

# KOF Dissertation Series

## Essays in Economic Geography

Nicole Loumeau

Diss. ETH No. 26790

KOF Dissertation Series, No. 43, 2020

# Imprint

## **Publisher**

KOF Swiss Economic Institute, ETH Zurich

© 2020 Nicole Loumeau

DISS. ETH NO. 26790

# ESSAYS IN ECONOMIC GEOGRAPHY

A thesis submitted to attain the degree of  
DOCTOR OF SCIENCES of ETH ZURICH  
(Dr. sc. ETH Zurich)

presented by  
NICOLE LOUMEAU

M.A. in Philosophy and Economics, University of Bayreuth  
B.Sc. in International Economics, University of Tübingen

born on 20.04.1988  
citizen of  
Germany

accepted on the recommendation of

PROF. DR. PETER H. EGGER, examiner  
PROF. DR. DANIEL M. STURM, co-examiner

2020



*To my husband and my daughter.*



# Acknowledgments

First and foremost, I would like to express my sincere gratitude to Peter Egger for giving me the opportunity to write this dissertation under his supervision. I am indebted to him for his continuous support and advice over the years. I would also like to thank my thesis committee, Professor Daniel Sturm and Professor Marko Köthenbürger, for taking the time to read and comment on this dissertation. I am grateful for the financial support provided by the Swiss National Science Foundation.

I would like to thank all of my professors who taught me during my undergraduate and post-graduate studies. I am particularly grateful to Benedikt Heid, whose enthusiasm for science guided me towards pursuing a doctorate in economics. I would also like to thank my current and former colleagues at the Chair of Applied Economics: Innovation and Internationalization at ETH Zurich and KOF for their advice and their sustained open door policies. A special thanks goes to Sophia, who has been a wonderful and encouraging office mate throughout.

The path to complete this dissertation would not have been as smooth without the encouragement and guidance of my family and friends. Without them I would not be where I am today. I am especially grateful to my parents, Kerstin and Dietmar, for their continued patience and unconditional support. Thank you to my dearest friends - Miriam, Kira, Carmen and Sören - who kept believing in me throughout the ups and downs of this experience.

Last but certainly not least, I owe an enormous debt of gratitude to my husband, Gabriel. His endless patience, continuous encouragement, and most importantly his enduring love has kept me strong and made this dissertation possible. I cannot thank him enough for the invaluable support he has given me as a co-author, husband and father to our daughter Selma, who makes our lives so much brighter.

*Nicole Loumeau, April 2020*





# Contents

<b>Acknowledgments</b>	<b>i</b>
<b>Abstract</b>	<b>xiii</b>
<b>Zusammenfassung</b>	<b>xv</b>
<b>Introduction</b>	<b>1</b>
<b>1 The Economic Geography of Innovation</b>	<b>7</b>
1.1 Introduction . . . . .	7
1.2 The Model . . . . .	10
1.2.1 Innovation and Production . . . . .	10
1.2.2 Innovation and Total Employment . . . . .	12
1.2.3 Utility and Consumption . . . . .	13
1.2.4 Equilibrium in Each Period . . . . .	14
1.2.5 Balanced Growth Path . . . . .	15
1.3 Calibration of Key Model Parameters . . . . .	16
1.3.1 Delineation of Regions and Land Endowments . . . . .	18
1.3.2 Trade-cost-function Parameters . . . . .	19
1.3.3 Initial Efficiency Distribution . . . . .	20
1.3.4 Estimation of the Productivity Shifter for R&D Workers . . . . .	20
1.3.5 Technology and Efficiency-evolution Parameters . . . . .	26
1.3.6 Patented vs. Non-patented Innovations . . . . .	27
1.3.7 Estimating Amenity-function Parameters . . . . .	31
1.4 Counterfactual Analysis . . . . .	32
1.4.1 Economic Outcomes and R&D-policy Instruments . . . . .	33

1.4.2	The Role of Treatment Size, Remoteness, and Amenities for Welfare Responses . . . . .	35
1.4.3	The Role of the Patented Innovation Weight for Innovation Responses . . . . .	37
1.5	Conclusion . . . . .	39
	<b>Appendix</b> . . . . .	41
1.6	Estimation Table: Robustness . . . . .	41
1.7	Initial Efficiency and Amenity Distribution . . . . .	41
1.8	Equilibrium: Existence and Uniqueness . . . . .	43
1.9	Balanced Growth Path: Derivation . . . . .	46
1.9.1	Uniqueness and Existence Condition in the BGP . . . . .	46
1.9.2	Growth Rate of Aggregate Welfare . . . . .	48
1.10	Data Aggregation . . . . .	49
1.10.1	Population Data from SEDAC . . . . .	49
1.10.2	Population and GDP from G-Econ Project . . . . .	49
1.10.3	Transportation Costs based on the Fast-marching Algorithm . . . . .	50
<b>2</b>	<b>Natural City Growth in the People’s Republic of China</b>	<b>51</b>
2.1	Introduction . . . . .	51
2.2	Natural City Borders in China, 1992-2013 . . . . .	54
2.3	Drivers of Natural City Growth . . . . .	59
2.3.1	Data . . . . .	59
2.3.2	Descriptive Statistics . . . . .	62
2.3.3	Econometric Approach . . . . .	64
2.3.4	Results . . . . .	66
2.4	Conclusion . . . . .	75
	<b>Appendix</b> . . . . .	77
2.5	Supplement Tables . . . . .	77
<b>3</b>	<b>Decomposing the Economic Effects of Transport Infrastructure</b>	<b>87</b>
3.1	Introduction . . . . .	87
3.2	Theoretical Framework . . . . .	92
3.2.1	Setup . . . . .	92

3.2.2	Individual Preferences . . . . .	93
3.2.3	Technology and Production . . . . .	94
3.2.4	Trade . . . . .	95
3.2.5	Migration . . . . .	96
3.2.6	Equilibrium . . . . .	96
3.3	Transport Network . . . . .	99
3.3.1	Chinese Transport Infrastructure Network . . . . .	99
3.3.2	Optimal Transport Network . . . . .	102
3.4	Calibration . . . . .	107
3.4.1	Elasticity-of-substitution Parameter $\sigma$ . . . . .	110
3.4.2	Land Endowments $(H_i^R, H_i^C)$ . . . . .	110
3.4.3	Infrastructure Network Data and Network (Spillover) Weights $(d_{ijt}, \mathbb{W}_{ijt}^R, \mathbb{W}_{ijt}^C)$ . . . . .	111
3.4.4	Cobb-Douglas Shares in Preferences and Production $(\alpha, \mu)$ .	112
3.4.5	Migration-cost Parameters $(\Omega, \kappa_{ijt}, \phi_1)$ . . . . .	113
3.4.6	Trade-cost Parameters $(\theta, \zeta_{ijt}, \phi_2)$ . . . . .	115
3.4.7	Parameters Governing the Distribution of Endogenous Tech- nology $(T_{it}, \gamma_1, \gamma_2)$ . . . . .	118
3.4.8	Parameters Governing the Distribution of Endogenous Amenities $(a_{it}, \lambda_1, \lambda_2)$ . . . . .	120
3.5	Counterfactual Analysis . . . . .	121
3.5.1	Changes in Trade and Migration Costs and Initial Model State	122
3.5.2	Simulating Benchmark and Counterfactual Equilibrium Outcomes . . . . .	124
3.5.3	Long-run Economic Effects of a Counterfactual Change in the Infrastructure Between Years 2000 and 2013 . . . . .	126
3.5.4	Decomposing the Long-run Equilibrium Effects by Channel .	129
3.5.5	Decomposing the Long-run Equilibrium Effects by Road Type	133
3.6	Conclusion . . . . .	135
	<b>Appendix</b> . . . . .	137
3.7	Derivation: Equilibrium Equations . . . . .	137
3.8	Equilibrium: Existence and Uniqueness . . . . .	138

3.9	Equilibrium: Stability Condition . . . . .	143
3.10	Calibration: Trade-cost Parameters . . . . .	144
3.11	Data Source: Transportation Atlases . . . . .	147
<b>4</b>	<b>Capital Cities and Road Network Integration: Evidence from the</b>	
	<b>U.S.</b>	<b>149</b>
4.1	Introduction . . . . .	149
4.2	Historical Background . . . . .	154
4.2.1	U.S. American State Capitals . . . . .	154
4.2.2	The U.S. National Highway System . . . . .	159
4.3	Data Construction . . . . .	160
4.4	Empirical Strategy . . . . .	167
4.4.1	Identification . . . . .	167
4.4.2	Instrumental Variable Design . . . . .	168
4.4.3	Results . . . . .	172
4.5	Discussion . . . . .	174
4.6	Conclusion . . . . .	177
	<b>Appendix</b> . . . . .	<b>179</b>
4.7	Supplement Tables . . . . .	179
	<b>Bibliography</b>	<b>185</b>
	<b>Curriculum Vitae</b>	<b>199</b>

# List of Figures

1.1	REGPAT REGIONS . . . . .	18
1.2	KERNEL DENSITY OF THE ESTIMATED LOG R&D-WORKER-SPECIFIC PRODUCTIVITY SHIFTER . . . . .	26
1.3	LOG OVERALL INNOVATIVE PRODUCTIVITY VS. ESTIMATED REGION- SPECIFIC IMPORTANCE WEIGHT OF PATENTED INNOVATIONS (2005) .	30
1.4	LOG OVERALL INNOVATIVE PRODUCTIVITY VS. LOG PATENTED IN- NOVATIONS (2005) . . . . .	31
1.5	DENSITY ESTIMATES OF COUNTERFACTUAL CHANGES, T=100 . . . . .	34
1.6	WELFARE CHANGE AT T=100 AND CHANGES IN $h_r$ . . . . .	36
1.7	WELFARE CHANGE AND AMENITY/REMOTENESS LEVELS (BY CONTINENTS) . . . . .	37
1.8	LONG-TERM LOG CHANGES IN OVERALL INNOVATION VS. ESTI- MATED REGION-SPECIFIC IMPORTANCE WEIGHT OF PATENTED IN- NOVATIONS (T=100) . . . . .	38
1.9	AGGREGATION FOR DATA WITH ONE DEGREE RESOLUTION . . . . .	50
2.1	CENTROIDS OF 300 BIGGEST ADMINISTRATIVE CITIES IN CHINA (BY POPULATION IN 2000) . . . . .	55
2.2	NATURAL CITY OVER TIME – BEIJING . . . . .	56
2.3	NATURAL CITY OVER TIME – SHANGHAI . . . . .	57
2.4	SHRINKING NATURAL CITIES FROM 2010 TO 2013 . . . . .	58
2.5	KERNEL DENSITY ESTIMATES OF NATURAL CITY SIZE ACROSS CITIES	59
2.6	KERNEL DENSITY ESTIMATES – OBSERVED VS. PREDICTED RADI- ANCE LEVELS, 2007 . . . . .	70

2.7	KERNEL DENSITY ESTIMATES: OBSERVED VS. PREDICTED RADI- ANCE LEVELS, ALL YEARS . . . . .	70
2.8	COUNTERFACTUAL ROAD AND RAIL – BEIJING OVER TIME . . . . .	73
2.9	COUNTERFACTUAL ROAD AND RAIL – SHANGHAI OVER TIME . . . . .	74
3.1	EVOLUTION OF UNNORMALIZED TRAVEL TIME BETWEEN ALL PRE- FECTURE PAIRS (2000-2013) . . . . .	100
3.2	CHINESE HIGHWAY NETWORK (2000 AND 2013) . . . . .	101
3.3	CHINESE REGIONAL ROAD NETWORK (2000 AND 2013) . . . . .	102
3.4	THE GEOGRAPHY OF CONNECTIVITY GROWTH (2000-2013) . . . . .	103
3.5	ILLUSTRATION OF THE SEARCH SPACE FOR DIJKSTRA’S ALGORITHM .	105
3.6	OPTIMAL NETWORK USING THE MONGE–KANTOROVICH TRANS- PORTATION PROBLEM . . . . .	107
3.7	PREFECTURE-LEVEL TRADE AND MIGRATION COSTS . . . . .	123
3.8	KEY VARIABLES IN THE INITIAL STATE . . . . .	124
3.9	KEY VARIABLES IN THE LONG-RUN EQUILIBRIUM CONSISTENT WITH YEAR-2000 ROAD INFRASTRUCTURE . . . . .	126
3.10	CHANGES IN LONG-RUN EQUILIBRIUM VARIABLES (2000-2013) . . . .	127
3.11	CHANGE IN LONG-RUN EQUILIBRIUM VARIABLES BY CONNECTIVITY GROWTH . . . . .	128
3.12	LONG-RUN EQUILIBRIUM WITH DIFFUSION OF AMENITIES . . . . .	130
3.13	LONG-RUN EQUILIBRIUM WITH DIFFUSION OF TECHNOLOGY . . . . .	131
3.14	LONG-RUN EQUILIBRIUM WITH REDUCED MIGRATION FRICTIONS . .	132
3.15	LONG-RUN EQUILIBRIUM WITH REDUCED GOODS-TRADE FRICTIONS .	132
3.16	CONNECTIVITY GROWTH BY ROAD TYPE (2000-2013) . . . . .	133
3.17	GROWTH IN LONG-RUN EQUILIBRIUM VARIABLES WITH IMPROVE- MENTS OF HIGHWAYS (2000-2013) . . . . .	134
3.18	GROWTH IN LONG-RUN EQUILIBRIUM VARIABLES WITH IMPROVE- MENTS OF REGIONAL ROADS (2000-2013) . . . . .	135
4.1	SPATIAL PATTERNS OF CAPITAL MOVEMENT . . . . .	157
4.2	STATE CAPITALS AND NATIONAL HIGHWAY SYSTEM . . . . .	160
4.3	ABSOLUTE DISTANCE MEASURES . . . . .	163

4.4	DIRECT VS. INDIRECT CONNECTION . . . . .	164
4.5	RELATIVE DISTANCE MEASURES . . . . .	165
4.6	RELATIVE DISTANCE MEASURES (AVERAGE BY STATE) . . . . .	166
4.7	PREDICTED VS. OBSERVED POPULATION DENSITY . . . . .	169
4.8	PREDICTED U.S. STATES AND HYPOTHETICAL CAPITAL LOCATIONS .	171
4.9	PREDICTED CAPITAL STATUS VS. RANK . . . . .	172





# List of Tables

1.1	CALIBRATION OVERVIEW . . . . .	17
1.2	SUMMARY STATISTICS (2005) . . . . .	23
1.3	ESTIMATION RESULTS (MARGINAL EFFECTS) . . . . .	25
1.4	AMENITY PARAMETER ESTIMATION RESULTS . . . . .	32
1.5	R&D POLICY INSTRUMENTS IN 2005 . . . . .	33
1.6	MOMENTS OF REAL GDP GROWTH . . . . .	35
1.7	ROBUSTNESS ESTIMATION RESULTS (MARGINAL EFFECTS) – SUB- SAMPLES . . . . .	41
2.1	TRANSITION MATRIX . . . . .	58
2.2	SUMMARY STATISTICS – IN AND OUT NATURAL CITY AND TOTAL . .	63
2.3	SUMMARY STATISTICS – AVERAGES FOR 1992, 1998, 2007, 2013 . . .	64
2.4	ESTIMATION RESULTS DYNAMIC TOBIT . . . . .	67
2.5	TRANSITION MATRIX OF COUNTERFACTUAL $\ln(\text{dist to road}_i)$ . . . . .	71
2.6	TRANSITION MATRIX OF COUNTERFACTUAL $\ln(\text{dist to rail}_i)$ . . . . .	71
2.7	LIST OF 300 BIGGEST CHINESE ADMINISTRATIVE CITIES (BY POPU- LATION IN 2000) . . . . .	78
2.8	LIST OF CHINESE NATURAL CITIES (BY SIZE IN 2000) . . . . .	80
2.9	ESTIMATION RESULTS FOR ALTERNATIVE NON-TOBIT MODELS . . . .	82
2.10	ESTIMATION RESULTS DYNAMIC TOBIT BY CATEGORIES . . . . .	84
2.11	ROBUSTNESS – TRANSITION MATRIX OF COUNTERFACTUAL $\ln(\text{dist to road}_i)$ . . . . .	86
2.12	ROBUSTNESS – TRANSITION MATRIX OF COUNTERFACTUAL $\ln(\text{dist to rail}_i)$ . . . . .	86
3.1	CALIBRATION OVERVIEW . . . . .	109

3.2	GRAVITY ESTIMATION RESULTS OF BILATERAL MIGRATION . . . . .	114
3.3	ENDOGENOUS TECHNOLOGY PARAMETER ESTIMATION RESULTS . . . . .	119
3.4	ENDOGENOUS AMENITY PARAMETER ESTIMATION RESULTS . . . . .	121
3.5	LONG-RUN REGIONAL CONVERGENCE . . . . .	128
3.6	SOURCES OF CHINESE TRANSPORTATION ATLASES . . . . .	147
4.1	DESCRIPTIVE STATISTICS . . . . .	158
4.2	GEOGRAPHY-BASED POPULATION DENSITY IN 1900 . . . . .	169
4.3	FIRST STAGE RESULTS . . . . .	172
4.4	SECOND STAGE RESULTS . . . . .	172
4.5	ESTIMATION RESULTS – DRIVERS . . . . .	175
4.6	ESTIMATION RESULTS – ALTERNATIVE INSTRUMENT . . . . .	176
4.7	ESTIMATION RESULTS – ALTERNATIVE MEASURE OF RELATIVE MAR- KET ACCESS . . . . .	177
4.8	STATE CAPITALS - BASIC FACTS . . . . .	180
4.9	SECOND STAGE RESULTS - FULL TABLE . . . . .	182
4.10	ESTIMATION RESULTS - DRIVERS (FULL TABLE) . . . . .	183

# Abstract

This dissertation is a collection of four essays in economic geography. The dissertation broadly examines how geography shapes the spatial distribution of economic activity and growth.

The first chapter, co-authored with Peter Egger, analyzes the importance of R&D-investment incentives in a quantitative general-equilibrium model at the level of 5,633 regions around the world. While incentives vary across countries, the responses are largely heterogeneous across regions within and across countries. The reason for this heterogeneity roots in average technology differences as well as in the geography and amenities across regions. Results indicate that the promotion of innovation has relatively long-lasting positive effects. About one-tenth of the economic growth around the world can be attributed to the use and implementation of innovation-promoting tax schemes. In particular, R&D-investment incentives benefit low-amenity, peripheral places, and ones where patenting is relatively less common than elsewhere. This suggests that the studied nation-wide investment incentives also work as place-based policies.

The second chapter, co-authored with Peter Egger and Gabriel Loumeau, documents patterns in the size and growth of natural cities in China between 1992 and 2013. The boundaries of a natural city are identified in terms of night-light radiance of connected subcity places. The analysis documents a rapid growth of natural cities in China between 1992 and 2009, which was followed by a slight reduction in the size of some natural cities between 2010 and 2013 in the aftermath of the recent global financial crisis. Institutional factors – such as the location of places near Special Economic Zones, the ramifications of legal migration from rural to urban areas following reforms to the hukou (household registration) system, and infrastructure accessibility – are found to be important drivers of the

integration of peripheral places into natural cities.

The third chapter, co-authored with Peter Egger and Gabriel Loumeau, investigates the economic effects of road network improvements. We propose a quantifiable general-equilibrium model in which transport infrastructure affects the long-run equilibrium along four channels: accessibility of amenities, technology diffusion, reduced migration frictions and reduced trade frictions. We calibrate the model using Chinese prefecture-level data and exploit the massive growth in the Chinese road network between 2000 and 2013 to inform the model. The analysis suggests that the network changes stimulated regional convergence of lagging-behind prefectures in terms of population density and real per-capita incomes. Road-network-induced technology spillovers and trade-cost reductions led to economically large gains in prefectures that benefited the most from connectivity improvements. Associated amenity spillovers and migration-cost reductions mitigated almost offsetting consequences for population density and real incomes. The results indicate that not only the extension of the highway network but also improvements of other road network layers were quantitatively important.

The final chapter quantifies the importance of capital city status on road network integration of U.S. micro/metropolitan statistical areas. Road network integration is defined as a class of measurements that evaluate how well a location is connected to all other locations through the National Highway System (NHS). Capital status is instrumented using a  $k$ -means clustering algorithm that predicts the boundaries of 48 U.S. states on the basis of historical county-level data and defines the geographical center as a hypothetical capital location. I find significant and robust evidence that capital cities are more *directly* integrated in the NHS than non-capital cities of similar characteristics. Two possible mechanisms are discussed: (i) the favorable geographical position of capital cities within their state and (ii) a political interest in connecting capital cities well to major urban areas around.

# Zusammenfassung

Diese Dissertation besteht aus vier Forschungsarbeiten im Bereich Wirtschaftsgeographie. In der Dissertation wird untersucht, wie geographische Aspekte die räumliche Verteilung der Wirtschaftsleistung und des Wirtschaftswachstums beeinflussen.

Das erste Kapitel wurde in Zusammenarbeit mit Peter Egger verfasst und analysiert die Bedeutung steuerpolitischer Maßnahmen zur Innovationsförderung. Dies geschieht anhand eines quantitativen Modells des allgemeinen Gleichgewichts und berücksichtigt 5.633 Regionen weltweit. Die steuerlichen Anreize variieren von Land zu Land. Wir können jedoch beobachten, dass Unterschiede in ihren ökonomischen Auswirkungen selbst zwischen Regionen innerhalb eines Landes und über Ländergrenzen hinweg bestehen. Der Grund für diese Heterogenität liegt darin, dass Regionen unterschiedlich produktiv sind und sich in ihren geographischen Gegebenheiten, sowie in den exogenen Faktoren, die zur Attraktivität einer Region beitragen, unterscheiden. Die Ergebnisse legen nahe, dass die Innovationsförderung relativ lang andauernde positive Wachstumseffekte hat. Etwa ein Zehntel des weltweiten Wirtschaftswachstums ist auf den Einsatz innovationsfördernder Steuersysteme zurückzuführen. Generell sind die positiven Auswirkungen nur zum Teil von patentierten Innovationen bestimmt. Der größere Anteil der Wachstumseffekte ist auf nicht patentierte Innovationen zurückzuführen, wobei die Bedeutung patentierter vs. nicht-patentierter Innovationen je nach Land variiert. Die Förderung von Innovationen kommt insbesondere Regionen zugute, die grundsätzlich weniger attraktiv sind und solchen wo der Zugang zu globalen Märkten schwieriger ist. Aus wirtschaftspolitischer Sicht sind diese steuerlichen Anreize zur Förderung von Innovationen somit vergleichbar mit anderen sogenannten *Place-based Policies*.

Das zweite Kapitel, das gemeinsam mit Peter Egger und Gabriel Loumeau ver-

fasst wurde, dokumentiert und analysiert das Wachstum von sogenannten *natural cities* in der Volksrepublik China im Zeitraum von 1992 bis 2013. Das Konzept der *natural city* definiert die Größe einer Stadt anhand ihrer organisch gewachsenen Ausdehnung. Die Grenzen einer *natural city* werden mit Hilfe von Daten identifiziert, die nächtliche Lichtemissionen von zusammenhängenden Teilflächen messen. Die Ergebnisse der Analyse deuten darauf hin, dass *natural cities* in China zwischen 1992 und 2009 sehr schnell gewachsen sind. Auf das schnelle Wachstum folgte jedoch zwischen 2010 und 2013 eine leichte Verkleinerung von manchen *natural cities*, welche auf die wirtschaftlichen Folgen der jüngsten globalen Finanzkrise zurückzuführen ist. Die Integration von peripheren Regionen in *natural cities* wird von verschiedenen institutionellen Faktoren positiv beeinflusst. Dazu gehören beispielsweise die Nähe einer Region zu Sonderwirtschaftszonen, die Auswirkungen der Reformen des Hukou-Systems<sup>1</sup>, sowie ein einfacher Zugang zu guter Verkehrsinfrastruktur.

Das dritte Kapitel wurde gemeinsam mit Peter Egger und Gabriel Loumeau verfasst und analysiert die wirtschaftlichen Auswirkungen einer Erweiterung des Straßennetzes. Basierend auf einem quantitativen Modell des allgemeinen Gleichgewichts analysieren wir den Einfluss von Verkehrsinfrastruktur auf das langfristige Gleichgewicht anhand von vier Mechanismen: dem vereinfachten Zugang zu Faktoren, die die Lebensqualität in einer Region steigern, dem vereinfachten Austausch von Technologien, sowie der Reduktion von Migrations- und Handelskosten. Das Modell wird mit Hilfe von Daten zu chinesischen Präfekturen kalibriert. Des Weiteren nutzen wir in unserem Modell Informationen bezüglich der massiven Erweiterung des chinesischen Straßennetzes zwischen 2000 und 2013. Die Analyse legt nahe, dass die Erweiterung des chinesischen Straßennetzes dazu führte, dass weniger entwickelte Präfekturen in Bezug auf die Bevölkerungsdichte und das reale Pro-Kopf-Einkommen aufholen konnten. Der vereinfachte Austausch von Technologien und die Senkung der Handelskosten haben die größten positiven Auswirkungen auf jene Präfekturen, die die stärkste Veränderung in Bezug auf eine verbesserte Netzwerkintegration erfahren haben. Der verbesserte Zugang zu lebensqualitätssteigernden Faktoren und die Reduzierung der Migrationskosten haben eine

---

<sup>1</sup>Das Hukou-System beschreibt ein offizielles System der Wohnsitzkontrolle der Bevölkerung in China.

ausgleichende Wirkung auf die Bevölkerungsdichte und das Realeinkommen. Im Allgemeinen, legen die Ergebnisse nahe, dass nicht nur Verbesserungen der Autobahnen, sondern auch die Erweiterung des regionalen Straßennetzes eine wichtige Rolle für die wirtschaftlichen Auswirkungen spielten.

Das letzte Kapitel analysiert die Bedeutung des Hauptstadtstatus für die Integration in das nationale Straßennetz am Beispiel von US-amerikanischen Städten. Die Integration eines Standortes in das nationale Straßennetz wird mit Hilfe von vier verschiedenen Variablen gemessen. Die Variablen bewerten in unterschiedlicher Weise, wie gut ein Standort über das National Highway System der USA mit allen anderen Standorten verbunden ist. Da die Wahl der Hauptstädte in den verschiedenen Staaten der USA keineswegs zufällig war, wird der Hauptstadtstatus mithilfe eines  $k$ -means-Clustering-Algorithmus instrumentiert. Dieser Algorithmus bestimmt die Grenzen der 48 Bundesstaaten auf Grundlage von historischen Regionaldaten und definiert das geographische Zentrum jedes Staates als *hypothetische* Hauptstadt. Die empirischen Ergebnisse legen nahe, dass Hauptstädte in den USA *direkter* in das National Highway System integriert sind als vergleichbare Städte, die keinen Hauptstadtstatus haben. Die Analyse diskutiert zwei mögliche Erklärungen für dieses Ergebnis. Zum einen sind Hauptstädte innerhalb ihres Staates geographisch sehr zentral und damit günstig gelegen. Und zum anderen ist es denkbar, dass es ein politisches Interesse daran gibt, die Hauptstädte besonders gut mit anderen wirtschaftlich wichtigen Städten in der Umgebung zu verbinden.





# Introduction

Economic activity is unevenly distributed across space. Individuals worldwide are increasingly concentrated in metropolitan regions, and it is expected that 68% of the world's population will be living in urban areas by 2050.<sup>2</sup> Similarly, economic output concentrates primarily in highly productive urban areas. In 2003, 38% of the total GDP in OECD countries was generated by only 10% of OECD regions.<sup>3</sup> A region's location in space and its proximity to economic clusters play a crucial role in the relative gains from (inter-)national trade and migration, the magnitude of investments, and the effectiveness of public policies. The overall aim of this dissertation is to examine the role of geography in the spatial distribution of economic activity. This will help to inform and design future public policies through understanding the forces at play and their consequences on regional well-being. Each chapter is briefly introduced below.<sup>4</sup>

Many national governments utilize R&D tax policy instruments to attract innovative activity and therefore stimulate regional growth and gain in international competition. This behavior affects the spatial distribution of innovative activity and provides important insights regarding the global economic growth path. The first chapter (co-authored with Peter Egger) outlines a quantitative global model to assess the importance of R&D investment incentives across 5,633 regions around the globe in a long-term economic growth analysis. The model is calibrated so that the main regional characteristics – including population density, per-capita income, and market accessibility – match those of the real world in the year 2005. In the counterfactual equilibrium analysis, we focus on the effects of R&D-investment incentives on three key variables – place-specific employment, productivity, and

---

<sup>2</sup>United Nations Population Division. World Urbanization Prospects: 2018 Revision.

<sup>3</sup>OECD Regions at a Glance: 2007 Edition.

<sup>4</sup>Each co-author made equal contributions to Chapters 1, 2 and 3.

welfare – in a scenario where investment incentives towards innovation are abandoned. The analysis suggests that the promotion of innovation has relatively long-lasting positive effects. About one-tenth of the economic growth around the world can be attributed to the use and implementation of innovation-promoting tax schemes. These effects are particularly strong in regions where innovation-promoting policies are in place. However, regions that are not exposed to such policies nonetheless benefit from them being in place elsewhere, given that improvements in technology dissipate and become publicly available with time. On average, only a relatively small part of the benefits, in terms of real consumption gains, materializes through patenting. To a varying degree across regions, most of the gains are due to unpatented innovations. Overall, innovation-promoting policy instruments are most productive in regions where amenities are weak and global markets are harder to access. From an economic policy point of view, R&D investment incentives act as place-based policies, which are commonly adapted to enhance the economic performance of disadvantaged regions.

Apart from innovation clusters, the concentration of economic activity is reflected by the existence of densely populated urban areas. The growth of urban settlements is particularly large in fast-growing economies. For instance, in China almost 25% of the population has moved to urban areas during the past two decades. The second chapter (co-authored with Peter Egger and Gabriel Loumeau) documents patterns in the size and growth of Metropolitan Statistical Areas in China between 1992 and 2013. We employ a definition of city boundaries based on what we call natural borders. Natural city borders are identified in terms of night-light radiance – a measure of economic activity – of connected subcity places using a City Clustering Algorithm. The analysis suggests that the number of distinct natural city centers decreased between 1992 and 2013 due to the absorption of some natural cities by others. This was particularly the case for larger cities, such as Shanghai or Beijing, that formed natural super-cities during that time period. Moreover, we document that Chinese natural cities grew considerably beyond the administrative boundaries. The rapid growth between 1992 and 2009 was followed by a slight reduction in the size of some natural cities between 2010 and 2013 in the aftermath of the recent global financial crisis. Institutional factors – such as

the proliferation of Special Economic Zones, provisions of the hukou (household registration) system and infrastructure accessibility – are found to be important drivers of the integration of peripheral places into natural cities. In particular, the results indicate that the enormous investments in transport infrastructure in China can be expected to induce relatively rapid adjustments in natural city size over the next 20 years; a finding that inspired us to further research.

Investments in transport infrastructure are among the few public policy interventions that have a direct impact on the determinants of trade and migration costs. The vast majority of research on the economic effects of infrastructure improvements has focused on the accessibility of goods and factors through highway networks. The third chapter (co-authored with Peter Egger and Gabriel Loumeau) takes a broader view on transport infrastructure. It considers multiple channels through which infrastructure improvements affect the region-specific well-being of households beyond the standard transport of goods and factors, and decomposes the heterogeneous economic effects across network types. We propose a quantifiable general-equilibrium model that is amenable to studying the decomposition of economic effects of transport infrastructure across four channels: the accessibility of amenities, technology diffusion, reduced migration frictions and reduced trade frictions. Our analysis builds on a calibrated version of the open Chinese economy in which we use novel, hand-collected, annual data on the road network for the period between 2000 and 2013. Regarding the endogeneity of road infrastructure placements, we propose a novel Instrumental Variable approach that is based on the classical Monge-Kantorovich transportation problem. In the counterfactual analysis, we decompose the overall effects of road infrastructure improvements on population densities and regional income across Chinese prefectures. The results indicate that the extension of the Chinese road network between 2000 and 2013 fostered regional convergence in population density and real income, as population relocated from large centers to mid-sized prefectures in central China. Increased accessibility of technology and reduced trade frictions have had large positive real income effects in prefectures that gained the most in connectivity. Associated amenity spillovers and migration-cost reductions mitigated almost offsetting consequences for population density and real incomes. Regarding network types, the

results indicate that highway improvements have strong integration effects on formerly remote places. Regional road improvements reduce former differences in local connectivity and alleviate the differences that are generated by the highway network.

Beyond recent policies, there exist long-lasting institutional features that shape the economic geography of regions and countries. One prominent example of such an institutional feature is the political status of urban areas. The last chapter links the political status of U.S. micro/metropolitan statistical areas to their integration in the National Highway System (NHS) in order to understand whether there is a *capital premium* in road network provision. Historically, most state capital cities in the U.S. embodied their role of political power by being centrally located within their state. I use this common feature of geographical centrality to construct an instrument for the endogenous capital location. The Instrumental Variable design is based on a  $k$ -means clustering algorithm that predicts the boundaries of 48 U.S. states and defines the geographical center as a hypothetical capital location. I then estimate the causal effect of capital status on four outcomes of road network integration. Two outcomes (connectivity and market access) evaluate the strength of integration based on the aggregate proximity to all other locations. The other two outcomes (relative connectivity and relative market access) measure how directly connected a location is to all others. I find significant and robust evidence that U.S. state capital cities are more *directly* integrated in the NHS compared to non-capital cities of similar characteristics. The reason for this finding is a combination of two aspects. First, capital cities have a favorable geographical position within their state. This makes them a natural candidate for a direct road network integration according to the central place theory. Second, as capital cities are places of political power and decision-making, there seems to be a governmental interest in establishing direct connections to other major urban areas. The decision on the location of the U.S. federal highway network was subject to inter-governmental negotiations, which likely favored state capital cities.

The dissertation contributes to the literature in several ways. First, it provides new tools, both theoretical and empirical, to analyze key spatial interactions in the context of heterogeneous geography. Second, it details mechanisms through

which various public policies affect the spatial distribution of economic activity. Third, it proposes new approaches to address the endogenous location choice of policy interventions, such as transport infrastructure placements or capital city location. Finally, it discusses implications for the design of future spatial policies.

The four chapters follow below.



# Chapter 1

## The Economic Geography of Innovation

*This is the original manuscript of an article published as CEPR Discussion Paper No. DP13338, available online: <https://ssrn.com/abstract=3290531>*

### 1.1 Introduction

Technology and productivity are key drivers not only of production potential of places but also of the attractiveness for mobile factors to locate there and, hence, of demand potential and well-being. The technological capabilities of production factors located in a place are influenced to a major extent by local innovation and the capability of absorbing innovations generated elsewhere. Policy makers have a number of instruments at hand which are particularly aimed at stimulating innovation for exactly that reason. Earlier research concerned with the effect of innovation incentives – where innovations are commonly measured by patent filings and other patenting behavior – on economic outcomes focuses largely on reduced-form effects, which abstract from general-equilibrium repercussions. One related strand of reduced-form work focuses on the effect of R&D tax incentives on innovation (see Jong and Verhoeven, 2007, for the Netherlands; Ernst and Spengel, 2011, for multiple EU countries; Westmore, 2013, for 19 OECD countries; Aralica and Botrić, 2013, for Croatia; Knoll et al., 2014, for European countries; Czarnitzki and Lopes-Bento, 2014, for Canada; Bösenberg and Egger, 2017, for 106

countries). Another strand of reduced-form work highlights the effect of R&D tax incentives on productivity (see Caiumi, 2011, for Italy; Hallépée and Houlou, 2012, for France; and Cappelen et al., 2012, for Norway). The shortcoming of reduced-form analysis is that quantitatively potentially important interdependencies of outcomes, markets, and places are ignored by assumption. Also, the heterogeneity of places in their response to even a homogeneous treatment of economic policy is beyond the reach of a reduced-form analysis, at least to the degree that is suggested by general-equilibrium theory.

The present paper adopts a structural approach, which permits accounting for direct and indirect (spillover plus general-equilibrium) effects of customary innovation-stimulating policies. For this purpose, it formulates, estimates key parameters of, and calibrates a quantitative, multi-region model of trade and factor mobility among places in order to assess the economic value of innovation incentives and their consequences for the location of supply and demand across places as well as for the well-being of consumers there. With this agenda, the paper particularly relates to three lines of work. The one on the social cost-benefit and aggregate analysis of individual tax incentives towards R&D (see Cornet, 2001, and Lokshin and Mohnen, 2011, for the Netherlands; Parsons and Phillips, 2007, for Canada; and Bloom et al., 2013, for the United States) which is based on reduced-form estimates but aims at accounting for effects on various outcomes. The structural aggregate (macro-economic) modeling approach in Atkeson and Burstein (2019), which permits gauging global effects of innovation policies, abstracting from the multi-region structure of the world economy. And a host of studies with a focus on structural-quantitative, multi-region models with mobile goods and factors without a deeper consideration of innovation policies (see Allen and Arkolakis, 2014; Ahlfeldt et al., 2015; Donaldson and Hornbeck, 2016; Nagy, 2017; Allen and Donaldson, 2018; Monte et al., 2018; Caliendo and Parro, 2015; Caliendo et al., 2018; Desmet et al., 2018; Donaldson, 2018; Redding and Rossi-Hansberg, 2017, for an extensive review of that line of work).

The model we propose builds on Allen and Arkolakis (2014), Desmet et al. (2018), and Allen and Donaldson (2018) and describes a world in which each place is unique in terms of amenities, productivity, and geography. Firms have an



incentive to innovate as it improves their productivity and competitiveness. However, the benefits from innovation which are exclusive to the firm are short-lived, and knowledge about any newly-invented technology becomes public after one period. The technology available to firms in a place evolves through an endogenous dynamic process. Innovation is produced under constant returns to scale, using research labor for each unit of innovation produced. In contrast to Allen and Arkolakis (2014) and Desmet et al. (2018), total factor productivity consists of a random and a chosen part through (optimal) investments in innovation. The parametrization and estimation of the endogenous productivity component as well as of the dynamic technology process are at the heart of the paper's interest. Firms benefit from R&D investment incentives in places, *ceteris paribus*, as they reduce the costs of generating innovations all else equal. Firms use patented as well as non-patented innovations in doing so.

Our analysis considers 5,633 regions in 213 countries around the globe, where the delineation of regions follows the definition by the Organization of Economic Cooperation and Development (OECD) and their Regional Patent-statistical Database (REGPAT). For the estimation of the R&D-worker-specific productivity shifter, we use region-specific efficiency levels that are recovered from the model structure and five country-specific indicators on R&D investment incentives which are geared towards innovations from Bösenberg and Egger (2017).

In the counterfactual equilibrium analysis we focus on the effects on three key variables – place-specific employment, productivity, and welfare – in a scenario where investment incentives towards innovation are abandoned. There are three main take-aways from the analysis. First, the use of policy instruments which are designed to stimulate private R&D are globally beneficial in terms of productivity and welfare. In other words, also countries and their regions who do not use such instruments benefit from their use elsewhere due to technology spillovers. Second, the long-run relocation effects due to a hypothetical abolishment of R&D tax incentives are substantial and lead to a re-shifting of the population towards high-density areas (i.e., centrally-located ones with great exogenous amenities). Hence, transport accessibility and good exogenous amenities work as a quasi-insurance against adverse innovation policy shocks. Analogously, the quantitative analysis

suggests that a nation-wide innovation policy works indirectly as a place-based policy, where low-amenity, peripheral regions benefit, *ceteris paribus*, relatively more than high-amenity, centrally located ones. This effect is particularly driven by a relative gain in international competitiveness from national R&D policies due to the cross-border mobility of labor.

Furthermore, the quantitative analysis suggests that about one-tenth of the long-run growth rate of real GDP on the globe can be attributed to the use of R&D policy instruments as used in the year 2005 alone. The findings also imply that only a relatively small fraction of that should be attributed to the stimulus on patenting, but the share of non-patented innovations triggered by such policy instruments is relatively large.

The remainder of the paper is organized as follows. Section 1.2 presents the model, states the equilibrium conditions for each period and defines the underlying assumptions for a unique balanced growth path to exist. Section 1.3 discusses the calibration of key model parameters, including a methodology to determine or estimate them. Section 1.4 presents the results of our counterfactual analysis. Section 1.5 concludes.

## 1.2 The Model

We consider a world where  $S$  is a set of regions  $r$  on a two-dimensional surface, i.e.,  $r \in S$ . Region  $r$  has land density  $G_r > 0$ , where  $G_r$  is exogenously given and normalized by the average land density of all regions in the world. The world is inhabited by a measure  $\bar{L}$  of workers, who are freely mobile between regions and endowed with one inelastically-supplied unit of labor each. Each region is unique in terms of geography, amenities and productivity.

### 1.2.1 Innovation and Production

In each region, firms produce product varieties  $\omega$ , innovate, and trade subject to iceberg transport costs. A firm's production of  $\omega$  per unit of land in the intensive form is defined as

$$q_{rt}(\omega) = \phi_{rt}(\omega)^{\gamma_1} z_{rt}(\omega) L_{rt}(\omega)^\mu, \quad \gamma_1, \mu \in (0, 1]. \quad (1.1)$$

Output depends on production labor per unit of land,  $L_{rt}(\omega)$ , and the firm's total factor productivity, which is determined by two components: an endogenous innovation component,  $\phi_{rt}(\omega)$ , and an exogenous, product-specific productivity factor,  $z_{rt}(\omega)$ , which is drawn from a Fréchet distribution with location parameter  $T_{rt} = \tau_{rt} \bar{L}_{rt}^\alpha$  and shape parameter  $\theta$ , where  $\alpha \geq 0$  and  $\theta > 0$ . Where in the productivity distribution a firm is located depends on the total workforce at region  $r$  in period  $t$ ,  $\bar{L}_{rt}$ , and the region's level of efficiency,  $\tau_{rt}$ .

The value of  $\tau_{rt}$  is determined by an endogenous dynamic process, which depends on past investments into local innovations, and the capability of absorbing innovations that were generated elsewhere and now diffuse globally. Assuming a first-order autoregressive process about efficiency,<sup>1</sup> we postulate

$$\tau_{rt} = \phi_{rt-1}^{\gamma_1 \theta} \left[ \int_S W_{rs} \tau_{st-1} ds \right]^{1-\gamma_2} \tau_{rt-1}^{\gamma_2}, \quad (1.2)$$

where  $\gamma_1, \gamma_2 \in (0, 1)$  and  $W_{rs}$  is an  $r$ - $s$ -specific technology diffusion weighting scalar.<sup>2</sup> The value of  $\gamma_2$  determines the strength of technological diffusion. The higher  $\gamma_2$ , the more a region benefits from own investments in technology. In return, low levels of  $\gamma_2$  imply that the aggregate level of investment into technology in a region is relatively more important than local investments.

Firms have an incentive to invest into innovation as it improves their productivity in (1.1). This allows them to post a higher bid for the regionally fixed factor of production, land. However, due to a decreasing marginal product of labor, the innovation effort will be finite. The latter is guaranteed by the parameter configuration where land intensity is larger than the cost normalized innovation intensity in production,  $[1 - \mu] > \gamma_1/\xi$ .

Innovation,  $\phi_{rt}(\omega)$ , is produced under Cobb-Douglas technology and with constant returns to scale, such that a firm has to employ  $\nu \phi_{rt}(\omega)^\xi h_{rt}^{-1}$  additional units

<sup>1</sup>Allowing for a longer memory in the process would be technically straightforward. E.g., Allen and Donaldson (2018) consider a second-order process. However, the available data for the present paper do not permit doing so, as the time series available for each region is extremely short, as will become clear below.

<sup>2</sup>The definition of  $W_{rs}$  is discussed in further detail in Section 1.2.5.

of labor in order to innovate, where  $h_{rt} \geq 1$  is a region-time-specific R&D-worker productivity shifter, which reduces the cost of innovation per unit of innovation produced. The latter will be key to the analysis here, as it captures the influence of R&D tax incentives.

Firms enjoy the benefit of their innovation for only one period. In the next period all entrants to the market have the same access to technology. This simplifies the dynamic profit maximization to a sequence of static problems. After learning their productivity draw  $z_{rt}(\omega)$ , firms maximize their profits with choosing the level of employment and innovation through

$$\max_{L_{rt}(\omega), \phi_{rt}(\omega)} p_{rt}(\omega) \phi_{rt}(\omega)^{\gamma_1} z_{rt}(\omega) L_{rt}(\omega)^\mu - w_{rt}[L_{rt}(\omega) + \phi_{rt}(\omega)^\xi h_{rt}^{-1}] - b_{rt},$$

where  $p_{rt}$  is the price a firm charges for a product that is sold in region  $r$  and period  $t$ . A firm's productivity affects prices without changing unit costs,  $o_{rt}$ , such that  $p_{rt}(\omega) = o_{rt}/z_{rt}(\omega)$ , with

$$o_{rt} \equiv \left[ \frac{1}{\mu} \right]^\mu \left[ \frac{\nu \xi}{\gamma_1} \right]^{1-\mu} \left[ \frac{b_{rt} \gamma_1}{w_{rt} \nu (\xi(1-\mu) - \gamma_1)} \right]^{(1-\mu) - \frac{\gamma_1}{\xi}} h_{rt}^{-\frac{\gamma_1}{\xi}} w_{rt}. \quad (1.3)$$

Each firm considers their production unit costs as given, which is why  $o_{rt}$  is not product-specific.  $b_{rt}$  reflects the firms' bid rent for land, which can be derived from the first-order conditions as a function of the per-unit costs of innovation  $w_{rt} \phi_{rt}(\omega)^\xi h_{rt}^{-1}$ , so that

$$b_{rt} = \left[ \frac{\xi(1-\mu)}{\gamma_1} - 1 \right] \nu w_{rt} \phi_{rt}(\omega)^\xi h_{rt}^{-1}. \quad (1.4)$$

### 1.2.2 Innovation and Total Employment

Total employment in region  $r$  at period  $t$  is the sum of production workers,  $L_{rt}(\omega)$ , and innovation workers,  $\nu \phi_{rt}(\omega)^\xi h_{rt}^{-1}$ , so that

$$\bar{L}_{rt}(\omega) = L_{rt}(\omega) + \nu \phi_{rt}(\omega)^\xi h_{rt}^{-1} = L_{rt}(\omega) \left[ 1 + \frac{\gamma_1}{\mu \xi} \right], \quad (1.5)$$

where the last equality follows from the first-order relation between production labor and innovation labor,

$$\frac{\xi}{\gamma_1} \nu \phi_{rt}(\omega)^\xi h_{rt}^{-1} = \frac{L_{rt}(\omega)}{\mu} \Rightarrow \nu \phi_{rt}(\omega)^\xi h_{rt}^{-1} = \frac{\gamma_1}{\xi[\mu + \gamma_1/\xi]} \bar{L}_{rt}(\omega). \quad (1.6)$$

### 1.2.3 Utility and Consumption

When choosing residence in region  $r$ , a representative worker in period  $t$  derives utility from local amenities,  $a_{rt}$ , and from consuming a set of differentiated product varieties  $\omega$  with CES preferences according to

$$u_{rt} = a_{rt} C_{rt} = a_{rt} \left[ \int_0^1 c_{rt}(\omega)^{\frac{\sigma-1}{\sigma}} d\omega \right]^{\frac{\sigma}{\sigma-1}} \quad \text{with } a_{rt} = \bar{a}_{rt} \bar{L}_{rt}^{-\lambda}, \quad (1.7)$$

where  $a_{rt}$  are amenities at  $r$  in  $t$ , with  $\bar{a}_{rt}$  being an exogenous amenity attribute and  $\lambda \geq 0$  being a congestion externalities parameter.  $C_{rt}$  is the real consumption bundle, and  $\sigma \in (1, \infty)$  is the elasticity of substitution between products  $\omega$ .

Consumer-workers earn income from work,  $w_{rt}$ , and from local ownership of land. Local land rents are uniformly distributed among all residents in a region, i.e., the land rent per resident is  $b_{rt}/\bar{L}_{rt}$ . As we assume that agents cannot write debt contracts with each other and there is perfect local competition, it follows that each consumer-worker spends all her income. Hence, the indirect utility is defined as

$$u_{rt} = a_{rt} y_{rt} = a_{rt} \frac{w_{rt} + b_{rt}/\bar{L}_{rt}}{P_{rt}}, \quad (1.8)$$

where  $P_{rt} = \Gamma\left(\frac{1-\sigma}{\theta} + 1\right)^{\frac{1}{1-\sigma}} \left[ \int_S T_{kt} [o_{kt} \zeta_{ks}]^{-\theta} dk \right]^{-\frac{1}{\theta}}$  is the price index in region  $r$  and period  $t$ . As in Eaton and Kortum (2002), the share of consumption in region  $r$  of products produced in region  $s$  is determined by

$$\pi_{rst} = \frac{T_{rt} [o_{rt} \zeta_{rs}]^{-\theta}}{\int_S T_{kt} [o_{kt} \zeta_{ks}]^{-\theta} dk}, \quad \forall r, s \in S, \quad (1.9)$$

where  $\zeta_{rs} > 1$  denote the iceberg costs of transporting a product from  $r$  to  $s$ .

### 1.2.4 Equilibrium in Each Period

Profits and utility are maximized within each period, as neither firms nor consumers are forward-looking; see also Desmet et al. (2018) and Allen and Donaldson (2018).

The equilibrium population density will be evaluated as a measure of the location specific utility,  $u_{rt}$ , such that

$$\bar{L}_{rt} = \frac{\bar{L}}{G_r} \frac{u_{rt}^{1/\Omega}}{\int_S u_{kt}^{1/\Omega} dk}, \quad \text{with } \int_S G_r L_{rt} dr = \bar{L}, \quad (1.10)$$

where  $\Omega$  is a Fréchet dispersion parameter of a location-specific preference shock as in Desmet et al. (2018).<sup>3</sup> Overall, population mobility is restricted by the location-specific preference parameter ( $\Omega$ ), an amenity-reducing congestion parameter ( $\lambda$ ) and the land-intensity in production ( $1 - \mu$ ).

Product-market clearing requires total revenues in region  $r$  to be equal to total expenditures on products from region  $r$ . Hence,

$$w_{rt} G_r \bar{L}_{rt} = \int_S \pi_{rst} w_{st} G_s \bar{L}_{st} ds \quad \forall r, s \in S, \quad (1.11)$$

where  $L_{rt}$  can be replaced with  $\bar{L}_{rt}$  as production labor is proportional to total labor across all regions.

In equilibrium, population density in each region is determined by (1.10), replacing  $u_{rt}$  by the indirect utility in (1.8). The product-market clearing pins down wages, with substituting (1.4) into (1.3) and using this expression to replace it into the trade share (1.9), which in return can be substituted in (1.11).

An equilibrium exists and is unique if dispersion forces are greater than agglomeration forces. Hence,

$$\underbrace{\frac{\alpha}{\theta} + \frac{\gamma_1}{\xi}}_{\text{Static agglomeration forces}} \leq \underbrace{\lambda + 1 - \mu + \Omega}_{\text{Static dispersion forces}}. \quad (1.12)$$

---

<sup>3</sup>Notice that location-time-specific utility,  $u_{rt}$ , and, more specifically, the amenity parameter,  $a_{rt}$ , is proportional to average migration costs in region  $r$ . In general, residence-region-specific migration costs are isomorphic to location-specific amenities. Hence, the population-share specification in (1.10) accounts for such costs.

A detailed proof of the uniqueness condition can be based on the insights from Allen and Arkolakis (2014), Desmet et al. (2018) and Allen and Donaldson (2018), and it is presented in Appendix 1.8.

### 1.2.5 Balanced Growth Path

In a balanced growth path (BGP), technology growth rates are constant and identical across regions at constant fundamentals, implying that  $\frac{\tau_{rt+1}}{\tau_{rt}}$  is constant over time and space, and  $\frac{\tau_{st}}{\tau_{rt}}$  is constant over time. Firms' investment decisions into innovation are constant, but they differ across regions. In order for a BGP to materialize, we assume that the R&D-worker-specific productivity shifter is constant over time,  $h_{rt} = h_r$ , in the BGP as well as in the transition towards it. Rewriting the endogenous dynamic process in (1.2), the growth rate of  $\tau_{rt}$  can then be expressed as

$$\frac{\tau_{rt+1}}{\tau_{rt}} = \phi_{rt}^{\theta\gamma_1} \left[ \int_S \frac{W_{rs}\tau_{st}}{\tau_{rt}} ds \right]^{1-\gamma_2}. \quad (1.13)$$

That growth rate relative to region  $s'$  is

$$\frac{\frac{\tau_{rt+1}}{\tau_{rt}}}{\frac{\tau_{st+1}}{\tau_{st}}} = \left[ \frac{\tau_{st}}{\tau_{rt}} \right]^{1-\gamma_2} \left[ \frac{\phi_{rt}}{\phi_{st}} \right]^{\theta\gamma_1} \left[ \frac{\int_S W_{rs}\tau_{st} ds}{\int_S W_{sr}\tau_{rt} dr} \right]^{1-\gamma_2}. \quad (1.14)$$

In a BGP,  $\left( \frac{\tau_{rt+1}/\tau_{rt}}{\tau_{st+1}/\tau_{st}} \right) = 1$ . Furthermore, for a BGP to exist, technological diffusion, which is governed by  $W_{rs}$ , needs to be uniform across space, implying that  $[\int_S W_{rs}\tau_{st} ds / \int_S W_{sr}\tau_{rt} dr] = 1$ , see Egger and Pfaffermayr (2006).<sup>4</sup> Desmet et al. (2018) propose to specify  $W_{rs} = \left[ \frac{1}{N} \right], \forall rs$ , where  $N$  describes the total number of locations considered in the model, and we follow them in this regard. With the latter assumption, (1.14) reduces to

$$\frac{\tau_{rt}}{\tau_{st}} = \left[ \frac{\phi_{rt}}{\phi_{st}} \right]^{\frac{\theta\gamma_1}{1-\gamma_2}} = \left[ \frac{\bar{L}_{rt}h_r}{\bar{L}_{st}h_s} \right]^{\frac{\theta\gamma_1}{(1-\gamma_2)\xi}}. \quad (1.15)$$

---

<sup>4</sup>To see this, consider the following thought experiment. Suppose each region  $r$  would receive the same time-invariant, common growth impulse. If  $[\int_S W_{rs}\tau_{st} ds \neq \int_S W_{sr}\tau_{rt} dr]$ , the same impulse would have region-specific consequences due to the importance of the regions' location in the spillover network. Then, the same impulse would be amplified (or moderated) to a heterogeneous degree, and regional growth would be heterogeneous in the BGP, as a result.

Then, there exists a unique BGP of the system, if

$$\underbrace{\frac{\alpha}{\theta} + \frac{\gamma_1}{\xi}}_{\text{Static agglomeration forces}} + \underbrace{\frac{\gamma_1}{[1 - \gamma_2]\xi}}_{\text{Dynamic agglomeration forces}} \leq \underbrace{\lambda + 1 - \mu + \Omega}_{\text{Static dispersion forces}}, \quad (1.16)$$

which is the same as in Desmet et al. (2018); see Appendix 1.9.1 for a proof.

In the BGP, aggregate welfare and real consumption growth depends on the population density, the R&D-worker-specific productivity shifter and their distribution in space, according to

$$\frac{u_{rt+1}}{u_{rt}} = \frac{C_{rt+1}}{C_{rt}} = \left[ \frac{1}{N} \right]^{\frac{1-\gamma_2}{\theta}} \left[ \frac{\gamma_1/\nu}{\gamma_1 + \mu\xi} \right]^{\frac{\theta\gamma_1}{\xi}} \left( \int_S (\bar{L}_s h_s)^{\frac{\theta\gamma_1}{[1-\gamma_2]\xi}} ds \right)^{\frac{1-\gamma_2}{\theta}}, \quad (1.17)$$

where  $a_{rt} = a_{rt+1}$ , as the population density in each region is constant over time in the BGP.<sup>5</sup>

### 1.3 Calibration of Key Model Parameters

To compute the quantitative multi-region equilibrium for each time period from a given year to the steady state, we need the parameters contained in the equations above and summarized in Table 1.1. Apart from parameters that are common to all regions and region-specific land endowments which are given in the data, these are initial efficiency levels in production and exogenous amenity levels for all regions, the R&D-worker-specific productivity shifter as well as trade costs between all pairs of regions. Table 1.1 alludes to the sources of these parameters, some of which are collected from other work and some of which are derived (computed or estimated) here.

We organize the remainder of this section in subsections which pertain to important model blocks based on which estimating equations are formulated or key parameters can be backed out.

---

<sup>5</sup>A detailed derivation of the growth rate of aggregate welfare is presented in Appendix 1.9.2.



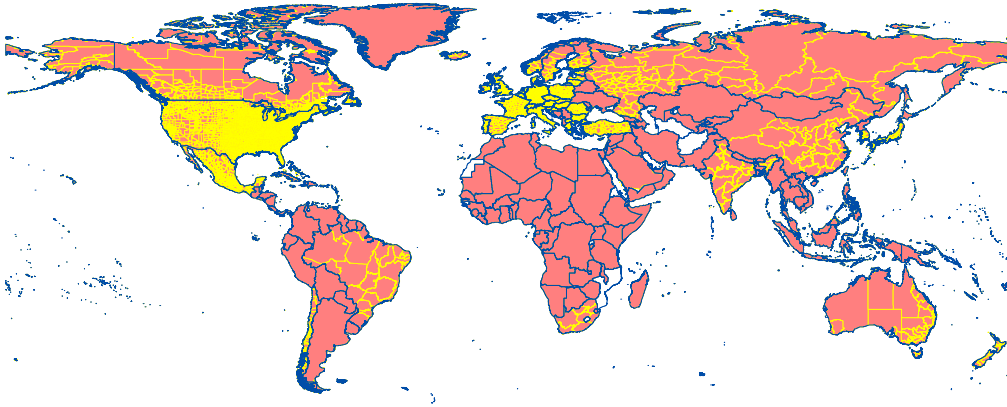
Table 1.1: CALIBRATION OVERVIEW

PARAMETER COMMON TO ALL REGIONS		
<b>1. Preferences</b>		
$\sigma = 4$	Elasticity of substitution.	Bernard et al. (2003)
$\lambda = 0.596$	Relation between amenities and population.	Own estimation, Section 1.3.7
$\Omega = 0.5$	Elasticity of migration flows w.r.t. income.	Monte et al. (2018)
<b>2. Technology</b>		
$\alpha = 0.06$	Elasticity of productivity w.r.t. pop. density.	Carlino et al. (2007)
$\theta = 6.5$	Trade elasticity and dispersion of productivity.	Eaton and Kortum (2002), Simonovska and Waugh (2014)
$\mu = 0.8$	Labor share in production (non-land share).	Greenwood et al. (1997); Desmet and Rappaport (2017)
$\gamma_1 = 0.234$	Elasticity of tomorrow's productivity w.r.t. today's innovation.	Own estimation, Section 1.3.5
<b>3. Evolution of Productivity</b>		
$\gamma_2 = 0.979$	Elasticity of tomorrow's productivity w.r.t. today's productivity.	Own estimation, Section 1.3.5
$\xi = 125$	Elasticity of innovation costs w.r.t. innovation.	Desmet and Rossi-Hansberg (2015)
$\nu = 0.15$	Intercept parameter in innovation cost function.	Desmet et al. (2018)
REGION-SPECIFIC PARAMETER		
<b>1. Land Endowments</b>		
$G_r$	Extract land mass for each region. ( $G_r$ is normalized by $\frac{1}{N} \sum_S G_r$ )	ArcGIS Software
<b>2. Initial Efficiency in 2005</b>		
$\tau_{rt}$	Initial efficiency distribution.	Own estimation, Section 1.3.3
<b>2. Amenities in 2005</b>		
$a_{rt}$	Initial amenity distribution.	Own estimation, Section 1.3.7
$\bar{a}_{rt}$	Exogenous amenity attribute.	Own estimation, Section 1.3.7
<b>3. Productivity-shifter for R&amp;D workers in 2005</b>		
$h_r$	Estimation using binary R&D policy indicators $h_r = \exp(\mathbf{D}_r \hat{\beta} +  lat_r  \mathbf{D}_r \hat{\gamma} + \hat{\delta}  lat_r )$	Own estimation, Section 1.3.4.
<b>4. Transport Costs</b>		
$\zeta_{rs}$	Based on Allen and Arkolakis (2014) and Fast Marching Algorithm.	
<b>5. Other Trade Costs</b>		
$tariff_{rs}$	Weighted applied import tariffs for manufactures	World Development Indicator (WDI)

### 1.3.1 Delineation of Regions and Land Endowments

The delineation of regions used in our analysis is dictated by the definitions used in the Regional Patent-Statistical Database (REGPAT) of the Organization of Economic Cooperation and Development (OECD).<sup>6</sup> In 2005, REGPAT distinguishes 5,633 regions across 213 countries on the globe. The size of regions by land mass (somewhat less so by income or patenting) differs to a large extent. In some countries the granularity of regions is very fine, while it is coarse in others. In some cases, even a whole country is a region, e.g., in some African, Asian or South American countries. This pattern is related to the intensity of patenting in a country: economies with more patents tend to be organized in a more fine-grained fashion, while the ones with less patenting tend to be more coarsely captured. Figure 1.1 shows a world map of all regions that indicates all countries in the sample with a red color and countries not part of the sample with a white color. In the figure, country borders are drawn in blue and regional borders in yellow. Whenever region and country borders coincide, the yellow region borders are not visible.

Figure 1.1: REGPAT REGIONS



The map shows that REGPAT regions are relatively small (and numerous) in North America (United States, Canada, and Mexico) and Europe. We can link the REGPAT regions with spatial information from two sources: (i) the Geographical Information and Maps (GISCO) database from Eurostat for spatial information

---

<sup>6</sup>The REGPAT database links the Worldwide Statistical Patent Database (PATSTAT) from the European Patent Office (EPO) to 5,633 regions across the globe, utilizing the addresses of the applicants and inventors.

on European countries (NUTS3 regions, 2010), and (ii) the Global Administrative Areas (GADM) spatial database on administrative boundaries for all other countries. We extract the land mass for each region using ArcGIS software after excluding water sheds within the boundaries of a region and normalize the region-specific land mass by the average landmass,  $\frac{1}{N} \sum_S G_r$ .

### 1.3.2 Trade-cost-function Parameters

In constructing trade costs, we employ three ingredients: (i) fast-marching-algorithm-based transportation costs between pairs of  $1^\circ$  grid cells along the lines of Desmet et al. (2018), using passing-through parameters from Allen and Arkolakis (2014)<sup>7</sup>; (ii) a correspondence of these transport costs to the level of REGPAT regions by weighted averaging them within regions as explained in Appendix 1.10.3; (iii) the consideration of discontinuities in trade costs at national borders due to tariffs and linguistic proximity. Tariffs and common language are among the most important factors which are used in parameterizing the international trade-cost function beyond mere transportation costs. We follow the customary approach to specify the trade-elasticity-scaled trade costs as a product of their scaled ad-valorem ingredients – here a transport-cost factor, a tariff factor, and a language factor. We specify the tariff factor between regions  $r$  and  $s$  as  $(1 + tariff_{rs})^{-\theta}$ , where  $tariff_{rs}$  is the weighted applied import tariff on manufactures in 2005 (which differs between most-favored-nation partners and customs-union or free-trade-area members). To acknowledge the language factor in trade costs we follow Melitz and Toubal (2014) and use  $\exp(\rho \times proxl_{rs})$ , where  $proxl_{rs} \in [0, 1]$  is the linguistic proximity and  $\rho = 0.078$  is the corresponding parameter estimate favored in Melitz and Toubal (2014, p.357, Table 3, column 6) on their Automated Similarity Judgment Program (ASJP) measure, which we use here.

---

<sup>7</sup>We modify those costs by making them symmetric (using the average for costs from  $r$  to  $s$  and  $s$  to  $r$ ) and by assuming that intra-cell transport costs are (essentially) zero as is customary in quantitative Ricardian work (see Eaton and Kortum, 2002; Donaldson, 2018).

### 1.3.3 Initial Efficiency Distribution

Simulating the model requires knowledge on the spatial distribution of the initial efficiency ( $\tau_{rt}$ ) in the benchmark year 2005. We use the product-market clearing condition in (1.11) to rewrite  $\tau_{rt}$  as a function of observables, replacing the R&D-worker-specific productivity shifter, which is unknown at this point, by the BGP relationship,  $\tau_{rt} \propto (\bar{L}_r h_r)^{\frac{\theta\gamma_1}{(1-\gamma_2)\xi}}$ . Hence, we can express a scaled version of  $\tau_{rt}$  as follows

$$\tau_{rt}^{(2-\gamma_2)} = \frac{\bar{L}_{rt}^{1-\iota_1} G_r w_{rt}^{1+\theta}}{\int_S w_{st} \bar{L}_{st} G_s \zeta_{rs}^{-\theta} \left[ \int_S \tau_{kt}^{(2-\gamma_2)} \bar{L}_{rt}^{\iota_1} \zeta_{rk}^{-\theta} w_{kt}^{-\theta} dk \right]^{-1} ds}, \quad (1.18)$$

where  $\iota_1 \equiv \alpha - (1 - \mu)\theta$ . We numerically solve for  $\tau_{rt}^{2-\gamma_2}$  by applying a standard contraction mapping procedure and using observed levels of population densities,  $\bar{L}_{rt}$ , and wages,  $w_{rt}$ , for the benchmark year 2005. Population levels are from SEDAC and wage levels from the G-Econ Project, which are both aggregated to the regional level as described in Appendices 1.10.1 and 1.10.2, respectively.<sup>8</sup>

### 1.3.4 Estimation of the Productivity Shifter for R&D Workers

To estimate the R&D-worker-specific productivity shifter governing the BGP and the transition towards it,  $h_r$ , we use (1.15) along with the derived (scaled) initial efficiency distribution from the previous section.<sup>9</sup>

<sup>8</sup>Technical details on the derivation of (1.18) are presented in Appendix 1.7.

