrastructure networks strengthen.

Lastly, we perform a residual analysis to investigate the robustness of the instruments used in the model. This step ensures that the residuals—the unexplained portions of the model—do not exhibit systematic patterns, such as those indicative of omitted variable bias, autocorrelation, or other specification issues. By verifying that the residuals are randomly distributed as expected, we confirm that our model is well-specified and that the instruments do not

introduce bias or distort the relationships between variables.

2.4.1 Weight Matrix

In our study, we employ adjacent weight matrices to construct synthetic instruments and estimate spatial relationships in a subsequent section. Weight matrices define the spatial relationships among different regions, enabling us to capture the influence of one region on another. These relationships are typically based on criteria such as contiguity, specifically whether regions share a border.

We specifically use geographic weight matrices represented as binary adjacency matrices. In these matrices, each element W_{ij} is defined as follows:

$$W_{ij} = \begin{cases} 1, & \text{if } i \text{ and } j \text{ are neighbors,} \\ 0, & \text{not neighbors.} \end{cases} \quad (2.1)$$

This binary adjacency matrix indicates whether two regions share a border (1) or not (0). Using such matrices is a common practice in the construction of eigenvectors for spatial analysis (Tiefelsdorf and Boots, 1995; Griffith, 1996; Fingleton, 2022; Juhl, 2022).

To minimise spatial bias in instrumental variables, we standardise the spatial weights within the model and construct instrumental variables accordingly. By incorporating the spatial relationships defined by the weight matrices, we create synthetic instruments that preserve the spatial dependencies inherent in the data.

2.4.2 Synthetic Instruments Construction

We adopt the methodologies proposed by Griffith and Chun (2014) and Juhl (2022) to construct synthetic instruments using a supervised spatial filter. This process begins by integrating endogenous variables with an exogenous weight matrix to derive eigenvectors.

The selection of eigenvectors is systematically guided by the application of a supervised spatial filter, which employs the Moran Coefficient (MC) as a benchmark. As noted by Juhl (2022), the MC quantifies the spatial autocorrelation of each eigenvector, providing a robust basis for selecting those most suitable for our model. By applying this supervised filter, we ensure that only eigenvectors containing meaningful spatial information are included in the final model.

Eigenvectors Construction

The eigenvectors construction is based on the eigenfunction decomposition of a transformed connectivity matrix, encapsulated by the equation from Juhl (2022)

$$MVM = E\Lambda E' \tag{2.2}$$

In this context, M represents a symmetric and idempotent projection matrix. V denotes the predetermined connectivity matrix, with the eigenvectors in E being the orthogonal outputs derived from MVM , and Λ being the diagonal matrix of eigenvalues. These eigenvectors capture unique spatial patterns based on the spatial arrangement of units.

The projection matrix M is constructed as follows:

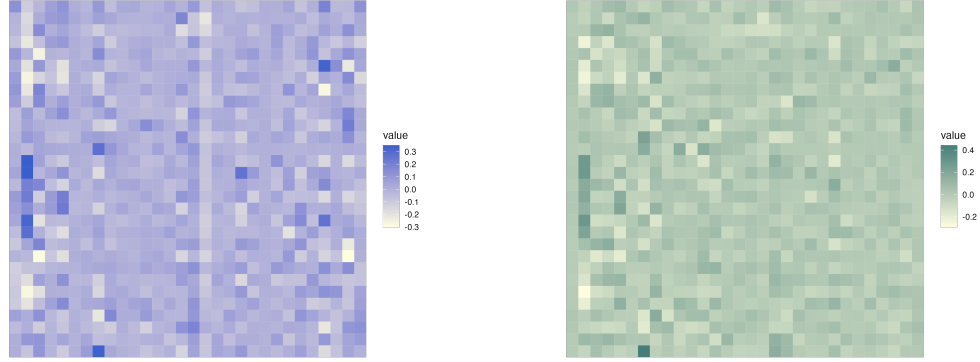
$$M = I - X(X'X)^{-1}X' \quad (2.3)$$

where I is the identity matrix and X is the design matrix of covariates. The resulting eigenvectors are orthogonal to the design matrix, ensuring no correlation with the covariates. As established in the literature, the selection of eigenvectors for the regression model is restricted to a subset due to their number matching the count of observations (Juhl, 2022).

While the instruments employed by Gallo and Páez (2013) are suitable for cross-sectional studies, identifying appropriate analogues for panel data analyses presents greater complexity. Instrumental variables that are highly correlated with the regressors often become unintentionally correlated with the error term, thereby compromising the validity of the model. Conversely, instruments with low correlation to the endogenous variables tend to be weak and lack relevance.

Fingleton (2022) addresses the validation of such synthetic instruments, particularly through the application of the two-step GMM in dynamic panel regression models. To adapt this approach for panel data analysis, we construct eigenvectors on an annual basis and subsequently aggregate them. This process ensures that the synthetic instruments remain relevant and valid over time, enabling more accurate estimation within dynamic panel models.

Figure 2.1: Visualization of All Eigenvectors



(a) Eigenvectors for Railservices

(b) Eigenvectors for Roads-services

Note: The heat map illustrates a distribution of all eigenvectors relative to endogenous variables.

To provide a comprehensive overview, we calculate the average values of the eigenvectors and present these data in Figure 2.1. This figure illustrates the distribution of all eigenvectors across various endogenous variables.

The figure reveals that both sets of eigenvectors exhibit a random distribution pattern. Randomly distributed eigenvectors indicate that these vectors are statistically independent of potentially confounding variables not included in the model. This independence ensures that any correlation observed between the instruments and the dependent variables reflects genuine causal relationships, rather than spurious associations arising from omitted variable bias.

Supervised Eigenvector Selection

Selecting appropriate eigenvectors is a critical and flexible component of the eigenfunction spatial filtering methodology. As outlined by Juhl (2022), the selection process can be di-

vided into two filtering strategies, both of which are based on the spatial autocorrelation inherent in the eigenvectors. The first strategy employs a direct threshold of spatial autocorrelation to determine relevance, while the second strategy selects eigenvectors based on the regression equation of residuals.

In both strategies, the initial step involves creating a candidate subset $E^C \subseteq E$ defined according to established criteria. Subsequently, the final selection of eigenvectors is made, which are then incorporated into the model.

The first approach aims to distinguish the value of eigenvectors which is associated with the spatial autocorrelation measured by the MC. Griffith (2003) introduces a quantitative threshold to delineate this subset by evaluating the ratio $\frac{MC_i}{MC_{max}}$ for all eigenvectors i in $\{1, 2, \dots, n\}$, including in E^C those with a ratio exceeding 0.25. The exclusion of eigenvectors linked with negative spatial autocorrelation is standard to prevent the confounding effects of mixed autocorrelation signals.

Alternatively, Chun et al. (2016) propose an unsupervised spatial filter, which employs a non-linear calculation to determine the optimal size of the candidate set based on the spatial autocorrelation level and the quantity of eigenvectors with positive correlation. This method is applicable only when the residuals exhibit positive spatial autocorrelation. In generalised linear models, it is essential to ensure the uncorrelatedness of eigenvectors, analogous to uncorrelated weights in linear models, to address multicollinearity concerns.

Dynamic spatial models often deviate from the assumptions of linear regression due to the presence of spatial dependencies. To evaluate the spatial capabilities of our synthetic instruments, we utilise threshold-based supervised spatial filters. This method enables us to systematically examine the performance of instruments by analysing geographic agglomera-

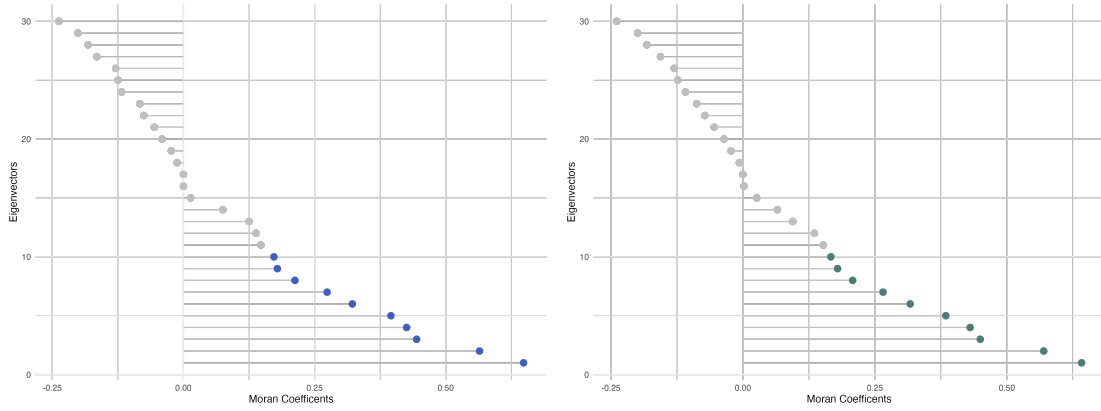
tion trends and spatial autocorrelation, assessing how each factor distinctly influences spatial dynamics within the model.

Therefore, we follow the MC approach, which was established by Griffith (1996) and Tiefelsdorf and Boots (1995). The Moran coefficient, a measure utilized in spatial analysis, can be computed for each eigenvector using a specific formula that incorporates the eigenvalue and the number of observations. The MC for an eigenvector is given by Equation 2.4 . Where λ_i is the eigenvalue associated with the i^{th} eigenvector, n is the number of observations, and $1'V1$ represents the sum of the elements in the spatial weights matrix V , pre-multiplied and post-multiplied by a row vector of ones :

$$MC_i = \frac{\lambda_i n}{1'V1} \quad (2.4)$$

As Juhl (2022) indicates, the first eigenvector depicts the spatial pattern permitted by W with the largest possible degree of positive spatial autocorrelation, and the last eigenvectors in the set capture patterns of negative autocorrelation. We present the mean MC for all instruments over the examined years as illustrated in Figure 2.2. We highlight points with a threshold ratio $\frac{MC_i}{MC_{\max}} \geq 0.25$, which are prominently marked in the graph.

Figure 2.2: Moran Coefficients for All Eigenvectors



(a) Rail Services Eigenvectors

(b) Road Services Eigenvectors

Notes: The x-axis presents MC for all eigenvectors. The y-axis is the sequence number of the eigenvectors. Points in colour are eigenvectors over the threshold: $\frac{MC_i}{MC_{max}} \geq 0.25$.

In our analysis, we sought to effectively address the issue of weak instruments. When the residuals of a model exhibit positive spatial autocorrelation, it is essential to exclude eigenvectors with negative spatial autocorrelation to maintain analytical integrity. This approach is based on the principles outlined by Tiefelsdorf and Griffith (2007), who suggest that combining eigenvectors associated with both positive and negative spatial autocorrelation can complicate the analytical process and introduce inconsistencies.

Given that the underlying variables in our study displayed positive spatial autocorrelation, we carefully selected eigenvectors whose spatial autocorrelation patterns aligned with this overall positive trend. This deliberate selection ensured that our analysis remained robust by harmonising the characteristics of our instruments with the inherent spatial properties of the data.

To further enhance the effectiveness of the instruments, we specifically selected the top ten eigenvectors that exceeded the threshold in each case, as illustrated in Figure 2.2. These

eigenvectors were chosen because they surpassed the threshold ratio and exhibited positive MC, a measure of spatial autocorrelation. Selecting eigenvectors with positive MC ensures consistency with the positive spatial autocorrelation present in the data, thereby reinforcing the robustness and reliability of our instruments.

2.5 Econometric Analysis

This section presents the theoretical verification and spatial empirical analysis of the implications of synthetic instruments. Our simulation is based on the dynamic spatial Durbin model.

We construct a simple dynamic SDM with a one-period lag on the dependent variable. The SDM extends the Spatial Lag Model by incorporating both the spatially lagged dependent variable and the spatially lagged independent variables within the regression equation (Elhorst, 2014). The dynamic SDM model is expressed as Equation 2.5:

$$Y_{it} = \gamma Y_{it-1} + \delta W_{ij} Y_{jt} + \eta W_{ij} Y_{jt-1} + \beta_1 X_{it} + \beta_2 W_{ij} X_{it} + \nu_{it} \quad (2.5)$$

The analysis encompasses N units, which may be interpreted as regions, locations, or individuals, observed across T temporal intervals. The matrix element W_{ij} represents the exogenous, spatial interaction in the i, j th position of the $N \times N$ spatial connectivity matrix. Parameters γ , δ , η , β_1 , and β_2 await estimation within the model framework. We assume that the instruments could remove the correlation in the residual term in the model, so the residual term as ν_{it} is assumed as exogenous in the ordinary equation. To test this assumption, we introduce more details in the next subsection.

2.5.1 Two-step GMM Estimation

Instrumental variables are essential tools for addressing endogeneity. When the set of available instruments exceeds the number of parameters to be estimated, the 2SLS approach is typically employed. As noted by Anselin (1988) and Elhorst (2014), MLE relies on assumptions about the distribution of errors, which must be carefully considered in spatial analyses.

MLE has several limitations when addressing spatial dependence and nonlinearity. First, it assumes that error terms are independent and identically distributed (i.i.d.), an assumption often violated in spatial economic models. Spatial dependence implies interactions between regions, leading to spatial autocorrelation in the error terms. When such autocorrelation exists, the independence assumption of MLE is invalidated, resulting in biased estimates. This issue is particularly pronounced in dynamic models or those with spatial autocorrelation, where MLE's assumptions are too restrictive to account for complex dependence structures.

Another limitation of MLE is its reliance on specific distributional assumptions. While it can be adapted for various distributions, such as Poisson or negative binomial, it requires the data's distribution to be specified in advance. In models with complex nonlinear or spatial structures, accurately specifying the distribution of the error term can be challenging, increasing the risk of estimation bias. If the assumed distribution does not align with the actual data, MLE estimates may deviate significantly from the true values.

The GMM estimator provides a robust alternative. GMM is obtained by minimising a criterion function that measures the distance between sample moments and population moments. It is consistent and asymptotically normal if the GMM assumptions hold, including the validity of the instruments and the correct specification of the model. Moreover, GMM is

efficient if the optimal weighting matrix is employed, ensuring that the moment conditions are satisfied and equal to zero.

Another advantage of GMM is its ability to use the two-step approach. In the second step, an optimal weighting matrix enhances the efficiency and accuracy of estimation while addressing issues such as heteroscedasticity and serial correlation in the model. This makes GMM more robust than MLE in complex spatial econometric models, as it provides consistent estimation results without relying on specific distributional assumptions.

For spatial regression models, where geographic or spatial dependencies introduce endogeneity, the application of GMM is particularly effective. Liu and Saraiva (2015), along with further tests by Fingleton, demonstrate that GMM estimators can handle spatially lagged dependent variables effectively. These estimators mitigate bias and improve efficiency, even in the presence of numerous synthetic instruments derived from the spatial weight matrix, which can lead to over-identification.

Gallo and Páez (2013) highlight the challenges of evaluating instrumental variables within dynamic spatial models and emphasise the validity of GMM in instrument estimation. Supporting this, Jin and Wang (2021) demonstrate the consistency and asymptotic normality of GMM estimators under heteroscedasticity and spatial autocorrelation. As Baltagi (2013) observes, GMM circumvents the reliance on the normality assumption of error terms and strict initial conditions, offering a robust framework for complex models with endogenous variables.

GMM estimation minimises the distance between sample and population moments, yielding consistent and asymptotically normal estimators when its conditions are satisfied, including instrument validity and accurate model specification. Liu and Saraiva (2015) build on

Lee (2007)’s GMM estimator, leveraging moment conditions that link the spatially lagged dependent variable to the error term and instruments. The reliability of GMM estimation has been validated through Monte Carlo simulations in various spatial regression settings, as explored by Fingleton. Following Fingleton’s approach, we employ the two-step GMM to address heteroscedasticity and obtain more robust standard errors.

The literature highlights that geographic weight matrices, while useful for measuring spatial effects, can introduce spatial dependencies that cause bias and inefficiency in the ordinary least squares (OLS) estimator. This issue renders traditional endogeneity tests less reliable, as emphasised by Fingleton with respect to the Sargan-Hansen tests. The number of synthetic instruments, often derived from the weight matrix, frequently exceeds the number of endogenous variables. GMM provides an effective solution for handling multiple endogenous variables and instruments.

The Arellano-Bond test (Arellano and Bond, 1991) for serial correlation evaluates whether the error term is independent over time, a crucial assumption for the validity of the GMM estimator. Autocorrelation can reduce the efficiency of the GMM estimator and bias the standard errors. The test uses the residuals from the GMM regression to check for correlation with their own lags, with the null hypothesis positing no serial correlation.

In econometric studies involving panel data, small sample sizes often present challenges that compromise the reliability of statistical inference. Our dataset comprises 570 observations, for 30 provinces over 19 years. One key concern is the potential for downward bias in the standard errors of coefficient estimates when using two-step GMM (Windmeijer, 2005). This bias arises because the second-step GMM estimator relies on an optimal weighting matrix derived from the first-step residuals, introducing additional variability that asymptotic

standard error formulas do not adequately account for in small samples (Windmeijer, 2005).

To address this, we employ the `xtabond2` command in Stata, which incorporates the finite sample correction proposed by Windmeijer (Roodman, 2009). This correction adjusts the standard errors of the two-step GMM estimator to account for the estimation uncertainty of the optimal weighting matrix. By mitigating the downward bias in standard errors, the Windmeijer correction ensures that the results from two-step GMM are more reliable, particularly for hypothesis testing and constructing confidence intervals (Windmeijer, 2005). We provide an explanation in Appendix B.2.

Following Roodman (2009), we apply the two-step GMM estimation technique, incorporating endogenous variables. This approach addresses endogeneity by modelling the relationship between the endogenous variables and the outcome while using instrumental variables to correct for potential bias. To validate our instruments and assumptions, we adjust the dynamic SDM with endogenous variables, as recommended by Fingleton (2022).

$$Y_{it} = \gamma Y_{it-1} + \delta W_{ij} Y_{jt} + \eta W_{ij} Y_{it-1} + \beta_1 X_{1it} + \beta_2 W_{ij} X_{jt} + \beta_2 \tilde{X}_{2it} + \beta_2 W_{ij} \tilde{X}_{2jt} + \epsilon_{it} \quad (2.6)$$

$$\epsilon_{it} = \mu_i + \nu_{it} \quad (2.7)$$

Consider the general model, $\tilde{X}_{2it}$ represents the endogenous variables, specifically rail and road services. We assume that ν_{it} is an exogenous error term, as shown in Equation 2.5. We adopt the error term structure proposed by Fingleton (2022)), which allows each endogenous

variable to be influenced by a distinct error process. Thus, ϵ_{it} includes μ_i as an endogenous component, which is correlated with $\tilde{X}_{2it}$.

We also emphasise the importance of testing for heteroscedasticity, which examines whether the variance of the error term remains constant across observations. If heteroscedasticity is present, the GMM estimator may become inefficient, and the standard errors may be biased. To address this issue, we employ the two-step GMM method to correct for heteroscedasticity, following Fingleton (2022).

In the spatial Durbin model, we treat the lagged dependent variables and the spatially lagged components as endogenous variables, in addition to the other endogenous variables. Synthetic instruments are selected based on the highest positive spatial autocorrelation. After this selection, we compare results obtained using additional instruments exceeding the threshold ratio. This comparison enables us to evaluate the impact of instrument selection on GMM estimation.

Table 2.1 presents the results of our two-step GMM tests using Stata’s `xtabond2` command. The table displays models with 4, 7, and 10 synthetic instruments, illustrating how increasing the number of instruments affects the model’s complexity and robustness. A total of 144 instruments is tested, reflecting extensive examination across different instrument sets. By comparing these GMM results, we evaluate the effectiveness of the synthetic instruments.

Additionally, we provide supplementary information in the Appendix, including negative coefficients and total synthetic instruments for validity estimation, as shown in Table B.3. This additional analysis validates our instrument selection and further reinforces the robustness of our findings.

The Arellano-Bond test is designed to detect autocorrelation in the first-differenced residuals

Table 2.1: Two-Step GMM Diagnostic Tests

Synthetic Instrument in MC	with 4	with 7	with 10	Total
Number of instruments	92	98	104	144
Arellano-Bond Test				
AR(1)	-2.19 (0.028)	-2.45(0.014)	-2.42(0.016)	-1.55 (0.122)
AR(2)	1.03(0.302)	0.81(0.416)	0.82(0.410)	0.83 (0.404)
Instrument Validity				
Sargan-Hansen test of overid	20.36	20.42	20.42	16.68
p-value	1.000	1.000	1.000	1.000
d. of freedom	76	82	88	128

Notes: Note: This table presents GMM results for 4, 7, and 10 synthetic instruments with a MC over the threshold ratio as 0.25. Total instruments contain all instruments with both negative and positive MC. P-value presented in parentheses for Arellano-Bond tests.

of panel data models. Autocorrelation occurs when residuals are correlated with their own lagged values, potentially biasing both the estimates and the standard errors. As outlined by Arellano and Bond (1991), the null hypothesis of the Arellano-Bond test is that there is no autocorrelation of a specified order. For instance, the null hypothesis for the AR(1) test posits no first-order autocorrelation, while for the AR(2) test it posits no second-order autocorrelation.

The results presented in Table 2.1 indicate significant AR(1) autocorrelation but insignificant AR(2) autocorrelation. This outcome aligns with the assumptions of the first-difference GMM estimator, which assumes that the error term follows a first-order autoregressive process. Specifically, this implies that the error term in the current period is correlated with the error term in the previous period but uncorrelated with error terms from earlier periods. Consequently, lagged values of the endogenous variables can serve as valid instruments, as they are correlated with the first-differenced endogenous variables but not with the first-differenced error term. The absence of second-order autocorrelation, confirmed by p-values exceeding 0.30, supports the reliability of the instruments and indicates that the model is

not affected by higher-order autocorrelation.

We employed a supervised instrument selection strategy based on the Moran Coefficient. Instruments were ranked from highest to lowest MC values and matched against a predefined threshold ratio. The Arellano-Bond tests for AR(1) and AR(2) across all models confirm acceptable levels of serial correlation. The significance of AR(1) indicates the expected presence of first-order correlation in differenced GMM models, while the lack of significance for AR(2) aligns with the model's specifications, confirming the absence of second-order serial correlation.

We also conducted tests using instruments with exclusively positive and negative MC values, as detailed in Table B.3 in the Appendix. The results remained valid. When all instruments were included, the AR(1) results failed to reject the null hypothesis of no autocorrelation, demonstrating effectiveness only when the MC signs were consistent. Mixed selection did not yield significant results.

The Hansen J-test assesses whether the instruments used in IV estimation are valid and uncorrelated with the error term (Wooldridge, 2010). The test statistic is derived from the difference between the GMM criterion function evaluated at the IV estimator and at the optimal GMM estimator (Hansen, 1982). This statistic follows a chi-square distribution, with degrees of freedom equal to the number of over-identifying restrictions—that is, the difference between the number of instruments and the number of endogenous variables. A large test statistic and a small p-value lead to rejection of the null hypothesis, indicating that the instruments are not exogenous or are correlated with the error term.

The results of the Sargan-Hansen over-identification test, showing uniform p-values of 1.000 across different instrument sets, provide no evidence of invalid over-identifying restrictions.

This supports the hypothesis that the instruments employed in the model are valid.

We incorporate the spatial lag into the GMM estimation equation to test the instruments further. Given the potential for spatial autocorrelation, the Sargan-Hansen test for over-identifying restrictions remains a focal point in our analysis. The null hypothesis of the Hansen test is that the instruments are valid, meaning they are both uncorrelated with the error term and relevant for the endogenous variables.

High p-values from the Hansen test indicate that we cannot reject the null hypothesis, affirming the validity of the synthetic instruments, even as their number increases. This result implies non-rejection of the null hypothesis of instrument validity. Moreover, the range of instruments—from 92 to 144—does not compromise the robustness of the Hansen test, suggesting that the instruments remain valid and do not overfit the model. The consistent degrees of freedom and high p-values further reinforce this interpretation.

When considering the Moran Coefficient value of the instruments, we find that those exceeding the threshold ratio exhibit consistency. However, including instruments with negative MC values in the total test reduces the significance level in the Arellano-Bond tests and the chi-squared statistic of the Sargan-Hansen tests. This analysis highlights the importance of a rigorous instrument selection process and validates the use of the MC as a criterion for guiding this process.

The results from our analyses confirm the robustness and validity of the instruments employed in our GMM estimations. The Arellano-Bond test results indicate consistent first-order autocorrelation and the absence of second-order autocorrelation. This alignment with the expectations of the first-difference GMM estimator supports the use of lagged values as valid instruments. Furthermore, the Hansen J and Sargan-Hansen tests substantiate the

integrity of our instrument set, showing no evidence of invalid over-identifying restrictions and confirming the correct specification of the model.

Our approach to instrument filtering, guided by the MC, effectively adjusts the correlation of the residual term in the context of endogeneity. This reinforces the assumption that supervised filtered instruments provide an exogenous relationship between the residual term and the endogenous variables.

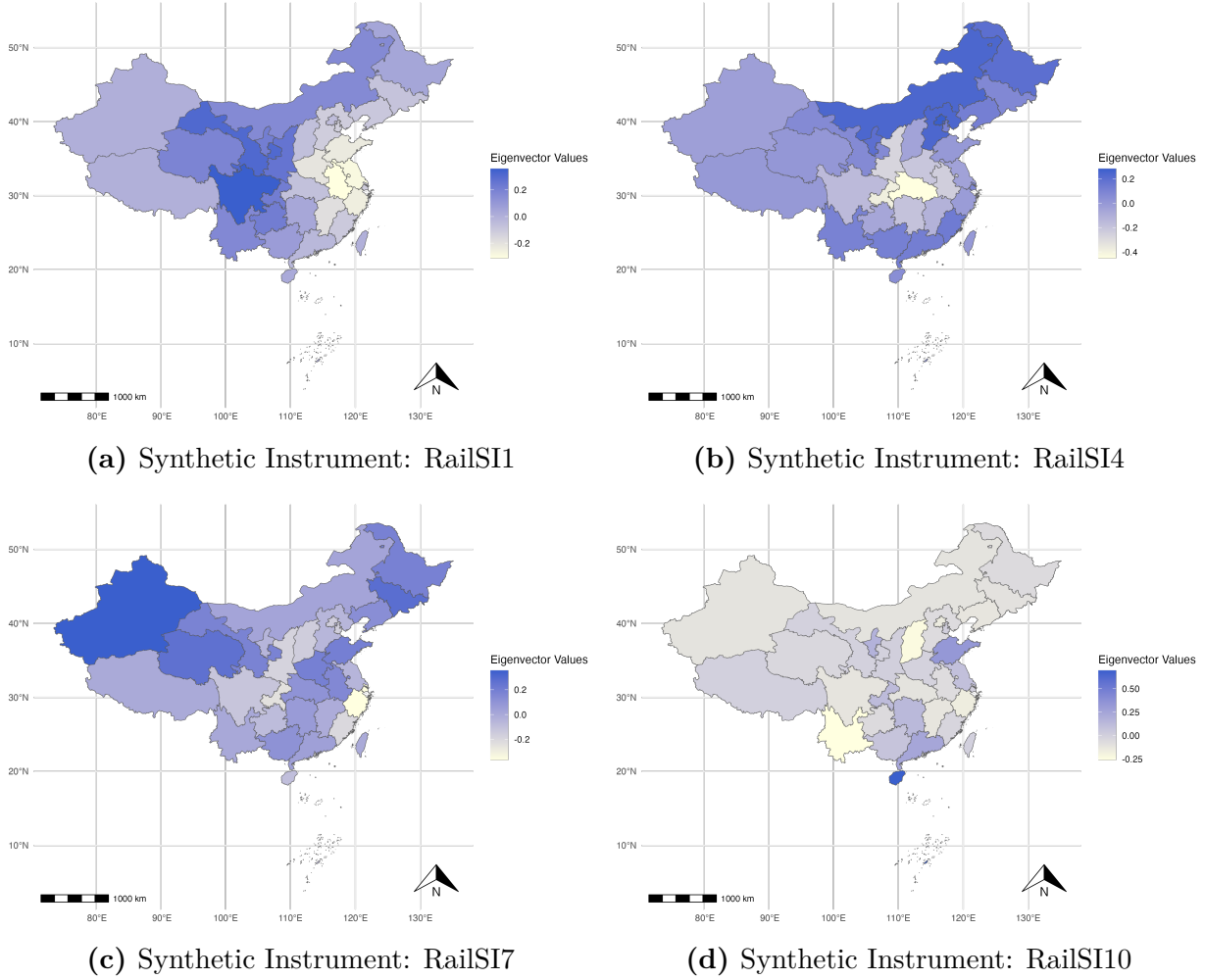
In conclusion, the inclusion of spatial lag and rigorous instrument testing confirms the robustness and validity of our GMM estimations. The Arellano-Bond test results validate the use of lagged values as instruments, while the Hansen J and Sargan-Hansen tests confirm the overall integrity of the instrument set. The application of the MC in instrument selection proves effective in maintaining instrument relevance and mitigating bias.

2.5.2 Spatial Analysis in Synthetic Instruments

Geographic Distribution

Spatial analysis typically begins with an examination of the geographic distribution to assess whether the raw data exhibit notable spatial characteristics. We initiated our analysis by exploring geographic patterns to analyse spatial trends across various synthetic instruments.

Figure 2.3: Geographic Distribution on Rail Services and Instruments



Note: SI refers to synthetic instruments. For presentation purposes, the values for Hong Kong, Macau, Taiwan, and Tibet have been set to 0.

Our objective is to determine whether these instruments exhibit similar spatial distributions and to examine how the MC influences these patterns. To support this analysis, we visualised the geographic distribution of synthetic instruments, organised by different MC values, as shown in Figure 2.3.

Firstly, each figure demonstrates distinct geographic distributions, highlighting the diversity of these instruments. This differentiation is crucial, as similarity among instruments could lead to multicollinearity, potentially undermining the robustness of the analysis.

