Institute for Economic Studies, Keio University

Keio-IES Discussion Paper Series

From Battlefield to Marketplace: Industrialization via Interregional Highway Investments in the Greater Mekong Sub-Region

Manabu Nose, Yasuyuki Sawada

30 May, 2025
DP2025-010
https://ies.keio.ac.jp/en/publications/26050/

Keio University

CALAMVS GLADIO FORTIOR
1858

Institute for Economic Studies, Keio University
2-15-45 Mita, Minato-ku, Tokyo 108-8345, Japan
ies-office@adst.keio.ac.jp
30 May, 2025

From Battlefield to Marketplace: Industrialization via Interregional Highway Investments in the Greater Mekong Sub-Region

Manabu Nose, Yasuyuki Sawada

Keio-IES DP2025-010
30 May, 2025
JEL Classification: O14, O18, O22, O25, R12, R32, R58
Keywords: spatial structural transformation, market access, treatment spillover, agglomeration, core-periphery

Abstract

This paper examines the nonlinear effects of a large-scale highway construction project in the Greater Mekong Subregion, which connects the historically conflict-affected borderlands of northern Vietnam to the country's industrial core. Employing a market access framework with geo-coded highway network and firm-level panel data, we estimate the causal impact of improved interregional connectivity, while accounting for spillovers via production input-output linkages. To address endogeneity issues arising from non-random route placements, we construct least-cost path spanning tree networks. Our instrumental variable estimates reveal that enhanced market access spurred manufacturing firm agglomeration and employment growth, particularly in peripheral rural areas. We further explore the underlying sources of polycentric development patterns, finding pronounced effects in second-tier cities characterized by less intense competition and better access to national road networks. Our findings are robust to controls for industrial zones, underscoring the pivotal role of the upgraded highway connectivity in transforming previously marginalized regions and supporting economy-wide industrialization over the past decade.

Manabu Nose
Keio University, Faculty of Economics
mnose@keio.jp

Yasuyuki Sawada
University of Tokyo, Faculty of Economics, Graduate School of Economics
sawada@e.u-tokyo.ac.jp

Acknowledgement: Authors thank Yasuyuki Todo, Tomoya Mori, Gilles Duranton, Yuhei Miyauchi, Hisaki Kono, Yuki Higuchi, Katsuya Takii, Shuhei Kitamura, Hirokazu Ishise, Takuma Kamada, Miki Kohara, Ken Miura, and seminar participants at Osaka, Keio, Kyoto, ADB-UTokyo Urban Development Conference, JEA Spring meeting, WEAI, IIPF, and ADB LEAP seminar for their useful comments. We thank Hiroyuki Miyazaki and Dan Tran at GLODAL Inc. for geospatial data analytical support. The project is supported by the ADB, the Joint Research Program No. 970 at the Center for Spatial Information Science, The University of Tokyo, and JSPS KAKENHI Grants-in-Aid for Scientific Research, Japan (International Leading Research, Number 22K21341). All errors are our own.

From Battlefield to Marketplace: Industrialization via Interregional Highway Investments in the Greater Mekong Sub-Region*

Manabu Nose†
Yasuyuki Sawada‡
Tung Nguyen§

May 30, 2025

Abstract

This paper examines the nonlinear effects of a large-scale highway construction project in the Greater Mekong Subregion, which connects the historically conflict-affected borderlands of northern Vietnam to the country's industrial core. Employing a market access framework with geo-coded highway network and firm-level panel data, we estimate the causal impact of improved interregional connectivity, while accounting for spillovers via production input-output linkages. To address endogeneity issues arising from non-random route placements, we construct least-cost path spanning tree networks. Our instrumental variable estimates reveal that enhanced market access spurred manufacturing firm agglomeration and employment growth, particularly in peripheral rural areas. We further explore the underlying sources of polycentric development patterns, finding pronounced effects in second-tier cities characterized by less intense competition and better access to national road networks. Our findings are robust to controls for industrial zones, underscoring the pivotal role of the upgraded highway connectivity in transforming previously marginalized regions and supporting economy-wide industrialization over the past decade.

Keywords: spatial structural transformation, market access, treatment spillover, agglomeration, core-periphery.