<sup>9</sup>Using the BGP relationship to estimate the R&D-worker specific productivity shifter implicitly assumes that all regions in the sample are growing at the steady-state rate already in 2005. Notice that also Desmet et al. (2018, pp. 927 and 929) have to assume that the data are characterized by a BGP in order to determine the relative importance of technology inertia and diffusion parameters. We have to make this assumption also in order to calibrate the productivity levels across regions in the benchmark year and to link it to the R&D-policy variables. For robustness regarding the latter, we ran the analysis only for OECD member countries, including Singapore, and found that the parameter estimates in estimating (1.19) are very similar when running the regression for the mentioned sub-sample relative to the full data-set (see Table 1.7 in the Appendix). However, we should admit that solving for  $\tau_{rt}$  in Section 1.3.3 inevitably requires assuming all places to grow at the BGP rate. Otherwise, the market-clearing condition for goods for all places depends on both  $\{\tau_{rt}, h_{rt}\}$  for every  $\{rt\}$ . Hence, the mentioned robustness analysis should be taken with a grain of salt. We also reran the analysis for all non-OECD member countries (excluding Singapore). In this subsample of 434 regions, only two out of five R&D-policy instruments are used, and the most important instrument in this sample are tax holidays. We present the table summarizing the corresponding results for OECD countries plus Singapore on the one hand (5,199 regions) and the other 434 regions as Table 1.7 in the Appendix.

Taking logs of the BGP relationship and multiplying both sides with the exponent  $(2 - \gamma_2)$  obtains

$$\log \left[ \tau_{rt}^{2-\gamma_2} / \tau_{st}^{2-\gamma_2} \right] = \frac{\theta\gamma_1(2-\gamma_2)}{(1-\gamma_2)\xi} \log \left[ \bar{L}_r / \bar{L}_s \right] + \frac{\theta\gamma_1(2-\gamma_2)}{(1-\gamma_2)\xi} \log \left[ h_r / \bar{h}_s \right]. \quad (1.19)$$

We parameterize  $h_r$  as

$$h_r = \exp(\mathbf{D}_r\beta + |\text{lat}_r|\mathbf{D}_r\gamma + \delta|\text{lat}_r|), \quad (1.20)$$

where  $\mathbf{D}_r$  describes a vector of binary R&D-policy indicators from Bösenberg and Egger (2017) which are measured in the same year as  $h_r$  (here, 2005). The indicators in  $\mathbf{D}_r$  are country-specific and pertain to all regions in a country. Specifically,  $\mathbf{D}_r$  includes a binary indicator variable for partial exemptions of returns on R&D investments, also known as patent boxes ( $D_{\text{patentbox}_r}$ ), R&D investment related grants from the government which act akin to subsidies ( $D_{\text{grants}_r}$ ), tax credit on R&D investments ( $D_{\text{taxcredit}_r}$ ), tax holidays for firms with R&D investment ( $D_{\text{taxholiday}_r}$ ), and any form of deductions of R&D investments from profits other than super deductions ( $D_{\text{deduc}_r}$ ). In any case, a binary indicator is set to unity, if the respective kind of provision is in place in the year 2005 and zero else. We use binary indicators for R&D instruments for one specific reason: combining these instruments into specific rates requires information about the detailed investment structure of firms in each region and time. These data are not available globally.

Additionally, we include an interaction term of each binary R&D policy indicator with the absolute value of the latitude of the region's centroid ( $|\text{lat}_r|\mathbf{D}_r$ ) for two reasons. First, it allows us to account for differences on how productively a region can use an adopted R&D policy instrument depending on its distance to the equator. Notable contributions that have highlighted a relation between a firm's ability to adopt new technologies and its distance to the equator are Theil and Galvez (1995) and Hall and Jones (1997), among others. And, second, it adds variation in the marginal effect of country-level policy instruments across regions. Clearly, the absolute value of the latitude is a better representation for between- rather than within-country variations when considering the ability to adopt new

technologies. However, there is also evidence for a within-country variation as a number of economies in our sample display a clear north-south divide in economic activity, e.g., Italy, France or the US.<sup>10</sup> The specification which takes interactions of the national binary R&D policy environment with latitudes into account allows for general latitude-related patterns of the unobserved investment structure of firms, which affects the bite of the R&D-policy environment for R&D-cost reductions.

We refrain from explicitly modeling any budgetary effects of the considered R&D policy instruments for the following reason. The employed instruments affect the marginal tax rate on returns generated from R&D in a highly nonlinear way. However, as countries do not report specific tax revenues generated from such investments, it is not possible to validate a structural form of the associated nonlinear relationship. From this perspective, it appears customary to resort to a reduced-form nexus between the instruments and innovation and consider treatment effects of the instruments based on this reduced-form nexus by embedding it in the structure of the general equilibrium model.<sup>11</sup>

As the location decision of individuals is endogenous to the productivity potential of a region, we instrument  $\log(\bar{L}_r)$  in the year 2005 with a region-specific remoteness index in logs,  $\log(R_r) = \log(\text{areashare}_r) + \log\left(\frac{1}{N} \sum_S \zeta_{rs}\right)$ .<sup>12</sup> After substituting  $\log(h_r)$  with  $(\mathbf{D}_r\beta + |\text{lat}_r|\mathbf{D}_r\gamma + \delta|\text{lat}_r|)$  according to (1.20), we estimate (1.19) with two-stage least squares (2SLS) to obtain the parameter estimates  $\left\{\frac{\theta\gamma_1(2-\gamma_2)}{(1-\gamma_2)\xi}, \hat{\beta}, \hat{\gamma}, \hat{\delta}\right\}$  based on data for the baseline year  $t = 2005$ .

In Table 1.2, we summarize all variables which inform this procedure. The table is organized in three vertical blocks: the one at the top summarizes moments of the scaled initial efficiency, the land and population distribution as well

---

<sup>10</sup>We tried to estimate (1.19) using a region-specific remoteness index being interacted with each binary R&D-policy indicator. The results are comparable to the ones presented here, but remoteness interactions are less statistically significant than latitude interactions.

<sup>11</sup>Modeling tax revenue effects more explicitly with the R&D tax instruments at hand would require detailed information in the structure of a region's capital stock (the ratio of buildings versus machinery in that stock and its financing, etc.), see Egger and Loretz (2010). As such information is not available for the regions at hand, we resort to the parsimonious approach adopted here.

<sup>12</sup>Notice that trade frictions are among the few exogenous parameters in the model. Hence, they are natural candidate instruments for endogenous variables in the model. Any (highly nonlinear) reduced form of the model would involve trade costs as a determinant of every one of the endogenous variables in the model.

as  $\log(\bar{L}_r)$ , which combines the two and the remoteness index; the one in the center summarizes the elements of  $\mathbf{D}_r$  as well as  $|lat_r|$  used in  $|lat_r|\mathbf{D}_r$ , underlying the parametrization of  $h_r$ ; and the block at the bottom provides information on registered patents per unit of land in region  $r$  from REGPAT, which will be further discussed in the decomposition exercise in Section 1.3.6.

Table 1.2: SUMMARY STATISTICS (2005)

Variable	Mean	Std. Dev.	Min.	Max.
Scaled initial efficiency ( $\tau_{rt}^{2-\gamma_2}$ )	11.487	413.097	5.43e-09	30,142
Population ( $\bar{L}_{rt}G_r$ )	1,108,491	7,637,692	5	2.17e+08
Normalized land ( $G_r$ )	1	11.83	1.9e-04	624.27
$\log(\bar{L}_{rt})$	9.89	2.01	-0.428	16.44
Remoteness ( $R_r$ )	0.171	1.668	0	69.754
<b>R&amp;D-policy indicators</b>				
Dtaxcredit <sub>r</sub>	0.715	0.452	0	1
Dtaxholiday <sub>r</sub>	0.023	0.151	0	1
Dgrants <sub>r</sub>	0.081	0.273	0	1
Dpatentbox <sub>r</sub>	0.022	0.147	0	1
Ddeduc <sub>r</sub>	0.029	0.169	0	1
Absolute latitude ( $ lat_r $ )	40.205	9.583	0.2	74.728
<b>Patents per norm. unit of land</b>				
$Patents_{rt}^{(inv)}$ (2005)	1,278.1	8,648.6	0	297,026.4
$Patentstock_{rt}^{(inv)}$ (1995-2005)	10,366.8	67,655.1	0	2,474,476.5
$Patents_{rt}^{(app)}$ (2005)	1,749.2	20,186.2	0	832,164.6
$Patentstock_{rt}^{(app)}$ (1995-2005)	9,330.5	105,772.2	0	4,488,536.5

Notes:  $Patents_{rt}^{(inv)}$  and  $Patentstock_{rt}^{(inv)}$  refer to a regional denomination of patents in 2005 and patent stocks from 1995-2005, respectively, according to the residence of inventors (*inv*).  $Patents_{rt}^{(app)}$  and  $Patentstock_{rt}^{(app)}$  refer to a regional denomination of patents in 2005 and patent stocks from 1995-2005, respectively, according to the residence of applicants (*app*).

While the information about the population and land data may be interesting to some readers, we suppress a discussion here for the sake of brevity and rather focus on the R&D-policy instruments used in the parametrization of  $h_r$ . The respective indicators suggest that more than two-thirds of the regions operated under a regime with tax credits ( $Dtaxcredit_r$ ), while other R&D-policy instruments were used much less frequently (by fewer countries or by countries with not very fine-grained regions) in 2005. For example, a grants system was applied in only about eight percent of the regions, and deductions, tax holidays, and patent boxes were used in only about two to three percent of the regions.

The parameter estimates and some other statistics based on the aforementioned procedure and data are summarized in Table 1.3. There, we report on marginal effects of the covariates in (1.19) for three specifications. The first column presents the ordinary least squares (OLS) results, while columns (2) and (3) show the results from the 2SLS regression, which takes the potential endogeneity of an individual location decision into account. The second and third columns differ, as the specification in column (3) controls for continent fixed effects, while the one in column (2) does not.<sup>13</sup> Apart from marginal effects of  $\log(\bar{L}_r)$  as well as individual elements in  $\mathbf{D}_r$  and the absolute value of the latitude we report the overall model fit through the correlation of the data with the model prediction as well as the number of observations (regions) used for estimation.<sup>14</sup> As key variables of interest are measured at the country level, all standard errors and test statistics are robust to clustering at the country level.

We document in the upper block of Table 1.3 that the proposed instrument is highly relevant. The OLS and second-stage 2SLS results suggest that more densely populated regions (i.e., ones with higher values of  $\log(\bar{L}_r)$ ) have higher efficiency values, as predicted by the model. Comparing the parameter on  $\log(\bar{L}_r)$  in column (1) to those in column (2) and (3), it becomes evident that accounting for endogeneity not only reduces the importance of population density on efficiency levels but also reveals significant effects of country-specific R&D policy instruments on regional efficiency. The latter is concealed by the bias of the OLS estimates. In particular, tax holidays ( $Dtaxholidays_r$ ) and grants ( $Dgrants_r$ ) tend to raise efficiency according to columns (2) and (3), while patent boxes ( $Dpatentbox_r$ ; a back-end incentive which primarily promotes the ownership but not the invention of patents) reduce efficiency levels.<sup>15</sup> Also regular deductions ( $Ddeduc_r$ ) of R&D investments from profits display a positive effect on efficiency levels. The explanatory power of the model is relatively high, as can be seen from the overall

---

<sup>13</sup>Continent fixed effects inter alia capture the heterogeneity in the granularity of regions as classified in REGPAT. Moreover, they capture a heterogeneity at the macro-regional level in terms of the desirability of patenting among innovative firms.

<sup>14</sup>Clearly, as the elements in  $\mathbf{D}_r$  are binary, what we report is the average effect of an indicator being unity versus zero for the considered R&D tax-policy instruments.

<sup>15</sup>Patent box is the only policy instrument in our analysis for which the invention does not need to have taken place at the same location as where the tax incentive would be enjoyed. This is why the point estimate is likely to differ in sign compared to other instruments.



Table 1.3: ESTIMATION RESULTS (MARGINAL EFFECTS)

	(1)	(2)	(3)
$\log(\tau_r^{2-\gamma_2})$	OLS	2SLS	2SLS
<b>First Stage</b>		$\log(\bar{L}_r)$	$\log(\bar{L}_r)$
$\log(R_r)$		-0.576***	-0.688***
		(0.102)	(0.074)
<b>Second Stage</b>			
$\widehat{\log(\bar{L}_r)}$	1.154***	0.620***	0.593***
	(0.104)	(0.096)	(0.087)
Dtaxcredit <sub>r</sub>	0.216	-0.464	0.915**
	(0.342)	(0.313)	(0.368)
Dtaxholiday <sub>r</sub>	0.873	1.931***	1.299**
	(0.738)	(0.598)	(0.628)
Dgrants <sub>r</sub>	0.602	1.552**	1.838***
	(0.645)	(0.614)	(0.525)
Dpatentbox <sub>r</sub>	-0.301	-0.572	-1.679**
	(0.715)	(0.569)	(0.650)
Ddeduc <sub>r</sub>	0.763	1.300***	0.715**
	(0.502)	(0.352)	(0.410)
$ lat_r $	0.076***	0.039***	0.017*
	(0.008)	(0.011)	(0.009)
continent FE	NO	NO	YES
# obs	5,633	5,633	5,633
Corr. coeff. $\{\widehat{\log(\tau_r^{2-\gamma_2})}; \widehat{\log(\tau_r^{2-\gamma_2})}\}$	0.734	0.707	0.708

Notes: Robust and country-level clustered standard errors in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

fit measured by the correlation coefficient between the data and the model prediction as reported at the bottom of the table. Overall, these results document that, as postulated and hypothesized, a favorable so-called front-end R&D-policy environment indeed appears to have cost-reducing effects on innovation and productivity – which is the very intention of the associated policies – and, hence, boosts productivity as intended in a way which is measurable at the regional level.

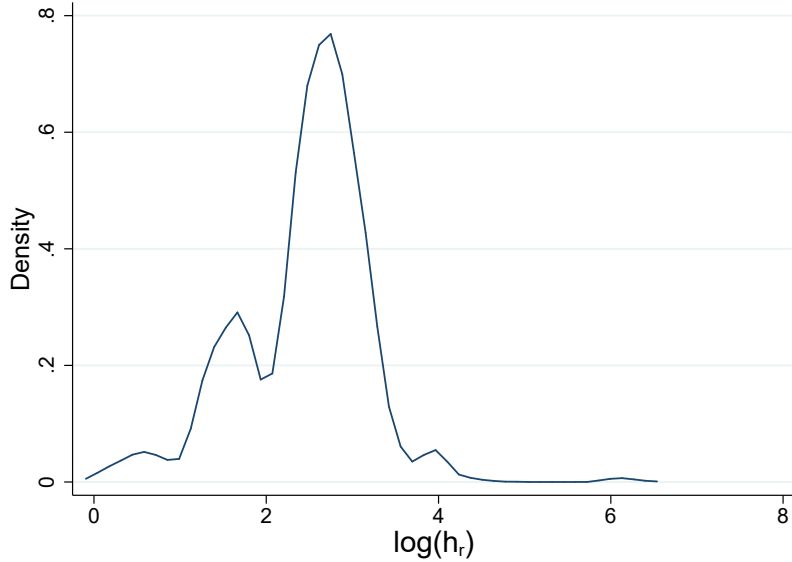
In what follows, we will use the specification in column (3) as the preferred model, since its explanatory power is relatively highest among the two 2SLS mod-

els, and the parameters on R&D policy instruments are all statistically relevant predictors of regional efficiency levels. Given the parameter estimates, we obtain an estimate of  $h_r$  for each region  $r$  in 2005 and the transition towards as well as the BGP as

$$\hat{h}_r = \exp(\mathbf{D}_r \hat{\beta} + |lat_r| \mathbf{D}_r \hat{\gamma} + \hat{\delta} |lat_r|). \quad (1.21)$$

The R&D-policy instruments included in  $\mathbf{D}_r$  jointly contribute to a sizable variation of  $\log(\hat{h}_r)$  in the data. We illustrate the latter by way of a kernel density plot in Figure 1.2.

Figure 1.2: KERNEL DENSITY OF THE ESTIMATED LOG R&D-WORKER-SPECIFIC PRODUCTIVITY SHIFTER



### 1.3.5 Technology and Efficiency-evolution Parameters

Table 1.1 summarizes the assumed values of the technology parameters  $\{\alpha, \theta, \mu\}$  and the efficiency-evolution parameters  $\{\xi, \nu\}$  which we take from others' work. Here, we focus on the two remaining parameters  $\{\gamma_1, \gamma_2\}$  which are elemental but for which existing estimates are not available given the adopted model structure. Specifically, the BGP implies that welfare grows according to (1.17). Taking logs and expressing (1.17) for a finite number of regions obtains

$$\begin{aligned} \log(u_{rt+1}) - \log(u_{rt}) &= \log(y_{rt+1}) - \log(y_{rt}) \\ &= \frac{(1 - \gamma_2)}{\theta} \log\left(\frac{1}{N}\right) + \frac{\gamma_1}{\xi} \log(\Psi) + \frac{1 - \gamma_2}{\theta} \log\left(\sum_S (\bar{L}_s h_s)^{\frac{\theta \gamma_1}{(1 - \gamma_2) \xi}}\right), \end{aligned} \quad (1.22)$$

where  $\Psi \equiv \frac{\gamma_1/\nu}{\gamma_1+\mu\xi}$  and  $N = 5,633$ . Note that equation (1.22) depends on both population density ( $\bar{L}_{rt}$ ) and on real-income ( $y_{rt+1}, y_{rt}$ ). Either type of data is available at the  $1^\circ \times 1^\circ$  resolution from the G-Econ 4.0 Research Project at Yale University. However, as the estimation is informed by parameter values established in the estimation of Section 1.3.4, we employ the population data from SEDAC for consistency.<sup>16</sup>

For identification of the parameters it is useful to see that the left-hand side of (1.22) is indexed by  $t$ , whereas none of the parameters and variables on the right-hand side is. Moreover,  $\gamma_1$  can be expressed as a function of  $\gamma_2$  (and vice versa), and all the other parameters are known at this point. Hence, for a single year,  $\gamma_2$  could be exactly solved for. For identification we pool the mentioned data for  $t \in \{1990, 1995, 2000\}$  and  $t + 5 \in \{1995, 2000, 2005\}$  and approximate the log difference between years  $t + 1$  and  $t$  by the average annual change within any five-year interval. We use the estimated parameter  $\frac{\theta\widehat{\gamma_1(2-\gamma_2)}}{(1-\gamma_2)\xi}$  of Section 1.3.4 and rearrange all parameters dependent on  $\gamma_1$  in (1.22) to express them as a function of  $\gamma_2$ . Noting that  $\gamma_1, \gamma_2 \in (0, 1)$ , we can search for the optimal value of  $\gamma_2$  by doing a grid search on the unit interval with an objective function that minimizes the sum of squared residuals between the left-hand side and the right-hand side of (1.22) for the mentioned three year tuples  $\{t, t + 1\}$  together. Adopting this procedure obtains the grid-search estimates  $\hat{\gamma}_2 = 0.979$  and the implied  $\hat{\gamma}_1 = 0.234$  as listed in Table 1.1.

### 1.3.6 Patented vs. Non-patented Innovations

Patenting is often used as a measure of innovation (see e.g., Griliches, 1990; Nagaoaka et al., 2010). However, not all innovations are patented. In fact, non-patented innovations appear much more common than patented ones on average (see more details on this in the discussion below). The model structure allows us to obtain a measure of the overall innovation level for each region,  $\phi_{rt}$ , and data on patenting permit attributing it to patented innovations versus (residual)

---

<sup>16</sup>Whereas SEDAC provides gridded population data with an output resolution of 30 arc-seconds (approx. 1 km at the equator), the G-Econ project provides the same data on an aggregated  $1^\circ \times 1^\circ$  resolution. We use population data from SEDAC directly to avoid measurement error from aggregation. However, we reran the analysis with population data from the G-Econ Project as a robustness check, and the parameter estimates do not change significantly, when doing so.

non-patented ones. For this, we use data on patent registrations, assuming a Cobb-Douglas relationship:  $\phi_{rt} = (\phi_{rt}^{Patent})^{\alpha_{1r}} (\phi_{rt}^{Rest})^{1-\alpha_{1r}}$ . Taking logs we obtain

$$\log(\phi_{rt}) = \alpha_{1r} \log(\phi_{rt}^{Patent}) + (1 - \alpha_{1r}) \log(\phi_{rt}^{Rest}), \quad (1.23)$$

where  $\phi_{rt} \propto \tau_{rt}^{\frac{1-\gamma_2}{\theta\gamma_1}}$  according to the BGP relationship in (1.15),  $\phi_{rt}^{Patent}$  is a measure of patent registrations from REGPAT,  $\phi_{rt}^{Rest}$  is a measure of non-patented (unobserved) innovations, and  $\alpha_{1r} \in (0, 1)$  is a region-specific Cobb-Douglas weight.

Table 1.2 reports figures on patent registrations at the bottom, which are expressed in normalized units of land,  $G_r$ . The two lines at the top of the respective block pertain to a regional denomination of patents according to the residence of inventors (*inv*), whereas the two lines at the bottom of the respective block pertain to a regional denomination of patents according to the residence of applicants (*app*). For each concept, we report the average normalized patent registration counts for 2005 as well as the patent stock counts from 1995 to 2005. The respective figures suggest that inventions are more dispersed than applications (i.e., applications are more concentrated). This pattern shows in higher first and second moments of patent applications as well as in a higher frequency of zeros across regions in the applications data than the inventions data, which is not obvious from the table.

It is useful to introduce a parametrization of  $\alpha_{1r} \log(\phi_{rt}^{Patent})$  in order to gauge the relative importance of observable patented innovations and unobservable non-patented ones. In particular, we parameterize  $\alpha_{1r} \log(\phi_{rt}^{Patent})$  as a weighted average of the log of the normalized patent stock in a region ( $\log(Patentstock_{rt})$ ) and an interaction term thereof with the log normalized land mass ( $\log(G_r)$ ). The reason for an inclusion of the latter is that REGPAT regions tend to be larger in areas of the globe where patenting is relatively rare, and the mentioned interaction term captures this pattern. Then, using inventor-based patent data, the suggested parametrization reads

$$\alpha_{1r} \log(\phi_{rt}^{Patent(inv)}) \equiv \alpha_2 \log(Patentstock_{rt}^{(inv)}) + \alpha_{3r} \log(Patentstock_{rt}^{(inv)}) \times \log(G_r). \quad (1.24)$$

Based on this, we can replace  $\alpha_{1r} \log(\phi_{rt}^{Patent})$  in (1.23) by the expression on the right-hand side of (1.24) and obtain  $(1 - \alpha_{1r}) \phi_{rt}^{Rest}$  as a residual, in order to yield

region-specific shares for patent-related innovations as  $\hat{\alpha}_{1r} = \hat{\alpha}_2 + \hat{\alpha}_{3r} \log(G_r)$ .<sup>17</sup> According to the data and estimates, (inventor-based) patent-related innovation stocks explain about 42 percent of the variation in  $\log(\phi_{rt})$  (in terms of the  $R^2$ ), and their Cobb-Douglas share  $\hat{\alpha}_{1r}$  ranges from 0.005 to 0.014, with an average of 0.009 and a standard deviation of 0.0008. Hence, the cost share of patented innovations in the generation of all innovations is with about one percent on average relatively small, and it does not vary too starkly in the data.

One may assume that this low cost share is driven by regions in which the patent law is such that the patent stock would over-represent inventive activities and, hence, bias our estimates. Nagaoka et al. (2010) mention this problem by reference to the Japanese patent law. Including a binary indicator for Japan in the analysis, however, does not reveal any significant effect, which we take as evidence that the institutional differences do not seem to play a significant role for our results (when conditioning on the included factors determining endogenous innovation,  $\phi_{rt}$ ).

In fact, the notion of a relatively low cost share of patented innovations in all innovations squares with earlier evidence. For instance, Bloom and Reenen (2000) find a relatively low elasticity of total-factor productivity with respect to patents of about 0.03. Moreover, the evidence in Danguy et al. (2010) suggests that an increase in R&D expenditures raises patents at an elasticity of only 0.12. Moser (2016) documents that, using historical exhibition data, the share of inventors who chose to patent their innovations varied between 5 and 20 percent across industries. Sierotowicz (2015) finds that the average number of patents per million euros of R&D expenditures in leading European Union countries varied between 0.03 in Spain and 0.26 in Germany. Nagaoka et al. (2010) summarize the reasons for why innovations may not be patented. Clearly, in the proposed model a micro foundation of the choice of patenting is absent, and firms are characterized as to rely on both patented and non-patented innovations for technological reasons.<sup>18</sup>

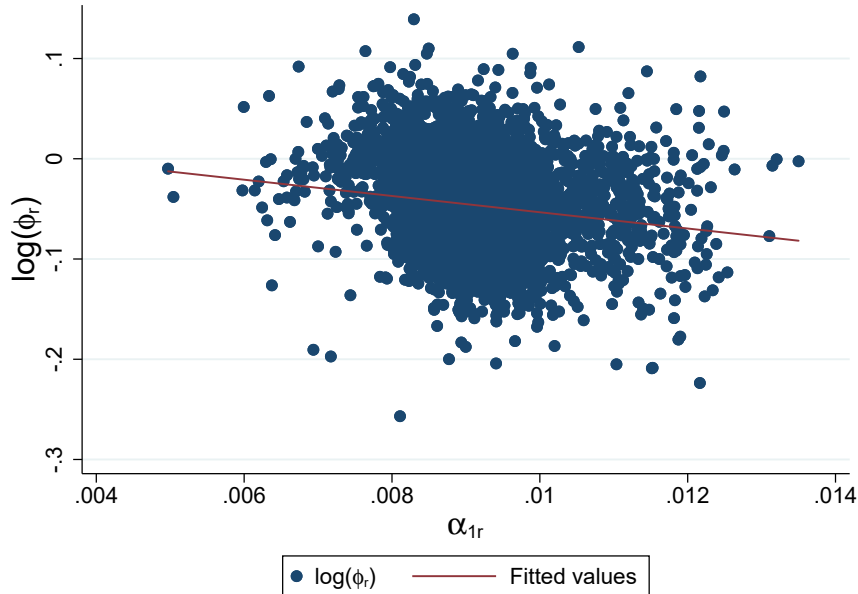
In Figure 1.3, we display the relationship between calibrated log overall innovative productivity in the benchmark year 2005 ( $\log(\phi_r)$ ) and the estimated

---

<sup>17</sup>The results are similar when estimating (1.23) with any other measure of patent registrations that is listed in Table 1.2.

<sup>18</sup>This is consistent with a notion of patented innovations to be technologically different from non-patented ones.

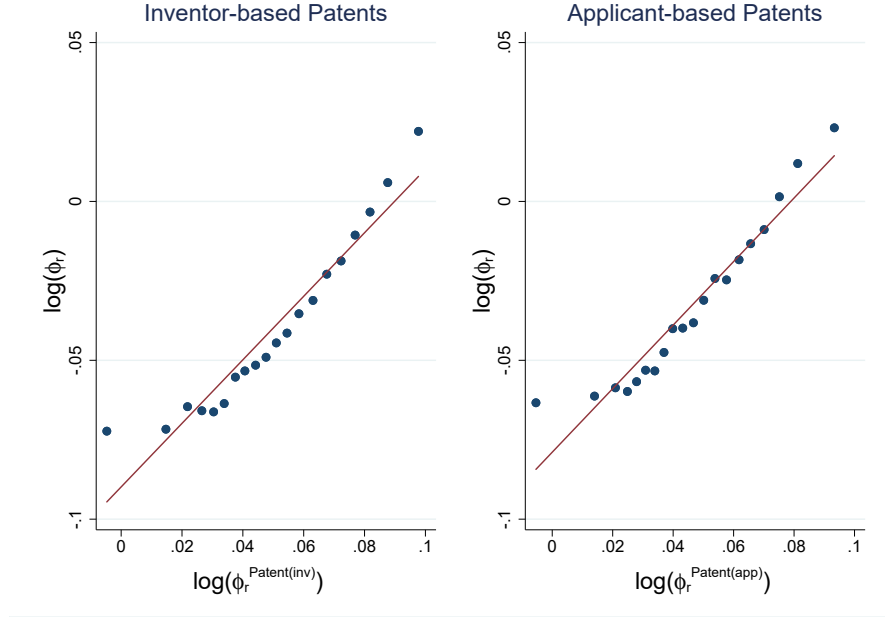
Figure 1.3: LOG OVERALL INNOVATIVE PRODUCTIVITY VS. ESTIMATED REGION-SPECIFIC IMPORTANCE WEIGHT OF PATENTED INNOVATIONS (2005)



region-specific importance weight of patented innovations therein ( $\hat{\alpha}_{1r}$ ). Interestingly, this relationship is negative, though weakly so. This means that in larger regions – where patents are, on average, relatively rare and  $\hat{\alpha}_{1r}$  is relatively high – the overall productivity is relatively lower than on average, in spite of the higher weight of the (fewer) patents. Hence, larger regions enjoy on average a lesser degree of non-patented innovations. However, we should acknowledge that the  $R^2$  underlying the linear relationship in Figure 1.3 is as low as 0.04.

It is worth mentioning that (1.23) postulates a relationship between  $\log(\phi_{rt})$  and  $\log(\phi_{rt}^{Patent})$  which does not vary too starkly around 0.009. Using the estimates  $\{\hat{\alpha}_2, \hat{\alpha}_{3r}\}$  and data on  $\log(Patentstock_{rt}^{(inv)})$  and, alternatively,  $\log(Patentstock_{rt}^{(app)})$  as well as  $\log(G_r)$ , we can plot  $\log(\phi_{rt})$  against the estimates  $\log(\hat{\phi}_{rt}^{Patent})$ . Figure 1.4 does so by way of scatter plots using binned data, where we group regions into 20 equally-sized bins and compute averages within bins for inventor-based (left panel) and applicant-based patents (right panel). The result is a non-parametric visualization of the conditional expectation function, and the figure suggests that the data support relatively well a low variability of the log-linear relationship between  $\log(\phi_{rt})$  and  $\log(\hat{\phi}_{rt}^{Patent})$ , as expected.

Figure 1.4: LOG OVERALL INNOVATIVE PRODUCTIVITY VS. LOG PATENTED INNOVATIONS (2005)



### 1.3.7 Estimating Amenity-function Parameters

Before we can simulate the model and do counterfactual analyses, we need to estimate the amenity-function parameters. We postulate and expect overall amenities to decrease with population density as described in equation (1.7). Taking logs of  $a_{rt} = \bar{a}_{rt} \bar{L}_{rt}^{-\lambda}$  obtains

$$\log(a_{rt}) = -\lambda \log(\bar{L}_{rt}) + \text{const.} + \varepsilon_{rt}^a, \quad (1.25)$$

where  $\log(\bar{a}_{rt})$  is specified as a common constant (*const.*, which measures the average of  $\log(\bar{a}_{rt})$  across all regions) plus a deviation from it ( $\varepsilon_{rt}^a$ , i.e., a disturbance term). Clearly, as population density  $\bar{L}_{rt}$  depends on people's location choice in the model which itself depends on  $a_{rt}$ , it should be treated as endogenous in estimating the region-specific exogenous amenity parameter  $\bar{a}_{rt}$  and the congestion parameter  $\lambda$  based on (1.25). Therefore, we estimate (1.25) by two-stage least squares (2SLS) for the baseline year 2005, instrumenting  $\bar{L}_{rt}$  with a region-specific area-weighted remoteness index,  $R_r = \text{weight}_r^{\text{area}} \left( \frac{1}{N} \sum_S \zeta_{rs} \right)$ , which does not depend on individual location decisions.<sup>19</sup> In order to measure  $\bar{L}_{rt}$  we use gridded

<sup>19</sup>See Footnote 12 for a reasoning regarding this instrumentation strategy.

population data from the Socioeconomic Data and Application Center (SEDAC) which we aggregate to the required (non-gridded) regional level.<sup>20</sup>

To construct the dependent variable based on  $a_{rt}$  in (1.25), we use the structure of the model, substitute the indirect utility (1.8) into (1.10) and solve for  $a_{rt}$  as in equation (1.27) in Appendix 1.7.

Table 1.4: AMENITY PARAMETER ESTIMATION RESULTS

<b>First Stage</b> Dep. Var. $\log(\bar{L}_r)$			<b>Moments of <math>\hat{a}_r \equiv \exp(\widehat{const.} + \hat{\varepsilon}_r^a)</math></b>				
			<b>Mean</b>		<b>Std.Dev.</b>		
$\log(R_r)$	$\rho_1$	-0.473*** (0.014)	60,107		352,390		
<b>Second Stage</b> Dep. Var. $\log(a_{rt})$			<b>5%</b>	<b>10%</b>	<b>50%</b>	<b>90%</b>	<b>95%</b>
$\log(\bar{L}_r)$	$-\lambda$	-0.596*** (0.033)	194.8	440.0	6,272.7	77,787.9	158,529.7
#obs		5,633					

Notes: Robust standard errors in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table 1.4 reports the estimation results from estimating (1.25), with the congestion parameter estimated at a value of  $\hat{\lambda} = 0.596$ . Furthermore, the table reports first-order and second-order moments of  $\hat{a}_{rt}$ . As described above, the region-specific exogenous amenity attribute is defined as  $\hat{a}_{rt} \equiv \exp(\widehat{const.} + \hat{\varepsilon}_{rt}^a)$ . In the general-equilibrium analysis,  $\hat{a}_{rt}$  is kept constant at its level of the year  $t = 2005$  for all subsequent time periods.

## 1.4 Counterfactual Analysis

In the counterfactual equilibrium analysis we focus on the effects on three key variables – place-specific employment, productivity, and welfare – in a scenario where investment incentives towards innovation – except for patent boxes – are abandoned. Effectively, this means that in the counterfactual analysis the R&D-worker-specific productivity shifter equals  $h_r^c = \exp(\hat{\beta}_{PB}DPatentbox_r + \hat{\gamma}_{PB}|lat_r|DPatentbox_r + \hat{\delta}|lat_r|)$ ,  $\forall r \in S$ . We split the analysis in three parts.

First, we investigate how economic outcomes react in response to abandoning incentives towards innovation and distinguish between regions in policy-adopting

<sup>20</sup>Technical details on this aggregation are described in Appendix 1.10.1.



vs. policy-non-adopting countries. Table 1.5 lists all policy-adopting countries for each instrument in the year 2005, according to Bösenberg and Egger (2017). The second part of the analysis concentrates on the role of the treatment-size, exogenous amenities, and remoteness for welfare responses. Lastly, we investigate the role of the patented innovation weight for innovation responses.

Table 1.5: R&D POLICY INSTRUMENTS IN 2005

R&D Policy Instrument	Description	Adopting Countries (in 2005)
Dtaxcredit <sub>r</sub>	Tax credits on R&D investments.	Austria, Canada, China, France, Ireland, Japan, Mexico, Netherlands, Norway, Portugal, South Korea, Spain, Taiwan, US, Venezuela.
Dtaxholiday <sub>r</sub>	Tax holidays for firms with R&D investments.	France, Malaysia, Singapore, Switzerland.
Dgrants <sub>r</sub>	R&D investment related grants from the government.	Germany, Hungary, Ireland, Israel.
Dpatentbox <sub>r</sub>	(Partial) exemption of returns on R&D investments.	France, Hungary.
Ddeduc <sub>r</sub>	Deductions on R&D investments other than super deductions.	Australia, Belgium, Ireland, Japan, South Korea.

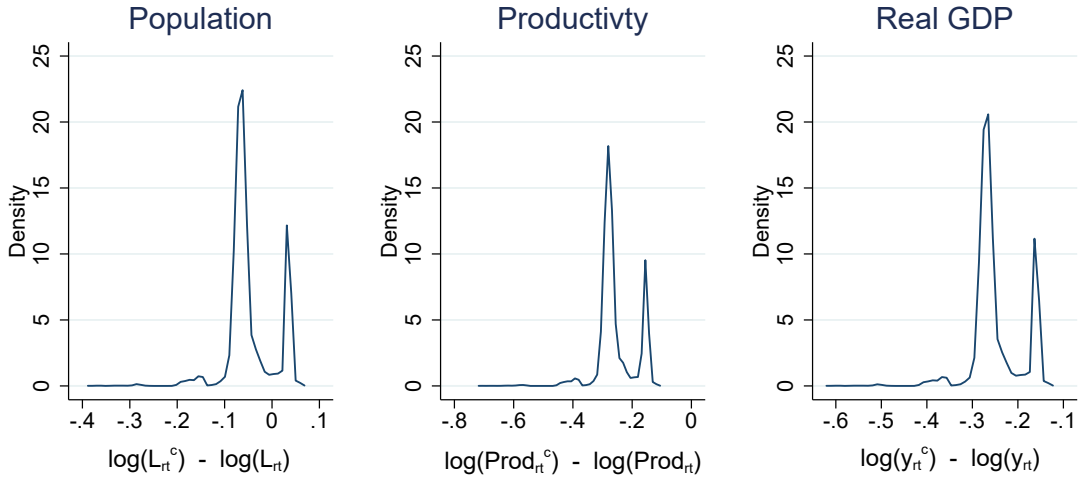
*Notes:* France incl. Guadeloupe, French Guiana, Martinique, Reunion; Netherlands incl. Bonaire; US incl. American Samoa, US Minor Outlying Islands; Australia incl. Cocos Islands; UK incl. Falkland Islands, Gibraltar, Montserrat, Pitcairn, St. Helena.

### 1.4.1 Economic Outcomes and R&D-policy Instruments

In Figure 1.5 we display the variation in long-run ( $T = 100$ ) counterfactual changes in important economic outcomes across all regions in the data. These three outcomes are log population levels ( $\log(\bar{L}_{rt}G_r)$ ), log (overall) productivity levels of the Fréchet location parameter ( $\theta^{-1} \log(\tau_{rt}\bar{L}_{rt}^\alpha)$ ), and log welfare levels of the representative household as expressed in real GDP ( $\log(y_{rt}) = \log(u_{rt}/a_{rt})$ ).

The three panels in the figure suggest that all three economic aggregates are reduced on average when abolishing the considered R&D-policy instruments. However, a non-trivial mass of regions gains population – mainly due to a loss in competition for workers from otherwise less attractive regions that could compete for mobile workers through the use of R&D-policy instruments. The ( $T = 100$ ) long-run changes are quite substantial: some regions gain about eight percent in population while others lose more than 30 percent due to the hypothetical policy change in the long run. Note that the distribution of log changes in population

Figure 1.5: DENSITY ESTIMATES OF COUNTERFACTUAL CHANGES, T=100



*Note:* Superscript  $c$  describes counterfactual values.

levels does not integrate to one. Given the logarithmic transformation of the displayed population change, the hypothetical abolishment of R&D-policy indicators implies that workers move to high-density places (large agglomerations) away from (previously competitive) low-density places. Hence, country-wide R&D tax incentives increase the competitiveness of less attractive low-density places in comparison to similar low-density places abroad, where these policy indicators are not adopted.

Accordingly, these nationally adopted instruments indirectly work as place-based policies in an international context for two reasons. First, they raise the attractiveness of low-density (peripheral and low-amenity) places relative to high-density places at a national level, where the policy is adopted, and, second, they raise the attractiveness of low-density places in policy-adopting countries relative to such places in non-adopting economies.

The effects of abandoning R&D-policy instruments on overall productivity are detrimental throughout and even larger than on population changes; also the welfare changes are negative throughout and almost as large as the productivity changes. The fact that welfare and productivity changes are negative throughout the distribution implies that also countries and their regions which do not use such instruments benefit from their use elsewhere due to technology spillovers.

Table 1.6 presents moments of the real GDP growth rate in the short- (T=10),

Table 1.6: MOMENTS OF REAL GDP GROWTH

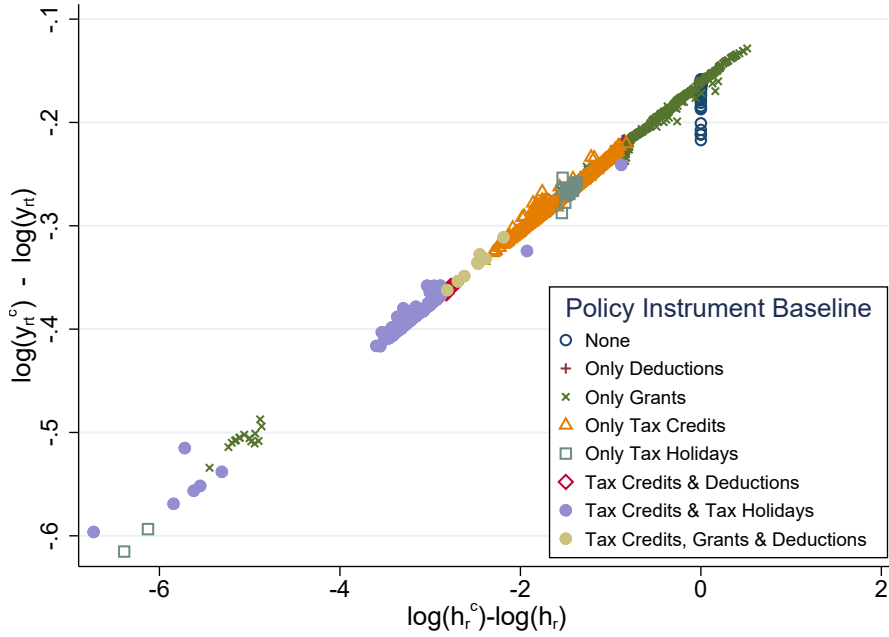
Period	Min	Max	Mean	Std
<b>Baseline in %</b>				
T=10	0.9	6.5	3.6	0.53
T=50	1.6	4.1	2.8	0.24
T=100	2.1	3.1	2.6	0.09
<b>Counterfactual in %</b>				
T=10	0.8	6.5	3.4	0.52
T=50	1.4	3.9	2.5	0.24
T=100	1.9	2.8	2.3	0.09
<b>Counterfactual-Baseline in %pts</b>				
T=10	-0.94	-0.00	-0.23	0.10
T=50	-0.58	-0.16	-0.26	0.05
T=100	-0.38	-0.22	-0.26	0.02

medium- (T=50), and long-run (T=100). The table shows that regions converge towards a model-induced balanced growth path, as the dispersion of growth rates declines with time. The lower panel of the table presents counterfactual-minus-benchmark growth rate differences in percentage points. The corresponding panel suggests that one-tenth of the average long-run real GDP growth can be attributed to the R&D policy instruments alone.

### 1.4.2 The Role of Treatment Size, Remoteness, and Amenities for Welfare Responses

In Figure 1.6, we focus on the welfare changes as in the third panel of Figure 1.5 and plot them against the size of the direct treatment changes – i.e., the change in  $h_r$  induced by abolishing R&D-policy instruments. We differentiate between all possible combinations of R&D-policy instruments that were in place in 2005. Figure 1.6 suggests that the relationship between the treatment change and the associated change in utility is almost linear. Hence, the direct (or partial) effect entails a strong signal for the long-run response. There are indirect effects, which are most obvious for the non-adopting regions in 2005 (about one-percent of the regions displayed in blue circles in the upper-right corner of the figure). The indirect effects on the other regions materialize inter alia as deviations of the data points from the latent linear relationship in Figure 1.6.

Figure 1.6: WELFARE CHANGE AT T=100 AND CHANGES IN  $h_r$

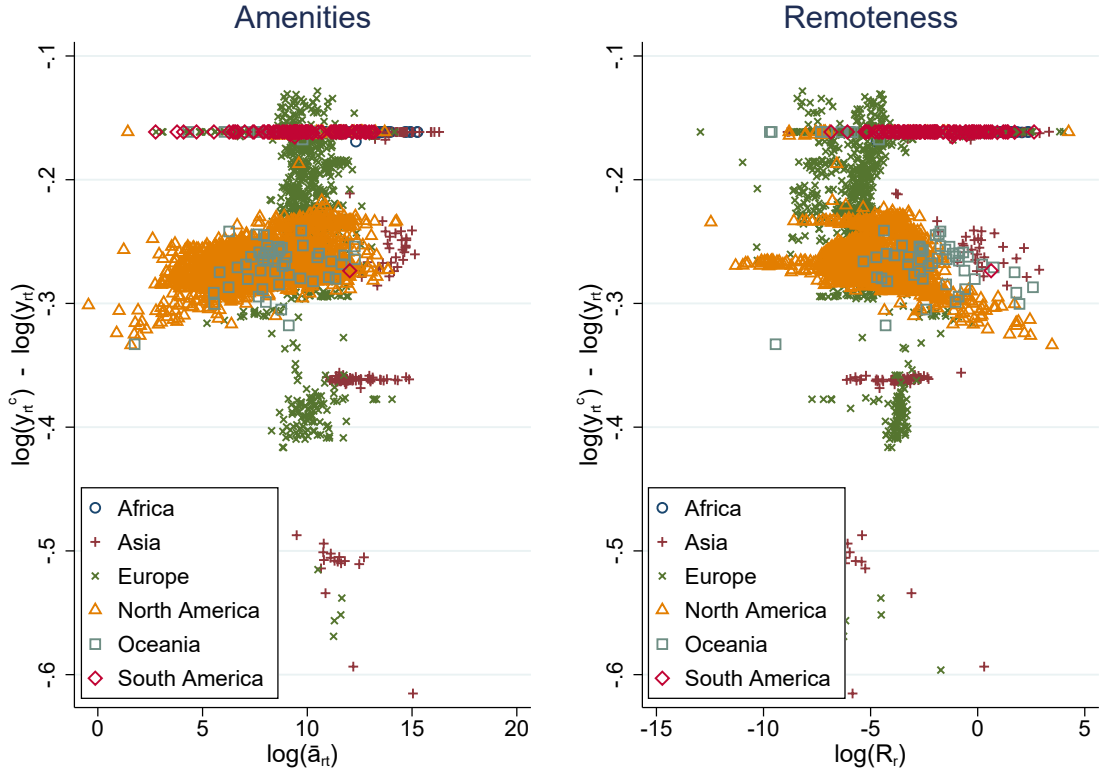


Note: Superscript  $c$  describes counterfactual values.

In Figure 1.7, we shed further light on potentially important mediators of the general-equilibrium treatment effect on welfare changes. While Figure 1.6 alluded to the nexus between the treatment signal and the welfare response, we focus now on the role of exogenous amenities in 2005 ( $\log(\bar{a}_{rt})$ ; in the left panel) and a region’s remoteness ( $\log(R_r)$ ; in the right panel).

In the two panels of Figure 1.7, we use different color to plot the relationships for different continents. Interestingly, the left panel reveals a positive relationship between amenities and the welfare change for regions in North America, Europe and Oceania (including Australia). Hence, a better endowment with good amenities provides for a better insurance against adverse effects from the global abolishment of R&D-stimulating policies. That relationship is still positive but weaker for regions in South America, while it is negative for regions in Africa and Asia. The right panel in Figure 1.7 reveals a negative relationship between remoteness and the welfare change (i.e., more remote regions lose less on welfare from the global abolishment of R&D-promoting policies) for regions in North America, Oceania (including Australia), and also South America, while for regions on other continents this relationship tends to be positive.

Figure 1.7: WELFARE CHANGE AND AMENITY/REMOTENESS LEVELS  
(BY CONTINENTS)



Note: Superscript  $c$  describes counterfactual values.

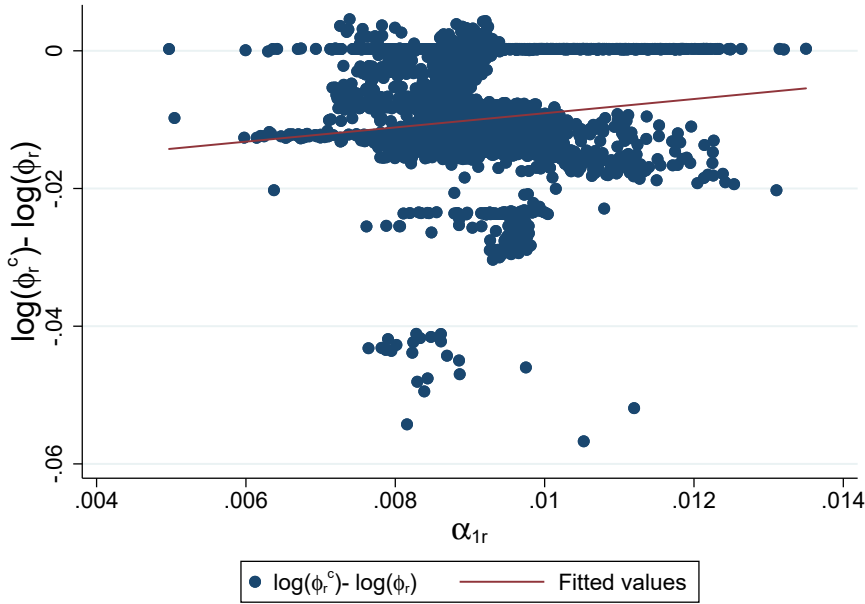
In summary, greater (exogenous) amenities and a higher degree of centrality of a place in the transport network provides for a better quasi-insurance against adverse effects from weak R&D-policy institutions, on average. Moreover, an R&D policy at the national level with a homogeneous direct effect of treatment of all innovations across the places there has indirect place-based effects which are *ceteris paribus* stronger for more peripheral places with less attractive amenities.

### 1.4.3 The Role of the Patented Innovation Weight for Innovation Responses

In Section 1.3.6 we discussed the relative importance of patented and non-patented innovations for overall innovation in the data. In the model, the weight of patented in all innovations is  $\alpha_{1r}$ ; see equation (1.23). The respective parameter is indexed by region, because the importance of patenting depends on the land mass of a region, according to equation (1.24). The latter was introduced to capture the

fact that the delineation of regional borders in the REGPAT database was done according (with region size being inversely related) to the frequency of patenting. However, the overall role of innovation in a region is not a simple function of land mass only but also depends on other fundamentals, such as amenities, market access, etc.

Figure 1.8: LONG-TERM LOG CHANGES IN OVERALL INNOVATION VS. ESTIMATED REGION-SPECIFIC IMPORTANCE WEIGHT OF PATENTED INNOVATIONS (T=100)



*Note:* Superscript  $c$  describes counterfactual values.

In this subsection, we shed light on how the innovation responses of general R&D incentives depend on the relative importance of patenting in a region, as captured by  $\alpha_{1r}$ . In Figure 1.8 we plot the log change in overall (patented plus non-patented) innovation as induced by the counterfactual change in R&D tax instruments against  $\alpha_{1r}$ . There could be a pattern in this relationship, if the land mass of the regions were related to the latitude (as the effectiveness of R&D incentives may vary with the latitude of a region) or to the actual use of instruments (e.g., through the more intensive use of the instruments by countries where patenting is common and, hence, the average land mass of a region is small). The figure suggests that the relationship between the counterfactual-to-benchmark change in  $\log(\phi_r)$  and  $\alpha_{1r}$  is weak: recall that the  $R^2$  of a linear regression of  $\log(\phi_r)$  on  $\alpha_{1r}$  was 0.04, and the one of a linear regression of  $\log(\phi_r^c) - \log(\phi_r)$  on  $\alpha_{1r}$  is 0.02 when considering the change after  $T = 100$  periods. However, the slope of the

regression line for the change is positive. Hence, larger regions – i.e., ones with a lower patent count on average in the outset which are also the ones where the overall innovation level  $\log(\phi_r)$  was low and  $\alpha_{1r}$  was high – are the ones which gain more in overall innovation than on average. It turns out that this relationship is mainly driven by changes in Asia and not on other continents.

## 1.5 Conclusion

This paper outlines a multi-regional model of innovation, production, trade, and factor mobility with a dynamic technology diffusion process. The key parameters of the model are estimated and the model is otherwise calibrated to 5,633 REGPAT regions. One of the main goals of the paper is to provide a quantitative account of the consequences and the value of innovation for regional and national economies as well as the global economy. Since nationally implemented policy instruments towards firm-level R&D are particularly important, we put emphasis on quantifying the role of such incentives. We document that, in spite of their national inception, these instruments affect regions between but also within adopting and non-adopting countries heterogeneously. The degree of heterogeneity depends on the extent of the treatment – how many and which instruments are used and how productively (in terms of its absorptive capacity) a region can use them. Moreover, the degree of heterogeneity depends on other fundamentals such as a region’s integration in the national and international transport network as well as its attractiveness for the location of mobile labor in terms of the available amenities.

One important insight is that the use of policy instruments which are designed to stimulate private R&D are globally beneficial in terms of productivity and welfare. In other words, also countries and their regions who do not use such incentives benefit from their use abroad due to technology spillovers. Also, the long-run relocation effects from a hypothetical abolishment of R&D investment incentives are substantial and lead to a re-shifting of the population towards high-density areas. This is mainly due to a loss in competition for workers from otherwise less attractive regions, which could gain in international competitiveness for mobile

factors through the use of R&D policy instruments.

In line with the previous result, the quantitative analysis suggests that particularly low-amenity, peripheral places – and, on average, ones where the patenting of innovations is less common than elsewhere – benefit relatively more strongly from R&D investment incentives than others. The latter implies that these instruments work as place-based policies. This result is especially true for regions in North America and Oceania, whereas the effect is less predominant in Europe or Asia. Overall, R&D-policy instruments affect endogenous innovations primarily through non-patented innovations, as the estimated range of weights of patented innovations in all innovations is relatively small around the globe.



# Appendix

## 1.6 Estimation Table: Robustness

Table 1.7: ROBUSTNESS ESTIMATION RESULTS (MARGINAL EFFECTS) – SUBSAMPLES

	OECD Countries (with Singapore)			Non-OECD Countries (without Singapore)		
	(1) OLS	(2) 2SLS	(3) 2SLS	(4) OLS	(5) 2SLS	(6) 2SLS
$\log(\tau_r^{2-\gamma_2})$						
<b>First Stage</b>						
$\log(R_r)$		$\log(\bar{L}_r)$ -0.793*** (0.074)	$\log(\bar{L}_r)$ -0.814*** (0.065)		$\log(\bar{L}_r)$ -0.442*** (0.070)	$\log(\bar{L}_r)$ -0.517*** (0.063)
<b>Second Stage</b>						
$\widehat{\log(\bar{L}_r)}$	1.200*** (0.088)	0.620*** (0.063)	0.606*** (0.087)	0.608** (0.274)	0.350 (0.434)	0.283 (0.389)
Dtaxcredit <sub>r</sub>	0.090 (0.421)	-0.650 (0.421)	1.150** (0.448)	-0.080 (0.541)	0.265 (0.669)	-1.103 (0.894)
Dtaxholiday <sub>r</sub>	-0.262 (0.441)	1.098*** (0.320)	0.418* (0.201)	8.843*** (0.648)	8.752*** (0.576)	8.626*** (0.830)
Dgrants <sub>r</sub>	0.265 (0.706)	1.293** (0.636)	1.787*** (0.489)	(omitted)	(omitted)	(omitted)
Dpatentbox <sub>r</sub>	0.710* (0.381)	0.225 (0.308)	-1.238*** (0.419)	(omitted)	(omitted)	(omitted)
Ddeduc <sub>r</sub>	0.564 (0.514)	1.176*** (0.392)	0.660* (0.324)	(omitted)	(omitted)	(omitted)
$ lat_r $	0.079*** (0.009)	0.039*** (0.011)	0.014** (0.007)	0.033 (0.021)	0.023 (0.020)	0.060* (0.033)
continent FE	NO	NO	YES	NO	NO	YES
# obs	5,199	5,199	5,199	434	434	434
Corr. coeff. $\{\log(\tau_r^{2-\gamma_2}); \widehat{\log(\tau_r^{2-\gamma_2})}\}$	0.736	0.704	0.709	0.549	0.417	0.120

Notes: Robust and country-level clustered std. errors in parentheses. In columns (4)-(6) the binary indicators Dgrants<sub>r</sub>, Dpatentbox<sub>r</sub> and Ddeduc<sub>r</sub> are omitted because none of these policy instruments was in place in any of the non-OECD countries in 2005.

## 1.7 Initial Efficiency and Amenity Distribution

To identify the initial efficiency distribution, we need to derive an expression for  $\tau_{rt}$ , using the model structure. To do so, we replace unit costs (1.3) into the bilateral trade share in (1.9), plug it into the product-market clearing (1.11)

and solve for a scaled  $\tau_{rt}$ . At this point, we do not have any information on the R&D-worker-specific productivity shifter  $h_r$ . However, we can use the BGP relationship,  $\tau_{rt} \propto \left(\bar{L}_r h_r\right)^{\frac{\theta\gamma_1}{(1-\gamma_2)\xi}}$ , and replace  $h_r$  as a function of population density and efficiency levels. Then,

$$\tau_{rt}^{(2-\gamma_2)} = \frac{\bar{L}_{rt}^{1-\iota_1} G_r w_{rt}^{1+\theta}}{\int_S w_{st} \bar{L}_{st} G_s \zeta_{rs}^{-\theta} \left[ \int_S \tau_{kt}^{(2-\gamma_2)} \bar{L}_{rt}^{\iota_1} \zeta_{rk}^{-\theta} w_{kt}^{-\theta} dk \right]^{-1} ds}, \quad (1.26)$$

where  $\iota_1 \equiv \alpha - (1 - \mu)\theta$ . Now, we numerically solve for the scaled  $\tau_{rt}$  by applying a standard contraction mapping procedure as it is described in Appendix B.7 in Desmet et al. (2018), and using observed levels of population densities,  $\bar{L}_{rt}$  and wages,  $w_{rt}$  for the benchmark year 2005. Population levels come from SEDAC and wage levels come from the G-Econ Project, which are aggregated to the regional level as described in 1.10.1 and 1.10.2, respectively. Note that  $\bar{L}_{rt}$  represents population density, hence, population levels are divided by normalized land  $G_r$  to obtain  $\bar{L}_{rt}$ .

After learning  $h_r$  and parameters values  $\gamma_1$  and  $\gamma_2$  as described in Section 1.3.4 and 1.3.5, respectively, we identify the initial distribution of amenities,  $a_{rt}$  in the year 2005. To do so, we replace the unit costs (1.3) in the price index and plug the price index into the indirect utility function in (1.8). Then we replace the utility in (1.10) and solve for amenities,  $a_{rt}$ . Then, after defining

$$\Pi_{st} \equiv \bar{L}_{st}^{\iota_1} G_s w_{st}^{-\theta} h_s^{-\theta\gamma_1/\xi} \tau_{st} \zeta_{rs}^{-\theta},$$

$$a_{rt} = \left( \frac{\bar{L}_{rt} G_r}{\bar{L}} \right)^\Omega \frac{1}{w_{rt}} \left[ \int_S (a_{kt} w_{kt})^{1/\Omega} \left( \int_S \Pi_{st} ds \right)^{1/\Omega\theta} dk \right]^\Omega \left[ \int_S \Pi_{kt} dk \right]^{-1/\theta}. \quad (1.27)$$

Again, we apply an iterative procedure to solve for the initial amenity distribution  $a_{rt}$  using observed population densities and wages. With  $a_{rt}$  we estimate the exogenous region-specific amenity-shock  $\bar{a}_{rt}$  as described in Section 1.3.7.

## 1.8 Equilibrium: Existence and Uniqueness

The uniqueness condition in (1.12) can be derived along the lines of Desmet et al. (2018) (see their Section B.3). We can manipulate the system of equations that defines an equilibrium as follows. For the first set of equations, we substitute (1.4) into (1.3) and replace that expression in the price index. Then,

$$P_{rt} = \kappa_0 \left[ \int_S \tau_{st} \bar{L}_{st}^{\alpha-(1-\mu-\gamma_1/\xi)} w_{st}^{-\theta} \zeta_{rs}^{-\theta} h_{st}^{\theta\gamma_1/\xi} ds \right]^{-\frac{1}{\theta}}, \quad (1.28)$$

where  $\kappa_0 = \bar{p} \left(\frac{1}{\mu}\right)^\mu \left(\frac{\xi\nu}{\gamma_1}\right)^{\gamma_1/\xi} \left(\frac{\xi\mu+\gamma_1}{\xi}\right)^{-(1-\mu-\gamma_1/\xi)}$  and  $\bar{p} = \Gamma\left(\frac{1-\sigma}{\theta} + 1\right)^{\frac{1}{1-\sigma}}$ .

Substituting (1.28) into (1.8) gives

$$\left[ \frac{\bar{a}_r}{u_{rt}} \right]^{-\theta} \bar{L}_{rt}^{\theta\lambda} w_{rt}^{-\theta} = \kappa_1 \int_S \tau_{st} \bar{L}_{st}^{\alpha-(1-\mu-\gamma_1/\xi)\theta} w_{st}^{-\theta} \zeta_{rs}^{-\theta} h_{st}^{\theta\gamma_1/\xi} ds, \quad (1.29)$$

where  $\kappa_1 = \left(\kappa_0 \frac{\mu\xi+\gamma_1}{\xi}\right)^{-\theta}$ . For the second set of equations, we insert (1.9) and the price index into the product-market clearing (1.11) so that

$$w_{rt} G_r \bar{L}_{rt} = \bar{p}^{-\theta} \int_S T_{rt} [o_{rt} \zeta_{sr}]^{-\theta} P_{st}^\theta w_{st} G_s \bar{L}_{st} ds. \quad (1.30)$$

Substituting unit costs (1.3) and  $T_{rt} = \tau_{rt} \bar{L}_{rt}^\alpha$ , as well as replacing the price index with the indirect utility in the previous equation yields

$$\tau_{rt}^{-1} w_{rt}^{1+\theta} G_r h_{rt}^{-\frac{\theta\gamma_1}{\xi}} \bar{L}_{rt}^{1-(\alpha-(1-\mu-\gamma_1/\xi)\theta)} = \kappa_1 \int_S \left[ \frac{\bar{a}_s}{u_{st}} \right]^\theta \zeta_{sr}^{-\theta} w_{st}^{1+\theta} G_s \bar{L}_{st}^{1-\lambda\theta} ds. \quad (1.31)$$

Assuming symmetric trade costs, we follow the proof of Theorem 2 in Allen and Arkolakis (2014), which is based on Theorem 2.19 in Zabreyko et al. (1975). Let us introduce the following function  $\bar{f}_r$ , which is the ratio of LHS's of (1.29) and (1.31):

$$\bar{f}_r = \frac{\tau_{rt}^{-1} w_{rt}^{1+\theta} G_r h_{rt}^{-\frac{\theta\gamma_1}{\xi}} \bar{L}_{rt}^{1-(\alpha-(1-\mu-\gamma_1/\xi)\theta)}}{\left[ \frac{\bar{a}_r}{u_{rt}} \right]^{-\theta} \bar{L}_{rt}^{\theta\lambda} w_{rt}^{-\theta}}. \quad (1.32)$$

Equivalently,  $\bar{f}_r$  also equals the RHS's of (1.29) and (1.31) that is

$$\bar{f}_r = \frac{\int_S \left[ \frac{\bar{a}_s}{u_{st}} \right]^\theta \zeta_{sr}^{-\theta} w_{st}^{1+\theta} G_s \bar{L}_{st}^{1-\lambda\theta} ds}{\int_S \tau_{st} \bar{L}_{st}^{\alpha-(1-\mu-\gamma_1/\xi)\theta} w_{st}^{-\theta} \zeta_{rs}^{-\theta} h_{st}^{\theta\gamma_1/\xi} ds}. \quad (1.33)$$

Applying symmetric trade costs,  $\zeta_{rs} = \zeta_{rs}$ , we can rewrite  $\bar{f}_r$  as follows

$$\bar{f}_r = \frac{\int_S \bar{f}_s^{-\lambda} \bar{f}_{sr} ds}{\int_S \bar{f}_s^{-(1+\lambda)} \bar{f}_{sr} ds}, \quad (1.34)$$

where

$$\bar{f}_{sr} = \left[ \frac{\bar{a}_s}{u_{st}} \right]^{\theta(1+\lambda)} \tau_{st}^{-\lambda} G_s^{1+\lambda} \zeta_{sr}^{-\theta} h_{st}^{-\lambda \frac{\theta\gamma_1}{\xi}} w_{st}^{1+\theta+(1+2\theta)\lambda} \bar{L}_{st}^{1-\lambda\theta-\lambda[\alpha-1+(\lambda+\frac{\gamma_1}{\xi}-(1-\mu))\theta]}. \quad (1.35)$$

Rewrite (1.34) as

$$\bar{\bar{f}}_r = \frac{\bar{f}_r^{-\lambda}}{\int_S \bar{f}_s^{-\lambda} \bar{f}_{sr} ds} = \frac{\bar{f}_r^{-(1+\lambda)}}{\int_S \bar{f}_s^{-(1+\lambda)} \bar{f}_{sr} ds}. \quad (1.36)$$

Then, changing the notation to

$$\bar{g}_r = \bar{f}_r^{-\lambda} \quad \text{and} \quad \bar{\bar{g}}_r = \bar{f}_r^{-(1+\lambda)}, \quad (1.37)$$

and rewrite both as follows

$$\bar{g}_r = \int_S \bar{\bar{f}}_r \bar{\bar{f}}_{sr} \bar{g}_s ds \quad \text{and} \quad \bar{\bar{g}}_r = \int_S \bar{\bar{f}}_r \bar{\bar{f}}_{sr} \bar{\bar{g}}_s ds. \quad (1.38)$$

Define  $\bar{\bar{f}}_r \bar{\bar{f}}_{sr}$  as kernel  $K_{sr}$ . Hence,  $\bar{g}_r$  and  $\bar{\bar{g}}_r$  are both solutions to the integral equation

$$x_r = \int_S K_{rs} x_s ds. \quad (1.39)$$

We have to ensure that  $K_{sr}$  is (i) non-negative, (ii) measurable and (iii) square-integrable. Non-negativity holds as  $\bar{\bar{f}}$  and  $\bar{\bar{f}}$  are non-negative. Measurability holds because it can be shown that  $\bar{\bar{f}}$  and  $\bar{\bar{f}}$  are approximately continuous everywhere. Square-integrability holds as long as population at any given location is bounded from below and above. The former is true because by construction population

cannot shrink to zero unless nominal wages are zero or amenities are infinitely high. The latter is true because population at any given location cannot exceed the level of world population  $\bar{L}$ .