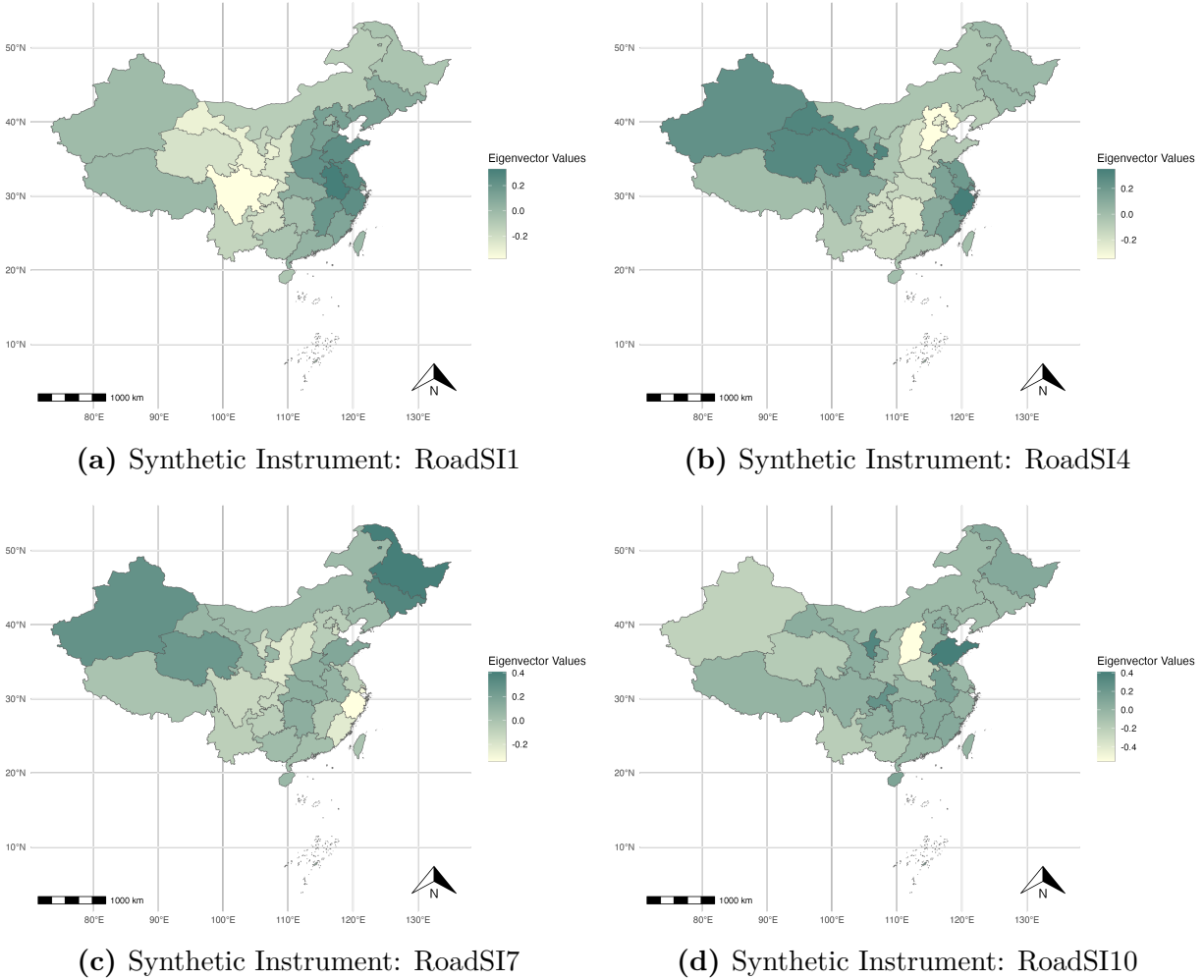
The distinct patterns observed for the instruments related to rail and road services underscore their effectiveness in capturing unique geographic variations. This uniqueness is essential for ensuring that the spatial models accurately represent true variability rather than redundant information. Such geographic specificity is particularly important in models where spatial factors are critical to predicting outcomes.

Secondly, the figures reveal that synthetic instruments with higher MC values tend to exhibit more pronounced spatial trends. Instruments with well-defined and concentrated geographic patterns indicate robust spatial autocorrelation, suggesting their effectiveness in capturing spatially clustered phenomena.

In contrast, instruments with lower MC values, such as RoadSI10 and RailSI10, display flatter and more dispersed geographic distributions. This pattern suggests a more stochastic nature or reduced spatial dependency in the variables these instruments measure.

The less pronounced spatial clustering in these cases may indicate that the variables are influenced by broader, non-spatial factors or that these instruments are less effective in capturing spatial dependencies. This finding highlights the importance of selecting instruments with strong spatial properties when modelling geographic phenomena.

Figure 2.4: Geographic Distribution on Road Services and Instruments



Note: SI refers to synthetic instruments. For presentation purposes, the values for Hong Kong, Macau, Taiwan, and Tibet have been set to 0.

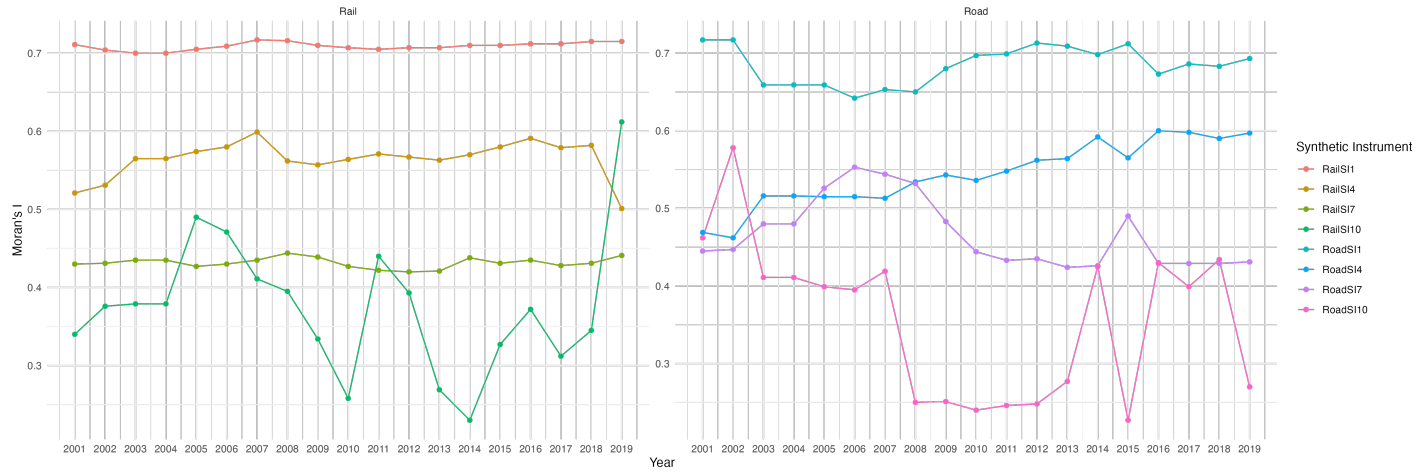
Our analysis highlights the importance of selecting instruments that align with the spatial characteristics of the data under study. Instruments exhibiting strong spatial trends are more effective at capturing clustered phenomena, while those with flatter distributions can offer valuable insights into more diffuse or random spatial processes. Striking this balance

ensures that spatial models remain robust and accurately reflect the true geographic trends in the data.

Spatial Autocorrelation

In the previous section, we screened out the synthetic variables based on the threshold of the average Moran coefficient. Next, we focused on the spatial autocorrelation transformation of the synthetic variables in the time trend. In Figure 2.5, we display panel Global Moran I data from the synthesis instruments, illustrating the level of spatial autocorrelation across time trends. SI indicates the synthetic instrument.

Figure 2.5: Spatial Autocorrelation in Synthetic Instruments



The data consistently reveal significant and positive spatial autocorrelation, highlighting two prominent trends.

First, synthetic instrumental variables with the highest Moran Coefficient, specifically RailSI1 and RoadSI1, exhibit significantly higher Moran index values compared to other instruments. Secondly, a distinct trend is observed in the performance of SI1 across the two types

of instrumental variables, where the time trend appears flatter. As the Moran Coefficient diminishes, this trend becomes increasingly volatile, suggesting potential instability in spatial relationships over time.

Overall, all synthetic variables exceeding the specified threshold demonstrate spatial dependence. We emphasise the persistence of spatial correlation, as it underpins the ongoing evaluation of spatial spillover effects within the context of a panel model. This suggests that the instruments exhibit a more pronounced spatial correlation, indicating stronger spatial dependence. Such dependence is typically associated with more substantial spatial spillover benefits in spatial econometric models.

Levin-Lin-Chu Test

In empirical research involving time series and panel data, testing the stationarity properties of the data is a fundamental step. A series is deemed stationary if its statistical properties, such as mean and variance, remain constant over time (Greene, 2012). If a series exhibits a unit root, it indicates non-stationarity, where shocks have persistent effects, potentially leading to unreliable inferences due to the risk of spurious regression (Wooldridge, 2010).

Panel unit root tests, such as the Levin-Lin-Chu test, are particularly advantageous as they leverage both cross-sectional and time series dimensions, thereby increasing the power of the test (Levin et al., 2002).

In this study, we apply the Levin-Lin-Chu (LLC) test to the variable $\ln Ind$ to determine whether it is stationary or contains a unit root. The variable $\ln Ind$ represents the natural logarithm of industrial output across 30 cross-sectional units over 19 periods. The null hypothesis of the LLC test posits that all panels contain unit roots, whereas the alternative

Table 2.2: Levin-Lin-Chu Test

Null Hypothesis (Ho):	Panels contain unit roots
Alternative Hypothesis (Ha):	Panels are stationary
Number of panels:	30
Number of periods:	19
Statistic	p-value
Unadjusted t-statistic:	-12.7983
Adjusted t-statistic:	-11.0390
p-value:	0.0000

hypothesis suggests that the panels are stationary.

The test was conducted with an autoregressive parameter assumed to be common across panels, including panel means but excluding a time trend. One lag was used in the regressions, and the long-run variance was estimated using the Bartlett kernel with an average of 8.00 lags. The results, presented in Table 2.2, indicate an unadjusted t-statistic of -12.7983 and an adjusted t-statistic of -11.0390, with a corresponding p-value of 0.0000. These results strongly reject the null hypothesis of a unit root at the 1

Spatial Estimation

The evaluation of the spatial model is organised into two primary segments. The first segment focuses on assessing direct and spillover effects.

This phase tests the efficacy of synthetic instruments in substituting endogenous variables to accurately capture spatial correlation. The second segment pertains to residual analysis, where we examine the ability of the synthetic instruments to enhance results concerning both time series and spatial characteristics.

Additionally, we incorporate the explanatory and control variables introduced in the previous

section. In the spatial models discussed, synthetic instruments are employed to replace endogenous explanatory variables, aiming to mitigate potential biases arising from omitted variables and endogeneity.

The results of our model are presented in Table 2.3. To improve the observational clarity of spatial effects, the model incorporates fixed effects. These fixed effects control for unobserved variables that remain constant over time but vary across regions, such as geographical or institutional characteristics. Furthermore, the model adjusts for annual variations, accounting for economic or policy changes that may influence industrial outputs across different regions. It ensures a more robust understanding of the spatial dynamics.

The lagged dependent variable consistently exhibits high and statistically significant coefficients across all model specifications. These values, ranging from 0.838 to 0.909, indicate a strong persistence effect, whereby industrial output in previous periods significantly influences output in the current period.

Similarly, the spatial autocorrelation coefficients are significant and substantial across all models, ranging from 0.301 to 0.336. This pronounced spatial dependence underscores the critical role of spatial effects in economic modelling, where developments in one region significantly affect neighbouring regions. The finding highlights the interconnected nature of regional economies and the necessity of incorporating spatial considerations into economic analysis. Furthermore, the high R-squared values, ranging from 0.8210 to 0.8567, suggest that the models are well-fitted, explaining a substantial portion of the variance in industrial outputs across regions.

For explanatory variables, synthetic instruments (SIs) are employed to replace endogenous variables such as rail and road services, denoted as RailSI and RoadSI, respectively. The

Table 2.3: Main Results with the Dynamic Spatial Durbin Model

Variables	SI1	SI4	SI7	SI10
Main Estimate				
Lag.Ind	0.882*** (41.620)	0.909*** (43.410)	0.838*** (39.950)	0.904*** (42.860)
RailSI	-0.106** (-2.550)	-0.009 (-0.240)	0.011 (0.350)	0.018 (0.890)
RoadSI	-0.082** (-2.460)	-0.039 (-0.920)	-0.126*** (-2.910)	-0.037* (-1.720)
lnGOVE	-0.010 (-0.400)	-0.008 (-0.320)	0.000 (0.010)	-0.007* (-0.300)
lnPOP	-0.058 (-1.030)	-0.049 (-0.860)	-0.053 (-0.910)	-0.064 (-1.140)
lnEDU	0.049* (1.900)	0.054* (2.140)	0.046 (1.810)	0.046* (1.800)
lnLabour	0.030 (0.920)	0.048 (1.450)	0.041 (1.250)	0.029 (0.890)
GDPgr	0.003 (1.350)	0.004* (1.850)	0.002 (0.820)	0.004 * (1.670)
Spatial Lag of X (W*)				
W*RailSI	0.187*** (3.380)	0.112* (1.940)	0.083 (1.270)	-0.014 (-0.330)
W*RoadSI	0.152*** (3.380)	0.071 (0.980)	0.145* (1.800)	0.145*** (2.820)
W*lnGOVE	0.287*** (3.020)	0.326*** (9.330)	0.292*** (8.120)	0.305*** (8.700)
W*lnPOP	0.522*** (1.420)	0.588*** (5.640)	0.578*** (5.950)	0.518*** (5.410)
W*lnEDU	-0.213*** (0.050)	-0.240*** (-7.570)	-0.234*** (-7.090)	-0.227*** (-7.070)
W*lnLabour	0.142** (2.290)	0.185*** (3.150)	0.195*** (3.260)	0.146*** (2.540)
W*GDPgr	-0.016*** (-0.240)	-0.020*** (-5.370)	-0.011*** (-2.850)	-0.016*** (-4.240)
Spatial Autocorrelation (rho)	0.311*** (14.750)	0.319*** (15.240)	0.336*** (16.040)	0.301*** (14.000)
R-squared	0.8567	0.8382	0.8210	0.8535

Note: *SI* indicates synthetic instruments. *Z*-statistics in parentheses. Significance levels: *** 1%, ** 5%, * 10%.

spatial lags of these instruments exhibit varied impacts across regions and models. The associated coefficients reveal complex and mixed spillover effects, which vary significantly depending on the specific context and model parameters. These nuances are detailed in Table 2.3, where the contributions of freight services to regional dynamics are further explored.

With regard to control variables, most display insignificant effects across all instruments in the main estimation, suggesting limited direct influence on the dependent variable. However, the weighted estimation, which incorporates the spatial lags of control variables, reveals significant effects across all models. This finding highlights the importance of accounting for spatial dependencies when assessing the impact of control variables.

In summary, the model demonstrates the dynamic and spatial interactions shaping regional economic output. The dynamic SDM model addresses strong temporal and spatial dependencies within the data. The significance of infrastructure and socioeconomic variables, both direct and spatially lagged, varies across different synthetic instruments. Notably, the robust performance of the spatial lags indicates that the characteristics of neighbouring regions significantly influence the dependent variable, underscoring the interconnectedness of regions in the dataset.

Further analysis, as Table 2.4 and Table 2.5 explores the direct and indirect effects more deeply to delineate the specific pathways through which these influences manifest.

SI1 exhibits significantly more spatial effects than other instruments in both short-term and long-term scenarios. SI1 not only shows the most substantial effects but also achieves the highest level of significance across all spatial analyses. It effectively captures and interprets the significant effects observed in other instruments. For example, while railSI4 and railSI7 each demonstrate positive short-term spillover effects, SI1 comprehensively incorporates

Table 2.4: Spatial Analysis in Short-term Effects

Synthetic Instruments	SI1	SI4	SI7	SI10
Short Term				
Direct Effect				
RailSI	-0.068** (-2.190)	0.018 (0.720)	0.033 (1.550)	0.016 (1.060)
RoadSI	-0.047 * (-1.920)	-0.021 (-0.750)	-0.097 *** (-3.290)	-0.003** (-0.190)
lnGOVE	0.061 * (3.020)	0.072 (3.520)	0.075*** (3.650)	0.067*** (3.330)
lnPOP	0.069 * (1.420)	0.095 (1.950)	0.091* (1.840)	0.060*** (1.230)
lnEDU	0.001 (0.050)	-0.001 (-0.070)	-0.009 (-0.420)	-0.006*** (-0.260)
lnLabour	0.068 (2.290)	0.098 (3.330)	0.095*** (3.100)	0.068 (2.280)
GDPgr	-0.000 (-0.240)	-0.000 (-0.160)	-0.001 (-0.340)	0.000*** (0.150)
Indirect Effect				
RailSI	0.190*** (3.700)	0.137** (2.390)	0.115 (1.640)	-0.006 (-0.120)
RoadSI	0.150*** (3.500)	0.068 (0.940)	0.120 (1.390)	0.154*** (2.770)
lnGOVE	0.341*** (8.820)	0.394*** (9.950)	0.366*** (8.830)	0.358*** (9.180)
lnPOP	0.608*** (5.300)	0.707*** (5.670)	0.704*** (5.940)	0.594** (5.240)
lnEDU	-0.239*** (-7.530)	-0.273*** (-8.750)	-0.275*** (-8.320)	-0.254 (-8.160)
lnLabour	0.183*** (2.880)	0.245 *** (3.850)	0.262*** (4.000)	0.183** (3.010)
GDPgr	-0.018 *** (-4.380)	-0.023*** (-5.540)	-0.013*** (-3.020)	-0.018 (-4.360)
Total Effect				
RailSI	0.122*** (3.720)	0.156*** (3.460)	0.148** (2.330)	0.011 (0.240)
RoadSI	0.103*** (3.460)	0.047 (0.870)	0.023 (0.310)	0.151*** (2.720)
lnGOVE	0.402*** (10.280)	0.466 *** (11.770)	0.441 *** (10.320)	0.426*** (10.830)
lnPOP	0.677*** (5.580)	0.802 *** (5.940)	0.795*** (6.170)	0.654*** (5.440)
lnEDU	-0.238*** (-8.170)	-0.274 *** (-9.340)	-0.284*** (-8.930)	-0.260*** (-9.050)
lnLabour	0.251*** (3.680)	0.343 *** (5.130)	0.357*** (4.980)	0.251*** (3.910)
GDPgr	-0.018 *** (-4.430)	-0.023 *** (-5.460)	-0.013*** (-3.080)	-0.017*** (-4.200)

Note: Z-statistics in parentheses. ***, **, and * show the significance at of 1%, 5% and 10%, respectively.

these outcomes, highlighting its capability to aggregate and reflect broader impacts.

This integrative feature of SI1 is similarly evident in the analysis of long-term effects, where it again encapsulates and extends the significant effects of other instruments. Besides, the coefficients representing significant effects for SI4, SI7, and SI10 closely align with those of SI1, with minimal variation observed among them. This uniformity is particularly encouraging as it suggests consistent spatial performance across all evaluated instruments, reinforcing their stability and reliability.

These observations align with our initial hypothesis and are supported by our analysis of spatial autocorrelation, which suggests that instruments with a higher MC are more adept at comprehensively interpreting spatial effects within the dynamic SDM. This finding is crucial for our research, emphasizing the importance of selecting appropriate synthetic tools when multiple options are available. Such choices are integral, as they significantly influence the effectiveness and clarity of spatial analyses.

In previous tests, we confirmed the effectiveness and spatial representation capabilities of the synthetic tool for endogenous variables, respectively. Therefore, we illustrate the impact of freight service efficiency on regional industrial production through an empirical analysis using synthetic instruments.

In the short-term analysis, the negative direct effects of RailSI and RoadSI suggest that initially, an increase in freight volume relative to the labour force in both rail and road transportation systems is associated with a decrease in industrial production. This could be interpreted as inefficiencies or imbalances in the transport sector affecting industrial productivity negatively when labour cannot effectively handle the freight volume increases.

However, the significant positive indirect effects observed for both RailSI and RoadSI high-

light that the increased freight volume relative to the labour force may have beneficial spillover effects to nearby areas. It could involve the diffusion of economic activity through increased connectivity and availability of goods, fostering industrial growth in regions adjacent to those directly affected by transport inefficiencies. Thus, while the local transport sectors struggle with efficiency, the adjacent regions might benefit from the increased movement of goods, which supports the significant positive total effects at 1% level

In the long term, the continued negative direct effects of RailSI and RoadSI, which are stronger than in the short term, suggest that the issues of inefficiency or labour underutilization become more pronounced or remain unaddressed over time. The lack of significant indirect effects in the long term indicates that the initial beneficial spillovers do not sustain; possibly, the neighbouring regions adapt or the saturation of the transport systems limits further regional economic benefits. It can be seen as a signal that while short-term economic gains from increased freight activity can be realized despite inefficiencies. Without improvements in transport freight service efficiency, infrastructure, or freight labour force scaling, the transport system may act as a bottleneck, hindering broader economic development and the ability to capitalize on potential indirect benefits.

To sum up, the dynamics act as the critical role of transport efficiency in regional economic development. While the short-term impacts can mask underlying inefficiencies through temporary economic boosts from increased connectivity and goods availability, the long-term sustainability of industrial growth is heavily dependent on resolving these inefficiencies. Policy interventions aimed at improving freight handling efficiency, investing in transport infrastructure, and scaling the freight labour force in line with freight volumes are essential to convert short-term gains into long-term growth in industry value-added.

Table 2.5: Spatial Analysis in Long-term Effects

Synthetic Instruments	SI1	SI4	SI7	SI10
	Long-term			
Direct Effects				
RailSI	-0.601** (-2.460)	-0.219 (-0.670)	-12.739 (-0.050)	0.066 (0.420)
RoadSI	-0.475*** (-2.700)	-0.314 (-0.840)	13.060 (0.040)	-0.456 ** (-2.360)
lnGOVE	-0.972*** (-2.430)	-0.497 (-1.100)	-34.807 764.663 (-0.050)	-0.828 (-2.820)
lnPOP	-1.755*** (-2.570)	-1.126 (-1.560)	-53.016 (-0.050)	-1.527 (-3.200)
lnEDU	0.705*** (4.150)	0.690*** (3.250)	13.322 (0.050)	0.721 (4.420)
lnLabour	-0.509 (-1.510)	0.070 (0.120)	-34.830 (-0.050)	-0.321 (-1.070)
GDPgr	0.052*** (3.440)	0.053** (2.090)	1.497 (0.050)	0.051 (3.150)
Indirect Effect				
RailSI	0.156 (0.860)	-0.253 (-0.930)	12.160 (0.040)	-0.103 (-0.990)
RoadSI	0.100 (0.880)	0.171 (0.610)	-13.153 (-0.040)	-0.071 (-0.690)
lnGOVE	-0.494 (-1.080)	-0.916** (-2.280)	33.085 (0.040)	-0.646 *** (-3.770)
lnPOP	-0.708 (-0.950)	-1.303* (-1.890)	49.917 (0.040)	-0.734 (-2.150)
lnEDU	0.164 (0.840)	0.142 (0.630)	-12.212 (-0.040)	0.178 (1.310)
lnLabour	-0.407 (-1.170)	-1.111** (-2.000)	-34.830 (-0.050)	-0.549** (-2.370)
GDPgr	0.014 (0.940)	0.016 (0.650)	-1.445 (0.050)	0.009 (0.730)
Total Effect				
RailSI	-0.445*** (-3.430)	-0.471*** (-3.400)	-0.579** (-2.230)	-0.037 (-0.240)
RoadSI	-0.374*** (-3.230)	-0.143 (-0.860)	-0.093 (-0.320)	-0.527** (-2.530)
lnGOVE	-1.466*** (-7.120)	-1.414*** (-8.510)	-1.721 (-6.760)	-1.473*** (-7.680)
lnPOP	-2.463*** (-5.070)	-2.429*** (-5.600)	-3.099 (-5.270)	-2.261*** (-5.000)
lnEDU	0.869*** (6.070)	0.832*** (7.190)	1.111 (6.110)	0.899*** (6.630)***
lnLabour	-0.916*** (-3.400)	-1.041*** (-4.680)	-1.396 (-4.270)	-0.871 *** (-3.650)
GDPgr	0.065*** (4.410)	0.069*** (5.520)	0.051 (3.100)	0.060*** (4.300)

Note: Z-statistics in parentheses. ***, **, and * show the significance at of 1%, 5% and 10%, respectively.

2.6 Empirical Findings: Spillover Effects of Freight Productivity

As demonstrated in the previous section, the most spatially effective synthetic variable proves to be the most stable within the spatial model. Consequently, in this section, we shift our focus to the empirical analysis and the economic significance of using Z1 as an instrumental variable. Z1 is a matrix composed of synthetic instruments associated with both rail and road infrastructure.

2.6.1 Data and Background

To begin with the empirical background, over the past few decades, particularly following the economic reforms, China's industry has made historic advancements. Industrial added value exceeded 1 trillion yuan in 1992, 10 trillion yuan in 2007, 20 trillion yuan in 2012, and 30 trillion yuan in 2018. By 2018, China's manufacturing value accounted for more than 28% of the global total, establishing it as a crucial driver of global industrial growth⁴. According to World Bank data, the share of the manufacturing industry in GDP has gradually declined from approximately 30% at the start of the 21st century but has consistently remained above 20%, underscoring its continued importance in the economy⁵.

In parallel with China's remarkable industrial expansion, the country's railway and highway freight systems underwent extensive modernisation and expansion. These developments were not just reactive measures but strategic enhancements designed to meet the escalating

⁴The statistics come from the State Council of China, available from https://www.gov.cn/xinwen/2019-09/21/content_5431841.htm

⁵The data from World Bank, available from <https://data.worldbank.org/indicator/NV.IND.MANF.ZS?locations=CN>

demands of China’s rapidly growing economy and its logistics needs.

Driven by substantial infrastructure and industrialisation projects, railway freight volume surged from approximately 2 billion tons in 2000 to over 4 billion tons by 2019⁶. Similarly, the highway sector experienced robust growth, with extensive upgrades leading to an expansion of the road network to over 4 million kilometres by 2010. This growth in freight capacity reflects China’s strategic focus on enhancing its transport infrastructure to support and sustain industrial growth.

Amidst these substantial infrastructure advancements, the efficiency of freight services has become increasingly crucial. These improvements are pivotal in ensuring that the expanding networks can effectively support the demands of a growing economy and optimise the flow of goods across regions, thereby enhancing overall productivity and economic connectivity. Our research focuses on the spatial relationship between industrial production and freight services in rail and highway transportation across China.

2.6.2 Empirical Result

The results change significantly when spatial dependence is taken into account. This means that freight efficiency has obvious spatial effects.

The spatial lag variables, $W \times RailSI$ and $W \times RoadSI$, show a positive and significant impact on regional industry production. The coefficients for $W \times RailSI$ and $W \times RoadSI$ are 0.187 and 0.152, respectively, both significant at the 1% level. This indicates positive spillover effects from neighbouring regions’ freight productivity. In other words, improvements in rail and road freight productivity in one region benefit the industry production in surrounding

⁶Statistical bulletin on the development of the transport sector in 2020, Ministry of Transport, Retrieved from: https://www.gov.cn/xinwen/2021-05/19/content_5608523.htm

places. This could be due to enhanced connectivity and reduced transportation costs, facilitating the flow of goods and resources across regions and thus boosting production in the interconnected regions.

Control variables provide additional context to the analysis. For example, the positive and significant coefficient for education indicates that higher education levels are positively correlated with regional industrial production. This emphasises the role of human capital and skills in driving economic productivity. Conversely, variables such as government expenditure, population, and industrial labour do not show significant direct effects on regional industrial production in this model, suggesting that their impacts may be more complex or indirect. For instance, government expenditure typically influences the regional economy through various channels, such as infrastructure development, investments in education and training, and social protection. The effects of these expenditures may not be immediately reflected in short-term changes in industrial production.

In the short term, the analysis of SI1 reveals that both rail and road freight productivity exhibit negative direct effects on regional industry production. One plausible explanation for these negative direct effects is the reallocation of resources. As freight productivity improves, resources such as labour and capital might move out of the regions to take advantage of new opportunities elsewhere, temporarily reducing local industry production.

Despite the negative direct effects, the analysis uncovers positive and significant indirect effects of freight productivity on regional industry production. For *RailSI*, the coefficient is 0.190, significant at the 1% level, indicating that improvements in rail freight productivity in neighbouring regions positively impact local industry production. This positive spillover effect suggests that enhanced rail productivity leads to better connectivity and resource

flow between regions, ultimately benefiting regional industry production. Similarly, *RoadSI* has an indirect effect coefficient of 0.150, also significant at the 1% level, emphasizing that increased road freight productivity in adjacent areas boosts local production. These positive externalities underscore the importance of regional interconnectivity and the broader benefits of efficient transportation networks.

In the long term, the negative effect also could be due to the reallocation of resources, where industries might relocate to regions with better rail connectivity, resulting in a decline in local production. As road transport becomes more efficient, resources such as labour and capital may move out of the region to areas with better opportunities, causing a decrease in local industry output. For the indirect effects, these findings imply that the positive spillovers from enhanced freight productivity in adjacent areas are relatively weak in the long-term context.

Over time, as freight systems become more efficient, resources such as labour and capital may migrate to regions with better connectivity or greater economic opportunities, leading to a decline in local industry production.

Overall, the findings emphasize the need for a nuanced approach to infrastructure improvements. Policymakers should recognize the potential short-term disruptions and long-term resource reallocations that can accompany enhanced freight productivity.

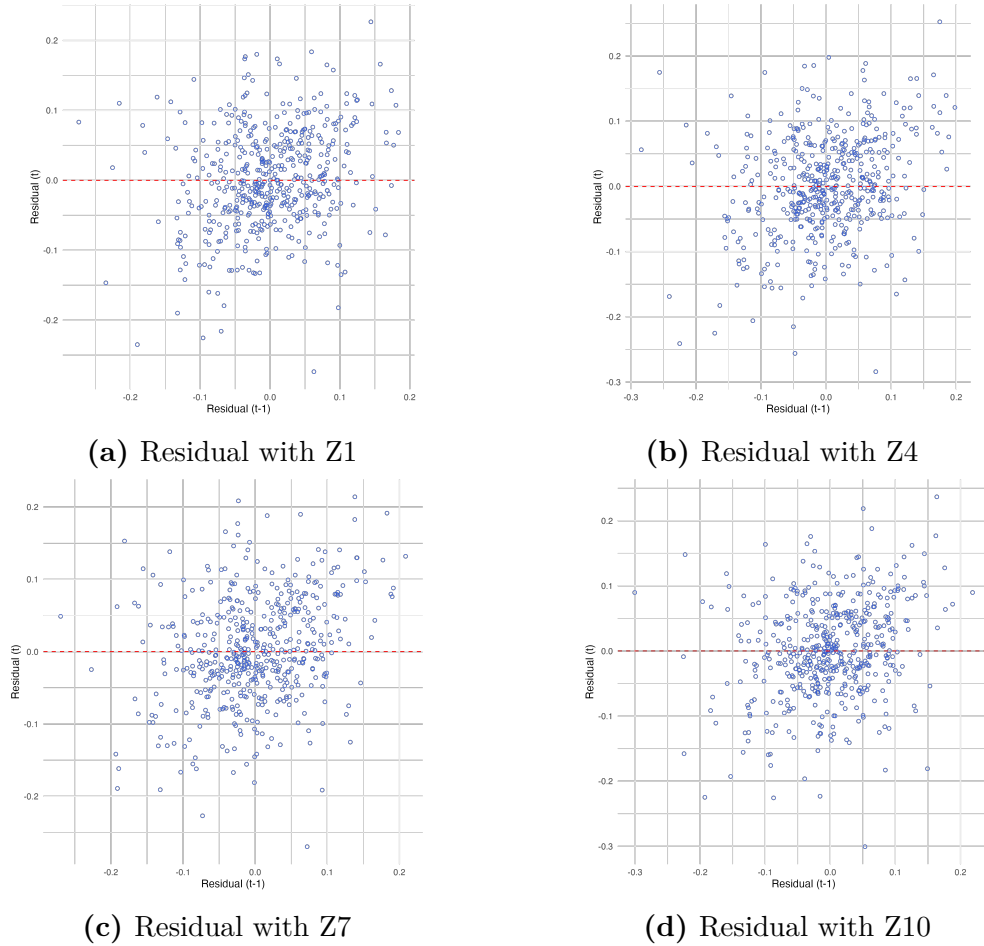
2.7 Residual Analysis

As introduced in the previous section, instrumental variables are expected to eliminate the residual bias that arises from endogenous variables. This section analyses the residuals of

the model, examining them from both time series and spatial dimensions.