Given the properties of  $K_{sr}$ , Theorem 2.19 in Zabreyko et al. (1975) guarantees that there exists a unique (to scale) strictly positive function that satisfies the system of equations in (1.39). Hence,

$$\bar{g}_r = \varpi \bar{g}_r \Rightarrow \bar{f}_r^{-\lambda} = \varpi \bar{f}_r^{-(1+\lambda)} \Rightarrow \bar{f}_r = \varpi, \quad (1.40)$$

where  $\varpi$  is a constant. Therefore, we have

$$\frac{\tau_{rt}^{-1} w_{rt}^{1+\theta} G_r h_{rt}^{-\frac{\theta\gamma_1}{\xi}} \bar{L}_{rt}^{1-(\alpha-(1-\mu-\gamma_1/\xi)\theta)}}{\left[\frac{\bar{a}_r}{u_{rt}}\right]^{-\theta} \bar{L}_{rt}^{\theta\lambda} w_{rt}^{-\theta}} = \varpi, \quad (1.41)$$

and solving for  $w_{rt}$  gives

$$w_{rt} = \bar{w} \left[\frac{\bar{a}_r}{u_{rt}}\right]^{-\frac{\theta}{1+2\theta}} \tau_{rt}^{\frac{1}{1+2\theta}} G_r^{-\frac{1}{1+2\theta}} \bar{L}_{rt}^{\frac{\alpha-1+\left[\lambda+\frac{\gamma_1}{\xi}-[1-\mu]\right]\theta}{1+2\theta}} h_{rt}^{\frac{\theta\gamma_1/\xi}{1+2\theta}}, \quad (1.42)$$

where  $\bar{w} = \varpi^{\frac{1}{1+2\theta}}$ . Substituting (1.42) into (1.29) gives

$$\begin{aligned} & \left[\frac{\bar{a}_r}{u_{rt}}\right]^{-\frac{\theta(1+\theta)}{1+2\theta}} \tau_{rt}^{-\frac{\theta}{1+2\theta}} G_r^{\frac{\theta}{1+2\theta}} \bar{L}_{rt}^{-\lambda\theta-\frac{\theta}{1+2\theta}[\alpha-1+\left[\lambda+\frac{\gamma_1}{\xi}-[1-\mu]\right]\theta]} h_{rt}^{-\frac{\theta(\theta\gamma_1/\xi)}{1+2\theta}} \\ &= \kappa_1 \int_S \left[\frac{\bar{a}_s}{u_{st}}\right]^{\frac{\theta^2}{1+2\theta}} \tau_{st}^{\frac{1+\theta}{1+2\theta}} G_s^{\frac{\theta}{1+2\theta}} \zeta_{rs}^{-\theta} \bar{L}_{st}^{-1-\lambda\theta+\frac{1+\theta}{1+2\theta}[\alpha-1+\left[\lambda+\frac{\gamma_1}{\xi}-[1-\mu]\right]\theta]} h_{st}^{\frac{(1+\theta)(\theta\gamma_1/\xi)}{1+2\theta}} ds. \end{aligned} \quad (1.43)$$

Inserting (1.10) into (1.43) gives

$$\begin{aligned} & \bar{B}_{rt} \hat{u}_{rt}^{\frac{1}{\Omega}[\lambda\theta-\frac{\theta}{1+2\theta}[\alpha-1+\left[\lambda+\frac{\gamma_1}{\xi}-[1-\mu]\right]\theta]]+\frac{\theta(1+\theta)}{1+2\theta}} \\ &= \kappa_1 \int_S \hat{u}_{st}^{\frac{1}{\Omega}[1-\lambda\theta+\frac{1+\theta}{1+2\theta}[\alpha-1+\left[\lambda+\frac{\gamma_1}{\xi}-[1-\mu]\right]\theta]]-\frac{\theta^2}{1+2\theta}} \bar{B}_{st} \zeta_{rs}^{-\theta} ds, \end{aligned} \quad (1.44)$$

where

$$\bar{B}_{rt} = \bar{a}_r^{-\frac{\theta(1+\theta)}{1+2\theta}} \tau_{rt}^{-\frac{\theta}{1+2\theta}} G_r^{\frac{\theta}{1+2\theta}[\alpha+\left[\lambda+\frac{\gamma_1}{\xi}-(1-\mu)\right]\theta]-\lambda\theta} h_{rt}^{-\frac{\theta(\theta\gamma_1/\xi)}{1+2\theta}},$$

and

$$\bar{B}_{st} = \bar{a}_s^{\frac{\theta^2}{1+2\theta}} \tau_{st}^{\frac{1+\theta}{1+2\theta}} G_s^{\frac{\theta}{1+2\theta}-1+\lambda\theta-\frac{1+\theta}{1+2\theta}[\alpha-1+\left[\lambda+\frac{\gamma_1}{\xi}-(1-\mu)\right]\theta]} h_{st}^{\frac{(1+\theta)(\theta\gamma_1/\xi)}{1+2\theta}},$$

and

$$\hat{u}_{rt} = u_{rt} \left[ \frac{\bar{L}}{\int_S u_{kt}^{1/\Omega} dk} \right]^\Omega \left[ 1 - \frac{\theta}{\frac{1}{\Omega} \left[ \left[ \lambda + (1-\mu) - \frac{\gamma_1}{\xi} \right] \theta - \alpha \right] + \theta} \right] \quad (1.45)$$

Rewrite (1.44) as

$$\bar{B}_r f_r^{\tilde{\gamma}_1} = \kappa_1 \int_S \bar{\bar{B}}_s \zeta_{rs}^{-\theta} f_s^{\tilde{\gamma}_2} ds, \quad (1.46)$$

and apply Theorem 2.19 in Zabreyko et al. (1975), then the solution  $f_{(\cdot)}$  to equation (1.46) exists and is unique if (a) the function  $\kappa_1 \bar{\bar{B}}_r^{-1} \bar{\bar{B}}_s \zeta_{rs}^{-\theta}$  is strictly positive and continuous, and (b)  $\left| \frac{\tilde{\gamma}_2}{\tilde{\gamma}_1} \right| \leq 1$ . The latter implies

$$\frac{\frac{1}{\Omega} \left[ 1 - \lambda\theta + \frac{1+\theta}{1+2\theta} \left[ \alpha - 1 + \left[ \lambda + \frac{\gamma_1}{\xi} - [1-\mu] \right] \theta \right] \right] - \frac{\theta^2}{1+2\theta}}{\frac{1}{\Omega} \left[ \lambda\theta - \frac{\theta}{1+2\theta} \left[ \alpha - 1 + \left[ \lambda + \frac{\gamma_1}{\xi} - [1-\mu] \right] \theta \right] \right] + \frac{\theta(1+\theta)}{1+2\theta}} \leq 1,$$

which after some simplification can be written as the uniqueness condition (1.12) as stated in Section 1.2.4

$$\frac{\alpha}{\theta} + \frac{\gamma_1}{\xi} \leq \lambda + 1 - \mu + \Omega.$$

## 1.9 Balanced Growth Path: Derivation

### 1.9.1 Uniqueness and Existence Condition in the BGP

The BGP uniqueness and existence condition is derived along the lines of Desmet et al. (2018). Efficiency evolves according to a endogenous dynamic process in (1.2) and, hence, the growth rate of  $\tau_{rt}$  is given by

$$\frac{\tau_{rt+1}}{\tau_{rt}} = \phi_{rt}^{\theta\gamma_1} \left[ \int_S \frac{W_{rs} \tau_{st}}{\tau_{rt}} ds \right]^{1-\gamma_2}, \quad (1.47)$$

where we define  $W_{rs} \equiv 1/N, \forall rs$  as described in Section 1.2.5. Divide both sides by the corresponding equation for region  $s$ , and rearrange, knowing that  $\frac{\tau_{rt+1}}{\tau_{rt}}$  is constant over time and space and  $\frac{\tau_{st}}{\tau_{rt}}$  is constant over time. Hence,

$$\underbrace{\frac{\frac{\tau_{rt+1}}{\tau_{rt}}}{\frac{\tau_{st+1}}{\tau_{st}}}}_{=1} = \left[ \frac{\tau_{st}}{\tau_{rt}} \right]^{1-\gamma_2} \left[ \frac{\phi_{rt}}{\phi_{st}} \right]^{\theta\gamma_1} \underbrace{\left[ \frac{\int_S \tau_{st} ds}{\int_S \tau_{rt} dr} \right]^{1-\gamma_2}}_{=1} \Rightarrow \frac{\tau_{st}}{\tau_{rt}} = \left[ \frac{\phi_{st}}{\phi_{rt}} \right]^{\frac{\theta\gamma_1}{1-\gamma_2}} = \left[ \frac{\bar{L}_s h_s}{\bar{L}_r h_r} \right]^{\frac{\theta\gamma_1}{(1-\gamma_2)\xi}}, \quad (1.48)$$

where the last equality follows from (1.6). We drop the time subscript to demonstrate that population density remains constant in the BGP. Rewrite the last equation as

$$\bar{L}_s = \left[ \frac{\tau_{st}}{\tau_{rt}} \right]^{\frac{(1-\gamma_2)\xi}{\theta\gamma_1}} \bar{L}_r \frac{h_r}{h_s},$$

and integrate both sides over  $s$  and apply the labor market clearing condition,

$$\int_S G_s \bar{L}_{st} ds = \bar{L} \text{ such that}$$

$$\int_S G_s \bar{L}_s ds = \tau_{rt}^{-\frac{(1-\gamma_2)\xi}{\theta\gamma_1}} \bar{L}_r h_r \int_S G_s \tau_{st}^{\frac{(1-\gamma_2)\xi}{\theta\gamma_1}} h_s^{-1} ds \Rightarrow \tau_{rt} = \tilde{\kappa}_t (h_r \bar{L}_r)^{\frac{\theta\gamma_1}{(1-\gamma_2)\xi}}, \quad (1.49)$$

where  $\tilde{\kappa}_t$  depends on time but not on location. Take the last equation and substitute it into (1.43) such that

$$\begin{aligned} & \left[ \frac{\bar{a}_r}{u_{rt}} \right]^{-\frac{\theta(1+\theta)}{1+2\theta}} G_r^{\frac{\theta}{1+2\theta}} \bar{L}_r^{\lambda\theta - \frac{\theta}{1+2\theta}} \left[ \alpha - 1 + \left[ \lambda + \frac{\gamma_1}{\xi} - [1-\mu] \right] \theta + \frac{\theta\gamma_1}{(1-\gamma_2)\xi} \right] h_r^{-\frac{\theta(\theta\gamma_1/\xi)}{1+2\theta} \left( 1 + \frac{1}{1-\gamma_2} \right)} \\ &= \kappa_1 \tilde{\kappa}_t \int_S \left[ \frac{\bar{a}_s}{u_{st}} \right]^{\frac{\theta^2}{1+2\theta}} G_s^{\frac{\theta}{1+2\theta}} \zeta_{rs}^{-\theta} \bar{L}_s^{1-\lambda\theta + \frac{1+\theta}{1+2\theta}} \left[ \alpha - 1 + \left[ \lambda + \frac{\gamma_1}{\xi} - [1-\mu] \right] \theta + \frac{\theta\gamma_1}{(1-\gamma_2)\xi} \right] \\ & \quad h_s^{\frac{(1+\theta)(\theta\gamma_1/\xi)}{1+2\theta} \left( 1 + \frac{1}{1-\gamma_2} \right)} ds. \end{aligned} \quad (1.50)$$

Inserting (1.10) in (1.50) and rearranging conveniently, yields

$$\begin{aligned} & \bar{D}_r \hat{u}_{rt}^{\frac{1}{\Omega}} \left[ \lambda\theta - \frac{\theta}{1+2\theta} \left[ \alpha - 1 + \left[ \lambda + \frac{\gamma_1}{\xi} - [1-\mu] \right] \theta + \frac{\theta\gamma_1}{(1-\gamma_2)\xi} \right] \right] + \frac{\theta(1+\theta)}{1+2\theta} \\ &= \kappa_1 \tilde{\kappa}_t \int_S \hat{u}_{st}^{\frac{1}{\Omega}} \left[ \lambda\theta + \frac{1+\theta}{1+2\theta} \left[ \alpha - 1 + \left[ \lambda + \frac{\gamma_1}{\xi} - [1-\mu] \right] \theta + \frac{\theta\gamma_1}{(1-\gamma_2)\xi} \right] \right] - \frac{\theta^2}{1+2\theta} \bar{D}_s \zeta_{rs}^{-\theta} ds, \end{aligned} \quad (1.51)$$

where

$$\bar{D}_r = \bar{a}_r^{-\frac{\theta(1+\theta)}{1+2\theta}} G_r^{\frac{\theta}{1+2\theta}} \left[ \alpha + \left[ \lambda + \frac{\gamma_1}{\xi} - [1-\mu] \right] \theta + \frac{\theta\gamma_1}{(1-\gamma_2)\xi} \right] h_r^{-\frac{\theta(\theta\gamma_1/\xi)}{1+2\theta} \left( 1 + \frac{1}{1-\gamma_2} \right)},$$

and

$$\bar{D}_s = \bar{a}_s^{\frac{\theta^2}{1+2\theta}} G_s^{\frac{\theta}{1+2\theta}} \left[ \alpha - 1 + \left[ \lambda + \frac{\gamma_1}{\xi} - [1-\mu] \right] \theta + \frac{\theta\gamma_1}{(1-\gamma_2)\xi} \right] h_s^{\frac{(1+\theta)(\theta\gamma_1/\xi)}{1+2\theta} \left( 1 + \frac{1}{1-\gamma_2} \right)},$$

are exogenously given, and

$$\hat{u}_{rt} = u_{rt} \left[ \frac{\bar{L}}{\int_S u_{kt}^{1/\Omega} dk} \right]^{\Omega} \left[ 1 - \frac{\theta}{\frac{1}{\Omega} \left[ \left[ \lambda + (1-\mu) - \frac{\gamma_1}{\xi} \right] \theta - \alpha - \frac{\theta\gamma_1}{(1-\gamma_2)\xi} \right] + \theta} \right]. \quad (1.52)$$

Analogously to the existence and uniqueness proof in Section 1.8, we can rewrite (1.51) as

$$\bar{D}_r g_r^{\tilde{\gamma}_1} = \kappa_1 \tilde{\kappa}_t \int_S \bar{D}_s \zeta_{rs}^{-\theta} g_s^{\tilde{\gamma}_2} ds. \quad (1.53)$$

According to Theorem 2.19 in Zabreyko et al. (1975)  $g_{(\cdot)}$  is a solution to the system of equations in (1.53) that is unique if  $\left| \frac{\tilde{\gamma}_2}{\tilde{\gamma}_1} \right| \leq 1$ . This condition implies

$$\frac{\frac{1}{\Omega} \left[ \alpha - 1 + \left[ \lambda + \frac{\gamma_1}{\xi} - [1 - \mu] \right] \theta + \frac{\theta \gamma_1}{(1 - \gamma_2) \xi} \right] + \frac{\theta(1 + \theta)}{1 + 2\theta}}{\frac{1}{\Omega} \left[ 1 - \lambda \theta + \frac{1 + \theta}{1 + 2\theta} \left[ \alpha - 1 + \left[ \lambda + \frac{\gamma_1}{\xi} - [1 - \mu] \right] \theta + \frac{\theta \gamma_1}{(1 - \gamma_2) \xi} \right] \right] - \frac{\theta^2}{1 + 2\theta}} \leq 1,$$

from which, after some rearrangement, we get the uniqueness condition in the balanced growth path (1.16) as stated in Section 1.2.5

$$\frac{\alpha}{\theta} + \frac{\gamma_1}{\xi} + \frac{\gamma_1}{[1 - \gamma_2] \xi} \leq \lambda + 1 - \mu + \Omega.$$

## 1.9.2 Growth Rate of Aggregate Welfare

To derive the growth rate of aggregate welfare, we follow again the procedure in Desmet et al. (2018). To start, rewrite (1.42) as follows

$$\tau_{rt} = \bar{w}^{-(1+2\theta)} \left[ \frac{\bar{a}_r}{u_{rt}} \right]^\theta w_r^{1+2\theta} G_r \bar{L}_r^{\frac{1-\alpha + \left[ \lambda + \frac{\gamma_1}{\xi} - [1 - \mu] \right] \theta}{1+2\theta}} h_r^{-\frac{\theta \gamma_1}{\xi}}. \quad (1.54)$$

Substituting the previous equation into (1.49) and solving for  $u_{rt}$  gives

$$u_{rt} = \tilde{\kappa}_t^{\frac{1}{\theta}} E_r, \quad (1.55)$$

where  $E_r$  is only dependent on the location and not on time. Hence,

$$\frac{u_{rt+1}}{u_{rt}} = \left( \frac{\tilde{\kappa}_{t+1}}{\tilde{\kappa}_t} \right)^{\frac{1}{\theta}} = \left( \frac{\tau_{rt+1}}{\tau_{rt}} \right)^{\frac{1}{\theta}}, \quad (1.56)$$

where the last equality follows from (1.49). From (1.47) and (1.48) we know

$$\frac{\tau_{rt+1}}{\tau_{rt}} = \phi_{rt}^{\theta \gamma_1} \left[ \frac{1}{N} \int_S \frac{\tau_{st}}{\tau_{rt}} ds \right]^{1-\gamma_2} = \left( \frac{\gamma_1 / \nu}{\gamma_1 + \mu \xi} \bar{L}_r h_r \right)^{\frac{\theta \gamma_1}{\xi}} \left[ \frac{1}{N} \int_S \left( \frac{\bar{L}_s h_s}{\bar{L}_r h_r} \right)^{\frac{\theta \gamma_1}{(1-\gamma_2) \xi}} ds \right]^{1-\gamma_2}. \quad (1.57)$$



Rearranging the previous equation and substituting it into (1.56) gives

$$\frac{u_{rt+1}}{u_{rt}} = \left[ \frac{1}{N} \right]^{\frac{1-\gamma_2}{\theta}} \left[ \frac{\gamma_1/\nu}{\gamma_1 + \mu\xi} \right]^{\frac{\theta\gamma_1}{\xi}} \left( \int_S (\bar{L}_s h_s)^{\frac{\theta\gamma_1}{[1-\gamma_2]\xi}} ds \right)^{\frac{1-\gamma_2}{\theta}}.$$

## 1.10 Data Aggregation

Our unit of interest are REGPAT regions. We use gridded data with different resolution for which we need an aggregation strategy to the regional level. Hereafter, we discuss the aggregation strategy for each data source separately.

### 1.10.1 Population Data from SEDAC

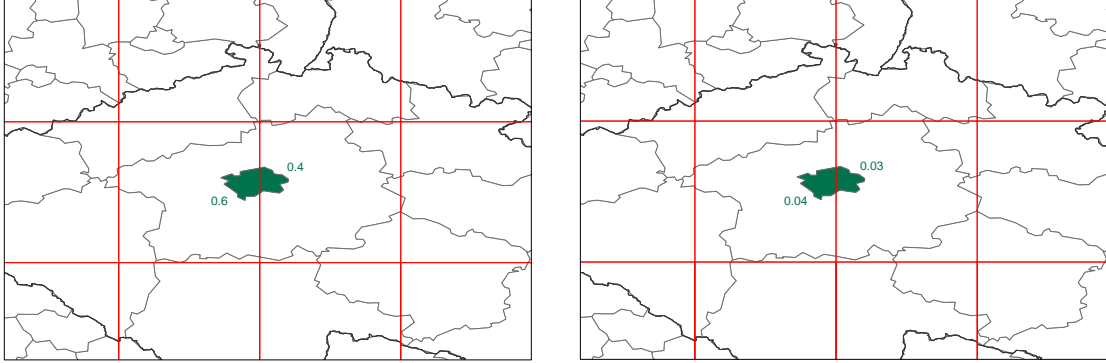
The Socioeconomic Data and Application Center (SEDAC) provides gridded population data with an output resolution of 30 arc-seconds (approximately 1 km at the equator). As the size of each grid cell is smaller than the smallest region in our data, we simply sum up the population count over all grid cells falling within the regional border.

### 1.10.2 Population and GDP from G-Econ Project

The Geographically based Economic Data (G-Econ) project at Yale University provides SEDAC gridded population data aggregated to the 1° by 1° resolution (approximately 100km by 100km at the equator), which is about the same size as second level political entities in most countries. Besides population data, the G-Econ project offers gridded GDP data (gross cell product at purchasing power parity (PPP)) at the 1° by 1° resolution. We assign population and GDP values to each region through an area-weighted average aggregation. Figure 1.9 illustrates how the area-weights are assigned in the case of GDP data (left panel) and population count data (right panel). In both panels, the green area represents the urban region of Prague in the Czech Republic, which falls into two different grid cells (bordered in red). Therefore, the GDP value of Prague is equal to six-tenth of the left grid cell plus four-tenth of the right grid cell. In the case of population count data, we construct the area-weight as the part of Prague that falls into the grid cell relative to the overall area of the grid cell. Hence, the population count

of Prague is four-hundreds of the left grid cell plus three-hundreds of the right grid cell.

Figure 1.9: AGGREGATION FOR DATA WITH ONE DEGREE RESOLUTION



### 1.10.3 Transportation Costs based on the Fast-marching Algorithm

We derive the fast-marching-algorithm-based transportation costs between pairs of  $1^\circ$  grid cells along the lines of Desmet et al. (2018). To find a correspondence of these transportation costs to the level of REGPAT regions, we employ an area-weighted average assignment. The area-weights are constructed as the share of regional area falling into a grid cell relative to the total regional area (see left panel of Figure 1.9). Our averaging procedure can be best explained using matrix notation. Let  $W_{nx1}$  be the vector of area-weights for  $n$  sub-regions, where a sub-region refers to an intersection between a REGPAT region and a one-degree grid cell area. Furthermore, we define the fast-marching transportation costs matrix as  $T_{n \times n}$ , which is blown up from the number of one-degree grid cells to the number of  $n$  sub-regions, using information on sub-region intersections with one-degree grid cells from ArcGIS. Lastly, we need a correspondence of sub-region to the final set of REGPAT regions  $r$  and define a selector matrix  $S_{n \times r}$  using ArcGIS, where  $r$  is equal to 5,633. Then the regional transportation costs  $T_{r \times r}$  can be obtained as follows

$$T_{r \times r} = W'_{n \times r} T_{n \times n} W_{n \times r}, \quad (1.58)$$

where  $W_{m \times r} = (W_{m \times 1} l'_{m \times 1}) \circ S_{m \times r}$ .

## Chapter 2

# Natural City Growth in the People's Republic of China

*This is the original manuscript of an article published by MIT Press in the Asian Development Review 34(2), available online: [https://doi.org/10.1162/adev\\_a\\_00095](https://doi.org/10.1162/adev_a_00095)*

### 2.1 Introduction

With the increase in global population, the change in urbanization rates around the world is a startling dynamic phenomenon. While in 1994 only 30% of the world population lived in cities as defined by national statistical offices, about 54% of the population did in 2014.<sup>1</sup> In the People's Republic of China (henceforth China), which has been among the most dynamic economies over the last quarter of a century, almost 25% of the population has moved to urban areas within the past two decades. China's National New-type Urbanization Plan, 2014-2020 targets an urbanization rate of 60% by 2020. While urbanization is often measured as the increase in the population within the administrative boundaries of cities, urbanization in a broad sense is driven by three phenomena: (i) the increase in population density (and economic activity) within the administrative boundaries of existing urban zones, (ii) the increase in population density (and economic activity) in areas in the vicinity of administrative urban zones through the growth of Metropolitan Statistical Areas (MSAs), and (iii) (to a lesser extent) the physical

---

<sup>1</sup>World Bank. 2015. *World Development Indicators*.

growth of the administrative areas of cities.<sup>2</sup> This paper focuses on the first two phenomena, which are objects of interest in the theoretical and empirical urban economics literature focusing on city growth; urban sprawl, which goes hand in hand with the formation of densely populated urban subcenters; and the decentralization of economic activity (see, for example, Fujita and Ogawa, 1980, 1982; Henderson and Mitra, 1996; Glaeser and Kahn, 2001, 2004; McMillen and Smith, 2003; Burchfield et al., 2006; Garcia-Lopez et al., 2017).

Unlike in many other countries, China's city growth in the recent past has been governed by regulations. The country's one-child policy, which had been instituted in its most restrictive form between 1978 and 2015, led to a slump in overall population growth, reduced the growth rate of cities, and slowed the average urbanization rate. Furthermore, the *hukou* (household registration) system has restricted the internal migration of people to urban centers by limiting access to public goods such as health care, schools, universities, and official housing. Finally, the inception of Special Economic Zones (SEZs) has ensured the protection of the private property rights of foreign investors, alleviated taxes and tariffs, regulated the policy of land usage, and liberalized economic and labor laws in geographically confined zones. According to Wang (2013), most major cities in China's 326 municipalities hosted some sort of SEZ by 2006. A consideration of these regulatory provisions – apart from factors capturing the economic attractiveness and amenities in cities – appears relevant as they may lead to a gap between actual and optimal city size in China, thereby affecting the associated economies of scale and scope (see, for example, Au and Henderson, 2006a,b; Desmet and Rossi-Hansberg, 2013), and resulting in potentially significant output losses.

China's extensive investments in transport infrastructure, particularly road and railway networks, have fundamentally reshaped the structure of its urban areas. In the early 1990s, the Chinese government began to renew and upgrade its transport infrastructure, which caused previously underdeveloped regions to grow faster as

---

<sup>2</sup>The term MSA is mostly used in the context of the study of cities in the United States. In Europe, the literature primarily refers to a Functional Urban Area, which essentially describes the same concept of agglomerations measured by a minimum density of the population according to census data. In this paper, we utilize the term natural cities to indicate something similar, though it is based on the measurement of a city by remote-sensing (night-light radiance) data in conjunction with the City Clustering Algorithm.

industries started to decentralize (see Banerjee et al., 2012; Faber, 2014; Baum-Snow et al., 2016, 2017). For example, Baum-Snow et al. (2017) find that suburban ring roads have displaced an average of about 50% of central city industrial gross domestic product (GDP) to the outskirts of cities, while marginal radial railroads have displaced an additional 20%. Similarly, Baum-Snow et al. (2016) argue that expanded regional highway networks in China have had a negative average effect on local population density, causing a reallocation of economic activity and altering the structure of the country's cities.

The focus of this paper is on the growth of *natural cities*, which are defined as connected places with a minimum night-light radiance as a measure of place- and time-specific economic activity (Henderson et al., 2012), and which are associated with China's 300 largest administrative cities over the period 1992-2013. One major merit of using remote-sensing data to define cities is that such data are available at much higher frequency than population census data. Furthermore, the data collection itself is much more homogeneous in terms of timing and concept. The data suggest that China's natural cities grew rapidly between 1992 and 2010 before shrinking to some extent in the last few years of the review period, which might be attributable to the detrimental effects of the recent global financial crisis. We document this phenomenon for all cities in terms of descriptive statistics and illustrate it for two major agglomerations, Beijing and Shanghai. This paper explores these developments using econometric analysis and identifies institutional factors – as reflected in the proliferation of SEZs and the provisions of the hukou system – and infrastructure accessibility as being important determinants of natural city growth. We highlight the effects of road and railway accessibility, and illustrate that shocks to infrastructure can be expected to induce relatively rapid adjustments in natural city size over the next 20 years.

The remainder of the paper is organized as follows. Section 2.2 introduces the definition of a natural city employed in this paper and outlines the measurement thereof. The data and their descriptive statistics, empirical strategy, and results are presented in Section 2.3. Section 2.4 concludes.

## 2.2 Natural City Borders in China, 1992-2013

In this paper, we employ a definition of city boundaries based on what we call natural borders. Natural city borders relate to the well-known concepts of MSAs and Functional Urban Areas (FUAs), which measure city size by activity rather than administrative boundaries (see, for example, Zipf, 1949; Krugman, 1996; Eaton and Eckstein, 1997; Harris Dobkins and Ioannides, 2001; Ioannides and Overman, 2003; Eeckhout, 2009; Rozenfeld et al., 2011). A general motivation to use a city definition based on either MSA or FUA is that they capture more accurately the extent of urban units, going beyond – and sometimes integrating several units with – administrative boundaries. When looking at emerging urban areas, especially in transition economies such as China, the study of MSAs and FUAs follows an economic rather than an administrative logic. We define the boundaries of natural cities based on the City Clustering Algorithm (CCA) (Rozenfeld et al., 2008, 2011), which we apply to remote-sensing (night-light radiance) data collected from satellites (Burchfield et al., 2006; Henderson et al., 2012). We measure the average night-light radiance in places that are 3 kilometers (km) in length by 3 km in width.<sup>3</sup> We are facing a trade-off between portraying and approximating the boundaries of small cities, especially in the early phases of the sample period, and the tractability of the data, particularly the application of the CCA.<sup>4</sup> The former requires sufficiently small places and the latter sufficiently few places. For those reasons, the consideration of 3 km x 3 km places was the finest-grained grid we could use given the time constraints. In general, one major advantage of using remote-sensing data to define natural cities is that annual data are available between 1992 and 2013, while MSA and FUA data are based on population censuses and therefore only available at lower frequency.

We consider the 300 biggest administrative cities in China by population as of the year 2000.<sup>5</sup> Figure 2.1 shows a map of China and the location of the centroids

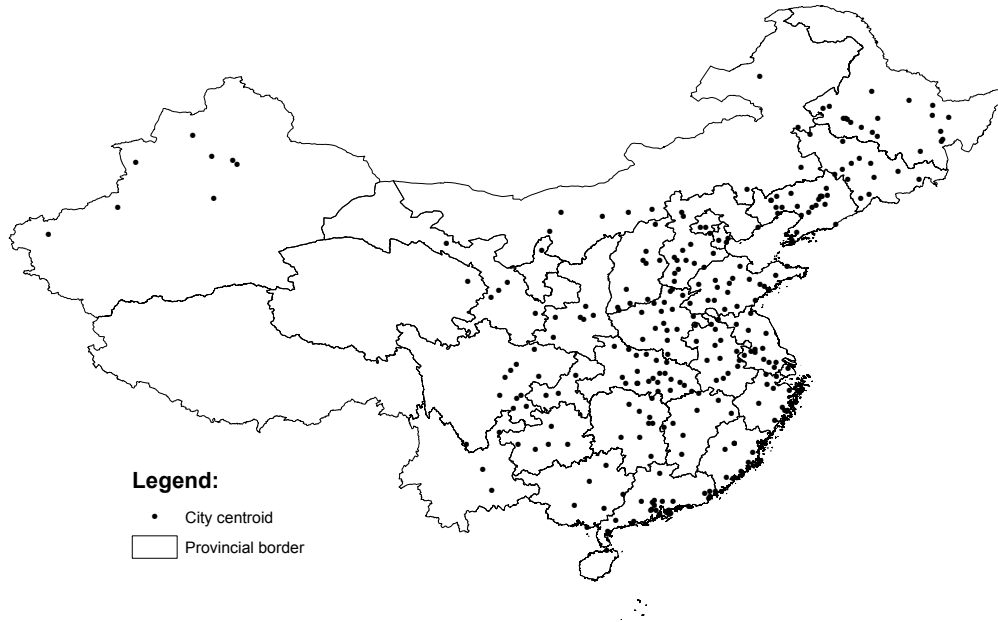
---

<sup>3</sup>Individual places in China bordering water or other boundaries may be smaller in size than 3 km x 3km.

<sup>4</sup>The distribution of city sizes in the sample is presented in Figure 2.5.

<sup>5</sup>A list of the 300 biggest Chinese cities by population in 2000 is presented in Table 2.7 in the Appendix. Moreover, we report a list of all natural cities by size in 2000 in Table 2.8. There are three different administrative levels of cities in the Chinese urban system: municipalities, prefecture-level cities, and county-level cities. With regard to the empirical analysis we use administrative boundary information on the county-level only. For further information on this

Figure 2.1: CENTROIDS OF 300 BIGGEST ADMINISTRATIVE CITIES IN CHINA  
(BY POPULATION IN 2000)



of all 300 cities covered. Very few cities are located in western China, while there is a particularly high density in the vicinity of the coastal belt, which is not surprising provided the high degree of economic activity through international trade in that area.

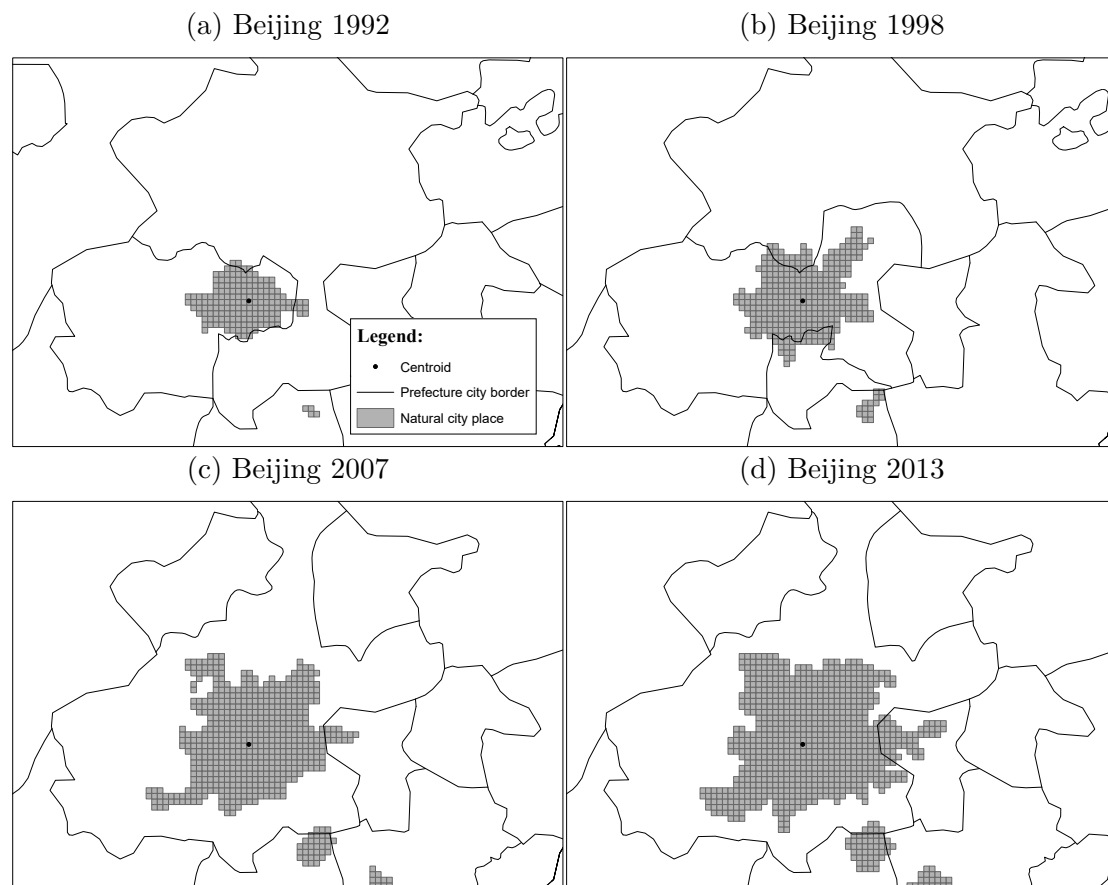
The object of interest in this study are the aforementioned 3 km x 3km places. We define natural city borders on a uniform grid of such places for all cities in the sample. On this grid, we assign a place to a *natural city* in a year if (i) the average night-light radiance on the square exceeds a value of 40; and (ii) it is located near a cluster of places with average night-light radiance over 40, including the place that contains the city centroid (based on the CCA algorithm). We employ Version 4 of the Defense Meteorological Satellites Program Operational Linescan System to measure night-light radiance at the pixel level (Croft, 1978). The remote-sensing (night-light radiance) data therein take on values between 0 (no light) and 63 (maximum light). Night-light radiance data per pixel are available for all years between 1992 and 2013 based on pictures from six different satellites (F10, F12, F14, F15, F16, and F18)<sup>6</sup>, with some years covered by two satellites. We chose

point please proceed to Section 2.3.1.

<sup>6</sup>The satellite identifiers correspond to those used by the Defense Meteorological Satellites Program. For further information, please see National Oceanic and Atmospheric Administration. <https://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html>

the data such that the number of satellites they come from is minimized (F10 for 1992-1993, F12 for 1994-1999, F15 for 2000-2004, F16 for 2005-2009, and F18 for 2010-2013). The data comprise a raw-data version as well as a stable-data version, where the latter ensures that the data are not conflated by fire or firework incidents, or clouds or any other weather conditions. In this paper, we use the stable-light data version and compute the mean of radiance across all pixels within each place. In the final data set, we include all those places that were assigned to be in a natural city in any year between 1992 and 2013, and we track these places over the entire review period.

Figure 2.2: NATURAL CITY OVER TIME – BEIJING

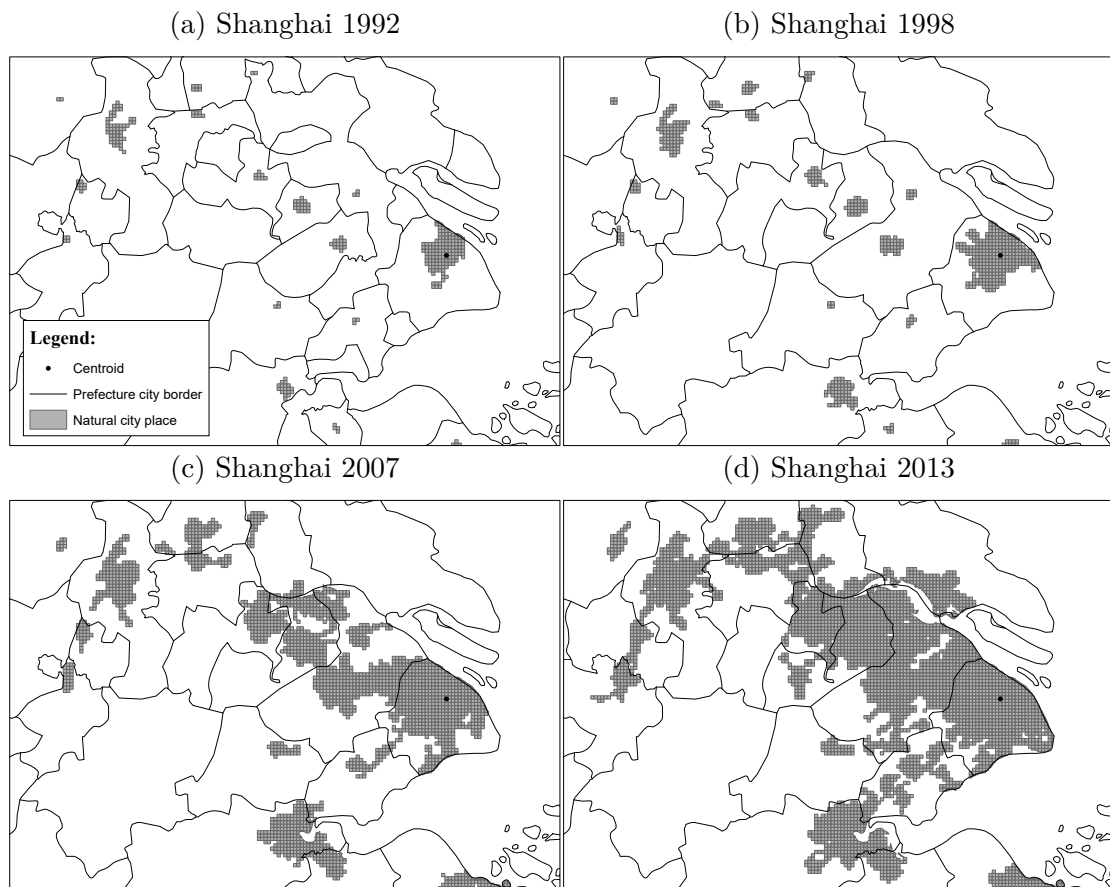


Figures 2.2 and 2.3 delineate the natural city with its city centroid (black dot) and administrative boundaries for Beijing and Shanghai for the years 1992, 1998, 2007, and 2013. In every panel, gray grids represent places that constitute the natural city in that particular year. Prefecture-level administrative city boundaries are indicated in black.<sup>7</sup> In the case of Beijing, we observe that its natural city size

<sup>7</sup>Figures 2.2 and 2.3 show prefecture-level administrative boundaries. However, in the es-



Figure 2.3: NATURAL CITY OVER TIME – SHANGHAI



grew remarkably over the entire sample period. Especially from 1998 onward, the natural city of Beijing grew outward toward the northeast, which could be partly related to the 1993 opening of the Airport Expressway linking central Beijing to the Beijing Capital International Airport. Additional infrastructure investments to improve airport connectivity (e.g., Airport Express Subway) in preparation for the 2008 Olympic Games may have also contributed to the northeast developing more rapidly than other parts of Beijing.

Similar to Beijing, Shanghai’s natural city grew over the entire review period and mostly integrated urban areas along the downstream part of the Yangtze River. The example of Shanghai illustrates that, especially toward the end of the review period, several administrative cities merged into one natural super-city. The natural city of Shanghai in 1992 contained only one administrative centroid, while by 2013 it had incorporated a number of formerly distinct administrative

---

timination, all variables that include information on administrative boundaries rely on county-level boundaries as those boundaries represent the city-size distribution in a better way than prefecture-level boundaries.

and natural cities along the Yangtze River into one natural super-city of Shanghai. However, in spite of the general growth of natural cities through 2007-2013, many natural cities, including Beijing and Shanghai, shrank between 2010 and 2013, most likely as a consequence of the global financial crisis (Figure 2.4).

Figure 2.4: SHRINKING NATURAL CITIES FROM 2010 TO 2013

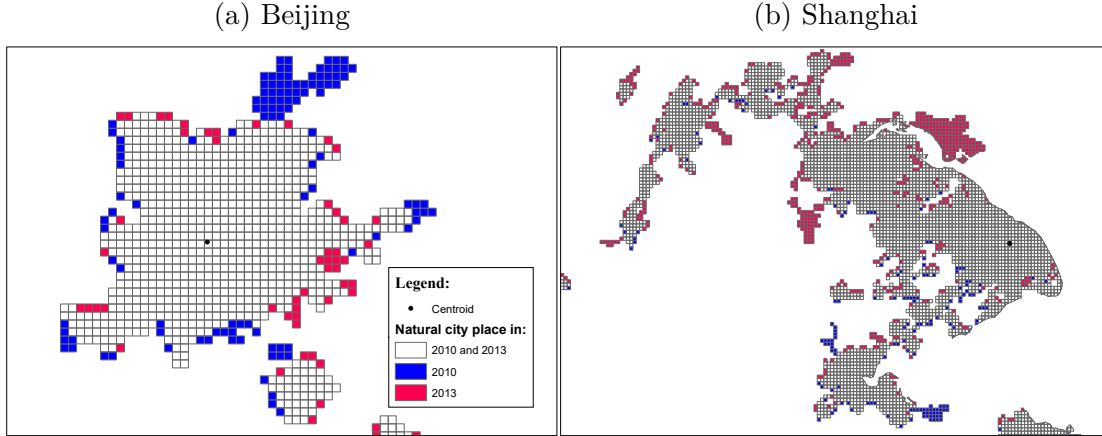


Table 2.1 reports average unconditional transition probabilities for natural city places for the whole sample of places considered. The table suggests that there is a high degree of persistence from one year to another: 92% of all natural city places keep their status, while about 90% of all places outside the natural city boundary remain outside that boundary from one year to another. The probability of acquiring natural city status amounts to 10%, while losing natural city status occurs in 7% of all cases from one year to another. The latter development is almost entirely driven by transitions during 2010-2013, reflecting China's economic downturn in the aftermath of the global financial crisis.

Table 2.1: TRANSITION MATRIX

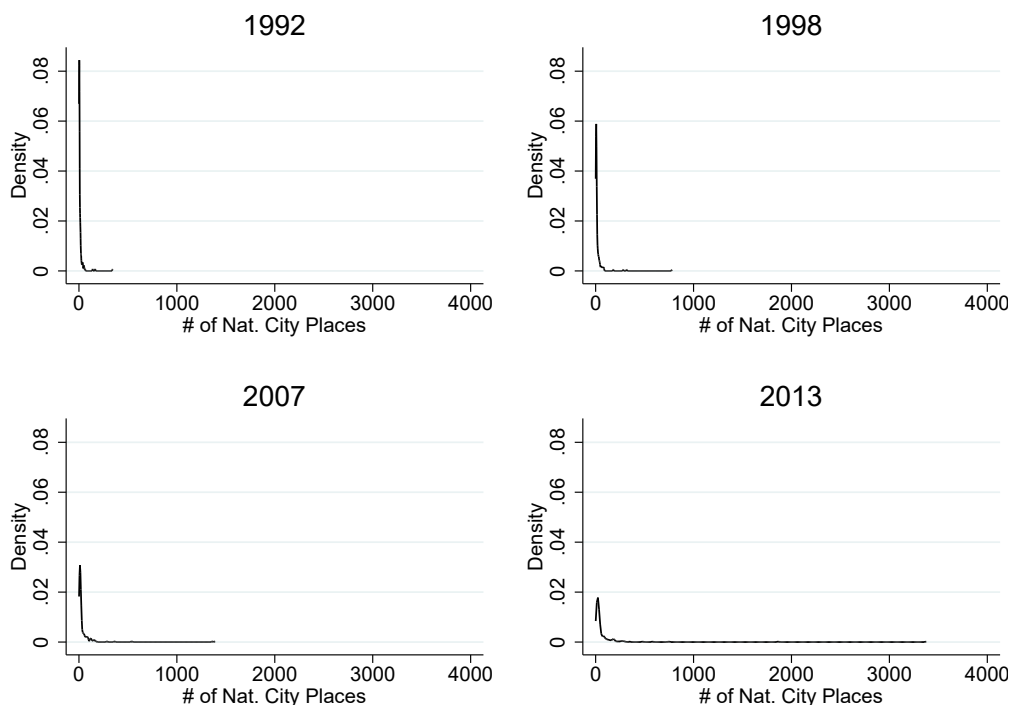
		<u>Target</u>		
		<i>Nat</i> = 1	<i>Nat</i> = 0	<b>Total</b>
Origin	<i>Nat</i> = 1	92.63	7.37	100
	<i>Nat</i> = 0	10.06	89.94	100
<b>Total</b>		39.65	60.35	100

*Abbreviation: Nat* = natural city.

Per Table 2.1, the average natural city size is expected to grow over the sample period. Figure 2.5 draws kernel density estimates of natural city sizes for the years

1992, 1998, 2007, and 2013. In each of the four panels of Figure 2.5, the horizontal axis shows the number of 3 km x 3 km places in a natural city. We observe that the average natural city size, reflected in the total number of places covered, increases remarkably with time. Especially in the beginning of the review period, the density mass is concentrated in the left tail of the distribution, indicating a great number of relatively small natural cities and only a small number of very large super-cities in the sample. Later in the review period, the degree of dispersion in terms of natural city size increases and the density mass in the left tail of the distribution gets smaller.

Figure 2.5: KERNEL DENSITY ESTIMATES OF NATURAL CITY SIZE ACROSS CITIES



## 2.3 Drivers of Natural City Growth

In this section, we introduce all variables included in the subsequent empirical analysis.

### 2.3.1 Data

We use average (night-light) radiance data in a 3 km x 3 km place  $i$  at period  $t$  as the dependent variable to measure economic activity in that area. The variable

$radiance_{it}$  is continuous and censored from below as well as from above, ranging from 0 (no light) to 63 (maximum light). Information on the source and the processing of the radiance data can be found in Section 2.2.

We identify three key categories of variables that drive natural city growth: geographical, climate, and institutional variables. The geographical variables include the following distance measures, some of which are time-variant (indexed by both  $i$  and  $t$ ) and others are not (indexed by  $i$  only): distance to the administrative city center ( $dist\ to\ center_i$ ), distance to the administrative city border ( $dist\ to\ adborder_{it}$ ), distance to the nearest waterway ( $dist\ to\ water_i$ ), distance to the ocean ( $dist\ to\ ocean_i$ ), distance to the nearest road ( $dist\ to\ road_i$ ) and distance to the nearest railway line ( $dist\ to\ rail_i$ ). The geographical variables include a binary indicator that is unity if a place lies within the administrative boundary of the city centroid, and zero otherwise ( $within\ admin\ boundary_{it}$ ). Except for  $dist\ to\ adborder_{it}$  and  $within\ admin\ boundary_{it}$ , which utilize annual information on administrative boundaries (at the county-level) from the China Data Center at University of Michigan, all distances of places are taken from OpenStreetMap using ArcGIS software.<sup>8</sup> Furthermore, we utilize topological information in the form of a measurement of altitude ( $altitude_i$ ) from WorldClim Global Climate Data, and we control for the geographical location of each centroid by using information on its longitude and latitude from ArcGIS. For instance, Chinese cities near the coast grew faster due to better accessibility to sea transport, which attracted foreign direct investment and was further stimulated by the formation of SEZs.

We use the following time-invariant climate data: the average annual rainfall during the period of observation ( $rain_i$ ), the average annual temperature ( $temperature_i$ ), and the average annual temperature variation ( $sd\ temperature_i$ ; as measured by the standard deviation). Gridded climate data are available from WorldClim Global Climate Data.

The institutional variables represent two types of institutional changes that governed China’s urban growth: reforms in the *hukou* system and the formation

---

<sup>8</sup>OpenStreetMap information is based on the most recent network information available only. Distance to the nearest road includes all types of different roads (e.g., private roads, lower-capacity highways, higher-capacity highways, and limited-access highways). Distance to the nearest railway line includes all types of railway lines (e.g., subway lines and inter-provincial railway lines).

of SEZs. Between the late 1970s and mid-2000s, a period which is referred to as the first wave of *hukou* reforms, restrictions on movement and work were lifted, which led to a large inflow of rural workers into urban areas. In most provinces, the scale of reforms varied with city size. Generally, reforms have had little impact on institutions in the most attractive urban areas such as provincial capitals and large cities along the coastal belt. To capture the different effects, we introduce three binary indicators –  $small_{it}$ ,  $medium_{it}$  and  $large_{it}$  – which are unity if a province applied their latest *hukou* reform to small, medium, and large cities, respectively, and zero otherwise. A combined effect of these reforms is captured in the binary indicator  $hukou_{it}$ , which is unity if either of the three, two out of three, or all three city size indicator variables are unity, and zero otherwise. Time-variant information on the extent of the latest *hukou* reform by province during the period 1998-2008 is available in the Organisation for Economic Co-operation and Development Economic Surveys: China (2013).

SEZs are geographic regions that are typically characterized by liberal economic policies designed to attract foreign investors and enhance economic activity. In this paper, we use the term SEZ as a generic term for all types of special economic zones and open areas, including Free Trade Zones, Economic and Technology Development Zones, and open coastal cities, among others. Wang (2013) characterizes four big waves in the formation of SEZs in China (1979-1985, 1986-1990, 1991-1995, and 1996-2007) and lists the corresponding municipalities that were designated as SEZs in each of the first three waves. This allows us to code three different binary indicator variables –  $firstwave_{it}$ ,  $secondwave_{it}$ , and  $thirdwave_{it}$  – of which the former two are time-variant because of the time variation in administrative city boundaries. The third variable is time-variant because in our coding there is no treatment of places and cities prior to 1995. We also include the combined effect of the three waves that is captured in the binary indicator  $SEZ_{it}$ , which is unity if any one of the three, two of the three, or all three SEZ wave indicator variables are unity, and zero otherwise. Since the information on SEZs provided in Wang (2013) pertains to the municipality level, and while data utilized here vary by place, we assume that all places within the treated municipalities were affected by SEZs in the same way.

As an additional control variable we include the population density ( $popdens_{i1990}$ ) in 1990.<sup>9</sup>

### 2.3.2 Descriptive Statistics

Table 2.2 summarizes descriptive features of all variables by natural city status (within a natural city,  $Nat=1$ ; outside of a natural city,  $Nat=0$ ; Average) and reports the mean and standard deviation for each variable.

Table 2.2 indicates that places within a natural city are on average 1.4 times closer to the city centroid than places outside of a natural city. Similarly, places inside are 1.1 times closer to the coast, 1.3 times closer to waterways, 1.6 times closer to the nearest road, and 1.6 times closer to the nearest railway line. As expected, places are on average much closer to the nearest road (0.3 km) than to the nearest railway line (3.5 km). We also observe that places inside a natural city are closer to the nearest administrative border since administrative areas close to the considered city centroids are smaller in the average year than areas outside of the considered administrative city centers. Places within and outside of natural cities do not differ in terms of their average location in terms of longitude and latitude, but they differ in terms of altitude: places inside natural cities have an average altitude 1.2 times lower than places outside. Only about 30% of all places in the data lie within the administrative boundaries of one of the 300 major city centroids in our sample. By comparison, 62% of all places are located inside natural cities in the average year. Finally, places inside and outside natural cities do not significantly differ in terms of average precipitation and temperature.

Table 2.2 further suggests that places inside natural cities are more densely populated and more luminous in the beginning of our study period (1.5 times and 2.3 times, respectively). Over the entire study period, both places inside and outside of natural cities have a higher radiance level than they did in 1992. Places inside of a natural city appear to experience a relatively stronger increase in radiance during the study period. These places are also an average of 2.5 times more luminous than places outside of a natural city over the entire study period.

---

<sup>9</sup>Gridded population density data for 1990 by 2.5 arc-minute grid cells are available from the Socioeconomic Data and Applications Center. <http://dx.doi.org/10.7927/H4XK8CG2>.

Table 2.2: SUMMARY STATISTICS – IN AND OUT NATURAL CITY AND TOTAL

	<u>Nat=1</u>		<u>Nat= 0</u>		<u>Average</u>	
	Mean	SD	Mean	SD	Mean	SD
<b>Geography</b>						
<i>dist to road<sub>i</sub></i> (in km)	0.23	0.32	0.37	0.44	0.32	0.40
<i>dist to rail<sub>i</sub></i> (in km)	2.50	3.26	4.05	4.68	3.46	4.26
<i>dist to ocean<sub>i</sub></i> (in km)	261.61	413.37	293.31	426.36	281.14	421.70
<i>dist to water<sub>i</sub></i> (in km)	1.65	1.94	2.17	2.45	2.00	2.28
<i>dist to center<sub>i</sub></i> (in km)	12.39	9.60	17.28	11.48	15.40	11.06
<i>dist to adborder<sub>it</sub></i> (in km)	3.15	3.07	3.76	3.35	3.52	3.26
<i>within admin boundary<sub>it</sub></i>	0.33	0.47	0.26	0.44	0.29	0.45
<i>altitude<sub>i</sub></i> (in m)	152.4	327.5	176.3	359.7	167.1	347.9
<i>longitude<sub>i</sub></i>	116.30	6.94	116.30	7.19	116.30	7.10
<i>latitude<sub>i</sub></i>	33.47	7.03	33.65	6.31	33.58	6.59
<b>Climate</b>						
<i>rain<sub>i</sub></i> (in mm)	95.37	47.32	92.17	41.83	93.40	44.04
<i>temperature<sub>i</sub></i> (in °C)	14.52	5.40	14.24	4.93	14.35	5.12
<i>sd temperature<sub>i</sub></i> (in °C)	9.13	2.72	9.23	2.48	9.19	2.57
<b>hukou<sub>it</sub></b>	0.69	0.46	0.38	0.48	0.50	0.50
<i>small<sub>it</sub></i>	0.65	0.48	0.37	0.48	0.48	0.50
<i>medium<sub>it</sub></i>	0.56	0.50	0.30	0.46	0.40	0.49
<i>large<sub>it</sub></i>	0.54	0.50	0.29	0.45	0.39	0.49
<b>SEZ<sub>it</sub></b>	0.69	0.46	0.54	0.50	0.60	0.49
<i>firstwave<sub>it</sub></i>	0.04	0.19	0.02	0.14	0.03	0.16
<i>secondwave<sub>it</sub></i>	0.46	0.50	0.37	0.48	0.41	0.49
<i>thirdwave<sub>it</sub></i>	0.39	0.49	0.27	0.45	0.32	0.47
<b>Miscellaneous</b>						
<i>popdens<sub>i1990</sub></i> (in ppl/km <sup>2</sup> )	1209	1837	812	775	964	1305
<i>radiance<sub>i1992</sub></i>	29.67	17.79	13.02	8.93	19.41	15.37
<i>radiance<sub>it</sub></i> (0-63)	53.93	6.95	21.78	11.58	34.12	18.59
# observations	266,613		166,061		432,674	

Abbreviations: °C = degree Celsius, km = kilometer, m = meter, mm = millimeter, Nat = natural city, SD= standard deviation, ppl/km<sup>2</sup> = people per squared kilometer, SEZ= Special Economic Zone.

Table 2.3 summarizes descriptive statistics (mean and standard deviation) for all time-variant variables by year, 1992, 1998, 2007, and 2013. Table 2.3 suggests that the latest wave of *hukou* started applying in small cities – 7.5% of all places in the sample were treated in 1998 – before reaching medium-sized and large cities after 1998. Given that the *hukou* data are coded at the provincial level and that we consider the 300 biggest administrative cities in China, it is not surprising that by 2013 almost 93% of all places in the sample had experienced some degree of *hukou* reform. Concerning the SEZ indicators, the first wave of reforms (1979-

Table 2.3: SUMMARY STATISTICS – AVERAGES FOR 1992, 1998, 2007, 2013

	<u>1992</u>		<u>1998</u>		<u>2007</u>		<u>2013</u>	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
<b>hukou<sub>it</sub></b>	0.041	0.198	0.075	0.263	0.842	0.364	0.928	0.258
<i>small<sub>it</sub></i>	0.041	0.198	0.075	0.263	0.842	0.364	0.850	0.357
<i>medium<sub>it</sub></i>	0.000	0.000	0.000	0.000	0.724	0.447	0.773	0.419
<i>large<sub>it</sub></i>	0.000	0.000	0.000	0.000	0.698	0.459	0.742	0.437
<b>SEZ<sub>it</sub></b>	0.415	0.493	0.623	0.485	0.625	0.484	0.610	0.488
<i>firstwave<sub>it</sub></i>	0.026	0.160	0.027	0.161	0.027	0.161	0.027	0.161
<i>secondwave<sub>it</sub></i>	0.389	0.488	0.410	0.492	0.410	0.492	0.395	0.489
<i>thirdwave<sub>it</sub></i>	0.000	0.000	0.347	0.476	0.349	0.477	0.349	0.477
<i>dist to adborder<sub>it</sub></i> (in km)	3.61	3.35	3.54	3.29	3.50	3.23	3.52	3.22
<i>within admin boundary<sub>it</sub></i>	0.296	0.456	0.285	0.451	0.288	0.453	0.298	0.457
<i>radiance<sub>it</sub></i> (0-63)	19.41	15.37	27.58	17.18	41.68	15.19	53.88	8.72
# observations	19,667		19,667		19,667		19,667	

Abbreviations: SD = standard deviation, km = kilometer.

1985) included a relatively small number of places, with only 2.6% of all places treated during this wave, whereas the second (1986-1990) and third (1991-1995) waves applied to more than one third of all places in the sample. Consequently, about 62.3% of all places were assigned to an SEZ as of 1995. Finally, Table 2.3 indicates that the average night-light radiance ( $radiance_{it}$ ) increased from 19.4 in 1992 to 53.9 in 2013.

### 2.3.3 Econometric Approach

In this subsection, we outline the econometric model used to estimate coefficients on the suspected determinants of the (night-light) luminosity of place  $i$  in year  $t$ ,  $radiance_{it}$ . Two features of the dependent variable are worth mentioning: (i) it is censored from below at 0 and from above at 63, and (ii) it appears to be serially correlated.<sup>10</sup>

<sup>10</sup>Even though the original night-light radiance data take on integer values only, the dependent variable used here is continuous over the entire range of the data as we take the average of the night-light radiance across the pixels within a 3 km x 3km place.



To respect both the double-censoring and auto-correlation through equicorrelation (accruing to the repeated observation of places over time and the presence of place-specific effects) and through inertia, we postulate a dynamic Tobit model with double-censoring and random effects. We account for dynamic adjustment by letting  $radiance_{it}$  be a function of its first-, second-, and third-lagged values  $\mathbf{R}_{it}=(radiance_{it-1}, radiance_{it-2}, radiance_{it-3})$ , respectively, and estimate it along the lines of Wooldridge (2005). Accordingly, the endogeneity of the lagged dependent variables on the right-hand side of the model – through the presence of time-invariant random shocks,  $\mu_i$ , in the models – can be acknowledged by properly specifying the initial conditions of the process Hsiao (2014).

Subsume all exogenous drivers of  $radiance_{it}$  in the common vector  $\mathbf{X}_{it}$ , and let  $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)'$  be the unknown parameters on  $\mathbf{R}_{it}$  and  $\boldsymbol{\beta}$  be the unknown parameters on  $\mathbf{X}_{it}$ . Furthermore, let  $\varepsilon_{it}$  be the (normalized) remainder disturbances in the processes. Then, we may introduce a latent, uncensored, normal counterpart to  $radiance_{it}$ ,  $radiance_{it}^*$ , and relate the two of them as follows:

$$radiance_{it} = \begin{cases} 0 & \text{if } radiance_{it}^* \leq 0 \\ radiance_{it}^* & \text{if } 0 < radiance_{it}^* \leq 63 \\ 63 & \text{if } radiance_{it}^* > 63 \end{cases} \quad . \quad (2.1)$$

Moreover, we may specify the latent variable  $radiance_{it}^*$  in a linear fashion as a function of the parameters of interest through

$$radiance_{it}^* = \mathbf{R}_{it}\boldsymbol{\alpha} + \mathbf{X}_{it}\boldsymbol{\beta} + \mu_i + \varepsilon_{it}. \quad (2.2)$$

For estimation of equation (2.2), we employ two alternative sets of initial conditions for  $\mathbf{R}_{it}$ . One involves the observed radiance in the initial year of the data,  $radiance_{i1992}$ , and the other one additionally involves the time averages of all time-variant variables in  $\mathbf{X}_{it}$ . Since the functional form of the dynamic Tobit model with double-censoring is nonlinear and  $\mathbf{X}_{it}$  includes squared values of some of the determinants, we will report marginal effects only, as is customary with nonlinear models.

### 2.3.4 Results

Table 2.4 summarizes the estimated effects of the lagged dependent variables associated with  $\hat{\alpha}$ , but only a subset of the effect estimates associated with  $\hat{\beta}$ .<sup>11</sup> For instance, we do not report the effects pertaining to variables used for the modeling of the initial condition with averages of the time-variant variables. Since the models are dynamic, the reported estimates should be interpreted as short-run effects materializing within a 3-year time window. Moreover, for the binary variables in  $\mathbf{X}_{it}$  (e.g., the four variables relating to either *hukou* or SEZ each), we compare the average of the conditional mean when the variable takes on a value of unity for all places with the one when the variable takes on a value of zero for all places (Greene, 2017). In Column (1) of Table 2.4, we model the initial condition as a function of the radiance in the initial year,  $radiance_{i1992}$ . In Column (2), the initial condition additionally includes the time averages of all time-variant variables. On a final note, the magnitudes of the total short-run effects of continuous variables in Table 2.4 should only be compared across such variables after normalization (e.g., by scaling them with the standard deviation of the respective variables in Table 2.3).

As the signs of significant effects do not differ qualitatively between Columns (1) and (2), and since the estimation of Column (2) is less efficient than for Column (1), we focus on the effects in Column (1). While we observe that *hukou* and SEZ variables induce significant effects on  $radiance_{it}$ , we skip discussion of those effects here for the sake of brevity. Similarly, we forego discussion of the effects of geography and climate that are also reported in Table 2.4. In what follows, we focus on the effects of infrastructure, particularly roads and railways, near a place.

Two things stand out regarding these effects: (i) greater distance to transport infrastructure – such as roads, railway lines, and waterways – reduces the

---

<sup>11</sup>Table 2.9 provides effects estimates akin to the dynamic Tobit model in Table 2.4 based on three alternative specifications which ignore censoring. These alternative models are linear models which always include satellite fixed effects, and otherwise, apart from the infrastructure variables of interest, they are specified as follows: the model in Column (1) does not include any other variables besides place fixed effects; the model in Column (2) is as the one in Column (1) but includes control variables; the model in Column (3) is as the one in Column (2) but includes lags of the dependent variable and is an immediate linear counterpart to the dynamic Tobit model in Table 2.4. In any case, it turns out that the results across these models and the dynamic Tobit in Table 2.4 are very robust. As with the dynamic Tobit model, the fixed effects are parameterized in terms of averages of the time-variant explanatory variables.

Table 2.4: ESTIMATION RESULTS DYNAMIC TOBIT

	(1)	(2)	(3)	(4)	(5)	(6)
	$radiance_{it}$	$radiance_{it}$	$radiance_{it}$	$radiance_{it}$	$radiance_{it}$	$radiance_{it}$
	Full Sample	Full Sample	$Nat = 1$	$Nat = 0$	$Admin = 1$	$Admin = 0$
$radiance_{it-1}$	0.603*** (0.002)	0.602*** (0.002)	0.237*** (0.002)	0.524*** (0.002)	0.607*** (0.003)	0.598*** (0.002)
$radiance_{it-2}$	0.194*** (0.002)	0.194*** (0.002)	0.0950*** (0.002)	0.149*** (0.002)	0.206*** (0.004)	0.189*** (0.002)
$radiance_{it-3}$	0.100*** (0.002)	0.100*** (0.002)	0.114*** (0.002)	0.028*** (0.002)	0.088*** (0.003)	0.106*** (0.002)
$radiance_{i1992}$	0.039*** (0.001)	0.039*** (0.001)	0.088*** (0.002)	0.190*** (0.003)	0.035*** (0.002)	0.044*** (0.001)
$\ln(popdens_{i1990})$	0.149*** (0.011)	0.158*** (0.011)	0.098*** (0.019)	0.115*** (0.023)	0.110*** (0.024)	0.166*** (0.013)
$\ln(dist\ to\ road_i)$	-0.219*** (0.007)	-0.217*** (0.007)	-0.360*** (0.013)	-0.317*** (0.014)	-0.219*** (0.013)	-0.221*** (0.009)
$\ln(dist\ to\ rail_i)$	-0.041*** (0.008)	-0.038*** (0.008)	-0.066*** (0.014)	-0.091*** (0.015)	-0.003 (0.015)	-0.056*** (0.009)
$\ln(dist\ to\ ocean_i)$	-0.116*** (0.010)	-0.108*** (0.010)	-0.272*** (0.018)	-0.276*** (0.018)	-0.165*** (0.021)	-0.092*** (0.011)
$\ln(dist\ to\ water_i)$	-0.079*** (0.008)	-0.077*** (0.008)	-0.085*** (0.015)	-0.140*** (0.015)	-0.089*** (0.015)	-0.080*** (0.010)
$\ln(dist\ to\ center_i)$	-0.199*** (0.016)	-0.236*** (0.016)	-0.055* (0.028)	-0.095*** (0.031)	-0.105*** (0.027)	-0.260*** (0.021)
$\ln(dist\ to\ adborder_i)$	-0.025*** (0.009)	-0.059 (0.044)	-0.015 (0.017)	0.056*** (0.017)	-0.022 (0.017)	-0.035*** (0.012)
$within\ admin\ boundary_{it}$	-0.047* (0.024)	-0.704*** (0.118)	-0.027 (0.043)	-0.065 (0.045)		
$\ln(altitude_i)$	-0.010*** (0.016)	-0.085*** (0.016)	-0.306*** (0.032)	0.072** (0.029)	-0.051* (0.028)	-0.133*** (0.020)
$longitude_i$	0.006* (0.003)	-0.001 (0.003)	0.014** (0.006)	-0.001 (0.006)	-0.018*** (0.006)	0.022*** (0.004)
$latitude_i$	0.054*** (0.011)	0.093*** (0.012)	0.087*** (0.020)	0.078*** (0.020)	0.119*** (0.022)	0.026** (0.013)
$\ln(rain_i)$	-0.288*** (0.055)	-0.103* (0.057)	-1.134*** (0.107)	-0.719*** (0.099)	0.089 (0.102)	-0.494*** (0.067)
$temperature_i$	0.076*** (0.012)	0.079*** (0.012)	0.035 (0.023)	0.197*** (0.021)	0.077*** (0.020)	0.052*** (0.016)
$sd\ temperature_i$	-0.003 (0.018)	-0.081*** (0.019)	-0.089*** (0.033)	-0.016 (0.032)	-0.082** (0.035)	-0.033 (0.022)
$hukou_{it}$	-0.750*** (0.069)	-1.316*** (0.081)	-0.036 (0.070)	-2.627*** (0.137)	0.596*** (0.207)	-0.963*** (0.082)
$small_{it}$	1.128*** (0.066)	1.506*** (0.078)	0.592*** (0.067)	2.746*** (0.132)	-0.408** (0.190)	1.556*** (0.080)
$medium_{it}$	0.497*** (0.054)	1.093*** (0.074)	0.405*** (0.062)	1.226*** (0.095)	-0.441*** (0.165)	0.550*** (0.064)
$large_{it}$	-0.696*** (0.053)	-0.927*** (0.065)	-0.809*** (0.063)	-0.809*** (0.088)	0.018 (0.153)	-0.785*** (0.064)

Continued.