Initially, we focus on the time series aspect by analysing the relationship between the residuals and their one-period lags. Figure 2.6 demonstrates the independence of the residuals, a key assumption in time series modelling. It assesses whether past values influence current residuals, which could indicate autocorrelation issues not accounted for by the model.

Figure 2.6: Residual Analysis with the Dynamic Spatial Durbin Model



Note: Z refers as a matrix for synthetic instruments. Z1 involves instruments with highest MC

The residual plots from the Dynamic Spatial Durbin Model provide insights into the rela-

tionships between the residuals and their one-period lags for different synthetic instrument practices. Z indicates an instrument matrix that contains synthetic instruments both in rail service and road service. Plot (a) shows a generally central clustering of residuals with mild dispersion, suggesting minimal autocorrelation issues under $Z1$. Plot (b) and Plot (c) reveal a wider spread in residuals, indicating greater variability and potentially less effective control of endogeneity.

The variation across these plots indicates no severe autocorrelation, suggesting well-specified model parameters. However, the differences in residual dispersion among the instruments point to varying degrees of effectiveness in capturing relevant variability and controlling for endogeneity.

On the spatial front, our goal is to address any residual spatial autocorrelation. Residual spatial autocorrelation often indicates that the model has not fully captured spatial effects, as noted by Elhorst (2014) . Furthermore, Juhl (2022) demonstrates how synthetic variables can mitigate spatial autocorrelation in the residuals. In our analysis, we compare results from ordinary least squares regression with weighted matrices, the GMM, and the SDM. We include the results in Table B.4 with the highest Moran coefficient.

We found that although synthetic instruments significantly rectify spatial autocorrelation in the residuals of GMM and OLS models, Moran’s I-statistic remains significant in the spatial Durbin model. This indicates that while synthetic instruments are effective in some contexts, the spatial Durbin model may require further adjustments or alternative approaches to adequately address spatial dependencies. Table B.4 in the appendix provides results for additional synthetic instruments, with consistent findings.

2.8 Conclusion

In this chapter, we introduced a novel geographic approach for constructing instrumental variables using exogenous geographical weight matrices. Initially, we generated eigenvectors from these matrices and selected suitable synthetic instrumental variables through a supervised spatial filter. The effectiveness of these synthetic variables was then evaluated.

Building on the research of Juhl (2022), we created multiple eigenvectors based on the inter-provincial boundary weight matrix. These eigenvectors, which match the observation count of the endogenous variables, exhibited varying degrees of spatial autocorrelation. This result is consistent with the test result of Griffith (2003) Juhl (2022), that is, the original spatial correlation of the synthetic instrumental variable needs to be consistent with the spatial trend of the original data. This method is predominantly applied to geographically oriented cross-sectional data.

Given that panel data offers a larger number of observations, few studies have explored the application of this method in panel data. Our study highlights the importance of selecting appropriate spatial filters and synthetic tools in contexts with extensive observational data, providing valuable insights for constructing and screening synthetic instruments in panel data settings.

The effectiveness of our selected synthetic instrument was demonstrated through a two-step GMM test, which confirmed that the synthetic tool significantly reduces bias induced by endogenous variables. Although increasing the number of synthetic instruments does not undermine the test, it does influence the effectiveness and efficiency of the estimations.

Our spatial analysis revealed that the choice of synthetic instrumentation is crucial for

capturing spatial effects in the dynamic SDM model. The results show that synthetic instruments with higher spatial dependencies facilitate more comprehensive and significant spatial effects. In contrast, instruments with low spatial correlation fail to capture the full spatial dynamics, underscoring the importance of selecting instruments with adequate spatial properties.

We also examined the spatial relationship between freight productivity and industrial production, marked by significant spatial effects. Improvements in rail and road freight productivity in one region positively affect industrial production in neighbouring provinces through enhanced connectivity and reduced transportation costs, demonstrating positive spillover effects. However, in the short term, direct effects suggest that increased freight productivity may lead to a temporary reduction in local industrial production due to the reallocation of resources such as labour and capital to more advantageous areas. Despite this, indirect effects indicate that neighbouring regions benefit from these productivity gains. In the long term, while positive spillovers persist, they tend to be weaker.

There are several interesting avenues for future research. Our findings suggest that, while synthetic tools effectively mitigate residual bias in time series, they do not entirely eliminate spatial autocorrelation in the residuals. This issue may arise from the heightened sensitivity of spatial models to geographic dependencies, further exacerbated by the secondary geographical weighting applied to synthetic instruments, which already incorporate geographic factors. Future research could explore ways to enhance the spatial properties of synthetic tools to address these complexities more effectively, potentially offering a more robust approach to handling spatial autocorrelation in econometric modelling.

Chapter 3

All Roads Lead to Rome: The Role of Road Networks in the Economic Resilience of Cities

Abstract

This study investigates the determinants of urban resilience, with a particular focus on the role of market accessibility in urban specialisation. We argue that effective resilience strategies should address regional transport mobility and integrate resilience planning to capitalise on the strengths of interconnected urban areas. We differentiate industrial agglomeration among cities using machine learning algorithms and construct composite weights to measure spatial dependencies and interactions between cities. Our results show that market accessibility to developed cities, particularly via road networks as determined through network analysis algorithms, can facilitate rapid economic recovery from recessions. Industrial agglomeration among cities can further amplify these positive effects and benefit surrounding regions. The significant spatial autocorrelation observed underscores the interconnectedness of urban resilience and suggests that regional cooperation and coordinated policies are essential for fostering resilience.

Keywords: Urban Resilience, Network Analysis, Market Accessibility, K-means Cluster, Spatial Autoregressive Model.

3.1 Introduction

Cities are the engines of economic growth and innovation, serving as hubs of industry, commerce, and agglomeration (Collier et al., 2018). As urban areas continue to expand and evolve, the concept of urban resilience has gained prominence in regional economics and urban planning. Urban resilience refers to the capacity of cities to withstand, adapt to, and recover from various shocks and stresses, such as economic downturns, natural disasters, and social disruptions (Sutton et al., 2023).

Does the connection to wealthier areas bring additional resources and benefits? Can these factors enhance regional resilience? We explore this question by examining the road network of China, a large-scale system that connects 288 cities across the country and covers the third-largest land area in the world.

Inspired by the research conducted by Banerjee et al. (2020), we aim to investigate the economic activity mechanisms between developed cities and their surrounding smaller, developing cities. We focus on the significant disparity between urban areas in China, particularly the stark contrast between the highly developed eastern coastal cities and the less developed central and western regions. Developed cities in the east benefit from their strategic spatial location, resulting in higher GDP per capita, better employment opportunities, and improved living standards. For instance, Shanghai’s GDP per capita was approximately 175,420 yuan in 2021, significantly higher than the national average of 81,370 yuan.¹

This disparity is also evident in the transportation infrastructure. China has established a national highway network comprising expressways, trunk roads, and rural roads. As of

¹National Bureau of Statistics of China in 2022. Retrieved from: <https://www.stats.gov.cn/sj/nd-sj/2022/indexch.htm> and <https://data.stats.gov.cn/easyquery.htm?cn=E0103>

April 2020, the total length of China's highway network reached 5.28 million km². At the city level, there are notable differences in road network density. For instance, Guangzhou, a major city in the economically advanced Guangdong province, has a road network density of 1.56 kilometres per square kilometre. In contrast, Lhasa, the major prefecture-level city of Tibet, has a road network density of only 0.18 kilometres per square kilometre³.

We explore the benefits of improving transportation infrastructure in developed regions based on China's road network, specifically in relation to regional economic resilience. By using regional population as a measure of market size and incorporating road costs based on exogenous geographical distance, we calculate the cost to market and address the issue of endogeneity. Furthermore, we examine whether cities with similar industrial structures can benefit from improved transportation conditions, drawing on the geoeconomics hypothesis of industrial agglomeration by Marshall (1920) and Krugman (1999), which suggests that companies form industrial clusters to reduce transportation costs.

Our results demonstrate that enhancing market accessibility to large cities significantly improves urban economic resilience. This impact extends beyond the local area, revealing spatial spillover effects, as the transportation network facilitates market entry for surrounding regions. Furthermore, we found that cities with lower overall development levels negatively affect the economic resilience of neighbouring cities. It suggests that underdeveloped cities may hinder the economic vitality of their region.

However, by improving market access, these underdeveloped cities can alleviate sluggish economic conditions. This phenomenon is also evident in the context of industrial agglomeration, where improved transportation infrastructure supports the formation of industrial

²The State Council of China, Retrieved from: https://english.www.gov.cn/statecouncil/ministries/202204/26/content_W

³National Bureau of Statistics of China, Retrieved from: <https://data.stats.gov.cn/easyquery.htm?cn=E0103>

clusters, thereby enhancing overall economic stability and resilience.

The following chapter is structured as follows: Section 3.2 introduces the related literature. Section 3.3 presents the data description and background information. Section 3.4 outlines the methodology for key variables. Section 3.5 constructs a dynamic spatial panel model. Section 3.6 presents our estimation results. Section 3.7 concludes and discusses future work.

3.2 Literature Review

This paper contributes to three broad strands of literature: (1) Market Accessibility in Road Infrastructure, (2) Urban Economic Resilience, and (3) Industry Agglomeration and Urban Clusters.

3.2.1 Market Accessibility with Road Network

First, our research contributes to the literature on the role of transportation. The impact of transportation infrastructure on regional economies has long been debated, particularly regarding its effects on underdeveloped areas. On the one hand, improved transportation can stimulate economic growth by providing access to opportunities and resources. On the other hand, it can also introduce greater competition and lead to resource outflows.

Yu et al. (2013) demonstrate that the impacts of transportation investment across Chinese regions are not uniform; less developed areas tend to experience adverse effects. Similarly, Faber (2014) reveals that network connectivity in peripheral counties can lead to diminished GDP growth, primarily due to a decline in industrial output. His research suggests that large-scale transportation projects, intended to link distant regions, paradoxically slow economic progress in smaller, outlying areas when compared to those that remain discon-

nected. The reduction in trade costs between major cities and smaller regions increases the latter's dependence on the former, stifling the development of local industries.

In a more recent and prominent study, Banerjee et al. (2020) examines the effects of transportation infrastructure connecting smaller towns to major cities in China. Their findings underscore that improved connectivity to large cities does not significantly enhance GDP per capita in smaller cities. While the new infrastructure might have bolstered the overall economy, its benefits were unevenly distributed. This disparity arises because the movement of people, capital, and resources—the very elements essential for maximising economic gains—was not sufficiently fluid. Consequently, although the infrastructure provided some advantages, its positive effects were constrained by the limited mobility of these crucial factors.

From a market access perspective, researchers generally view transportation improvements more positively. Donaldson (2018) quantified the positive effects of the Indian railway network on regional economies by reducing trade costs, which in turn increased economic welfare, explaining over half of the observed income gains from the expansion of railroads. Cuberes et al. (2021) provided a dynamic perspective, suggesting that while large cities initially overshadow surrounding areas, reduced transportation costs eventually promote economic growth in regions near major urban centres. Cuberes's study found that proximity to large urban clusters initially hindered growth due to urban shadows, but in the post-period, this proximity positively influenced growth, although the effect has weakened in recent decades. This shift coincided with a decrease in commuting costs.

Initially, when commuting costs were high, people tended to migrate from smaller cities to larger urban centres to be closer to their jobs and other opportunities. This migration led

to smaller cities losing both population and economic activity to the nearby larger cities. However, as commuting costs decreased over time due to improvements in transportation and infrastructure, it became easier and cheaper for people to live in smaller cities and commute to work in larger urban centres. This shift allowed smaller cities to recover and benefit from their proximity to larger urban centres.

To summarise, most studies measure transportation infrastructure based on transportation investments or static cross-sections of the existing network. Nevertheless, the stock of infrastructure is not only a key factor in economic activity but also needs to be maintained to sustain its continued impact on the economy through effective maintenance. We contribute to this literature by providing a more dynamic perspective on road choice and accessibility. By integrating the concepts of stock and flow in road infrastructure through network analysis, we map the current road network as a stock and combine it with dynamic traffic data, providing a detailed and time-sensitive understanding of transportation accessibility.

3.2.2 Urban Economic Resilience

Our paper evaluates urban economic resilience, linking it to the literature on the definition and measurement of regional economic resilience (RER).

Scholars have significantly developed the concept of RER, creating a well-structured framework where nearly every aspect of resilience is clearly defined and understood (Christopherson et al., 2010; Martin, 2012; Meerow et al., 2016; Sutton and Arku, 2022; Sutton et al., 2023). In economic research, resilience is commonly defined as the ability of regional economies to resist, adapt to, or transform in response to shocks, subsequently recovering to maintain or improve their pre-shock economic performance (Sutton et al., 2023). From an

evolutionary perspective, the theory of regional economic resilience suggests that actively restructuring and transforming an economy can help a region embark on new development paths (Simmie and Martin, 2010). Sutton et al. (2023) provide a definition of resilience that directly informs its measurement and estimation. In this study, we focus on engineering resilience, a concept frequently discussed in regional economics. This form of resilience refers to how an area responds to an economic shock, with an emphasis on the speed at which it recovers to its pre-shock state (Sutton et al., 2023).

In the empirical analysis, regional gross production is used as a key indicator of resilience. For instance, Hundt and Holtermann (2020) examine the impact of national settings on regional economic resilience in Europe, focusing on national GDP recessionary shocks from 1990 to 2014. Doran and Fingleton (2018) investigate the resilience of U.S. metropolitan areas following the 2008 financial crisis, using pre-shock regional production and employment agglomeration as indicators. They highlight how regions with higher levels of industrial diversity were able to recover more rapidly due to their inherent economic structures. Similarly, Zhang et al. (2023) incorporate expected regional production in the final year into their resilience measurements, providing insights into how provinces perform economically over extended periods.

Consequently, we build upon previous research to measure urban resilience in economic performance, with a specific focus on the role of mobility. Transport infrastructure presents a trade-off. On the negative side, road infrastructure development can initially lead to economic displacement, as businesses and residents may migrate from smaller towns to larger cities in search of better opportunities, more amenities, and improved services. This migration can weaken the economic base of smaller towns and cities, resulting in a concentration

of economic activity in larger urban areas.

On the positive side, road infrastructure supports long-term recovery and growth by enhancing connectivity between regions, reducing transportation costs, and improving access to markets, jobs, and services. Better road networks can enable smaller cities to integrate more effectively into broader regional economies, facilitating the movement of goods, people, and capital.

Furthermore, road infrastructure can influence the spread of economic shocks. Improved roads can accelerate the transmission of economic shocks across regions, as disruptions in one area may ripple through connected cities.

Conversely, efficient road networks can also facilitate faster recovery by enabling the rapid mobilisation of resources and workforce, spreading the benefits of economic rebound from larger to smaller cities. In this study, we explore the mechanisms through which road infrastructure affects urban resilience, considering both its potential to transmit economic shocks and its ability to support recovery and growth across cities of varying sizes.

3.2.3 Industry Agglomeration and Urban Cluster

Finally, we also examine the literature on industrial agglomeration and urban specialisation. Industrial specialisation refers to the concentration of specific industries within a particular urban area or region, where businesses and resources cluster around a particular sector.

This creates a focused industrial environment that benefits from shared expertise, specialised labour, and infrastructure tailored to the needs of that industry (Krugman, 1991). Such clustering enhances efficiency, innovation, and productivity, as firms in the same industry can collaborate, share knowledge, and attract skilled workers and suppliers. Over time,

industrial specialisation can strengthen regional competitiveness, making the area a hub for specific sectors. However, it can also increase vulnerability to sector-specific shocks, as regions become more dependent on the success of a single industry.

The positive relationship between firm clusters and regional economic performance can be traced back to Marshall (1890), who introduced the concept of external economies of scale. According to Marshall, firms within a specialised cluster can share resources, infrastructure, and services, reducing costs and increasing operational efficiency. Krugman (1991) further elaborated on this in economic geography, emphasising the role of increasing returns and transportation costs in forming economic agglomerations. Firms benefit from economies of scale in logistics, as coordinated efforts can streamline supply chains and distribution networks. Empirical studies, such as those by Audretsch and Feldman (1996) and Moretti (2021), demonstrate that clusters also enhance innovation and productivity through knowledge spillovers and collaboration among firms. When companies within the same industry are geographically clustered, they benefit from the exchange of ideas, technologies, and best practices (Moretti, 2021).

However, the benefits of industrial specialisation come with significant risks, particularly in terms of economic resilience. A major challenge is that heavy reliance on a single industry can make a region more susceptible to downturns in that sector over time. As research by Potter and Watts (2011) indicates, agglomeration economies develop and transform over time and across geographical space, thus forming an agglomeration life cycle. Initially, agglomeration economies can drive growth and productivity, but over time, they may lead to decreased performance as the industry matures and the negative aspects of clustering, such as the development of bottlenecks, emerge (Potter and Watts, 2011). Urban specialisation

can make regions more vulnerable to economic shocks that affect the dominant industry, reducing overall economic resilience (Frenken et al., 2007).

Therefore, our research aims to explore whether transportation improvements can have a positive impact on regions within the context of industrial agglomeration. The focus is on understanding whether enhanced transportation infrastructure will lead to increased competition or co-prosperity among cities with similar industrial structures. We will examine whether reducing transportation costs to large cities could improve the resilience of regional economies in the context of industrial agglomeration. We are particularly interested in whether the presence of transportation infrastructure fosters competition or co-prosperity in cities with similar industrial structures.

Finally, we briefly summarise the ongoing debate in the existing literature. While transportation infrastructure is often seen as a catalyst for economic development, its benefits for underdeveloped regions are not universally guaranteed. One of the main controversies revolves around the potential loss of factor resources, such as labour and capital, which may be drawn away from these regions toward more developed urban centres. This issue is particularly critical when considering economic resilience. In the face of economic shocks, the mobility facilitated by improved roads exacerbates the concentration of resources in large cities, thereby weakening the resilience of underdeveloped regions. For cities with similar industrial structures, it remains uncertain whether enhanced mobility increases competition or fosters co-prosperity.

The literature provides mixed evidence on this matter, with some studies suggesting that better connectivity can promote regional collaboration and shared growth, while others indicate that it might intensify competitive pressures, potentially disadvantaging less com-

petitive cities. Drawing on the insights from the existing literature, we propose the following research questions to guide our investigation:

- To what extent can improved market accessibility enhance regional economic resilience, particularly in underdeveloped areas?
- How does increased market accessibility in major cities impact surrounding smaller cities, and what are the implications of this spillover effect?
- In the context of industrial agglomeration, can improved market accessibility enhance the regional resilience of cities with similar industrial structures?

3.3 Background and Data

In this section, we describe the selection of regional data and provide an overview of Chinese cities.

3.3.1 Data Description

This section outlines the categories of empirical data used in our analysis. We integrate empirical data from various sources into two main categories: urban economic fundamentals and urban geographic information. The urban geographic information data includes the identification of geographical locations and the structure of transportation networks.

Our regional division is primarily based on the Administrative Division Management Regulations ⁴ and the China City Statistical Yearbook ⁵. In China, cities are categorised into

⁴The official website of the People’s Government of China, in the section of State Council documents, administrative region regulations, Retrieved from: https://www.gov.cn/zhengce/content/2018-11/01/content_5336379.htm

⁵National Bureau of Statistics. China City Statistical Yearbook Retrieved from: https://www.stats.gov.cn/zs/tjwh/tjkw/tjzl/202302/t20230220_1913734.html

three levels based on administrative divisions: provincial-level cities, prefectural-level cities, and county-level cities. To ensure greater connectivity of the transport network, our study collected data from Chinese cities at the prefectural and provincial levels.

Provincial-level cities include major provincial capitals and municipalities such as Beijing, Shanghai, Tianjin, and Chongqing, which are directly governed by the central government. These cities serve as the economic and political centres of their respective provinces and typically have more developed transportation systems and greater government funding.

Prefectural-level cities are subordinate to provinces and generally encompass multiple districts and counties, with substantial economic and population sizes. These cities are governed by provincial governments and include both urban and rural areas, often incorporating several counties or county-level cities within their jurisdiction.

County-level cities are the smallest administrative units and are governed by prefectural-level cities. Since county-level cities often include forest areas, special zones, and autonomous administrative regions with significant ethnic minority populations, some of these areas experience low accessibility and limited economic activity. Consequently, our data selection focuses on prefectural-level and provincial-level administrative regions.

This classification is primarily based on administrative hierarchy, population size, and economic development levels. Our data captures a broad spectrum of urban environments, facilitating a thorough examination of urban economic resilience and transportation metrics across various city types and sizes. Further discussions on the transport environment will be provided in subsequent sections.

Hong Kong, Macau, and Taiwan differ significantly from Mainland China in terms of policies and administration, which may affect the consistency and accuracy of the study results.

These regions have distinct transport policies, economic systems, and development levels. Hainan Province, primarily connected to the mainland by sea routes, lacks direct land transport links. This makes conventional road access assessment methods unsuitable for Hainan, potentially introducing bias into the overall evaluation of transport accessibility in China. Excluding these regions ensures dataset consistency and accuracy, avoiding potential errors from unique policy and geographic conditions.

We integrate various data sources and categories for this study. Regional economic data are sourced from the China City Statistical Yearbook for 288 cities spanning 2013 to 2021.⁶ This dataset includes city-level regional gross output, total resident population, government expenditure, and the stock of household savings in financial institutions. Geographic data are collected as randomised location points for city longitude and latitude. Road network shapefiles are obtained from OpenStreetMap for access calculations.

A brief statistical analysis is provided in Table C.1 in the appendix. The methodology section will specifically address how this data was utilised and how the index was constructed. Additionally, a summary of the statistics is presented in Table C.2 in the appendix.

3.3.2 Background of Chinese Cities

In this section, we briefly summarise the current economic development and geographic characteristics of China’s cities.

Since the late 20th century, driven by economic reforms and the transition from an agricultural economy to an industrial and service-based economy, China has experienced one of the fastest rates of urban growth in history. According to the World Bank, China’s urban

⁶National Bureau of Statistics. China City Statistical Yearbook Retrieved from: https://www.stats.gov.cn/zs/tjwh/tjkw/tjzl/202302/t20230220_1913734.html

population increased from 26% in 1990 to 61% in 2020.⁷

Initially, China's urbanisation was concentrated in a few large cities, led by Beijing, Shanghai, and Guangzhou. These cities became centres of economic activity, attracting substantial domestic and foreign investment. They rapidly developed infrastructure, expanded their industrial base, and became the primary economic drivers of their regions. The success of these cities stimulated further urbanisation, leading to the development of second- and third-tier cities across the country. According to the National Bureau of Statistics of China, the GDP of Beijing, Shanghai, and Guangzhou collectively accounted for 15% of the national GDP in 2020.⁸

Despite impressive growth, Chinese cities face several challenges. Urban areas struggle with issues related to the unequal allocation of resources between regions. The development gap between coastal and inland cities remains a significant issue, with inland cities often lagging in terms of economic opportunities and infrastructure. Data from the National Bureau of Statistics of China indicate that coastal provinces such as Guangdong and Jiangsu have GDP per capita levels more than double those of inland provinces like Gansu and Qinghai.⁹

It is evident that economically developed cities in China are not limited to provincial capitals. For example, while Guangzhou is the provincial capital of Guangdong, Shenzhen, also located within the province, has a higher economic output.

⁷World Bank. (2021). Urban population (% of total population) of China. Retrieved from: <https://data.worldbank.org/indicator/SP.URB.TOTL.IN.ZS?locations=CN>

⁸National Bureau of Statistics of China. (2021). GDP data. Retrieved from: <https://data.stats.gov.cn/easyquery.htm?cn=E0103>

⁹National Bureau of Statistics of China. (2020). Regional GDP per capita data. Retrieved from: <https://data.stats.gov.cn/easyquery.htm?cn=E0103>

Figure 3.1: Geographic Distribution of Chinese Cities



(a) Big Cities



(b) All Cities

Note: These two pictures include the geographical distribution of big cities and small cities. Among them, big cities include 81 prefecture-level cities. All cities include a total of 288 cities, both big cities and small cities.

In fact, Shenzhen’s GDP surpassed that of Guangzhou, reaching approximately 30,664.85 billion yuan in 2021, compared to Guangzhou’s 28,231.97 billion yuan, highlighting its significant economic influence.¹⁰

Recognising this, we expand the definition of large cities in our research. Based on the average production levels of cities from 2014 to 2021, we selected the top three cities in each province as economically developed large cities. These selections were made using comprehensive economic indicators, including GDP, industrial output, and foreign investment. Other cities within each province were categorised as underdeveloped areas. Figure 3.1 presents our city selection from a geographic perspective.

To conclude, the geographical distribution of Chinese cities shows a trend of high concentration in the eastern and south-eastern coastal regions, with a decrease in the number of cities in the western and northern regions. This underscores the vastness of the land and the clear regional imbalance. Our subsequent delineation of regional industrial agglomeration and the geographical distribution of the transport network is also based on these maps of city distribution.

3.4 Methodology

In this section, we introduce the three primary indices utilised in our model. The first is the industry clustering index, where we evaluate the concentration of similar industries across different cities using K-means clustering analysis. The second index is the regional economic resilience index, which we compute to assess the economic resilience of various regions. The

¹⁰National Bureau of Statistics of China. (2021). GDP data. Retrieved from: <https://data.stats.gov.cn/easyquery.htm?cn=E0103>

third index involves measuring market access through an analysis of the road network.

3.4.1 Industry Cluster Analysis

Industrial agglomeration is often reflected in the geographical clustering of similar industries. We capture this phenomenon by examining both the industry structure and geographic distribution. First, we assess how similar industries are organised and concentrated within specific regions, identifying patterns of specialisation. We then examine the physical proximity of these industries and use geographic distance to identify clusters.

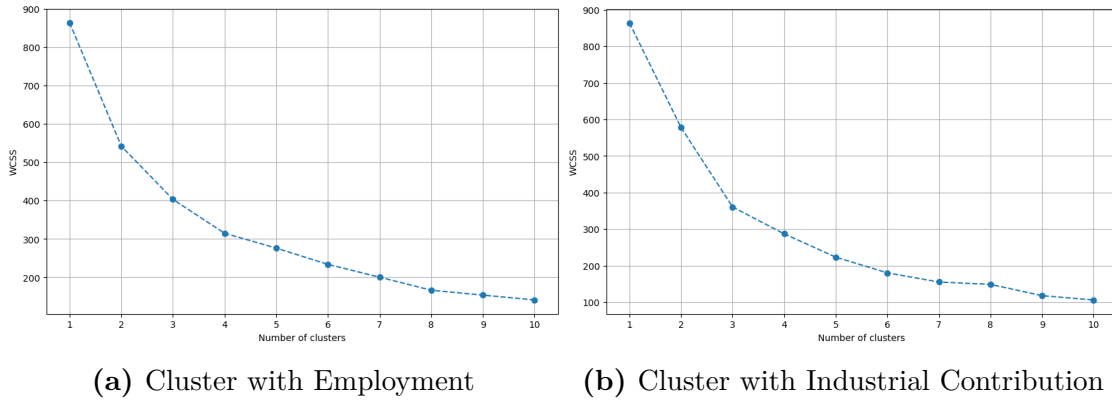
Our construction of the industrial structure is based on the classification of Chinese industries by the National Bureau of Statistics of China. These categories are derived from the National Bureau of Statistics' notice on the issuance of the Regulations on the Division of the Three Industries¹¹. In China, industries are classified into three categories. The primary industry encompasses the agricultural sector, including agriculture, forestry, animal husbandry, fisheries, and aquaculture, all primarily related to the extraction and collection of natural resources. The secondary sector focuses on manufacturing, encompassing mining, manufacturing, construction, and utilities, with core activities involving the processing of raw materials into final products. The tertiary sector, also known as the service sector, includes a wide range of services such as wholesale and retail trade, transportation, food and beverage services, financial services, real estate, education, health and social work, and culture, sports, and recreation, which provide non-material goods and services.

We primarily use employment in the three main industries of the city and the share of

¹¹The official website of the People's Government of China, in the section of State Council documents, Industrial Classification of the National Economy, Retrieved from: https://www.gov.cn/gongbao/content/2003/content_62360.htm

industrial output in total gross regional production to determine the industrial structure of each city.

Figure 3.2: Elbow Plots in Industry Cluster



We employ K-means clustering in machine learning methods. K-means clustering is one of the most widely used unsupervised learning algorithms in the field of machine learning and data mining. It aims to partition a set of observations into k clusters, where each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster (MacQueen, 1967). The algorithm iteratively refines the cluster centres, also referred to as centroids, in order to minimise the within-cluster variance, providing a straightforward and efficient method for identifying underlying structures in data (MacQueen, 1967; Lloyd, 1982). Additionally, K-means assumes spherical clusters of similar sizes, which makes it effective for identifying robust clusters (Jain, 2010).

We also use Principal Component Analysis (PCA), a statistical technique employed for dimensionality reduction while preserving as much variance as possible in the data. PCA transforms the original variables into a new set of uncorrelated variables, known as principal components, which are ordered by the amount of variance they capture (Jolliffe, 2002). Ding

and He (2004) discuss how combining PCA with K-means clustering improves clustering results and visualisation by preprocessing the data with PCA, leading to more meaningful and interpretable clusters. This reduction facilitates more efficient and accurate clustering.

To determine the optimal number of clusters for our dataset, we employed the elbow method, a widely recognised technique in cluster analysis. The elbow method involves plotting the within-cluster sum of squares against the number of clusters (k). The point at which the rate of decrease sharply slows—forming an “elbow”—is considered the inflection point, indicating the most appropriate number of clusters (Thorndike, 1953; Ketchen and Shook, 1996).

The elbow graph, shown in Figure 3.2, based on employment and production data, demonstrates a significant reduction in the Within-Cluster Sum of Squares (WCSS) as the number of clusters increases from 1 to 3. This indicates that dividing the data into more clusters up to this point results in more compact and well-defined clusters, thus better capturing the underlying structure of the employment data. However, after reaching 3 clusters, the rate of decrease in WCSS begins to slow down significantly. This suggests that while adding more clusters beyond this point continues to reduce WCSS, the improvement in clustering quality becomes marginal. Therefore, selecting 3 clusters strikes a good balance between accurately representing the data structure and maintaining simplicity.

In our analysis, we determined this inflection point to be at $k = 3$, suggesting that three clusters offer a balance between explanatory power and simplicity. The detailed plots and calculations are presented in the appendix.

Once the number of clusters was established, we proceeded with the classification of data points into these three clusters. This classification was further evaluated using the silhouette

coefficient, a measure of how similar an object is to its own cluster compared to other clusters (Rousseeuw, 1987). The silhouette coefficient ranges from -1 to 1, with higher values indicating better-defined clusters. Our results yielded a silhouette coefficient that affirmed the suitability of the three-cluster solution.

Figure 3.3: K-Means Cluster Analysis in Cities

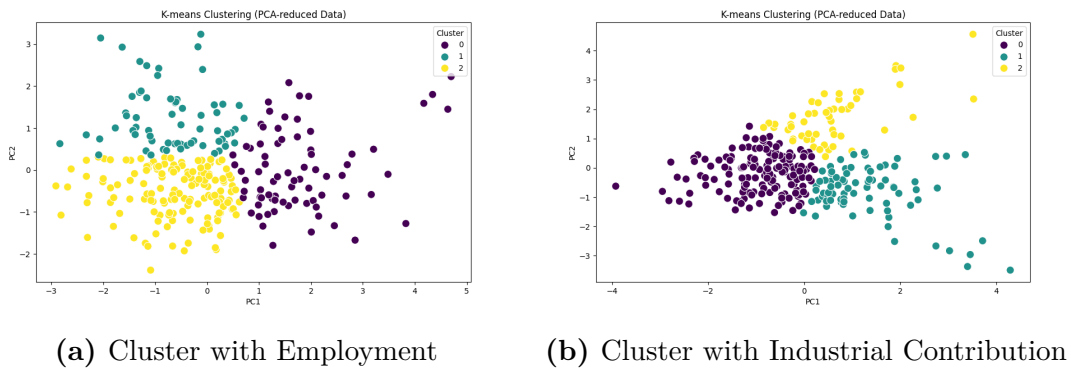


Figure 3.3 presents the cluster results. The first graph shows the outcomes of applying K-means clustering to employment data, which has been reduced to two principal components through PCA. These principal components are linear combinations of the original employment data, capturing the maximum variance in the data. PC1 reflects the overall employment size or industry scale, while PC2 captures secondary employment trends or growth rates. The clusters are visually distinguished by colours: purple for cluster 0, teal for cluster 1, and yellow for cluster 2. Cluster separation in this graph reveals three distinct groups. Cluster 0 (purple) has a wide distribution, overlapping somewhat with Cluster 1 (teal) and Cluster 2 (yellow), indicating potential commonalities or transitional industries between these clusters. The broad spread of Cluster 0 across both principal components suggests significant variability in employment characteristics within this cluster.

The second graph illustrates the K-means clustering results for industrial production data, also reduced to two principal components. Similar to the employment data graph, the principal components here reflect the main sources of variance in the industrial outcome ratio. The cluster separation shows three distinct groups, but with more compact groupings, particularly between clusters 0 and 1. This suggests less variability compared to the employment data. Cluster 0 exhibits a broad spread along PC1 but is narrow along PC2, indicating that the primary variance in production is captured well by the first principal component. This cluster likely includes industries with varying contributions but similar secondary characteristics. Cluster 2, positioned distinctly from the other two clusters, represents industries with higher contributions or different economic dynamics.

In addition to visual inspection, we evaluated the quality of the clusters using the silhouette score, a key metric for assessing the performance of clustering algorithms such as K-means. The silhouette score measures how similar an object is to its own cluster compared to other clusters, providing insights into the separation distance between clusters (Kaufman and Rousseeuw, 2009). The score ranges from -1 to 1, with higher scores indicating better-defined clusters (Rousseeuw, 1987).

The silhouette score, introduced by Rousseeuw (1987), measures the quality of clustering by comparing the average intra-cluster distance with the average nearest-cluster distance. The formula is as follows:

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}$$

where $a(i)$ is the average distance between the i -th sample and all other points in the same

cluster, and $b(i)$ is the average distance between the i -th sample and points in the nearest cluster (Rousseeuw, 1987).

According to Rousseeuw (1987), positive silhouette scores, ranging from 0 to 1, indicate that an object is well-matched to its own cluster. A score close to 1 suggests that the object is very similar to other objects in its cluster and distinct from those in other clusters, implying a well-defined cluster. Conversely, negative silhouette scores, ranging from -1 to 0, suggest that an object may be misclassified, indicating overlapping clusters or noisy data.

The silhouette score for industry GDP ratio clustering is 0.40278, while the score for employment clustering is 0.31226. Both silhouette scores are positive, confirming that the clusters are better defined than if the data points were randomly assigned. The higher silhouette score for industry GDP ratio clustering suggests a clearer separation between clusters compared to the employment clustering. This indicates that the economic structure of cities, as defined by the ratio of industry contributions to GDP, forms more distinct groups than their employment distributions.

Then, we measure the distance between the cities to their respective centroids with Euclidean distance. In the context of K-means clustering, it is used to measure the distance between each data point (city) and the cluster centroids. The Euclidean distance between two points $\mathbf{x} = (x_1, x_2, \dots, x_n)$ and $\mathbf{y} = (y_1, y_2, \dots, y_n)$ in 2-dimensional space is given in the Appendix C, as section C.2.

As a result, we construct spatial weight matrices that account for both the similarity of industrial structures between cities and their geographical proximity, resulting in composite weights. The similarity of industrial structures is determined by comparing the types and proportions of industries present in each city, with cities exhibiting more similar industrial

structures receiving higher similarity scores. Additionally, we incorporate the geographical proximity of the cities, with closer cities receiving higher proximity scores. These two factors are combined to form the composite weight, where a higher composite weight indicates that the two cities are both industrially similar and geographically close. This means that if City A and City B have very similar industrial profiles and are geographically close, their composite weight will be high, reflecting a strong linkage in our spatial weight matrix. Conversely, if City A and City B have different industrial structures and are geographically distant, their composite weight will be low, indicating a weaker linkage.

3.4.2 Regional Economic Resilience

As highlighted by Sutton et al. (2023), regional economic resilience refers to a region's ability to withstand, adapt to, and recover from economic shocks while maintaining or enhancing its long-term development potential.

This study adopts the resilience calculation methodology proposed by Doran and Fingleton (2018) and Zhang et al. (2023), which measures economic resilience by calculating the ratio of actual to expected variance in gross regional production for the region from 2014 to 2021. The analysis spans the pandemic period in 2020 and includes a recovery year in 2021.

The resilience of region i is calculated as follows. Where Δy_i^{t+k} is the actual production change in region i during the $t+k$ period, and $(\Delta y_i^{t+k})^{\text{expected}}$ is the expected gross production change in region i during the $t+k$ period. National gross domestic production is as y_n^{t+k} . The expected production change is based on the national gross domestic production growth rate:

$$(\Delta res_i^{t+k})^{\text{exp}} = y_i^t * (\Delta y_n^{t+k}) \quad (3.1)$$

The expected change in region i during the $t + k$ period is calculated as:

$$Res_i = \frac{(\Delta res_i^{t+k}) - (\Delta res_i^{t+k})^{\text{exp}}}{|(\Delta res_i^{t+k})^{\text{exp}}|} \quad (3.2)$$

We calculate economic resilience for a region or entity over a specific time period, from t to $t + k$. The resilience score, as Res_i , is based on the difference between the observed change in an economic indicator of urban production and the expected change over that period.

The observed change, as Δres_i^{t+k} , reflects the actual change in the economic condition of the region during the time period. On the other hand, the expected change, as $(\Delta res_i^{t+k})^{\text{exp}}$, is what would be anticipated based on national trends.

The interpretation of the resilience score is straightforward. If Res_i is greater than 0, it means the region performed better than expected, showing strong resilience. If the score is 0, the region performed as expected. However, if Res_i is negative, it indicates that the region performed worse than expected, showing lower resilience.

To further investigate the relationship between gross production and resilience, we provide a linear fit view of the data. The scatter plot, shown in Figure 3.4, reveals a positive correlation between regional GDP growth rate and economic resilience. The red regression line in the plot highlights this positive trend, indicating that regions with higher GDP growth rates tend to exhibit greater economic resilience.

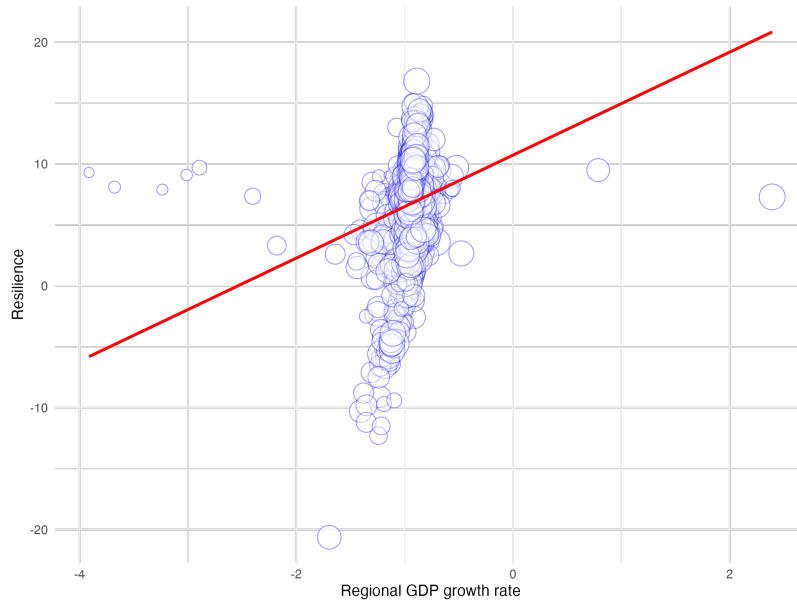


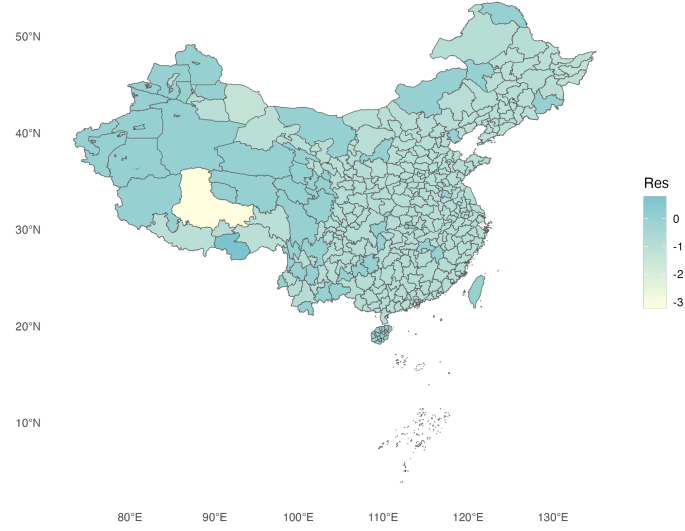
Figure 3.4: Comparison of Resilience and Regional GDP Growth Rate

Notes: The x-axis represents resilience, while the y-axis indicates the regional GDP growth rate. The size and colour of the circles reflect the actual GDP, after applying the logarithmic transformation.

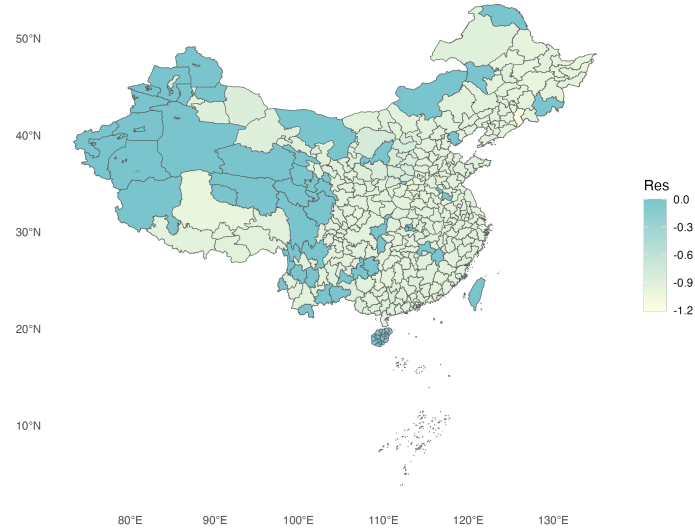
The scatter plot also reveals a central cluster of data points, indicating that the majority of regions exhibit similar GDP growth rates and resilience levels. This clustering suggests that while there is a general trend of higher resilience corresponding to higher growth rates, many regions fall within a relatively narrow range for both metrics. The size of the bubbles, representing the logarithm of GDP, reflects the variability in the economic scale of different regions.

Furthermore, the relative uniformity in bubble sizes around the central cluster implies that most regions have similar overall economic sizes, reinforcing the notion that regions with comparable economic scales tend to display similar levels of growth and resilience.

Figure 3.5: Geographic Distribution of Resilience



(a) 2014



(b) 2021

Note: The map illustrates a geographic distribution of resilience.

We also discuss the geographical cross-sectional distribution of economic resilience. Figure 3.5 presents a series of maps illustrating the geographic distribution of economic resilience

across 288 cities for the years 2014 and 2021. A colour gradient is used in each map to depict resilience levels, with lighter shades representing lower resilience and darker shades indicating higher resilience.

It is evident that the resilience of most cities is negative relative to the national GDP. The 2014 map displays a relatively uniform distribution of economic resilience across the country. The overall light to medium shades of blue suggest that most cities exhibited moderate levels of resilience, with a few areas standing out for their higher stability.

The 2021 map highlights even more pronounced differences in economic resilience across China. Several areas in the central and southern regions show increased resilience, with darker colours indicating stronger economic stability and recovery. However, following the pandemic in 2020, the gap in resilience has widened significantly, with other cities showing noticeably lower resilience.

3.4.3 Access to Big Markets

We define market accessibility following the research by Geurs and van Wee (2004), which emphasizes that accessibility to opportunities is influenced by both the associated travel costs and the availability of opportunities. The distribution of such opportunities and the associated travel costs usually involve the number of opportunities or the potential purchasing power contained in an area, while the travel costs include the time or financial costs required to reach the area.

Research by Evenett and Keller (2002) indicates that the traditional gravity model is frequently applied in the study of trade dynamics between regions and nations. Similarly, in the context of trade, the gravity model suggests that larger populations generate more

significant economic interactions due to greater demand and supply capabilities.

The interaction between two locations is directly related to their respective sizes, often quantified by population (Lall et al., 2004). Building on this foundation, they construct an accessibility index measure based on population-weighted access to measure the Indian industrial market.

Moreover, the gravity model accounts for the distance between trading partners, with greater distances generally reducing the likelihood and volume of trade due to increased transportation costs and time. However, advancements in technology and transportation infrastructure can mitigate these distance effects, so we use network analysis to measure the actual travel cost between two places.

Network Analysis

In this study, we define travel cost as the time required to travel from one city to any point in another city. We refer to Khanna et al. (2022) for detailed calculations and methodologies applied in network analysis, particularly with respect to the role of roads in shaping cities. This framework enables us to capture traffic paths within and between urban areas.

For intra-city journeys, we account for the complexity of navigating dense street networks and the potential for bottlenecks. Our analysis focuses on route selection after traffic restrictions, particularly on inner-city roads with one-way lanes.

We obtain road network data from OpenStreetMap for the years 2014 to 2021 and use Geographic Information Systems (GIS) to calculate the network costs between cities. Specifically, we consider all 288 cities as starting points and calculate the travel costs required for each city to reach various destinations. This dataset includes nodes, such as intersections,

junctions, and edges, representing road segments and pathways. Each edge in the road network has attributes such as length, travel time, speed limit, and road type.

We work with network analysis tools in ArcGIS and QGIS to calculate both the shortest road and fastest routes. These tools enable us to calculate the most efficient paths based on various criteria, such as distance and travel time, while incorporating real-world factors like road types, speed limits, and traffic conditions. In this study, we standardise the speed limits to enable the calculation of the shortest path based on travel time and road speed limits.

The calculation considers different costs. The total cost is the sum of the entry cost, network cost, and exit cost for each origin-destination pair in the road network. The entry cost is the cost of starting a trip from an origin point, the network cost is the cost of travelling along the road segments between the origin and destination, and the exit cost is the cost of reaching the destination point. We consider the maximum speed on different roads to constrain the road selection.

Start Points

Specifically, we employ all 288 cities as starting points and calculate the travel costs required for each of these cities to reach various destinations. To ensure a more realistic and valid analysis, we select random points within each city as starting locations instead of using a single central point. This method is chosen to capture the true variability and complexity of urban areas.

Cities are not homogeneous entities; they consist of diverse neighbourhoods, each with different levels of accessibility and infrastructure. By selecting random points, we account for these variations and avoid the bias that might result from only considering central locations,

which often have better infrastructure and accessibility than other parts of the city.

Furthermore, in selecting these random points, we carefully avoid locations that are inaccessible, such as bodies of water, mountains, and other obstacles that would not be practical starting points for travel. This step ensures that our analysis reflects realistic travel scenarios and avoids skewed results that could arise from including points where travel is not feasible.

Destination Points

In our background analysis of Chinese cities, we select developed cities in each province based on their economic performance. The destinations are defined as the developing cities, specifically the top three cities by average urban gross production in each province.

Unlike previous studies, we do not lock in the choice of large cities within provinces. This is because some cities located on the provincial boundary lines do not prioritize their own provincial megacities in terms of trade flows. As mentioned earlier, there is a clear geographical division in China's urban development. The areas closer to the coast have more developed cities. Based on Evenett and Keller (2002) theory of gravity, we similarly argue that more developed cities attract trade partners across provincial borders. Therefore, we choose all developed cities as endpoints.

Road Categories

We consider including motorways, primary roads, secondary roads, and tertiary roads. The road shapefiles provide the levels of the road. According to OpenStreetMap, the Road Traffic Safety Law of the People's Republic of China¹² and the Highway Law of the People's

¹²Transportation Department, 2005. Retrieved from: https://www.gov.cn/banshi/2005-05/26/content_996.htm.

Republic of China¹³, China's highways are divided into five technical levels: expressways, primary highways, secondary highways, third-class highways, and fourth-class highways. Different levels of highways have distinct technical indicators, such as design speed, number of lanes, traffic volume, and pavement structure.

Based on administrative levels, these are classified into national highways, provincial highways, county highways, township highways, and special-purpose highways. National highways include expressways and ordinary national highways, while provincial highways include provincial-level expressways and ordinary provincial highways. National and provincial highways are categorized as trunk roads, whereas county and township highways are considered branch roads. We provide a description of the road types in our analysis. Below is a description of the major road types in our analysis.

- **Motorways:** Motorways, also known as national highways, are the highest classification of roads and are designed for fast, long-distance travel. They are typically multi-lane roads with limited access points, allowing vehicles to travel at high speeds without interruptions from traffic signals or intersections. Motorways are often referred to as highways or free ways in different regions and are crucial for connecting major cities and regions.
- **Primary roads:** Primary roads are the main roads that connect important towns and cities. They are usually single or dual carriageways and can accommodate significant traffic volumes. Primary roads are essential for regional connectivity and often serve as major arteries within the road network, facilitating the movement of goods and people between urban centres.
- **Secondary roads:** Secondary roads serve as connectors between primary roads and

¹³National Standardization Administration, 2015, Available from: <https://www.sac.gov.cn/sacen/index.html>

smaller towns or rural areas. While they are not as heavily trafficked as primary roads, they are still important for regional transportation. Secondary roads typically have fewer lanes and lower speed limits compared to motorways and primary roads, but they play a vital role in local accessibility and mobility.

- **Tertiary roads:** Tertiary roads are local roads that provide access within towns, cities, and rural areas. These roads are generally narrower and have lower speed limits. They facilitate local traffic, connecting residential areas, small businesses, and local amenities to the broader road network. Tertiary roads are crucial for daily commutes and local transportation needs.
- **Impact on Network Costs and Accessibility:** By incorporating these different types of roads into our analysis, we are able to calculate accurately the network costs between cities and assess the impact of road infrastructure on accessibility. Each type of road contributes differently to the overall connectivity and travel efficiency within the network, influencing the accessibility of various regions.

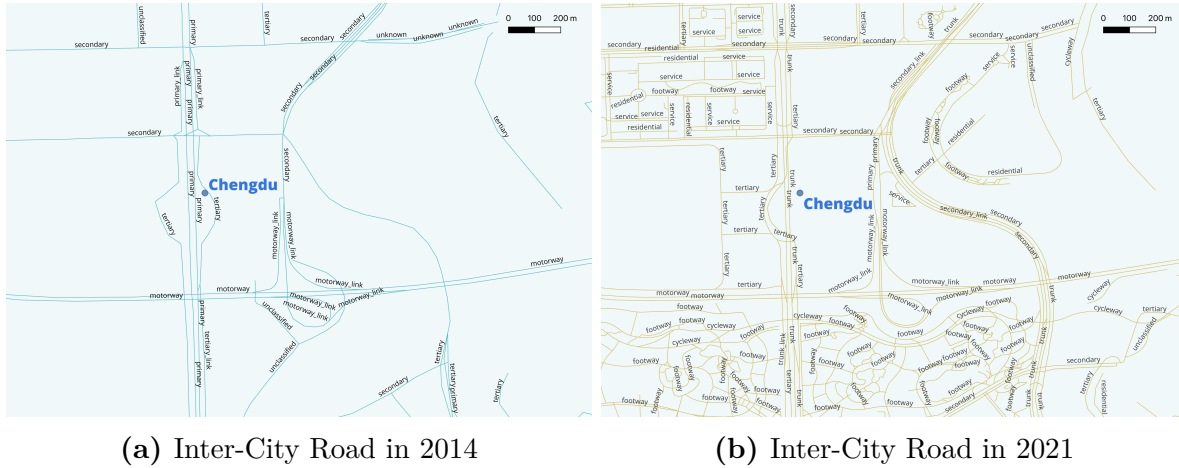
By incorporating these different types of roads into our analysis, we are able to accurately calculate the network costs between cities and assess the impact of road infrastructure on accessibility. Each road type contributes differently to the overall connectivity and travel efficiency within the network, thereby influencing the accessibility of various regions.

Intra-City and Inter-City Road Network: Chengdu

In this section, we compare the development of intra-city roads (Figure 3.6) with that of inter-city roads (Figure 3.7). To better visualize the categories of roads and the specific changes, we chose Chengdu as a sample for analysis. The intra-city roads refer to those in closer proximity to the city centre, while the inter-city roads are partly derived from those

crossing the city boundary line between Chengdu and Meishan.

Figure 3.6: Intra-City Road Network



Note: The map illustrates a part of inter-city road development in Chengdu.

Figure 3.6 illustrates the types and development of inner-city roads near Chengdu’s city centre. In 2014, the inner-city road network was primarily composed of primary, secondary, and tertiary roads, along with motorways and their respective links. Primary roads functioned as the main arteries of the city, connecting key areas and supporting significant traffic flow. Secondary roads complemented the primary network by providing critical connections and facilitating smoother traffic distribution. Tertiary roads linked local areas, residential neighbourhoods, and commercial zones to the broader road network. The presence of motorways and motorway links highlighted the level of road infrastructure development. Additionally, unclassified roads were present, reflecting areas with less structured or less developed planning.

By 2021, Chengdu’s inner-city road network had undergone significant development and expansion. The road types had diversified and increased in number, with the addition of new categories such as trunk roads, service roads, and footways. Primary and secondary

roads had expanded, indicating enhanced connectivity and urban growth. The introduction of trunk roads supported high-capacity traffic flow, reflecting substantial infrastructure investment. The expansion of tertiary roads, particularly in residential areas, suggested ongoing urban development and systematic planning. Residential and service roads had become more defined, and the networks of footways and cycleways had expanded significantly, promoting non-motorised transport and enhancing pedestrian and cyclist accessibility. The motorway network had also been extended, ensuring improved connectivity across the city.

Figure 3.7: Inter-City Road Network



Note: The map illustrates a part of intra-city road between Chengdu and Meishan. The grey line is the boundary between the two cities

The two inter-city road network maps between Chengdu and Meishan from 2014 and 2021 reveal significant changes in the region's transportation infrastructure over these years. The grey line represents the boundary between Chengdu and Meishan. In 2014, the road network was characterised by a limited number of primary roads, which served as major arterial routes. These primary roads were concentrated in central areas, reflecting the nascent stage of major route development. Secondary roads, which connect primary roads and facilitate

significant traffic flow, were moderately distributed, enhancing connectivity within the urban core. Tertiary roads were prevalent on the city's periphery, indicating the early stages of urban expansion. Additionally, unclassified roads were scattered throughout the network, reflecting areas with less structured road planning.

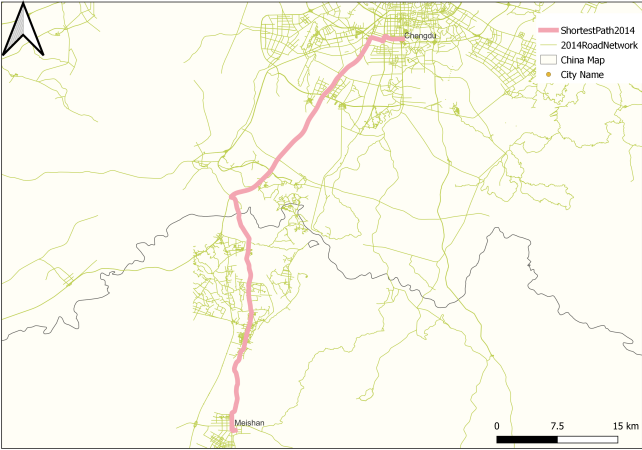
By 2021, Chengdu's road network exhibited substantial development and expansion. Trunk roads had been introduced, providing high-capacity urban routes. The number of primary roads had increased, connecting more parts of the city and signalling robust urban growth. Secondary roads had also expanded, further enhancing connectivity within the network. The growth of tertiary roads, particularly in residential areas, pointed to continued urban sprawl and the development of new housing zones. The addition of residential roads underscored the development of new neighbourhoods to support the city's growing scale. Compared to 2014, the 2021 network was denser and more interconnected, reflecting improved urban planning and strategic infrastructure investment.

Overall, the evolution of Chengdu's road networks from 2014 to 2021 highlights a period of robust urban and regional growth. Both intra-city and inter-city networks have been enhanced through the introduction of new road types and a focus on sustainable transport solutions.

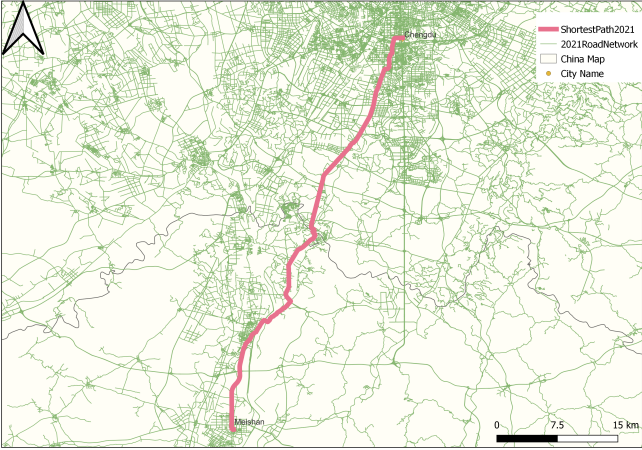
Shortest Path Across Cities: Meishan to Chengdu

Given the vast scale of the entire road network, we selected a fixed geographic area as a reference to illustrate the road distribution in China from 2014 to 2021. Figure 3.8 provides a clear and consistent comparison over time. Our analysis reveals that the road network in this reference area has become significantly more complex in 2021 compared to 2014. This development not only constructs more roads but also provides more travel options.

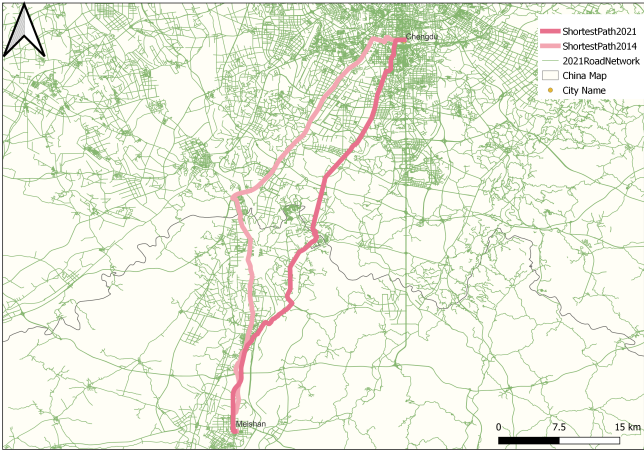
Figure 3.8: Road Network Analysis from 2014 to 2021



(a) 2014



(b) 2021



178
(c) 2014-2021

Note: The map illustrates the shortest pathway from Chengdu to Meishan. We compare the network analysis result from 2014 to 2021

The road network around Chengdu and Meishan underwent significant development between 2014 and 2021. The 2021 map illustrates a much denser network of roads, particularly around Chengdu, highlighting substantial urban and suburban growth. This expansion has increased the availability of route options, which, in turn, has influenced the determination of the shortest path between the two cities. The shift in the shortest path from 2014 to 2021 may reflect changes in the criteria used to identify the most efficient route.

3.5 Dynamic Spatial Model

This section provides a detailed discussion of how the above data was utilised to construct a dynamic spatial model. The process began with determining the spatial dependencies between cities, involving an analysis of how the locations and distances between cities influence various socio-economic and infrastructural interactions. Identifying spatial dependencies allows for the detection of patterns of connectivity, economic relationships, and potential areas for infrastructural development.

We employed Moran's I estimation to measure spatial autocorrelation, indicating the extent to which similar values occur in close proximity in space. Following this, we constructed a dynamic spatial autoregressive model using spatial weights to account for both temporal and spatial interdependencies between cities. This model captures how changes in one city influence neighbouring cities over time, providing insights into the spatial and temporal dynamics of urban and regional development.