Table 2.4: *Continued.*

	(1)	(2)	(3)	(4)	(5)	(6)
	$radiance_{it}$	$radiance_{it}$	$radiance_{it}$	$radiance_{it}$	$radiance_{it}$	$radiance_{it}$
	Full Sample	Full Sample	Nat = 1	Nat = 0	Admin = 1	Admin = 0
$SEZ_{it}$	0.167*** (0.040)	0.977*** (0.185)	0.159** (0.074)	0.561*** (0.074)	0.658*** (0.103)	0.048 (0.047)
$firstwave_{it}$	-0.258*** (0.071)	-0.315*** (0.075)	0.011 (0.125)	-0.475*** (0.142)	-1.761*** (0.179)	0.079 (0.081)
$secondwave_{it}$	-0.237*** (0.035)	-0.306*** (0.041)	-0.536*** (0.064)	-0.421*** (0.064)	-0.630*** (0.101)	-0.167*** (0.039)
$thirdwave_{it}$	0.319*** (0.031)	-1.649*** (0.277)	0.467*** (0.057)	0.069 (0.056)	0.175** (0.074)	0.421*** (0.036)
Constant	2.409*** (0.914)	0.990 (0.921)	18.02*** (1.644)	13.86*** (2.013)	4.793*** (1.558)	0.493 (1.660)
Time averages		YES				
Satellite effects	YES	YES	YES	YES	YES	YES
# observations	373,673	373,673	158,116	215,557	107,780	265,893
# places	19,667	19,667	19,656	16,756	6,144	14,187

*Notes:* Reported coefficients are marginal effects. Standard errors are reported in parentheses ; \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. All columns include squared terms for the following geography and climate variables:  $\ln(dist\ to\ road_i)$ ,  $\ln(dist\ to\ rail_i)$ ,  $\ln(dist\ to\ ocean_i)$ ,  $\ln(dist\ to\ water_i)$ ,  $\ln(dist\ to\ adborder_{it})$ ,  $\ln(dist\ to\ center_i)$ ,  $\ln(altitude_i)$ ,  $\ln(rain_i)$ ,  $temperature_i$ ,  $sd\ temperature_i$ . Column (2) includes time averages for all time variant variables. All distance measures in the empirical estimation are in units of meters rather than kilometers. Abbreviations: Nat= natural city, Admin = within administrative boundary.

night-light radiance of a place; and (ii) the magnitude of the marginal effect of  $\ln(dist\ to\ road_i)$  is around five times larger than the one of  $\ln(dist\ to\ rail_i)$ . Clearly, these effects on  $radiance_{it}$  reflect the importance of transport infrastructure, particularly roads, for local economic growth across all places in the sample.

In Columns (3)-(6), we estimate the same model as in Column (1) for various subsamples of the data. Columns (3) and (4) divide the sample between places inside and outside of natural cities, while Columns (5) and (6) separate places inside and outside of the administrative borders of the major cities in our sample. Interestingly, the effect of  $\ln(dist\ to\ road_i)$  in Column (3) is larger than in Column (4), while the opposite is observed for  $\ln(dist\ to\ rail_i)$ . Similarly, Column (6) shows a significant negative impact of  $\ln(dist\ to\ rail_i)$ , while the corresponding estimate in Column (5) is much smaller and not significant. The differences in the effects between Columns (3) and (4) on one hand and Columns (5) and (6) on the other – both in absolute terms and in comparison to Column (1) – reflect differences in the opportunity costs of certain types of transport infrastructure depending

on the relative centrality or peripherality of places relative to the natural city or the administrative city center. In general, these results indicate that a marginal decline in distance to the road network leads places inside the natural city to grow relatively faster than places outside of it. However, a marginal decline in the distance to railway lines benefits peripheral areas more than central ones.<sup>12</sup>

Using the estimated effects from Column (1) in Table 2.4, we can predict the radiance level of all places from period to period and the change associated with an infrastructure improvement to the road or railway networks. We do so by reducing the distance to roads and railway lines by one standard deviation. We use 2007 as the benchmark year for this thought experiment since it is the year in which there are almost as many places outside (49%) as inside natural cities (51%). We predict the radiance level of all places in 2007 given the estimated coefficients associated with Column (1) of Table 2.4 and the variables in  $\mathbf{R}_{it}$  and  $\mathbf{X}_{it}$  as observed. We plot the kernel density estimates of observed and predicted radiance levels in 2007 in Figure 2.6. Then, we shock  $\ln(\text{dist to road}_i)$  and  $\ln(\text{dist to rail}_i)$  alternatively by one standard deviation in 2007 and let the process run to see how such shocks impact radiance levels in the short- and long-term. Following the definition of a natural city used in this paper, we assume that any place will be part of a natural city in the counterfactual scenario if (i) its predicted radiance level amounts to at least 40, and (ii) it is connected to other places in the natural city with a radiance level of at least 40.

---

<sup>12</sup>Table 2.10 presents the results of the estimation of the baseline model (Column (1) in Table 2.4) on further subsamples, namely places in all four domains of the binary classification divided into administrative and natural city boundaries.

Figure 2.6: KERNEL DENSITY ESTIMATES – OBSERVED VS. PREDICTED RADIANCE LEVELS, 2007

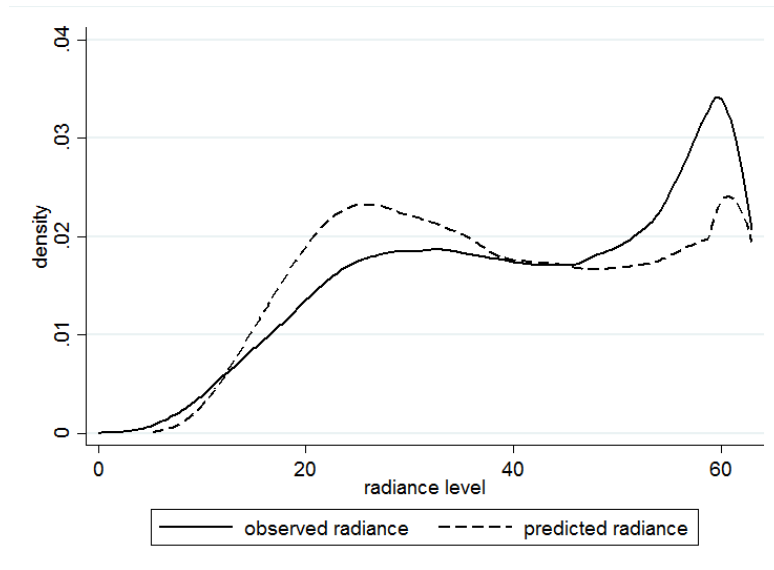
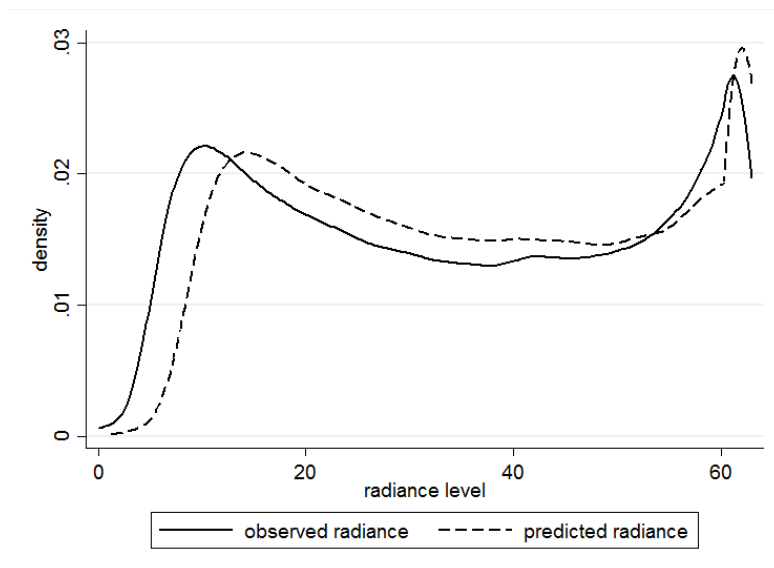


Figure 2.7: KERNEL DENSITY ESTIMATES: OBSERVED VS. PREDICTED RADIANCE LEVELS, ALL YEARS



In Tables 2.5 and 2.6, we report effects of these shocks on radiance levels in 2007 – as well as after 5, 10, 15, and 20 years – compared to the baseline predictions. Table 2.5 shows the effect of a shock on road infrastructure. Most places predicted to lie inside a natural city in the baseline case remain inside it in the counterfactual scenario after 5 years (99.7% in 2012) and after 20 years (100.0% in 2027). However, the share of places in the sample that are predicted to lie outside of the natural city in the baseline but inside of it in the counterfactual scenario steadily increases over time in response to the shock from 0.7% in 2007

Table 2.5: TRANSITION MATRIX OF COUNTERFACTUAL  $\ln(\text{dist to road}_i)$

Definition:  $Nat = 1$  if radiance  $\geq 40$ ,  $Nat = 0$  if radiance  $< 40$

		Counterfactual: $\ln(\text{dist to road}_i)$		
		$Nat = 1$	$Nat = 0$	<b>Total</b>
Baseline	$Nat = 1$	99.95 <sub>(2007)</sub>	0.05 <sub>(2007)</sub>	100
		99.72 <sub>(2012)</sub>	0.28 <sub>(2012)</sub>	100
		99.80 <sub>(2017)</sub>	0.20 <sub>(2017)</sub>	100
		99.95 <sub>(2022)</sub>	0.05 <sub>(2022)</sub>	100
		99.99 <sub>(2027)</sub>	0.01 <sub>(2027)</sub>	100
	$Nat = 0$	0.66 <sub>(2007)</sub>	99.34 <sub>(2007)</sub>	100
		6.18 <sub>(2012)</sub>	93.82 <sub>(2012)</sub>	100
		30.15 <sub>(2017)</sub>	69.85 <sub>(2017)</sub>	100
		51.85 <sub>(2022)</sub>	48.15 <sub>(2022)</sub>	100
		83.39 <sub>(2027)</sub>	16.61 <sub>(2027)</sub>	100
<b>Total</b>		43.50 <sub>(2007)</sub>	56.50 <sub>(2007)</sub>	100
		61.43 <sub>(2012)</sub>	38.57 <sub>(2012)</sub>	100
		85.57 <sub>(2017)</sub>	14.43 <sub>(2017)</sub>	100
		96.91 <sub>(2022)</sub>	3.09 <sub>(2022)</sub>	100
		99.73 <sub>(2027)</sub>	0.27 <sub>(2027)</sub>	100

*Notes:* Additional to the radiance threshold, the CCA condition is a necessary condition for a place to be assigned *natural city* ( $Nat = 1$ ). The CCA condition implies that a place has to be in the neighborhood of a cluster of places with average radiance greater or equal the threshold. Abbreviation: Nat = natural city.

Table 2.6: TRANSITION MATRIX OF COUNTERFACTUAL  $\ln(\text{dist to rail}_i)$

Definition:  $Nat = 1$  if radiance  $\geq 40$ ,  $Nat = 0$  if radiance  $< 40$

		Counterfactual: $\ln(\text{dist to rail}_i)$		
		$Nat = 1$	$Nat = 0$	<b>Total</b>
Baseline	$Nat = 1$	100.00 <sub>(2007)</sub>	0.00 <sub>(2007)</sub>	100
		99.92 <sub>(2012)</sub>	0.08 <sub>(2012)</sub>	100
		99.91 <sub>(2017)</sub>	0.09 <sub>(2017)</sub>	100
		99.96 <sub>(2022)</sub>	0.04 <sub>(2022)</sub>	100
		99.98 <sub>(2027)</sub>	0.02 <sub>(2027)</sub>	100
	$Nat = 0$	0.13 <sub>(2007)</sub>	99.87 <sub>(2007)</sub>	100
		1.02 <sub>(2012)</sub>	98.98 <sub>(2012)</sub>	100
		4.18 <sub>(2017)</sub>	95.82 <sub>(2017)</sub>	100
		10.47 <sub>(2022)</sub>	89.53 <sub>(2022)</sub>	100
		15.34 <sub>(2027)</sub>	84.66 <sub>(2027)</sub>	100
<b>Total</b>		43.22 <sub>(2007)</sub>	56.78 <sub>(2007)</sub>	100
		59.43 <sub>(2012)</sub>	40.57 <sub>(2012)</sub>	100
		80.36 <sub>(2017)</sub>	19.64 <sub>(2017)</sub>	100
		94.31 <sub>(2022)</sub>	5.69 <sub>(2022)</sub>	100
		98.64 <sub>(2027)</sub>	1.36 <sub>(2027)</sub>	100

*Notes:* Additional to the radiance threshold, the CCA condition is a necessary condition for a place to be assigned *natural city* ( $Nat = 1$ ). The CCA condition implies that a place has to be in the neighborhood of a cluster of places with average radiance greater or equal the threshold. Abbreviation: Nat = natural city.

to 30.2% in 2017 and to 83.4% in 2027. The magnitude of the effect is considered to be relatively high because the actual number of places not in a natural city after 2017 is relatively small by construction of the data set.<sup>13</sup> Figures 2.8 and 2.9 illustrate the examples of Beijing and Shanghai, respectively.

The picture is similar, albeit of a smaller magnitude, when looking at the effect of a shock on rail infrastructure as shown in Table 2.6. The extreme majority of places predicted inside a natural city in the baseline are also predicted to lie inside in the counterfactual analysis after 5 years (99.9% in 2012) and 100.0% in 2027). The share of places predicted to lie outside in the baseline but inside in the counterfactual (Nat = 0 in baseline, Nat = 1 in counterfactual) is also increasing over time from 0.1% in 2007 to 4.2% in 2017 and to 15.2% in 2027. The smaller magnitude of the effect reflects the smaller magnitude of the coefficient of  $\ln(\text{dist to rail}_i)$  compared to the coefficient of  $\ln(\text{dist to road}_i)$  estimated in Column (1) of Table 2.4.<sup>14</sup>

Our finding that transport infrastructure has a positive effect on local economic activity is well-aligned with the findings in Banerjee et al. (2012). They indicate that transport networks lead to higher levels of GDP per capita, even though the effect reported is small in magnitude. In line with Baum-Snow (2007) and Baum-Snow et al. (2016), the results in our paper also suggest that better transport connectivity increases local economic activity in suburban areas. Considering the population density in city centers versus suburban areas, Baum-Snow et al. (2016, p. 2) suggest that “each additional radial highway displaced about 4% of [the] central city population to suburban regions and that the existence of some ring road capacity in a city reduced city population by about 20%”. Contrary to these findings, we observe a positive effect of transport infrastructure on natural city growth, with a positive effect on both central and more peripheral areas of an average natural city. These results contrast with Faber (2014), who, looking

---

<sup>13</sup>The data set includes only those places that were in a natural city at some point in time between 1992 and 2013. This implies that all places that are not yet in a natural city in 2007 have a high probability of becoming part of a natural city within a few years

<sup>14</sup>In Tables 2.11 and 2.12, we report the dynamic responses to an infrastructure shock when fixing the night-light-radiance threshold for a place to be inside a natural city to 50 instead of 40. As expected, the main message of the results holds, even though the share of places predicted to lie outside the natural city in the baseline and inside in both counterfactuals is lower than in Tables 2.5 and 2.6. This simply reflects the distribution of night-light radiance across places as shown in Figure 2.7.



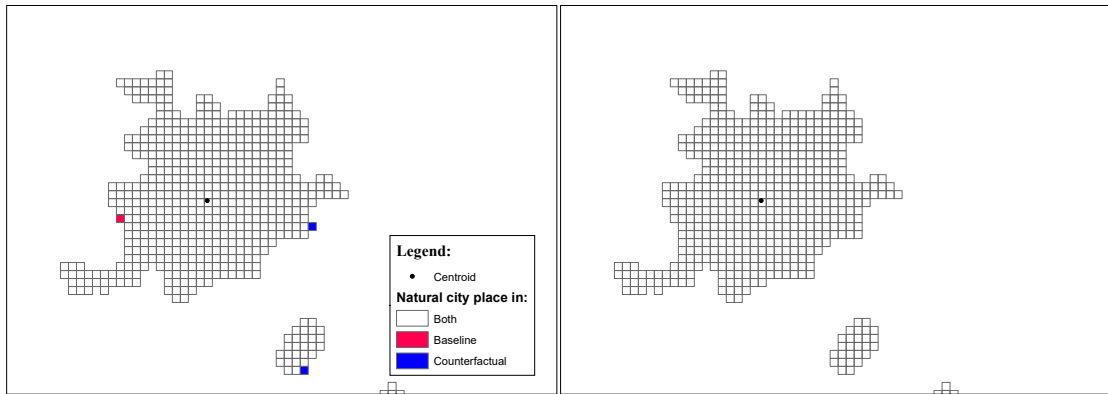
Figure 2.8: COUNTERFACTUAL ROAD AND RAIL – BEIJING OVER TIME

**Counterfactual Road**

**Counterfactual Rail**

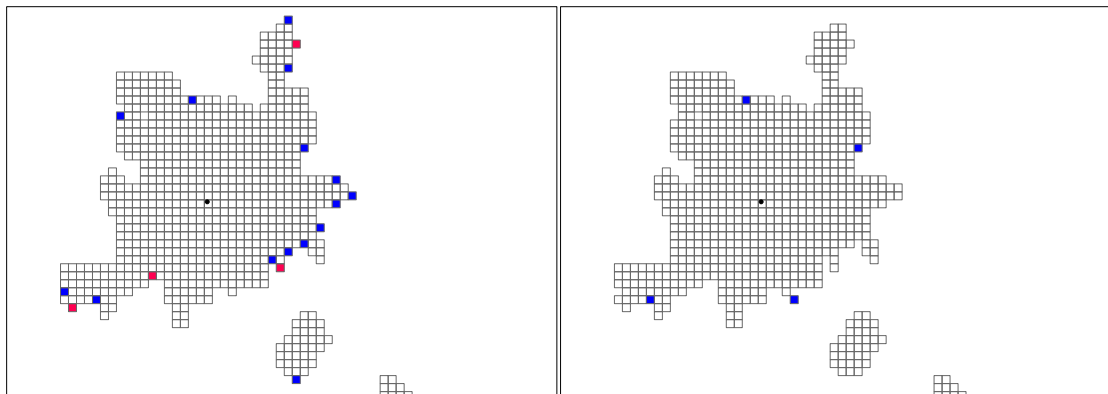
(a) Beijing 2007

(b) Beijing 2007



(c) Beijing 2012

(d) Beijing 2012



(e) Beijing 2017

(f) Beijing 2017

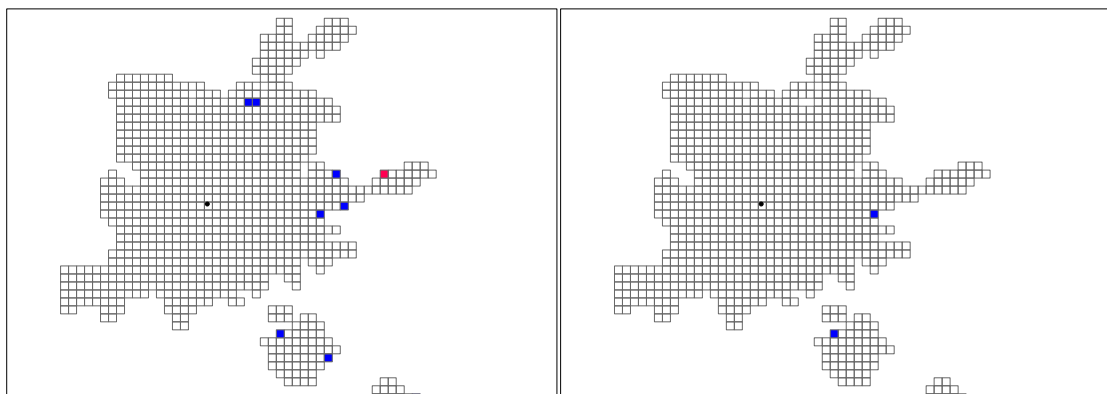


Figure 2.9: COUNTERFACTUAL ROAD AND RAIL – SHANGHAI OVER TIME

**Counterfactual Road**

**Counterfactual Rail**

(a) Shanghai 2007

(b) Shanghai 2007



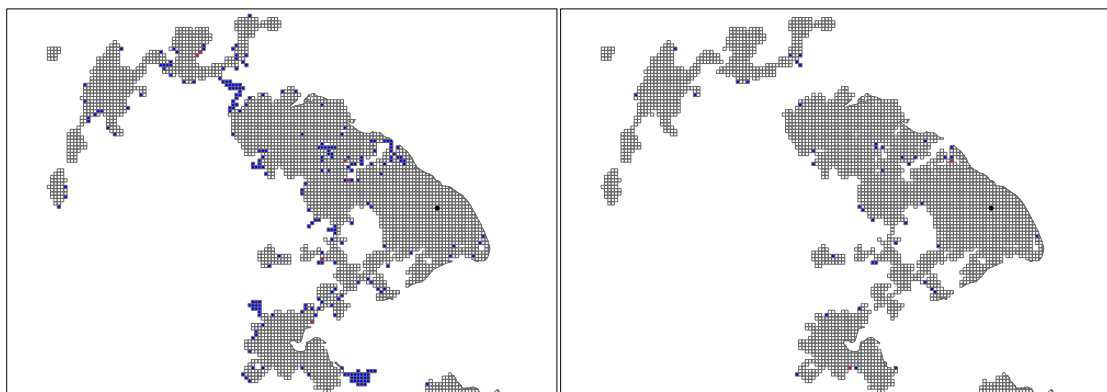
(c) Shanghai 2012

(d) Shanghai 2012



(e) Shanghai 2017

(f) Shanghai 2017



at peripheral counties outside the commuting zones of metropolitan areas, finds that highway network connections have led to lower GDP growth among peripheral counties. This difference in findings suggests that transport networks have different effects on economic activity in remote areas than in metropolitan areas.

## 2.4 Conclusion

This paper documents patterns in the size and growth of natural cities in China for the 300 largest urban entities between 1992 and 2013. Rather than using administrative data on economic outcomes and their determinants, the paper identifies the boundaries of a natural city, which is related more closely to the notion of MSAs or Functional Urban Zones, in terms of the night-light radiance of connected places that measure 3 km x 3 km. Ultimately, the boundaries of natural cities are determined by applying the CCA to remote-sensing data for those places during the review period.

The key results of our analysis include the following. First, the number of distinct natural city centers decreased during the review period due to the absorption of some natural cities by others. This was particularly the case for larger cities, such as Shanghai, that formed natural super-cities during the review period. Second, we detected rapid growth for the average natural city, which is in accordance with population census data that are only available at less frequent time intervals than night-light data, and adheres to China's goal of fostering the rate of urbanization. The results suggest that natural cities grew considerably beyond the administrative boundaries of cities, which calls into question policies that target urbanization rates and other related development objectives based on administrative city boundaries. Third, the global financial crisis at the end of the last decade left its marks on natural city growth as many Chinese natural cities in our sample shrank between 2010 and 2013. Fourth, infrastructure improvements to the road and railway networks benefit agglomerations, although railway network improvements are expected to mainly benefit peripheral areas of cities more so than road improvements.

In future work, we plan to focus more explicitly on the difference between time-variant administrative versus natural city boundaries, and shed further light on the dynamic process of responses to the two exogenous shocks. While we mainly used institutional variables related to the *hukou* system and SEZs as control variables, we will scrutinize their effects more closely after having coded them at a greater level of detail in order to understand these effects with greater precision and a broader scope than was possible in the current paper. Finally, we will investigate the effects of changes in China's infrastructure networks, which was not possible with the data at hand, to better identify the associated effects on economic and other outcomes.

# Appendix

## 2.5 Supplement Tables

Table 2.7: LIST OF 300 BIGGEST CHINESE ADMINISTRATIVE CITIES (BY POPULATION IN 2000)

Rank	City Name	Population	Rank	City Name	Population	Rank	City Name	Population	Rank	City Name	Population
1	Shanghai	13 278 500	76	Shatian	648 400	151	Maoming	346 500	226	Zouxian	196 500
2	Beijing	7 209 900	77	Mudanjiang	646 700	152	Jixi	345 200	227	Chenzhou	193 300
3	Wuhan	4 104 300	78	Zhenjiang	643 000	153	Zhoukou	345 000	228	Badaojiang	192 900
4	Chengdu	4 064 700	79	Yancheng	638 800	154	Jinxi	337 300	229	Wanxiang	191 700
5	Tianjin	3 945 900	80	Shaoyang	636 700	155	Shiongshui	336 300	230	Shangqiu	191 300
6	Shenyang	3 527 800	81	Jinzhou	630 800	156	Zhongshan	334 100	231	Haibowan	190 000
7	Xian	3 480 600	82	Taizhou	622 400	157	Zhaoqing	331 400	232	Ezhou	189 300
8	Chongqing	3 378 900	83	Taian	618 900	158	Dongchang	328 500	233	Quanzhou	188 200
9	Guangzhou	3 244 900	84	Dandong	604 600	159	Tieling	324 100	234	Zaoyang	187 900
10	Harbin	3 129 300	85	Luancheng	598 500	160	Daipo	322 300	235	Kaiyuan	186 200
11	Nanjing	2 870 200	86	Panjin	582 200	161	Xinpu	322 200	236	Dongling	184 600
12	Taiyuan	2 690 500	87	Guilin	580 200	162	Dezhou	317 900	237	Zaozhuang	184 000
13	Changchun	2 337 000	88	Kaifeng	577 400	163	Luzhou	317 200	238	Jiutai	183 800
14	Zhengzhou	2 052 700	89	Zhangjiakou	573 800	164	Jingdezhen	314 400	239	Baiyin	183 500
15	Jiulong	2 040 200	90	Yingkou	571 400	165	Chifeng	314 200	240	Yizheng	183 100
16	Jinan	1 961 500	91	Haikou	568 900	166	Shishi	308 800	241	Jingzhou	181 200
17	Dalian	1 925 200	92	Zhanjiang	568 000	167	Tongling	307 600	242	Qincheng	176 700
18	Changsha	1 891 000	93	Huayin	564 400	168	Dongying	303 800	243	Nanping	176 300
19	Hangzhou	1 881 500	94	Neijiang	562 700	169	Yanji	302 900	244	Hangu	175 600
20	Shijiazhuang	1 683 800	95	Puyang	561 300	170	Suizhou	302 700	245	Laiyang	174 900
21	Nanchang	1 657 900	96	Bengbu	561 200	171	Xuanhua	302 600	246	Luan	174 700
22	Jilin	1 625 700	97	Shihezi	553 400	172	Baicheng	292 300	247	Xianning	173 900
23	Tangshan	1 600 600	98	Yangzhou	548 600	173	Hulan ergi	283 200	248	Jingmen	170 900
24	Qingdao	1 584 300	99	Jiamusi	548 400	174	Anshun	283 100	249	Weinan	170 800
25	Urumqi	1 424 300	100	Yueyang	542 800	175	Quanwan	277 700	250	Jiaozhou	170 200
26	Luoyang	1 417 200	101	Maanshan	536 700	176	Saumenxia	274 100	251	Liupanshui	169 600
27	Xinyang	1 412 300	102	Xiamen	535 600	177	Linyi	268 500	252	Jian	169 400
28	Lanzhou	1 409 200	103	Shaoguan	529 300	178	Jiujiang	268 200	253	Bozhou	169 300
29	Fushun	1 409 000	104	Zhangzhou	519 700	179	Huizhou	265 700	254	Fuling	168 900
30	Hefei	1 362 900	105	Wuhu	512 500	180	Nanyang	265 500	255	Honghu	168 500
31	Xianggangdao	1 345 800	106	Xingtai	512 500	181	Wuzhou	264 200	256	Huanggang	168 300
32	Baotou	1 226 600	107	Zhannmen	499 600	182	Yingcheng	263 600	257	Zhaodong	167 800
33	Anshan	1 226 200	108	Jiaozuo	497 800	183	Aksu	263 500	258	Yuhong	167 600
34	Shantou	1 221 300	109	Foshan	496 300	184	Mianyang	261 800	259	Beipiao	166 900
35	Guiyang	1 160 700	110	Yuanlong	494 800	185	Jincheng	259 600	260	Hengshan	166 700
36	Suzhou	1 150 200	111	Tanggu	490 000	186	Wafangdian	258 400	261	Wulanhaote	166 500
37	Handan	1 129 800	112	Zhangye	489 600	187	Shangrao	258 300	262	Linhe	166 400

*Continued.*

Table 2.7: *Continued.*

Rank	City Name	Population	Rank	City Name	Population	Rank	City Name	Population	Rank	City Name	Population
38	Fuzhou	1 120 200	113	Siping	486 800	188	Tongliao	257 500	263	Huzhou	165 500
39	Xuzhou	1 099 200	114	Kuiqing	480 100	189	Suihua	256 700	264	Fuyang	165 300
40	Datong	1 053 600	115	Yichang	478 200	190	Heze	256 600	265	Mentougou	165 200
41	Wuxi	1 010 500	116	Panzhuhua	473 900	191	Sucheng	254 200	266	Longfeng	164 200
42	Xianyang	976 200	117	Xiangfan	471 500	192	Jining	248 400	267	Deyang	163 600
43	Kunming	967 900	118	Jiaojiang	470 500	193	Laohekou	248 300	268	Xiaogan	163 400
44	Benxi	918 900	119	Cangzhou	470 400	194	Rizhao	246 800	269	Yulin	162 600
45	Changzhou	917 400	120	Liaoyuan	470 400	195	Yibin	246 400	270	Zhicheng	162 300
46	Pingdingshan	906 700	121	Jiaxing	466 900	196	Kashi	240 400	271	Nanpiao	161 300
47	Baoding	888 100	122	Yinchuan	465 900	197	Yining	240 000	272	Sujiatun	159 500
48	Nanning	880 400	123	Zhuhai	460 600	198	Hebi	239 300	273	Shanwei	158 600
49	Wenzhou	867 200	124	Changde	457 600	199	Mianchang	238 900	274	Korla	158 400
50	Qiqihar	860 000	125	Jiangmen	457 500	200	Beihai	238 200	275	Beian	158 300
51	Huainan	859 300	126	Aomen	453 300	201	Ganzhou	236 500	276	Yichun	157 800
52	Huaipei	808 100	127	Shashi	446 100	202	Xiantao	234 800	277	Acheng	157 300
53	Xining	782 700	128	Chengde	442 300	203	Jinzhou	231 900	278	Daliang	156 700
54	Hengyang	780 700	129	Hengshui	432 500	204	Ranghulu	229 800	279	Boshan	156 600
55	Hohhot	762 700	130	Luqiao	428 600	205	Chuzhou	228 500	280	Dunhua	156 600
56	Anyang	759 900	131	Baoji	415 200	206	Linfen	226 000	281	Qianjiang	156 000
57	Xinxiang	757 800	132	Yangquan	402 100	207	Liaocheng	225 700	282	Leshan	155 800
58	Shenzhen	752 200	133	Zunyi	401 200	208	Saertu	225 700	283	Gaomi	155 500
59	Liuzhou	748 200	134	Jining	399 900	209	Tongchuan	224 500	284	Linhai	154 600
60	Zhaotong	742 400	135	Shiyan	398 000	210	Xintai	223 600	285	Guangshui	154 300
61	Zhuzhou	729 000	136	Dongguan	380 700	211	Guangyuan	223 100	286	Kaoli	153 200
62	Hegang	724 700	137	Pingxiang	379 900	212	Yuci	221 700	287	Xinzhou	152 900
63	Langfang	721 800	138	Weifang	376 700	213	Tianmen	220 100	288	Yichun	152 700
64	Ningbo	721 100	139	Putian	371 700	214	Nanchong	219 700	289	Dingzhou	152 000
65	Zigong	709 900	140	Xuchang	371 200	215	Wuxue	219 100	290	Karamay	150 200
66	Qinhuangdao	707 200	141	Saigong	369 900	216	Yiyang	218 800	291	Xinji	149 400
67	Xiangtan	692 900	142	Xigong	365 600	217	Shuangyashan	209 300	292	Gongzhuling	147 900
68	Fuxin	687 800	143	Changji	365 400	218	Zhumadian	209 000	293	Huangyan	147 000
69	Changzhi	683 600	144	Yangjiang	365 000	219	Haicheng	206 400	294	Huadian	145 900
70	Zhangdian	679 200	145	Qitaihe	356 500	220	Sanning	205 500	295	Jinhua	145 800
71	Nantong	677 200	146	Chaozhou	354 200	221	Linxia	203 200	296	Hanzhong	145 700
72	Huangshi	671 400	147	Luobe	350 900	222	Suzhou	199 500	297	Yushan	145 700
73	Rongcheng	669 500	148	Shaoying	350 700	223	Yuncheng	197 500	298	Fuyu	145 500
74	Yantai	652 200	149	Chaoyang	350 100	224	Hailar	197 300	299	Huicheng	145 300
75	Liaoyang	650 100	150	Anqing	346 600	225	Anda	197 000	300	Chizhou	144 300

Source: <http://www.tageo.com/index-e-ch-cities-CN.htm>

Table 2.8: LIST OF CHINESE NATURAL CITIES (BY SIZE IN 2000)

Rank	City Name	Merged	Area	Rank	City Name	Merged	Area	Rank	City Name	Merged	Area
1	Guangzhou	8	8,473.02	73	Xingtai		117.00	145	Yichang		54.00
2	Beijing	1	3,042.00	74	Guiyang		108.00	146	Yining		54.00
3	Shanghai		3,027.42	75	Jinxi		108.00	147	Yizheng		54.00
4	Tianjin		1,125.00	76	Mianyang		108.00	148	Yuci		54.00
5	Ranghulu	2	1,044.00	77	Siping		108.00	149	Zaozhuang		54.00
6	Shenyang	3	990.00	78	Xiangfan		108.00	150	Acheng		45.00
7	Harbin		756.00	79	Xianyang		108.00	151	Anqing		45.00
8	Nanjing		738.00	80	Xinxiang		108.00	152	Badaojiang		45.00
9	Wuhan		711.00	81	Haikou		100.68	153	Haibowan		45.00
10	Changchun		684.00	82	Cangzhou		99.00	154	Huainan		45.00
11	Qingdao		640.41	83	Huaiyin		99.00	155	Laiyang		45.00
12	Dalian	1	638.59	84	Korla		99.00	156	Qitaihe		45.00
13	Hangzhou		621.42	85	Nanchang		99.00	157	Shaoguan		45.00
14	Urumqi		594.00	86	Pingdingshan		99.00	158	Xiangtan		45.00
15	Xian		558.00	87	Shihezi		99.00	159	Xinyang		45.00
16	Jinan		477.00	88	Xinpu		99.00	160	Yangquan		45.00
17	Taiyuan		459.00	89	Yingkou		99.00	161	Yueyang		45.00
18	Chengdu		450.00	90	Zhaoqing		99.00	162	Yulin		45.00
19	Shijiazhuang		423.00	91	Zhanjiang		92.56	163	Yuncheng		45.00
20	Shishi	1	404.31	92	Benxi		90.00	164	Zhuzhou		45.00
21	Baotou		396.00	93	Liaocheng		90.00	165	Aksu		36.00
22	Zhengzhou		387.00	94	Maanshan		90.00	166	Baiyin		36.00
23	Kunming		369.00	95	Zhangzhou		90.00	167	Boshan		36.00
24	Suzhou (JS)		360.00	96	Rizhao		86.51	168	Chuzhou		36.00
25	Chongqing		351.00	97	Bengbu		81.00	169	Fuyang		36.00
26	Shantou		346.83	98	Dandong		81.00	170	Gaomi		36.00
27	Jilin		342.00	99	Fuxin		81.00	171	Gongzhuling		36.00
28	Zhuhai	1	337.01	100	Hailar		81.00	172	Putian		36.00
29	Hohhot		324.00	101	Huzhou		81.00	173	Suihua		36.00
30	Anshan		315.00	102	Jiaozuo		81.00	174	Xinzhou		36.00
31	Yantai		307.56	103	Jining (SD)		81.00	175	Xuchang		36.00
32	Dongying		306.00	104	Panzhuhua		81.00	176	Zhaodong		36.00
33	Fuzhou		306.00	105	Yancheng		81.00	177	Zhoushou		36.00
34	Wuxi		306.00	106	Yushan		81.00	178	Zhumadian		36.00
35	Hefei		297.00	107	Zhenjiang		81.00	179	Anda		27.00
36	Changsha		279.00	108	Changji		72.00	180	Dunhua		27.00
37	Ningbo		270.00	109	Chaoshou		72.00	181	Huangyan		27.00
38	Fushun		243.00	110	Guilin		72.00	182	Jingmen		27.00
39	Wenzhou		238.53	111	HuaiBei		72.00	183	Jining (NM)		27.00

*Continued.*



Table 2.8: *Continued.*

Rank	City Name	Merged	Area	Rank	City Name	Merged	Area	Rank	City Name	Merged	Area
40	Tanggu		233.44	112	Jixi		72.00	184	Kaiyuan		27.00
41	Changzhou		225.00	113	Kashi		72.00	185	Linhai		27.00
42	Datong		225.00	114	Linfen		72.00	186	Suzhou (AH)		27.00
43	Xuzhou		225.00	115	Nanyang		72.00	187	Tongling		27.00
44	Baoding		216.00	116	Shangqiu		72.00	188	Wafangdian		27.00
45	Handan		207.00	117	Shaoxing		72.00	189	Weinan		27.00
46	Nanning		207.00	118	Taizhou		72.00	190	Wuzhou		27.00
47	Zhangdian		180.00	119	Yanji		72.00	191	Yichun (HL)		27.00
48	Huizhou		171.00	120	Baicheng		63.00	192	Bozhou		18.00
49	Liaoyang		171.00	121	Baoji		63.00	193	Deyang		18.00
50	Linyi		171.00	122	Changde		63.00	194	Fuling		18.00
51	Luoyang		171.00	123	Chifeng		63.00	195	Hangu		18.00
52	Panjin		171.00	124	Dongchang		63.00	196	Huadian		18.00
53	Weifang		171.00	125	Hengshui		63.00	197	Jiujiang		18.00
54	Yinchuan		171.00	126	Huangshi		63.00	198	Maoming		18.00
55	Jiaojiang	1	169.39	127	Hulan		63.00	199	Mianchang		18.00
56	Langfang		162.00	128	Jinhua		63.00	200	Sucheng		18.00
57	Qinhuangdao		159.86	129	Karamay		63.00	201	Xiaogan		18.00
58	Fuyu		153.00	130	Luohe		63.00	202	Xinji		18.00
59	Anyang		144.00	131	Tieling		63.00	203	Zunyi		18.00
60	Puyang		144.00	132	Tongliao		63.00	204	Shanwei		16.93
61	Qiqihar		144.00	133	Zouxian		63.00	205	Chengzhou		9.00
62	Xiamen		140.68	134	Beihai		62.86	206	Dingzhou		9.00
63	Xining		135.00	135	Yangjiang		61.26	207	Huanggang		9.00
64	Yangzhou		135.00	136	Chengde	1	54.00	208	Jiutai		9.00
65	Dezhou		126.00	137	Haicheng		54.00	209	Luancheng		9.00
66	Jiamusi		126.00	138	Heze		54.00	210	Suizhou		9.00
67	Jinzhou (LN)		126.00	139	Jiaozhou		54.00	211	Wanxian		9.00
68	Mudanjiang		126.00	140	Jincheng		54.00	212	Xintai		9.00
69	Jiaying		117.00	141	Jingzhou	1	54.00	213	Zhangye		9.00
70	Rongcheng		117.00	142	Liaoyuan		54.00				
71	Taian		117.00	143	Wulanhaote		54.00				
72	Wuhu		117.00	144	Xuanhua		54.00				

*Notes:* Columns *Merged* report the total number of cities, if any, that got merged to the city named in the respective Column *City Name*. We specify the merged cities in rank order hereafter. 1. Guangzhou: Daliang, Dongguan, Foshan, Huicheng, Jiangmen, Shenzhen, Shitongshui, Zhongshan. 2. Beijing: Mentougou. 5. Ronghulu: Longfeng, Saertu. 6. Shenyang: Dongling, Sujiatun, Yuhong. 12. Dalian: Jinzhou (Dalian). 20. Shishi: Quanzhou. 28. Zhuhai: Aomen. 55. Jiaojiang: Luqiao. 136. Chengde: Chaoyang. 141. Jingzhou: Sashi. The total number of natural cities and merged natural cities may be below 300 as some cities do not meet the threshold of 40 in that year. The area of the natural cities is calculated using the number of places per city, where one grid has a maximum size of 3 km x 3 km (and less if it is located at the country border). Area in  $km^2$ . Abbreviations: AH = Anhui Province, HL = Heilongjiang Province, JS = Jiangsu Province, LN = Liaoning Province, NM = Inner Mongolia Autonomous Region, SD = Shandong Province.

Table 2.9: ESTIMATION RESULTS FOR ALTERNATIVE NON-TOBIT MODELS

	(1)	(2)	(3)
	OLS	OLS	OLS
	$radiance_{it}$	$radiance_{it}$	$radiance_{it}$
$radiance_{it-1}$			0.609*** (0.002)
$radiance_{it-2}$			0.194*** (0.002)
$radiance_{it-3}$			0.100*** (0.002)
$radiance_{i1992}$			0.019*** (0.001)
$\ln(popdens_{i1990})$			0.098*** (0.008)
$\ln(dist\ to\ road_i)$	-2.851*** (0.066)	-2.195*** (0.063)	-0.185*** (0.006)
$\ln(dist\ to\ rail_i)$	-2.915*** (0.068)	-2.405 (0.067)	-0.044*** (0.007)
$\ln(dist\ to\ ocean_i)$		-1.434*** (0.092)	-0.090*** (0.008)
$\ln(dist\ to\ water_i)$		-1.417*** (0.072)	-0.080*** (0.007)
$\ln(dist\ to\ center_i)$		-7.130*** (0.143)	-0.159*** (0.013)
$\ln(dist\ to\ adborder_i)$		-0.404*** (0.138)	-0.094** (0.046)
$\ln(altitude_i)$		-0.261* (0.145)	-0.066*** (0.013)
$longitude_i$		-0.026 (0.034)	-0.004 (0.003)
$latitude_i$		1.183*** (0.131)	0.044*** (0.011)
$\ln(rain_i)$		-2.387*** (0.561)	-0.160*** (0.047)
$temperature_i$		1.183*** (0.120)	0.060*** (0.011)
$sd\ temperature_i$		-1.633*** (0.186)	-0.066 (0.013)
$hukou_{it}$		-1.966*** (0.253)	-1.502*** (0.079)
$small_{it}$		3.059*** (0.257)	1.667*** (0.079)
$medium_{it}$		2.094*** (0.231)	1.093*** (0.071)
$large_{it}$		-0.480** (0.231)	-0.807*** (0.071)

Continued.

Table 2.9: *Continued.*

	(1)	(2)	(3)
	OLS	OLS	OLS
	$radiance_{it}$	$radiance_{it}$	$radiance_{it}$
$SEZ_{it}$		0.520*** (0.179)	0.769*** (0.118)
$firstwave_{it}$		-2.192*** (0.311)	-0.245*** (0.059)
$secondwave_{it}$		0.003 (0.220)	-0.214*** (0.032)
$thirdwave_{it}$		0.401*** (0.144)	-1.468*** (0.222)
Constant	64.25*** (1.652)	140.6*** (10.22)	3.105*** (0.684)
Place fixed effects	YES	YES	YES
Satellite effects	YES	YES	YES
Control variables		YES	YES
Lagged variables			YES
# observations	432,674	432,674	373,673
# places	19,667	19,667	19,667

*Notes:* Reported coefficients are marginal effects. Standard errors are reported in parentheses ; \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. Squared terms are included for the following geography and climate variables:  $\ln(dist\ to\ road_i)$ ,  $\ln(dist\ to\ rail_i)$ ,  $\ln(dist\ to\ ocean_i)$ ,  $\ln(dist\ to\ water_i)$ ,  $\ln(dist\ to\ adborder_{it})$ ,  $\ln(dist\ to\ center_i)$ ,  $\ln(altitude_i)$ ,  $\ln(rain_i)$ ,  $temperature_i$ ,  $sd\ temperature_i$ . Contrary to the summary statistics in Table 2.2, all distance measures in the empirical estimation are in units of meters rather than kilometers.

Table 2.10: ESTIMATION RESULTS DYNAMIC TOBIT BY CATEGORIES

	(1)	(2)	(3)	(4)
	$radiance_{it}$	$radiance_{it}$	$radiance_{it}$	$radiance_{it}$
	$Nat = 1,$	$Nat = 1,$	$Nat = 0,$	$Nat = 0,$
	$Admin = 1$	$Admin = 0$	$Admin = 1$	$Admin = 0$
$radiance_{it-1}$	0.487*** (0.005)	0.530*** (0.003)	0.263*** (0.003)	0.226*** (0.002)
$radiance_{it-2}$	0.167*** (0.005)	0.142*** (0.003)	0.106*** (0.003)	0.090*** (0.002)
$radiance_{it-3}$	0.024*** (0.005)	0.030*** (0.003)	0.090*** (0.003)	0.124*** (0.002)
$radiance_{i1992}$	0.161*** (0.006)	0.207*** (0.004)	0.079*** (0.003)	0.092*** (0.002)
$\ln(popdens_{i1990})$	0.180*** (0.044)	0.084*** (0.027)	0.103** (0.040)	0.094*** (0.022)
$\ln(dist\ to\ road_i)$	-0.324*** (0.025)	-0.315*** (0.016)	-0.294*** (0.022)	-0.387*** (0.016)
$\ln(dist\ to\ rail_i)$	-0.004 (0.028)	-0.131*** (0.018)	-0.083*** (0.025)	-0.046*** (0.017)
$\ln(dist\ to\ ocean_i)$	-0.342*** (0.040)	-0.245*** (0.021)	-0.320*** (0.035)	-0.236*** (0.022)
$\ln(dist\ to\ water_i)$	-0.118*** (0.027)	-0.140*** (0.018)	-0.113*** (0.026)	-0.061*** (0.018)
$\ln(dist\ to\ center_i)$	-0.268*** (0.055)	0.123*** (0.040)	0.219*** (0.044)	-0.290*** (0.038)
$\ln(dist\ to\ adborder_i)$	-0.003 (0.031)	0.036 (0.022)	0.025 (0.029)	-0.026 (0.021)
$\ln(altitude_i)$	0.159*** (0.048)	0.039 (0.037)	-0.188*** (0.052)	-0.364*** (0.039)
$longitude_i$	-0.040*** (0.010)	0.014* (0.008)	-0.012 (0.010)	0.025*** (0.008)
$latitude_i$	0.050 (0.037)	0.090*** (0.024)	0.275*** (0.038)	-0.009 (0.025)
$\ln(rain_i)$	-0.561*** (0.173)	-0.778*** (0.122)	-0.636*** (0.189)	-1.503*** (0.130)
$temperature_i$	0.147*** (0.034)	0.220*** (0.029)	0.151*** (0.035)	-0.101*** (0.031)
$sd\ temperature_i$	0.031 (0.060)	-0.056 (0.040)	-0.272*** (0.062)	-0.159*** (0.043)

Continued.

Table 2.10: *Continued.*

	(1)	(2)	(3)	(4)
	$radiance_{it}$	$radiance_{it}$	$radiance_{it}$	$radiance_{it}$
	$Nat = 1,$	$Nat = 1,$	$Nat = 0,$	$Nat = 0,$
	$Admin = 1$	$Admin = 0$	$Admin = 1$	$Admin = 0$
$hukou_{it}$	-1.068*** (0.302)	-2.844*** (0.165)	1.398*** (0.272)	-0.122 (0.090)
$small_{it}$	0.985*** (0.281)	3.073*** (0.162)	-0.559** (0.252)	0.822*** (0.085)
$medium_{it}$	0.156 (0.223)	1.573*** (0.112)	-0.708*** (0.243)	0.331*** (0.077)
$large_{it}$	-0.188 (0.203)	-1.034*** (0.102)	0.190 (0.227)	-0.930*** (0.078)
$SEZ_{it}$	1.323*** (0.192)	0.476*** (0.086)	1.076*** (0.173)	-0.244*** (0.089)
$firstwave_{it}$	-3.040*** (0.388)	-0.198 (0.159)	-2.047*** (0.293)	0.835*** (0.142)
$secondwave_{it}$	-0.810*** (0.185)	-0.471*** (0.072)	-1.362*** (0.169)	-0.362*** (0.072)
$thirdwave_{it}$	-0.380*** (0.135)	0.224*** (0.066)	0.534*** (0.120)	0.688*** (0.069)
Constant	11.90*** (3.227)	21.32*** (3.970)	25.09*** (2.600)	15.14*** (2.921)
Satellite effects	YES	YES	YES	YES
# observations	51,995	55,785	159,772	106,121
# places	5,957	4,781	12,330	14,040

*Notes:* Reported coefficients are marginal effects. Standard errors are reported in parentheses ; \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. All columns include squared terms for the following geography and climate variables:  $\ln(dist\ to\ road_i)$ ,  $\ln(dist\ to\ rail_i)$ ,  $\ln(dist\ to\ ocean_i)$ ,  $\ln(dist\ to\ water_i)$ ,  $\ln(dist\ to\ adborder_{it})$ ,  $\ln(dist\ to\ center_i)$ ,  $\ln(altitude_i)$ ,  $\ln(rain_i)$ ,  $temperature_i$ ,  $sd\ temperature_i$ . All columns include satellite effects. Contrary to the summary statistics in Table 2.2, all distance measures in the empirical estimation are in units of meters rather than kilometers. Abbreviations: Nat = natural city, Admin = within administrative boundary.

Table 2.11: ROBUSTNESS – TRANSITION MATRIX OF COUNTERFACTUAL  
 $\ln(\text{dist to road}_i)$

Definition:  $Nat = 1$  if radiance  $\geq 50$ ,  $Nat = 0$  if radiance  $< 50$

		Counterfactual: $\ln(\text{dist to road}_i)$		
		$Nat = 1$	$Nat = 0$	<b>Total</b>
Baseline	$Nat = 1$	99.87 <sub>(2007)</sub>	0.13 <sub>(2007)</sub>	100
		99.70 <sub>(2012)</sub>	0.30 <sub>(2012)</sub>	100
		99.51 <sub>(2017)</sub>	0.49 <sub>(2017)</sub>	100
		99.48 <sub>(2022)</sub>	0.52 <sub>(2022)</sub>	100
		99.50 <sub>(2027)</sub>	0.50 <sub>(2027)</sub>	100
	$Nat = 0$	0.49 <sub>(2007)</sub>	99.51 <sub>(2007)</sub>	100
		2.77 <sub>(2012)</sub>	97.23 <sub>(2012)</sub>	100
		7.94 <sub>(2017)</sub>	92.06 <sub>(2017)</sub>	100
		18.58 <sub>(2022)</sub>	81.42 <sub>(2022)</sub>	100
		35.03 <sub>(2027)</sub>	64.97 <sub>(2027)</sub>	100
<b>Total</b>		28.71 <sub>(2007)</sub>	71.29 <sub>(2007)</sub>	100
		39.02 <sub>(2012)</sub>	60.98 <sub>(2012)</sub>	100
		51.50 <sub>(2017)</sub>	48.50 <sub>(2017)</sub>	100
		67.11 <sub>(2022)</sub>	32.89 <sub>(2022)</sub>	100
		82.16 <sub>(2027)</sub>	17.84 <sub>(2027)</sub>	100

*Notes:* Additional to the radiance threshold, the CCA condition is a necessary condition for a place to be assigned *natural city* ( $Nat = 1$ ). The CCA condition implies that a place has to be in the neighborhood of a cluster of places with average radiance greater or equal the threshold. Abbreviation:  $Nat =$  natural city.

Table 2.12: ROBUSTNESS – TRANSITION MATRIX OF COUNTERFACTUAL  
 $\ln(\text{dist to rail}_i)$

Definition:  $Nat = 1$  if radiance  $\geq 50$ ,  $Nat = 0$  if radiance  $< 50$

		Counterfactual: $\ln(\text{dist to rail}_i)$		
		$Nat = 1$	$Nat = 0$	<b>Total</b>
Baseline	$Nat = 1$	100.00 <sub>(2007)</sub>	0.00 <sub>(2007)</sub>	100
		99.93 <sub>(2012)</sub>	0.07 <sub>(2012)</sub>	100
		99.88 <sub>(2017)</sub>	0.12 <sub>(2017)</sub>	100
		99.82 <sub>(2022)</sub>	0.18 <sub>(2022)</sub>	100
		99.84 <sub>(2027)</sub>	0.16 <sub>(2027)</sub>	100
	$Nat = 0$	0.06 <sub>(2007)</sub>	99.94 <sub>(2007)</sub>	100
		0.52 <sub>(2012)</sub>	99.48 <sub>(2012)</sub>	100
		1.54 <sub>(2017)</sub>	98.46 <sub>(2017)</sub>	100
		3.65 <sub>(2022)</sub>	96.35 <sub>(2022)</sub>	100
		6.31 <sub>(2027)</sub>	93.69 <sub>(2027)</sub>	100
<b>Total</b>		28.43 <sub>(2007)</sub>	71.57 <sub>(2007)</sub>	100
		37.70 <sub>(2012)</sub>	62.30 <sub>(2012)</sub>	100
		48.32 <sub>(2017)</sub>	51.68 <sub>(2017)</sub>	100
		61.34 <sub>(2022)</sub>	38.66 <sub>(2022)</sub>	100
		74.68 <sub>(2027)</sub>	25.32 <sub>(2027)</sub>	100

*Notes:* Additional to the radiance threshold, the CCA condition is a necessary condition for a place to be assigned *natural city* ( $Nat = 1$ ). The CCA condition implies that a place has to be in the neighborhood of a cluster of places with average radiance greater or equal the threshold. Abbreviation:  $Nat =$  natural city.

# Chapter 3

## Decomposing the Economic Effects of Transport Infrastructure

### 3.1 Introduction

In the wide range of determinants of trade and migration costs, few are actionable by policy makers. Of those, infrastructure investments, as a *transport cost* variable, appear most important, both sub-nationally and internationally. The need of transport infrastructure investments appears particularly pertinent, whenever economies are economically large or fast-growing while facing high transport costs due to bad transport networks. A prime example of such a country is China. Since the early 1990s and particularly since its membership in the World Trade Organization (WTO) in 2001, we did see an enormous surge in transport infrastructure projects in China. Of all the transport infrastructure projects in the world worth \$1,1 billion funded by the World Bank in 2017, \$450 million were invested in China.<sup>1</sup> In the same year, the estimated private investments into transport infrastructure in China amounted to \$13 billion.<sup>2</sup> Similarly, India saw its share

---

<sup>1</sup>The World Bank funded 29 projects on the development of roads and highways in China since 1990 and the process is still ongoing. Projects & Operations, The World Bank <http://projects.worldbank.org>

<sup>2</sup>Private Participation in Infrastructure Project Database, The World Bank. <http://ppi.worldbank.org>

of infrastructure investment to GDP double between 2004 and 2016.<sup>3</sup> Overall, we see a widespread use of infrastructure improvements to make places (countries or regions) more accessible for goods and factors.

Motivated by this line of thought, earlier research documented the role of infrastructure for *international* goods trade and migration (Michaels, 2008; Duranton et al., 2014; Donaldson, 2018) as well as for *interregional* goods trade (Tombe and Zhu, 2015; Caliendo et al., 2018) and migration (Baum-Snow, 2007; Duranton and Turner, 2012; Fretz et al., 2017). Economic models suggest that infrastructure effects on economic outcomes tend to be asymmetric across economic agents and jurisdictions due to their inherent heterogeneity in many dimensions. Additionally, it has been documented that infrastructure improvements induce heterogeneous effects at the regional and even at the national level (Chandra and Thompson, 2000; Faber, 2014). The latter poses the question of identifying the optimal size and spatial design of infrastructure networks (Fajgelbaum and Schaal, 2017). Not surprisingly in view of the rapid changes witnessed in China, economists have looked into the development there, and we have now good evidence of Chinese transport infrastructure effects on economic growth (Banerjee et al., 2012; Yu et al., 2012; Egger et al., 2017; Baum-Snow et al., 2018) and real consumption (Baum-Snow et al., 2017; Baum-Snow and Turner, 2017) as well as the heterogeneity of these effects (Faber, 2014; Qin, 2017; Lovely et al., 2019). Moreover, there is recent evidence on the specific effects on routing, volume, and congestion in affecting broader measures of welfare beyond mere consumption (Ma and Tang, 2020; Barwick et al., 2018; Allen and Arkolakis, 2019).

What economists tend to have in mind with accessibility is more or less exclusively: *for goods and factors*. Moreover, the vast majority of research on the economic effects of infrastructure improvements has focused on *highway networks*. The purpose of this paper is to take a broader view on transport infrastructure, and consider the complexity and multiplicity of the economic effects of infrastructure improvements. In particular, we study: (i) multiple channels through which the extension of the road network affected region-specific well-being of households beyond the standard transport of goods and factors, and (ii) the heterogeneous

---

<sup>3</sup>Infrastructure Investment, OECD.  
<https://data.oecd.org/transport/infrastructure-investment.htm>



economic effects across network types (i.e., highway or regional networks).

Beyond the standard transport of goods and factors, this paper considers as well the *dissipation of technology* and *accessibility of amenities beyond locally available ones*. Of those two, only the former has been on the radar of economic models (see, among others, Coe and Helpman, 1995; Rodriguez-Clare, 1996; Feenstra and Kee, 2008; Fracasso and Vittucci Marzetti, 2015).<sup>4</sup> However, in most quantitative multi-regional models, amenities – as an unobservable parameter to characterize the desirability of a place to locate in – are a strictly local attribute,<sup>5</sup> and technology is either strictly local or fully global.<sup>6</sup> We propose a framework that allows us to analyze the economic effects of transport infrastructure through each channel separately, as well as their interactions.

Transport networks in general and road networks in particular are multi-layered objects. In most countries, national road networks are composed of a highway network as well as multiple layers of integrated regional road networks. The economic literature on quantitative effects of road networks has focused almost exclusively on highways (see, among others, Baum-Snow, 2007; Michaels, 2008; Duranton and Turner, 2012; Faber, 2014). However, in most countries highway networks connect only a small subset of the micro-regions directly. Regional road networks, on the other hand, establish important access nodes to the highway networks and they connect many such regions, where highways do not. For instance, according to the data used in this paper, 32% of the reduction in inter-prefectural travel times in China between 2000 and 2013 accrue to changes in regional road networks and not highways.<sup>7</sup>

We propose a quantitative multi-region model of the (open) Chinese economy

---

<sup>4</sup>The discussion emerged around the debate whether goods trade or technology should be mainly held responsible for changes in factor demand, and the mentioned authors proposed and found that technology “travels” with, or at least along, the same routes as goods trade.

<sup>5</sup>Even though in a different context and using a static approach, let us note that recent works in urban economics analyzing amenity spillovers as part of local agglomeration forces are similar in spirit to our approach of amenity diffusion (see, among others, Ahlfeldt et al., 2015).

<sup>6</sup>For instance, in static multi-country quantitative models of international trade, technology (and productivity) is characterized by fixed country- or region-specific parameters (Eaton and Kortum, 2002; Caliendo and Parro, 2015). In dynamic multi-region quantitative models, innovation is a local activity that depends on the factors available there, but technology dissipates completely to all regions within a relatively short period (see Desmet and Rossi-Hansberg, 2009, 2014; Desmet et al., 2018; Allen and Donaldson, 2018).

<sup>7</sup>This number is based on the sum of bilateral times between all prefecture pairs in China. See Figure 3.1.

which features the following aspects. First, both goods and individuals are mobile between regions up to some frictions as is customary in quantitative regional general-equilibrium work. Second, the production potential of regions depends on the local endogenous labor supply and the exogenous availability of land as well as on productivity. Productivity consists of two parts: one that is exogenous to a region and one that is endogenous and a function of local labor supply and spillovers from other regions. Similarly, local amenities are determined by two components: one that is exogenous to the region and one that is endogenous and a function of the local population and spillovers from other regions. Third, mobility and trade frictions *inter alia* depend directly on the transport infrastructure. Accordingly, the improvement of the transport network affects regional well-being through four channels: accessibility of goods markets (trade); accessibility of residence places (migration); accessibility of productivity (transportation-bound technology spillovers); and accessibility of amenities (transportation-bound amenity spillovers).

In conducting a structural quantitative analysis for 330 prefectures in China and the rest of the world (RoW), we build on three pillars. First, we use hand-collected, and digitized data on the Chinese road network between 2000-2013, distinguishing between three hierarchical layers, namely highways, province-level roads, and prefecture-level roads. The sources of these data are 14 detailed road maps covering the entire Chinese road network. This data-set permits tracking the changes in the road network for each layer and it allows computing the connectivity between prefectures in terms of travel times. The magnitude of China's road infrastructure improvements between 2000 and 2013 is nothing else than startling: the total road network length increased from 371,385 kilometers to 515,480 kilometers. Hence, by 2013 about 28% of all road-kilometers were connections which did not exist in 2000. The total length of highways alone almost tripled (from 50,127 kilometers in 2000 to 142,983 kilometers in 2013), and even the total length of Chinese regional (non-highway) roads increased by 77,855km, implying a net increase of 21%.

Second, we develop a novel approach in dealing with the endogeneity of the realized road infrastructure placements in a given year. We address this endo-

geneity through an Instrumental Variable (IV) approach in which the observed travel time between Chinese prefectures is instrumented by the travel time on an optimal road network.<sup>8</sup> The optimal road network is obtained by solving a modified version of the classical Monge-Kantorovich transportation problem (Monge, 1781; Kantorovitch, 1958) using historical population data and estimated costs of building roads on a fine geographical grid of China. This approach has three key advantages: (i) it offers an high predictive power to the (dense) Chinese road network; (ii) it permits a derivation of an optimal road typology – between highways, national roads, and secondary roads – in a straightforward manner; and (iii) it solves for the global optimum road network.

Third, we conduct a counterfactual analysis, where we compare the evolution of the Chinese economy since the year 2000 in terms of two long-run equilibria: one where the road infrastructure stays constant at its level of 2000, and one where it is changed in the year 2000 to its level of 2013. We then decompose the overall effects of road infrastructure improvements in China on population densities and regional income across Chinese prefectures into four components: one accruing to the increased accessibility of local amenities for the population in a prefecture; one pertaining to the increased accessibility of technology levels elsewhere in China by firms in a prefecture; one capturing transport cost reductions for goods; and one reflecting reduced migration costs. Moreover, we decompose the overall effects into their components which accrue to changes in the highway- versus other road network components.

There are three main take-away messages from the corresponding analysis. First, the road network improvements made between 2000 and 2013 foster overall regional convergence in population and real income by relocating population from large centers to mid-sized prefectures in central China. Second, dissipation of technology and reduced trade frictions have large positive real income effects in prefectures that gain the most in connectivity. Amenity diffusion and reduced migration frictions, however, work in opposite directions with a similar order of magnitude for both population and real income. Finally, whereas highway improvements have strong integration effects on formerly remote places, regional

---

<sup>8</sup>See, among others, Faber (2014) and Alder and Kondo (2018) for examples of the use of an optimal transport network in an instrumental variable approach.

road improvements relieve former differences in local connectivity and mitigate the differences that are generated by the highway network.

The paper proceeds as follows. In Section 3.2, we present our theoretical framework. We describe the evolution of the Chinese road network between 2000 and 2013 and detail our instrumentation strategy in Section 3.3. Before presenting the counterfactual analysis in Section 3.5, we calibrate the model using Chinese prefecture-level data in Section 3.4. Section 3.6 concludes.

## 3.2 Theoretical Framework

In this section, we outline a multi-region general equilibrium model that is amenable to studying the decomposition of economic effects of road infrastructure improvements in China within the last two decades. Locations differ according to (time-varying) amenity and productivity fundamentals. These fundamentals evolve according to a dynamic amenity- and technology-diffusion process that crucially depends on each location's integration in the transport network. Locations interact in product markets through (costly) trade in goods and in labor markets through (costly) migration. Trade and migration frictions are directly affected by the transport network, which implies feedback effects in prices and consumption patterns through changes in trade volumes and individual mobility. In what follows, we describe the elements of the theoretical framework in detail.