3.5.1 Weight Matrices and Spatial Autocorrelation

We constructed three spatial weights, all based on geographic distance, to capture different dimensions of spatial dependencies. The first type is the inverse distance weight, which considers only the geographical proximity between cities, assigning greater influence to nearer cities while diminishing the influence as distance increases.

The second type comprises two composite weights that incorporate industrial aggregation. These weights are designed to reflect not only geographic proximity but also economic similarities between cities. To create these composite weights, we utilised K-means clustering as described in the previous section, grouping cities based on their industrial characteristics. These weights were constructed by combining the K-means industrial clusters with two specific economic indicators: employment share and industrial production share. By incorporating these economic variables, the spatial weights account for both physical geographic distance and economic interdependencies between cities.

Inverse Distance Weight Matrix

Inverse Distance Weighting (IDW) is a spatial interpolation technique frequently used in geostatistics and spatial analysis. The IDW method assigns weights to nearby observations based on their distance from the point of interest, with higher weights given to closer observations and lower weights to those farther away (Shepard, 1968). The weight matrix, W , for inverse distance weighting can be expressed as:

$$W_{ij} = \begin{cases} \frac{1}{d_{ij}}, & \text{if } i \neq j \text{ and } d_{ij} \neq 0 \\ 0, & \text{if } i = j \end{cases} \quad (3.3)$$

In this equation, W_{ij} represents the weight between the i -th and j -th locations, and d_{ij} denotes the distance between these two locations. The weight W_{ij} is inversely proportional to the distance d_{ij} , meaning that as the distance increases, the weight decreases. d_{ij} is obtained by calculating the longitude and latitude distance between city i and city j .

This ensures that closer observations have a more significant impact on the interpolated value than those further away. If i equals j or the distance d_{ij} is zero, the weight is set to zero to avoid singularity and to ensure that a location does not contribute to its own weight in the matrix.

Economic Cluster Weight Matrix

In the previous sections, we estimated industry agglomeration using K-means cluster analysis. To measure the distance between industry similarities, we constructed two matrices. Here, we recall these matrices and apply an inverse transformation to ensure that closer locations have higher weights. We then combine the industry similarity matrices, denoted as w_{ij} , with inverse distance matrices to build two weighted cluster weight matrices as follows:

$$W_{ij} = \begin{cases} w_{ij} \frac{1}{d_{ij}}, & \text{if } i \neq j \text{ and } d_{ij} \neq 0 \\ 0, & \text{if } i = j \end{cases} \quad (3.4)$$

Spatial Autocorrelation

We analysed spatial autocorrelation in conjunction with spatial weights to evaluate the extent to which similar values in our dataset are spatially clustered or dispersed. Spatial autocorrelation quantifies the degree to which the presence of an attribute at one location correlates with its presence in neighbouring locations.

To estimate spatial autocorrelation, we employed Global Moran's I, a widely used statistic in spatial analysis. Global Moran's I measures spatial dependence among spatial units and is mathematically defined to compare the value of a variable at each location with its values at other locations, weighted by a spatial weights matrix. Positive values of Moran's I indicate positive spatial autocorrelation, where similar values tend to cluster, whereas negative values suggest negative spatial autocorrelation, indicating a dispersed spatial pattern.

To test the significance of the observed Moran's I, we conducted a two-tailed test. This involved calculating the expected value and variance of Moran's I under the null hypothesis of no spatial autocorrelation. By standardising the observed Moran's I into a Z-score and comparing it against a standard normal distribution, we assessed whether the spatial autocorrelation was statistically significant. In a two-tailed test, if the absolute value of the Z-score exceeds the critical value, we reject the null hypothesis, confirming significant spatial autocorrelation (Cliff and Ord, 1981; Anselin, 1995).

The results are summarised in Table 3.1, which presents Moran's I values alongside their corresponding Z-scores and significance levels. These results span the period from 2014 to 2021, using three types of spatial weights: Inverse Distance, Industrial Production Cluster, and Industrial Employment Cluster.

The IDW matrix primarily captures the effect of geographical proximity on spatial autocorrelation. In 2014, Moran's I was -0.003, with a Z-score of 0.096, indicating no significant spatial autocorrelation. However, beginning in 2015, there was a noticeable increase in positive spatial autocorrelation, with Moran's I values turning positive and significant. For instance, in 2015, Moran's I was 0.016, with a Z-score of 4.272, reflecting strong spatial clustering at the one percent significance level. This trend continued and intensified over

Table 3.1: Global Moran's I

Weight Matrix Year	IDW	Industrial Production	Industrial Employment
2014	-0.003 (0.096)	0.002 (0.574)	-0.009 (-1.006)
2015	0.016*** (4.272)	0.009 (1.422)	0.014 *** (3.247)
2016	0.035*** (11.648)	0.039 *** (6.970)	0.035 *** (9.699)
2017	0.009*** (3.520)	0.008* (1.660)	0.008** (2.554)
2018	0.006 ** (2.338)	0.003 (0.830)	0.003 (1.350)
2019	0.119 *** (19.532)	0.123*** (11.045)	0.108 *** (15.180)
2020	0.042*** (7.572)	0.051*** (4.911)	0.057 *** (8.579)
2021	0.074*** (12.477)	0.063*** (5.894)	0.065*** (9.422)
Note: Z-value in parentheses. ***, **, and * show the significance at 1%, 5%, and 10%, respectively..			

subsequent years, peaking in 2019 with Moran's I at 0.119, demonstrating very strong positive spatial autocorrelation. Significant values persisted through 2020, when Moran's I was 0.042, and 2021, when Moran's I was 0.074. This pattern suggests that cities in China are increasingly clustering geographically, likely as a result of enhanced connectivity and regional development strategies.

The industrial production cluster weight matrix incorporates economic factors, specifically industrial GDP output, alongside geographical proximity. In 2014, Moran's I was 0.002, indicating no significant spatial autocorrelation. From 2015 onwards, there were intermittent periods of significant spatial autocorrelation. For instance, in 2016, Moran's I reached 0.039, indicating significant clustering at the one percent level. This significance persisted into 2019,

when Moran's I reached 0.123, reflecting strong positive spatial autocorrelation.

These findings indicate that cities with similar industrial structures tend to cluster geographically. This clustering reflects economic interdependence and industrial linkages between these cities. Cities sharing similar industries often develop close ties through trade relationships, technology spillovers, and labour market mobility. For instance, industries in these clustered cities may benefit from more efficient goods exchange, shared technological advancements, and access to a common pool of skilled labour.

However, during the period from 2017 to 2018, the analysis reveals weak or insignificant spatial autocorrelation, reflecting the degree to which economic activity in one city is correlated with that in nearby cities. This change may indicate several underlying shifts in regional economies. One possible explanation is industrial relocation, where businesses move operations from one city to another, potentially driven by lower costs or improved infrastructure in alternative locations. Another explanation could involve industrial transformation, where cities transition from traditional industries to new sectors, resulting in changes to the geographic distribution of industries.

Although the physical locations of cities remain fixed, their industrial structures and economic performance can evolve over time, leading to shifts in the spatial distribution of industries. For instance, a city may diversify its industrial base, transition from manufacturing to service industries, or emerge as a technology hub, thereby altering its role in regional industrial clustering. Such changes in industrial composition can weaken previously established industrial linkages between cities, resulting in reduced spatial autocorrelation and altered patterns of economic interdependence.

The industrial employment cluster weight matrix combines geographical proximity with em-

ployment data in industries and reflects how employment patterns influence spatial clustering. In 2014, Moran's I was -0.009, indicating no significant spatial autocorrelation. Similar to the other weight matrices, 2015 marked the onset of significant positive spatial autocorrelation, with Moran's I reaching 0.014 and indicating significance at the one percent level. This positive trend continued, peaking in 2019 with Moran's I at 0.108, demonstrating very strong spatial clustering. The results for 2020, with Moran's I at 0.057, and for 2021, with Moran's I at 0.065, indicate continued significant positive spatial autocorrelation.

This weight matrix, which focuses on the industrial employment structure, suggests that cities with similar distributions of workers across industries are increasingly clustering together. Cities with comparable employment structures tend to have workforces concentrated in the same industries, such as manufacturing, technology, or services. Consequently, these cities are more likely to develop similar industrial profiles, as their workforce demands and industrial needs align.

The clustering of cities with similar employment structures indicates that they may benefit from shared labour market dynamics, such as common pools of skilled workers and a similar demand for industry-specific talent. Moreover, this clustering reflects the specialisation of cities in particular industries. As certain cities concentrate their workforce in specific sectors, they develop specialised labour markets, attract further investment in those industries, and promote regional economic integration.

Similar employment structures across cities can also result in interdependent labour markets. When one city experiences economic growth, it can drive increased demand for workers in neighbouring cities with similar industries. Conversely, economic downturns may have ripple effects across these cities, disrupting their interconnected labour market dynamics.

Overall, the inverse distance weight matrix shows a consistent increase in geographical clustering over the years, highlighting growing spatial interdependencies driven by physical proximity. The industrial production cluster weight matrix reveals significant economic clustering based on industrial production, particularly in the later years, reflecting the economic integration of cities. Similarly, the industrial employment cluster weight matrix demonstrates increasing clustering based on industrial employment, suggesting shared labour market characteristics and economic activities. These results underscore the dynamic nature of spatial autocorrelation, influenced by both geographical proximity and economic factors. The increasing positive spatial autocorrelation over time indicates that cities are becoming more interconnected, both physically and economically, fostering regional development and integration.

3.5.2 Model Setting

This section introduces a dynamic Spatial Autoregression (SAR) model that incorporates three weight matrices. The SAR model is characterised by the inclusion of a spatial lag of the dependent variable, which captures the influence of neighbouring observations. This spatial modelling approach effectively accounts for spillover effects and interdependencies commonly observed in spatial data (Anselin, 1988). The SAR model is particularly advantageous in contexts where spatial dependence in the dependent variable is hypothesised or empirically observed.

We follow the dynamic SAR model as described by Baltagi et al. (2019). The model equation is given by:

$$Y_{it} = \alpha Y_{it-1} + \rho_1 \sum_{j=1}^N w_{ij} Y_{jt} + X_{it} \beta + \epsilon_{it} \quad (3.5)$$

In the context of a dynamic spatial autoregressive data model, Y_{it} represents the resilience for region i at time t , while X_{it} is a vector of explanatory variables. The parameters α and β correspond to the dimensions 1×1 and $K \times 1$, respectively, and are to be estimated. The element w_{ij} in the matrix W_N represents the spatial weight between regions i and j . The spatial weights matrix W_N is a known $N \times N$ matrix with zero diagonal elements and is row-normalized.

The spatial lag coefficient, ρ_1 , is a critical parameter and is constrained within the interval $[1/\alpha_{\min}, 1]$, where $\alpha_{\min}$ denotes the most negative purely real characteristic root of the row-normalized W_N . This constraint ensures spatial stationarity. For the model to remain stationary across space, it is essential that the stationarity assumption $|\alpha| < 1$ holds.

To mitigate the risk of omitted variable bias in our analysis, we include a range of regional characteristics as control variables. Specifically, we incorporate variables such as population density, education level, and household savings in financial institutions as a proxy for the local financial level. Population density is included as it influences both the demand for and the development of road infrastructure. Education level is considered because it affects the availability of human capital in a region, which in turn can impact economic development and infrastructure utilisation. The local financial level is also controlled, as it reflects the economic resources available for infrastructure investment and maintenance.

3.6 Empirical Finding

3.6.1 Main Result

The analysis presented in Table 3.2 employs a dynamic spatial autoregressive model to examine the impact of market access, population density, financial savings, and educational expenditures on urban resilience. The model incorporates three distinct spatial weight matrices, as previously discussed, to capture spatial dependencies effectively.

Table 3.2: Main Results with Dynamic Spatial Autoregressive Model

Weight Matrix	IDW	Industrial Production	Industrial Employment
Main Estimate			
L.Res	-0.119*** (0.021)	-0.117 *** (0.020)	-0.114 *** (0.021)
lnMarketAcc	0.023*** (0.008)	0.022 *** (0.008)	0.023 *** (0.008)
lnPopDensity	-0.002 (0.004)	-0.002 (0.004)	-0.001 (0.004)
lnFinancesav	-0.002 (0.006)	-0.002 (0.006)	-0.004 (0.006)
lnEduExp	0.005 (0.009)	0.006 (0.009)	0.009 (0.009)
Spatial Autocorrelation (rho)	1.980*** (0.1680)	57.047 *** (6.105)	426.009 *** (117.033)

Note: standard errors in parentheses. ***, **, and * show the significance at 1%, 5%, and 10%, respectively..

Using the Inverse Distance Weight matrix, the model shows that the lagged resilience, for L.Res, as Res(t-1), has a significant negative coefficient, as -0.119, at the 1% significance level, indicating a mean reversion effect where cities with past negative resilience tend to experience lower current resilience.

However, the positive and significant coefficient for market accessibility to big cities, as logMarketAcc, at 0.023 implies that better connectivity and access to larger markets signif-

icantly enhance urban economic resilience. However, population density, financial savings, and educational expenditure do not show significant impacts on resilience. The spatial autocorrelation coefficient (ρ) is 1.980, highly significant, indicating strong spatial dependencies and suggesting that a city's resilience is heavily influenced by the resilience of neighbouring cities.

When using the Industrial Production weight matrix, the results are similar. The lagged resilience remains significantly negative, as -0.117, reinforcing the mean reversion effect observed with IDW. The coefficient for market accessibility to big cities is also positive and significant at 0.022, highlighting the importance of economic connectivity.

Again, population density, financial savings, and educational expenditure are not significant predictors of resilience. The spatial autocorrelation coefficient is much higher at 57.047, indicating extremely strong spatial dependencies. This suggests that cities with similar industrial production levels are highly interconnected in terms of their economic resilience, due to similar industrial structures.

The industrial-employment weighted matrix also shows consistent results, with the lagged resilience coefficient being significantly negative, as -0.114. Market accessibility to big cities remains a significant positive factor, as 0.023, underscoring its critical role in enhancing resilience. Similar to the other matrices, population density, financial savings, and educational expenditure do not significantly impact resilience.

The spatial autocorrelation coefficient is exceptionally high at 426.009, indicating very strong spatial dependencies based on employment patterns. This suggests that cities with similar industrial employment structures exhibit pronounced interdependencies in their economic resilience, reflecting shared labour market dynamics and economic interactions.

In conclusion, the analysis across different weight matrices reveals consistent patterns. The negative coefficients for lagged resilience indicate a mean reversion effect, while the positive coefficients for market accessibility to major cities underscore the critical role of connectivity in enhancing urban resilience. The high spatial autocorrelation coefficients across all matrices highlight the strong spatial dependencies in urban economic resilience. This finding aligns with regional economic theories, which emphasise the significance of integration into larger economic networks for urban development and resilience. Furthermore, the clustering of industries within production structures and labour markets is likely to amplify the interdependencies between cities.

3.6.2 Spatial Analysis

We then proceed to analyse the long-term and short-term benefits of spatial influence based on the results of the dynamic model. Our aim is to understand how connectivity and economic interdependencies shape urban economic resilience over different time horizons.

Short-term effects refer to the immediate impact of changes in explanatory variables on the dependent variable within a single time period. These effects capture the direct response of economic variables to shocks or policy changes without considering the propagation of these changes over time. In spatial models, short-term effects also account for immediate spatial interactions and dependencies between neighbouring regions or entities.

The mechanisms driving short-term effects include direct impacts within the same region, such as improved market access, enhanced economic activities due to better connectivity, and reduced transportation costs. Additionally, indirect or spatial spillover effects capture how changes in one region influence neighbouring regions through increased trade, investment,

and economic activities, thereby enhancing their economic resilience.

Long-term effects in dynamic spatial economics encompass the cumulative impact of changes in explanatory variables over multiple time periods, taking into account both the temporal and spatial propagation of these changes. These effects illustrate the persistence and diffusion of economic shocks or policy interventions over time, demonstrating how regions adjust and stabilise in response to these changes.

Temporal dynamics in long-term effects consider the influence of past values of the dependent and explanatory variables on current and future outcomes, reflecting the persistence of economic trends and regional adjustments over time. Spatial dynamics, by contrast, emphasise the ongoing interactions between regions, where positive changes in one region can yield lasting benefits for neighbouring regions through sustained trade, investment, and economic linkages.

The spatial analysis presented in Table 3.3 examines the long-term and short-term effects of various determinants on urban resilience. This analysis distinguishes between direct, indirect, and total effects.

Table 3.3: Spatial Analysis in Short-term and Long-term Effects

	Short-term			Long-term		
	IDW	Industrial Production	Industrial Employment	IDW	Industrial Production	Industrial Employment
Direct Effect						
L.Res				-0.119 *** (0.020)	-0.116 *** (0.021)	-0.114 *** (0.020)
lnMarketAcc	0.026 *** (0.008)	0.027 *** (0.008)	0.027 *** (0.008)	0.023 *** (0.008)	0.022 *** (0.008)	0.023 *** (0.008)
lnPopDensity	-0.002 (0.004)	-0.003 (0.004)	-0.001 (0.004)	-0.002 (0.004)	-0.002 (0.005)	-0.001 (0.004)
lnFinancesav	0.002 (0.006)	0.001 (0.006)	-0.001 (0.006)	-0.002 (0.006)	-0.002 (0.006)	-0.004 (0.006)
lnEduExp	-0.003 (0.009)	-0.002 (0.009)	0.000 (0.009)	0.005 (0.009)	0.006 (0.009)	0.009 (0.009)
Indirect Effect						
L.Res				-3.073 *** (0.513)	-2.987 *** (0.528)	-2.947 *** (0.513)
lnMarketAcc	0.677 *** (0.207)	0.689 *** (0.207)	0.697 *** (0.213)	0.585 *** (0.215)	0.576 *** (0.210)	0.597 *** (0.211)
lnPopDensity	-0.064 (0.109)	-0.069 (0.108)	-0.027 (0.105)	-0.053 (0.110)	-0.051 (0.116)	-0.021 (0.115)
lnFinancesav	0.044 (0.150)	0.031 (0.152)	-0.021 (0.152)	-0.045 (0.154)	-0.045 (0.153)	-0.103 (0.156)
lnEduExp	-0.069 (0.229)	-0.043 (0.225)	0.011 (0.235)	0.141 (0.235)	0.160 (0.236)	0.219 (0.234)
Total Effect						
L.Res				-3.192 *** (0.532)	-3.103 *** (0.565)	-3.062 *** (0.533)
lnMarketAcc	0.704 *** (0.215)	0.715 *** (0.215)	0.724 *** (0.221)	0.608 *** (0.223)	0.598 *** (0.275)	0.620 *** (0.219)
lnPopDensity	-0.066 (0.114)	-0.072 (0.112)	-0.028 (0.109)	-0.056 (0.114)	-0.053 (0.120)	-0.022 (0.120)
lnFinancesav	0.045 (0.156)	0.032 (0.158)	-0.021 (0.158)	-0.047 (0.160)	-0.047 (0.159)	-0.107 (0.162)
lnEduExp	-0.072 (0.238)	-0.044 (0.234)	0.012 (0.244)	0.146 (0.244)	0.166 (0.245)	0.228 (0.243)

Note: Standard error values in parentheses. ***, **, and * show the significance at 1%, 5%, and 10%, respectively.

In spatial econometrics, it is common to analyse direct and indirect effects separately, as they provide detailed insights into the mechanisms of influence. Accordingly, we present a summary of the spatial effects in our model.

Direct effects refer to the immediate impact of an explanatory variable on the dependent variable within the same region. These effects capture the primary relationship between the variable of interest and the outcome under study. Direct effects are significant because they demonstrate how specific changes affect the outcome in the targeted region, without accounting for the broader spatial context.

Indirect effects, also referred to as spatial spillover effects, represent the impact of changes in an explanatory variable in one region on the dependent variable in neighbouring regions. These effects arise due to the interconnectedness of regions through trade, investment, labour mobility, and other economic linkages. Indirect effects are crucial for understanding the broader implications of policy changes, as they reveal how benefits or costs extend beyond the initially affected area.

Total effects combine both direct and indirect effects, providing a comprehensive measure of the overall impact of a change in an explanatory variable on the dependent variable across all regions. Total effects account for both the immediate impact within the target region and the spillover effects on neighbouring regions, offering a holistic perspective on spatial influence.

Short-term Effects

The direct effect of market access on urban economic resilience is positive and statistically significant at the 1% level, with a coefficient of 0.026. The significance of market access

indicates that better connectivity and access to markets enable local areas to better absorb and recover from economic disturbances.

The indirect effect of market access, with a coefficient of 0.677, is also significant at the 1% level, highlighting the presence of spatial spillover effects. This means that improvements in market access in one region can enhance the economic resilience of neighbouring regions as well. This is consistent with the literature and our hypothesis that increased market access can increase trade, investment flows, and shared economic activities.

The analysis using the industrial production cluster weight matrix reinforces the importance of market access in enhancing urban economic resilience. The direct effect of market access remains positive and statistically significant at the 1% level, with a coefficient of 0.027. This indicates that within industrial production clusters, cities with better market access are more resilient to economic shocks. The enhanced resilience can be attributed to the robust industrial networks and supply chains that benefit from improved market connectivity.

The indirect effect, with a coefficient of 0.689, underscores the strong spatial spillover effects within industrial production clusters. This finding suggests that policies aimed at improving market access in one region can generate widespread benefits, enhancing the economic resilience of other regions within the cluster through shared industrial linkages and economic activities.

In the context of the industrial employment cluster weight matrix, the analysis continues to emphasise the critical role of market access in promoting urban economic resilience. Regions with better market access within industrial employment clusters are more capable of withstanding and recovering from economic shocks. Improved market connectivity also fosters enhanced employment opportunities.

The indirect effect, with a coefficient of 0.697, highlights substantial spatial spillover effects in regions with similar industrial employment structures. Enhancements in market access in one region can positively influence the economic resilience of neighbouring regions through increased labour market mobility.

Control variables such as population density, financial savings, and educational expenditure do not exhibit significant short-term direct or indirect effects on urban economic resilience within the weight matrices. These factors do not appear to play a significant role in immediate resilience improvements in the spatial context.

In conclusion, the analysis of short-term effects across all three weight matrices consistently demonstrates the significant influence of market access on urban economic resilience. The direct effects of market access are positive and statistically significant, indicating that enhanced market access leads to immediate improvements in urban economic resilience. Furthermore, the significant indirect effects highlight the presence of strong spatial spillover effects, suggesting that improvements in market access benefit not only the immediate region but also neighbouring regions through diverse economic linkages.

Long-term Effects

Regions with sustained improvements in market access within industrial employment clusters experience a measurable enhancement in urban economic resilience. Specifically, a 1% increase in market access is associated with a 0.023 increase in the resilience index, significant at the 1% level. While the elasticity may appear modest, in practical terms, this translates into higher survival rates of firms, more stable employment, and improved recovery trajectories in the aftermath of economic shocks.

Our findings align conceptually with Doran and Fingleton (2018), who emphasize the role of industrial structure in shaping resilience outcomes. Their study is from 2001 to 2014, with 377 metropolitan statistical areas. They use the Krugman Dissimilarity Index to measure how different a metropolitan area's industry mix is from the national average. During the crisis phase, regions with more specialized industry structures suffered disproportionately, as the coefficient is -0.077, but they also rebounded more quickly in the recovery phase as 0.0865.

While Doran and Fingleton (2018) emphasize the role of industrial structure specialization—measured through the Krugman Dissimilarity Index—in determining regional resilience, our study builds on this foundation by examining the role of spatial connectivity within industrial clusters. Specifically, we focus on how market access, conditional on industrial agglomeration, contributes to the ability of regions to withstand and recover from shocks.

Their results highlight that more specialized regions tend to suffer greater immediate losses but recover faster, due to stronger internal linkages and adaptive specialization. Our results suggest that market connectivity within these specialized or clustered regions plays a complementary and amplifying role. To sum up, resilience is not solely a function of what industries exist, but also how well-connected those industries are to markets and to each other.

The indirect effect of market access, with a coefficient of 0.597, remains significant, indicating substantial long-term spatial spillover effects. This suggests that improvements in market access within one region continue to positively influence the economic resilience of neighbouring regions through persistent labour market dynamics and shared economic benefits.

Regarding the dynamic effects of a time lag, the findings reveal that past resilience is significantly related to the foundation of current resilience, with notable negative direct and indirect effects. As noted in our regional analysis of resilience, except for a few large cities, the resilience of most small cities is below zero. This indicates that the resilience levels of small cities are far lower than the expected national level. Consequently, the negative effects of past resilience are likely to persist into subsequent periods.

Consistent with the findings from the short-term analysis, population density, financial savings, and educational expenditure do not exhibit significant long-term direct or indirect effects on urban economic resilience.

In summary, our analysis demonstrates that market access is a crucial factor for enhancing urban resilience in both short-term and long-term contexts, emphasising the substantial benefits of connectivity to major urban centres. The negative long-term effects of lagged resilience highlight the need for past efforts to be carefully managed and continuously adapted to mitigate detrimental impacts on current and future resilience. The strong and significant spatial autocorrelation underscores the interconnected nature of urban resilience. These findings align with our hypothesis that market access enhances resilience in both local and neighbouring regions, an effect amplified by agglomeration benefits derived from industrial structures and labour markets. This leads to greater spillover and direct benefits, particularly for agglomeration-based weight matrices.

3.7 Conclusion

This study provides a comprehensive analysis of the determinants of urban resilience using a dynamic spatial autoregressive model. The findings underscore the critical role of market

access in enhancing regional economic resilience across various spatial and temporal contexts.

The analysis consistently demonstrates that market access significantly enhances urban resilience. Both short-term and long-term effects of market access are positive and statistically significant, indicating that cities with better connectivity to major urban centres are more resilient. Additionally, the results reveal significant negative effects of lagged resilience on current resilience, both directly and indirectly, across all weight matrices. This finding is similar to the studies of Doran and Fingleton (2018) and Zhang et al. (2023), which suggests that previous resilience efforts may have been overestimated or mismanaged, adversely affecting current resilience levels.

To capture spatial effects, the spatial analysis further highlights the spillover effects of market access and resilience efforts. The positive indirect effects of market access reveal that improvements in one city's connectivity can benefit neighbouring cities, emphasising the importance of regional cooperation and coordinated policy interventions. Conversely, the negative indirect effects of lagged resilience suggest that inadequate local resilience efforts can also have detrimental impacts on surrounding regions.

When examining the industrial structure and labour market of cities, the analysis finds that industrial agglomeration amplifies the positive impact of transportation on resilience. Regions with high industrial production clusters benefit significantly from improved market access, which strengthens the robustness of industrial networks and supply chains.

This interconnectedness within industrial clusters facilitates the diffusion of economic benefits and stability, enhancing resilience against economic shocks. This is consistent with Krugman (1991) explanation of regional industrial agglomeration, namely that regions with industrial clusters can benefit from transportation. Similarly, regions with concentrated

industrial employment experience heightened resilience due to improved employment opportunities and economic stability provided by enhanced connectivity.

These findings indicate that policies aimed at improving transportation and market access should be integrated with efforts to support industrial agglomeration. By fostering robust industrial networks and labour market connectivity, regions can achieve higher levels of economic resilience. Such integrated policy approaches ensure that the benefits of enhanced market access and transportation infrastructure are maximised, contributing to sustained economic stability and growth across urban areas.

Conclusion

In this final section, we briefly summarize how our research builds on and extends the existing literature in terms of theory, policy implications, and empirical results.

Our three papers investigate the relationship between regional economic development and transport, offering new insights into the role of transport infrastructure across various regional and industrial settings. By examining multimodal transport networks, freight efficiency, and market access, we highlight the pivotal role of transport systems in promoting regional economic integration, enhancing industrial productivity, and strengthening the economic resilience of cities.

Our research addresses two key questions: How does economic activity in one region influence neighbouring regions through trade, transport, and labour mobility? And, can lagging regions benefit from improvements in transport infrastructure in other regions? These questions guide our exploration of how transport systems foster interregional connectivity and economic growth, particularly for less developed areas, by facilitating the flow of goods, services, and labour.

Our study makes two key contributions to the existing literature and research field. The introduction of synthetic instrumental variables allows for more precise control over endo-

geneity and enables the identification of the long-term dynamic effects of improved freight efficiency on economic output across different regions. This research bridges a gap in traditional methods for addressing endogeneity in economic models.

Secondly, we use a more detailed data framework and constructed a composite spatial weight matrix to simulate the influence of transport infrastructure on regional development. By incorporating machine learning techniques, we can extract detailed economic spatial features from extensive geographic data, accurately reflecting transport connections between regions. The composite matrix not only enhances the accuracy in capturing spatial correlations but also serves as a more flexible tool for assessing the influence of transport infrastructure on regional economies at various spatial scales.

New Economic Geography and Spatial Economics

We begin by outlining our contributions to the literature. A key concept is spatial dependence, where local factors shape the economic activity in one region but also affect neighbouring regions through transport, trade, and geographic proximity. This concept, central to economic geography and spatial economics, forms the foundation of our study.

First, Krugman introduced the core-periphery model, which explains the dynamics of regional economic agglomeration and uneven development. His model points out that economic activity tends to concentrate in a few core regions due to better market access, lower transport costs, and economies of scale. This agglomeration effect leads to the ‘core-periphery’ phenomenon, where core areas continue to attract more investment, labour, and firms, while peripheral areas may be left behind due to inaccessibility and market constraints. This model provides a foundational theoretical framework for understanding urbanization, industrial clustering, and regional inequality.

Subsequent studies by Banerjee, Donaldson, Masahisa Fujita, and Faber extended the understanding of spatial agglomeration and diffusion processes driven by regional transport. They emphasized the roles of increasing returns to scale, decreasing transport costs, and labour migration in urban and industrial concentration. Their research suggests that improvements in transport networks can reduce inter-regional frictions, promote the spatial diffusion of economic activities over a wider area, and alleviate over-agglomeration in core regions, thereby fostering more balanced development. On this basis, we further explore the role of transport infrastructure in regional economic development, in particular how it continues to influence industrial distribution and regional growth by optimising resource allocation and enhancing economic linkages between regions.

Meanwhile, work in spatial econometrics by researchers such as Anselin, Elhorst, Fingleton, and Martin has further advanced these concepts. Through the use of methods such as SAR and SDM, these scholars have quantified the direct and indirect effects of economic interactions between regions, especially those driven by geographic proximity and infrastructure improvements. Spatial economics highlights that regional economies are not only shaped by industrial activities within their borders but are also significantly impacted by spatial spillovers, making spatial econometric models particularly valuable for analysing the external benefits of transport infrastructure.

These studies emphasize the strong interdependence of spatial economic activities and the significant role that transportation and infrastructure networks play in shaping regional resource allocation and economic performance. They provide a solid theoretical and empirical foundation for understanding the dynamic processes of regional economic integration.