### 3.2.1 Setup

Consider a world composed of a finite number of locations  $i \in S = \{1, \dots, s\}$  on a lattice. Time is discrete and indexed by  $t$ , which refers to years. Available land in each location is allocated to residential ( $R$ ) and commercial ( $C$ ) use and measured by the land densities  $H_i^R, H_i^C > 0$ , respectively. We assume that income from land is owned by immobile landlords, who receive expenditures on commercial and residential land as income and consume only local goods.<sup>9</sup> The world economy is populated by a total of  $\bar{L}$  individuals who are endowed with one unit of labor

---

<sup>9</sup>This is a customary assumption in regional economics, and it ensures that neither the individuals' nor firms' location decisions are affected by the distributional assumptions of local land rents (see, e.g., Ahlfeldt et al., 2015; Monte et al., 2018).

each, which they supply inelastically. After deciding on their residence location, individuals produce in the location in which they live and consume.

### 3.2.2 Individual Preferences

Each period, the utility of an individual  $\omega$  residing in  $i$  at period  $t$  is defined over the consumption of a set of differentiated products ( $C_{it}$ ), residential land use ( $H_i^R$ ), local endogenous amenities ( $a_{it}$ ), an idiosyncratic utility shock ( $\varepsilon_{it}^U$ ), and bilateral moving costs ( $\kappa_{ijt}$ ). Formally, individual utility takes the following Cobb-Douglas form:<sup>10</sup>

$$U_{ijt}(\omega) = \left( \frac{C_{it}(\omega)}{\alpha} \right)^\alpha \left( \frac{H_i^R(\omega)}{1 - \alpha} \right)^{1 - \alpha} \frac{a_{it}}{\kappa_{ijt}} \varepsilon_{it}^U(\omega), \quad (3.1)$$

where  $\alpha \in (0, 1)$  denotes the Cobb-Douglas share of goods consumption in utility. Endogenous amenities,  $a_{it}$ , are determined by a dynamic amenity process, as follows

$$a_{it} = \left( \frac{L_{it}}{H_i^R} \right)^{-\lambda_1} \exp \left( \int_S \mathbb{W}_{ijt-1}^R \log a_{jt-1} dj \right)^{\lambda_2} \varepsilon_{it}^R. \quad (3.2)$$

Endogenous amenities decrease with residential population density  $\left( \frac{L_{it}}{H_i^R} \right)$  as long as  $\lambda_1 > 0$ , and  $\lambda_1$  denotes a congestion-externality parameter. It is further characterized by the past weighted amenity level of all other locations,  $\exp \left( \int_S \mathbb{W}_{ijt-1}^R \log(a_{jt-1}) dj \right)$ . The amenity-diffusion process is governed by an  $ijt$ -specific weighting scalar,  $\mathbb{W}_{ijt-1}^R$  and  $\lambda_2 \in (0, 1)$ . The definition of the weighting scalar  $\mathbb{W}_{ijt-1}^R$  crucially depends on the road network and will be discussed in detail in Section 3.4.3. The parameter  $\varepsilon_{it}^R$  is a location- and time-specific amenity shifter.

Individuals choose a residence that maximizes their utility in (3.1), after learning their individual preference shock,  $\varepsilon_{it}^U(\omega)$ . The individual preference shock governs the dispersion of utility and, hence, absorbs unobserved preferences for location choices that cannot be explained by income and price differences. We assume that  $\varepsilon_{it}^U(\omega)$  is drawn from a Fréchet distribution with shape parameter  $1/\Omega$ .<sup>11</sup>

Individuals consume residential housing,  $H_i^R$ , and a continuum of tradable

<sup>10</sup>We do not use a time index on  $H_i^R$  to indicate that it will be fixed in the data.

<sup>11</sup>The probability that individual  $\omega$  in location  $i$  at time  $t$  draws an idiosyncratic preference smaller than  $z$  is given by  $\Pr(\varepsilon_{it}^U(\omega) \leq z) = \exp(-z^{-(1/\Omega)})$ . We assume that  $\varepsilon_{it}^U(\omega)$  is i.i.d. across locations, individuals and time.

varieties  $\rho$ , which are combined in a consumption bundle,  $C_{it}$ .<sup>12</sup> Their indirect utility is defined as

$$u_{ijt}(\omega) = \tilde{u}_{it} \frac{\varepsilon_{it}^U(\omega)}{\kappa_{ijt}} \quad \text{with} \quad \tilde{u}_{it} = \frac{a_{it} w_{it}}{P_{it}^\alpha r_{it}^{1-\alpha}} = a_{it} y_{it}, \quad (3.3)$$

where  $\tilde{u}_{it}$  is the indirect utility that an individual derives from residing in location  $i$ , which is independent of factors that determine the ex-ante location choice to reside in  $i$ .<sup>13</sup>  $\tilde{u}_{it}$  is a function of local amenities ( $a_{it}$ ), income ( $w_{it}$ ) and the corresponding housing and consumer prices ( $r_{it}^R$  and  $P_{it}$ ). We define the goods-consumption price index at location  $i$  and time  $t$  as  $P_{it} = \left[ \int_S p_{it}(\rho)^{1-\sigma} d\rho \right]^{\frac{1}{1-\sigma}}$ . For later convenience, we introduce  $y_{it}$  to denote real income per capita.

### 3.2.3 Technology and Production

Firms in  $i$  at  $t$  use two production factors, labor ( $L_{it}$ ) and commercial land ( $H_i^C$ ), to produce output units ( $Q_{it}$ ) of product varieties  $\rho$  under a Cobb-Douglas technology.<sup>14</sup>

$$Q_{it}(\rho) = z_{it}(\rho) T_{it} L_{it}(\rho)^\mu H_i^C(\rho)^{1-\mu}, \quad (3.4)$$

where  $\mu \in (0, 1)$  denotes the labor share in production. Output depends on production inputs and a firm's total-factor productivity, which is determined by two components. The first component is a variety-specific exogenous productivity parameter,  $z_{it}$ , that is drawn from a Fréchet distribution with shape parameter  $\theta > 0$ . The second component is an endogenous technology part,  $T_{it}$ , which is determined as

$$T_{it} = \left( \frac{L_{it}}{H_i^C} \right)^{\gamma_1} \exp \left( \int_S \mathbb{W}_{ijt-1}^C \log T_{jt-1} dj \right)^{\gamma_2} \varepsilon_{it}^C. \quad (3.5)$$

Endogenous technology increases with commercial population density  $\left( \frac{L_{it}}{H_i^C} \right)$ , and  $\gamma_1 > 0$  is an agglomeration parameter in production. Analogously to

<sup>12</sup>We assume that  $C_{it} = \left[ \int_0^1 c_{it}(\rho)^{\frac{\sigma-1}{\sigma}} d\rho \right]^{\frac{\sigma}{\sigma-1}}$  according to constant-elasticity-of-substitution (CES) preferences, where  $\sigma > 0$  is the elasticity of substitution.

<sup>13</sup>In other words, this assumes that bilateral moving costs,  $\kappa_{ijt}$ , and the idiosyncratic preference for that location,  $\varepsilon_{it}^U(\omega)$ , are irrelevant for the location-specific indirect utility  $\tilde{u}_{it}$ .

<sup>14</sup>We use a time index on  $L_{it}$  but not on  $H_i^C$  to indicate that the former will change from period to period, while the latter will be fixed in the data.

amenities, it is further characterized by the past-weighted aggregate technology levels elsewhere,  $\exp\left(\int_S \mathbb{W}_{ijt-1}^C \log(T_{jt-1}) dj\right)$ .  $\gamma_2 \in (0, 1)$  governs the technology diffusion process.  $\mathbb{W}_{ijt-1}^C$  is an  $ijt$ -specific weighting scalar, which crucially depends on the road network (see Section 3.4.3 for further details) and  $\varepsilon_{it}^C$  is a location- and time-specific technology shifter.

After learning their exogenous productivity draw,  $z_{it}$ , firms maximize their profits by choosing the level of employment and commercial land through

$$\max_{L_{it}(\rho), H_i^C(\rho)} p_{it}(\rho) z_{it}(\rho) T_{it} L_{it}(\rho)^\mu H_i^C(\rho)^{(1-\mu)} - w_{it} L_{it}(\rho) - r_{it}^C H_i^C(\rho),$$

where  $p_{it}(\rho)$  is the price a firm charges for a product that is sold in  $i$  and period  $t$ . The model structure implies that  $p_{it}(\rho)$  is directly proportional to unit costs and inversely proportional to a location's total-factor productivity, which follows the basic price-productivity relation as outlined in Eaton and Kortum (2002). Then,

$$E[p_{it}(\rho)] = \frac{o_{it}}{T_{it} z_{it}} \quad \text{with} \quad o_{it} = \left[ \frac{(1-\mu)^{\mu-1}}{\mu^\mu} \right] r_{it}^{C^{1-\mu}} w_{it}^\mu = \frac{1}{\mu} w_{it} \left( \frac{L_{it}}{H_i^C} \right)^{1-\mu}, \quad (3.6)$$

where  $o_{it}$  denotes the unit costs in location  $i$  at time  $t$  which firms take as given.

### 3.2.4 Trade

Bilateral sales from location  $i$  to location  $j$  are subject to iceberg transportation costs,  $\zeta_{ijt} \geq 1$ . The price of a good produced in  $i$  and consumed in  $j$  is  $p_{ijt} = p_{it} \zeta_{ijt} = \frac{o_{it} \zeta_{ijt}}{T_{it} z_{it}}$ . Then, bilateral trade shares take on the well-known gravity form as in Eaton and Kortum (2002):

$$\pi_{ijt}^C = \frac{T_{it} [o_{it} \zeta_{ijt}]^{-\theta}}{\int_S T_{kt} [o_{kt} \zeta_{kjt}]^{-\theta} dk}, \quad \forall i, j \in N, \quad (3.7)$$

where  $\pi_{ijt}^C$  is the share of expenditures in  $j$  on  $i$  and the price index in a location  $j$  is given by

$$P_{jt} = \bar{p} \left[ \int_S T_{kt} [o_{kt} \zeta_{kjt}]^{-\theta} dk \right]^{-\frac{1}{\theta}}, \quad (3.8)$$

where  $\bar{p}$  is a normalizing constant.

### 3.2.5 Migration

Bilateral migration from location  $i$  to location  $j$  is subject to migration costs  $\kappa_{ijt} \geq 1$ . The indirect utility in (3.3) is monotonically increasing with the idiosyncratic Fréchet-distributed preference shock  $\varepsilon_{it}^U(\omega)$ . Then, the indirect utility for an individual residing in  $i$  at  $t$  is itself Fréchet-distributed.<sup>15</sup> Analogously to trade shares, the share of individuals coming from  $j$  and choosing to reside in  $i$  can be expressed in the well-known gravity form as

$$\pi_{ijt}^R = \frac{(a_{it}w_{it})^{1/\Omega}(\kappa_{ijt}P_{it}^\alpha r_{it}^{R1-\alpha})^{-1/\Omega}}{\int_S (a_{kt}w_{kt})^{1/\Omega}(\kappa_{kjt}P_{kt}^\alpha r_{kt}^{R1-\alpha})^{-1/\Omega} dk} = \frac{\tilde{u}_{it}^{1/\Omega} \kappa_{ijt}^{-1/\Omega}}{\int_S \tilde{u}_{kt}^{1/\Omega} \kappa_{kjt}^{-1/\Omega} dk}. \quad (3.9)$$

Summing  $\pi_{ijt}^R$  across migration locations  $j$  for a given residence, we obtain the probability that individuals reside in location  $i$ :

$$\pi_{it}^R = \frac{\int_S L_{ijt} dj}{\bar{L}} = \frac{L_{it}}{\bar{L}} = \frac{\tilde{u}_{it}^{1/\Omega} \int_S \kappa_{ijt}^{-1/\Omega} dj}{\int_S \int_S \tilde{u}_{kt}^{1/\Omega} \kappa_{kjt}^{-1/\Omega} dk dj}. \quad (3.10)$$

### 3.2.6 Equilibrium

Profits and utility are maximized within each period, as neither individuals nor firms are forward-looking. This implies that both firms and individuals are uninformed about the endogenous dynamic process that guides the amenity and technology diffusion. Furthermore, we assume that decisions on the extension of transport network, which directly impact the amenity and technology diffusion as well as trade and migration costs, are taken from an absentee central planner and, hence, are not anticipated.

An equilibrium is characterized such that goods and factor markets clear in each period. More formally, for any initial amenity, technology, population, and land vectors  $\{a_{it}, T_{it}, L_{it}, H_i^R, H_i^C\}$  for all  $\{i, t\}$ , any trade- and migration-cost vectors  $\{\kappa_{ijt}, \zeta_{ijt}\}$ , as well as any weight scalars  $\{\mathbb{W}_{ijt}^R, \mathbb{W}_{ijt}^C\}$  for all  $\{i, j, t\}$ , a competitive equilibrium is a set of endogenous vectors  $\{L_{it}, w_{it}, u_{it}, P_{it}, a_{it}, T_{it}, r_{it}^R, r_{it}^C\}$  for all  $\{i, t\}$ , such that for all locations  $i$  in each time period  $t$  we have:

<sup>15</sup>Given that, utility is distributed as  $G_{it}(u) = \exp(-\Phi_{it}u^{-1/\Omega})$ , where  $\Phi_{ijt} = (a_{it}w_{it})^{1/\Omega}(\kappa_{ijt}P_{it}^\alpha r_{it}^{R1-\alpha})^{-1/\Omega}$ .



1. **Goods-market Clearing:** Total revenues in a location  $i$  are equal to the value of all location's purchases from it. Thus,

$$w_{it}L_{it} = \int_S \pi_{ijt}^C w_{jt}L_{jt} dj. \quad (3.11)$$

2. **Population Accounting:** The number of residents in all locations is equal to the total population  $\int_S L_{it} di = \bar{L}$ . In view of equation (3.10), this implies

$$L_{it} = \int_S \pi_{ijt}^R L_{jt} dj. \quad (3.12)$$

3. **Land-market Clearing:** The residential and commercial land markets in each location  $i$  are in equilibrium, so land is assigned to the highest bidder. Hence, equilibrium land rents are given as

$$r_{it}^R = [1 - \alpha]w_{it} \frac{L_{it}}{H_i^R} \quad \text{and} \quad r_{it}^C = \left[ \frac{1 - \mu}{\mu} \right] w_{it} \frac{L_{it}}{H_i^C}. \quad (3.13)$$

We can manipulate the system of equations presented above and reduce it to a system that determines the equilibrium distribution of wages,  $w_{it}$ , and employment levels,  $L_{it}$ , in all locations. Conditional on  $\{a_{it}, T_{it}, H_i^R, H_i^C, \kappa_{ijt}, \zeta_{ijt}, \mathbb{W}_{ijt}^R, \mathbb{W}_{ijt}^C\}$ , and parameter values, the equilibrium wage and equilibrium employment satisfy the following equations:

**Equilibrium Wages** using goods-market clearing in (3.11)

$$w_{it} = L_{it}^{\frac{(\gamma_1 - \theta(1-\mu)) - 1}{1+\theta}} H_i^C \frac{-\theta(\mu-1) - \gamma_1}{1+\theta} \tau_{it}^{\frac{1}{1+\theta}} \left[ \int_S \frac{w_{jt}L_{jt}\zeta_{ijt}^{-\theta}}{\int_S \Pi_{kt}\zeta_{kjt}^{-\theta} dk} dj \right]^{\frac{1}{1+\theta}}, \quad (3.14)$$

where  $\tau_{it} \equiv \exp \left( \int_S \mathbb{W}_{ijt-1}^C \log T_{jt-1} dj \right)^{\gamma_2}$  and  $\Pi_{kt} \equiv \tau_{kt} \left( \frac{L_{kt}}{H_k^C} \right)^{\gamma_1 - \theta(1-\mu)} w_{kt}^{-\theta}$ .

**Equilibrium Employment** using migration shares in (3.10)

$$L_{it} = \frac{(w_{it}^\alpha (\bar{L} \tilde{\kappa}_{it})^\Omega H_i^{R(1-\alpha)+\lambda_1} \tilde{a}_{it})^{\frac{1}{\tilde{\Omega}}} \left( \int_S \Pi_{it} \zeta_{ijt}^{-\theta} dj \right)^{\frac{\alpha}{\theta \tilde{\Omega}}}}{\left[ \int_S \int_S \left( \frac{\tilde{a}_{kt} w_{kt}^\alpha H_k^{R(1-\alpha)+\lambda_1}}{\kappa_{kjt} L_{kt}^{(1-\alpha)+\lambda_1}} \right)^{\frac{1}{\tilde{\Omega}}} \left( \int_S \Pi_{jt} \zeta_{kjt}^{-\theta} dj \right)^{\frac{\alpha}{\theta \tilde{\Omega}}} dk dj \right]^{-\frac{\Omega}{\tilde{\Omega}}}}, \quad (3.15)$$

where  $\tilde{\kappa}_{it} = \int_S \kappa_{ijt}^{-1/\Omega} dj$ ,  $\tilde{a}_{it} \equiv \exp \left( \int_S \mathbb{W}_{ijt-1}^R \log a_{jt-1} dj \right)^{\lambda_2}$  and  $\tilde{\Omega} \equiv \Omega + \lambda_1 + (1 - \alpha)$ .<sup>16</sup> Given these two sets of equilibrium values, all other endogenous variables of the model are determined.

To guarantee that the solution to the system of equations exists, is stable and unique, we establish two crucial conditions.<sup>17</sup>

**Uniqueness** There exists a unique (up-to-scale) solution of the vectors  $\{\mathbf{w}_t, \mathbf{L}_t\}$  to satisfy (3.14) and (3.15), if

$$\frac{\alpha \gamma_1}{\theta} \leq (1 - \mu \alpha) + \lambda_1 + \Omega. \quad (3.16)$$

This condition is intuitive. It states that agglomeration forces expressed in local production externalities ( $\frac{\alpha \gamma_1}{\theta}$ ) may not dominate the three dispersion forces: land availability restrictions ( $1 - \mu \alpha$ ), congestion externalities in amenities ( $\lambda_1$ ) and the variance of local preferences ( $\Omega$ ).

**Stability** An equilibrium requires that both the amenity diffusion in (3.2) as well as the technology diffusion in (3.5) reach an equilibrium level in the long run upon a shock in infrastructure. For either process, this is the case whenever (i) the infrastructure-accessibility weights  $\mathbb{W}_{ijt}^R$  and  $\mathbb{W}_{ijt}^C$  are properly normalized so that their integral for each prefecture and time period is bounded, and (ii) the adjustment-cost parameters,  $\lambda_2$  and  $\gamma_2$ , are suitably bounded. To address the former, we normalize both  $\mathbb{W}_{ijt}^R$  and  $\mathbb{W}_{ijt}^C$  such that their diagonal elements are zero and the maximum of their row sums equals unity, i.e.,  $\max_i \int_S \mathbb{W}_{ijt}^{(\cdot)} dj = 1$  for  $(\cdot) \in \{C, R\}$ . We adopt a maximum-row-sum normalization over the more

<sup>16</sup>Section 3.7 of the Appendix presents the derivation of (3.14) and (3.15) in further detail.

<sup>17</sup>Section 3.8 and 3.9 of the Appendix present a detailed derivation of the uniqueness and stability conditions.

standard row-sum normalization as it retains differences in absolute connectivity of the locations (see Kelejian and Prucha, 2010). To address the latter, we need to assume  $|\lambda_2|, |\gamma_2| \in ]0, 1[$ .<sup>18</sup>

### 3.3 Transport Network

In the last three decades, the development of the transport infrastructure network has been a key focus for the Chinese government. A major part of the transport infrastructure investments was spent on the extension and technological improvements of the national road system. In what follows, this section describes a hand-collected novel dataset on the Chinese road network (Section 3.3.1) and explains in detail in which way we treat it as endogenous and overcome potential associated biases in estimating key model parameters when using it in estimating equations (Section 3.3.2).

#### 3.3.1 Chinese Transport Infrastructure Network

We collected 14 road atlases covering the entire Chinese road network for the years 2000-2013.<sup>19</sup> All books were digitalized in high-resolution portable document format (PDF). After digitalization, we geo-referenced all maps and classified the road features into three different categories: highways (including the National Trunk Highway System built between 1992-2007), provincial-level roads, and prefecture-level roads using ArcGIS software.<sup>20</sup> Furthermore, as a significant fraction of the transport infrastructure investments was spent on upgrading the existing network, our classification scheme pays particular attention to such upgrading. As a result, we obtained a highly comprehensive, novel database on the Chinese road network (including highways, provincial-level roads, and prefecture-level roads) between 2000 and 2013. It permits observing changes in infrastructure which can inform

---

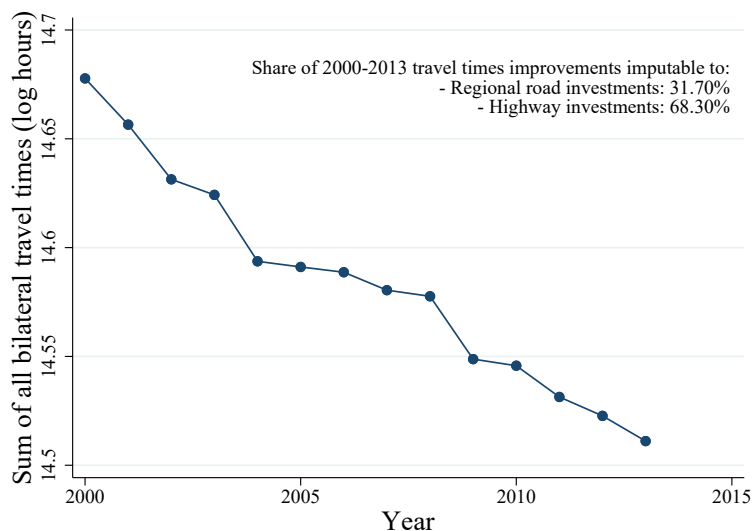
<sup>18</sup>We detail the stability condition for the amenity- and technology-diffusion processes in further detail in Section 3.9 of the Appendix.

<sup>19</sup>We list the sources of all road atlases in Table 3.6 of Section 3.11 in the Appendix.

<sup>20</sup>In the road atlases, provincial-level roads are sometimes also referred to as national roads, and prefecture-level roads are dubbed secondary roads. The main difference between these road categories is the road capacity and allowed driving speed. The official speed limits for these roads in China are 120km/h, 80km/h, and 50km/h, respectively; see Ma and Tang (2020).

a regional quantitative analysis that investigates the overall economic effects and various components thereof.

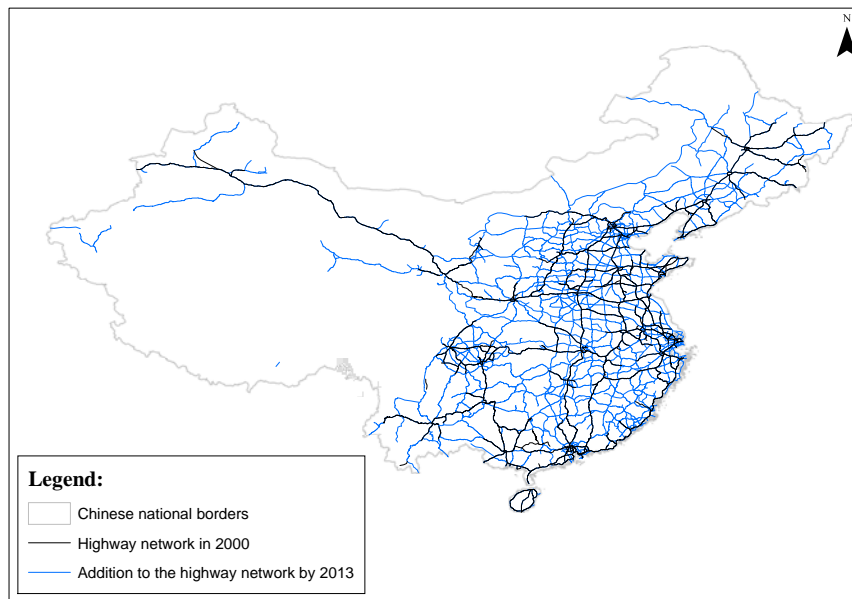
Figure 3.1: EVOLUTION OF UNNORMALIZED TRAVEL TIME BETWEEN ALL PREFECTURE PAIRS (2000-2013)



Between 2000 and 2013, the total road network length increased from 371,385 kilometers to 515,480 kilometers. Hence, by 2013 about 28% of all road-kilometers were connections which did not exist in 2000. Figure 3.1 shows the evolution of Chinese road connections for the period 2000 and 2013. In particular, it displays the sum of bilateral travel times between all 330 considered Chinese prefectures for each year between 2000 and 2013. We observe an almost linear decrease in the sum of bilateral travel times in the given period. This decrease is imputable to the extension of the highway network, but also to the densification of the regional road network. About one third (31.70%) of the decrease in the sum of bilateral travel times was achieved by developing the regional road network.

One noticeable improvement of the Chinese transport network between 2000 and 2013 is the concentration of the highway network. The total number of kilometers of highway (including the National Trunk Highway System) grew from 50,127 kilometers in 2000 to 142,983 kilometers in 2013. Figure 3.2 maps the evolution of the highway network between 2000 and 2013. Consistent with population-density patterns, most new highway segments were built in the eastern and northern parts of China.

Figure 3.2: CHINESE HIGHWAY NETWORK (2000 AND 2013)

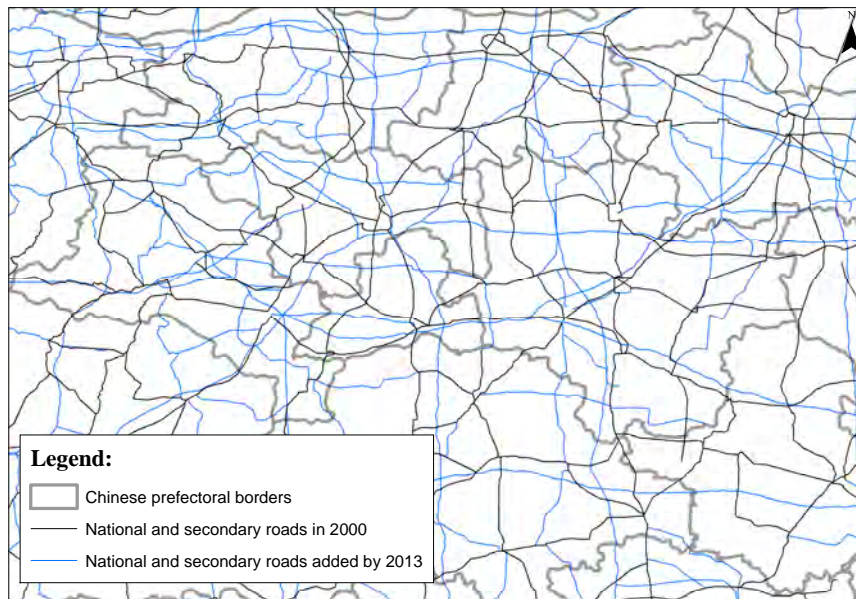


However, looking solely at the improvements of the highway network would not capture the full extent of the Chinese road network extension as already indicated in Figure 3.1. Figure 3.3 displays the evolution of the regional road network (i.e., the provincial-level and prefecture-level roads together), using gray lines for prefecture borders. For better display, we focus on a few prefectures in mid-eastern China here (namely, Pingdingshan, Luohe, and Zhoukou, all located in Henan Province). In 2013, Chinese provincial-level roads totaled 124,594 kilometers and prefecture-level roads amounted to 246,146 kilometers. Between 2000 and 2013, the total length of Chinese regional roads increased by 21%.

Whereas Figure 3.2 highlights the importance of the highway network improvements on a nation-wide scale, Figure 3.3 reveals that regional roads are crucial to accurately capture bilateral connections at the prefecture level. This is especially true for short distance travels and when it comes to providing shortest-distance access to the highway system. Prefectures are relatively small geographical units, and for analyzing the road network improvement effects on them – even when focusing on highways – it is elemental to also consider regional road network connections and their improvements.

Finally, Figure 3.4 displays the geography of connectivity growth between 2000 and 2013 in terms of aggregate inverse travel times in a map for each prefecture

Figure 3.3: CHINESE REGIONAL ROAD NETWORK (2000 AND 2013)



*Notes:* Regional roads include both provincial-level and prefecture-level roads. The figure zooms in on the prefectures of Pingdingshan, Luohe, and Zhoukou, all in Henan Province.

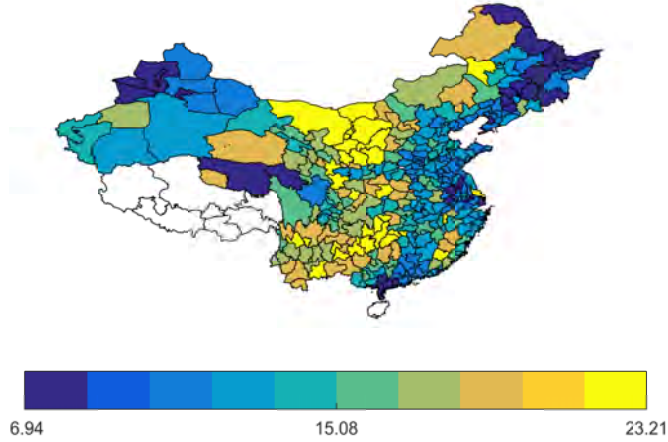
(a) and in a graph by longitude bins of  $5^\circ$  (b). It suggests that prefectures in mid-eastern China (at a longitude of around  $105^\circ$  E) had the largest gains in connectivity. Moreover, prefectures in the Far West of China gained more than on average in road connectivity, mostly because they started from a very low connectivity level in 2000.

### 3.3.2 Optimal Transport Network

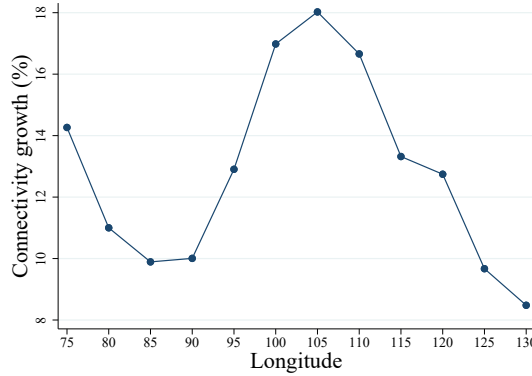
One key concern when analyzing transport networks is that connections between places are not random. Central planners decide on the location of new road segments to maximize some objective function (e.g., connecting large centers, opening up isolated regions, etc.). The explicit modeling of this objective function is outside of the focus of the present paper. For estimating the impact of travel time on moving and trade costs, we employ an Instrumental Variable (IV) strategy to avoid a bias associated with endogenous road-transport infrastructure placements. Travel time in the observed network is instrumented with the solution of the classical Monge-Kantorovich optimal-transport problem as proposed in Monge (1781) and Kantorovitch (1958). This approach has several advantages. First, it

Figure 3.4: THE GEOGRAPHY OF CONNECTIVITY GROWTH (2000-2013)

(a) Prefecture-level Connectivity Growth



(b) Connectivity Growth by Longitude



Note: Connectivity growth is measured as  $\frac{\sum_{j=1}^{331} d_{ij}^{-1} - \sum_{j=1}^{331} d_{ij}^{-1}}{\sum_{j=1}^{331} d_{ij}^{-1}}$ , where  $d_{ijt}$  denotes the bilateral travel time in year  $t$ . Figure 3.4b displays the average connectivity growth by longitude bins of  $5^\circ$ .

predicts significantly more connections (or edges) between the nodes (here, prefectures) in a network than a minimum-spanning-tree approach. This can lead to a higher predictive power for the Monge-Kantorovich predicted network than for a minimum-spanning-tree network in the first stage, where the actual road network is explained. Second, it allows to predict an intensity-of-connection usage and, hence, a connection typology (here, of highways, provincial-level, and prefecture-level roads), which would not be the case for the minimum-spanning-tree network.

We derive the optimal transport network in three steps. First, using a grid of 4,000,000 cells over China, we predict the cost of building a road in each cell using observed geographical, historical and year-2007 network data (this is the year in the middle of our study period). Second, using Dijkstra's (1959) algorithm,

we compute a path of minimum costs between any two Chinese prefectures. Finally, we select among all cost-minimizing paths the ones that solve the Monge-Kantorovich optimal-transport problem using historical population data. In the remainder of this section, we outline each step of deriving the optimal transport network in more detail.

**Step 1: Grid and Estimation of Construction Costs** — We build a grid of  $2,000 \times 2,000 = 4,000,000$  cells covering all of mainland China. This leads to cells of about  $3.1 \times 3.1$  kilometers in size. To each cell, we attribute the average slope (land gradient) in the cell, the area of the cell covered by water, and the population density of the prefecture in which the cell lies according to 1953 population data.<sup>21</sup> Note that these three variables are not affected by the construction of new roads in 2007. We also defined an indicator variable which equals unity, if a cell does *not* contain (is not crossed by) a road connection in 2007, and zero otherwise. We use the latter as a binary dependent variable in a cross-sectional probit model, where the underlying latent continuous variable should be interpreted as to reflect the costs of building roads in a cell. The index which informs the propensity of road construction is specified as a linear function of a constant and the three aforementioned variables (the average slope, the extent of water areas, and the population density of 1953).

**Step 2: Computing the Minimum Costs of Building a Connection Between Each Prefecture Pair** — Based on the construction costs obtained from the previous step, we determine the minimum-cost path between any two prefectures using Dijkstra's (1959) algorithm. This algorithm searches for the cost-minimizing path by sequentially adding legs (here, cells) to the best current path and comparing it to all alternative paths. Searching the full space for each bilateral connection requires an inefficiently large amount of time as extreme routes are unlikely to be cost minimizing. Hence, we restrict the search space to all cells within a  $7^\circ$  (about 750km) bandwidth around the great-circle-distance connec-

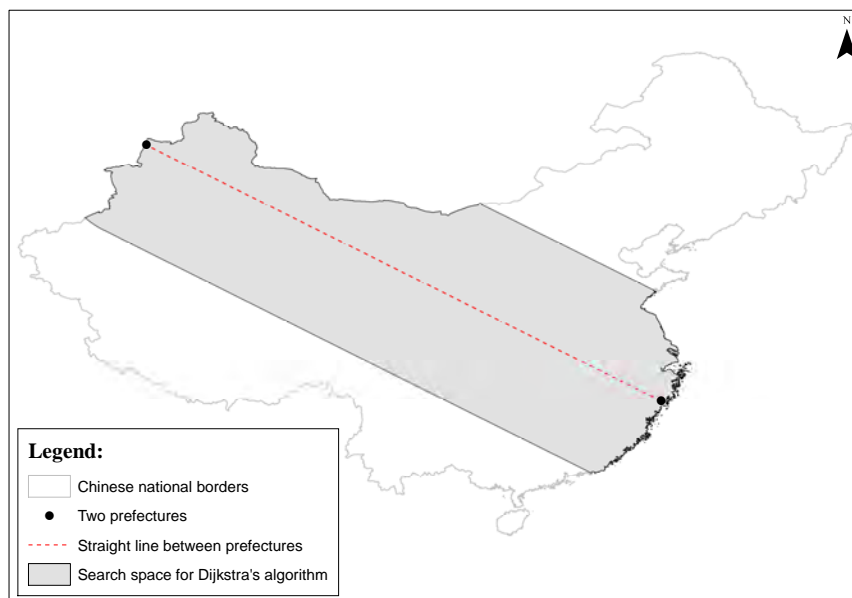
---

<sup>21</sup>The 1953 Population Census for China is available in pdf format and contains information on population counts at the provincial and county levels. We digitalized the data using optical character recognition (OCR) methods. Then, we aggregated all county-level population data to the prefecture level using prefecture borders as of the year 2000.



tion between any two prefectures. Figure 3.5 illustrates the search space for an exemplary pair of prefectures.

Figure 3.5: ILLUSTRATION OF THE SEARCH SPACE FOR DIJKSTRA’S ALGORITHM



**Step 3: Selecting Connections** — In the last step, we select the connections to be included in the optimal network. Importantly, to obtain a valid instrument, it is crucial that the selection process does not depend on shocks and other time-specific components in variables (specifically, population density) which are measured during the period of investigation and, in particular, in 2000, which is the benchmark year for the state of the road infrastructure network (see Section 3.4 for more details). Solutions based on a cost-benefit analysis using contemporary data would typically not be suited. Moreover, given the relatively high density of the observed transport network and the relatively large number of prefectures, a minimum-spanning tree connecting all prefectures is likely to have low predictive power and underestimate the connectivity of prefectures during the period of investigation.

To obtain a valid instrument with a sufficiently high predictive power and a reasonable degree of predicted connectivity among prefectures, we formulate and solve the classical Monge-Kantorovich optimal-transport problem<sup>22</sup> proposed

<sup>22</sup>See Villani (2003), for a textbook treatment of the mathematical problem.

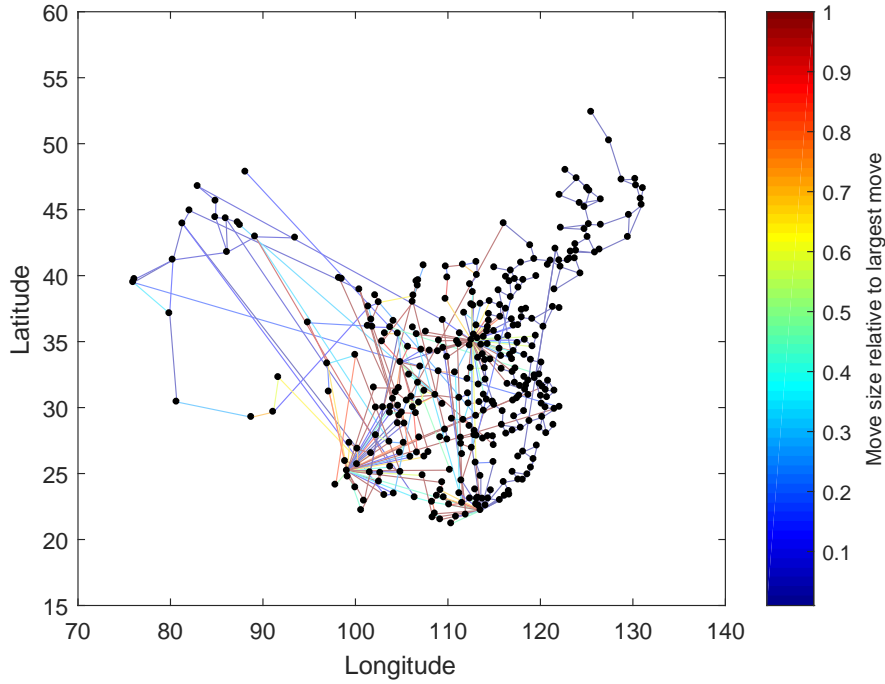
in Monge (1781) and Kantorovitch (1958) in the context of the Chinese road transport network. Specifically, we formulate the problem as a standard linear program in which *all* individuals as of 1953 must move to another prefecture at minimized overall moving costs. Costs between each pair of prefectures are given by the previous steps. Clearly, this leads to a denser network, when large prefectures are surrounded by smaller prefectures than when prefectures are of similar sizes. After defining,  $c_{ij}$  as the cost between prefectures  $i$  and  $j$  and  $h_{ij}$  as the population flow between them, and  $\mathbf{h}$  is the stacked vector of  $s(s-1)$  elements  $h_{ij}$  with  $i \neq j$ , the linear program can be formulated as follows:

$$\begin{aligned}
\min_{\mathbf{h}} f(\mathbf{h}) &= \sum_{i=1}^s \sum_{j \neq i} c_{ij} h_{ij} \\
\text{s.t. } \sum_i h_{ij} &= \sum_j h_{ij} \\
h_{ij} &\geq 0, \forall \{i \neq j\}.
\end{aligned} \tag{3.17}$$

To insure that the solution of the optimal-transport problem at hand does not lead to disconnected sub-networks within the big network of Chinese prefectures, we complement the respective solution by the minimum-spanning-tree connections. This is done using the cost-minimizing path from Dijkstra's (1959) algorithm and Kruskal's (1956) algorithm. Hence, all unconnected sub-networks (islands) based on the Monge-Kantorovich algorithm will be connected by minimum-spanning-tree least-cost road paths. The latter complement is similar to the strategy employed in Faber (2014).

The overall optimal-transport network is illustrated in Figure 3.6. Note that each connection is defined by a number of movers. This can be interpreted as a connection intensity and permits defining a road typology. Sorting connections by the number of movers, we assume the largest third to be highways, the middle third to be provincial-level roads and the bottom third to be prefecture-level roads. Assuming that movers travel at the speed limits of 120km/h on highways, 80km/h on provincial-level roads, and 50km/h on prefecture-level roads (see Ma and Tang, 2020), we obtain the projected travel time between any two prefectures based on the optimal network and the adopted assumptions. We use the resulting bilateral travel time as an instrument for the observed bilateral travel time.

Figure 3.6: OPTIMAL NETWORK USING THE MONGE–KANTOROVICH TRANSPORTATION PROBLEM



### 3.4 Calibration

In general, for calibrating the proposed model, one needs data for two consecutive years. In principal, one could of course calibrate the model to any sequence of consecutive year pairs  $t$  and  $t + 1$  for which data are available. Following the adopted convention, baseline road network data will be measured in 2000 and other variables associated with this baseline will be measured in the same year or the one thereafter, 2001. We choose 2000 for the baseline, because it is the earliest year in the period for which we collected road-transport-network data, and the network changed quite substantially up until more recent years. When calibrating the model, we will close the gap between key model variables and the corresponding data in 2000 and 2001. We will compute long-run equilibrium values which are informed by the exogenous factors as of 2000. These long-run equilibrium values will be reached only after the adjustment processes about amenities in (3.2) and the one about technology in (3.5) played out. In any case, for what follows in this section it should be borne in mind that whenever referring to the baseline year of the calibration we address 2000.

In the counterfactual analysis, we will shock China’s domestic road-transport network as of 2000 *ceteris paribus*. For this, we will impose actual road-connectivity data from 2013 in 2000. Hence, the outcomes realized in the baseline model in 2000 will adjust in response to this shock up until they stay constant between any pair of years when letting the adjustment process run forward. We will then compare the long-run-equilibrium vectors of outcome variables for all considered prefectures in China (and, in the background, the rest of the world, RoW) between the baseline road network as of 2000 and the one of 2013. Other exogenous factors are held constant at the values of the year 2000.

In the benchmark year 2000, China was composed of 349 prefecture-level regions. We use information on administrative boundaries of Chinese prefectures for that year from the China Data Center at University of Michigan.<sup>23</sup> Throughout the analysis, we hold prefecture-level borders constant as of the year 2000 and calibrate all variables accordingly. In the analysis, we focus on a sub-sample of 330 Chinese prefectures. With this strategy, we dropped 19 prefectures which include the following: ones that are not part of what is often called *mainland China* (i.e., Hongkong, Taiwan, and Macao); prefectures on Hainan Island (as there is no road connection to mainland China); and all prefectures in the Tibet Autonomous Region (mainly for data restrictions). Moreover, we merged prefectures to one greater-city prefecture within the following five cities, namely Beijing, Baoshan, Chongqing, Shanghai, and Tianjin. Overall, we observe that the land mass of prefectures tends to be larger in the West and smaller in the East, which clearly reflects the relatively greater population density in the East.

While the focus of this study is on Chinese prefectures, we decided to model China as a large open economy, given the importance of its foreign trade. However, there is no need for going into much regional detail with regard to the rest of the world for the present purpose. That means, apart from the 330 Chinese prefecture-level regions, our data include one additional region representing the rest of the world, which accounts for all the economic activity outside of the considered 330 regions. Hence, the total number of regions we cover is 331 in this paper. The rest of the world population count is obtained by subtracting the mainland Chinese

---

<sup>23</sup>Note that the Chinese Data Center website has been closed and is no longer available since September 2018. We acquired the data on Chinese administrative boundaries in November 2016.

population count in 2000 and 2001 from the world population in those years. To determine the per-capita wage in the rest of the world in 2000 and 2001, we use the following data moments. First of all, we use the prefecture-population-weighted average nominal wages to obtain a per-capita wage level for mainland China in 2000 and 2001. Second, we compute the ratio of the OECD to Chinese price level given OECD data.<sup>24</sup> We then multiply this ratio to the Chinese per-capita average wage and thereby back out the per-capita wage level for the rest of the world.

Table 3.1: CALIBRATION OVERVIEW

PARAMETER COMMON TO ALL LOCATIONS		
<b>1. Preferences &amp; Evolution of Amenities</b>		
$\sigma = 4$	Elasticity of substitution.	Bernard et al. (2003)
$\alpha = 0.850$	Consumption share in utility (non-land share).	Own estimation
$\lambda_1 = 0.690$	Elasticity of end. amenities w.r.t. pop. density.	Own estimation
$\lambda_2 = 0.494$	Elasticity of tomorrow's amenities w.r.t. today's aggregate amenities.	Own estimation
$\Omega = 0.5$	Elasticity of migration flows w.r.t. income.	Monte et al. (2018)
<b>2. Technology &amp; Evolution of Productivity</b>		
$\mu = 0.800$	Labor share in production (non-land share).	Own estimation
$\theta = 3.570$	Trade elasticity and dispersion of technology.	Own estimation
$\gamma_1 = 0.683$	Elasticity of technology w.r.t. pop. density.	Own estimation
$\gamma_2 = 0.728$	Elasticity of tomorrow's technology w.r.t. today's aggregate technology.	Own estimation
<b>3. Migration &amp; Trade Cost Elasticities</b>		
$\phi_1 = 0.055$	Elasticity of travel time to migration costs.	Own estimation
$\phi_2 = 0.044$	Elasticity of travel time to trade costs.	Own estimation
LOCATION-SPECIFIC PARAMETER		
<b>1. Land Endowments</b>		
$H_i^R$	Residential land mass in location $i$ .	Geofabrik, OSM
$H_i^C$	Commercial land mass in location $i$ .	Geofabrik, OSM
<b>2. Initial Distributions</b>		
$T_{it}$	Initial technology distribution.	Own estimation
$a_{it}$	Initial amenity distribution.	Own estimation
<b>3. Network Parameters &amp; Weighting Scalars</b>		
$d_{ijt}$	Bilateral travel time (Road)	Network Analyst, ArcGIS
$\mathbb{W}_{ijt}^R$	Weighting scalar governing amenity diffusion.	Own calibration
$\mathbb{W}_{ijt}^C$	Weighting scalar governing technology diffusion.	Own calibration
<b>4. Migration &amp; Trade Costs</b>		
$\kappa_{ijt}$	Migration costs	Own estimation
$\zeta_{ijt}$	Trade costs	Own estimation

*Note:* OSM refers to Open Street Map.

To compute the quantitative multi-region long-run general equilibrium for any baseline calibration, we need the parameters contained in the equations above and

<sup>24</sup>Source: <https://data.oecd.org/price/price-level-indices.htm>

summarized in Table 3.1. Apart from parameters that are common to all regions in the model, these are residential and commercial land endowments  $\{H_i^R, H_i^C\}$ , distributions of amenities and technology for years 2000 as well as 2001,  $\{T_{it}, a_{it}\}$  and  $\{T_{it+1}, a_{it+1}\}$ , network transmission (or spillover) weights captured by the scalars  $\{\mathbb{W}_{ijt}^R, \mathbb{W}_{ijt}^C\}$  measured in 2000 (and, in the counterfactual in 2013), as well as interregional trade- and migration-cost parameters  $\{\kappa_{ijt}, \zeta_{ijt}\}$ , also measured in 2000 (and 2013). Table 3.1 lists the parameters and gives a brief explanation of how they are assigned and chosen. Two of the parameter values are based on previous work, namely  $\{\sigma, \Omega\}$ , while all others are either estimated or calibrated by us. In what follows, we briefly describe our choices for the adopted parameter values in more detail.

### 3.4.1 Elasticity-of-substitution Parameter $\sigma$

We assume a value for the elasticity of substitution between varieties of  $\sigma = 4$  in the model. This value is motivated by earlier work. For instance, Bernard et al. (2003) report a value of  $\sigma = 3.8$ , and Broda and Weinstein (2006) an average one of  $\sigma = 4$  (based on price and expenditure data measured at the three-digit trade-classification product level).

### 3.4.2 Land Endowments $(H_i^R, H_i^C)$

The total area of each prefecture is measured using ArcGIS software. However, as our theoretical framework distinguishes between residential and commercial land markets, we need additional information on land use in each prefecture. This information is provided by the Open Street Map (OSM) GeoFabrik Database.<sup>25</sup> Land-use data for the rest of the world are obtained by generalizing the land-use patterns observed in the EU28 countries to all countries in the rest of the world (RoW). For this, we use data from the European Union buildings database.<sup>26</sup>

<sup>25</sup>The GeoFabrik database is updated every day and we extracted the information on November 11<sup>th</sup>, 2018. Hence, residential and commercial land endowments  $(H_i^R, H_i^C)$  are dated to the year 2018 and assumed constant for the entire analysis. <https://download.geofabrik.de/asia/china.html>.

<sup>26</sup><https://ec.europa.eu/energy/en/eu-buildings-database>

### 3.4.3 Infrastructure Network Data and Network (Spillover) Weights ( $d_{ijt}$ , $\mathbb{W}_{ijt}^R$ , $\mathbb{W}_{ijt}^C$ )

We use hand-collected and digitized information on the size and exact location of all highways, provincial-level roads, and prefecture-level roads in China, in this paper specifically for the years 2000 and 2013. Based on this comprehensive road network information, we derive the travel time on the network between any two prefectures  $i$  and  $j$  in hours for each year,  $d_{ijt}$ , using the Network Analyst from ArcGIS and customary speed limits.

For the purpose of measuring inter-prefectural distances, we define the location of prefecture  $i$  by the main city center as measured by the highest density within a prefecture using population-density data from the Socioeconomic Data and Application Center (SEDAC) for the year 2000.

Inter-prefectural travel time is obtained by assuming that commuters drive at the speed limits. Speed limits are 120km/h on highways, 80km/h on provincial-level roads, and 50km/h on prefecture-level roads (see, Ma and Tang, 2020). Accordingly, we assume that travel times are symmetric between regions  $i$  and  $j$ , whereby for any generic year  $d_{ij.} = d_{.ji.}$ . The travel time between each prefecture and the RoW is defined as the travel time on the network to the nearest of the six largest Chinese ports – Shanghai, Shenzhen, Ningbo, Guangzhou, Qingdao, and Tianjin – plus the travel time from that port to the centroid of the population-weighted RoW countries.<sup>27</sup> By this token, the connectivity to the RoW is impacted by Chinese road network improvements.

The cross-prefecture diffusion or accessibility of amenities and technology through the network are governed by the weights  $\mathbb{W}_{ijt}^R$  and  $\mathbb{W}_{ijt}^C$ , respectively. We conjecture that the diffusion decreases with distance and assume that  $\mathbb{W}_{ijt}^R$  and  $\mathbb{W}_{ijt}^C$  both are zero for intra-regional relations ( $i = j$ ). Otherwise, each element ( $i \neq j$ ) is inversely related to  $d_{ijt}$ . Moreover, we follow the literature on spatial and network econometrics and normalize both  $\mathbb{W}_{ijt}^R$  and  $\mathbb{W}_{ijt}^C$  (Kelejian and Prucha, 2010) such that the maximum row sum equals unity (see the stability condition in Section 3.2.6). To summarize, for any  $V \in \{R, C\}$  of the connectivity or spillover

---

<sup>27</sup>For the travel time from any Chinese port to the centroid of the population-weighted RoW we assume an average travel speed of 25km/h.

scalars  $\mathbb{W}_{ijt}^V$ , we obtain  $\mathbb{W}_{ijt}^V = 0$  for all  $i = j$  and  $\mathbb{W}_{ijt}^V = \frac{d_{ijt}^{-1}}{\max_i \sum_{i=1}^{331} d_{ijt}^{-1}}$  for  $i \neq j$ .

### 3.4.4 Cobb-Douglas Shares in Preferences and Production ( $\alpha, \mu$ )

We compute the consumption share in utility ( $\alpha$ ) and the labor share in production ( $\mu$ ) using data on residential and commercial land rents. Rearranging equilibrium land rents (3.13) gives

$$\alpha = 1 - \left[ \frac{r_{it}^R H_i^R}{w_{it} L_{it}} \right] \quad \text{and} \quad \mu = \left[ 1 + \left[ \frac{r_{it}^C H_i^C}{w_{it} L_{it}} \right] \right]^{-1}.$$

We measure employment ( $L_{it}$ ) by combining population data from the Socioeconomic Data and Application Center (SEDAC), time-varying labor-force participation rates from the International Labor Organization Statistics Database (ILO-STAT), and time-varying labor-force-in-population shares from the World Bank's World Development Indicators (WDI) for the years 1990-2015.<sup>28</sup> Hence, we obtain the employment measurement by multiplying the population count from SEDAC with the labor-force participation rate from ILOSTAT and the share of population in employment age (15-64) from WDI. Data on the average annual wage per capita ( $w_{it}$ ) in RMB for the years 2000-2009 come from the Chinese Annual Survey of Industrial Firms (CASIF) provided by the National Bureau of Statistics China. As these data are geo-referenced, we can compute average wages by prefecture for each year covered.

Residential and commercial land use data ( $H_i^R, H_i^C$ ) for all 330 prefectures and the RoW come from the sources described in Section 3.4.2. Residential land rents ( $r_{it}^R$ ) in RMB per square meter for 105 prefectures between 2008-2016 are provided by the Ministry of Natural Resources in China (Department of Land Use Management). Commercial land rents ( $r_{it}^C$ ) in RMB per square meter are available for 340 prefectures between 1992-2014 from the National Bureau of Statistics China. Both residential and commercial land-rent data represent purchasing prices and

---

<sup>28</sup>SEDAC provides gridded population data with an output resolution of 30 arc-seconds (corresponding to approximately  $1 \times 1$  kilometer at the equator) for a five-year interval between 1990-2015 for the whole world (including China). We fill in data for the missing years within the intervals using linear interpolation. The data from ILOSTAT and WDI vary by year and country and are assumed to be identical across regions within a country.



need to be converted to annual rental-price data. We assume that annual rental prices correspond to a hypothetical annual mortgage payment if one had to borrow the total purchasing price. The generic annual mortgage payment  $M$  is calculated using the formula  $M = 12 \times \left( P \frac{\varsigma(1+\varsigma)^n}{(1+\varsigma)^n - 1} \right)$ , where  $P$  is the generic total purchasing price,  $\varsigma$  is the monthly interest rate, and  $n$  is the total tenure of the mortgage, which we assume to be  $n = 240$  months (or 20 years) for all regions and years. We use data on annual interest rates for China from the World Bank database.

Adopting this procedure and using the mentioned data obtains the parameters  $\alpha = 0.850$  and  $\mu = 0.800$  as listed in Table 3.1.

### 3.4.5 Migration-cost Parameters ( $\Omega$ , $\kappa_{ijt}$ , $\phi_1$ )

We follow Desmet et al. (2018) and Allen and Donaldson (2018) and assume a common parameter about the Fréchet dispersion of stochastic migration costs,  $\Omega$ . The latter pertains to the heterogeneity in individuals' tastes for location choices (with higher values of  $\Omega$  indicating larger taste heterogeneity). More formally,  $\Omega$  describes the inverse of the elasticity of migration flows to location-specific utility levels (net bilateral migration costs), which becomes apparent in equation (3.10). We assume a value of  $\Omega$  on the basis of previous estimates that have been derived in the within-country context, i.e., without formal migration restrictions. We follow Desmet et al. (2018), who set a value of  $\Omega = 0.5$ , based on studies by Ortega and Peri (2016), Diamond (2016), Fajgelbaum et al. (2019), and Monte et al. (2018). This value is informed by data based on the European Union. Two remarks are in order with regard to that choice. First of all, assuming a common value for  $\Omega$  should appear less problematic here than in global regional studies because the regional entities considered in this study are – with the exception of one large entity capturing the RoW – all located within one country, China. Second, the residence choice in China is to some extent impeded by the household registration system called *hukou*. The *hukou* system was aimed at decreasing population pressure in large urban centers and Special Economic Zones (SEZs). However, at the same time China aims at increasing the share of the population living in urban centers (see, Egger et al., 2017). Between the 1970s and the mid-2000s, the *hukou* system has undergone numerous reforms towards greater flexibility. In any case, while

there is the *hukou* system, there are (almost) no cultural and only minor to no language barriers for internal migrants in China, unlike for internal migrants in the European Union. In view of these arguments, we follow earlier quantitative work in regional economics and use the parameter value  $\Omega = 0.5$ .

As is customary in the quantitative economic-geography literature, we model mobility frictions  $\kappa_{ijt}$  as an exponential function of bilateral travel times.  $\phi_1$  governs the translation of bilateral travel times into migration frictions.

$$\kappa_{ijt} = \exp(\phi_1 d_{ijt}). \quad (3.18)$$

Accordingly, there is a semi-log gravity equation for mobility flows between residence  $i$  and origin  $j$  in terms of travel time between  $i$  and  $j$  in year  $t$ :

$$\log \pi_{ijt}^R = -\xi d_{ijt} + a_{it} + b_{jt} + \epsilon_{ijt}, \quad (3.19)$$

where  $a_{it}$  are residence-region-time fixed effects capturing residence characteristics,  $b_{jt}$  are destination-region-time characteristics. The denominator of (3.9) is captured by  $b_{jt}$  as it is constant across residence regions  $i$ . The parameter  $\xi$  is the semi-elasticity of migration flows with respect to travel time. It is defined as  $\xi = \phi_1/\Omega$ , where  $\phi_1$  is the travel-cost parameter and  $\Omega = 0.5$  is the heterogeneity parameter associated with the Fréchet-distributed shock on individuals' utility.

Table 3.2: GRAVITY ESTIMATION RESULTS OF BILATERAL MIGRATION

Dep. Var.	$\log \pi_{ij}^R$ in 2000
$d_{ij}(= -\phi_1/\Omega)$	-0.109*** (0.003)
Estimation	2SLS
Fixed effects	Yes
Observations	10'547
$R^2$	0.21
Weak instr. (F statistic)	5,679

*Notes:* Standard errors in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. For one endogenous regressor, one instrumental variable and a maximum relative bias of 5%, the critical value for the weak instrumentation F-Statistic is 16.38 (Stock and Yogo, 2005).

To empirically estimate the semi-elasticities of migration flows, we measure bilateral migration shares from the Fifth National Population Census (2000). The prided data are a random sample of the census with a coverage of 0.95 per 1,000 people. Overall, there are 1,180,111 individuals covered. The data include, for each individual, their origin and current location at the prefecture level, and the year they moved, if they did. We consider all individuals in 2000 that have changed their residential location between 1995 and 2000 as migrants. This is the case for 10.5% of the sample. We aggregate all individuals to the prefecture level (for origin and destination) and get a bilateral measure of migration:  $L_{ijt}$ , for the year 2000. We instrument the travel time between two locations by the Monge-Kantorovich procedure-based travel time ( $d_{ij}^*$ ). Table 3.2 displays the estimation of (3.19) using these data. Given  $\Omega = 0.5$ , we obtain  $\phi_1 = 0.055$  based on these estimates.

### 3.4.6 Trade-cost Parameters ( $\theta$ , $\zeta_{ijt}$ , $\phi_2$ )

Chinese prefecture-to-prefecture trade data are not available, as it is the case in other data-sets used in quantitative multi-region general equilibrium work (see, e.g., Desmet et al., 2018). However, in contrast to many other countries, China reports prefecture-level trade aggregates with various regional aggregates within China and with foreign countries which can be used for parameter identification. In any case, to determine the trade-cost parameters, we cannot adopt the same strategy as for the migration-cost parameters. In this section, we propose a novel, model-guided approach which can be used with the available trade data.

The Investment Climate Survey (ICS) provides information on domestic and international trade volume for China. This survey is part of the 2005 World Bank Enterprise Survey, and it collects a wide range of information on 12,400 establishments in 123 prefectures in mainland China (for the year 2004). For instance, the survey reports the percentage of sales of an establishment (i) within the prefecture of its location, (ii) to other prefectures within the province of its location, (iii) to the rest of China (i.e., to all other provinces together), and (iv) to foreign countries (including Taiwan, Hongkong, and Macao). The ICS also offers information on the firms' performance from their income statements, which we use in order to aggregate sales of the sampled firms to the prefecture level.

Let us define  $\mathbf{p}^i$  as the set of all prefectures within the same province as prefecture  $i$ , and  $\mathbf{r}^i$  as the set of prefectures in China outside of  $i$ 's province, and let us use the index  $p^i$  to refer to this aggregate as an aggregate destination region. The share of aggregate purchases from firms in  $i$  in the total expenditures of the province that  $i$  is located in (denoted by  $p^i$ ) according to the model may be written as:

$$\pi_{ip^i t}^C = \frac{\int_{\mathbf{p}^i} T_{it}(o_{it})^{-\theta} \zeta_{ijt}^{-\theta} dj}{\int_{p^i} \int_S T_{kt}(o_{kt})^{-\theta} \zeta_{kjt}^{-\theta} dk dj}. \quad (3.20)$$

Analogously, let us define  $\mathbf{r}^i$  as the set of all Chinese prefectures not contained in (geographically, outside of the) set  $\mathbf{p}^i$ , and let us use the index  $r^i$  to refer to this aggregate as an aggregate destination region. Then, the share of aggregate purchases from firms in  $i$  in the total expenditures of other Chinese provinces than  $i$ 's is:

$$\pi_{ir^i t}^C = \frac{\int_{\mathbf{r}^i} T_{it}(o_{it})^{-\theta} \zeta_{ijt}^{-\theta} dj}{\int_{r^i} \int_S T_{kt}(o_{kt})^{-\theta} \zeta_{kjt}^{-\theta} dk dj}. \quad (3.21)$$

Based on the ICS data, we can measure both  $\pi_{ip^i t}^C$  and  $\pi_{ir^i t}^C$ , which we will use to obtain both  $\theta$  and  $\zeta_{ijt}$ . For identification of these parameters, it will be useful to define  $\mathfrak{M}_{it}$  as the Mahalanobis distance of a prefecture  $i$ . It measures how distant a prefecture  $i$  is to all others relative to the Chinese average. As such it involves the predicted row vector  $\hat{d}_{it}$  using the Monge-Kantorovich procedure-based travel time ( $d_{ij}^*$ ) for prefecture  $i$ . Furthermore, it depends on the  $s \times s$  variance-covariance matrix  $\Sigma_t$  of travel times among the Chinese prefectures. Then, the Mahalanobis distance for prefecture  $i$  based on year-2004 data is:

$$\mathfrak{M}_{i2004} = \sqrt{\hat{d}_{i2004} \Sigma_{2004}^{-1} \hat{d}'_{i2004}}. \quad (3.22)$$

**Trade Elasticity and Dispersion of Technology** — We first estimate the trade-elasticity and dispersion-of-technology parameter  $\theta$ . Consider two prefectures  $i$  and  $i'$  in the same province with similar Mahalanobis distance values  $\mathfrak{M}_{it} \cong \mathfrak{M}_{i't}$ . The latter means that these prefectures have approximatively the same trade costs to all other prefectures.

Thus,<sup>29</sup>

$$\frac{\pi_{ip^i t}^C}{\pi_{i'p^i t}^C} \cong \frac{T_{it}(o_{it})^{-\theta}}{T_{i't}(o_{i't})^{-\theta}}, \quad \text{if } \mathfrak{M}_{it} \cong \mathfrak{M}_{i't}. \quad (3.23)$$

Similarly, considering the expenditure shares on these two prefectures to other provinces than their own province, we have:

$$\frac{\pi_{ir^i t}^C}{\pi_{i'r^i t}^C} \cong \frac{T_{it}(o_{it})^{-\theta}}{T_{i't}(o_{i't})^{-\theta}}, \quad \text{if } \mathfrak{M}_{it} \cong \mathfrak{M}_{i't}. \quad (3.24)$$

Knowing that we can express  $o_i$  as a function of  $L_{it}$ ,  $w_{it}$ ,  $H_i^C$ , and  $\mu$  following (3.6), we can rewrite (3.23) and (3.24) for prefectures with  $\mathfrak{M}_{it} \cong \mathfrak{M}_{i't}$  by taking logs as:

$$\begin{aligned} \log \frac{\pi_{ip^i t}^C}{\pi_{i'p^i t}^C} &\cong -\theta \log \left( \frac{w_{it} \left( \frac{L_{it}}{H_i^C} \right)^{1-\mu}}{w_{i't} \left( \frac{L_{i't}}{H_{i'}^C} \right)^{1-\mu}} \right) + \varepsilon_{ii^i p^i t} \\ \log \frac{\pi_{ir^i t}^C}{\pi_{i'r^i t}^C} &\cong -\theta \log \left( \frac{w_{it} \left( \frac{L_{it}}{H_i^C} \right)^{1-\mu}}{w_{i't} \left( \frac{L_{i't}}{H_{i'}^C} \right)^{1-\mu}} \right) + \varepsilon_{ii^i r^i t}. \end{aligned} \quad (3.25)$$

Note that in equation (3.25) the two terms  $\varepsilon_{ii^i p^i t}$  and  $\varepsilon_{ii^i r^i t}$  are treated as residual terms which are functions of the endogenous technology terms  $T_{it}$  and  $T_{i't}$ . As outlined in Section 3.4.4, the vectors  $L_{it}$ ,  $w_{it}$ ,  $H_i^C$ , and  $\mu$  are observable. We estimate (3.25) using only those prefectures in a province, for which the absolute difference  $|\mathfrak{M}_{it} - \mathfrak{M}_{i't}|$  is minimal. We instrument the ratio of unit costs,  $o_{it}/o_{i't}$ , with the natural logarithm of the ratio of land areas<sup>30</sup> for prefectures  $i$  and  $i'$ . Adopting this procedure, we obtain a value of  $\theta = 3.570$ .

**Elasticity of Travel Times to Trade Costs** — Given the value of  $\theta$ , we employ a brute-search approach to determine the parameters of trade frictions  $\zeta_{ijt}$ . Similar to the migration frictions, we model  $\zeta_{ijt}$  as an exponential function of bilateral travel times,  $d_{ijt}$ .  $\phi_2$  governs the mapping of travel time into trade costs.

$$\zeta_{ijt} = \exp(\phi_2 d_{ijt}). \quad (3.26)$$

<sup>29</sup>The detailed derivation is presented in Section 3.10 of the Appendix.

<sup>30</sup>Land area is measured in squared kilometers. Note that the availability of land in a prefecture is relevant and exogenous in the model.

Consider the ratio of ratios of expenditure on two prefectures  $i$  and  $i'$  from the same and other provinces in China. Ensure again that  $i$  and  $i'$  belong to the same province but now use only ones for which  $\mathfrak{M}_{it} \neq \mathfrak{M}_{i't}$ . For such prefectures, this ratio of ratios is defined as:<sup>31</sup>

$$\frac{\pi_{ip^i t}^C \pi_{i'r^i t}^C}{\pi_{i'p^i t}^C \pi_{ir^i t}^C} = \frac{\int_{p^i} \zeta_{ijt}^{-\theta} dj \int_{t^i} \zeta_{i'jt}^{-\theta} dj}{\int_{p^i} \zeta_{i'jt}^{-\theta} dj \int_{t^i} \zeta_{ijt}^{-\theta} dj}. \quad (3.27)$$

Note that the left-hand-side variable of equation (3.27) is observed, and, regarding the right-hand side,  $\theta$  is known from above, and so is  $\zeta_{ijt}$  up to the scalar  $\phi_2$  in equation (3.26). Hence, we can minimize the sum of squared distances between the left-hand side and the right-hand side in equation (3.27). In fact, this procedure is best informed for those prefectures  $i$  and  $i'$ , for which the absolute difference  $|\mathfrak{M}_{it} - \mathfrak{M}_{i't}|$  takes on the maximum value within a province. Using only those prefectures in a province, for which the absolute difference  $|\mathfrak{M}_{it} - \mathfrak{M}_{i't}|$  is maximal, the proposed procedure obtains a parameter of  $\phi_2 = 0.044$ . From (3.26) and the observed  $d_{ijt}$ , we then obtain  $\zeta_{ijt}$ .

### 3.4.7 Parameters Governing the Distribution of Endogenous Technology ( $T_{it}$ , $\gamma_1$ , $\gamma_2$ )

Solving the model for a reference year  $t$  requires knowledge on the prevailing technology in that year,  $T_{it}$  for all regions  $i$ , and the respective parameters governing the technology-adjustment process,  $\gamma_1$  and  $\gamma_2$ . We start by numerically solving for the endogenous technology  $T_{it}$  through a standard contraction-mapping procedure, using the model structure and observed levels of population, wages, and land endowments for the benchmark year 2000. For this, we use the goods-market-clearing condition in (3.11) to solve for  $T_{it}$  as a function of observables, substituting the trade shares with the expression in equation (3.7) and the unit costs with the one in equation (3.6).

---

<sup>31</sup>The detailed derivation is presented in Section 3.10 of the Appendix.

Hence, we obtain

$$T_{it} = \frac{w_{it}^{1+\theta} L_{it}^{1+\theta(1-\mu)} H_i^{C\theta(\mu-1)}}{\int_S w_{jt} L_{jt} \zeta_{ijt}^{-\theta} \left( \int_S T_{kt} L_{kt}^{-\theta(1-\mu)} w_{kt}^{-\theta} H_k^{C-\theta(\mu-1)} \zeta_{kjt}^{-\theta} dk \right)^{-1} dj}. \quad (3.28)$$

To inform this system of equations, we use the parameter values for  $\{\mu, \theta, \zeta_{ijt}\}$  for all regions  $\{i, j, k\}$  in the year 2000 from the earlier subsections, population data from the Socioeconomic Data and Application Center (SEDAC), annual average wages per capita from the Chinese Annual Survey of Industrial Firms (CASIF), and commercial land endowments from the OSM GeoFabrik database.

Table 3.3: ENDOGENOUS TECHNOLOGY PARAMETER ESTIMATION RESULTS

<b>First Stage</b>		
<b>Dep. Var.</b>	$\log(L_{it+1}/H_i^C)$	$\int_S \mathbb{W}_{ijt}^C \log(T_{jt}) dj$
$\log(\text{Longitude}_i)$	4.817*** (0.708)	-0.849*** (0.135)
$\int_S \mathbb{W}_{ijt}^* \log(T_{jt}) dj$	-0.187 (0.116)	0.702*** (0.066)
<b>Second Stage</b>		
<b>Dep. Var.</b>	$\log(T_{it+1})$	
$\log(L_{it+1}/H_i^C)$	0.683*** (0.075)	
$\int_S \mathbb{W}_{ijt}^C \log(T_{jt}) dj$	0.728*** (0.157)	
Observations	1,324	
Weak Instr. (F-statistic)	43.71	

*Notes:* Robust standard errors in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. For two endogenous regressors, two instrumental variables and a maximum relative bias of 5%, the critical value for the weak instrumentation F-Statistic is 7.03 (Stock and Yogo, 2005).

Based on the respective model solutions for  $T_{it}$ , we can estimate  $\gamma_1$  and  $\gamma_2$  using the technology-evolution equation (3.5). Taking logs of (3.5) obtains

$$\log T_{it+1} = \gamma_1 \log \left( \frac{L_{it+1}}{H_i^C} \right) + \gamma_2 \left( \int_S \mathbb{W}_{ijt}^C \log T_{jt} dj \right) + \log \varepsilon_{it+1}^C, \quad (3.29)$$

where  $\log \varepsilon_{it+1}^C$  is the log of the technology shifter that includes a common constant. We estimate (3.29) by 2SLS, instrumenting  $\left( \int_S \mathbb{W}_{ijt}^C \log T_{jt} dj \right)$  and log population density  $\left( \frac{L_{it+1}}{H_i^C} \right)$  on the right-hand side. The included instruments are  $\left( \int_S \mathbb{W}_{ijt}^* \log T_{jt} dj \right)$  and the longitude coordinate of each location  $i$ .  $\mathbb{W}_{ijt}^*$  is based on the normalized, inverse Monge-Kantorovich procedure-based travel time ( $d_{ij}^*$ )

and otherwise obtained in the same way as  $\mathbb{W}_{ijt}^C$  in Section 3.4.3. We use  $\{T_{it}, T_{it+1}\}$  together with population data  $\{L_{it}, L_{it+1}\}$ , where  $t \in \{2000, 2001, 2002, 2003\}$  and  $t + 1 \in \{2001, 2002, 2003, 2004\}$ .<sup>32</sup> Adopting this procedure obtains a parameter value for  $\gamma_1 = 0.683$  and  $\gamma_2 = 0.728$ . Table 3.3 reports the results from estimating (3.29).

### 3.4.8 Parameters Governing the Distribution of Endogenous Amenities ( $a_{it}$ , $\lambda_1$ , $\lambda_2$ )

Endogenous amenities are determined by an adjustment-cost process akin to amenities. The process for amenities from  $t$  onwards depends on the initial (or baseline) distribution of amenities in year  $t$  as well as on the congestion parameter  $\lambda_1$  and the amenity diffusion parameter  $\lambda_2$ .

The adopted procedure to obtain these parameters is analogous to the one that pins down initial technology levels  $T_{it}$  and the technology parameters  $\gamma_1$  and  $\gamma_2$  in Section 3.4.7. First, we numerically solve for  $a_{it}$  through a contraction mapping approach using parameter values  $\{\alpha, \mu, \Omega, \theta, \kappa_{ijt}, \zeta_{ijt}, T_{it}\}$  for the year 2000 that are known at this point as well as observed levels of population, wages and land endowments for the same year.<sup>33</sup> To do so, we rewrite migration shares in (3.10), replacing the location-specific indirect utility with (3.3) and the price index with (3.8). Then, after some rearrangements,  $a_{it}$  can be expressed as

$$a_{it} = \frac{L_{it}^{\Omega+(1-\alpha)} \left(\bar{L}\tilde{\kappa}_{it}\right)^{-\Omega} w_{it}^{-\alpha} H_i^{R(\alpha-1)}}{\left[ \int_S \int_S \left( \frac{a_{kt} w_{kt}^\alpha H_k^{R(1-\alpha)}}{\kappa_{kjt} L_{kt}^{(1-\alpha)}} \right)^{\frac{1}{\Omega}} \left( \int_S \Pi_{kt} \zeta_{kjt}^{-\theta} dj \right)^{\frac{\alpha}{\theta\Omega}} dk dj \right]^{-\Omega} \left[ \int_S \Pi_{it} \zeta_{ijt}^{-\theta} dj \right]^{\frac{\alpha}{\theta}}}, \quad (3.30)$$

where  $\tilde{\kappa}_{it} = \int_S \kappa_{ijt}^{-1/\Omega} dj$  and  $\Pi_{kt} \equiv T_{kt} w_{kt}^{-\theta} (L_{kt} H_k^C)^{-\theta(1-\mu)}$ .

Having obtained a measure for  $a_{it}$  from this procedure, we can estimate  $\lambda_1$  and  $\lambda_2$  using the endogenous amenity equation (3.2). Taking logs of (3.2) obtains

$$\log a_{it+1} = -\lambda_1 \log \left( \frac{L_{it+1}}{H_i^R} \right) + \lambda_2 \left( \int_S \mathbb{W}_{ijt}^R \log a_{jt} dj \right) + \log \varepsilon_{it+1}^R, \quad (3.31)$$

<sup>32</sup> $\{T_{it}, T_{it+1}\}$  for  $t \in \{2001, 2002, 2003, 2004\}$  are readily obtained from solving the system in (3.28) in the same way as done for the year 2000 and based on data from the same sources.

<sup>33</sup>Observed levels of population, wages, and land endowments are taken from the same sources as described in Section 3.4.7.



Table 3.4: ENDOGENOUS AMENITY PARAMETER ESTIMATION RESULTS

<b>First Stage</b>		
<b>Dep. Var.</b>	$\log(L_{it+1}/H_i^R)$	$\int_S \mathbb{W}_{ijt}^R \log(a_{jt}) dj$
$\log(\text{Longitude}_i)$	2.704*** (0.358)	0.008 (0.005)
$\int_S \mathbb{W}_{ijt}^* \log(a_{jt}) dj$	0.633** (0.289)	1.257*** (0.045)
<b>Second Stage</b>		
<b>Dep. Var.</b>	$\log(a_{it+1})$	
$\log(L_{it+1}/H_i^R)$	-0.690*** (0.279)	
$\int_S \mathbb{W}_{ijt}^R \log(a_{jt}) dj$	0.494* (0.297)	
Observations	1,324	
Weak Instr. (F-statistic)	34.11	

*Notes:* Robust standard errors in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. For two endogenous regressors, two instrumental variables and a maximum relative bias of 5%, the critical value for the weak instrumentation F-Statistic is 7.03 (Stock and Yogo, 2005).

where  $\log \varepsilon_{it+1}^R$  is the log of the amenity shifter that includes a common constant. As for endogenous technology, we estimate (3.31) by 2SLS, instrumenting  $\left(\int_S \mathbb{W}_{ijt}^R \log a_{jt} dj\right)$  and  $\left(\frac{L_{it+1}}{H_i^R}\right)$  by  $\left(\int_S \mathbb{W}_{ijt}^* \log a_{jt} dj\right)$  and the longitude coordinate of each location  $i$ . We use  $\{a_{it}, a_{it+1}\}$  together with population data  $\{L_{it}, L_{it+1}\}$ , where  $t \in \{2000, 2001, 2002, 2003\}$  and  $t + 1 \in \{2001, 2002, 2003, 2004\}$ .<sup>34</sup> This procedure leads to parameter values of  $\lambda_1 = 0.690$  and  $\lambda_2 = 0.494$ . Table 3.4 summarizes the estimation results.

### 3.5 Counterfactual Analysis

In this section, we investigate the economic effects of the road infrastructure improvements that took place in China between 2000 and 2013. Given the parameters as determined in the previous section, we can simulate the model forward by solving the system of two equations (3.14) and (3.15).

The remainder of this section is organized as follows. We first present the main aspects of the initial state (Section 3.5.1), i.e., the level of trade and migration costs which are consistent with the road infrastructure of the year 2000. We report on some outcome variables – population, nominal wages relative to Bei-

<sup>34</sup>Again,  $\{a_{it}, a_{it+1}\}$  for  $t \in \{2001, 2002, 2003, 2004\}$  is readily obtained from solving the system in (3.30) in the same way as done for year 2000 and based on data from the same sources.

jing’s, amenities, and technology across Chinese prefectures – that are measured or estimated in the year 2000. In Section 3.5.2, we outline the counterfactual analysis and present evidence on the distribution of key long-run-equilibrium values across Chinese prefectures. These values include population, nominal wages relative to Beijing’s, amenities, and technology and are obtained when letting the model run based on the initial conditions and road infrastructure levels of the year 2000. In Section 3.5.3, we discuss the overall long-run effects on population and real income in the model when changing the road infrastructure in 2000 to 2013. We then decompose these effects by channel and road type. We first focus on the decomposition along four channels in Section 3.5.4 into: (1) increased accessibility of amenities, (2) increased technology diffusion, (3) reduced migration frictions, and (4) reduced trade frictions. Finally, we decompose the economic effects of transport infrastructure improvements by road type: (1) highway network improvements and (2) regional road network improvements in Section 3.5.5.

### 3.5.1 Changes in Trade and Migration Costs and Initial Model State

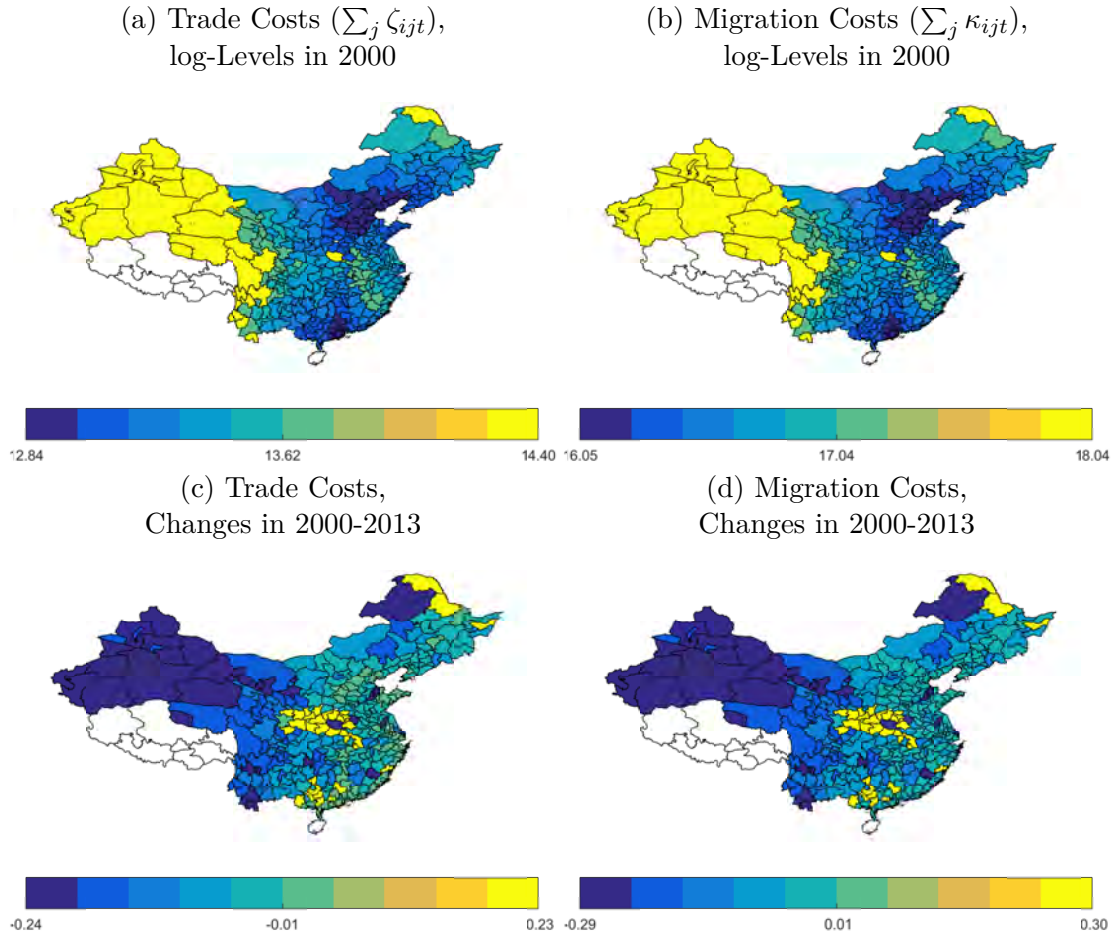
In Figure 3.7, we provide prefecture-level maps<sup>35</sup> to highlight the implicit trade and migration costs associated with the road infrastructure in the initial period in 2000 as well as their evolution between 2000 and 2013. In the top two panels of Figure 3.7, we illustrate the summed levels of trade costs (left panel) and migration costs (right panel) across prefectures. A larger value implies larger trade and migration costs.<sup>36</sup>

The top two panels should be viewed as to summarize the status quo of relative aggregate remoteness across the prefectures as of the year 2000. The two panels suggest that, not surprisingly, the domestic transport network was particularly dense – and associated trade and mobility costs were low – in what may be called central-eastern China (the blue areas in the two panels at the top of Figure 3.7). This is particularly true for the greater region around Beijing and the provinces surrounding it. Moreover, frictions were particularly low in the south

<sup>35</sup>The prefecture-level maps use the delineation of prefectures as of the year 2000.

<sup>36</sup>To illustrate both larger values on the extreme and smaller values within the distribution, we plot the average value within decile for each prefecture.

Figure 3.7: PREFECTURE-LEVEL TRADE AND MIGRATION COSTS



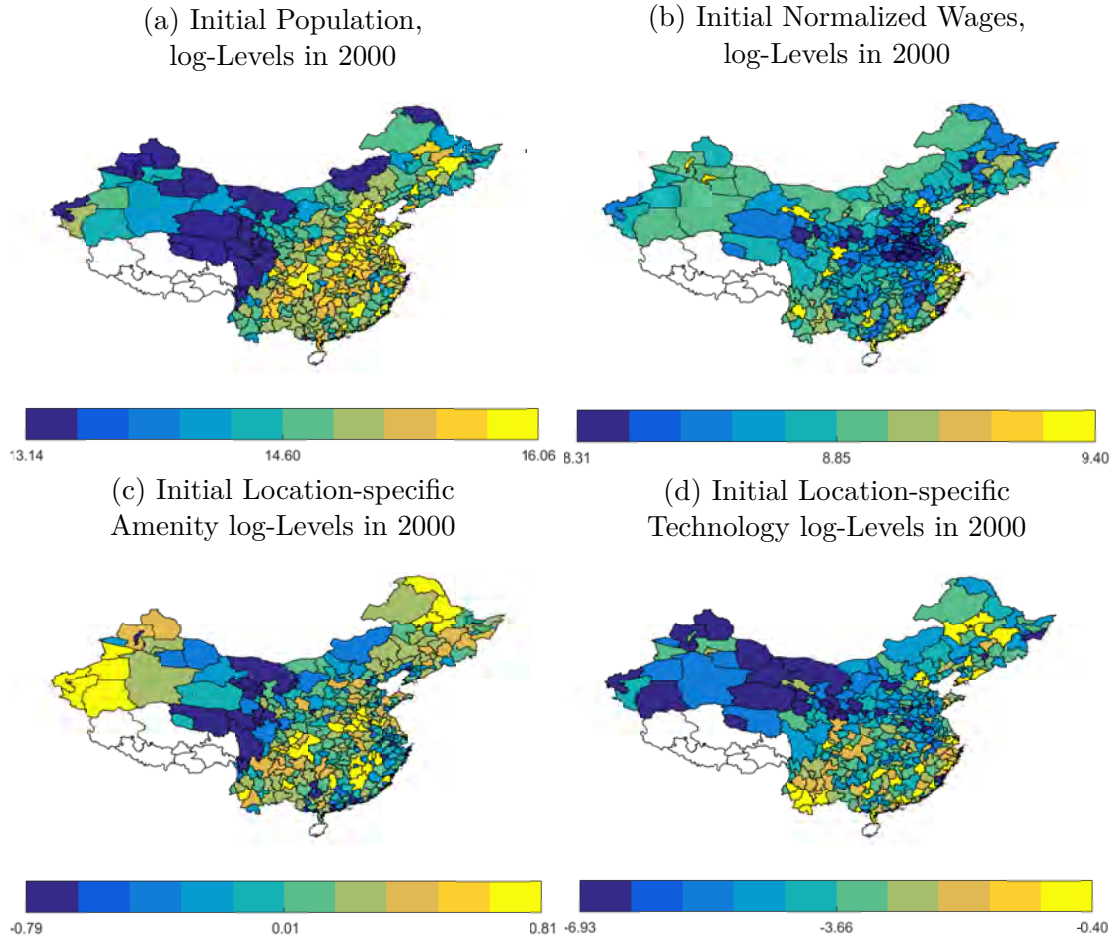
*Notes:* White prefectures are omitted for network accessibility reasons, as indicated in Section 3.4.

Location-specific migration and trade costs are defined as the sum of associated frictions over all destinations.

in the bigger area of the Pearl River Basin around Guangdong province. In the bottom two panels of Figure 3.7, we report on changes in these remoteness variables due to the road infrastructure changes between the years 2000 and 2013. The figure suggests that large reductions in trade and migration costs happened especially in peripheral areas which are remote from the coast. The western and northern prefectures (along the Russian border), experienced a particularly important reduction in trade and migration costs. Moreover, some few prefectures in central-eastern China without direct access to the coast experienced an increase in connectivity.

In Figure 3.8, we report on the distribution of the population, normalized wages, location-specific amenities and location-specific technology in 2000. Normalized wages are expressed relative to the wages in Beijing. Clearly, these two

Figure 3.8: KEY VARIABLES IN THE INITIAL STATE



Notes: White prefectures are omitted for network accessibility reasons, as indicated in Section 3.4.

maps illustrate that the population was dense in central-eastern China, especially, in the greater region around Beijing and bordering provinces. Wages were high along the coast and in central prefectures with little population. Initial location-specific amenities were particularly high in remote and less connected prefectures, whereas initial location-specific technology was high along the coast and in large urban centers, such as Beijing or Shanghai.

### 3.5.2 Simulating Benchmark and Counterfactual Equilibrium Outcomes

For the counterfactual analysis, we start with simulating the model forward to obtain long-run vectors of the endogenous variables of interest – population and real income – based on the initial conditions established for the year 2000 and

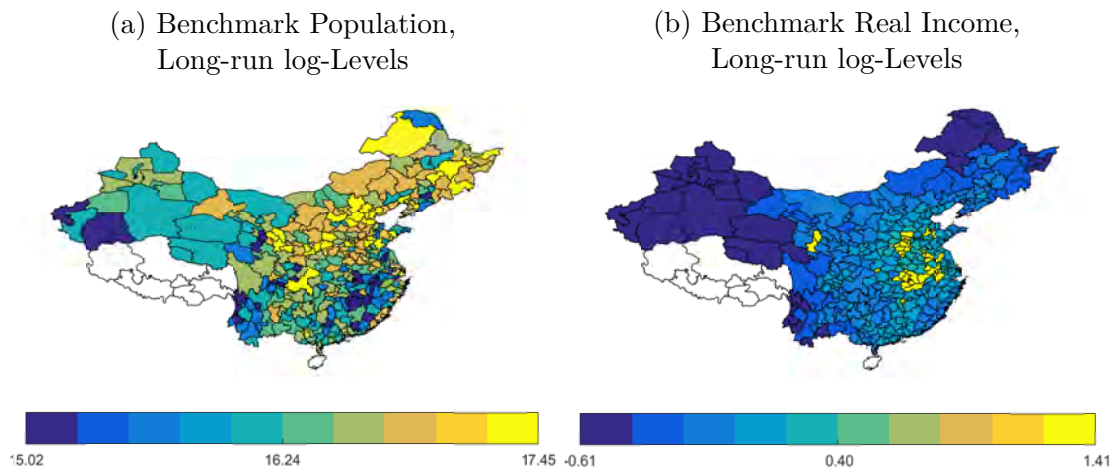
keeping road infrastructure values constant at that year’s level. Population and real income summarize the key mechanisms at play given our framework. All other effects can be derived from these.

The lagged weighting scalars on amenities and technology on the right-hand-side of equations (3.2) and (3.5) are kept at their levels of the year 2000 for all years thereafter, and trade and migration costs will also stay constant at their levels from 2000 onwards. Then, technology and amenities, both of which depend on past-year values, and also the population and prices (normalized wages but also output and rental prices) will be updated in periods from the year 2000 onwards. We apply a standard contraction-mapping procedure to solve for equilibrium wages and employment levels jointly using (3.14) and (3.15).

After updating the weighted technology and amenity levels on the right-hand-side of equations (3.2) and (3.5) as lagged values for the next period, the simulation runs forward for  $P$  time periods until it reaches a long-run general equilibrium. How many time periods it needs to reach this equilibrium depends on the distance between  $\{L_{it-1}, L_{it}\}$  and  $\{w_{it-1}, w_{it}\}$  for all  $\{i, t\}$ . We assume that as soon as the maximum absolute difference for either equilibrium outcome is lower than  $1e-03$ , the system has reached an equilibrium state  $\{L_{it}^*, w_{it}^*\}$ . Knowing the equilibrium distribution of employment and wages, the one of technology and amenity levels,  $T_{it}^*$  and  $a_{it}^*$ , is also determined. One can then solve for the location-specific equilibrium utility  $\{\tilde{u}_{it}^*\}$  through a contraction mapping using (3.10). From there, we determine equilibrium real income  $y_{it}^* = \tilde{u}_{it}^*/a_{it}^* = \frac{w_{it}^*}{P_{it}^\alpha r_{it}^{1-\alpha}}$ . With the mentioned stopping criteria, we reach the baseline long-run equilibrium after  $P = 40$  periods. We illustrate the initial long-run equilibrium values of the mentioned outcomes by way of Figure 3.9. We observe a decentralization pattern for population, especially in favor of northern prefectures. Real income, however, concentrates in central prefectures slightly of the coast between Beijing and Shanghai.

Regarding the counterfactual equilibrium, the lagged weighting scalar for amenities and technology on the right-hand-side of equations (3.2) and (3.5) will be updated and kept constant at their levels of the year 2013, and also trade and migration costs will be kept constant at the corresponding levels. The annual and long-run equilibrium values of all endogenous variables are then determined in the

Figure 3.9: KEY VARIABLES IN THE LONG-RUN EQUILIBRIUM CONSISTENT WITH YEAR-2000 ROAD INFRASTRUCTURE



*Note:* White prefectures are omitted for network accessibility reasons, as indicated in Section 3.4.

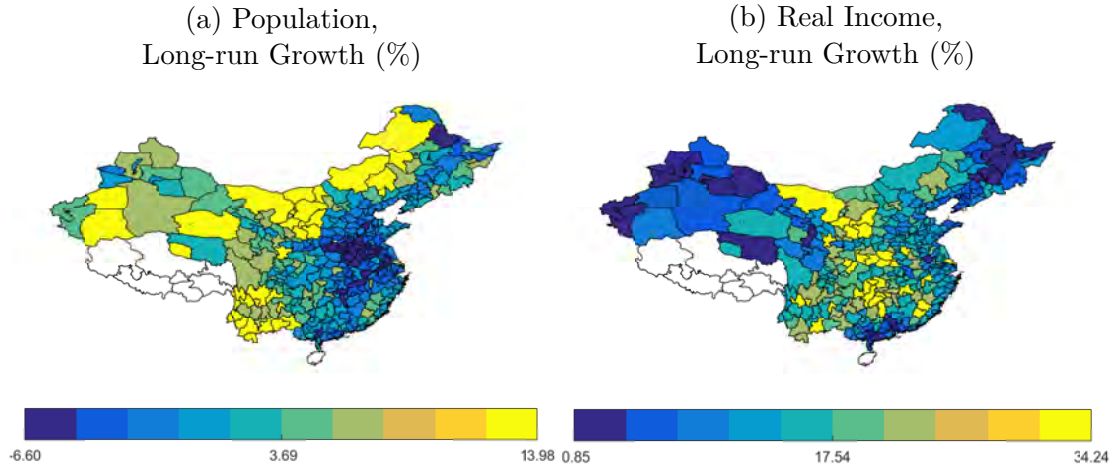
same way as for the benchmark equilibrium path. We achieve the counterfactual long-run equilibrium based on the chosen stopping criteria after  $P = 37$  periods.

### 3.5.3 Long-run Economic Effects of a Counterfactual Change in the Infrastructure Between Years 2000 and 2013

To assess the overall economic effects of China’s road network improvements between 2000 and 2013, we use the long-run equilibrium with the road network as of year-2013 versus year-2000 data. We illustrate the overall changes in the two outcome variables of particular interest – population and real income – in Figure 3.10 by way of maps.

We illustrate the changes across prefectures of mainland China. All changes of variables are defined for any variable as the difference between counterfactual and benchmark divided by the benchmark value (i.e., as a growth rate). Hence, larger values imply larger changes. It suggests that peripheral prefectures (along the Mongolian border and around Tibet) particularly benefited in terms of population. Their population growth comes at the expense of coastal or near-coastal prefectures. The largest reductions in the sum of migration frictions, displayed in Figure 3.7, appear to explain the largest population growth. Real income growth

Figure 3.10: CHANGES IN LONG-RUN EQUILIBRIUM VARIABLES (2000-2013)



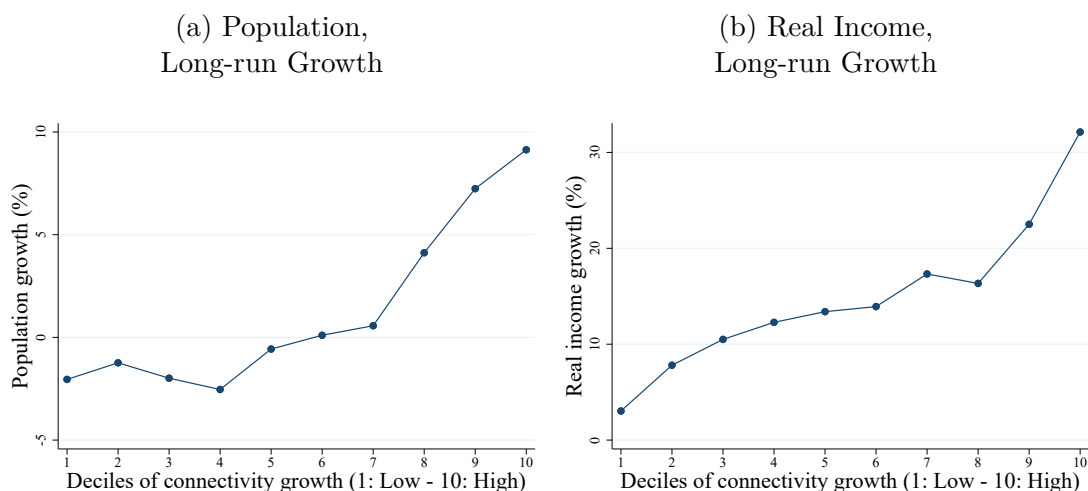
Note: White prefectures are omitted for network accessibility reasons, as outlined in Section 3.4.

appears particularly large for central prefectures on the longitude line south of Mongolia ( $105^\circ$  E). These prefectures benefited from important gains in connectivity, while enjoying a strategic geographical location (i.e., being a mid-way between China's large centers and more peripheral regions in the Far West).

Figure 3.11 shows how population and real income grew as a function of road connectivity improvements. Such improvements are measured in a straightforward manner by taking, for every prefecture, the inverse of the sum of travel time to all other prefectures. Formally, connectivity growth is measured as  $\frac{\sum_{j=1}^{331} d_{ij2013}^{-1} - \sum_{j=1}^{331} d_{ij2000}^{-1}}{\sum_{j=1}^{331} d_{ij2000}^{-1}}$  here. The horizontal axis of Figure 3.11 displays deciles of connectivity growth between 2000 and 2013, with the tenth decile referring to the largest gains in connectivity. Naturally, Figure 3.11 reveals the same patterns as Figure 3.10. A larger connectivity growth leads to a larger population growth (up to 9%), and an increase in real income (up to 32%). Note, however, that the increase in population only concerns the top-half of the prefectures in terms of connectivity growth. Individuals relocate from the bottom-half and the rest of the world (i.e., historically large centers) towards prefectures with improved connectivity (i.e., historically small and mid-sized prefectures).

In Table 3.5, we consider the overall effects of the transport network improvements on the long-run regional convergence of the prefectures in terms of the considered outcome variables. To compare convergence between the baseline and the counterfactual appropriately given normalizations, we look at the difference in

Figure 3.11: CHANGE IN LONG-RUN EQUILIBRIUM VARIABLES BY CONNECTIVITY GROWTH



the Herfindhal-Hirschmann index ( $hhi$ ) between the counterfactual and the baseline long-run equilibria. Formally, denoting the counterfactual (baseline) variables with a subscript  $c$  ( $b$ ) and taking population as an example, the  $hhi$  index is defined as:

$$hhi^L = \sum_i \left( \frac{L_i^c}{\sum_i L_i^c} \right)^2 - \sum_i \left( \frac{L_i^b}{\sum_i L_i^b} \right)^2. \quad (3.32)$$

Consequently, a negative value implies a smaller concentration in the counterfactual than in the baseline scenario; hence, convergence. Overall, the road network improvements between 2000 and 2013 appear to have led prefectures to converge in population levels and real income. Convergence in real income is particularly important. These findings are in line with the patterns observed in Figure 3.11.

Table 3.5: LONG-RUN REGIONAL CONVERGENCE

	(1) Population	(2) Real Income
$hhi$	-0.03	-6.75

Notes: A negative value indicates convergence.  
 $hhi \times 10^3$  is reported.



### 3.5.4 Decomposing the Long-run Equilibrium Effects by Channel

The framework in this paper models the impact of transport infrastructure along four channels: (i) increased accessibility of (or spillovers in) amenities, (ii) increased technology diffusion, (iii) reduced migration frictions, and (iv) reduced trade frictions. Each channel affects the long-run equilibrium in its own way. For instance, if prefectures were symmetric in terms of all exogenous fundamentals, an equalization of mobility would lead individuals to locate such that prefectures are equally large. This would also maximize trade flows. Conversely, making migration costs extremely asymmetric would lead individuals to relatively concentrate in the prefecture with the lowest cost. This would reduce the demand and need for trade, as fewer people elsewhere would need to be served with goods from the agglomeration. Similar arguments can be made for amenities and technology. For instance, making amenities and technology elsewhere better accessible reduces the importance of workers to locate, where local amenities and where expected local productivity draws are the best. Hence, all of these channels affect the attractiveness of locations on their own, and they matter for long-run equilibrium outcome across prefectures in an interdependent way.

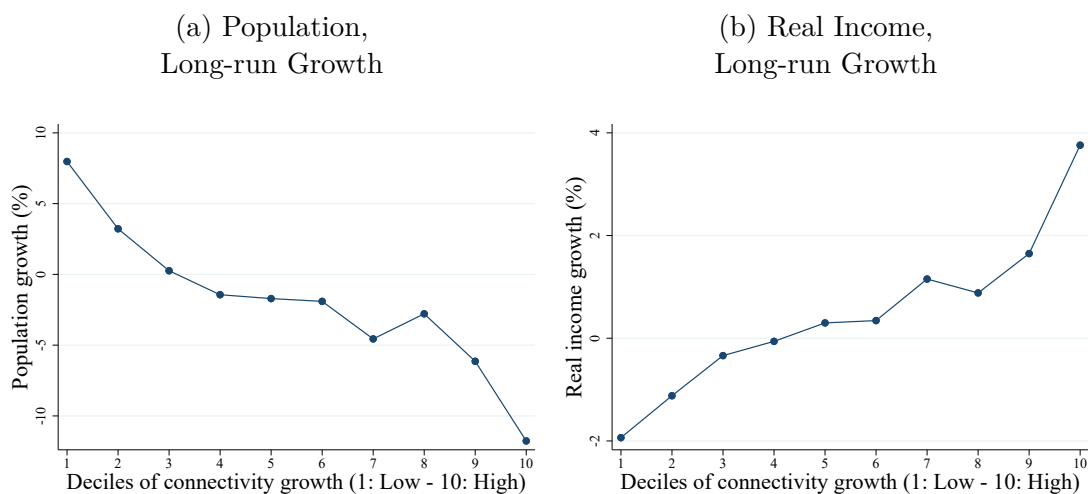
To disentangle the relative importance of each channel through China's road network expansion between 2000 and 2013, we run four different counterfactual exercises. In each of them, we replace the 2000 transport network by the 2013 network as before, but now *only through the channel under investigation*. Hence, while the change in road infrastructure alters the accessibility of amenities, the diffusion of technology, the migration costs of individuals, and the transport costs of goods simultaneously, we will consider only one of those changes at a time in this subsection. This will give us an insight into the qualitative as well as the quantitative differences of the effects through the four channels. Clearly, since the model structure is nonlinear, the sum of the individual effects across the four channels will principally add up to more (with a positive correlation of the channel-specific effects) or less (with a negative correlation) than the total. However, the generic insights are not fundamentally affected by this feature.<sup>37</sup>

---

<sup>37</sup>By way of analogy, in nonlinear regression work, one would decompose the overall explana-

**Channel 1: Diffusion of Amenities** ( $\mathbb{W}_{ij,2000}^R \rightarrow \mathbb{W}_{ij,2013}^R$ ) — Many factors determine the long-run level of local amenities ( $a_{it}$ ) as they depend on local population density and inverse-distance-weighted lagged amenities of other prefectures. Hence, convergence to the long-term level is a sluggish adjustment process. That being said, allowing only the diffusion of amenities to adjust to the considered road network change provides interesting insights. Akin to Figure 3.11, Figure 3.12 shows population growth, and real income growth in the long-run with improved diffusion of amenities only. The figure suggests that places with the largest connectivity growth lose the most in population (about -10%). As access to these places increases, it becomes less important to locate precisely there relative to nearby places. As population declines in places with the largest connectivity growth, real income follows a reverse trend due to reduced congestion. In these prefectures, real income grows by almost 4%.

Figure 3.12: LONG-RUN EQUILIBRIUM WITH DIFFUSION OF AMENITIES



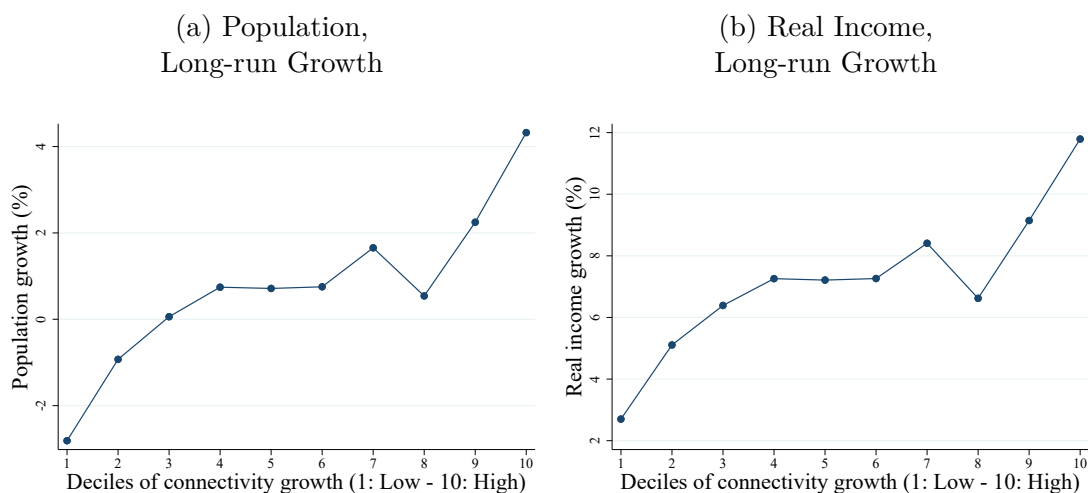
**Channel 2: Diffusion of Technology** ( $\mathbb{W}_{ij,2000}^C \rightarrow \mathbb{W}_{ij,2013}^C$ ) — Figure 3.13 presents the changes in population and real income that are triggered by the increased technology diffusion in the quantitative model. According to the figure, prefectures enjoying the largest connectivity growth benefit the most from the dif-

---

tory power of all variables in a model together by using the contributions of component-specific explanatory power to decompose the overall level of explanatory power. For instance, the variance contribution to the overall explanatory power could be approximated by the share of contributed variance by one component in the sum of all individual components, where the latter is normalized to the joint level of explanatory power.

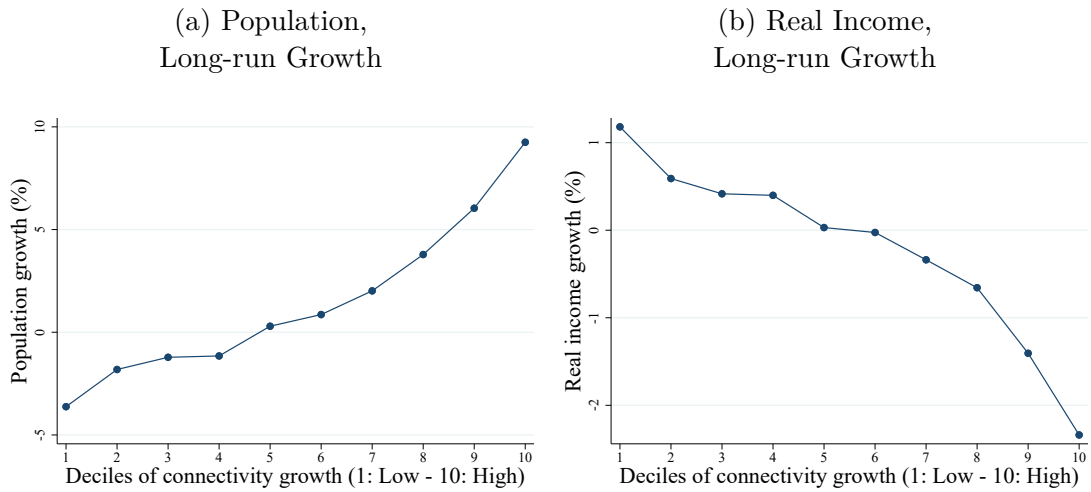
fusion of technology. This is especially the case, since the road network improvements benefited most mid-eastern prefectures, which had lower initial productivity levels than large urban centers such as Beijing or Shanghai. In response, population moves towards better-connected places (with a maximum growth of 4% for the top decile). Only the prefectures in the bottom-two deciles of connectivity growth face a decline in population, indicating that China’s road-infrastructure-improvement-induced productivity growth particularly benefits China relative to the rest of the world. The gains in connectivity are also associated with gains in real income across the board (from 3% and up to 12% for the top decile).

Figure 3.13: LONG-RUN EQUILIBRIUM WITH DIFFUSION OF TECHNOLOGY



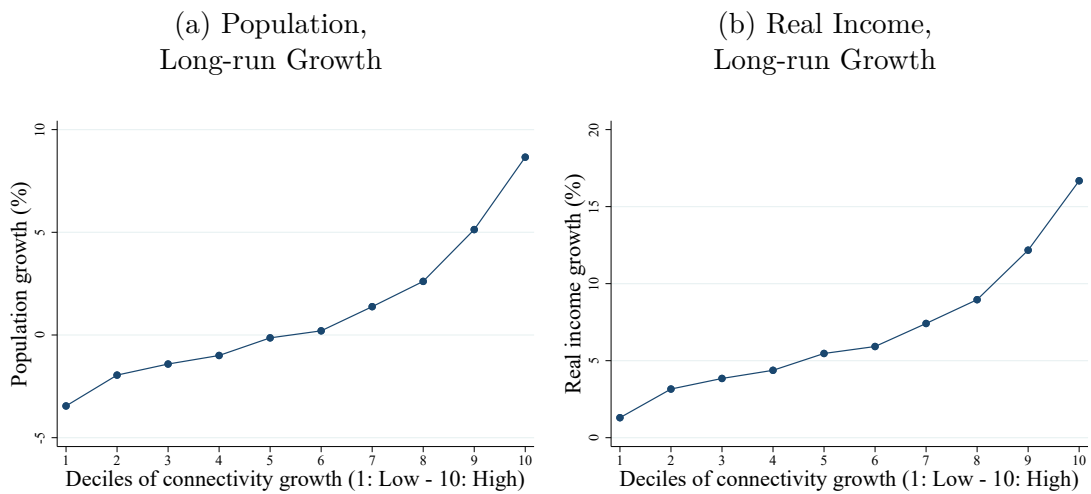
**Channel 3: Reduced Migration Frictions ( $\kappa_{ij,2000} \rightarrow \kappa_{ij,2013}$ )** — After isolating the economic effects of amenity and technology diffusion, we turn our attention to the effects of reduced migration and trade frictions. We start by examining the case of reduced migration frictions alone. Figure 3.14 presents the changes in population and real income that are triggered by changes in migration costs alone. Apparently, the effects are opposite to the ones of amenity diffusion. Population increases in the top six deciles of connectivity growth at the expense of the other Chinese prefectures and of the rest of the world. Population growth is larger in prefectures that experience a larger connectivity growth (up to 9%). Due to the increasing pressure on the local housing market which is a consequence of population growth and relatively less gains in productivity, real income drops (but by a smaller magnitude, down to -2%).

Figure 3.14: LONG-RUN EQUILIBRIUM WITH REDUCED MIGRATION FRICTIONS



**Channel 4: Reduced Goods-trade Frictions** ( $\zeta_{ij,2000} \rightarrow \zeta_{ij,2013}$ ) — Finally, we analyze the impact of reduced trade frictions alone that are consistent with connectivity changes in China between the years 2000 and 2013. We summarize the associated effects by way of Figure 3.15. Population grows in the top half of the prefectures in terms of connectivity growth, with a maximum of 9%. In prefectures with the largest connectivity growth, real income grows by about 17%. Even in prefectures with lower connectivity growth and negative population growth, real income increases as congestion costs decline mostly via a reduced demand on the housing market.

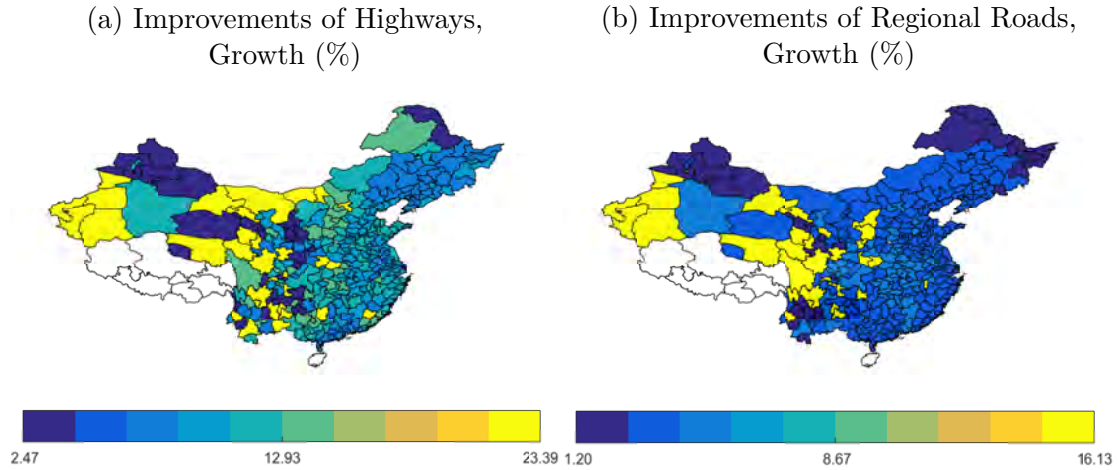
Figure 3.15: LONG-RUN EQUILIBRIUM WITH REDUCED GOODS-TRADE FRICTIONS



### 3.5.5 Decomposing the Long-run Equilibrium Effects by Road Type

So far, we have considered road network changes at any level of roads, jointly. As outlined above, the Chinese road network can be characterized by at least two layers, and the improvements between these layers had been heterogeneous between the years 2000 and 2013. Highways are particularly relevant to connect large centers over long distances. Regional roads connect all centers, large and small, as directly as possible, and they establish an access to the highway system. Figure 3.16 displays the deciles of connectivity growth for highways and regional roads between 2000 and 2013 separately. Improvements of the highway network led to large connectivity gains along the Mongolian border as well as along the 105° E longitude line (i.e., south of Central Mongolia). Improvements of the regional (non-highway) roads led to a more smooth picture of connectivity growth. This is particularly true in central China, off the coastal regions. Both types of network improvements have benefited the far-western prefectures. Overall, given these simple observations, it appears reasonable to expect that each road type affects the long-run equilibrium in its own way.

Figure 3.16: CONNECTIVITY GROWTH BY ROAD TYPE (2000-2013)

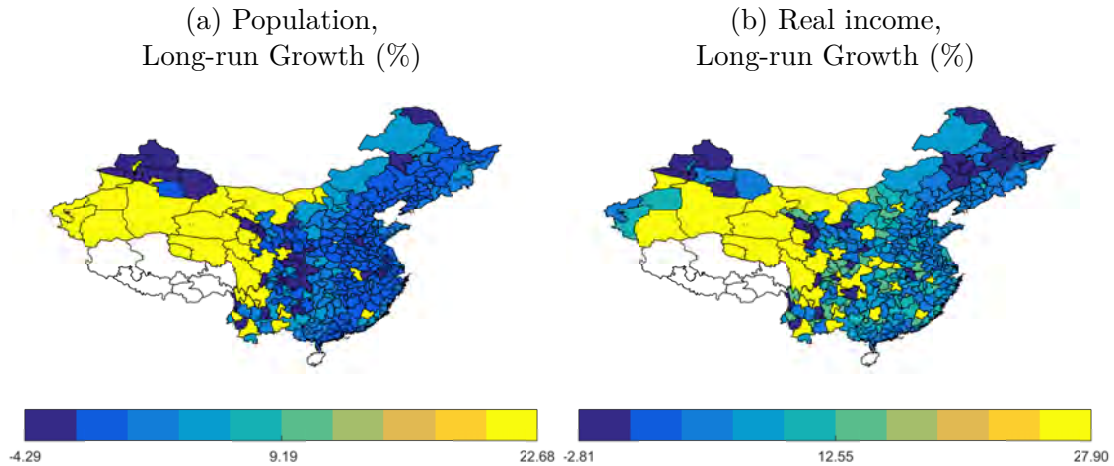


In the remainder of this section, we investigate the effects of both road types – highways versus regional roads – on long-run equilibrium variables. To do so, we derive  $\mathbb{W}_{ij,2013}^{R,\text{high}}$ ,  $\mathbb{W}_{ij,2013}^{C,\text{high}}$ ,  $\zeta_{ij,2013}^{\text{high}}$  and  $\kappa_{ij,2013}^{\text{high}}$  by adding *only* the highway-network improvements made between 2000 and 2013 to  $\mathbb{W}_{ij,2000}^R$ ,  $\mathbb{W}_{ij,2000}^C$ ,  $\zeta_{ij,2000}$  and  $\kappa_{ij,2000}$ .

We proceed analogously for regional road improvements.

**Road Type 1: Improvements of Highways** — We start by looking at the role of highway improvements on the long-run equilibrium. Figure 3.17 reveals that highway improvements led to a larger population growth in the western prefectures as well as along the Mongolian border. This is in line with the connectivity effects of highway improvements observed in Figures 3.2 and 3.16. The picture is similar for real income. We observe larger real income growth in the western prefectures, but mid-eastern prefectures also gain relative to coastal ones. Hence, for both population and real income, highways improvements have mostly acted as a tool to connect the western prefectures to central China, and by extension, to the rest of the world.

Figure 3.17: GROWTH IN LONG-RUN EQUILIBRIUM VARIABLES WITH IMPROVEMENTS OF HIGHWAYS (2000-2013)

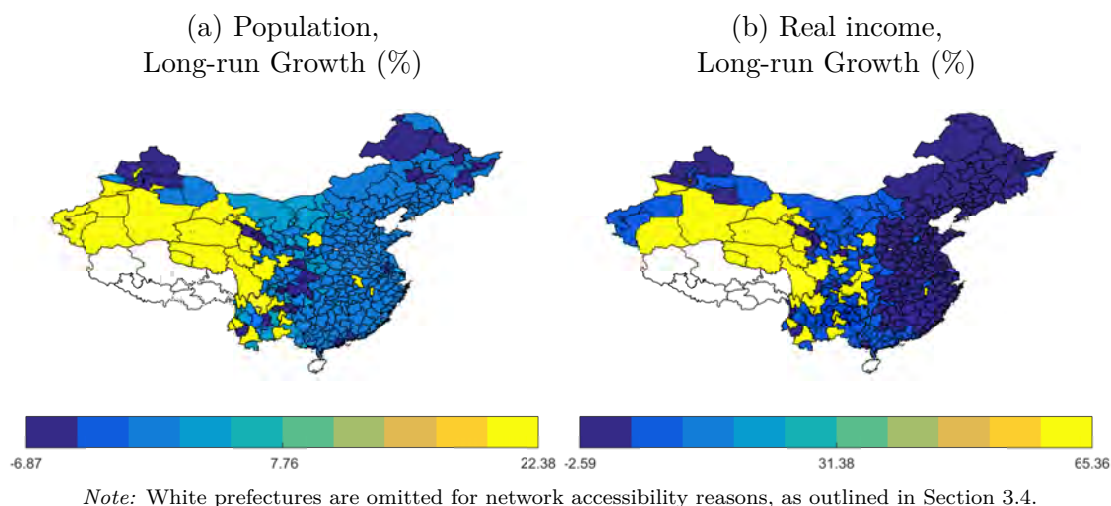


*Note:* White prefectures are omitted for network accessibility reasons, as outlined in Section 3.4.

**Road Type 2: Improvements of Regional Roads** — We now turn to the role of improvements in the regional road network on the long-run equilibrium. Figure 3.18 suggests that regional road improvements have benefited mostly the western prefectures, similarly to highway improvements. There are two main reasons for this result: these regions were characterized by a particularly low level of road infrastructure in 2000, and the regional roads are important means of access to the highway system. As indicated above, regional roads have a less heterogeneous effect across the regions than the highway system does, because the

network has grown very densely. The smooth increase in regional network density is particularly apparent when looking at real income. Real income growth appears to increase in three steps from coastal prefectures (small decline), via mid-eastern prefectures (small growth) to prefectures in the Far West (large growth).

Figure 3.18: GROWTH IN LONG-RUN EQUILIBRIUM VARIABLES WITH IMPROVEMENTS OF REGIONAL ROADS (2000-2013)



In sum, the extension of the highway network has led to a strong integration of formerly remote prefectures into the Chinese economy. The very patchy picture observed in panel (b) of Figure 3.17 indicates that improvements in highways have benefited a given subset of prefectures which have enjoyed large connectivity gains. Conversely, regional roads improvements largely relieved the former differences in local connectivity, since they led to a very dense network of roads between the years 2000 and 2013.

### 3.6 Conclusion

This paper contributes to the literature on the quantitative effects of transport infrastructure in open regions in three ways. First, it compiles and digitizes detailed annual data on China’s road network over the period 2000-2013. Specifically, it does so using a taxonomy which distinguishes between highways, province-level roads, and prefecture-level roads. This data-set permits a detailed view on the changes in road connectivity, e.g., in terms of shortest travel time, at a detailed

regional level and it supports an attribution of the changes to specific layers of the road network.

Second, the paper proposes a quantitative regional model of 330 Chinese prefectures plus a rest of the world, where the transport network may affect regions through four direct channels. Of those, two are customary and at the heart of traditional quantitative models of regions, and these are the goods-trade-cost and the moving-cost (or mobility-cost) channels. Two other ones are not customary, and they relate to the technology-spillover and amenity-spillover effects of transport networks. By this token, we consider that technology will dissipate more easily between better-connected places, and we allow peoples' residence choices not only to depend on locally available amenities but also on ones in the well-connected neighborhood.

Third, we delineate the quantitative effects of China's road network improvements over the considered time span along two lines: distinguishing between the four considered channels on the one hand and between highways and other roads on the other hand. Estimates of key model parameters and simulation results of the model suggest that all four channels of effects and both types of roads are important in their own right. For instance, road network improvements stimulate technology spillovers and reduce trade costs both of which fostered the convergence of Chinese prefectures in terms of population density and real income. The same improvements stimulated amenity spillovers – a dispersion force – and reduced mobility costs – an agglomeration force – in a relatively offsetting way. The improvements of the highway system led *ceteris paribus* to a heterogeneous change in connectivity of the prefectures, but the simultaneous expansion of the provincial- and prefecture-level road network layers equalized these changes by connecting some prefectures directly and by granting highway access to others.



# Appendix

## 3.7 Derivation: Equilibrium Equations

This section derives the set of equations that solves equilibrium wages and employment levels in each location  $i$ .

For simplicity, we rewrite the equations that define amenities in (3.2) and endogenous technology in (3.5) as follows:

$$a_{it} = \left( \frac{L_{it}}{H_i^R} \right)^{-\lambda_1} \tilde{a}_{it}, \quad \text{with} \quad \tilde{a}_{it} = \exp \left( \int_S \mathbb{W}_{ijt-1}^R \log a_{jt-1} dj \right)^{\lambda_2},$$

and

$$T_{it} = \left( \frac{L_{it}}{H_i^C} \right)^{\gamma_1} \tau_{it}, \quad \text{with} \quad \tau_{it} = \exp \left( \int_S \mathbb{W}_{ijt-1}^C \log T_{jt-1} dj \right)^{\gamma_2}.$$

Note that in equilibrium, the location- and time-specific amenity and technology shifter,  $\varepsilon_{it}^R \varepsilon_{it}^C$ , cancel as their expected value is unity for every location  $i$ . Furthermore, as  $\tilde{a}_{it}$  and  $\tau_{it}$  are only dependent on past variables, they are exogenously given in period  $t$ .

**Equilibrium Wages** We start by deriving equilibrium wages in (3.14). To do so, take the good-market-clearing condition (3.11) and substitute (3.7). Then,

$$w_{it} L_{it} = \int_S \frac{T_{it} [o_{it} \zeta_{ijt}]^{-\theta}}{\int_S T_{kt} [o_{kt} \zeta_{kjt}]^{-\theta} dk} w_{jt} L_{jt} dj. \quad (3.33)$$

Inserting  $T_{it} = \tau_{it} \left( \frac{L_{it}}{H_i^C} \right)^{\gamma_1}$  and replacing unit costs (3.6) obtains

$$w_{it} L_{it} = \tau_{it} w_{it}^{-\theta} \left( \frac{L_{it}}{H_i^C} \right)^{-\theta(1-\mu)+\gamma_1} \int_S \frac{\zeta_{ijt}^{-\theta} w_{jt} L_{jt}}{\int_S \tau_{it} \left( \frac{L_{it}}{H_i^C} \right)^{-\theta(1-\mu)+\gamma_1} w_{kt}^{-\theta} \zeta_{kjt}^{-\theta} dk} dj. \quad (3.34)$$

Solving the previous equation for wages gives the first set of equilibrium equations that is detailed in (3.14)

$$w_{it} = L_{it}^{\frac{(\gamma_1 - \theta(1-\mu)) - 1}{1+\theta}} H_i^C \frac{-\theta(\mu-1) - \gamma_1}{1+\theta} \tau_{it}^{\frac{1}{1+\theta}} \left[ \int_S w_{jt} L_{jt} \zeta_{ijt}^{-\theta} \left( \int_S \Pi_{kt} \zeta_{kjt}^{-\theta} dk \right)^{-1} dj \right]^{\frac{1}{1+\theta}},$$

where  $\Pi_{kt} \equiv \tau_{kt} \left( \frac{L_{it}}{H_i^C} \right)^{\gamma_1 - \theta(1-\mu)} w_{kt}^{-\theta}$ .

**Equilibrium Employment** To derive equilibrium employment in (3.15) take migration shares in (3.10) and substitute location-specific utility ( $\tilde{u}_{it}$ ). Then,

$$\frac{L_{it}}{\bar{L}} = \frac{\left[ \frac{a_{it} w_{it}}{P_{it}^\alpha r_{it}^{1-\alpha}} \right]^{1/\Omega} \tilde{\kappa}_{it}}{\int_S \int_S \left[ \frac{a_{kt} w_{kt}}{P_{kt}^\alpha r_{kt}^{1-\alpha}} \right]^{1/\Omega} \kappa_{kjt}^{-1/\Omega} dk dj}, \quad (3.35)$$

where  $\tilde{\kappa}_{it} = \int_S \kappa_{ijt}^{-1/\Omega} dj$ .

Substituting  $a_{it} = \tilde{a}_{it} \left( \frac{L_{it}}{H_i^R} \right)^{-\lambda_1}$ , replacing the residential land rents with (3.13) and the price index with (3.8) gives, after some rearrangements, the second set of equilibrium equations (3.15)

$$L_{it} = \frac{\left( w_{it}^\alpha (\bar{L} \tilde{\kappa}_{it})^\Omega H_i^R (1-\alpha) + \lambda_1 \tilde{a}_{it} \right)^{\frac{1}{\tilde{\Omega}}} \left( \int_S \Pi_{it} \zeta_{ijt}^{-\theta} dj \right)^{\frac{\alpha}{\theta \tilde{\Omega}}}}{\left[ \int_S \int_S \left( \frac{\tilde{a}_{kt} w_{kt}^\alpha H_k^R (1-\alpha) + \lambda_1}{\kappa_{kjt} L_{kt}^{(1-\alpha) + \lambda_1}} \right)^{\frac{1}{\tilde{\Omega}}} \left( \int_S \Pi_{jt} \zeta_{kjt}^{-\theta} dj \right)^{\frac{\alpha}{\theta \tilde{\Omega}}} dk dj \right]^{-\frac{\Omega}{\tilde{\Omega}}}}, \quad (3.36)$$

where  $\tilde{\Omega} \equiv \Omega + \lambda_1 + (1 - \alpha)$ .

### 3.8 Equilibrium: Existence and Uniqueness

The uniqueness condition in (3.16) can be derived along the lines of Desmet et al. (2018) (see their Section B.3). The rewriting of amenities and endogenous technology as stated at the beginning of Section 3.7 applies. We can manipulate the system of equations that defines an equilibrium as follows.

**First Set of Equations** We derive a first set of equations that uses the location-specific indirect utility,  $\tilde{u}_{it}$ , in (3.3) as a basis. Take the price index in (3.8) and substitute the unit costs (3.6), commercial land rents (3.13) and  $T_{it} = \tau_{it} \left( \frac{L_{it}}{H_i^C} \right)^{\gamma_1}$ , then

$$P_{it} = \psi_0 \left[ \int_S \tau_{jt} L_{jt}^{\gamma_1 - \theta(1-\mu)} w_{jt}^{-\theta} H_j^{C - \theta(\mu-1) - \gamma_1} \zeta_{jit}^{-\theta} dj \right]^{-\frac{1}{\theta}}, \quad (3.37)$$

where  $\psi_0 = \bar{p}/\mu$  and  $\bar{p} = \Gamma \left( \frac{1-\sigma}{\theta} + 1 \right)^{\frac{1}{1-\sigma}}$ . Substituting (3.37), residential land rents (3.13) and  $a_{it} = \tilde{a}_{it} \left( \frac{L_{it}}{H_i^R} \right)^{-\lambda_1}$  into (3.3) gives

$$\left[ \frac{\tilde{a}_{it}}{\tilde{u}_{it}} \right]^{-\frac{\theta}{\alpha}} \left[ \frac{L_{it}}{H_i^R} \right]^{-\frac{\theta(\alpha-1-\lambda_1)}{\alpha}} w_{it}^{-\theta} = \psi_1 \int_S \tau_{jt} \left[ \frac{L_{jt}}{H_j^C} \right]^{\gamma_1 - \theta(1-\mu)} w_{jt}^{-\theta} \zeta_{jit}^{-\theta} dj, \quad (3.38)$$

where  $\psi_1 = \psi_0^{-\theta} ((1-\alpha)^{\alpha-1})^{\theta/\alpha}$ .