A key challenge in spatial economics is addressing the endogeneity of transport factors in

regional economic activities. In Chapter 2, we introduce synthetic instrumental variables and spatial filtering techniques, offering new solutions to the endogeneity problem in econometrics, particularly in panel data and large-scale observations. These synthetic instruments not only provide a novel approach for future empirical studies but also enhance the accuracy of capturing spatial interactions and dependencies between regions, using dynamic spatial Durbin models.

In Chapter 3, we discuss how cities' economic resilience can be improved through market access after an economic shock. We examine how existing transport infrastructure can help struggling cities recover by connecting regional resources and equalizing opportunities. Additionally, we extend the urban shadow theory, which typically focuses on the negative effects of large cities on smaller neighbouring areas. Our study highlights the positive effects large cities can have on smaller regions during crises, emphasizing the greater resilience of modern urban agglomerations. This offers a fresh perspective on industrial agglomeration and spatial diffusion in the context of new economic geography.

Our study builds on this foundation by refining the concept of spatial dependence through spatial weighting matrices and dynamic spatial modelling. We address the question posed at the outset: how does transportation cause economic factors to flow between unequal regions, creating both opportunities for regional development and resource redistribution? Our research expands on the role of transport infrastructure in industrial agglomeration and regional development, analysing different spatial dimensions such as distance, proximity, and economic similarity. In doing so, we contribute to the literature by enriching the understanding of regional economic spaces and how transportation links them.

Regional Inequalities and Transport

At the policy level, our research focuses on addressing regional inequalities and fostering the development of remote and underdeveloped regions. In China, transport infrastructure investment is primarily driven by the central government, which provides significant financial subsidies and support, while provincial and local governments contribute relatively less. This funding structure reflects the central government's dominant role in advancing national infrastructure projects, ensuring that remote areas also receive critical investment. However, the limited financial contributions from inter-provincial and local governments can result in uneven resource allocation, leading to disparities in development across regions. Our study offers valuable insights for policymakers on how to better optimize the allocation of funds, promoting more balanced and equitable regional economic development.

By revealing the differing impacts of various types of transport infrastructure—such as roads, railways, and inland waterways—on regional economies, our research provides new insights into addressing regional development inequalities. Remote and underdeveloped areas often suffer from inadequate transport infrastructure, limiting market access and resource flow, which in turn widens the economic gap with more developed regions.

We argue that targeted investments in transport infrastructure can help bridge this gap by improving market access and connectivity in developing regions. By identifying each region's specific transport needs, the government can formulate accurate regional development policies, with a focus on prioritizing infrastructure investments in areas where improved connectivity can significantly boost economic activity. We aim to help balance development between core and peripheral regions, mitigating economic disparities.

In our multimodal transport network, We analyse the differences between regions by categorizing provinces into eastern, central, and western areas. We identify how each regional

industry responds differently to multiple transport infrastructures.

Based on our findings, policymakers can adopt more targeted strategies for allocating investment funds. Rather than evenly distributing limited resources to less effective projects, funds should be concentrated on initiatives that significantly boost regional economic dynamism. For instance, prioritizing investments in rail networks in regions where rail infrastructure has a substantial positive impact on industrial production would not only enhance the efficiency of the investment but also ensure that transport projects contribute to sustainable regional economic growth.

We also emphasize that transport investment should focus not only on infrastructure development but also on improving freight efficiency. By assessing freight productivity, we have shown how efficient freight services enhance industrial production, regional competitiveness, and connectivity. A well-functioning freight transport system increases the economy by improving resource mobility and market accessibility. Therefore, to maximize economic return, transport investments should consider both infrastructure and freight efficiency improvements.

Good connectivity in transport networks strengthens the links between small cities and larger cities or other economic centres, expanding access to markets and facilitating the flow of resources, labour, and capital. For small cities, a strong transport network not only increases their integration into regional markets but also allows them to become part of the wider economy, enabling them to benefit from economies of scale and greater development opportunities. By improving access to larger economic centres, small cities can experience enhanced growth prospects and increased opportunities for industrial and commercial expansion.

In times of external shocks, urban agglomerations with robust transport connectivity are better positioned to quickly recover from economic disruptions. Cities with well-developed transport systems can re-establish connections to external markets, restore supply chains, and resume industrial production and business activities more efficiently. Without rapid restoration of transport and supply chain links, smaller cities may face long term economic downturns and more serious consequences.

However, transport connectivity can also have unintended negative effects. In some cases, regions with highly developed transport networks may not only aid in economic recovery but also accelerate the spread of economic shocks. For example, during an epidemic outbreak or economic crisis, the same transport networks can facilitate the rapid transmission of shocks to neighbouring cities and regions. Well-connected urban areas can become conduits for the spread of disruptions, affecting regions that may have otherwise been insulated from negative impacts.

Therefore, our findings provide valuable insights for the authorities to think of more resilient urban and regional infrastructure strategies. For developing regions that are more sensitive to shocks, investing in transport networks in advance can accelerate recovery by enhancing connectivity to external markets.

In Chapter 2, we measured labour productivity in the transport industry using freight volume per transport worker, which effectively captured transport efficiency at the provincial level. However, applying this measure at the city level posed several challenges. We notice that many small cities lacked freight employment statistics, limiting data representativeness and comparability. Additionally, variations in industrial structure affected transport demand—industrial hubs like Tangshan and Shenyang rely heavily on freight transport,

whereas service-driven cities like Beijing and Shenzhen have higher passenger transport volumes. Another key issue was traffic congestion, which significantly impacts transport efficiency in large cities. Despite developed road networks, cities like Beijing, Shanghai, and Guangzhou experience reduced per-worker transport efficiency during peak hours due to congestion, whereas smaller cities may appear more efficient simply due to lower traffic volumes. However, their higher efficiency does not necessarily reflect better infrastructure but rather lower transport demand. These challenges guide us to realize that it is not only the spatial scale that affects regional spatial research but also many other invisible factors. This is what we need to think about in subsequent research.

To sum up, our study provides theoretical support for building cross-regional transport networks, particularly in countries or regions with pronounced economic imbalances. Optimizing transport networks can promote the efficient allocation and flow of resources. Governments in lagging regions can drive neighbouring economic development by strengthening cross-regional transport connectivity, particularly in rail, road, and logistics. These policies not only promote regional economic integration but also foster cross-regional industrial output and reduce economic disparities.

Looking Back and Moving Forward

Over the past four years, we have dedicated much effort to this research, exploring the interactions between regional economies, transport infrastructure, and spatial effects. We have also refined our empirical methodology to capture better these dynamic processes. While our research has covered multiple aspects of transport infrastructure and freight efficiency, several key factors remain under explored, particularly regarding transport mobility and its side pollution.

Existing Problems

Current research in transport economics shows a clear tendency to focus more on investment in transport infrastructure while paying less attention to the actual operation of transport systems. One reason for this imbalance is the lack of detailed actual transport data. Many studies rely heavily on macro-level data, such as road construction funding or the budget of infrastructure projects, which often fail to reflect the operations of transport systems and their specific impact on economic activity.

In reality, the complexity of modern transport systems goes far beyond infrastructure investments. These systems consist of various components, including different road types—such as highways, urban roads, and rural roads—as well as data on vehicle flows, congestion patterns, and the types and purposes of vehicles in use. These factors provide a comprehensive picture of traffic flows, road-use efficiency, transport times, or costs, but difficult to obtain.

This operational data becomes particularly important when considering regional economic inequalities. Different regions have varying levels of dependence on and demand for transport infrastructure. Developed regions typically have more complete and efficient transport networks, while less developed regions may struggle with inadequate infrastructure, congestion, and inefficient road use. These imbalances in transport resources not only limit market access and economic growth in less developed areas but also exacerbate existing economic disparities between regions.

In our study, we also realize the limitations posed by the lack of detailed traffic data. We were unable to fully assess the physical capacity of roads, such as the number of vehicles that the infrastructure can support under actual operating conditions. Despite growing traffic pressures in rapidly developing urban areas—due to population growth and increased

industrial activity—we lack the data needed to accurately evaluate road capacity.

Furthermore, we also noticed the lack of maintenance data. The varying impact of roads on regional economies is influenced not only by the economic development level of the region but also by the condition and maintenance of the roads themselves. A road's physical condition and maintenance level have a direct effect on its longevity and operational efficiency. Even if two roads are constructed to the same design standards, their performance can diverge significantly if maintenance efforts differ. This can lead to varying economic outcomes across regions.

Economically developed regions are typically able to allocate more resources to maintaining and upgrading their transport infrastructure, ensuring long-term road efficiency. Conversely, less developed regions, with limited financial resources, often struggle to maintain or upgrade their infrastructure regularly. This leads to a decline in road quality, increasing transport costs and travel times, which, in turn, weakens the logistics efficiency and market competitiveness of these regions. The unequal state of infrastructure not only hinders the economic development of less developed areas but also widens the economic gap between regions, further impeding their integration into broader economic networks.

Lastly, our ability to connect all regions into a comprehensive transport network was hindered by differences in the statistical methods used by various regions and the lack of data for certain areas. Notably, data is often missing for less developed regions, making it difficult to fully assess the transport linkages between these areas and their more developed counterparts.

Looking ahead, future research will prioritize better access to and integration of missing data to construct a more complete and coherent regional transport network. We will also

explore new methods to expand and enhance data sources, particularly in understanding spatial dimensions.

Future Research Directions

Based on the existing literature and our research findings, our outlook for future research focuses on two key areas: expanding data dimensions and exploring econometric models of spatial relationships.

Exploration of High-Dimensional Data

First, we have acknowledged the limitations posed by the lack of traffic and regional data. A major issue in traditional transport economics and economic geography studies is the insufficient data granularity and dimensionality, especially in capturing the complexities of transport networks and their interactions with economic development. However, emerging studies are starting to incorporate more high-dimensional data, which not only reveal the efficiency of transport infrastructure but also capture its dynamics across both temporal and spatial scales.

One important category of high-dimensional data is real-time data, which integrates machine learning and big data networks. This includes real-time traffic flow, road quality, congestion patterns, vehicle types, and usage frequency. While this data offers a comprehensive understanding of the transport system's operations, its vast volume presents challenges related to data cleaning, dimensionality reduction, and complex processing.

Another significant source of high-dimensional data is geographic information derived from technologies like satellite remote sensing, sensor networks, and mobile device tracking. This data provides more detailed spatial information, allowing for a finer analysis of transport

systems in relation to economic geography. Increasingly, researchers—including us—are applying GIS to economic geography research. By integrating multidimensional data, we can analyze the precise impacts of transport infrastructure on regional economies, taking into account factors such as local climate, disaster frequency, and geomorphological variations. This type of data will support more informed policymaking by offering a holistic view of regional dynamics.

Our future research will utilize detailed remote sensing data combined with big data to investigate how spatial spillovers and economic geography manifest in energy use and land use. For example, by monitoring infrastructure investments through remote sensing, we can assess their impacts on surrounding regions in terms of economic activity, energy demand, and land use changes.

Remote sensing will allow governments and researchers to track land use changes over time and across space more accurately. This information can help create balanced regional development plans that align economic growth with environmental sustainability. Furthermore, remote sensing can quantify the scale and intensity of spillover effects, helping to identify which regions benefit and which bear greater environmental costs.

Spatial Interactions in Economics

Our future research outlook centres on the further exploration of spatial relationships in economics, with a particular focus on regional inequalities and the development paths of developing regions. We aim to explore how spatial relations explain regional disparities, economic agglomeration, and diffusion, with the perspectives of development economics and economic geography.

In development economics, the classical convergence hypothesis suggests that less developed

regions should experience faster growth than more developed ones, gradually narrowing the economic gap. However, spatial factors are often responsible for slowing down this convergence process. Research by Professor Lin Guangping and colleagues on China's regional economies highlights the slower rate of convergence due to spatial dynamics. In theory, less developed regions can catch up with their more developed neighbours through learning effects, technology diffusion, and policy interventions. Yet in practice, spatial barriers, market failures, and geographic constraints frequently slow or even halt this catch-up process.

To address these challenges, future research will focus on how to optimize spatial relationships to overcome these obstacles. Specifically, we will explore strategies to help less developed regions better integrate into broader economic networks through upgrades to transport infrastructure, enhanced regional connectivity, and improved market access. By constructing stronger spatial interactions, these regions can accelerate the absorption and spreading of technology and increase their participation in economic activities, thereby promoting more balanced regional development.

In economic geography, spatial relations are core concepts that explain the distribution, agglomeration, and diffusion of regional economic activities. Economic activities are not randomly scattered across space; they are shaped by geographical conditions, resource distribution, infrastructure, and interregional economic connections. The study of spatial agglomeration, such as urban agglomerations and industrial clusters, and economic diffusion, like technological spillovers and capital flows, has long been a classic topic in economic geography. Future research will place greater emphasis on dynamic spatial relationships, exploring how economic activities cluster or diffuse within specific spatial and temporal contexts.

With the rapid advancement of big data, artificial intelligence, and remote sensing technologies, we now have access to richer, multidimensional spatial data that allow us to better understand spatial activities. Future research will leverage these spatial dynamics to address pressing challenges in development economics and economic geography. In particular, spatial relationships offer new perspectives for explaining regional inequality, economic agglomeration, and technology diffusion. It helps to reveal the complex patterns of interaction between different regions.

Today's spatial analyses have evolved from traditional spatial matrices to more sophisticated models, such as spatial heterogeneity modelling, which introduces multiple weight matrices to capture region interactions. It provides valuable insights into regional heterogenous. Additionally, by combining traditional spatial metrics with network analysis, spatial network models are able to capture non-physical spatial interactions, such as trade flows, capital exchanges, and other economic relationships. By analysing key nodes and corridors in these networks, the authorities can more effectively identify regions of strategic importance within the economic landscape.

Our future research will continue to focus on the realities of regional development inequalities, particularly within the context of sustainable development, urban land use, and public health disparities. By combining spatial measurement tools with economic theory, we aim to tackle practical policy problems and help create more effective solutions.

Appendix A

Appendix to Chapter 1

This Appendix includes detailed descriptions of the data sources and variables used in the analysis and presents additional regression results for robustness checks.

A.1 Descriptive Statistics

Table A.1 provides a brief description of key variables used in the analysis, including measures of industrial production, infrastructure density (railways, roads, and rivers), government expenditure, foreign direct investment, and labour force. All data represents year-end stock values, capturing the final amounts for each variable.

The table provides a summary of the key variables, offering insights into the characteristics of various factors influencing regional economies, for 570 observations from 2001 to 2019. These variables include measures of industrial output, infrastructure, both rail and road, government expenditure, foreign direct investment, and labour force size. The descriptive

Table A.1: Data Description

Data	Description
Industrial Production (<i>Ind</i>)	total final results of industrial production activities [in 100 million yuan].
Railway density (<i>RailD</i>)	total actual length of railway ten thousand kilometres per actual area of each province
Road density (<i>RoadD</i>)	total actual length of road ten thousand kilometres per actual area of each province
Inner river per Boat (<i>IRB</i>)	total actual length of Inner river kilometres per freight boat
Government Expenditure (<i>GOVE</i>)	total local general budget expenditures [in 100 million yuan]
Foreign direct investment (<i>FDI</i>)	Total investment of foreign-invested enterprises [one million US dollars]
Labour (<i>Labour</i>)	Total number of labour force at the end of a year

Note: All data is stock data which is the final amount at the end of the year.

statistics present the mean and standard deviation of each variable, alongside the minimum and maximum values.

Table A.2: Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Ind	570	5680.663	6519.889	61.9	39063.673
roadD	570	0.751	0.483	0.032	2.126
railD	570	0.022	0.018	0.002	0.1
IRB	570	166.912	607.290	0.001	9800
GOVE	570	2538.095	2435.894	0.113	15654.163
FDI	570	563.331	509.311	10.169	2507.557
labor	570	839.889	972.139	40.9	5571.7
GDPgr	570	110.121	2.772	100.5	119.6

A.2 Main Regression

Table A.3: Baseline Main Results of SDM with Different Matrices

Weight Matrix Spatial Lag of X (W^*)	Adjacency-W1	Distance-W2	Economic-W3
RoadD	-0.016 (-0.480)	0.052** (2.320)	0.026 (1.090)
RailD	-0.168*** (-4.460)	-0.183*** (-5.530)	-0.091*** (-2.650)
IRB	-0.007 (-1.170)	-0.007 (-1.010)	-0.006 (-0.760)
lnGOVE	0.132*** (3.620)	-0.047 (-1.140)	0.071 (1.620)
lnFDI	-0.218*** (-4.680)	-0.046 (-0.990)	-0.162*** (-3.520)
lnLabour	-0.070** (-2.260)	-0.073** (-2.380)	-0.063* (-1.890)
Observation	570	570	570
Spatial.rho	0.743*** (29.630)	0.781*** (22.560)	0.585*** (20.650)
R-squared	0.9235	0.9157	0.9217
Region FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes

Note: Z-statistics in parentheses. All the spatial matrices are row-standardized. ***, **, and * show the significance at of 1%, 5% and 10%, respectively.

A.3 Robust Results

To further test the robustness of our model, we introduced the GDP growth rate (GDPgr) as a control variable. The rationale for including GDPgr is to account for macroeconomic fluctuations that could influence regional economic performance and, by extension, affect the relationship between transport infrastructure and industrial output. By controlling for overall economic growth, we aim to isolate the specific effects of road and rail density, ensuring that observed relationships are not confounded by broader cyclical trends.

The inclusion of GDPgr did not substantially alter the coefficients of the key variables. This suggests that our model is robust, as the estimates for road and rail density remained stable and statistically significant. The stability of these coefficients reinforces the interpretation that transport infrastructure—particularly road networks—plays a consistent and significant role in shaping regional economic outcomes, even after controlling for macroeconomic dynamics.

In a second robustness check, we excluded the financial crisis year (2008) from our dataset. As discussed in the literature review, periods of exceptional economic volatility—such as global shocks and financial crises—can lead to distorted or even negative spillover effects from transport infrastructure. Removing 2008 ensures that our estimates are not driven by anomalies associated with this extraordinary year.

After re-estimating the model without 2008, we found that the key coefficients remained largely unchanged and retained their statistical significance. This indicates that the observed relationships between infrastructure density and industrial output are not overly sensitive to the presence of outlier events. The robustness of these findings further strengthens the reliability of our empirical results and supports our main conclusions regarding the spatial

and economic significance of transport infrastructure.

We also tested the robustness of the spillover effects, and found that these effects remained stable and statistically significant across model specifications. This consistency reinforces the conclusion that the model reliably captures the spatial interdependence of regional economies, as well as the dynamic flow of labour, goods, and resources across administrative and geographic boundaries.

These findings further emphasize that infrastructure investment should not be evaluated solely based on local impacts. Instead, policymakers must also account for the broader spatial spillovers that influence neighbouring regions, promote economic integration, and enhance the overall functionality of the transport network.

Table A.4: Robust Estimation - with GDP growth rate

Matrix	W1	W2	W3
Main			
RoadD	-0.009 (-0.290)	0.056** (2.520)	0.030 (1.270)
RailD	-0.147*** (-3.900)	-0.177*** (-5.370)	-0.069** (-1.980)
IRB	-0.008 (-1.230)	-0.007 (-0.990)	-0.005 (-0.700)
lnGOVE	0.199*** (5.000)	0.004 (0.100)	0.138*** (2.930)
lnFDI	-0.247*** (-5.300)	-0.074 (-1.560)	-0.218*** (-4.540)
lnlabor	-0.046 (-1.470)	-0.053* (-1.670)	-0.040 (-1.190)
GDPgr	0.018*** (3.860)	0.013** (2.590)	0.017*** (3.420)
W_x			
RoadD	-0.065 (-1.000)	0.143** (2.710)	0.207*** (6.670)
RailD	0.118* (2.460)	-0.450*** (-3.230)	0.008 (0.200)
IRB	0.034* (2.180)	0.009 (0.270)	0.017* (1.710)
lnGOVE	-0.685*** (-8.920)	-0.848*** (-3.680)	0.111 (0.900)
lnFDI	0.465*** (6.040)	0.540*** (2.920)	-0.165* (-1.790)
lnlabor	0.029 (0.850)	0.080** (2.190)	-0.023 (-0.660)
GDPgr	-0.009 (-1.310)	-0.014 (-1.440)	0.003 (0.420)
Spatial rho	0.737*** (28.170)	0.784*** (22.240)	0.567*** (19.270)

Note: Z-statistics in parentheses. All the spatial matrices are row-standardized. SDM analysis with fixed-effect. ***, **, and * show the significance at the 1%, 5% and 10% levels, respectively.

Table A.5: Robust Estimation - without financial shock

Matrix	W1	W2	W3
Main			
RoadD	-0.016 (-0.480)	0.052** (2.320)	0.026 (1.090)
RailD	-0.168*** (-4.460)	-0.183*** (-5.530)	-0.091*** (-2.650)
IRB	-0.007 (-1.170)	-0.007 (-1.010)	-0.006 (-0.760)
lnGOVE	0.132*** (3.620)	-0.047 (-1.140)	0.071 (1.620)
lnFDI	-0.218*** (-4.680)	-0.046 (-0.990)	-0.162*** (-3.520)
lnlabor	-0.070** (-2.260)	-0.073** (-2.380)	-0.063* (-1.890)
W_x			
RoadD	-0.058 (-0.880)	0.147** (2.880)	0.220*** (7.100)
RailD	0.127** (2.620)	-0.452*** (-3.330)	-0.024 (-0.580)
IRB	0.030* (1.910)	0.007 (0.210)	0.012 (1.190)
lnGOVE	-0.681*** (-9.580)	-0.801*** (-3.710)	0.083 (0.690)
lnFDI	0.502*** (7.230)	0.510*** (3.050)	-0.110 (-1.240)
lnlabor	0.043 (1.270)	0.102** (2.850)	-0.017 (-0.480)
Spatial rho	0.743*** (29.630)	0.781*** (22.560)	0.585*** (20.650)

Note: Z-statistics in parentheses. All the spatial matrices are row-standardized. SDM analysis with fixed-effect. ***, **, and * show the significance at of 1%, 5% and 10% levels, respectively.

Table A.6: Robust Tests - Spillover Effects1

Weight Matrix Spatial Lag of X (W*)		Adjacency-W1	Distance-W2	Economic-W3
Road Density	Direct	0.044 (1.660)	0.083*** (3.650)	0.098*** (3.990)
	Indirect	0.367*** (3.850)	0.856*** (3.900)	0.452*** (8.480)
	Total	0.412*** (4.480)	0.939*** (4.200)	0.550*** (8.780)
Rail Density	Direct	-0.230*** (-6.410)	-0.262*** (-6.190)	-0.079** (-2.050)
	Indirect	-0.571*** (-3.370)	-2.705*** (-3.150)	-0.054 (-0.650)
	Total	-0.801*** (-4.360)	-2.967*** (-3.350)	-0.133 (-1.230)
Inland Waterway	Direct	0.007 (0.780)	-0.005 (-0.580)	0.000 (0.010)
	Indirect	0.096* (1.920)	0.032 (0.190)	0.030 (1.400)
	Total	0.102 (1.820)	0.027 (0.160)	0.030 (1.080)
lnGOVE	Direct	-0.062 (-1.150)	-0.119* (-1.800)	0.194*** (3.360)
	Indirect	-1.831*** (-5.700)	-4.005*** (-2.780)	0.377 (1.570)
	Total	-1.893*** (-5.260)	-4.124*** (-2.760)	0.570** (2.040)
lnFDI	Direct	-0.109** (-2.180)	-0.002 (-0.030)	-0.302*** (-5.440)
	Indirect	0.975*** (3.750)	2.331** (2.140)	-0.572*** (-3.110)
	Total	0.866*** (3.010)	2.330** (2.060)	-0.874*** (-3.960)
lnLabour	Direct	-0.047 (-1.650)	-0.046 (-1.450)	-0.052* (-1.710)
	Indirect	-0.019 (-0.390)	0.171 (1.850)	-0.095** (-2.630)
	Total	-0.066 (-1.350)	0.126 (1.330)	-0.147*** (-4.220)
lnGDP Growth	Direct	0.020*** (4.050)	0.012* (2.350)	0.020*** (3.780)
	Indirect	0.013 (0.780)	-0.021 (-0.490)	0.024** (2.270)
	Total	0.033* (1.730)	-0.009 (-0.200)	0.044*** (3.280)

Note: Z-statistics in parentheses. All the spatial matrices are row-standardized. ***, **, and * show the significance at of 1%, 5% and 10%, respectively.

Table A.7: Robust Tests - Spillover Effects2

Weight Matrix Spatial Lag of X (W*)		Adjacency-W1	Distance-W2	Economic-W3
Road Density	Direct	0.041 (1.440)	0.078*** (3.340)	0.102*** (3.960)
	Indirect	0.390*** (4.110)	0.836*** (4.440)	0.495*** (8.780)
	Total	0.431*** (4.540)	0.914*** (4.690)	0.597*** (8.710)
Rail Density	Direct	-0.259*** (-7.450)	-0.268*** (-6.420)	-0.118*** (-3.160)
	Indirect	-0.629*** (-3.800)	-2.734*** (-3.410)	-0.166** (-2.120)
	Total	-0.888*** (-4.990)	-3.003*** (-3.620)	-0.284*** (-2.820)
Inland Waterway	Direct	0.005 (0.640)	-0.005 (-0.600)	-0.002 (-0.200)
	Indirect	0.087* (1.760)	0.030 (0.190)	0.021 (0.920)
	Total	0.092 (1.660)	0.024 (0.150)	0.019 (0.640)
lnGOVE	Direct	-0.152*** (-3.040)	-0.165** (-2.860)	0.110 (1.940)
	Indirect	-2.002*** (-6.690)	-3.840*** (-3.100)	0.263 (1.010)
	Total	-2.154*** (-6.410)	-4.005*** (-3.120)	0.373 (1.230)
lnFDI	Direct	-0.053 (-1.090)	0.021 (0.400)	-0.227*** (-4.090)
	Indirect	1.168*** (5.220)	2.177** (2.460)	-0.431** (-2.240)
	Total	1.115*** (4.510)	2.198** (2.400)	-0.658*** (-2.850)
lnLabour	Direct	-0.072*** (-2.610)	-0.065** (-2.150)	-0.078*** (-2.610)
	Indirect	-0.030 (-0.610)	0.201* (2.070)	-0.111*** (-2.900)
	Total	-0.102* (-2.190)	0.136 (1.400)	-0.189*** (-5.410)

Note: Z-statistics in parentheses. All the spatial matrices are row-standardized. ***, **, and * show the significance at 1%, 5%, and 10% levels, respectively.

Appendix B

Appendix to Chapter 2

B.1 Data Description

The Data Description table outlines key variables such as Industrial Production, Railway and Road Services, Government Expenditure, Population, Education, Labour, and GDP Growth Rate, quantifying each in specific metrics like millions of yuan or total counts. These variables assess the impact of socio-economic factors on regional industrial output.

Table B.2 offers a detailed statistical breakdown for these variables across 570 observations, providing mean values, standard deviations, and the range of data (minimum and maximum values). All economic data are adjusted for inflation and planned to be transformed using logarithms to facilitate more nuanced analysis, except for GDP growth rate percentages.

Table B.1: Data Description

Data	Description
Industrial Production (Ind)	Total final value-added result of industrial production activities [in 100 million yuan].
Railway Services (RailS)	Regional rail freight volume/ number of labour in railway freight system
Road Services (RoadS)	Regional road freight volume/ number of labour in road freight system
Government Expenditure (GOVE)	Total general budget expenditures in local government [in 100 million yuan]
Population (POP)	Total number of population at the end of a year
Education (EDU)	Total expenditure invested in education filed at the end of a year
Labour (Labour)	Total number of labour force at the end of a year
GDP Growth Rate (GDPgr)	The growth rate of Gross Domestic Product (GDP) is calculated based on the increase in value from the previous year

Notes: Economic data are reported in current prices for each year and are adjusted for inflation using the relevant price index.

Table B.2: Variables Statistic Summary

Variable	Obs	Mean	Std. Dev.	Min	Max
Ind	570	5680.663	6519.889	61.9	39063.673
RailS	570	0.182	0.154	0.005	1.114
RoadS	570	1.137	0.606	0.066	3.872
POP	570	4364.005	2808.287	7.04	12489
GOVE	570	2538.095	2435.894	0.113	15654.163
Labour	570	839.889	972.139	40.9	5571.7
EDU	570	97.484	433.537	0.26	2472
GDPgr	570	106.386	19.943	0.001	117.2

Notes: This table presents a summary of the raw data. All data values will be transformed using their logarithms lately. GDP growth rate is in percentage

B.2 Windmeijer's Small-Sample Correction

We are based on the small sample correction by Windmeijer (2005) and xtabond2 by Roodman (2009). Windmeijer's key idea is to account for the uncertainty of the weight matrix in the variance of parameter estimates. The correction proceeds in the following steps:

Windmeijer defines the GMM bias function $g(Y, \hat{\Omega})$ as:

$$g(Y, \hat{\Omega}) = \left[X'Z \left(Z'\hat{\Omega}Z \right)^{-1} Z'X \right]^{-1} X'Z \left(Z'\hat{\Omega}Z \right)^{-1} Z'E$$

The variable Y represents the data used in the model, while $\hat{\Omega}$ denotes the estimated weight matrix that plays a critical role in the GMM estimation process. Additionally, E refers to the true disturbance term, capturing the unobserved error in the system.

The bias function g depends on both the data Y and the estimated weight matrix $\hat{\Omega}$. To analyze its variance, Windmeijer performs a first-order Taylor expansion of g :

$$g(Y, \hat{\Omega}_{\beta_1}) \approx g(Y, \hat{\Omega}_{\beta}) + Dg(Y, H)$$

where:

$$D = \frac{\partial g(Y, \hat{\Omega}_{\beta})}{\partial \beta}$$

is the derivative matrix that captures the impact of the weight matrix on parameter estimates.

Finally, the corrected variance formula is given by:

$$\widehat{\text{Avar}}_c(\hat{\beta}_2) = \widehat{\text{Avar}}(\hat{\beta}_2) + \hat{D}\widehat{\text{Avar}}(\hat{\beta}_2) + \widehat{\text{Avar}}(\hat{\beta}_2)\hat{D}' + \hat{D}\widehat{\text{Avar}}(\hat{\beta}_1)\hat{D}'$$

where the first term represents the uncorrected two-step GMM variance estimate, while the last term incorporates the robust variance from the first-step estimate to correct for the second-step bias.

Therefore, in simulations using difference GMM regressions on panel data, Windmeijer (2005) demonstrated that two-step Efficient GMM outperforms one-step GMM in terms of coefficient estimation, yielding lower bias and smaller standard errors. Furthermore, the corrected two-step standard errors, proved to be highly accurate, making two-step estimation with corrected errors a slightly superior approach compared to cluster-robust one-step estimation.

B.3 Diagnostic Tests

The diagnostic tests presented in Table A.9 demonstrate that synthetic instruments, regardless of whether they have a positive or negative MC, are effective in addressing spatial autocorrelation in residuals for both GMM and OLS models. The Arellano-Bond tests indicate significant first-order autocorrelation (AR(1)) and no second-order autocorrelation (AR(2)) for both groups, which is consistent with expectations for a well-specified first-difference GMM model.

The Sargan-Hansen over identification tests also yield high p-values (1.000) for both positive and negative MC categories, suggesting that the instruments do not correlate with the error terms and hence are valid. This validation confirms the robustness of the model and supports the integrity of the instruments used.

Table B.3: Two-Step GMM Diagnostic Tests - with All Instruments

Supervised Instrument in MC	Postive	Negative
Number of instruments	131	131
Arellano-Bond Test		
AR(1)	-2.11(0.035)	-2.77(0.006)
AR(2)	0.73(0.465)	-0.00(0.997)
Instrument Validity		
Sargan-Hansen test of overid	18.19	18.96
p-value	1.000	1.000
d. of freedom	115	115

Note: "Over 0.25" indicates all instruments with a Moran coefficient over 0.25. Supervised instrument selection is based on the Moran coefficient as positive, negative and total value. P-value presented in parentheses for Arellano-Bond tests.

B.4 Residual Analysis for Spatial Autocorrelation

We estimate the spatial autocorrelation in residuals using various synthetic instruments. As shown in Table B.4, spatial autocorrelation persists in the instrument regression. With lower Moran Coefficients (MC), the results reveal more significant global Moran's I values, indicating that the MC influences the residual performance in spatial autocorrelation. Moreover, higher-ranking instruments tend to reduce spatial autocorrelation in residuals.

Table B.5 presents a residual analysis on Global Moran's I statistics for three different estimation methods: GMM, OLS, and SDM. If spatial autocorrelation is detected, it indicates that the model might not fully capture the spatial relationships between data points, leaving spatial patterns in the residuals.

For the GMM model, the results show relatively small values of Moran's I and Z-statistics, with p-values consistently above 0.1 for all years. This indicates that there is no significant spatial autocorrelation in the residuals of the GMM model.

Table B.4: Residual Analysis on Global Moran's I statistics - with Synthetic Instruments

	Z4		Z7		Z10	
Year	$I(Z)$	P-value	$I(Z)$	P-value	$I(Z)$	P-value
2001	.	.	.	.	.	.
2002	0.379(4.446)	0.000	0.307(3.664)	0.000	0.438(5.128)	0.000
2003	0.325(3.844)	0.000	0.275(3.261)	0.001	0.309(3.633)	0.000
2004	0.358(4.290)	0.000	0.361(4.356)	0.000	0.425(5.117)	0.000
2005	0.339(3.982)	0.000	0.338(3.962)	0.000	0.400(4.652)	0.000
2006	0.329(3.950)	0.000	0.283(3.516)	0.000	0.351(4.206)	0.000
2007	0.411(4.793)	0.000	0.352(4.110)	0.000	0.309(3.689)	0.000
2008	0.224 (2.771)	0.006	0.206(2.667)	0.008	0.157(2.097)	0.036
2009	0.494(5.564)	0.000	0.431(4.890)	0.000	0.379(4.323)	0.000
2010	0.403(4.708)	0.000	0.327(3.911)	0.000	0.406(4.763)	0.000
2011	0.431(4.965)	0.000	0.388(4.490)	0.000	0.475(5.571)	0.000
2012	0.334(3.911)	0.000	0.280(3.366)	0.001	0.356(4.187)	0.000
2013	0.245(3.043)	0.002	0.202(2.520)	0.012	0.275(3.344)	0.001
2014	0.582(6.998)	0.000	0.594(7.257)	0.000	0.504(5.975)	0.000
2015	0.320(4.070)	0.000	0.302(3.711)	0.000	0.362(4.470)	0.000
2016	0.167(2.151)	0.032	0.094(1.385)	0.166	0.206(2.567)	0.010
2017	0.247(3.038)	0.002	0.202(2.594)	0.009	0.261(3.207)	0.001
2018	0.392(4.532)	0.000	0.359(4.181)	0.000	0.383(4.432)	0.000
2019	0.316(3.862)	0.000	0.345(4.086)	0.000	0.303(3.706)	0.000

Note: Z-statistics in parentheses. Residual analysis with highest MC instruments in both rail service and road service

Similarly, the OLS model shows no significant spatial autocorrelation in the residuals. The Moran's I values and Z-statistics are comparable to those of the GMM model, and the p-values also remain consistently above 0.1. This suggests that the OLS model, despite not being specifically designed to account for spatial dependence, does not leave any significant spatial patterns in the residuals.

The Moran's I values for SDM are notably higher, and the Z-statistics are consistently large, with p-values below 0.05 in most years. This indicates that significant spatial autocorrelation

is present in the residuals of the SDM model, meaning that spatial dependencies remain even after using a spatial model.

Table B.5: Residual Analysis on Global Moran's I statistics - with Models

Year	GMM		OLS		SDM	
	$I(Z)$	P-value	$I(Z)$	P-value	$I(Z)$	P-value
2001	.	.	.	.	.	.
2002	0.083 (1.307)	0.191	0.103 (1.522)	0.128	0.280(3.347)	0.001
2003	0.083 (1.302)	0.193	0.102(1.510)	0.131	0.252(3.030)	0.002
2004	0.083 (1.306)	0.192	0.101(1.505)	0.132	0.303(3.627)	0.000
2005	0.083 (1.305)	0.192	0.103 (1.526)	0.127	0.321(3.783)	0.000
2006	0.083 (1.304)	0.192	0.105(1.546)	0.122	0.239 (3.015)	0.003
2007	0.082 (1.296)	0.195	0.104(1.538)	0.124	0.505 (5.716)	0.000
2008	0.082 (1.292)	0.196	0.104 (1.542)	0.123	0.225(2.856)	0.004
2009	0.081(1.282)	0.200	0.100(1.492)	0.136	0.461 (5.154)	0.000
2010	0.082 (1.288)	0.198	0.101 (1.510)	0.131	0.366(4.306)	0.000
2011	0.082(1.290)	0.197	0.103 (1.530)	0.126	0.301(3.558)	0.000
2012	0.082 (1.289)	0.197	0.104 (1.533)	0.125	0.190 (2.395)	0.017
2013	0.081(1.281)	0.200	0.101 (1.503)	0.133	0.212(2.649)	0.008
2014	0.081 (1.286)	0.199	0.105(1.544)	0.122	0.565(6.803)	0.000
2015	0.081(1.280)	0.201	0.104(1.542)	0.123	0.350(4.264)	0.000
2016	0.080(1.272)	0.203	0.101(1.503)	0.133	0.099 (1.435)	0.151
2017	0.080(1.269)	0.204	0.101(1.503)	0.133	0.206(2.613)	0.009
2018	0.080(1.276)	0.202	0.105 (1.551)	0.121	0.369(4.288)	0.000
2019	0.080 (1.276)	0.202	0.104 (1.538)	0.124	0.322 (3.875)	0.000

Note: Z-statistics in parentheses. Residual analysis with highest MC instruments in both rail service and road service

Appendix C

Appendix to Chapter 3

C.1 Data Description

For this Appendix to Chapter 3, we present the basic statistical description for major data.

Table C.1 present a summary of data. It presents key data variables used in the analysis, including regional economic resilience, market accessibility, government expenditure, population density, education expenditure, and finance savings. It is important to note that, in line with existing literature, we have not adjusted the data for price changes using a price index.

Table provides a summary of key economic variables across 2,304 observations from 2014 to 2021 for 288 cities, including measures of central tendency and variability. The variables analyzed are GDP residuals, market access, population density, financial figures, government expenditure, and education expenditure.

Resilience show a mean value slightly below zero, with low variability, indicating relatively

Table C.1: Data Description

Data	Description
Regional Economic Resilience (Res)	Resilience based on Gross Domestic Product (GDP).
Market Accessibility (MarketAcc)	Weighted cumulative cost of traversing through the network from the starting point to the destination.
Government Expenditure (GOVE)	Total general budget expenditures in local government [in 100 million yuan]
Population Density (POP)	Population of resident per square kilometre
Education Expenditure (EDU)	Total expenditure invested in education filed at the end of a year
Finance Saving (Finance)	Year-end Balance of Savings of Urban and Rural Residents (10,000 yuan)

consistent residual values across the observations. Market access exhibits significant variation, suggesting disparities in accessibility across different regions or entities. Similarly, control variables population distance, finance level, government expenditure, and education expenditure also varies widely.

Table C.2: Descriptive Statistics of Variables

Variable	Obs	Mean	Std. Dev.	Min	Max
GDPres	2,304	-0.9470199	0.1663316	-3.91286	2.39188
MarketAcc	2,304	255042.8	150610.4	67792.95	1197092
PopDensity	2,304	437.0278	339.8478	6	2648
Finance	2,304	2.07e+07	3.01e+07	583668	3.73e+08
GOVEExp	2,304	4940060	7327093	220571	8.43e+07
EduExp	2,304	824810.8	1063767	31233	1.15e+07

Notes: Economic data are reported in current prices for each year and are adjusted for inflation using the relevant price index.

C.2 Euclidean Distance

Here, we introduce how we measure the distance with Euclidean Distance in both geography and image. We follow the calculation in ArcGIS¹ and QGIS².

Euclidean Distance is derived from the principles of Euclidean geometry, which deals with the properties and relationships of points, lines, and surfaces in space. The concept is based on the Pythagorean theorem, which states that in a right-angled triangle, the square of the length of the hypotenuse (the side opposite the right angle) is equal to the sum of the squares of the lengths of the other two sides.

We consider Euclidean Distance in a GIS context. Euclidean Distance measures the straight-line distance between points in a spatial dataset, whether these points represent features in a vector dataset, as cities on a map in our city distance calculation.

The Euclidean Distance for vector data involves calculating the straight-line distance between two points in a our map. Consider two points P_1 , as city 1, and P_2 , as city 2, with coordinates (x_1, y_1) and (x_2, y_2) , respectively. The formula to compute the distance d between these points is given by:

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

In this formula, x_1 and y_1 represent the coordinates of the first point, while x_2 and y_2 represent the coordinates of the second point. The calculation involves finding the differ-

¹ArcGIS Pro, Euclidean Distance - Spatial Analyst. Retrieved from <https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-analyst/euclidean-distance.htm>

²QGIS 3.10, Distance calculation. Retrieved from <https://qgis3-10-geoanalysis-un.readthedocs.io/en/latest/vector/proximity.html?highlight=Euclidean%20Distance#distance-from-point-to-point>

ence between the corresponding coordinates of the two points, squaring these differences, summing them, and then taking the square root of this sum. This provides the direct, straight-line distance between the two points in a Cartesian plane.

We also calculate the cells in a raster dataset, as pixels in an image for our industry analysis. We measure the distance between each data point (city) and the cluster centroids in the cluster image.

For raster data, the Euclidean Distance calculation determines the straight-line distance between a specific cell and the nearest source cell within a raster grid. Let's consider a cell located at coordinates $(x_{\text{current}}, y_{\text{current}})$ and a source cell located at coordinates $(x_{\text{source}}, y_{\text{source}})$. The distance d between these two cells is calculated using the formula:

$$d = \sqrt{(x_{\text{source}} - x_{\text{current}})^2 + (y_{\text{source}} - y_{\text{current}})^2}$$

In this context, x_{current} and y_{current} are the coordinates of the current cell being evaluated, and x_{source} and y_{source} are the coordinates of the nearest source cell. The formula works by computing the differences between the x and y coordinates of the current cell and the source cell, squaring these differences, adding them together, and then taking the square root of the resultant sum. This yields the straight-line distance from the current cell to the nearest source cell within the raster grid.

C.3 Global Moran's I

We measure the spatial interaction with Global Moran's I which is widely used in the geographic estimation. We are based on the calculation in ArcGIS³.

Global Moran's I measures spatial autocorrelation, assessing whether a spatial pattern is clustered, dispersed, or random. The index considers both the locations and attribute values of features, calculating a statistic that indicates the degree of spatial clustering. Positive Moran's I values suggest clustering, negative values indicate dispersion and values near zero imply randomness. The result is interpreted using a z-score and p-value to determine statistical significance, helping identify meaningful spatial patterns.

$$I = \frac{N \sum_{i=1}^N \sum_{j=1}^N w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{W \sum_{i=1}^N (x_i - \bar{x})^2} \quad (\text{C.1})$$

In this equation, N represents the total number of spatial units, each indexed by i and j . The variable x_i is the observed value of the variable of interest at location i . The mean of the variable of interest across all spatial units is denoted by $\bar{x}$. The spatial weight between location i and location j is represented by w_{ij} , which reflects the spatial relationship between the units, such as distance or contiguity. The sum of all spatial weights is given by W , which is defined as $W = \sum_{i=1}^N \sum_{j=1}^N w_{ij}$.

The numerator of the equation, $N \sum_{i=1}^N \sum_{j=1}^N w_{ij} (x_i - \bar{x})(x_j - \bar{x})$, measures the covariance between spatial units, taking into account the spatial weights. This part of the equation sums the product of the deviations of each pair of observations from the mean, weighted by

³ArcGIS pro, How Spatial Autocorrelation (Global Moran's I) works, Retrieved from: <https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/h-how-spatial-autocorrelation-moran-s-i-spatial-st.htm>

the spatial relationship between them.

The denominator, $W \sum_{i=1}^N (x_i - \bar{x})^2$, is a normalization factor that accounts for the overall variation in the data and the sum of the spatial weights.

Appendix D

City-level Replication

D.1 Abstract

In the first chapter, we investigate the spatial effects of three types of transportation infrastructure on industrial output at the provincial level. Our findings, consistent with prior studies such as Rahman et al. (2020), Tong et al. (2013), and Sharif et al. (2019), reveal generally positive effects at the provincial or national scale. However, we also observe that several city-level studies report negative impacts of road infrastructure. We aim to know whether it is a potential scale-dependent variation in the effects of transportation development.

To investigate this phenomenon further, we replicate the analysis using city-level data, employing a more detailed spatial scale and a larger dataset to reassess the impact of roads on urban industrial output. At the city level, we develop a road accessibility measure that combines accessibility to major urban road networks with reliance on road freight trans-

portation. Our analysis indicates that, although the spatial effects of urban transportation networks exhibit a negative impact.

D.2 Introduction

We replicate the analysis using road data at the city level from a short-distance perspective. By incorporating more accurate urban geographic data, we enhance the spatial precision of the analysis. Comparing these results with broader regional analyses allows us to examine how the impact of roads varies across spatial scales, providing insights into the dynamics of localised versus regional spillover effects.

Negative effects of transportation infrastructure among cities are a common finding in urban research, especially in city-level research. These effects are often attributed to the resource dispersion effect. Several empirical studies have further validated these potential negative impacts of transportation infrastructure. For instance, Chandra and Thompson (2000) examined the economic effects of highways on cities in the United States and found that some industries declined as economic activities migrated away from certain cities. Similarly, Qin (2016) found that for non-major cities at the county level in China, the economic benefits of transportation improvements were often diminished. Baum-Snow et al. (2017) demonstrated that in China, radial highways dispersed service activities and railways led to the dispersion of industrial activities, ultimately reducing industrial output in some cities along transportation routes. Additionally, Li et al. (2020) confirmed that the positive economic impacts of transportation infrastructure were largely confined to more developed major cities, with little to no significant benefit for smaller cities.

The development of urban agglomerations and improvements in transportation infrastruc-

ture tend to draw more resources towards central cities, leaving peripheral cities and rural areas increasingly marginalised due to their lack of competitiveness. While transportation infrastructure can stimulate economic activity, it may also exacerbate regional imbalances and produce negative effects. In this chapter, we focus on comparing the varying spatial effects of transportation infrastructure across different spatial scales at the city level.

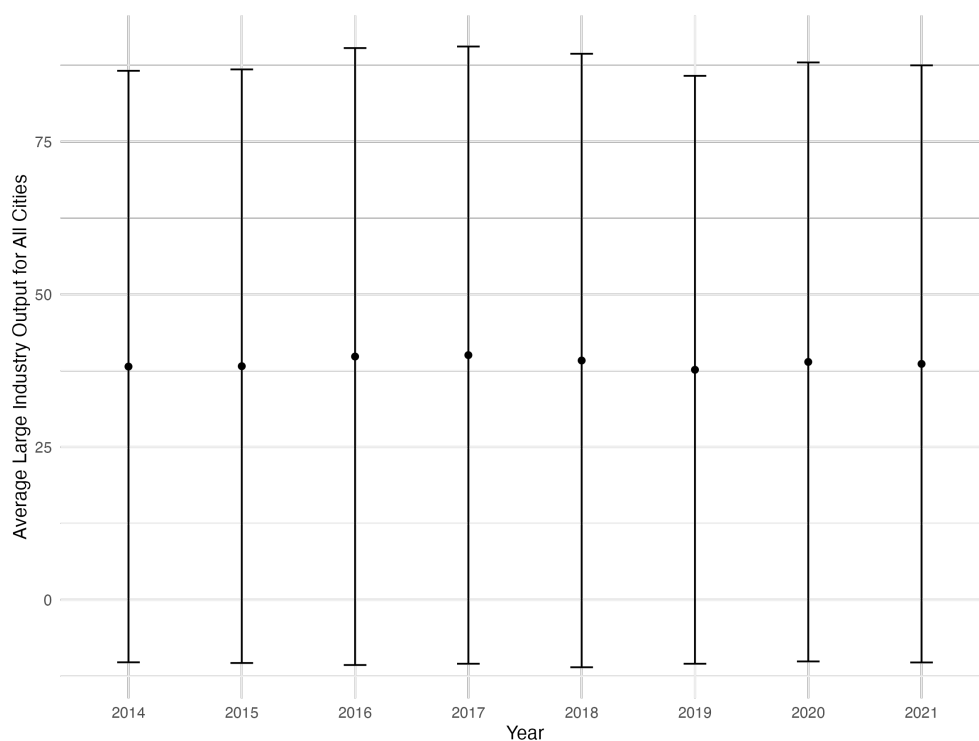
The remainder of this chapter is structured as follows. Section D.3 provides background information. Section D.4 describes the empirical data used in the analysis. Section D.5 presents the spatial autocorrelation results based on Moran's I tests. Section D.6 analyses the estimation results. Finally, Section D.7 discusses the findings and concludes this replication.

D.3 Background

City-level data offers higher spatial resolution, enabling a more granular analysis of economic heterogeneity within regions. This level of detail captures localised economic activities and provides a clearer picture of the dynamic characteristics of urban economies, including shifts in labour and capital flows. These dynamics are more readily observable at the city level. For instance, city-level data can elucidate variations in industrial activity between neighbouring cities and illustrate how road accessibility facilitates or constrains these interactions.

Moreover, city-level research is particularly effective at identifying spatial spillover effects. These effects, transmitted through mechanisms such as economic activity, population mobility, and infrastructure development, are more discernible when analysed at a finer scale. For example, industrial growth in one city may stimulate economic expansion in adjacent cities through market integration and labour redistribution.

Figure D.1: Average Industry Output



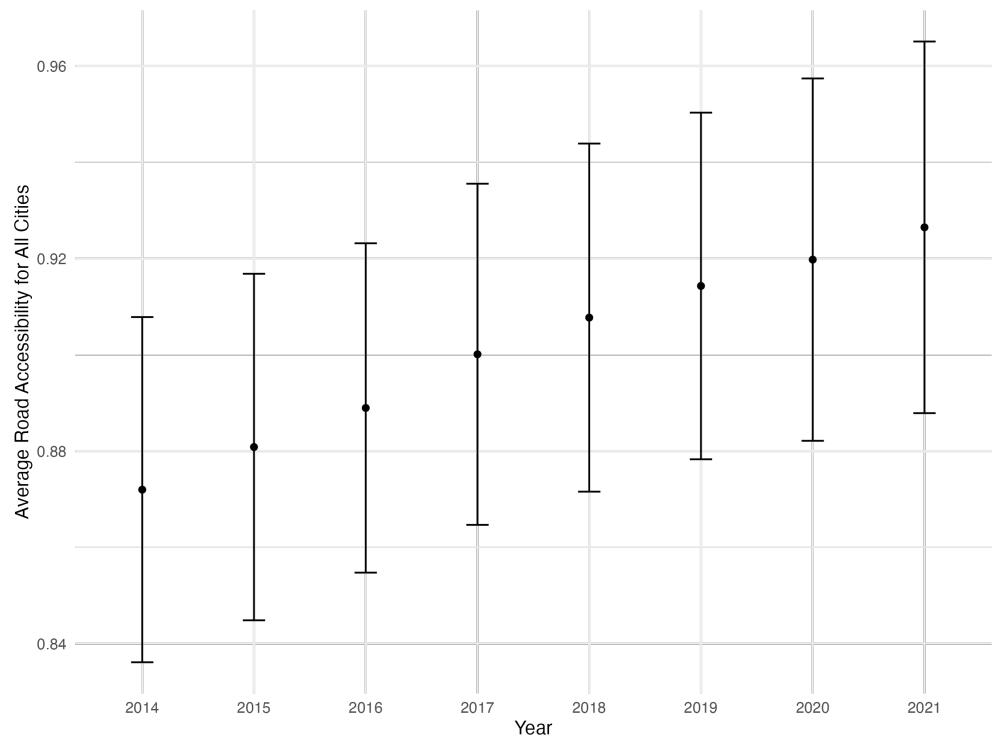
Note: This graph illustrates the average large industry output in 100 millions Yuan.

Similarly, labour spillovers from metropolitan areas can enhance productivity in surrounding regions. City-level analysis provides the detail necessary to examine these channels, offering valuable insights into why road infrastructure often lacks pronounced positive spillover effects in broader provincial-level analyses.

The accompanying chart is a time series line graph that illustrates the average industrial production across 288 cities from 2014 to 2021. Overall, industrial production exhibits a relatively steady upward trend. While the data varies slightly from year to year, this suggests that economic activity is relatively stable and exhibits limited volatility.

Although the annual growth appears consistent, certain years, such as 2020, show a slight slowdown in growth, potentially influenced by external economic shocks, such as the global pandemic, or other contributing factors.

Figure D.2: Average Road Accessibility to Big Cities



Note: This graph illustrates the average road accessibility to big cities road network

The chart illustrates road accessibility over the years from 2014 to 2021. The horizontal axis represents the timeline, while the vertical axis displays the accessibility index, which ranges from 0.84 to 0.96. The chart shows an overall gradual upward trend in road accessibility during the observed period, reflecting continuous improvement in transportation infrastructure.

The increase in road accessibility is steady, with no significant spikes or declines, suggest-

ing consistent progress in infrastructure development. However, the growth rate appears to plateau slightly after 2019, indicating a potential slowdown in the pace of further improvements.

D.4 Data Description

Our data is sourced from official records provided by the National Bureau of Statistics of China and the China City Yearbook.¹ Geographic and mapping data are derived from OpenStreetMap.

The distribution of industrial output exhibits substantial variation, with the maximum value approximately 836 times greater than the minimum. The mean and median values further highlight this disparity, indicating a heavily right-skewed distribution where a small number of regions demonstrate exceptionally high industrial output compared to others. This skewed distribution reflects pronounced regional inequality in industrial output, providing a diverse regional context for the analysis and enabling the exploration of how transportation accessibility exerts heterogeneous effects across regions with varying industrial profiles.

Most cities exhibit road accessibility costs close to the average, with limited deviation. However, extreme values highlight notable exceptions. The city with the lowest accessibility value of -62.321 incurs significantly higher costs than the average, reflecting poor accessibility. Conversely, the highest value of 25.424 corresponds to a city with exceptionally low road entry costs, indicating excellent accessibility to major urban networks. These extremes underline the heterogeneity in transportation conditions across cities, with some benefiting

¹National Bureau of Statistics. China City Statistical Yearbook Retrieved from: https://www.stats.gov.cn/zs/tjwh/tjkw/tjzl/202302/t20230220_1913734.html

from superior connectivity while others face considerable barriers.

We selected control variables to capture the potential impacts of regional economic and demographic characteristics on industrial output, ensuring the robustness and interpretability of the model results. Government expenditure and financial loans are key indicators of public service capacity and capital availability. Foreign investment inflows reflect the level of regional openness and access to industrial financing. Regional population and average employment figures serve as primary indicators of the labour market, illustrating the influence of labour supply and market demand on industrial output. Gross regional product, as a comprehensive measure of economic activity, represents the overall economic environment supporting industrial output. Education expenditure, meanwhile, reflects the level of human capital accumulation.

These control variables highlight significant differences in economic, demographic, and social development across regions. Variations in government expenditure and education expenditure underscore the uneven distribution of public resources, with some regions exhibiting substantially higher levels than others. The concentration of financial loans and foreign investment inflows demonstrates that economic activities are predominantly centralised in a few core areas, underscoring the dominant role of economic hubs in regional development. Similarly, the distribution of regional population and average employment figures reveals the spatial concentration of population and employment resources. The notable disparities in gross regional production further emphasise the imbalance in regional economic growth.

Table D.1: Summary Statistics of Key Variables

Variable Name	Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
Industry Output(in ten thousand)	395,436	10,961,799	20,616,633	38,510,490	42,667,611	330,379,747
Road Accessibility	-62.321	-0.423	-0.011	-0.007	0.429	25.424
Government Expenditure (in ten thousand)	216,246	2,163,509	3,198,776	4,846,060	4,850,729	83,556,553
Financeloan (in ten thousand)	1,059,064	7,899,064	13,882,538	40,884,324	29,869,327	874,727,453
Foreign Capital (in ten thousand)	3	6,428	27,257	96,605	86,206	3,022,121
Population (in ten thousand)	19.0	234.0	374.4	447.0	587.6	3,416.0
Average Employment	2.0	17.0	31.0	52.84	53.0	772.0
Regional Gross Production (in 100 Million yuan)	1,328,795	9,152,089	15,903,553	29,010,483	30,562,026	428,344,896
Education Expenditure (in ten thousand)	30,621	347,771	562,461	809,188	885,668	11,375,910

Note: All values are reported in their respective units. Variables are adjusted to account for CPI and other economic indicators where applicable.

D.5 Spatial Autocorrelation

The results of Moran’s I analysis for large industrial output and road accessibility across 288 cities over eight years from 2014 to 2021. We aim to understand the spatial dependence and clustering patterns of industrial production and road accessibility. Below is an interpretation of the findings.

The Moran’s I values are positive for all years and statistically significant at the 1% level. This indicates the presence of spatial autocorrelation in industrial production, meaning cities with similar levels of industrial output, either high or low, tend to cluster geographically. The Moran’s I values are relatively modest, ranging from 0.089 in 2014 to 0.042 in 2021. For instance, in 2014, a Moran’s I of 0.089 indicates moderate spatial dependence, while by 2021, the value drops to 0.042, reflecting a weaker clustering effect. The clustering presents the spatial dependence of industrial production, where neighbouring cities share similar economic or industrial characteristics.

The expected value of Moran’s I under spatial randomness is approximately -0.004, whereas the observed values are consistently positive. The high Z-scores, such as 14.903 in 2014 and 7.384 in 2021, confirm that the deviation from randomness is statistically significant.

The results show that the observed spatial clustering is highly unlikely to be due to random chance, and we need to consider the spatial effects in later estimation.

The Moran's I values are higher during these years from 0.089 to 0.099 during the first two years, indicating stronger spatial clustering. This could reflect industrial hubs and strong spillover effects in nearby regions during this period, possibly driven by concentrated investment or regional advantages. Then, there is a moderate decline in Moran's I values, as 0.064 in 2017 and 0.061 in 2018. A sharper decline to 0.044 in 2020 and 0.042 in 2021 may be linked to the economic disruptions caused by the COVID-19 pandemic. Lockdowns and supply chain disruptions likely weakened spatial spillovers.

To sum up, the Moran's I results reveal persistent but weakening spatial clustering in industrial production across cities over time. While geographic proximity continues to influence industrial patterns, its impact is diminishing due to economic decentralization, structural changes, and external disruptions.

Table D.2: Spatial Autocorrelation Test with Moran's I for Industry Output

Year	Moran's I (I)	Z-Statistic (Z)	Significance
2014	0.089	(14.903)	***
2015	0.097	(16.156)	***
2016	0.099	(16.547)	***
2017	0.064	(10.797)	***
2018	0.061	(10.352)	***
2019	0.064	(10.821)	***
2020	0.044	(7.719)	***
2021	0.042	(7.384)	***

Note: ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. Moran's I-statistics are based on a two-tailed test. The Z-statistics are provided in parentheses.

We also provide a spatial autocorrelation analysis for transportation accessibility. Moran's I result for road accessibility to the market across 288 cities over the same period, from

2014–2021, as following table.

Similar with industry production, the positive and statistically significant Moran’s I values at a 1% significant level for all years indicate the presence of spatial clustering in road accessibility. This means cities with similar levels of road accessibility are geographically clustered.

The Moran’s I values show a fluctuating but slightly increasing trend over time. Moran’s I start at 0.044 at 2014, indicating moderate clustering. A dip from 2015 to 2016 in Moran’s I to 0.025 suggests a temporary weakening of spatial dependence. From 2017 to 2021, an upward trend follows, with Moran’s I increasing to 0.044 by 2021. This is a return to stronger clustering. Also, the Z-scores are consistently high, such as 7.586 in 2014 and 7.658 in 2021, confirming that the observed clustering is statistically significant and unlikely to result from random spatial distribution.

Table D.3: Spatial Autocorrelation Test with Moran’s I for Road Accessibility

Year	Moran’s I (I)	Z-Statistic (Z)	Significance
2014	0.044	(7.586)	***
2015	0.025	(4.520)	***
2016	0.025	(4.446)	***
2017	0.034	(5.971)	***
2018	0.038	(6.528)	***
2019	0.042	(7.219)	***
2020	0.041	(7.152)	***
2021	0.044	(7.658)	***

Note: ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. Moran’s I-statistics are based on a two-tailed test. The Z-statistics are provided in parentheses.

Unlike industrial output, the spatial concentration of road accessibility, as measured by Moran’s I, appears to have remained largely unaffected by the COVID-19 pandemic. The Moran’s I values for road accessibility exhibit minimal variation during 2020 and 2021,

maintaining levels comparable to those observed in prior years. This contrasts with the significant weakening of spatial clustering in industrial output during the same period.

This stability is understandable, as road accessibility, being a physical and relatively fixed infrastructure system, is inherently less sensitive to short-term disruptions such as the pandemic. Unlike industrial output, which is influenced by fluctuating market demands, supply chain disruptions, and labour shortages, road networks retain a constant spatial distribution. This likely contributed to the sustained spatial patterns of road accessibility.

In this spatial autocorrelation analysis, the pandemic highlighted the relative resilience of road accessibility compared to industrial output. While industrial activity weakened and spatial clustering declined in response to economic shocks, the spatial concentration of road accessibility remained stable.