**Second Set of Equations** We derive a second set of equations that uses the equilibrium goods-market-clearing condition in (3.11). Insert the trade shares (3.7) and the price index into the goods-market-clearing condition so that

$$w_{it} L_{it} = \bar{p}^{-\theta} \int_S T_{it} [o_{it} \zeta_{ijt}]^{-\theta} P_{jt}^{\theta} w_{jt} L_{jt} dj. \quad (3.39)$$

Substituting unit costs (3.6) and  $T_{it} = \tau_{it} \left( \frac{L_{it}}{H_i^C} \right)^{\gamma_1}$ , as well as replacing the price index with the indirect utility into (3.39) yields

$$\begin{aligned} & \tau_{it}^{-1} w_{it}^{1+\theta} H_i^{C\theta(\mu-1)+\gamma_1} L_{it}^{1-(\gamma_1-\theta(1-\mu))} \\ & = \psi_1 \int_S \left[ \frac{\tilde{a}_{jt}}{\tilde{u}_{jt}} \right]^{\frac{\theta}{\alpha}} w_{jt}^{1+\theta} H_j^{R\frac{\theta(1-\alpha)+\lambda_1}{\alpha}} L_{jt}^{1+\frac{\theta(\alpha-1-\lambda_1)}{\alpha}} \zeta_{ijt}^{-\theta} dj. \end{aligned} \quad (3.40)$$

**Proof** We follow the uniqueness proof of Theorem 2 in Allen and Arkolakis (2014), which is based on Theorem 2.19 in Zabreyko et al. (1975). Let us introduce the following function  $\bar{f}_i$ , which is the ratio of LHS's of (3.38) and (3.40):

$$\bar{f}_i = \frac{\tau_{it}^{-1} w_{it}^{1+\theta} H_i^{C\theta(\mu-1)+\gamma_1} L_{it}^{1-(\gamma_1-\theta(1-\mu))}}{\left[ \frac{\tilde{a}_{it}}{\tilde{u}_{it}} \right]^{-\frac{\theta}{\alpha}} L_{it}^{-\frac{\theta(\alpha-1-\lambda_1)}{\alpha}} w_{it}^{-\theta} H_i^{R-\frac{\theta(1-\alpha+\lambda_1)}{\alpha}}}. \quad (3.41)$$

Equivalently,  $\bar{f}_i$  also equals the ratio of the RHS's of (3.38) and (3.40) that is

$$\bar{f}_i = \frac{\int_S \left[ \frac{\tilde{a}_{jt}}{\tilde{u}_{jt}} \right]^{\frac{\theta}{\alpha}} w_{jt}^{1+\theta} H_j^{R \frac{\theta(1-\alpha+\lambda_1)}{\alpha}} L_{jt}^{1+\frac{\theta(\alpha-1-\lambda_1)}{\alpha}} \zeta_{ijt}^{-\theta} dj}{\int_S \tau_{jt} L_{jt}^{\gamma_1 - \theta(1-\mu)} w_{jt}^{-\theta} H_j^{C - \theta(\mu-1) - \gamma_1} \zeta_{jit}^{-\theta} dj}. \quad (3.42)$$

Applying symmetric trade costs,  $\zeta_{ijt} = \zeta_{jit}$ , we can rewrite  $\bar{f}_i$  as follows

$$\bar{f}_i = \frac{\int_S \bar{f}_j^{-\beta} \bar{f}_{ij} \bar{f}_j \, dj}{\int_S \bar{f}_j^{-(1+\beta)} \bar{f}_{ij} \bar{f}_j \, dj}, \quad (3.43)$$

where

$$\bar{f}_{ij} = \frac{\left[ \frac{\tilde{a}_{jt}}{\tilde{u}_{jt}} \right]^{\frac{(1+\beta)\theta}{\alpha}} \tau_{jt}^{-\beta} \zeta_{ijt}^{-\theta} H_j^{C\beta(\theta(\mu-1) - \gamma_1)} H_j^{R \frac{(1+\beta)\theta(1-\alpha+\lambda_1)}{\alpha}} w_{jt}^{1+\theta+(1+2\theta)\beta}}{L_{jt}^{(1+\beta)(1+\frac{\theta(\alpha-1-\lambda_1)}{\alpha}) - \beta(\gamma_1 - \theta(1-\mu))}}. \quad (3.44)$$

Rewrite (3.43) as

$$\bar{\bar{f}}_i = \frac{\bar{f}_i^{-\beta}}{\int_S \bar{f}_j^{-\beta} \bar{f}_{ij} \bar{f}_j \, dj} = \frac{\bar{f}_i^{-(1+\beta)}}{\int_S \bar{f}_j^{-(1+\beta)} \bar{f}_{ij} \bar{f}_j \, dj}. \quad (3.45)$$

Then, changing the notation to

$$\bar{g}_i = \bar{f}_i^{-\beta} \quad \text{and} \quad \bar{\bar{g}}_i = \bar{f}_i^{-(1+\beta)}, \quad (3.46)$$

and rewrite both as follows

$$\bar{g}_i = \int_S \bar{\bar{f}}_i \bar{\bar{f}}_{ij} \bar{\bar{g}}_j \, dj \quad \text{and} \quad \bar{\bar{g}}_i = \int_S \bar{\bar{f}}_i \bar{\bar{f}}_{ij} \bar{\bar{g}}_j \, dj. \quad (3.47)$$

Define  $\bar{\bar{f}}_i \bar{\bar{f}}_{ij}$  as kernel  $K_{ij}$ . Hence,  $\bar{g}_i$  and  $\bar{\bar{g}}_i$  are both solutions to the integral equation

$$x_i = \int_S K_{ij} x_j \, dj, \quad (3.48)$$

where  $K_{ij}$  can be expressed as

$$K_{ij} = \frac{\tau_{it} L_{it}^{\gamma_1 - \theta(1-\mu)} w_{it}^{-\theta} H_i^{C - \theta(\mu-1) - \gamma_1} \zeta_{ijt}^{-\theta}}{\int_S \tau_{jt} L_{jt}^{\gamma_1 - \theta(1-\mu)} w_{jt}^{-\theta} H_j^{C - \theta(\mu-1) - \gamma_1} \zeta_{jit}^{-\theta} dj}, \quad (3.49)$$

or

$$K_{ij} = \frac{\left[ \frac{\tilde{a}_{it}}{\tilde{u}_{it}} \right]^{\frac{\theta}{\alpha}} w_{it}^{1+\theta} H_i^{R \frac{\theta(1-\alpha+\lambda_1)}{\alpha}} L_{it}^{1 + \frac{\theta(\alpha-1-\lambda_1)}{\alpha}} \zeta_{ijt}^{-\theta}}{\int_S \left[ \frac{\tilde{a}_{jt}}{\tilde{u}_{jt}} \right]^{\frac{\theta}{\alpha}} w_{jt}^{1+\theta} H_j^{R \frac{\theta(1-\alpha+\lambda_1)}{\alpha}} L_{jt}^{1 + \frac{\theta(\alpha-1-\lambda_1)}{\alpha}} \zeta_{jit}^{-\theta} dj}. \quad (3.50)$$

We have to ensure that  $K_{ij}$  is (i) non-negative, (ii) measurable and (iii) square-integrable. Non-negativity holds as  $\bar{f}$  and  $\bar{f}$  are non-negative. Measurability holds because it can be shown that  $\bar{f}$  and  $\bar{f}$  are approximately continuous everywhere. Square-integrability holds as long as population at any given location is bounded from below and above. The former is true because by construction population cannot shrink to zero unless nominal wages are zero or amenities are infinitely high. The latter is true because population at any given location cannot exceed the level of world population  $\bar{L}$ .

Given the properties of  $K_{ij}$ , Theorem 2.19 in Zabreyko et al. (1975) guarantees that there exists a unique (to scale) strictly positive function that satisfies the system of equations in (3.48). Hence,

$$\bar{g}_i = \varpi \bar{g}_i \Rightarrow \bar{f}_i^{-\beta} = \varpi \bar{f}_i^{-(1+\beta)} \Rightarrow \bar{f}_i = \varpi, \quad (3.51)$$

where  $\varpi$  is a constant. Hence,

$$\frac{\tau_{it}^{-1} w_{it}^{1+\theta} H_i^{C \theta(\mu-1) + \gamma_1} L_{it}^{1 - (\gamma_1 - \theta(1-\mu))}}{\left[ \frac{\tilde{a}_{it}}{\tilde{u}_{it}} \right]^{-\frac{\theta}{\alpha}} L_{it}^{-\frac{\theta(\alpha-1-\lambda_1)}{\alpha}} w_{it}^{-\theta} H_i^{R \frac{-\theta(1-\alpha+\lambda_1)}{\alpha}}} = \varpi, \quad (3.52)$$

and solving for  $w_{it}$  gives

$$w_{rt} = \bar{w} \left[ \frac{\tilde{a}_{it}}{\tilde{u}_{it}} \right]^{-\frac{\theta}{\alpha(1+2\theta)}} \tau_{it}^{\frac{1}{1+2\theta}} H_i^{C \frac{-\theta(\mu-1) - \gamma_1}{1+2\theta}} H_i^{R \frac{-\theta(1-\alpha+\lambda_1)}{\alpha(1+2\theta)}} L_{it}^{\frac{-\theta(\frac{\alpha-1-\lambda_1}{\alpha}) + (\gamma_1 - \theta(1-\mu)) - 1}{1+2\theta}}, \quad (3.53)$$

where  $\bar{w} = \varpi^{\frac{1}{1+2\theta}}$ . Substituting (3.53) into (3.38) gives

$$\begin{aligned}
& \left[ \frac{\tilde{a}_{it}}{\tilde{u}_{it}} \right]^{-\frac{\theta(1+\theta)}{\alpha(1+2\theta)}} \tau_{it}^{-\frac{\theta}{1+2\theta}} H_i^C \frac{\theta^2(\mu-1)+\gamma_1\theta}{1+2\theta} H_i^R \frac{-\theta(1+\theta)(1-\alpha+\lambda_1)}{\alpha(1+2\theta)} \\
& L_{it}^{-\theta\left(\frac{\alpha-1-\lambda_1}{\alpha}\right)+\frac{1}{1+2\theta}\left[\gamma_1-\theta(1-\mu)-\theta\left(\frac{\alpha-1-\lambda_1}{\alpha}\right)-1\right]} \\
& = \psi_1 \int_S \left[ \frac{\tilde{a}_{jt}}{\tilde{u}_{jt}} \right]^{\frac{\theta^2}{\alpha(1+2\theta)}} \tau_{jt}^{\frac{1+\theta}{1+2\theta}} H_j^C \frac{-\theta(1+\theta)(\mu-1)-\gamma_1}{1+2\theta} H_j^R \frac{\theta^2(1-\alpha+\lambda_1)}{\alpha(1+2\theta)} \\
& L_{jt}^{\gamma_1-\theta(1-\mu)-\frac{\theta}{1+2\theta}\left[\gamma_1-\theta(1-\mu)-\theta\left(\frac{\alpha-1-\lambda_1}{\alpha}\right)-1\right]} \zeta_{ijt}^{-\theta} dj.
\end{aligned} \tag{3.54}$$

Inserting (3.10) into (3.54) gives

$$\begin{aligned}
& \bar{B}_{it} \hat{u}_{it}^{\frac{1}{\Omega}\left[-\theta\left[\frac{\alpha-1-\lambda_1}{\alpha}\right]+\frac{1}{1+2\theta}\left[\gamma_1-\theta(1-\mu)-\theta\left(\frac{\alpha-1-\lambda_1}{\alpha}\right)-1\right]\right]+\frac{\theta(1+\theta)}{\alpha(1+2\theta)}} \\
& = \psi_1 \int_S \hat{u}_{jt}^{\frac{1}{\Omega}\left[\gamma_1-\theta(1-\mu)-\frac{\theta}{1+2\theta}\left[\gamma_1-\theta(1-\mu)-\theta\left(\frac{\alpha-1-\lambda_1}{\alpha}\right)-1\right]\right]-\frac{\theta^2}{\alpha(1+2\theta)}} \bar{B}_{jt} \zeta_{ijt}^{-\theta} dj,
\end{aligned} \tag{3.55}$$

where

$$\begin{aligned}
\bar{B}_{it} &= \tilde{a}_{it}^{-\frac{\theta(1+\theta)}{\alpha(1+2\theta)}} \tau_{it}^{-\frac{\theta}{1+2\theta}} H_i^C \frac{\theta^2(\mu-1)+\gamma_1\theta}{1+2\theta} H_i^R \frac{-\theta(1+\theta)(1-\alpha+\lambda_1)}{\alpha(1+2\theta)} \\
& \tilde{\kappa}_{it}^{-\theta\left[\frac{\alpha-1-\lambda_1}{\alpha}\right]+\frac{1}{1+2\theta}\left[\gamma_1-\theta(1-\mu)-\theta\left(\frac{\alpha-1-\lambda_1}{\alpha}\right)-1\right]},
\end{aligned}$$

and

$$\begin{aligned}
\bar{B}_{jt} &= \tilde{a}_{jt}^{\frac{\theta^2}{\alpha(1+2\theta)}} \tau_{jt}^{\frac{1+\theta}{1+2\theta}} H_j^C \frac{(1+\theta)(-\theta(\mu-1)-\gamma_1)}{1+2\theta} H_j^R \frac{\theta^2(1-\alpha+\lambda_1)}{\alpha(1+2\theta)} \\
& \tilde{\kappa}_{jt}^{\gamma_1-\theta(1-\mu)-\frac{\theta}{1+2\theta}\left[\gamma_1-\theta(1-\mu)-\theta\left(\frac{\alpha-1-\lambda_1}{\alpha}\right)-1\right]},
\end{aligned}$$

and

$$\hat{u}_{it} = \tilde{u}_{it} \left[ \frac{\bar{L}}{\int_S \int_S \tilde{u}_{kt}^{1/\Omega} \kappa_{kjt}^{-1/\Omega} dk dj} \right]^\Omega \left[ 1 - \frac{\theta}{\frac{1}{\Omega}\left[\gamma_1-\theta(1-\mu)-\theta\left(\frac{\alpha-1-\lambda_1}{\alpha}\right)\right]+\theta} \right] \tag{3.56}$$

Rewrite (3.55) as

$$\bar{B}_i f_i^{\beta_1} = \psi_1 \int_S \bar{B}_j \zeta_{ijt}^{-\theta} f_j^{\beta_2} dj, \tag{3.57}$$

and apply Theorem 2.19 in Zabreyko et al. (1975), then the solution  $f_{(\cdot)}$  to equation (3.57) exists and is unique if (a) the functions  $\psi_1 \bar{B}_i^{-1}$  and  $\bar{B}_j \zeta_{ijt}^{-\theta}$  are strictly positive and continuous, and (b)  $\left| \frac{\beta_2}{\beta_1} \right| \leq 1$ . The latter implies

$$\frac{\frac{1}{\Omega} \left[ \gamma_1 - \theta(1 - \mu) - \frac{\theta}{1+2\theta} \left[ \gamma_1 - \theta(1 - \mu) - \theta \left( \frac{\alpha-1-\lambda_1}{\alpha} \right) - 1 \right] \right]}{\frac{1}{\Omega} \left[ -\theta \left[ \frac{\alpha-1-\lambda_1}{\alpha} \right] + \frac{1}{1+2\theta} \left[ \gamma_1 - \theta(1 - \mu) - \theta \left( \frac{\alpha-1-\lambda_1}{\alpha} \right) - 1 \right] \right]} - \frac{\theta^2}{\alpha(1+2\theta)} \leq 1,$$

which after some simplification can be written as the uniqueness condition (3.16) that is stated in Section 3.2.6:

$$\frac{\alpha\gamma_1}{\theta} \leq (1 - \mu\alpha) + \lambda_1 + \Omega.$$

### 3.9 Equilibrium: Stability Condition

In the long run, the distribution of amenities as well as of technology is required to be stable. Here, we spell out the stability condition for amenities, knowing that the one for the technology process is analogous. For amenities, stability requires  $a_{it} = a_{it-1} = a_i^*$  and  $\mathbb{W}_{ijt}^R = \mathbb{W}_{ijt-1}^R = \mathbb{W}_{ij}^{R*}$  for all  $i$  and  $j$ . Using  $\mathbf{a}^* = (a_i^*)$  to denote the  $s \times 1$  stacked vector of amenities of  $s$  regions in a long-run equilibrium,  $\mathbf{W}^{R*} = (\mathbb{W}_{ijt}^R)$  for the respective  $s \times s$  equilibrium connectivity matrix,  $\mathbf{d}^* = (d_i^*)$  for the  $s \times 1$  vector of equilibrium population densities,  $\mathbf{I}_s$  for an  $s \times s$  identity matrix, and  $\boldsymbol{\varepsilon}^{a*}$  for the  $s \times 1$  vector of amenity shifters in logs in equilibrium, this means that the relationship for amenities in equation (3.2) in equilibrium needs to adhere to

$$\log \mathbf{a}^* = -\lambda_1 \log \mathbf{d}^* + \lambda_2 \mathbf{W}^{R*} \log \mathbf{a}^* + \boldsymbol{\varepsilon}^{a*}, \quad (3.58)$$

$$\Rightarrow (\mathbf{I}_s - \lambda_2 \mathbf{W}^{R*}) \log \mathbf{a}^* = \boldsymbol{\varepsilon}^{a*} - \lambda_1 \log \mathbf{d}^*, \quad (3.59)$$

$$\Rightarrow \log \mathbf{a}^* = (\mathbf{I}_s - \lambda_2 \mathbf{W}^{R*})^{-1} \boldsymbol{\varepsilon}^{a*} - (\mathbf{I}_s - \lambda_2 \mathbf{W}^{R*})^{-1} \lambda_1 \log \mathbf{d}^*. \quad (3.60)$$

Given that  $\mathbb{E}[\boldsymbol{\varepsilon}^{a*}] = \mathbf{0}$ , we have a unique mapping of amenity and population density in equilibrium:

$$\mathbb{E}[\log \mathbf{a}^*] = -(\mathbf{I}_s - \lambda_2 \mathbf{W}^{R*})^{-1} \lambda_1 \log \mathbf{d}^*. \quad (3.61)$$

As long as  $(\mathbf{I}_s - \lambda_2 \mathbf{W}^{R*})^{-1}$  as well as  $\lambda_1 \log \mathbf{d}^*$  exist and are finite, so are the  $s$  elements of  $\log \mathbf{a}^*$ . In the literature on network effects, it is customary to normalize  $\mathbf{W}^{R*}$  so that either all row sums or the maximum row sum of  $\mathbf{W}^{R*}$  is unity. The latter normalization then requires  $|\lambda_2| < 1$  so that  $(\mathbf{I}_s - \lambda_2 \mathbf{W}^{R*})^{-1} = \sum_{q=0}^{\infty} (\lambda_2 \mathbf{W}^{R*})^q$  is finite (see Kelejian and Prucha, 2010). We use row normalization about  $\mathbf{W}^{R*}$  which implies that all of its elements, which are real numbers with

$0 \leq \mathbb{W}_{ijt}^R \leq 1$ , are absolutely summable, and  $\sum_j \mathbb{W}_{ij}^{R*} \leq 1$ . Altogether, the assumptions about the elements of  $\mathbf{W}^{R*}$  and  $|\lambda_2| < 1$  guarantee that all elements of  $\log \mathbf{a}^*$  exist and are real numbers.

### 3.10 Calibration: Trade-cost Parameters

For the purpose of clarity, this section repeats significant parts of Section 3.4.6, and additionally includes the detailed derivation of each step.

**Trade Elasticity and Dispersion of Technology** Consider two prefectures  $i$  and  $i'$  in the same province with similar Mahalanobis distance values  $\mathfrak{M}_{it} \cong \mathfrak{M}_{i't}$ . The latter means that these prefectures have approximatively the same trade costs to all other prefectures. Thus,

$$\begin{aligned} \frac{\pi_{ip^{it}}^C}{\pi_{i'p^{it}}^C} &= \frac{\int_{\mathfrak{p}^i} T_{it}(o_{it})^{-\theta} \zeta_{ijt}^{-\theta} dj}{\int_{\mathfrak{p}^i} \int_S T_{kt}(o_{kt})^{-\theta} \zeta_{kjt}^{-\theta} dk dj} \left( \frac{\int_{\mathfrak{p}^i} T_{i't}(o_{i't})^{-\theta} \zeta_{i'jt}^{-\theta} dj}{\int_{\mathfrak{p}^i} \int_S T_{kt}(o_{kt})^{-\theta} \zeta_{kjt}^{-\theta} dk dj} \right)^{-1} \\ &\cong \frac{T_{it}(o_{it})^{-\theta}}{T_{i't}(o_{i't})^{-\theta}}, \end{aligned} \quad (3.62)$$

if  $\mathfrak{M}_{it} \cong \mathfrak{M}_{i't}$ .

Similarly, considering the expenditure shares on these two prefectures outside of others than their own province, we have:

$$\begin{aligned} \frac{\pi_{ir^{it}}^C}{\pi_{i'r^{it}}^C} &= \frac{\int_{\mathfrak{r}^i} T_{it}(o_{it})^{-\theta} \zeta_{ijt}^{-\theta} dj}{\int_{\mathfrak{r}^i} \int_S T_{kt}(o_{kt})^{-\theta} \zeta_{kjt}^{-\theta} dk dj} \left( \frac{\int_{\mathfrak{r}^i} T_{i't}(o_{i't})^{-\theta} \zeta_{i'jt}^{-\theta} dj}{\int_{\mathfrak{r}^i} \int_S T_{kt}(o_{kt})^{-\theta} \zeta_{kjt}^{-\theta} dk dj} \right)^{-1} \\ &\cong \frac{T_{it}(o_{it})^{-\theta}}{T_{i't}(o_{i't})^{-\theta}}, \end{aligned} \quad (3.63)$$

as long as  $\mathfrak{M}_{it} \cong \mathfrak{M}_{i't}$ .



Knowing that we can express  $o_i$  as a function of  $L_{it}$ ,  $w_{it}$ ,  $H_i^C$ , and  $\mu$  following (3.6), we can rewrite (3.63) for prefectures with  $\mathfrak{M}_{it} \cong \mathfrak{M}_{i't}$  by taking logs as:

$$\begin{aligned} \log \frac{\pi_{ip^it}^C}{\pi_{i'p^it}^C} &= -\theta \log \left( \frac{w_{it} \left( \frac{L_{it}}{H_i^C} \right)^{1-\mu}}{w_{i't} \left( \frac{L_{i't}}{H_{i'}^C} \right)^{1-\mu}} \right) + \varepsilon_{ii^i p^it} \\ \log \frac{\pi_{ir^it}^C}{\pi_{i'r^it}^C} &= -\theta \log \left( \frac{w_{it} \left( \frac{L_{it}}{H_i^C} \right)^{1-\mu}}{w_{i't} \left( \frac{L_{i't}}{H_{i'}^C} \right)^{1-\mu}} \right) + \varepsilon_{ii^i r^it}, \end{aligned} \quad (3.64)$$

Note that in equation (3.64) the two terms  $\varepsilon_{ii^i p^it}$  and  $\varepsilon_{ii^i r^it}$  are treated as residual terms which are functions of the endogenous technology terms  $T_{it}$  and  $T_{i't}$ . As outlined in Section 3.4.4, the vectors  $L_{it}$ ,  $w_{it}$ ,  $H_i^C$ , and  $\mu$  are observable. We estimate (3.64) using only those prefectures in a province, for which the absolute difference  $|\mathfrak{M}_{it} - \mathfrak{M}_{i't}|$  is minimal. We instrument  $o_{it}/o_{i't}$  with the natural logarithm of the ratio of land areas<sup>38</sup> of prefectures  $i$  and  $i'$ . Adopting this procedure, we obtain a value of  $\theta = 3.570$ .

**Elasticity of Travel Times to Trade Costs** Given the value of  $\theta$ , we can search for trade frictions  $\zeta_{ijt}$ . Similarly to the migration frictions, we model  $\zeta_{ijt}$  as an exponential function of travel time,  $d_{ijt}$ , where  $\phi_2$  governs the translation of travel time into trade frictions. Then,

$$\zeta_{ijt} = \exp(\phi_2 d_{ijt}). \quad (3.65)$$

Consider the ratio of ratios of expenditure on two prefectures  $i$  and  $i'$  from the same and other provinces in China. Let us again consider  $i$  and  $i'$  to belong to the same province but now use only ones for which  $\mathfrak{M}_{it} \neq \mathfrak{M}_{i't}$ .

---

<sup>38</sup>Land area is measured in squared kilometers. Note that the availability of land in a prefecture is relevant and exogenous in the model.

For such prefectures, this ratio of ratios is defined as

$$\begin{aligned}
\frac{\pi_{ip^i t}^C \pi_{i'r^i t}^C}{\pi_{i'p^i t}^C \pi_{i'r^i t}^C} &= \frac{\int_{\mathbf{p}^i} T_{it} [o_{it} \zeta_{ijt}]^{-\theta} dj}{\int_{\mathbf{p}^i} \int_S T_{kt} [o_{kt} \zeta_{kjt}]^{-\theta} dk} \frac{\int_{\mathbf{r}^i} T_{i't} [o_{i't} \zeta_{i'jt}]^{-\theta} dj}{\int_{\mathbf{r}^i} \int_S T_{kt} [o_{kt} \zeta_{kjt}]^{-\theta} dk} \\
&= \frac{\int_{\mathbf{p}^i} T_{i't} [o_{i't} \zeta_{i'jt}]^{-\theta} dj}{\int_{\mathbf{p}^i} \int_S T_{kt} [o_{kt} \zeta_{kjt}]^{-\theta} dk} \frac{\int_{\mathbf{r}^i} T_{it} [o_{it} \zeta_{ijt}]^{-\theta} dj}{\int_{\mathbf{r}^i} \int_S T_{kt} [o_{kt} \zeta_{kjt}]^{-\theta} dk} \\
&= \frac{\int_{\mathbf{p}^i} \zeta_{ijt}^{-\theta} dj \int_{\mathbf{r}^i} \zeta_{i'jt}^{-\theta} dj}{\int_{\mathbf{p}^i} \zeta_{i'jt}^{-\theta} dj \int_{\mathbf{r}^i} \zeta_{ijt}^{-\theta} dj}.
\end{aligned} \tag{3.66}$$

Note that the left-hand-side variable in equation (3.66) is observed, and, regarding the right-hand side,  $\theta$  is known from above, and so is  $\zeta_{ijt}$  up to the scalar  $\phi_2$  in equation (3.65). Hence, we can minimize the sum of squared distances between the left-hand-side and the right-hand side in equation (3.66). In fact, this procedure is best informed for those prefectures  $i$  and  $i'$ , for which the absolute difference  $|\mathfrak{M}_{it} - \mathfrak{M}_{i't}|$  takes on the maximum value within a province. Using only those prefectures in a province, for which the absolute difference  $|\mathfrak{M}_{it} - \mathfrak{M}_{i't}|$  is maximal, the proposed procedure obtains a parameter of  $\phi_2 = 0.044$ .

### 3.1.1 Data Source: Transportation Atlases

Table 3.6: SOURCES OF CHINESE TRANSPORTATION ATLASES

Year*	Title (English)	Title (pinyin)	Publisher (English)	Editor
2001	New Atlas of China's transportation	Xinbian zhongguo jiaotong dituce	SinoMaps Press	
2002	China Transportation Atlas	Zhongguo jiaotong tuce	Planet Map Press	
2003	General Transportation Atlas of China	Tongyong zhongguo jiaotong dituce	Hunan Map Publishing House	Han Jianzhong
2004	Practical China Transportation Atlas	Shiyong zhongguo jiaotong dituce	Xi'an Map Publishing House	
2005	China Transportation Atlas	Zhongguo jiaotong tuce	Planet Map Press	
2006	China Transportation Atlas	Zhongguo jiaotong tuce	Planet Map Press	
2007	China Transportation Atlas	Zhongguo jiaotong dituce	Shandong Province Map Publishing House	Wang Huaibao
2008	China Transportation Atlas	Zhongguo jiaotong dituce	Shandong Province Map Publishing House	Wang Huaibao
2009	China Transportation Atlas	Zhongguo jiaotong dituce	Shandong Province Map Publishing House	Wang Huaibao
2010	New Atlas of China's transportation	Xinbian zhongguo jiaotong dituce	Fujian Province Map Press	
2011	China Transportation Atlas	Zhongguo jiaotong dituce	Shandong Province Map Publishing House	Yuan Xinfang
2012	China Transportation Atlas	Zhongguo jiaotong dituce	SinoMaps Press	Chen Zhuoning
2013	China Transportation Atlas	Zhongguo jiaotong dituce	SinoMaps Press	Chen Zhuoning
2014	General Transportation Atlas of China	Tongyong zhongguo jiaotong dituce	Hunan Map Publishing House	Jiang Gonghe

*Note:* \*corresponds to the year of publication. A transportation atlas that is published in a given year contains maps of the previous year.



# Chapter 4

## Capital Cities and Road Network Integration: Evidence from the U.S.

### 4.1 Introduction

A significant share of world trade in goods happens within national borders, via national roads. In 2007, 80 percent of U.S. manufacturing production was traded domestically (Egger et al., 2019) and 44 percent of it using the national road network (U.S. Bureau of Transportation Statistics).<sup>1</sup> Thus, integration into the national transport system is key for any city’s economic prosperity. Cities, however, are heterogeneous. One important, yet understudied, dimension in which cities differ is their political status. Investigating whether city heterogeneity, such as political status, affects a city’s access to the transport network is important for local economic prosperity, and, therefore, for both researchers and policy makers.

This paper links the political status of U.S. urban areas to their integration in the national road network, in order to understand whether there is a *capital premium* – i.e., a premium to being a state capital city in terms of road infrastructure provision. Road network integration is defined as a class of measurements that evaluate how well a location is connected to other locations through the National

---

<sup>1</sup>Egger et al. (2019) report country-level information on own consumption in total production of manufacturing goods in US dollars. The U.S. Bureau of Transportation Statistics provides information on shipments by travel mode in U.S. ton-miles of freight.

Highway System (NHS). The main empirical result suggests that, indeed, U.S. state capitals have on average 14 percent larger levels of (population-weighted) road network integration compared to non-capital cities of similar characteristics.

The paper contributes to the literature in the following two ways. First, to the best of my knowledge, this is the first paper that quantifies the causal effect of political status on road network integration and applies the analysis to 920 U.S. Core Based Statistical Areas (CBSAs). The U.S. offer a unique variation to answer the question at hand. There is a large variation of state capital characteristics, which allow me to differentiate political status from size effects. Moreover, it is the only country in the world where such a large number of states, capital cities and urban areas are connected by a common national road network under the same institutional and cultural framework.

The second dimension in which the paper contributes to the literature is its instrumentation strategy. The location choice for most state capitals was closely related to the westward expansion of the U.S. along historical exploration roads, which are correlated with nowadays' transport network (see Duranton and Turner, 2012). To tackle this concern, I construct an instrument which captures the fact that state centrality – *independent of the transport network* – is a key feature of U.S. capital cities. Formally, I employ a  $k$ -means clustering algorithm – a concept that is widely applied in machine learning – that predicts the boundaries of 48 U.S. states and defines their geographical center as a hypothetical capital location. Then, capital status is predicted by the rank in distance to the respective hypothetical capital location.

For each CBSA, I determine the integration in the NHS by four different measures of road network integration: *connectivity* and *market access*, which are based on absolute distances between locations, and *relative connectivity* and *relative market access*, which are based on relative distances between locations. All measures have in common that their value for a given CBSA is the sum over connections to all other places. While connectivity and market access define a connection as inverse (absolute) distance on the network, direct connectivity and direct market access evaluate the distance on the network relative to the great-circle distance. Overall, I find a positive and significant effect of capital status on relative distance

measures, which is evidence for a more direct integration of capital cities in the National Highway System. The effect is driven by capital cities in large states and those with an above-median rate of urbanization (at the national level).

There are two reasons why U.S. state capitals are expected to be better integrated in the federal road network than comparably large non-capital cities. First, (most) state capitals embody their role of political power by being centrally located and easily accessible from other urban centers around them. This makes U.S. capital cities a natural candidate for a direct integration in the transport network, following Christaller's (1933) Central Place Theory (CPT). The transport principle in the CPT suggests that the most efficient and cost minimizing transport network is one that radially connects the most central place in the hierarchy to all other places in the jurisdiction.

Another reason why capital cities may have been favored in the provision of road network infrastructure is that the decision on new road locations is a highly political one. Highway spending, in particular, responds to strong interest groups, 'pork barrel' projects being an obvious example (see Evans, 1994). In the case of the *Interstate Highway System* (as part of the NHS), states were asked to submit proposals for their portion of the federal highway network in response to the recommended national plan (see Baum-Snow, 2007). The final proposal was quite certainly an outcome of inter-governmental lobbying, both from private and public sector interest groups. Of course, whenever the capital city was also the largest economic center, better road infrastructure provision has a straight-forward economic implication. However, for capital cities with little economic relevance this argument does not hold. In those cases, the political status itself could have been the main driver to attracting better access to road infrastructure – despite economic theory predicting the revers.

State capital centrality and the political interest to have state capitals well connected to major urban centers are the main mechanisms that motivate the empirical analysis in this paper. In the discussion, I provide further evidence for both mechanisms that underline an existing *capital premium* in direct road network integration.

This paper relates to several strands of the literature. It first relates to the new economic geography literature that has emphasized the importance of market access in explaining the spatial distribution of economic activity (starting with Krugman, 1991). Apart from theoretical contributions on market access (e.g., Helpman, 1998; Redding and Sturm, 2008), there is a vast empirical literature that focuses on the relationship between access to markets and economic development (e.g., Davis and Weinstein, 2003; Hanson and Xiang, 2004; Redding and Venables, 2004; Hanson, 2005). Most empirical contributions that specifically estimate the importance of transport infrastructure on urban development (e.g., Banerjee et al., 2012; Baum-Snow, 2007; Michaels, 2008; Donaldson, 2018) face an econometric challenge as changes in the transport infrastructure have both direct and indirect (i.e., general equilibrium induced) effects on the observed location. Donaldson and Hornbeck (2016) provide a methodology for measuring the *aggregate* impact of transport infrastructure changes using a reduced-form market access approach that is derived from general equilibrium trade theory. This paper builds on the insights from Donaldson and Hornbeck (2016) in creating the road network integration measures.

When analyzing the link between transport infrastructure and urban development, most studies differentiate cities according to economic characteristics such as city size, productivity or sector composition. Political status as an additional source of heterogeneity, however, has gained only recently more attention. In particular, there have been a few contributions that exploited the economic consequences of relocating or constructing national capitals. For instance, Becker et al. (2018) evaluate the impact of a public employment shock on private sector employment due to the relocation of the German capital from Berlin to Bonn and vice versa. Morten and Oliviera (2018) quantify the effect of an exogenous shock in transport infrastructure on trade and migration, succeeding the construction of Brazil's new capital Brasilia. Bai and Jia (2020) exploit the historical variation in changing provincial capitals to analyze the importance of administrative hierarchy on local development in China. They find evidence that Chinese regimes readjusted the transportation network in favor of prefectures that had capital status.

In contrast to urban economics, the importance of political status has been



widely studied in political sciences, more specifically, in the field of political economy. Most contributions in this literature concentrate on comparative analyses of national capitals in terms of locational policy agendas (see, i.e., Nagel, 2013, for capital cities of federations; Mayer et al., 2017, for capitals that are not the primary economic city of their nations; Rossman, 2018, for newly established capital cities).

Finally, this paper relates to a new strand of the economic geography literature that has applied algorithmic approaches in instrumentation strategies. For example, Faber (2014) instruments the endogenous placement of China’s National Expressway Network based on Kruskal’s (1956) minimum-spanning-tree algorithm. Egger et al. (2020) use a modified algorithm of the classical Monge-Kantorovich transportation problem to address the non-random highway and secondary road placements in China between 2000 and 2013. Moreover, machine learning approaches have recently been employed in the urban context as a tool to (optimally) define statistical borders of urban areas (see, among others, Arribas-Bel et al., 2019; de Bellefon et al., 2019). All of the mentioned studies have in common, that they use mathematical tools, including machine learning algorithms, to replicate key observed institutional features. In particular, they focus on modeling the main determinant that guides the institutional design. For example, transport networks are constructed along the least cost path (Faber, 2014; Alder and Kondo, 2018; Egger et al., 2020) or urban statistical areas unite locations with similar urban context (Arribas-Bel et al., 2019; de Bellefon et al., 2019). As long as the key determinant is well identified, these approaches will provide powerful instruments. The present paper adopts a similar approach and predicts hypothetical capital locations using the  $k$ -means clustering algorithm that is based on geographical and topological data. By minimizing the distance from the hypothetical capital to all points within a cluster, the  $k$ -means clustering algorithm exploits the fact that state capital cities occupy central locations within their jurisdictions.<sup>2</sup>

The structure of the paper is as follows. Section 4.2 provides a historical background on U.S. American state capitals and the U.S. National Highway System that motivates the empirical analysis. Section 4.3 presents the data and introduces

---

<sup>2</sup>A formal analysis of the performance of the algorithm is provided in Section 4.4.

the measurements of road network integration. Section 4.4 outlines the identification strategy and the instrumental variable design, and presents the main empirical result. Section 4.5 discusses drivers and possible mechanisms behind the results. Finally, Section 4.6 concludes.

## 4.2 Historical Background

### 4.2.1 U.S. American State Capitals

The historical geography of American state capitals is complex and diverse. This section attempts to summarize their spatial and historical evolution in five major patterns, which will inform the empirical analysis of this paper. The summarized facts build heavily on Christian Montès' (2014) comprehensive contribution on *American Capitals: A Historical Geography*. I start with a brief overview of the U.S. settlement and the evolution of urban centers in the 19<sup>th</sup> century.

After the creation of the United States of America in 1776, the U.S. territory expanded gradually toward the West, with its first great expansion being the Louisiana Purchase of 1803. The territorial expansion was followed by a substantial shift of population and economic activity from the coast to the center of the territory. U.S. population increased by a factor of eight between 1790 and 1860 and new cities formed, which led to a rapid urbanization during the 19<sup>th</sup> century. Two-thirds of the increase in urbanization can be attributed to new cities forming (predominantly) in the South and the Midwest (see Nagy, 2017, for a review on U.S. urban history in the 19<sup>th</sup> century).

The land was organized into territories and then states. Once established, states have generally retained their initial borders.<sup>3</sup> Capitals as symbolic places and embodiments of political power and decision making have participated in territorial structuring. Local elites tried to win state capital status not only for economic advantage but also for the political stability that such a status might provide. Most of the time, however, stability did not last, and state capitals mi-

---

<sup>3</sup>The exception are four states that have been created from land claimed by another state (Maine, Kentucky, Vermont, West Virginia) and four states (Louisiana, Missouri, Nevada and Pennsylvania) that expanded significantly after acquiring additional federal territory (Zandt, 1976).

grated – on average 3.8 times (Fact 1). The spatial patterns of capital movements were highly heterogeneous (Fact 2). Some states relocated their capitals following the general trends of the U.S. settlements, others experienced a rotation system of capital cities, and again others readjusted their capital location to balance economic and political forces. Eventually, all but eleven of the (present) state capitals were established in the 19<sup>th</sup> century, 35 of them before the American Civil War in 1860 (Fact 3). The final decision on capital selection in every state was as heterogeneous as the path that led toward it. However, one striking factor of capital selection was predominant: most states decided against the largest city of the time in the interest of economic and political balance (Fact 4). Some capitals have remained small, others evolved into bustling metropolises. While investigating the reasons for this different development path might warrant a deeper analysis, it is clearly beyond the scope of this paper. One important aspect that distinguished state capitals at the time, though, is whether or not they were strategically located at important trading routes (Fact 5).

In what follows, I explain the mentioned facts in more detail.

### **Fact 1: Migration of State Capitals**

Most first chosen capitals marked the *entry point* in the 'New World' or strategic defense spots that were built to “protect” the pioneers. However, political stability did not last long and capitals migrated – often westward, following the territorial conquest. Only eight states – Hawaii, Massachusetts, Minnesota, Nevada, New Mexico, Utah, Washington and Wyoming – never changed their capital. On average, American states have had 3.84 successive capitals. California, for instance, changed its capital seven times between 1849-1854. How often states have moved their capital does not follow a clear geographical pattern.

### **Fact 2: Heterogeneous Spatial Patterns of Capital Movements**

Montès (2014) identifies three major spatial patterns of capital movement that amount to 80 percent of all cases.

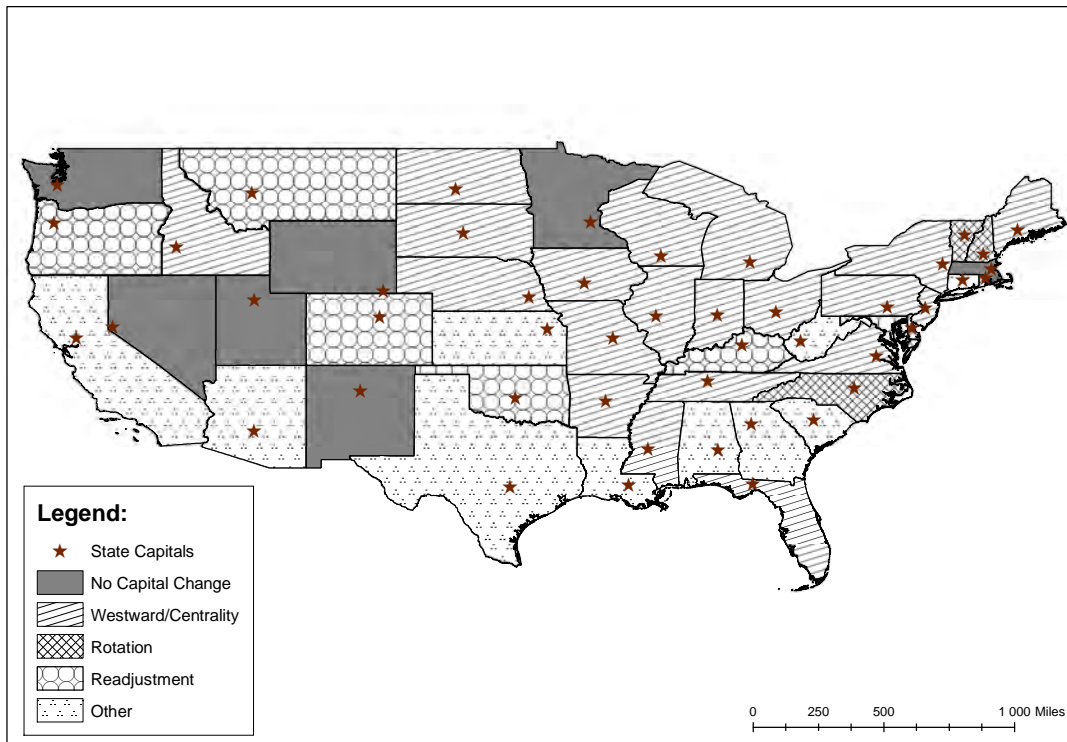
**Westward/Centrality** By far the most common reason to relocate the capital (44% of all cases) was due to the *western pull factor* that initiated the movement inland from the coast to the center of a territory. The process first occurred in the East, where coastal capitals had to yield to more centrally located cities, due to the westward expansion inside the state. For instance, New Jersey's first capital Elizabethtown – the port of entry – was relocated 50-miles south-westward and Trenton became the new capital in 1790.

**Rotation** Five states (10%) – mostly small ones – experienced a complex system of wandering capitals: Delaware, New Hampshire, North Carolina, Rhode Island and Vermont. This was often an outcome of political rivalries. Apart from Rhode Island, all other states that experienced a capital rotating system chose their permanent capital around the turn to the 19<sup>th</sup> century. Anecdotal evidence suggests that as the abrupt ending of rotation happened within two decades, the ultimate capital choice was a quasi-random outcome among all geographically central alternatives (Montès, 2014, p.71).

**Readjustment** Six states (14%) – Alaska, Colorado, Kentucky, Maryland, Montana, Oklahoma, Oregon – relocated their capital in a readjustment process to alter economic and political balance. The readjustment process typically implied a small relocation not far from the first capital choice towards the center of the state.

All three major spatial patterns for capital migration are somehow linked to state centrality. Nowadays, U.S. state capitals are on average located in a radius of 70 miles from the state centroid. Figure 2 summarizes the spatial patterns of capital change and shows a map of all states indicating the reason for capital relocation.

Figure 4.1: SPATIAL PATTERNS OF CAPITAL MOVEMENT



### Fact 3: Timing of Capital Selection

The majority of state capitals (79%) were selected during the 19<sup>th</sup> century. With New Mexico being the exception (Santa Fe was chosen already in 1610), the first states to select their permanent capital are – not surprisingly – those along the east coast: Massachusetts (1692), Maryland (1694), Delaware (1781), and Virginia (1779). In total, 35 states had made the decision for their permanent capital before 1860 – just when the American Civil War hit the country and created a large and long-lasting impact on the U.S. economy. With this timing in mind, one could argue that for those 35 states, the reconstruction and subsequent economic development after the Civil War happened with the capital city location as given. Table 4.8 in the Appendix summarizes the timing of capital selection for all states and adds additional information, including income and population statistics as well as the rank (by population) of each capital city now and then.

Table 4.1: DESCRIPTIVE STATISTICS

	Capitals	Non-Capitals	Top 50 Non-Capitals*
Av. Population (2010)	1,100,626	267,133	2,777,541
Av. Annual Wage (2018)	50,949	45,874	51,556
Av. Area (CBSA)	6,305	2,824	8,113
Av. Population Rank within State	2.75	15.06	2.40
Av. Annual Wage Rank within State	2.16	7.21	4.70
Av. Area within State	3.88	15.00	4.22
Observations	48	872	50

\*: by population in 2010. *Data Source:* Wage estimates (2018) for 376 CBSAs are taken from U.S. Bureau of Labor Statistics. Population records in 2010 and area in km<sup>2</sup> in 2017 by CBSA are provided by the U.S. Census Bureau. *Notes:* The average population rank within states ranges between 1-12 for capitals, 1-69 for non-capitals, and 1-9 for top 50 non-capitals. The average wage rank within states ranges between 1-7 for capitals, 1-26 for non-capitals, and 1-24 for top 50 non-capitals. The average area rank within state ranges between 1-18 for capitals, 1-69 for non-capitals, and 1-29 for top 50 non-capitals.

## Fact 4: Rejection of the Largest City

Even though states and their selection process for capital cities were highly heterogeneous, there is one striking pattern: the majority of states (70%) decided against the largest city (at the time) for their capital. The basic facts in Table 4.8 compare the population rank of all capital cities at the year of selection to the rank in 2010. Some capitals have remained small (e.g., Frankfort, Kentucky; Annapolis, Maryland; Carson City, Nevada), while others have evolved into the largest cities of their state (e.g., Jackson, Mississippi; Phoenix, Arizona; Atlanta, Georgia). Nowadays, most capitals are de facto large and economically important in their state, though, in 42 percent of the cases they are not *the* largest city.

Table 4.1 presents further descriptive statistics regarding recent population, wages and area for three categories: (present) capital cities, non-capital cities and the top 50 non-capitals cities (by population). It suggests that capital cities are half the size of the largest non-capital cities in terms of population, but four times larger than the average non-capital metropolitan area. Average annual wages in capital cities are similar to wages in the largest non-capital cities, however, capital cities are about 25% smaller in area than large non-capital agglomerations. Within their respective state, capital cities occupy an higher rank in population (i.e., they are relatively smaller in population), but a lower rank in wages and area relative to large non-capital cities.<sup>4</sup>

<sup>4</sup>The wage rank in Table 4.1 has not be interpreted with caution as data on wage estimates in metropolitan and non-metropolitan areas provided by the U.S. Bureau of Labor Statistics is only available for 376 out of 920 CBSAs.

## **Fact 5: Capitals as Mercantile Gateways**

Capital cities were of primary importance to the developing trading network. Almost all first capitals in the newly settled West were founded at important trading posts (Montès, 2014, p.119). Inland transportation in the 19<sup>th</sup> century relied heavily on trails and streams, while the importance of the road network declined drastically with technological change. However, intra-state and short-distance transportation still depended on the existing road network. In particular, centrally located capitals served as re-distribution hubs to all other populated places within their state.

### **4.2.2 The U.S. National Highway System**

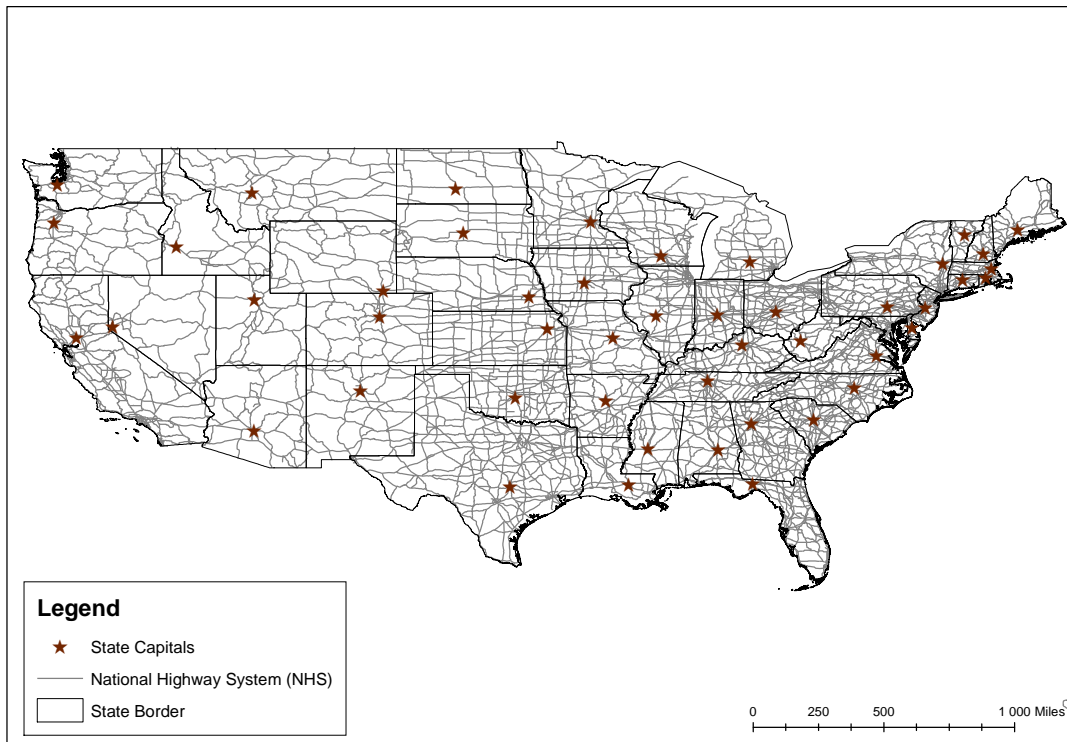
The National Highway System (NHS) constitutes the major federal road network, which strategically connects all states across the U.S. The first federal involvement in developing a national highway system came with the Federal Aid Highway Act of 1944 and the subsequent construction of the Interstate Highway System. By 2011, about 164,000 miles of national highways were completed, of which 47,000 miles comprise the Interstate Highway System. According to the U.S. Department of Transportation, all urban centers with a population of over 50,000 are within five miles of the network.

Figure 4.2 portrays the federal road network commonly known as the National Highway System and highlights the location of state capitals. It shows that the network is more dense in the Northeast – i.e., in proximity of a larger number of high-density urban areas – and less dense in the Midwest and West of the U.S. (except for California).<sup>5</sup> Moreover, the map suggests that in the present network capital cities are well connected. Especially in those states where capital centrality is predominant, the federal road network extends radially in all directions, suggesting a direct integration of capital cities (i.e., Arizona, Iowa, Indiana).

---

<sup>5</sup>The larger number of high-density urban areas in the Northeast is also a result of an average smaller state size in north-eastern U.S.

Figure 4.2: STATE CAPITALS AND NATIONAL HIGHWAY SYSTEM



*Note:* National Highway System as of 2017 (Natural Earth Data Version 4.0.0).

### 4.3 Data Construction

The units of analysis are Core Based Statistical Areas (CBSA) in the U.S. CBSAs include Metropolitan and Micropolitan Statistical Areas and consist of the county, counties or equivalent entities associated with at least one urban core of at least 10,000 people.<sup>6</sup> After excluding non-contiguous jurisdictions and off-shore territories (i.e., Alaska, Hawaii and Puerto Rico) the subsequent analysis includes a total of 920 CBSAs.

**Geographical Boundaries and Population** The U.S. Census Bureau provides information on geographical boundaries in 2015 and total population estimates between 2010 and 2017 for each CBSA. The geographical extent of a CBSA can be extracted using ArcGIS Software. CBSAs that belong to several

<sup>6</sup>Metropolitan Statistical Areas (MSAs) are based on urbanized areas of 50,000 people or more. Micropolitan Statistical Areas ( $\mu$ SAs) are based on urban clusters of at least 10,000 but less than 50,000 people. Adjacent counties become part of a larger urban entity if they have a high degree of social and economic integration with the core as measured by commuting ties.



U.S. states are attributed to the state in which the main urban center is located (i.e., New York-Newark-Jersey City is assigned to the state of New York even though it extends into New Jersey and Pennsylvania). Historical population records and county boundaries for each decade between 1790-1900 are provided by the National Historical Geographic Information System (NHGIS) Database at the University of Minnesota.<sup>7</sup>

**State Capitals** The data include a binary indicator for capital status that is unity if a CBSA is a state capital, and zero otherwise. Note that neither the capital of Vermont (Montpelier), nor the capital of Maryland (Annapolis) have an official CBSA definition. I add both to the data and use population levels of 2010 that correspond to the municipal population of Montpelier and Annapolis, respectively.

**Road Network** Geographical information on the U.S. road network in 2017 is provided by the Natural Earth database.<sup>8</sup> The road network includes major highways, secondary highways, minor roads and ferry routes. In the analysis, I concentrate on major and secondary highways, which broadly define the National Highway System (NHS). Quantifying distances between CBSAs requires a mapping of the geographical division to a single departure or destination point. In the economic geography literature it is customary to simply assume the centroid of an area. Given that the connection between CBSAs is of major interest to the analysis in this paper, I create a point measurement that represents the largest concentration of population in a CBSA (i.e., maximum density point) using ArcGIS. In comparison to the centroid of a CBSA, the maximum density point has the advantage that it represents the point from which most people (in expectation) commute or migrate from and therefore reduces a potential measurement error in establishing the distance between each CBSA pair.

## Measuring Road Network Integration

Road network integration is defined as a class of measurements that evaluate how

---

<sup>7</sup>Historical population records report information of any settlement above 2,500 inhabitants.

<sup>8</sup>Natural Earth is a public domain supported by the North American Cartographic Information Society. I use version 4.0.0. of the database, which got released in 2017.

well a location is integrated in the National Highway System. In total, I consider four different measures: connectivity and market access, which are based on absolute distances between locations, and relative connectivity and relative market access, which are based on relative distances between locations. All measures have in common that their value for a given CBSA is the sum over connections to all other locations. While connectivity and market access define a connection as inverse (absolute) distance on the network, relative connectivity and relative market access evaluate the distance on the network relative to the great-circle distance. Hence, relative distances measure network integration in terms of how directly two locations are connected.

In defining the four measures, I denote  $d_{od}$  as the shortest distance between an origin ( $o$ ) and a destination ( $d$ ) using the transportation network. Given that most CBSAs are not directly located on the transportation network,  $d_{od}$  is further defined as

$$d_{od} \equiv \varphi(d_{oN_o} + d_{dN_d}) + d_{N_oN_d}, \quad (4.1)$$

where  $d_{oN_o}$  (and  $d_{dN_d}$ ) indicate the straight-line distance from location  $o$  (and  $d$ ) to the transportation network  $N_o$  (and  $N_d$ ). Distances to the transportation network are adjusted by a common factor of  $\varphi = 1.4$ , adding an over-proportional cost if a location is far away from the transportation network (as in Donaldson and Hornbeck, 2016). The shortest distance through the transportation network is denoted as  $d_{N_oN_d}$ .

In the following, I discuss the theoretical foundation as well as the mathematical definition of each measurement.

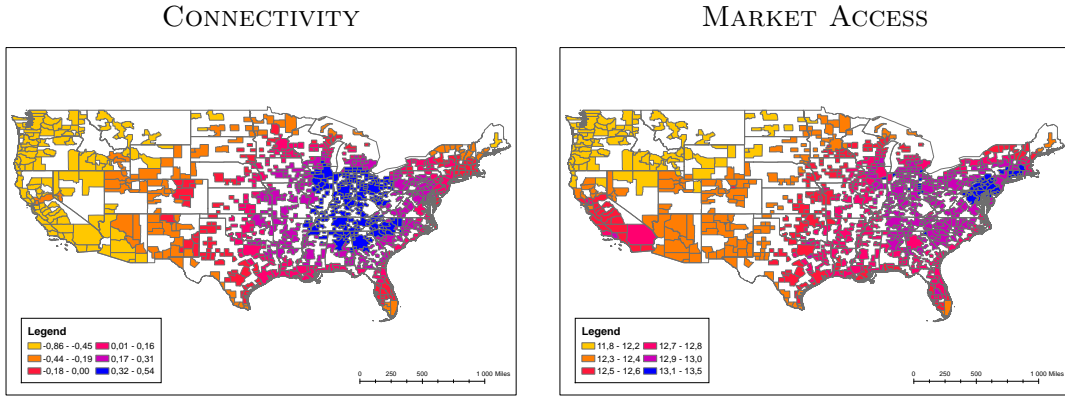
**Absolute Distance Measures** Absolute distances between locations matter for trade and migration. A first attempt to formally define how well a CBSA is integrated in the road network is to aggregate the inverse of all bilateral distances. Denote connectivity as  $Connect_o$ , then,

$$Connect_o = \sum_{d \neq o} (1/d_{od}). \quad (4.2)$$

Higher values of  $Connect_o$  indicate smaller aggregate distances to all locations

and, hence a better road network integration. A potential concern with the connectivity measure is that it is entirely dependent on the geographical position of a location in space. That is to say, connectivity is naturally larger for those CBSAs that are more centrally located in the road network, as compared to similar sized CBSAs at the border of the U.S. territory. To add another (important) dimension to the measure of network integration, one could account for the fact that some connections are economically more valuable than others. For instance, trade theory predicts that proximity to larger markets increases the probability to trade, and hence fosters economic growth. Consequently, considering a measure of market access addresses the economic value of transport connections.

Figure 4.3: ABSOLUTE DISTANCE MEASURES



*Note:* Both measures are reported in logs.

In the new economic geography literature, market access plays a major role in explaining the spatial distribution of economic activity (starting with Krugman, 1991). Donaldson and Hornbeck (2016) derive a first-order approximation for market access from general equilibrium trade theory, which offers an easy application in reduced form analysis. In particular, their approximation defines market access as the sum over the cost of trading with each other location and the other location's population. I follow them and define trade costs as the shortest distance on the network ( $d_{od}$ ), assuming the elasticity of distance to trade costs to be unity. Then market access,  $MA_o$  can be formulated as

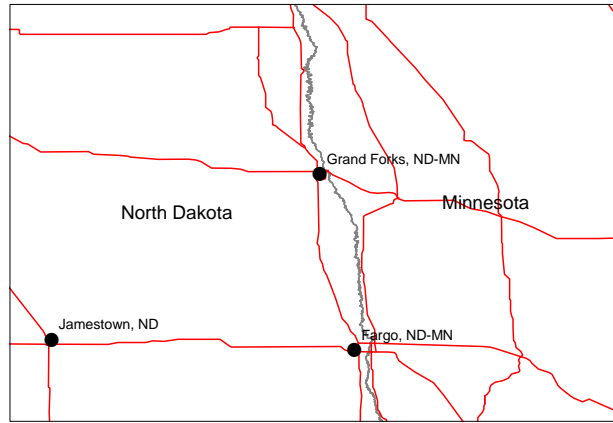
$$MA_o = \sum_{d \neq o} (L_d/d_{od}), \quad (4.3)$$

where  $L_d$  is the population level at destination  $d$ .

Figure 4.3 shows maps of both absolute distance measures at the CBSA level. As expected, connectivity levels are highest in central north-eastern CBSAs and gradually decrease as one moves towards the U.S. national border. The map on market access shows a similar pattern, however, the highest levels are shifted towards highly-populated CBSAs at the north-eastern coast (around Boston and New York). Also, market access levels are high along the south-western coast, due to a large number of high-density places in California.

**Relative Distance Measures** Contrary to absolute distance measures, relative distance measures evaluate a road network connection relative to the great-circle distance. Essentially, the relative distance captures how *direct* a connection is between a location pair. Figure 4.4 gives an example of a direct connection and an indirect connection. While the connection from Jamestown to Fargo is almost following a straight line, the connection from Jamestown to Grand Folks requires a detour via Fargo.

Figure 4.4: DIRECT VS. INDIRECT CONNECTION



The more a location is directly connected to others the better is its relative connectivity. Formally, relative connectivity is denoted as  $Connect_o^R$  and defined as

$$Connect_o^R = \sum_{d \neq o} (d_{od}^{GC} / d_{od}), \quad (4.4)$$

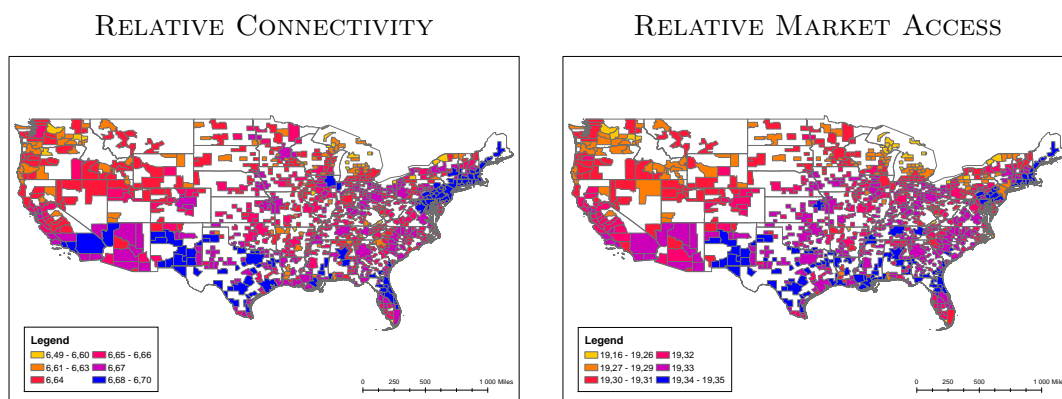
where  $d_{od}^{GC}$  describes the great-circle distance between origin  $o$  and destination  $d$ . The ratio of great-circle distance to network distance,  $(d_{od}^{GC} / d_{od})$ , ranges between

0 and 1 for all origin-destination pairs. A ratio close to unity implies that the network distance follows closely the straight-line between the connected places. Weighting the relative connection between two locations by size of the destination market adds an economic value to the relative connection. Hence, the fourth measurement combines market size with the relative distance and formally defines relative market access,  $MA_o^R$ , as

$$MA_o^R = \sum_{d \neq o} L_d (d_{od}^{GC} / d_{od}). \quad (4.5)$$

Figure 4.5 shows maps of relative distance measures at the CBSA level. Two things stand out. First, relative distance measures are less dependent on the geographical position of a CBSA. While there is still a concentration of high levels along the north-eastern coast, relative distance measures show a more significant within-state variation across all U.S. states. Second, both relative connectivity and relative market access identify road network transportation hubs within states. For instance, the centrally located CBSAs in Texas and Alabama are relatively better integrated in the federal road network as those closer to the state border.

Figure 4.5: RELATIVE DISTANCE MEASURES

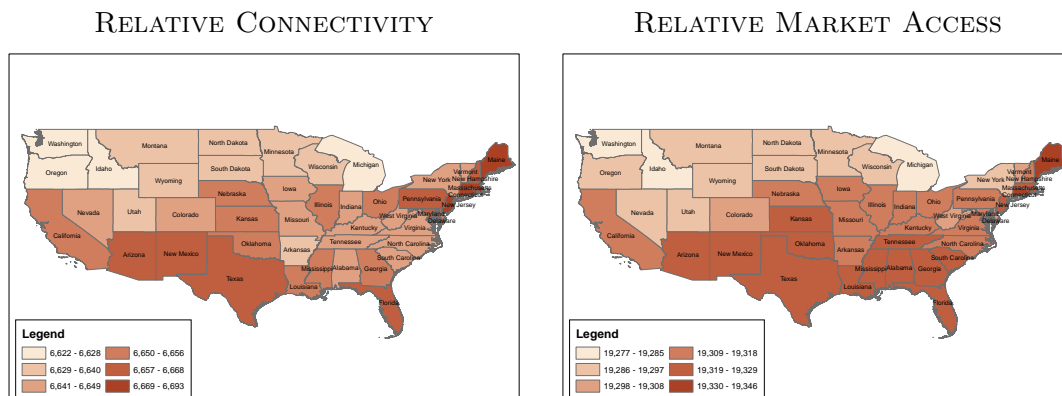


*Note:* Both measures are reported in logs.

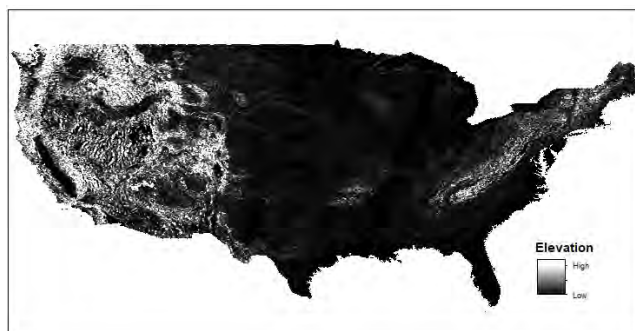
A potential concern with relative distance measures is that a direct connection between location pairs may be crucially dependent on natural features surrounding them. In other words, connecting CBSAs that are located in, say, the Rocky Mountains may require a large deviation from the great-circle distance solely due to terrain ruggedness. To show that this concern does not systematically affect

the relative distance measures, I calculate the average relative connectivity and average relative market access by state and plot the outcome in Figure 4.6.<sup>9</sup> For a better comparison, I add an elevation heat map in the lower panel of Figure 4.6.

Figure 4.6: RELATIVE DISTANCE MEASURES (AVERAGE BY STATE)



Note: Both measures are reported in logs.



The maps in Figure 4.6 suggest that terrain ruggedness is not a strong determinant of (average) relative distance measures. Neither states in the Rocky Mountains (e.g., Idaho, Montana, Wyoming, Colorado, New Mexico), nor those located along the Appalachian Mountains (e.g., Maine, New Hampshire, Pennsylvania, Virginia) have systematically lower levels in relative distance measures than neighboring states that are not located in the mountains. Instead, there is a clear north-south divide in relative distance measures. To some extent, this

<sup>9</sup>An alternative approach would be to use Dijkstra’s (1959) optimal route (algorithm) instead of the great-circle distance. The advantage of using Dijkstra’s algorithm would be that topological features enter as inputs into the design of the optimal route. However, the performance of the algorithm relies heavily on how building costs are specified, which is the topic of a large body of literature in engineering and transport design. The possibility to build tunnels, bridges, bypass segments, etc. typically complicates the choice of the appropriate specification. In comparison, the great-circle distance is a simple and intuitive measure. This modeling choice is further supported by the fact that that terrain ruggedness is indeed *not* a strong determinant of relative distance measures in the present application, as revealed by Figure 4.6.

captures the higher density of the National Highway System in the Northeast and the South, as compared to the Northwest (compare Figure 4.2).

## 4.4 Empirical Strategy

This section outlines the empirical strategy with a particular emphasis on the Instrumental Variable (IV) design, which allows – given its validity – a causal interpretation of capital status on road network integration.

### 4.4.1 Identification

I model the effect of capital status on road network integration using a log-linear specification, that is

$$\log Y_o = \beta \text{Capital}_o + \mathbf{X}_o \gamma + \varepsilon_o, \quad (4.6)$$

where  $Y_o$  is one of the road network integration outcomes at location  $o$ ,  $\text{Capital}_o$  describes a binary indicator that is one if a location is a state capital and zero otherwise,  $\mathbf{X}_o$  is a vector of covariates of interest, and  $\varepsilon_o$  is the error term. The vector of covariates includes log population levels in 2010 ( $\log L_o^{2010}$ ), log area size ( $\log A_o$ ), absolute values of longitude and latitude ( $|lat_o|, |lon_o|$ ), and binary indicators for the four large U.S. regions ( $Northeast_o, South_o, Midwest_o, West_o$ ).<sup>10</sup> Retrieving an unbiased estimate for  $\beta$  using Ordinary Least Squares (OLS) requires that capital status was randomly assigned. However, given the historical background of capital selection this assumption would be strong. Even though some capital locations could be defended as (quasi-) random, the location choice for most capitals was closely related to the westward expansion of the U.S. The main concern of endogeneity is that historical routes, such as exploration routes, were a strong determinant for both capital location and today’s transport network, which affects in turn the measures of road network integration. Duranton and Turner (2012) provide empirical evidence that historical exploration routes

---

<sup>10</sup>The mapping of each CBSA to one of the four large U.S. regions follows Caselli and Coleman (2001). For reasons of collinearity, only three out of four U.S. region indicators are included in the estimation.

in the U.S. (between 1528-1850) are a strong predictor for nowadays' transport connections. Montès (2014) provides anecdotal evidence that capitals were often located at mercantile gateways (Fact 5) along which the historical road network expanded. Consequently, the error term  $\varepsilon_o$  is likely correlated with capital status,  $Capital_o$ , leading to a biased estimate for  $\beta$ .