When compared to the provincial-level results, it becomes evident that the spatial autocorrelation of industrial production and road accessibility exhibits significantly weaker spatial connection effects at the city level. Whereas inter-provincial industrial production and transportation networks demonstrate stronger regional integration and cross-boundary influence, city-level dynamics are more localised and less influenced by neighbouring cities.

This difference can be attributed to several factors. Provincial-level analyses often capture broader patterns of economic clustering and infrastructure development, where provinces act as hubs of regional connectivity. Furthermore, provincial infrastructure frequently supports long-distance transportation and economic linkages, fostering stronger inter-regional dependencies. Conversely, city-level road networks primarily cater to local mobility needs, resulting in weaker spatial spillover effects. This observation is further corroborated by the subsequent results of the SDM test.

D.6 Spatial Estimation and Result

In the city-level analysis, we employ a geographic inverse distance weight matrix, consistent with the approach used in our provincial-level study. This weighting method captures spatial relationships by assigning greater influence to geographically closer cities while diminishing the impact of those farther away. For the modelling framework, we use the SDM model, which accounts for both direct and spillover effects, offering a comprehensive understanding of the spatial dependencies and interactions at the city level. This approach ensures comparability between the provincial and city-level analyses.

Table D.4: Regression Results

Variable	Estimate	Std. Error	t-value	Significance
Road Accessibility	-0.0116	0.0043	-2.7215	**
logGOVE	-0.1765	0.1310	-1.3478	
logFinanciaLoad	-0.0971	0.0810	-1.1992	
logForeignCapital	0.1098	0.0111	9.8540	***
logRegPop	0.0448	0.0897	0.4997	
logAvgEmp	0.1395	0.0732	1.9055	*
logRGDP	0.1465	0.0949	1.5438	
logEduExp	-0.1389	0.1276	-1.0886	
Wlag.Road Accessibility	0.0784	0.0577	1.3596	
Wlag.logGOVE	0.2781	0.5319	0.5227	
Wlag.logFinanciaLoad	-0.2626	0.2285	-1.1490	
Wlag.logForeignCapital	-0.0099	0.0496	-0.1998	
Wlag.logRegPop	0.1108	0.2402	0.4615	
Wlag.logAvgEmp	0.2215	0.2083	1.0634	
Wlag.logRGDP	0.1325	0.1834	0.7224	
Wlag.logEduExp	-0.1932	0.6067	-0.3185	
rho	-0.3684	0.2241	-1.6438	*

Note: ***, **, and * indicate significance at the 1%, 5%, and 10% levels.

The results indicate that road accessibility has a significant direct effect on regional industrial output. However, the estimated coefficient is -0.0116, significant at the 5% level, suggesting that improvements in road accessibility may negatively impact industrial output. This

finding aligns with the results observed in the provincial-level analysis.

Regarding the control variables, foreign capital inflows exhibit a coefficient of 0.1098, significant at the 1% level, indicating a strong positive impact of foreign investment on industrial output. The concentration of foreign capital appears to substantially enhance regional industrial productivity. Similarly, the coefficient for average employment is 0.1395, significant at the 10% level, highlighting the critical role of the labour market in supporting industrial development.

For spatial lag effects, the spatial lag of road accessibility has a coefficient of 0.0784 but is not statistically significant. This suggests that the indirect effects of road accessibility in neighbouring regions on local industrial output are relatively weak. The results imply that the spillover effects of road accessibility are not prominent within the current model. Other spatially lagged variables, such as foreign capital inflows and average employment, similarly fail to exhibit significant spatial spillover effects, indicating that industrial output across regions relies more on local characteristics than on interregional interactions.

Notably, the spatial autocorrelation coefficient, ρ , is -0.3684 and significant at the 10% level. This result reveals a degree of spatial dependency in regional industrial output, with the negative spatial autocorrelation indicating that regions with high industrial output are often surrounded by areas with relatively low output.

The spatial effects analysis based on the Spatial Durbin Model reveals distinct direct, indirect, and total effects of road accessibility, foreign capital inflows, and employment on regional industrial output.

First, the direct effect of road accessibility is -0.0116, significant at the 5% level, indicating that improved transportation accessibility may have a negative impact on local industrial

Table D.5: Spatial Analysis

Variable	Total		Direct		Indirect	
	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error
Road Accessibility	0.0504	0.0468	-0.0116	0.0043 **	0.0620	0.0465
logGOVE	0.0730	0.4304	-0.1790	0.1307	0.2521	0.4158
logFinanceloan	-0.2695	0.1968	-0.0942	0.0819	-0.1753	0.1852
logForeignCapital	0.0746	0.0426 *	0.1102	0.0111 ***	-0.0355	0.0415
logRegPop	0.1124	0.1968	0.0407	0.0916	0.0717	0.1844
logAvgEmp	0.2765	0.1775	0.1417	0.0736 *	0.1349	0.1712
logRGDP	0.2111	0.1634	0.1463	0.0951	0.0648	0.1470
logEduExp	-0.2528	0.4748	-0.1447	0.1240	-0.1080	0.4633

Note: ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

output. This result aligns with the provincial-level findings and may reflect the potential for improved road networks to lead to resource dispersion or a stronger reliance on alternative modes of transportation for industrial activities.

The spatial effects of foreign capital inflows are more pronounced. The direct effect is 0.1102, significant at the 1% level, indicating a strong positive influence of foreign investment on local industrial output. However, the indirect effect is -0.0355, though not statistically significant, suggesting that foreign capital inflows in neighbouring regions may create competitive pressures on local markets or resources. The total effect is 0.0746, significant at the 10% level, demonstrating that the overall impact of foreign capital inflows remains predominantly positive.

Average employment shows a direct effect of 0.1417, significant at the 10% level, highlighting the critical role of labour market expansion in boosting local industrial output. The indirect effect is 0.1349, which, although not statistically significant, indicates a potential positive spillover effect from improved employment levels in neighbouring regions on local industrial performance.

Among other variables, government expenditure and gross regional product do not exhibit significant direct or indirect effects. However, the positive signs of their indirect effects suggest the presence of potential spillover mechanisms between regions. Education expenditure also shows no significant effects. The short-term impact of these variables on industrial output is limited.

D.7 Conclusion

We conducted a systematic analysis of the impact of road transportation on regional industrial production across two spatial scales: the interprovincial and the city level. At the provincial scale, road transportation exhibited a significant positive effect, indicating that improved inter-regional connectivity facilitates market integration, reduces logistics costs, and enhances the spatial reallocation of industrial activities. These findings underscore the positive spillover effects of road infrastructure across regions, which are essential for promoting balanced regional growth.

In contrast, the city-level analysis, based on 2,304 observations from 288 prefecture-level cities, revealed a localized negative effect of urban road systems on industrial production. To ensure cross-scale comparability, we refined the measurement of transportation exposure by focusing on accessibility to major urban freight networks, rather than total road length. This measure captures the functional relevance of roads to industrial logistics.

This pattern further illustrates the scale-dependent mechanisms through which road infrastructure affects regional economies. At the urban level, roads primarily function as conduits for factor mobility, often reinforcing a “core-periphery” siphoning effect whereby labour and capital are drawn toward larger urban centres. This dynamic is consistent with findings from

Baum-Snow et al. (2017), who show that labour mobility in China is significantly shaped by road networks, with workers tending to move from smaller, disadvantaged cities to more developed areas. Similarly, Banerjee et al. (2020) emphasize the role of road connectivity in facilitating the direct movement of production factors between small and large cities within the same region, without necessarily generating proportional gains in per capita GDP growth rate. In contrast, at the interprovincial or national level, the role of road infrastructure becomes more integrative, supporting market consolidation, inter-regional trade, and productive specialization. These broader effects help explain why transportation investments tend to exhibit positive regional outcomes at larger spatial scales, while producing more polarized or uneven effects within urban systems. As such, the evaluation of transportation impacts must account for spatial scale, city hierarchy, and the structure of factor flows, rather than treating infrastructure as uniformly beneficial.

In sum, while interprovincial road infrastructure contributes positively to industrial output by enhancing regional-level economic integration, urban road expansion may generate localized inefficiencies that hinder productivity. These contrasting results across spatial scales highlight the dual nature of transportation infrastructure as a facilitator of macro-regional growth.

We consider this a particularly intriguing finding, and although we only provide a preliminary discussion here, we plan to explore the underlying mechanisms in greater depth in future research.

Bibliography

Gabriel M. Ahlfeldt and Arne Feddersen. From periphery to core: measuring agglomeration effects using high-speed rail. *Journal of Economic Geography*, 18(2):355–390, 2018. doi: 10.1093/jeg/lbx005.

Daniel Albalade and Xavier Fageda. High speed rail and tourism: Empirical evidence from spain. *Transportation Research Part A: Policy and Practice*, 85:174–185, 2016. doi: 10.1016/j.tra.2016.01.009.

Luc Anselin. *Spatial Econometrics: Methods and Models*. Springer Science & Business Media, 1988.

Luc Anselin. Local indicators of spatial association—lisa. *Geographical Analysis*, 27(2): 93–115, 1995. doi: 10.1111/j.1538-4632.1995.tb00338.x.

Manuel Arellano and Stephen Bond. Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *Review of Economic Studies*, 58: 277–297, 1991.

David Audretsch and Maryann P. Feldman. R&d spillovers and the geography of innovation and production. *American Economic Review*, 86(3):630–640, 1996. URL <http://links>.

jstor.org/sici?sici=0002-8282%2819960...O%3B2-Z&origin=repec. Access to full text is restricted to JSTOR subscribers.

Badi H. Baltagi. *Econometric Analysis of Panel Data*. John Wiley & Sons, 5th edition, 2013.

Badi H. Baltagi, Bernard Fingleton, and Alain Pirotte. A time-space dynamic panel data model with spatial moving average errors. *Regional Science and Urban Economics*, 76: 13–31, 2019.

Abhijit Banerjee, Esther Duflo, and Nancy Qian. On the road: Access to transportation infrastructure and economic growth in china. *Journal of Development Economics*, 2020. doi: 10.1016/j.jdeveco.2020.102442.

Nathaniel Baum-Snow, Loren Brandt, J. Vernon Henderson, Matthew A. Turner, and Qinghua Zhang. Roads, railroads and decentralization of chinese cities. *The Review of Economics and Statistics*, 99(3):435–448, 2017. doi: 10.1162/REST_a_00660.

K. Behrens and P. M. Picard. Transportation, freight rates, and economic geography. *Journal of International Economics*, 87(2):275–284, 2011. doi: 10.1016/j.jinteco.2011.03.005. URL <https://doi.org/10.1016/j.jinteco.2011.03.005>.

Roger S. Bivand. *Applied Spatial Data Analysis with R*. Springer, 2008.

John P. Blair and Michael C. Carroll. *Local Economic Development: Analysis, Practices, and Globalization*. SAGE Publications, Inc., 2 edition, 2009. doi: 10.4135/9781452275109. Online publication date: January 18, 2013.

Chiara F. Del Bo and Massimo Florio. Infrastructure and growth in a spatial framework:

- Evidence from the eu regions. *European Planning Studies*, 20:1393–1414, 2012. doi: 10.1080/09654313.2012.680587.
- G. Brancaccio, M. Kalouptsi, and T. Papageorgiou. Geography, transportation, and endogenous trade costs. *Econometrica*, 88(2):657–691, 2020. doi: 10.3982/ECTA14065. URL <https://doi.org/10.3982/ECTA14065>.
- Amitabh Chandra and Eric Thompson. Does public infrastructure affect economic activity? evidence from the rural interstate highway system. *Regional Science and Urban Economics*, 30(4):457–490, 2000. doi: 10.1016/S0166-0462(00)00040-5.
- Thanyaporn Chankrajang and Jittitip Vechbanyongratana. Canals and orchards: The impact of transport network access on agricultural productivity in nineteenth-century bangkok. *The Journal of Economic History*, 80(4):996–1030, 2020. doi: 10.1017/s0022050720000509.
- Susan Christopherson, Jonathan Michie, and Peter Tyler. Regional resilience: Theoretical and empirical perspectives. *Cambridge Journal of Regions, Economy and Society*, 3(1): 3–10, 2010. doi: 10.1093/cjres/rsq004.
- Yongwan Chun, Daniel A. Griffith, Minjee Lee, and Piyushimita Sinha. Eigenvector selection with stepwise regression techniques to construct eigenvector spatial filters. *Journal of Geographical Systems*, 18(1):67–85, 2016. URL <https://doi.org/10.1007/s10109-015-0225-3>.
- Andrew D. Cliff and J. K. Ord. *Spatial Processes: Models & Applications*. Pion, 1981.
- Jeffrey P. Cohen and Catherine J. Morrison Paul. Public infrastructure investment, interstate spatial spillovers, and manufacturing costs. *Review of Economic Statistics*, 86(2):551–560, 2004.

- Paul Collier, Peter Jones, and Susan Parnell. Cities as engines of economic growth: The case for investing in urban infrastructure, 2018. Retrieved from <https://thedocs.worldbank.org/en/doc/356001527999188468-0010022018/original/E1CitiesasEnginesofGrowthCollieretal28March2018.pdf>.
- David Cuberes, Klaus Desmet, and Jordan Rappaport. Urban growth shadows. *Regional Science and Urban Economics*, 90:Article 103701, 2021. doi: 10.1016/j.regsciurbeco.2021.103701. URL <https://doi.org/10.1016/j.regsciurbeco.2021.103701>.
- Tao Deng. Impacts of transport infrastructure on productivity and economic growth: Recent advances and research challenges. *Transport Reviews*, 33(6):686–699, 2013. doi: 10.1080/01441647.2013.851745.
- Chris Ding and Xiaofeng He. K-means clustering via principal component analysis. In *Proceedings of the Twenty-first International Conference on Machine Learning*, 2004. URL <https://api.semanticscholar.org/CorpusID:11356277>.
- D. Donaldson. Railroads of the raj: Estimating the impact of transportation infrastructure. *American Economic Review*, 108(4-5):899–934, 2018. doi: 10.1257/aer.20101199. URL <https://doi.org/10.1257/aer.20101199>.
- D. Donaldson and R. Hornbeck. Railroads and american economic growth: A “market access” approach. *The Quarterly Journal of Economics*, 131(2):799–858, 2016. doi: 10.1093/qje/qjw002. URL <https://doi.org/10.1093/qje/qjw002>.
- Justin Doran and Bernard Fingleton. Us metropolitan area resilience: Insights from dynamic spatial panel estimation. *Environment and Planning A*, 50(1):111–132, February 2018.
- Gilles Duranton and Matthew A. Turner. Urban growth and transportation. *The Review*

- of Economic Studies*, 79(4):1407–1440, 2012. doi: 10.1093/restud/rds010. Published: 12 March 2012.
- J. Paul Elhorst. *Spatial Econometrics: From Cross-sectional Data to Spatial Panels*. Springer, 2014.
- Simon J. Evenett and Wolfgang Keller. On theories explaining the success of the gravity equation. *Journal of Political Economy*, 110(2):281–316, 2002.
- Benjamin Faber. Trade integration, market size, and industrialization: Evidence from china’s national trunk highway system. *The Review of Economic Studies*, 81:1046–1070, 2014.
- Zhou Feng, Dong Liu, and Yi Yang. Evaluation of transportation ability of china: From county to province level. *Geographical Research*, 28(2):419–429, 2009.
- Bernard Fingleton. Externalities, economic geography, and spatial econometrics: Conceptual and modeling developments. *International Regional Science Review*, 26(2):197–207, 2003.
- Bernard Fingleton. Estimating dynamic spatial panel data models with endogenous regressors using synthetic instruments. *Environment and Planning A*, 45(10):2227–2242, 2022.
- Bernard Fingleton and Julie Le Gallo. Estimating spatial models with endogenous variables, a spatial lag and spatially dependent disturbances: Finite sample properties. *Papers in Regional Science*, 87(3):319–339, 2008. doi: 10.1111/j.1435-5957.2008.00187.x.
- Andrés Forero, Francisco A. Gallego, Felipe González, and Matías Tapia. Railroads, specialization, and population growth: Evidence from the first globalization. *Journal of Population Economics*, 34:1027–1072, 2020.

- Koen Frenken, Frank Van Oort, and Thijs Verburg. Related variety, unrelated variety and regional economic growth. *Regional Studies*, 41(5):685–697, 2007. doi: 10.1080/00343400601120296.
- Masahisa Fujita and Jacques-François Thisse. *Economics of Agglomeration: Cities, Industrial Location, and Regional Growth*. Cambridge University Press, Cambridge, 2002.
- Julie Le Gallo and Antonio Páez. Using synthetic variables in instrumental variable estimation of spatial series models. *Environment and Planning A*, 45:2227–2242, 2013. URL <https://doi.org/10.1068/a45457>.
- Andrew Gelman, Jennifer Hill, and Aki Vehtari. *Regression and Other Stories*. Cambridge University Press, 2021.
- Arthur Getis and Jared Aldstadt. Constructing the spatial weights matrix using a local statistic. *Geographical Analysis*, 36(2):90–104, 2004.
- Karst T. Geurs and Bert van Wee. Accessibility evaluation of land-use and transport strategies: Review and research directions. Technical report, Netherlands Environmental Assessment Agency, National Institute for Public Health and the Environment, Bilthoven BA 3720, The Netherlands; Delft University of Technology, Delft, The Netherlands, 2004.
- Stephen Gibbons, Teemu Lyytikäinen, Henry G. Overman, and Raquel Sanchis-Guarner. New road infrastructure: The effects on firms. *Journal of Urban Economics*, 110:35–50, 2019. doi: 10.1016/j.jue.2019.01.002.
- William H. Greene. *Econometric Analysis*. Prentice Hall, 7th edition, 2012.
- Daniel A. Griffith. Spatial autocorrelation and eigenfunctions of the geographic weights

- matrix accompanying geo-referenced data. *The Canadian Geographer*, 40(4):351–367, 1996. URL <https://doi.org/10.1111/j.1541-0064.1996.tb00462.x>.
- Daniel A. Griffith. *Spatial Autocorrelation and Spatial Filtering: Gaining Understanding Through Theory and Scientific Visualization*. Springer, Berlin, 2003.
- Daniel A. Griffith and Yongwan Chun. Spatial autocorrelation and spatial filtering. In Manfred M. Fischer and Peter Nijkamp, editors, *Handbook of Regional Science*, pages 1477–1507. Springer, Berlin, Heidelberg, 2014. URL https://doi.org/10.1007/978-3-642-23430-9_72.
- Kurt A. Hafner, Jörn Kleinert, and Julia Spies. Endogenous transport costs and international trade. *The World Economy*, 43:1333–1370, 2022. doi: 10.1111/twec.13337. First published: 29 September 2022.
- Lars Peter Hansen. Large sample properties of generalized method of moments estimators. *Econometrica*, 50:1029–1054, 1982.
- M. Hesse and J.-P. Rodrigue. The transport geography of logistics and freight distribution. *Journal of Transport Geography*, 12(3):171–184, 2004.
- Christian Hundt and Lea Holtermann. The role of national settings in the economic resilience of regions: Evidence from recessionary shocks in europe from 1990 to 2014. *Growth and Change*, 51(1):180–206, 2020.
- Anil K. Jain. Data clustering: 50 years beyond k-means. *Pattern Recognition Letters*, 31(8): 651–666, 2010. doi: 10.1016/j.patrec.2009.09.011. URL <https://doi.org/10.1016/j.patrec.2009.09.011>.
- F. Jin and Y. Wang. Gmm estimation of a spatial autoregressive model with autoregressive

- disturbances and endogenous regressors. *Econometric Reviews*, 41(6):652–674, 2021. doi: 10.1080/07474938.2021.2002521. Published online: 22 Nov 2021.
- Ian T. Jolliffe. *Principal Component Analysis*. Springer Series in Statistics. Springer, 2002.
- S. Juhl. spfilter: An r package for semiparametric spatial filtering with eigenvectors in (generalized) linear models. *Journal of Statistical Software*, 98(1):1–291, 2022.
- Leonard Kaufman and Peter J. Rousseeuw. *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley & Sons, 2009.
- P. L. Kent. *Freight Transport for Development Toolkit: Road Freight*. World Bank Group, Washington, D.C., 2010. URL <http://documents.worldbank.org/curated/en/649241468155983939/Freight-transport-for-development-toolkit-road-freight>. Retrieved from World Bank Group website.
- David J. Ketchen and Christopher L. Shook. The application of cluster analysis in strategic management research: An analysis and critique. *Strategic Management Journal*, 17(6): 441–458, 1996. doi: 10.1002/(SICI)1097-0266(199607)17:6<441::AID-SMJ819>3.0.CO;2-G. URL [https://doi.org/10.1002/\(SICI\)1097-0266\(199607\)17:6<441::AID-SMJ819>3.0.CO;2-G](https://doi.org/10.1002/(SICI)1097-0266(199607)17:6<441::AID-SMJ819>3.0.CO;2-G).
- Gaurav Khanna, Carlos Medina, Anant Nyshadham, Daniel Ramos-Menchelli, Jorge Tamayo, and Audrey Tiew. Spatial mobility, economic opportunity, and crime. Working paper, November 2022. URL https://scholar.harvard.edu/files/danielramosmenchelli/files/cablepaper_2022.pdf.
- Akihiro Konno, Hiroki Kato, Wataru Takeuchi, and Ryo Kiguchi. Global evidence on pro-

- ductivity effects of road infrastructure incorporating spatial spillover effects. *Transport Policy*, 103:167–182, 2021. doi: 10.1016/j.tranpol.2021.02.007.
- Paul Krugman. Increasing returns and economic geography. *Journal of Political Economy*, 99(3):483–499, 1991.
- Paul Krugman. The role of geography in development. *International Regional Science Review*, 22(2):142–161, 1999. doi: 10.1177/016001799761012307.
- Somik V. Lall, Zmarak Shalizi, and Uwe Deichmann. Agglomeration economies and productivity in indian industry. *Journal of Development Economics*, 73(2):643–673, 2004.
- L. F. Lee. Gmm and 2sls estimation of mixed regressive, spatial autoregressive models. *Journal of Econometrics*, 137(2):489–514, 2007. doi: 10.1016/j.jeconom.2005.10.004. URL <https://doi.org/10.1016/j.jeconom.2005.10.004>.
- James P. LeSage and R. Kelley Pace. *Introduction to Spatial Econometrics*. Taylor & Francis CRC Press, Boca Raton, 2009.
- Andrew Levin, Chien-Fu Lin, and Chia-Shang J. Chu. Unit root tests in panel data: Asymptotic and finite-sample properties. *Journal of Econometrics*, 108(1):1–24, 2002. doi: 10.1016/S0304-4076(01)00098-7.
- Xiaolong Li, Zongfa Wu, and Xingchen Zhao. Economic effect and its disparity of high-speed rail in china: A study of mechanism based on synthesis control method. *Transport Policy*, 99:262–274, December 2020.
- X. Liu and P. Saraiva. Gmm estimation of sar models with endogenous regressors. *Regional Science and Urban Economics*, 55(5–6):68–79, 2015.

- Stuart Lloyd. Least squares quantization in pcm. *IEEE Transactions on Information Theory*, 28(2):129–137, 1982. doi: 10.1109/TIT.1982.1056489.
- J. MacQueen. Some methods for classification and analysis of multivariate observations. In Lucien M. Le Cam and Jerzy Neyman, editors, *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 5.1, pages 281–297, 1967.
- Alfred Marshall. *Principles of Economics*. Macmillan, 1st edition, 1890.
- Alfred Marshall. *Principles of Economics*. Macmillan, London, 8th edition, 1920.
- Ron Martin. Regional economic resilience, hysteresis and recessionary shocks. *Journal of Economic Geography*, 12(1):1–32, 2012. doi: 10.1093/jeg/lbr019. URL <https://doi.org/10.1093/jeg/lbr019>.
- Sara Meerow, Joshua P. Newell, and Melissa Stults. Defining urban resilience: A review. *Landscape and Urban Planning*, 147:38–49, 2016.
- Guy Michaels. The effect of trade on the demand for skill: Evidence from the interstate highway system. *Review of Economics and Statistics*, 90(4):683–701, 2008. doi: 10.1162/rest.90.4.683.
- Enrico Moretti. The effect of high-tech clusters on the productivity of top inventors. *American Economic Review*, 111(10):3328–3375, 2021. doi: 10.1257/aer.20191277.
- National Bureau of Statistics. 2019 national economic and social development statistical bulletin of the people’s republic of china, 2020. Retrieved from: <http://www.stats.gov.cn/sj/zxfb/202302/t202302031900640.html>.
- R. Kelley Pace, James P. LeSage, and Shuang Zhu. Spatial dependence in regressors and its

- effect on performance of likelihood-based and instrumental variable estimators. In *30th Anniversary Edition*. Emerald Group Publishing Limited, 2012. ISBN 978-1-78190-309-4. Publication date: 19 December 2012.
- Antony Potter and H. Doug Watts. Evolutionary agglomeration theory: Increasing returns, diminishing returns, and the industry life cycle. *Journal of Economic Geography*, 11: 417–455, 2011. doi: 10.1093/jeg/lbq004. Advance Access Published on 24 March 2010.
- Diego Puga. European regional policies in light of recent location theories. *Journal of Economic Geography*, 2(4):373–406, 2002. doi: 10.1093/jeg/2.4.373. Published: 01 October 2002.
- Yu Qin. ‘no county left behind?’ the distributional impact of high-speed rail upgrades in china. *Journal of Economic Geography*, 2016. doi: 10.1093/jeg/lbw013.
- Iftikhar U. Rahman, B. P. Sharma, Elvis Fetuu, and Muhammad Yousaf. Do roads enhance regional trade? evidence based on china’s provincial data. *The Journal of Asian Finance, Economics and Business*, 7(12):657–664, 2020. doi: 10.13106/jafeb.2020.vol7.no12.657.
- Sergio J. Rey and Brett D. Montouri. Us regional income convergence: A spatial econometric perspective. *Regional Studies*, 33(2):143–156, 1999.
- Jean-Paul Rodrigue. Globalization and the synchronization of transport terminals. *Journal of Transport Geography*, 7:255–261, 1999.
- Jean-Paul Rodrigue. *The Geography of Transport Systems (Sixth Edition)*. Routledge, New York, 2024. ISBN 9781032380407. doi: 10.4324/9781003343196. URL <https://doi.org/10.4324/9781003343196>.
- Jean-Paul Rodrigue, Claude Comtois, and Brian Slack. *The Geography of Transport Systems*.

- Routledge, London and New York, 2006. URL https://geonas.at.ua/_ld/0/34_The_Geography_o.pdf. Retrieved from https://geonas.at.ua/_ld/0/34_The_Geography_o.pdf.
- David Roodman. How to do xtabond2: an introduction to difference and system gmm in stata. *Stata Journal*, 9(1):86–136, 2009.
- Peter J. Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53–65, 1987. doi: 10.1016/0377-0427(87)90125-7.
- A. Sharif, M. Shahbaz, and E. Hille. The transportation-growth nexus in usa: Fresh insights from pre-post global crisis period. *Transportation Research Part A: Policy and Practice*, 121:108–121, 2019.
- Donald Shepard. A two-dimensional interpolation function for irregularly-spaced data. In *Proceedings of the 1968 ACM National Conference*, pages 517–524, 1968. doi: 10.1145/800186.810616.
- James Simmie and Ron Martin. The economic resilience of regions: Towards an evolutionary approach. *Cambridge Journal of Regions, Economy and Society*, 3(1):27–43, 2010. doi: 10.1093/cjres/rsp029. Published: 06 January 2010.
- Jemima Sutton and Godwin Arku. Regional economic resilience: Towards a system approach. *Regional Studies, Regional Science*, 9(1):497–512, 2022. doi: 10.1080/21681376.2022.2092418. URL <https://doi.org/10.1080/21681376.2022.2092418>.
- Jesse Sutton, Alessia Arcidiacono, Gianpiero Torrissi, and Robert Nutifafa Arku. Regional economic resilience: A scoping review. *Progress in Human Geography*, 47(4):500–532,

2023. doi: 10.1177/03091325231174183. URL <https://journals.sagepub.com/doi/10.1177/03091325231174183>.

Robert L. Thorndike. Who belongs in the family? *Psychometrika*, 18(4):267–276, 1953. doi: 10.1007/BF02289263. URL <https://doi.org/10.1007/BF02289263>.

Michael Tiefelsdorf. *Modelling Spatial Processes: The Identification and Analysis of Spatial Relationships in Regression Residuals by Means of Moran's I*. Springer-Verlag, Berlin; New York, 2000.

Michael Tiefelsdorf and Barry Boots. The exact distribution of moran's i. *Environment and Planning A: Economy and Space*, 27(6):985–999, 1995.

Michael Tiefelsdorf and Daniel A. Griffith. Semiparametric filtering of spatial autocorrelation: The eigenvector approach. *Environment and Planning A: Economy and Space*, 39(5):1193–1221, 2007. URL <https://doi.org/10.1068/a37378>.

Waldo R. Tobler. A computer movie simulating urban growth in the detroit region. *Economic Geography*, 46(sup1):234–240, 1970.

Tiebin Tong, Tsai-Hsuan E. Yu, Seong-Hoon Cho, Kim Jensen, and Daniel De La Torre Ugarte. Evaluating the spatial spillover effects of transportation infrastructure on agricultural output across the united states. *Journal of Transport Geography*, 30:47–55, 2013. doi: 10.1016/j.jtrangeo.2013.03.001.

Samuel Washington, Matthew Karlaftis, Fred Mannering, and Panagiotis Anastasopoulos. *Statistical and Econometric Methods for Transportation Data Analysis*. CRC Press, 2020.

Frank Windmeijer. A finite sample correction for the variance of linear efficient two-step

gmm estimators. *Journal of Econometrics*, 126(1):25–51, 2005. doi: 10.1016/j.jeconom.2004.02.005. Published: January 2005.

Jeffrey M. Wooldridge. *Econometric Analysis of Cross Section and Panel Data*. MIT Press, 2010.

Ning Yu, Martin de Jong, Servaas Storm, and Jian Mi. Spatial spillover effects of transport infrastructure: Evidence from chinese regions. *Journal of Transport Geography*, 28:56–66, 2013.

Fang Zhang, Zhen Yang, Jie Jiao, Wei Liu, and Wei Wu. The effects of high-speed rail development on regional equity in china. *Transportation Research Part A: Policy and Practice*, 141:180–202, 2020. doi: 10.1016/j.tra.2020.09.013. URL <https://doi.org/10.1068/a270985>.

Jing Zhang, Tangwei Teng, Shengpeng Wang, and Gang Zeng. The structure of product space and regional economic resilience in china: Towards an evolutionary perspective. *Applied Economics*, 2023. doi: 10.1080/00036846.2023.2299219. Published: 30 December 2023.