#### 4.4.2 Instrumental Variable Design

To address the endogeneity concern, I construct an instrument for capital status. The instrument exploits the fact that (most) state capitals were chosen for their geographically central and easily accessible location relative to other population clusters in their jurisdiction.<sup>11</sup> I replicate this pattern by employing a heuristic algorithm that predicts the boundaries of 48 U.S. states based on historical U.S. county information and define the geographic center of each predicted U.S. state as the hypothetical capital location.

Formally, the construction of the instrument proceeds in three steps.

**Step 1: Predicting Geography-based Population Density** In order to inform the heuristic algorithm, I predict geography-based population density in 1900 at the county level.<sup>12</sup> Extracting the variation in population distribution that is due to geographical features addresses a potential endogeneity concern, in which the population distribution of 1900 was partially determined by the location of the historical road network. As geographical features I use three measures: (i) the average gradient of a county  $c$  ( $G_c$ ), (ii) the distance to the first arrival ( $F_c$ ), and (iii) the distance to a river ( $R_c$ ). The average gradient of a county is based on gridded elevation data from the U.S. Geological Survey. It addresses the feasibility of settlement given topographical constraints. The distance to first arrival is a (straight-line) distance between the county's centroid and the nearest first arrival point of European settlers on the eastern coast and on the western

---

<sup>11</sup>In the empirical analysis, I abstract from changes in capital city location prior to choosing the permanent capital.

<sup>12</sup>Historical population records are only available at the county level. In some cases historical counties were significantly larger than nowadays CBSAs. To avoid measurement error, I refrain from aggregating the historical county level information to the CBSA level.



coast, respectively.<sup>13</sup> It accounts for the gradual evolution of U.S. settlement from the coast to the center of the territory. The distance to the closest river is a (straight-line) distance between a county’s centroid and the nearest river based on Natural Earth data. It reflects the importance of population clusters close to good trading opportunities. To predict geography-based population density in 1900, I define the following log-linear specification:

$$\log \bar{L}_c^{1900} = \alpha_{1r} \log G_c + \alpha_{2r} \log F_c + \alpha_{3r} \log R_c + \epsilon_c \quad \text{with } G_c, F_c, R_c > 0, \quad (4.7)$$

where the subscript  $r$  on either coefficient  $\{\alpha_{1r}, \alpha_{2r}, \alpha_{3r}\}$  stands for region. It describes one of the four large U.S. regions: the Northeast, the South, the Midwest and the West.<sup>14</sup> I estimate (4.7) for each region separately by OLS and present the estimation outcome in Table 4.2. As expected, all geography measures are negatively related to population density in 1900. Moreover, all measures are highly relevant for counties located in the Northeast, the South and the Midwest. For counties in the West, only the distance to the first arrival is statistically significant.

Table 4.2: GEOGRAPHY-BASED POPULATION DENSITY IN 1900

	(1)	(2)	(3)	(4)
$\log \bar{L}_c^{1900}$	Northeast	South	Midwest	West
$\log G_c$	-0.471*** (0.090)	-0.147*** (0.041)	-0.475*** (0.063)	-0.128 (0.144)
$\log F_c$	-0.449*** (0.114)	-1.595*** (0.115)	-2.163*** (0.107)	-0.699*** (0.121)
$\log R_c$	-0.136** (0.055)	-0.262*** (0.038)	-0.225*** (0.035)	-0.086 (0.066)
Obs.	243	1253	1023	315
Adj R <sup>2</sup>	0.164	0.196	0.277	0.139

Notes: Robust standard errors in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Figure 4.7: PREDICTED VS. OBSERVED POPULATION DENSITY

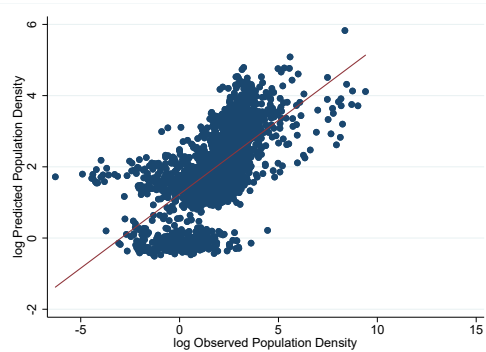


Figure 4.7 plots the predicted log population density against the observed log population density and highlights the fitted values as a red line. It suggests that the three geography measures replicate a large share of the variation in observed population density. The correlation coefficient between both is 0.64.

<sup>13</sup>The first arrival point on the east coast is Jamestown, Virginia. The first arrival point on the west coast is San Francisco, California.

<sup>14</sup>I assign each county to one of the four large U.S. regions (the Northeast, the South, the Midwest and the West) following the mapping of Caselli and Coleman (2001).

**Step 2:  $k$ -means Clustering Algorithm** I use the predicted (log) population density as weights in a  $k$ -means clustering algorithm. In my specific application, the  $k$ -means clustering provides an answer to the following question: If a central planner had to draw the borders of 48 U.S. states according to the (weighted) location of U.S. counties in 1900, where would she draw them? Mathematically,  $k$ -means clustering partitions observed counties  $c \in C$  into  $k = 48$  sets (with  $k \leq C$ ) based on (weighted) county coordinate information. The objective is to choose 48 clusters so as to minimize the within-cluster variance. Thereby, I assume the number of clusters as exogenously given, knowing that continental U.S. is composed of 48 contiguous states.<sup>15</sup> Formally, the algorithm solves the following optimization problem:

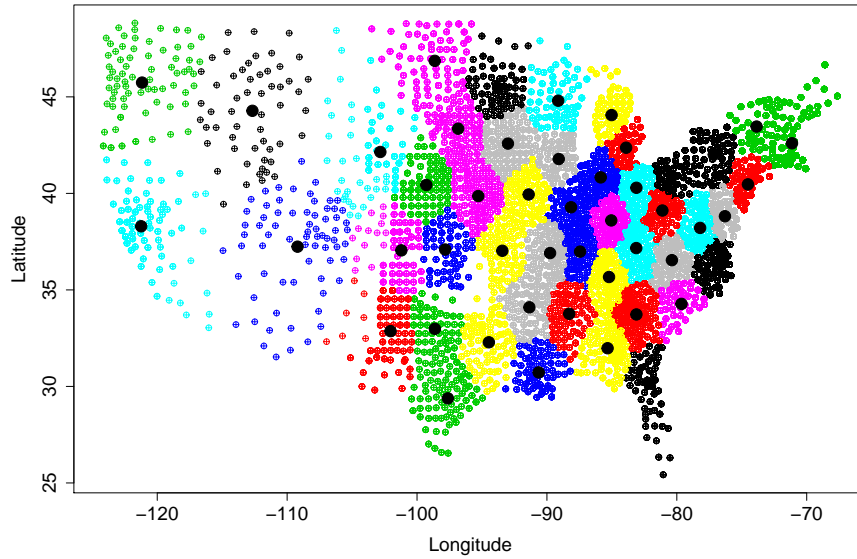
$$\operatorname{argmin}_S \sum_{i=1}^k \sum_{x_c \in S_i} \|x_c - x_{\mu_i}\|^2, \quad (4.8)$$

where  $x_c$  denotes the coordinate point of an observed county  $c$ , and  $x_{\mu_i}$  denotes the  $k$ -mean coordinate point of any set  $S = \{S_1, S_2, \dots, S_k\}$ , over which the algorithm optimizes.  $K$ -means clustering is a non-deterministic polynomial-time problem, which implies that it is computationally difficult (if not impossible) to determine a global optimum. However, applying an efficient heuristic algorithm allows converging quickly to a local optimum. The hypothetical capital locations are then defined as the resulting cluster centers (i.e.,  $k$ -mean coordinates). Figure 4.8 shows a cluster plot, which identifies the 48 U.S. states in different colors and marks the hypothetical capital location as the cluster centers in black dots.

---

<sup>15</sup>By construction, the  $k$ -means clustering algorithm improves within-cluster variance as the number of clusters  $k$  increases. Taking it to the extreme, the within-cluster variance is *optimal* if each data point is assigned to its own cluster. While this is not a desirable outcome, the data science literature has developed various methods to identify the *appropriate* number of clusters (see Kaufman and Rousseeuw, 1990). The most common methods (i.e., the elbow method or the silhouette method) are based on the idea that adding another cluster is only appropriate if the marginal gain in variance minimization is significantly large enough. In my data, the most appropriate number of clusters according to the silhouette method is  $k = 2$ .

Figure 4.8: PREDICTED U.S. STATES AND HYPOTHETICAL CAPITAL LOCATIONS



**Step 3: Construction of the Final Instrument** I spatially assign each CBSA in the data to the predicted U.S. state in which the CBSA lies.<sup>16</sup> Once each CBSA is mapped to a predicted U.S. state ( $k$ -cluster), I calculate the (straight-line) distance to the respective hypothetical capital. I then rank all CBSAs within a predicted U.S. state according to their (straight-line) distance to the hypothetical capital location; and denote this variable as  $Rank_o$ . In the first stage regression,  $Rank_o$  serves as an instrument that predicts the binary indicator  $Capital_o$ . Formally, I estimate the following first stage specification:

$$Capital_o = \rho Rank_o + \mathbf{X}_o \gamma + \varepsilon_o^{First}. \quad (4.9)$$

Table 4.3 presents the first stage results. The instrument is highly relevant and, as expected, it shows a negative sign, which implies that capital cities are in fact nearer located to hypothetical capital locations. The Kleinbergen-Paap F-Statistic for a weak instrument test can be rejected at the 5% significance level.<sup>17</sup> The strength of the instrument is further supported by Figure 4.9. It plots the

<sup>16</sup>The exact location of the CBSA is determined by its maximum density point (see Section 4.3).

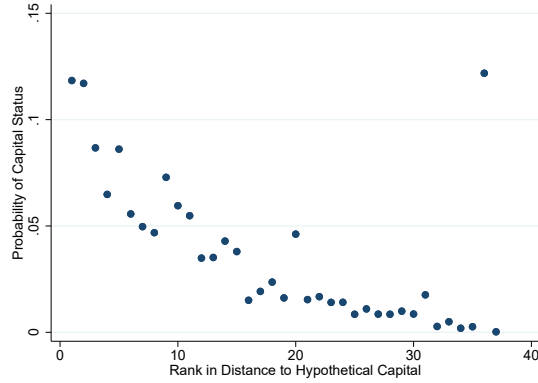
<sup>17</sup>Stock and Yogo (2005) report critical values at which the weak instrumentation test can be rejected. The critical value is a function of the number of included endogenous regressors, the number of instrumental variables, and the desired maximum bias relative to OLS. In my case, for one endogenous regressor, one instrumental variable and a maximum relative bias of 5%, the critical value is 16.38.

Table 4.3: FIRST STAGE RESULTS

	<i>Capital<sub>o</sub></i>
<i>Rank<sub>o</sub></i>	-0.003*** (0.001)
Number of <i>k</i>	48
Observations	920
Adj. R <sup>2</sup>	0.12
F-Stat Weak Inst	17.75

Notes: State clustered and robust standard errors in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The regression includes the list of covariates stated in Section 4.4.1.

Figure 4.9: PREDICTED CAPITAL STATUS VS. RANK



Note: The graph shows the average probability of capital status by rank.

probability of being a capital against its rank in distance to the hypothetical capital location and shows a clear negative correlation between the two.

### 4.4.3 Results

Table 4.4 reports the second stage results from estimating (4.6) for all network integration outcomes. For each outcome, I compare the results of the IV specification to the simple OLS estimate. For brevity, coefficients of all included covariates are suppressed in the main table. The interested reader can find the full table in the Appendix (see Table 4.9).

Table 4.4: SECOND STAGE RESULTS

	<b>Absolute Distances</b>				<b>Relative Distances</b>			
	log <i>Connect<sub>o</sub></i>		log <i>MA<sub>o</sub></i>		log <i>Connect<sub>o</sub><sup>R</sup></i>		log <i>MA<sub>o</sub><sup>R</sup></i>	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
<i>Capital<sub>o</sub></i>	0.022 (0.025)	0.298 (0.652)	-0.017 (0.026)	0.470 (0.484)	0.005** (0.002)	0.151*** (0.047)	0.008*** (0.002)	0.152*** (0.056)
Observations	920	920	920	920	920	920	920	920
Adj. R <sup>2</sup>	0.91	0.61	0.29	0.17	0.87	0.80	0.40	0.34

Notes: State clustered and robust standard errors in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . All regressions include the list of covariates stated in Section 4.4.1.

When estimating the model with OLS, the effect of capital status on road network integration outcomes is small in magnitude relative to the IV specification. This first finding is surprising. If the capital selection process did, as expected, favor cities that were already well connected, OLS should over-estimate the true capital effect. The small magnitude of the OLS estimates is actually more in line with

the historical balance-of-power hypothesis. This hypothesis stipulates that capital status was intentionally attributed to smaller, (initially) less well connected cities in order to spatially separate political and economic centers of power.

Once the endogenous binary indicator *Capital<sub>o</sub>* is instrumented, the effect gets larger in magnitude for all outcomes and highly significant for relative distance measures. Relative distances measures include relative connectivity and relative market access. The effect of capital status on both is very similar in terms of magnitude and significance. Hence, weighting each relative distance connection by size of the destination market does not change the capital effect overall. In particular, the IV regression result suggests that capital cities have on average about 14 percent larger levels of relative connectivity and relative market access as compared to non-capital cities of similar characteristics.<sup>18</sup> While this result does not have a straight-forward economic interpretation, it confirms that capitals are on average more directly integrated in the road network than non-capital cities.

Absolute distance measures, on the other hand, have a straight-forward economic interpretation. Trade theory suggests that larger levels of connectivity and – even more so of market access – imply better trading opportunities and hence economic prosperity. In the main empirical result, I find a positive though insignificant effect of capital status on connectivity and market access. A possible explanation for why the effect is insignificant could be due to the definition of the absolute distance measures. Both measures are very concentrated in some regions of the U.S. and their magnitude is heavily dependent on how centrally located the CBSA is in the overall National Highway System (see Figure 4.3). State capitals, however, are naturally very spread out across the entire country. Consequently, the spatial variation in capital city locations is not captured enough by the concentrated measures of connectivity and market access.

---

<sup>18</sup>Halvorsen and Palmquist (1980) provide a review on the interpretation of dummy variables in semilogarithmic equations. The effect is calculated as  $g = 100 \times (\exp(b - V(b)/2) - 1)$ , where  $g$  is the effect in percent,  $b$  is the coefficient on the dummy variable and  $V(b)$  is the variance of the coefficient of the dummy variable.

## 4.5 Discussion

The estimated effect of capital status on road network integration measures is an outcome of a pooled regression, which combines CBSAs of heterogeneous states and capital cities that have been selected for many different reasons. This section sheds further light on the drivers of the effect and provides evidence for plausible mechanisms behind the results. The subsequent analysis concentrates on relative market access as main outcome variable.

**Drivers** To understand the drivers of the capital effect on relative market access, I construct a number of state-level binary indicators that classify the sample according to general and historical characteristics.

The general characteristics include information on the state urbanization rate, state size and the size of the capital city. Regarding the urbanization rate and state size, I calculate the 50<sup>th</sup> percentile of the entire distribution and define each binary indicator as unity if a CBSA is located in a state with *below* median urbanization rate and state size, respectively. Regarding size of the capital city, I define the binary indicator as unity if a CBSA is located in a state in which the capital city is *not* the largest city.

The historical characteristics include information on the spatial patterns of capital migration (see historical background, Section 4.2). I define four binary indicators for either type of spatial pattern: Westward/Centrality, Rotation, Readjustment and Other.<sup>19</sup> Each indicator is unity if a CBSA is located in a state in which the respective spatial pattern was *not* prevalent.

In separate analyses, I use one binary indicator at the time as an interaction term with capital status, to single out its importance in the overall effect. The empirical specification is as follows

$$\log MA_o^R = \tilde{\beta} Capital_o + \delta_1 Indicator_o + \delta_2 Capital_o \times Indicator_o + \mathbf{X}_o \gamma + \epsilon_o, \quad (4.10)$$

where  $Indicator_o$  describes one of the previously mentioned binary indicators.

---

<sup>19</sup>The category *Other* includes all states that have never changed capital, and those that had other reasons for capital migration than westward/centrality, rotation or readjustment.

Then, the coefficient  $\tilde{\beta}$  is the effect of capital status on relative market access conditional on  $Indicator_o$  being zero. For this reason, I have defined each indicator in its reverse sense, implying that they are zero for the attribute they are analyzed for.

Table 4.5: ESTIMATION RESULTS – DRIVERS

PANEL A: GENERAL CHARACTERISTICS				
Dep. Var. $\log(MA_o^R)$	(1)	(2)	(3)	
	<b>Capital is Largest City</b>	<b>Above Median State Size</b>	<b>Above Median Urbanization Rate</b>	
$Capital_o$ ( $\tilde{\beta}$ )	0.297** (0.134)	0.454* (0.236)	0.189*** (0.0701)	
PANEL B: HISTORICAL CHARACTERISTICS				
Dep. Var. $\log(MA_o^R)$	(1)	(2)	(3)	(4)
	<b>Westward/Centrality</b>	<b>Rotation</b>	<b>Readjustment</b>	<b>Other</b>
$Capital_o$ ( $\tilde{\beta}$ )	0.255** (0.102)	5.230 (10.98)	2.008 (1.761)	1.105 (0.864)
Observations	920	920	920	920

*Notes:* State clustered and robust standard errors in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . All regressions include the list of covariates stated in Section 4.4.1.  $Capital_o$  and  $Capital_o \times Indicator_o$  are instrumented with  $Rank_o$  and  $Rank_o \times Indicator_o$ .

Table 4.5 presents the results from estimating (4.10) as IV regression. The table is divided in two panels and each column in a panel is named after the indicator that is analyzed.<sup>20</sup> When looking at general state characteristics, the results in panel A suggest that the capital effect is driven by large, urbanized states in which the capital is the largest city. Moreover, panel B suggests that the historical decision on state capital centrality within the state is a driver of the capital effect, while other spatial patterns are not.

**Centrality** Throughout the paper, centrality has played an important role in characterizing U.S. state capitals. By far the most common spatial pattern that decided on the capital location was (geographical) centrality. The employed instrument, that is highly relevant in predicting capital status, is fundamentally based on the idea of (demographic) centrality. The effect of capital status on relative market access is (partially) driven by large U.S. states, where centrality is key to governing the political jurisdiction. In short, centrality matters.

<sup>20</sup>For brevity, Table 4.5 shows only the estimate of interest,  $\tilde{\beta}$  (see Table 4.10 in the Appendix, for full information).

Additional evidence that underlines the centrality argument could be to rerun the analysis with an alternative instrument that is based on geographical centrality within the actual state borders. To do so, I rank each CBSA by its distance to the state centroid and denote this instrument as  $\widetilde{Rank}_o$ . Table 4.6 contrasts the main estimation result in column (1) to the estimation result with the alternative instrument in column (2). The results suggest that state centrality – based on actual U.S. state borders – is as relevant to predicting the capital location and the capital effect remains positive and significant, though smaller in magnitude.

Table 4.6: ESTIMATION RESULTS – ALTERNATIVE INSTRUMENT

<b>First Stage</b>				
		(1)	$\widetilde{Rank}_o$	(2)
Instrument	$Rank_o$	-0.003***		-0.003***
		(0.001)		(0.001)
Dep. Var. $\log(MA_o^R)$		<b>Second Stage</b>		
$Capital_o$		0.152***		0.039*
		(0.056)		(0.020)
Observations		920		920
F-Stat Weak Inst		17.75		11.94

*Notes:* Standard errors in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . All regressions include the list of covariates stated in Section 4.4.1. For one endogenous regressor, one instrumental variable and a maximum relative bias of 5%, the critical value for the weak instrumentation F-Statistic is 16.38 (Stock and Yogo, 2005).

Even though U.S. states and their capital cities are highly heterogeneous, their common feature of state capital centrality is the main mechanism that explains a better – and effectively more direct – road network integration. On the one hand, the Christaller’s (1933) Central Place Theory suggests that an efficient road network radially expands around the most central place on top of the hierarchy. On the other hand, even if the capital city is not the largest, most important urban center, its central location favors a better road network integration. This is because, a network that connects places across the entire jurisdiction passes on average more often by the geographic center. Along these lines, Faber (2014) provides evidence that some peripheral places in China have been comparatively well integrated in the National Trunk Highway System due to an *on-the-way* treatment between targeted metropolitan areas.



**Connection to Major Urban Centers** The second mechanism that may explain the capital effect on (direct) road network integration is related to political interest representation. Conceptually, this mechanism is hard to test for as political influence in road network provision is (nearly) impossible to measure. However, there is one plausible argument that interest groups could have defended. As capitals are places of power and decision making, it may be of interest to integrate capital cities well with economically important urban areas around. To test this hypothesis, I construct an alternative measurement of relative market access,  $\widetilde{MA}_o^R$ , which considers only connections to the 50 largest CBSAs – i.e., those with more than one million inhabitants.

Table 4.7: ESTIMATION RESULTS – ALTERNATIVE MEASURE OF RELATIVE MARKET ACCESS

	(1)	(2)
	$\log MA_o^R$	$\log \widetilde{MA}_o^R$
<i>Capital<sub>o</sub></i>	0.152*** (0.056)	0.147** (0.060)
Observations	920	920

*Notes:* Standard errors in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . All regressions include the list of covariates stated in Section 4.4.1.

Table 4.7 compares the main estimation result in column (1) with the effect on the alternative measurement of relative market access in column (2). The comparison shows that the capital effect still holds when only considering connections to the main urban centers, even though the coefficient is slightly smaller in magnitude.

## 4.6 Conclusion

This paper links the political status of U.S. urban areas to their integration in the National Highway System (NHS) in order to understand whether there is a *capital premium* in road network provision. I document historical patterns of U.S. state capital selection and use the common feature of geographical centrality to construct an instrument for the endogenous capital location. In particular, the IV

design is based on a  $k$ -means clustering algorithm that predicts the boundaries of 48 U.S. states and defines the geographical center as a hypothetical capital location. I then estimate the causal effect of capital status on four outcomes of road network integration. Two outcome measures (connectivity and market access) evaluate the strength of integration based on the aggregate proximity to all other locations. The other two outcomes (relative connectivity and relative market access) measure how directly connected a location is to all others. I find significant and robust evidence that capital cities are more *directly* integrated in the NHS compared to non-capital cities of similar characteristics. The reason for this finding is a combination of two aspects. First, (most) capital cities have a favorable geographical position within their state. This makes them a natural candidate for a direct road network integration according to the Central Place Theory. And second, as capital cities are places of political power and decision-making, there is a governmental interest in establishing direct connections to other major urban areas. Given that the decision on the location of the federal highway network was subject to inter-governmental negotiations, this interest likely played in favor of capital cities.

# Appendix

## 4.7 Supplement Tables

Table 4.8: STATE CAPITALS - BASIC FACTS

State	Capital	Year chosen	Population		Population (Census 2010)			Average Annual Wage (2018)		
			Rank when chosen	Rank in 2010	CBSA	State	% State	CBSA	% State	Rank in 2018
Alabama	Montgomery	1847	2	4	374,536	4,779,736	7.84	42,510	97.08	4
Alaska	Juneau	1900	2 or 3	3	31,275	710,231	4.40	N/A		
Arizona	Phoenix	1889	2	1	4,192,887	6,392,017	65.60	50,520	102.50	1
Arkansas	Little Rock	1820	1	1	699,757	2,915,918	24.00	45,020	108.38	2
California	Sacramento	1854	2	5	2,149,127	37,253,956	5.77	56,430	95.40	5
Colorado	Denver	1868	1 or 2	1	2,543,482	5,029,196	50.57	59,440	106.49	2
Connecticut	Hartford	1873	1	1	1,212,381	3,574,097	33.92	60,820	100.07	2
Delaware	Dover	1781	1 or 2	1	162,310	897,934	18.08	45,120	84.62	1
Florida	Tallahassee	1823	3	12	367,413	18,801,310	1.95	45,340	98.54	7
Georgia	Atlanta	1868	2	1	5,286,728	9,687,653	54.57	52,750	109.26	1
Hawaii	Honolulu	1900	1	1	953,207	1,360,301	70.07	54,870	103.72	
Idaho	Boise City	1864	New Town	1	616,561	1,567,582	39.33	45,470	104.58	1
Illinois	Springfield	1837	2	5	210,170	12,830,632	1.64	50,560	94.00	5
Indiana	Indianapolis	1825	2	1	1,887,877	6,483,802	29.12	49,380	109.03	1
Iowa	Des Moines	1857	7	1	569,633	3,046,355	18.70	52,220	113.15	1
Kansas	Topeka	1861	4	2	233,870	2,853,118	8.20	44,010	97.20	2
Kentucky	Frankfort	1793	2	9	70,706	4,339,367	1.63	N/A		
Louisiana	Baton Rouge	1882	2	2	802,484	4,533,372	17.70	45,200	105.95	1
Maine	Augusta	1832	2	3	122,151	1,328,361	9.20	N/A		
Maryland	Annapolis	1694	N/A	4	38,394*	5,773,552	0.66	N/A		
Massachusetts	Boston	1692	1	1	4,552,402	6,547,629	69.53	67,370	105.41	1
Michigan	Lansing	1847	New Town	3	464,036	9,883,640	4.69	49,320	99.62	5
Minnesota	Minneapolis	1849	1	1	3,348,859	5,303,925	63.14	57,420	105.94	2
Mississippi	Jackson	1826	3	1	567,122	2,967,297	19.11	43,180	109.54	1
Missouri	Jefferson City	1822	New Town	6	149,807	5,988,927	2.50	42,120	90.66	5

*Continued.*

Table 4.8: *Continued.*

State	Capital	Year chosen	Population Rank when chosen	Population (Census 2010)			Average Annual Wage (2018)			
				Rank in 2010	CBSA	State	% State	CBSA	% State	Rank in 2018
Montana	Helena	1874	1 or 2	6	74,801	989,415	7.56	N/A		
Nebraska	Lincoln	1867	New Town	2	302,157	1,826,341	16.54	46,800	100.19	2
Nevada	Carson City	1861	2 or 3	3	55,274	2,700,551	2.05	50,840	110.11	1
New Hampshire	Concord	1808	2	3	146,445	1,316,470	11.12	N/A		
New Jersey	Trenton	1790	1	1	366,513	8,791,894	4.17	63,700	109.43	1
New Mexico	Santa Fe	1610	1	3	144,170	2,059,179	7.00	46,260	101.89	2
New York	Albany	1797	2	4	870,716	19,378,102	4.49	54,400	87.93	3
North Carolina	Raleigh	1792	New Town	2	1,130,490	9,535,483	11.86	52,580	111.40	2
North Dakota	Bismarck	1889	4	2	114,778	672,591	17.07	50,390	101.55	1
Ohio	Columbus	1812	New Town	2	1,901,974	11,536,504	16.49	51,260	106.30	1
Oklahoma	Oklahoma City	1910	1	1	1,252,987	3,751,351	33.40	47,120	106.56	1
Oregon	Salem	1860	3	2	390,738	3,831,074	10.20	48,790	93.83	3
Pennsylvania	Harrisburg	1812	5	5	549,475	12,702,379	4.33	49,540	99.02	3
Rhode Island	Providence	1900	1	1	1,600,852	1,052,567	152.09	53,730	98.03	1
South Carolina	Columbia	1790	New Town	2	767,598	4,625,364	16.60	44,680	103.40	2
South Dakota	Pierre	1889	5	9	21,361	814,180	2.62	N/A		
Tennessee	Nashville	1843	1	1	1,670,890	6,346,105	26.33	48,370	108.31	1
Texas	Austin	1839	New Town	4	1,716,289	25,145,561	6.83	53,810	108.23	3
Utah	Salt Lake City	1856	1	1	1,087,873	2,763,885	39.36	50,920	106.26	1
Vermont	Montpelier	1808	6 at most	12	7,855*	625,741	1.26	N/A		
Virginia	Richmond	1779	Village	2	1,208,101	8,001,024	15.10	51,330	92.80	2
Washington	Olympia	1853	Village	4	252,264	6,724,540	3.75	54,050	90.98	4
West Virginia	Charleston	1885	5	2	227,078	1,852,994	12.25	44,680	105.45	2
Wisconsin	Madison	1836	New Town	2	605,435	5,686,986	10.65	52,890	111.70	1
Wyoming	Cheyenne	1869	1	1	91,738	563,626	16.28	48,550	99.84	2

*Source:* Information on the year of capital choice and rank when chosen are taken from Montès (2014, p.4-5). Population records come from the U.S. Census Bureau, average annual wage estimates come from U.S. Bureau of Labor Statistics. *Notes:* CBSA stands for Core Based Statistical Area. "New Town" means the town was intended to become the capital when founded, "Village" means the town was very small when designated as the capital. \*For Annapolis and Montpelier, the population is the municipal population because they are not part of any officially defined CBSA.

Table 4.9: SECOND STAGE RESULTS - FULL TABLE

	<u>Absolute Distances</u>				<u>Relative Distances</u>			
	log $Connect_o$		log $MA_o$		log $Connect_o^R$		log $MA_o^R$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
$Capital_o$	0.022 (0.025)	0.298 (0.652)	-0.017 (0.026)	0.470 (0.484)	0.005** (0.002)	0.151*** (0.047)	0.008*** (0.002)	0.152*** (0.056)
$\log(L_o^{2010})$	0.027** (0.011)	0.014 (0.035)	0.079*** (0.014)	0.055* (0.030)	0.004*** (0.001)	-0.003 (0.002)	0.002*** (0.001)	-0.005* (0.003)
$\log(A_o)$	-0.091*** (0.019)	-0.098*** (0.026)	-0.154*** (0.019)	-0.166*** (0.027)	-0.000 (0.002)	-0.004 (0.003)	-0.003* (0.001)	-0.007* (0.003)
$ lon_o $	-0.014*** (0.003)	-0.014*** (0.003)	-0.014*** (0.003)	-0.014*** (0.003)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
$ lat_o $	-0.003 (0.008)	-0.004 (0.006)	-0.011* (0.007)	-0.013** (0.005)	-0.002*** (0.000)	-0.003*** (0.000)	-0.003*** (0.000)	-0.003*** (0.000)
$Northeast$	-0.132 (0.147)	-0.142 (0.147)	-0.021 (0.099)	-0.040 (0.101)	0.023** (0.011)	0.018 (0.012)	0.016* (0.009)	0.010 (0.008)
$South$	0.161 (0.104)	0.152* (0.092)	-0.096 (0.074)	-0.112 (0.071)	-0.002 (0.009)	-0.007 (0.009)	0.005 (0.007)	-0.000 (0.006)
$Midwest$	0.250*** (0.080)	0.246*** (0.076)	0.007 (0.057)	-0.001 (0.059)	0.008 (0.006)	0.005 (0.007)	0.011** (0.004)	0.009** (0.004)
Cons.	1.712*** (0.534)	1.949*** (0.498)	14.682*** (0.381)	15.102*** (0.332)	6.687*** (0.032)	6.813*** (0.052)	19.378*** (0.031)	19.503*** (0.046)
Observations	920	920	920	920	920	920	920	920

Notes: State clustered and robust standard errors in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .  $\log(L_o^{2010})$  is log population in 2010,  $\log(A_o)$  is log area size,  $\log(|lon_o|)$  and  $\log(|lat_o|)$  are log absolute values of longitude and latitude, respectively.  $Northeast_o$ ,  $South_o$  and  $Midwest_o$  are binary indicators that are unity if a CBSA is located in a state that belongs to the Northeast, the South and the Midwest, respectively.

Table 4.10: ESTIMATION RESULTS - DRIVERS (FULL TABLE)

PANEL A: GENERAL CHARACTERISTICS			
Dep.Var. $\log(MA_o^R)$	(1)	(2)	(3)
	<b>Capital is Largest City</b>	<b>Above Median State Size</b>	<b>Above Median Urbanization Rate</b>
$Capital_o (\tilde{\beta})$	0.297** (0.134)	0.454* (0.236)	0.189*** (0.070)
$Indicator_o$	0.018* (0.010)	0.022** (0.010)	0.008 (0.005)
$Capital_o \times Indicator_o$	-0.274** (0.129)	-0.420* (0.226)	-0.174*** (0.067)
$\log(L_o^{2010})$	-0.007* (0.004)	-0.010 (0.006)	-0.004 (0.003)
$\log(A_o)$	-0.007** (0.003)	-0.004 (0.007)	-0.001 (0.003)
$ lon_o $	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
$ lat_o $	-0.003*** (0.000)	-0.004*** (0.001)	-0.003*** (0.000)
$Northeast_o$	0.001 (0.011)	0.030** (0.013)	0.014 (0.009)
$South_o$	0.000 (0.006)	0.006 (0.012)	0.005 (0.007)
$Midwest_o$	0.006 (0.006)	0.015 (0.010)	0.013*** (0.005)
Cons.	19.514*** (0.058)	19.530*** (0.108)	19.452*** (0.035)

*Continued.*

Table 4.10: *Continued.*

PANEL B: HISTORICAL CHARACTERISTICS					
Dep.Var. $\log(MA_o^R)$	(1)	(2)	(3)	(4)	
	<b>Westward</b>	<b>Centrality</b>	<b>Rotation</b>	<b>Readjustment</b>	<b>Other</b>
$Capital_o$ ( $\tilde{\beta}$ )	0.255** (0.102)	5.230 (10.982)	2.008 (1.761)	1.105 (0.864)	
$Indicator_o$	0.015*** (0.005)	0.538 (1.136)	0.127 (0.103)	0.048 (0.042)	
$Capital_o \times Indicator_o$	-0.240** (0.098)	-5.186 (10.920)	-1.965 (1.739)	-1.065 (0.842)	
$\log(L_o^{2010})$	-0.001 (0.002)	-0.012 (0.024)	-0.015 (0.016)	-0.012 (0.010)	
$\log(A_o)$	-0.010** (0.004)	-0.003 (0.012)	-0.008 (0.008)	-0.015 (0.011)	
$ lon_o $	0.000 (0.000)	-0.003 (0.006)	0.000 (0.000)	0.001 (0.001)	
$ lat_o $	-0.003*** (0.000)	0.000 (0.006)	-0.003*** (0.001)	-0.004*** (0.001)	
$Northeast_o$	0.007 (0.008)	-0.250 (0.543)	0.011 (0.019)	0.035* (0.019)	
$South_o$	-0.007 (0.006)	-0.020 (0.051)	-0.002 (0.012)	0.009 (0.016)	
$Midwest_o$	0.005 (0.005)	-0.076 (0.170)	0.003 (0.014)	0.014 (0.012)	
Cons.	19.487*** (0.044)	19.258*** (0.571)	19.517*** (0.153)	19.556*** (0.150)	
Observations	920	920	920	920	

*Notes:* State clustered and robust standard errors in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . All regressions include the list of covariates stated in Section 4.4.1.  $Capital_o$  and  $Capital_o \times Indicator_o$  are instrumented with  $Rank_o$  and  $Rank_o \times Indicator_o$ .  $\log(L_o^{2010})$  is log population in 2010,  $\log(A_o)$  is log area size,  $|lon_o|$  and  $|lat_o|$  are absolute values of longitude and latitude, respectively.  $Northeast_o$ ,  $South_o$  and  $Midwest_o$  are binary indicators that are unity if a CBSA is located in a state that belongs to the Northeast, the South and the Midwest, respectively.



# Bibliography

- AHLFELDT, G. M., S. J. REDDING, D. M. STURM, AND N. WOLF (2015): “The Economics of Density: Evidence From the Berlin Wall,” *Econometrica*, 83, 2127–2189.
- ALDER, S. AND I. KONDO (2018): “Political Distortions and Infrastructure Networks in China: A Quantitative Spatial Equilibrium Analysis,” 2018 Meeting Papers 1269, Society for Economic Dynamics.
- ALLEN, T. AND C. ARKOLAKIS (2014): “Trade and the Topography of the Spatial Economy,” *The Quarterly Journal of Economics*, 129, 1085–1140.
- (2019): “The Welfare Effects of Transportation Infrastructure Improvements,” NBER Working Papers 25487, National Bureau of Economic Research, Inc.
- ALLEN, T. AND D. DONALDSON (2018): “The Geography of Path Dependence,” Unpublished Manuscript.
- ARALICA, Z. AND V. BOTRIĆ (2013): “Evaluation of Research and Development Tax Incentives Scheme in Croatia,” *Economic Research-Ekonomska Istraživanja*, 26, 63–80.
- ARRIBAS-BEL, D., M.-A. GARCIA-LOPEZ, AND E. VILADECANS-MARSAL (2019): “Building(s and) cities: Delineating urban areas with a machine learning algorithm,” *Journal of Urban Economics*, 103217.
- ATKESON, A. AND A. BURSTEIN (2019): “Aggregate Implications of Innovation Policy,” *Journal of Political Economy*, 127, 2625–2683.

- AU, C.-C. AND J. V. HENDERSON (2006a): “Are Chinese Cities Too Small?” *Review of Economic Studies*, 73, 549–576.
- (2006b): “How migration restrictions limit agglomeration and productivity in China,” *Journal of Development Economics*, 80, 350–388.
- BAI, Y. AND R. JIA (2020): “The Economic Consequences of Political Hierarchy: Evidence from Regime Changes in China, AD1000-2000,” NBER Working Papers 26652, National Bureau of Economic Research, Inc.
- BANERJEE, A., E. DUFLO, AND N. QIAN (2012): “On the Road: Access to Transportation Infrastructure and Economic Growth in China,” NBER Working Papers 17897, National Bureau of Economic Research, Inc.
- BARWICK, P., D. DONALDSON, S. LI, AND Y. LIN (2018): “The Welfare Effects of Passenger Transportation Infrastructure: Evidence from China,” Unpublished Manuscript.
- BAUM-SNOW, N. (2007): “Did Highways Cause Suburbanization?” *The Quarterly Journal of Economics*, 122, 775–805.
- BAUM-SNOW, N., L. BRANDT, J. V. HENDERSON, M. A. TURNER, AND Q. ZHANG (2016): “Highways, Market Access and Urban Growth in China,” SERC Discussion Papers 0200, Spatial Economics Research Centre, LSE.
- (2017): “Roads, Railroads, and Decentralization of Chinese Cities,” *The Review of Economics and Statistics*, 99, 435–448.
- BAUM-SNOW, N., J. V. HENDERSON, M. A. TURNER, Q. ZHANG, AND L. BRANDT (2018): “Does investment in national highways help or hurt hinterland city growth?” *Journal of Urban Economics*.
- BAUM-SNOW, N. AND M. TURNER (2017): “Transport Infrastructure and the Decentralization of Cities in the People’s Republic of China,” *Asian Development Review*, 34, 25–50.

- BECKER, S. O., S. HEBLICH, AND D. M. STURM (2018): “The Impact of Public Employment: Evidence from Bonn,” CEPR Discussion Papers 12565, C.E.P.R. Discussion Papers.
- BERNARD, A. B., J. EATON, J. B. JENSEN, AND S. KORTUM (2003): “Plants and Productivity in International Trade,” *American Economic Review*, 93, 1268–1290.
- BLOOM, N. AND J. V. REENEN (2000): “Patents, productivity and market value: evidence from a panel of UK firms,” IFS Working Papers W00/21, Institute for Fiscal Studies.
- BLOOM, N., M. SCHANKERMAN, AND J. V. REENEN (2013): “Identifying Technology Spillovers and Product Market Rivalry,” *Econometrica*, 81, 1347–1393.
- BÖSENBERG, S. AND P. H. EGGER (2017): “R&D tax incentives and the emergence and trade of ideas,” *Economic Policy*, 32, 39–80.
- BRODA, C. AND D. WEINSTEIN (2006): “Globalization and the Gains From Variety,” *The Quarterly Journal of Economics*, 121, 541–585.
- BURCHFIELD, M., H. G. OVERMAN, D. PUGA, AND M. A. TURNER (2006): “Causes of Sprawl: A Portrait from Space,” *The Quarterly Journal of Economics*, 121, 587–633.
- CAIUMI, A. (2011): “The Evaluation of the Effectiveness of Tax Expenditures - A Novel Approach: An Application to the Regional Tax Incentives for Business Investments in Italy,” Tech. rep., OECD Taxation Working Papers, No. 5, OECD Publishing, Paris.
- CALIENDO, L. AND F. PARRO (2015): “Estimates of the Trade and Welfare Effects of NAFTA,” *Review of Economic Studies*, 82, 1–44.
- CALIENDO, L., F. PARRO, E. ROSSI-HANSBERG, AND P.-D. SARTE (2018): “The Impact of Regional and Sectoral Productivity Changes on the U.S. Economy,” *Review of Economic Studies*, 85, 2042–2096.

- CAPPELEN, A., A. RAKNERUD, AND M. RYBALKA (2012): “The effects of R&D tax credits on patenting and innovations,” *Research Policy*, 41, 334–345.
- CARLINO, G. A., S. CHATTERJEE, AND R. M. HUNT (2007): “Urban density and the rate of invention,” *Journal of Urban Economics*, 61, 389–419.
- CASELLI, F. AND W. J. COLEMAN (2001): “The U.S. Structural Transformation and Regional Convergence: A Reinterpretation,” *Journal of Political Economy*, 109, 584–616.
- CHANDRA, A. AND E. THOMPSON (2000): “Does public infrastructure affect economic activity?: Evidence from the rural interstate highway system,” *Regional Science and Urban Economics*, 30, 457–490.
- CHRISTALLER, W. (1933): *Die zentralen Orte in Süddeutschland: Eine ökonomisch-geographische Untersuchung über die Gesetzmässigkeit der Verbreitung und Entwicklung der Siedlungen mit städtischen Funktionen.*, Jena: Fischer.
- COE, D. AND E. HELPMAN (1995): “International R&D spillovers,” *European Economic Review*, 39, 859–887.
- CORNET, M. (2001): *De Maatschappelijke Kosten En Baten Van Technologiesubsidies Zoals De Wbso*, Den Haag: Centraal planbureau.
- CROFT, T. A. (1978): “Nighttime Images of the Earth from Space,” *Scientific American*, 239, 86–101.
- CZARNITZKI, D. AND C. LOPES-BENTO (2014): “Innovation Subsidies: Does the Funding Source Matter for Innovation Intensity and Performance? Empirical Evidence from Germany,” *Industry and Innovation*, 21, 380–409.
- DANGUY, J., G. DE RASSENFOSSE, AND B. VAN POTTELSBERGHE DE LA POTTERIE (2010): “The R&D-patent relationship: An industry perspective,” CEPR Discussion Papers 8145, C.E.P.R. Discussion Papers.

- DAVIS, D. R. AND D. E. WEINSTEIN (2003): “Market access, economic geography and comparative advantage: an empirical test,” *Journal of International Economics*, 59, 1–23.
- DE BELLEFON, M.-P., P.-P. COMBES, G. DURANTON, L. GOBILLON, AND C. GORIN (2019): “Delineating urban areas using building density,” *Journal of Urban Economics*, 103226.
- DESMET, K., D. K. NAGY, AND E. ROSSI-HANSBERG (2018): “The Geography of Development,” *Journal of Political Economy*, 126, 903–983.
- DESMET, K. AND J. RAPPAPORT (2017): “The settlement of the United States, 1800–2000: The long transition towards Gibrat’s law,” *Journal of Urban Economics*, 98, 50–68.
- DESMET, K. AND E. ROSSI-HANSBERG (2009): “Spatial growth and industry age,” *Journal of Economic Theory*, 144, 2477–2502.
- (2013): “Urban Accounting and Welfare,” *American Economic Review*, 103, 2296–2327.
- (2014): “Spatial Development,” *American Economic Review*, 104, 1211–1243.
- (2015): “On the spatial economic impact of global warming,” *Journal of Urban Economics*, 88, 16–37.
- DIAMOND, R. (2016): “The Determinants and Welfare Implications of US Workers’ Diverging Location Choices by Skill: 1980-2000,” *American Economic Review*, 106, 479–524.
- DIJKSTRA, E. W. (1959): “A note on two problems in connexion with graphs.” *Numerische Mathematik*, 1, 269–271.
- DONALDSON, D. (2018): “Railroads of the Raj: Estimating the Impact of Transportation Infrastructure,” *American Economic Review*, 108, 899–934.

- DONALDSON, D. AND R. HORNBECK (2016): “Railroads and American Economic Growth: A “Market Access” Approach,” *The Quarterly Journal of Economics*, 131, 799–858.
- DURANTON, G., P. MORROW, AND M. TURNER (2014): “Roads and Trade: Evidence from the US,” *Review of Economic Studies*, 81, 681–724.
- DURANTON, G. AND M. TURNER (2012): “Urban Growth and Transportation,” *Review of Economic Studies*, 79, 1407–1440.
- EATON, J. AND Z. ECKSTEIN (1997): “Cities and growth: Theory and evidence from France and Japan,” *Regional Science and Urban Economics*, 27, 443–474.
- EATON, J. AND S. KORTUM (2002): “Technology, Geography, and Trade,” *Econometrica*, 70, 1741–1779.
- ECKHOUT, J. (2009): “Gibrat’s Law for (All) Cities: Reply,” *American Economic Review*, 99, 1676–1683.
- EGGER, P. AND S. LORETZ (2010): “Homogeneous Profit Tax Effects for Heterogeneous Firms?” *The World Economy*, 33, 1023–1041.
- EGGER, P., G. LOUMEAU, AND N. PÜSCHEL (2017): “Natural City Growth in the People’s Republic of China,” *Asian Development Review*, 34, 51–85.
- EGGER, P. AND M. PFAFFERMAYR (2006): “Spatial convergence,” *Papers in Regional Science*, 85, 199–215.
- EGGER, P. H., N. LOUMEAU, AND G. LOUMEAU (2020): “Decomposing the Economic Effects of Transport Infrastructure,” Unpublished manuscript.
- EGGER, P. H., S. NIGAI, AND N. M. STRECKER (2019): “The Taxing Deed of Globalization,” *American Economic Review*, 109, 353–390.
- ERNST, C. AND C. SPENGLER (2011): “Taxation, R&D tax incentives and patent application in Europe,” ZEW Discussion Papers 11-024, ZEW - Leibniz Centre for European Economic Research.

- EVANS, D. (1994): “Policy and Pork: The Use of Pork Barrel Projects to Build Policy Coalitions in the House of Representatives,” *American Journal of Political Science*, 38, 894–917.
- FABER, B. (2014): “Trade Integration, Market Size, and Industrialization: Evidence from China’s National Trunk Highway System,” *Review of Economic Studies*, 81, 1046–1070.
- FAJGELBAUM, P. D., E. MORALES, J. C. S. SERRATO, AND O. ZIDAR (2019): “State Taxes and Spatial Misallocation,” *Review of Economic Studies*, 86, 333–376.
- FAJGELBAUM, P. D. AND E. SCHAAL (2017): “Optimal Transport Networks in Spatial Equilibrium,” NBER Working Papers 23200, National Bureau of Economic Research, Inc.
- FEENSTRA, R. AND H. L. KEE (2008): “Export variety and country productivity: Estimating the monopolistic competition model with endogenous productivity,” *Journal of International Economics*, 74, 500–518.
- FRACASSO, A. AND G. VITTUCCI MARZETTI (2015): “International trade and R&D spillovers,” *Journal of International Economics*, 96, 138–149.
- FRETZ, S., R. PARCHET, AND F. ROBERT-NICOUD (2017): “Highways, Market Access, and Spatial Sorting,” CESifo Working Paper Series 6770, CESifo Group Munich.
- FUJITA, M. AND H. OGAWA (1980): “Equilibrium Land Use Patterns in a Monocentric City,” *Journal of Regional Science*, 20, 455–475.
- (1982): “Multiple equilibria and structural transition of non-monocentric urban configurations,” *Regional Science and Urban Economics*, 12, 161–196.
- GARCIA-LOPEZ, M.-A., C. HEMET, AND E. VILADECANS-MARSAL (2017): “Next train to the polycentric city: The effect of railroads on subcenter formation,” *Regional Science and Urban Economics*, 67, 50–63.

- GLAESER, E. L. AND M. E. KAHN (2001): “Decentralized Employment and the Transformation of the American City,” NBER Working Papers 8117, National Bureau of Economic Research, Inc.
- (2004): “Sprawl and urban growth,” in *Handbook of Regional and Urban Economics*, ed. by J. V. Henderson and J. F. Thisse, Elsevier, vol. 4 of *Handbook of Regional and Urban Economics*, chap. 56, 2481–2527.
- GREENE, W. H. (2017): *Econometric Analysis (8th Edition)*, Pearson.
- GREENWOOD, J., Z. HERCOWITZ, AND P. KRUSELL (1997): “Long-Run Implications of Investment-Specific Technological Change,” *American Economic Review*, 87, 342–362.
- GRILICHES, Z. (1990): “Patent Statistics as Economic Indicators: A Survey,” *Journal of Economic Literature*, 28, 1661–1707.
- HALL, R. E. AND C. I. JONES (1997): “Levels of Economic Activity across Countries,” *American Economic Review*, 87, 173–177.
- HALLÉPÉE, S. AND A. HOULOU (2012): “Evaluation du dispositif JEI,” Tech. rep., DGCIS/ Ministère du Redressement Productif, Paris, France.
- HALVORSEN, R. AND R. PALMQUIST (1980): “The Interpretation of Dummy Variables in Semilogarithmic Equations,” *American Economic Review*, 70, 474–75.
- HANSON, G. H. (2005): “Market potential, increasing returns and geographic concentration,” *Journal of International Economics*, 67, 1–24.
- HANSON, G. H. AND C. XIANG (2004): “The Home-Market Effect and Bilateral Trade Patterns,” *American Economic Review*, 94, 1108–1129.
- HARRIS DOBKINS, L. AND Y. M. IOANNIDES (2001): “Spatial interactions among U.S. cities: 1900-1990,” *Regional Science and Urban Economics*, 31, 701–731.
- HELPMAN, E. (1998): *The Size of Regions*, Cambridge: Cambridge University Press, 33–54.



- HENDERSON, J. V., A. STOREYGARD, AND D. N. WEIL (2012): “Measuring Economic Growth from Outer Space,” *American Economic Review*, 102, 994–1028.
- HENDERSON, V. AND A. MITRA (1996): “The new urban landscape: Developers and edge cities,” *Regional Science and Urban Economics*, 26, 613–643.
- HSIAO, C. (2014): *Analysis of Panel Data*, Cambridge University Press.
- IOANNIDES, Y. M. AND H. G. OVERMAN (2003): “Zipf’s law for cities: an empirical examination,” *Regional Science and Urban Economics*, 33, 127–137.
- JONG, J. P. J. D. AND W. H. J. VERHOEVEN (2007): “Evaluatie WBSO 2001-2005: effecten, doelgroepbereik en uitvoering,” Ministerie van Economische Zaken, Netherlands.
- KANTOROVITCH, L. (1958): “On the Translocation of Masses,” *Management Science*, 5, 1–4.
- KAUFMAN, L. AND P. J. ROUSSEEUW, eds. (1990): *Finding Groups in Data*, John Wiley & Sons, Inc.
- KELEJIAN, H. H. AND I. PRUCHA (2010): “Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances,” *Journal of Econometrics*, 157, 53–67.
- KNOLL, B., M. BAUMANN, AND N. RIEDEL (2014): “The Global Effects of R&D Tax Incentives: Evidence from Micro-Data,” Annual conference 2014 (hamburg): Evidence-based economic policy, Verein fuer Socialpolitik / German Economic Association.
- KRUGMAN, P. (1991): “Increasing Returns and Economic Geography,” *Journal of Political Economy*, 99, 483–499.
- (1996): “Confronting the Mystery of Urban Hierarchy,” *Journal of the Japanese and International Economies*, 10, 399–418.

- KRUSKAL, J. B. (1956): “On the Shortest Spanning Subtree of a Graph and the Traveling Salesman Problem,” in *Proceedings of the American Mathematical Society*, 7.
- LOKSHIN, B. AND P. MOHNEN (2011): “How effective are level-based R&D tax credits? Evidence from the Netherlands,” *Applied Economics*, 44, 1527–1538.
- LOVELY, M. E., Y. LIANG, AND H. ZHANG (2019): “Economic geography and inequality in China: Did improved market access widen spatial wage differences?” *China Economic Review*.
- MA, L. AND Y. TANG (2020): “Geography, trade, and internal migration in China,” *Journal of Urban Economics*, 115, 103181, cities in China.
- MAYER, H., F. SAGER, D. KAUFMANN, AND M. WARLAND (2017): *The Political Economy of Capital Cities*, Routledge.
- McMILLEN, D. P. AND S. C. SMITH (2003): “The number of subcenters in large urban areas,” *Journal of Urban Economics*, 53, 321–338.
- MELITZ, J. AND F. TOUBAL (2014): “Native language, spoken language, translation and trade,” *Journal of International Economics*, 93, 351–363.
- MICHAELS, G. (2008): “The Effect of Trade on the Demand for Skill: Evidence from the Interstate Highway System,” *The Review of Economics and Statistics*, 90, 683–701.
- MONGE, G. (1781): *Mémoire sur la théorie des déblais et des remblais*, De l’Imprimerie Royale.
- MONTE, F., S. J. REDDING, AND E. ROSSI-HANSBERG (2018): “Commuting, Migration, and Local Employment Elasticities,” *American Economic Review*, 108, 3855–3890.
- MONTÈS, C. (2014): *American Capitals: A Historical Geography (University of Chicago Geography Research Papers Book 247)*, University of Chicago Press.

- MORTEN, M. AND J. OLIVIERA (2018): “The Effects of Roads on Trade and Migration: Evidence from a Planned Capital City.” Tech. rep., NBER Working Paper 22158.
- MOSER, P. (2016): “Patents and Innovation in Economic History,” *Annual Review of Economics*, 8, 241–258.
- NAGAOKA, S., K. MOTOHASHI, AND A. GOTO (2010): “Patent Statistics as an Innovation Indicator,” in *Handbook of the Economics of Innovation*, ed. by B. H. Hall and N. Rosenberg, Elsevier, vol. 2 of *Handbook of the Economics of Innovation*, chap. 0, 1083–1127.
- NAGEL, K.-J. (2013): *The Problem of the Capital City. New Research on Federal Capitals and Their Territory*, Col·leccio Institut d’Estudis Autonomics, Barcelona.
- NAGY, D. K. (2017): “City location and economic development,” Unpublished manuscript.
- ORTEGA, F. AND G. PERI (2016): “The effect of income and immigration policies on international migration,” in *The Economics of International Migration*, World Scientific Publishing Co. Pte. Ltd., chap. 11, 333–360.
- PARSONS, M. AND N. PHILLIPS (2007): “An evaluation of the federal tax credit for scientific research and experimental development,” Department of Finance, Economic and Fiscal Policy Branch.
- QIN, Y. (2017): “No county left behind? The distributional impact of high-speed rail upgrades in China,” *Journal of Economic Geography*, 17, 489–520.
- REDDING, S. AND D. STURM (2008): “The Costs of Remoteness: Evidence from German Division and Reunification,” *American Economic Review*, 98, 1766–97.
- REDDING, S. AND A. VENABLES (2004): “Geography and Export Performance: External Market Access and Internal Supply Capacity,” in *Challenges to Globalization: Analyzing the Economics*, National Bureau of Economic Research, Inc, NBER Chapters, 95–130.

- REDDING, S. J. AND E. ROSSI-HANSBERG (2017): “Quantitative Spatial Economics,” *Annual Review of Economics*, 9, 21–58.
- RODRIGUEZ-CLARE, A. (1996): “The role of trade in technology diffusion,” Discussion Paper / Institute for Empirical Macroeconomics 114, Federal Reserve Bank of Minneapolis.
- ROSSMAN, V. (2018): *Capital cities: Varieties and patterns of development and relocation*, London and New York: Routledge.
- ROZENFELD, H. D., D. RYBSKI, J. S. ANDRADE, M. BATTY, H. E. STANLEY, AND H. A. MAKSE (2008): “Laws of population growth,” *Proceedings of the National Academy of Sciences*, 105, 18702–18707.
- ROZENFELD, H. D., D. RYBSKI, X. GABAIX, AND H. A. MAKSE (2011): “The Area and Population of Cities: New Insights from a Different Perspective on Cities,” *American Economic Review*, 101, 2205–2225.
- SIEROTOWICZ, T. (2015): “Patent activity as an effect of the research and development of the business enterprise sectors in the countries of the European Union,” *Journal of International Studies*, 8, 101–113.
- SIMONOVSKA, I. AND M. E. WAUGH (2014): “Trade Models, Trade Elasticities, and the Gains from Trade,” NBER Working Papers 20495, National Bureau of Economic Research, Inc.
- STOCK, J. AND M. YOGO (2005): *Testing for Weak Instruments in Linear IV Regression*, New York: Cambridge University Press, 80–108.
- THEIL, H. AND J. GALVEZ (1995): “On Latitude and Affluence: The Equatorial Grand Canyon,” *Empirical Economics*, 20, 163–166.
- TOMBE, T. AND X. ZHU (2015): “Trade, Migration and Productivity: A Quantitative Analysis of China,” Working Papers tecipa-542, University of Toronto, Department of Economics.
- VILLANI, C. (2003): *Topics in Optimal Transportation*, Graduate studies in mathematics, American Mathematical Society.

- WANG, J. (2013): “The economic impact of Special Economic Zones: Evidence from Chinese municipalities,” *Journal of Development Economics*, 101, 133–147.
- WESTMORE, B. (2013): “R&D, Patenting and Growth: The Role of Public Policy,” OECD Economics Department Working Papers 1047, OECD Publishing.
- WOOLDRIDGE, J. M. (2005): “Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity,” *Journal of Applied Econometrics*, 20, 39–54.
- YU, N., M. D. JONG, S. STORM, AND J. MI (2012): “The growth impact of transport infrastructure investment: A regional analysis for China (1978-2008),” *Policy and Society*, 31, 25–38.
- ZABREYKO, P., A. KOSHELEV, M. KRASNOSEL'SKII, S. MIKHLIN, L. RAKOVSHCHIK, AND V. STETSENKO (1975): *Integral Equations: A Reference Text*, Noordoff International Publishing Leyden.
- ZANDT, F. K. V. (1976): *Boundaries of the United States and the several States*, US Geological Survey.
- ZIPF, G. K. (1949): *Human Behaviour and the Principle of Least Effort*, Addison-Wesley.



# N I C O L E L O U M E A U

ETH Zurich · KOF Swiss Economic Institute  
Leonhardstrasse 21, 8092 Zurich, Switzerland  
E [nicole.loumeau@kof.ethz.ch](mailto:nicole.loumeau@kof.ethz.ch) P +41 44 632 33 48

---

- PERSONAL INFORMATION Born in Berlin on April, 20 1988. German citizen.  
Married, one daughter.
- CURRENT POSITION Ph.D. Candidate in Economics, since March 2015  
Chair of Applied Economics, ETH Zurich  
Main supervisor: Prof. Peter H. Egger  
(*On maternity leave in 2019*)
- RESEARCH INTERESTS Economic Geography, Urban Economics, International Trade
- EDUCATION M.A. Philosophy & Economics (2014), University of Bayreuth, Germany  
B.Sc. International Economics (2012), University of Tübingen, Germany
- JOURNAL PUBLICATION Natural City Growth in the People's Republic of China (2017), *Asian Development Review* (special issue), 34:2, 51-85. (with P. H. Egger and G. Loumeau)
- WORKING PAPER The Economic Geography of Innovation (2018), CEPR Discussion Paper No. DP13338. (with P. H. Egger)

Non-technical summary featured in *Voxeu.org*,  
*Ökonomenstimme*

#### WORK IN PROGRESS

Decomposing the Economic Effects of Transport Infrastructure (with P. H. Egger and G. Loumeau)

Capital Cities and Road Network Integration: Evidence from the U.S.

#### CONFERENCES

**2018:** Trade and Geography Conference (NBER, Cambridge, US); Workshop on Quantitative International Economics (Villars, CH); Sinergia Workshop 2018- Inequality and Globalization (Sarnen, CH); European Economic Association (EEA Cologne, DE); European Trade Study Group (ETSG Warsaw, PL). **2016:** Seventh Research Workshop on International Trade (Villars, CH); Sinergia Workshop 2016- Inequality and Globalization (Stoos, CH).

#### ADDITIONAL TRAINING

**2018:** Quantitative Models of International Trade, Study Center Gerzensee (*S. Kortum*) **2016:** International Trade: Gravity and Geography, IfW Kiel (*G. Ottaviano*); IEB/UEA Summer school in Urban Economics, Barcelona; Trade, Migration and the Geography of Development, Study Center Gerzensee (*E. Rossi-Hansberg*). **2015:** Econometrics and Microeconomics for Research Students, Graduate School of Economics at University of Zurich.



TEACHING

**Spring 2013, 2014:** Introductory Econometrics (undergraduate), exercise class, University of Bayreuth.

**Fall 2019:** Managerial Economics (graduate), ETH Zurich (3× substitute lecturer).

LANGUAGES

German (native), English (fluent), French (intermediate)

COMPUTER SKILLS

(in alphabetical order) ArcGIS, L<sup>A</sup>T<sub>E</sub>X, Matlab, R, Stata

## KOF Dissertation Series

- No 1 Causes and Consequences of Democracy and Terror**  
Martin Gassebner
- No 2 The Impact of News on Expectation Formation in Monetary Policy, Business Cycles and Inflation Forecasts**  
Michael J. Lamla
- No 3 Micro Price Adjustments and Macro Inflation Dynamics**  
Sarah M. Lein
- No 4 Essays on Inflation Expectation Formation**  
Thomas Maag
- No 5 Essays in Political Economy and Fiscal Policy**  
Frank S. Somogyi
- No 6 Modern Statistical Methods Applied to Economic Time Series**  
Simone G. Elmer
- No 7 Analyzing Capital Investment and Unemployment Using Modern Statistical Methods**  
Rolf M. Schenker
- No 8 Capacity Utilisation and Macro Dynamics from a Micro Perspective**  
Eva M. Köberl
- No 9 Essays on the Production and Measurement of Knowledge Capital**  
Thomas Bolli
- No 10 Deviations from Rationality – Essays on Inflation Perceptions and Expectations**  
Lena C. R. Dräger
- No 11 Determinants of Training and the Performance of New Firms**  
Tobias Stucki
- No 12 Measuring Media Sentiment – Essays on Its Impact on the Economy and the Financial Markets**  
Matthias W. Uhl
- No 13 Rigidities, Institutions and the Macroeconomy**  
Heiner Mikosch
- No 14 Applying Modern Econometric Methods to Swiss Unemployment and Consumption**  
Martin Straub
- No 15 Determinants of Cross-Border Activity – Essays on Trade and Entrepreneurship**  
Andrea T. Lassmann
- No 16 Three Essays on International Trade Policy and Welfare**  
Sergey Nigai
- No 17 The Generation and Diffusion of Energy Efficient Technologies in Switzerland: An Empirical Study**  
Marius Ley
- No 18 Four Essays on the Economics of Crime**  
Marie Poprawe
- No 19 Examining the Special Case – Special Trends and Events at the Swiss Labor Market**  
Michael Siegenthaler
- No 20 Empirical Studies on Research Productivity, Labour Mobility, and the Swiss Hospitality Industry**  
Jörg Schläpfer
- No 21 The Globalization of Goods and Capital Flows**  
Benedikt Rydzek
- No 22 Taxation and Firms' Investment Decisions**  
Simon Bösenberg
- No 23 Essays on Applied Nonparametric Econometrics**  
Sabrina Maria Dorn
- No 24 Four Essays on the Estimation of Time-Invariant Variables in Panel Data Models**  
Michaela Kesina
- No 25 Survey Based Research in Economics – Essays on Methodology, Economic Applications and Long Term Processing of Economic Survey Data**  
Matthias Bannert
- No 26 News and the Economy – How to Measure Economic Trends by Using Media-Based Data**  
David Iselin
- No 27 Labor Income Taxation in a Globalizing World: 1980–2012**  
Nora Margot Strecker
- No 28 On the Heterogeneous Effects of Trade and Fiscal Policy**  
Katharina Margarete Klara Erhardt
- No 29 Forecasting at the Current Edge – Revisions and Mixed-Frequency Data**  
Stefan Gerald Neuwirth
- No 30 Employment Protection and Labor Market Outcomes**  
Johanna Mirka Kemper
- No 31 Broader Effects of Migration**  
Jan Ruffner
- No 32 The Relationship Between Knowledge Absorption and Innovation Outcomes**  
Florian Seliger
- No 33 Four Essays in Wine Economics: An Empirical Approach with Swiss Panel Data**  
Alexandre Mondoux
- No 34 The Macroeconomics of Uncertainty**  
Andreas Dibiasi
- No 35 Essays on Business Cycles and Sectoral Dynamics**  
Anna Pauliina Sandqvist
- No 36 Social Status, Typology, and Labor Market Outcomes of Vocational Education and Training**  
Ladina Rageth
- No 37 The Effects of Firm Innovation on the Economy**  
Andrin Spescha
- No 38 Essays in International Trade: Estimating the Impact of Economic Preferentialism**  
Mihai Filip Tarlea
- No 39 Essays on the Economic Geography of the Public Sector**  
Gabriel Solal Loumeau

# KOF Dissertation Series

- No 40** **Determinants of Quality & Appropriability of Patentable Invention**  
Mark James Thompson
- No 41** **Selection Models in the Presence of Networks: Theory and Application for Panel Data**  
Sophia Ding
- No 42** **Preferential Trade Agreements à la Carte: Design and Implications in an Interdependent World**  
Dominique Blümer
- No 43** **Essays in Economic Geography**  
Nicole Loumeau

# KOF

ETH Zurich  
KOF Swiss Economic Institute  
LEE G 116  
Leonhardstrasse 21  
8092 Zurich, Switzerland

Phone +41 44 632 42 39  
Fax +41 44 632 12 18  
[kof@kof.ethz.ch](mailto:kof@kof.ethz.ch)  
[www.kof.ethz.ch](http://www.kof.ethz.ch)  
#KOFETH