J.E.L. classification codes: O14, O18, O22, O25, R12, R32, R58

*Authors thank Yasuyuki Todo, Tomoya Mori, Gilles Duranton, Yuhei Miyauchi, Hisaki Kono, Yuki Higuchi, Katsuya Takii, Shuhei Kitamura, Hirokazu Ishise, Takuma Kamada, Miki Kohara, Ken Miura, and seminar participants at Osaka, Keio, Kyoto, ADB-UTokyo Urban Development Conference, JEA Spring meeting, WEAI, IIPF, and ADB LEAP seminar for their useful comments. We thank Hiroyuki Miyazaki and Dan Tran at GLODAL Inc. for geospatial data analytical support. The project is supported by the ADB, the Joint Research Program No. 970 at the Center for Spatial Information Science, The University of Tokyo, and JSPS KAKENHI Grants-in-Aid for Scientific Research, Japan (International Leading Research, Number 22K21341). All errors are our own.

†Corresponding author: Keio University. mnose@keio.jp
‡The University of Tokyo. sawada@e.u-tokyo.ac.jp
§Hitotsubashi University. tungnt.hit@gmail.com

1 Introduction

The Greater Mekong Sub-region (GMS) Economic Cooperation Program is a large-scale economic development program covering countries along the Mekong River in Southeast Asia which has played a critical role in transforming the region from battlefields in the 1970's into marketplaces.[1] The governments and aid agencies have supported the program by providing substantial funding for transport infrastructure to facilitate cross-border trade, investments, and inclusive growth. While the GMS program has been regarded as one of the earliest and most successful regional cooperation and integration (RCI) programs, often cited as a model "economic corridor" in the policy arena (Asian Development Bank, 2020), to the best of our knowledge, there has been no rigorous impact assessment of the program, uncovering possible mechanisms behind such success. Particularly, it is imperative to uncover the program's resulting pattern of spatial development quantitatively because, theoretically speaking, it is unclear whether it can lead to monocentric or polycentric development (Krugman, 1991). The overall welfare impact of large-scale regional policies can be rigorously evaluated only by assessing whether they foster balanced development or exacerbate regional disparities.

This paper aims at bridging this important lacuna by estimating the effect of the GMS program on firm's agglomeration patterns over a decade, shedding light on non-linear development between the core and peripheral cities. We examine how upgrading interregional transport infrastructure facilitated the industrialization of northern Vietnam, a region historically affected by wartime disruption and postwar underdevelopment, by enhancing market access and spatial integration between peripheral rural areas and urban economic centers. To this end, the paper places a particular focus on the GMS's North-South corridor that connects two major economic centers in the northern Vietnam (Hanoi and Haiphong) to two Chinese provinces (Yunnan Province and Guangxi Zhuang Autonomous Region) around the China-Vietnam border. Since the old national road that links China's Kunming city with Hanoi and Haiphong became obsolete and overloaded, the governments decided to invest in rehabilitating the road. The Vietnamese government invested approximately US$1.4 billion in upgrading the 244 km old road from Lao Cai (the border city with China) to Hanoi into a highway with multiple lanes. In parallel, the extension of the new highway to Haiphong and the expansion of the transport network to

[1] The GMS program comprises Cambodia, the People's Republic of China (PRC, specifically Yunnan and Guangxi Provinces), the Lao People's Democratic Republic (Lao PDR), Myanmar, Thailand, and Viet Nam, covering an area of 2.6 million square kilometers and a combined population of more than 330 million.

Lan Song (a border city with China's Guangxi province) have been implemented.

Theoretically, the construction of the interregional transport network significantly improves connectivity and facilitates the exchange of goods and people by reducing transportation costs. The improvement in infrastructure connectivity, in turn, increases local market potential, altering the landscape of economic activities over the space. Economic geography models predict non-linear relationship between connectivity infrastructure and the agglomeration of activities: improved connectivity can strengthen both *centripetal* forces to central (core) cities and *centrifugal* forces which may expand the city fringe, spreading out workers to the suburban (periphery) area (Krugman, 1991; Baum-Snow, 2007). Indeed, the existing empirical findings are mixed: transport investment has contributed to the diffusion of economic activities and suburbanization of the urban landscape in several specific contexts (Jedwab and Storeygard, 2020; Baum-Snow, 2007; Baum-Snow *et al.*, 2017), while other studies show the centralization of the urban system (Faber, 2014). When the transport investments connect large metropolitan areas and small peripheral cities like the case in China's national highway system (Faber, 2014), ex-ante spatial unevenness in local market size, amenity conditions, and the degree of transport network externality may lead to the asymmetric impact of the GMS program on urban development (Allen and Arkolakis, 2014).

This paper explores the question using hybrid commune-level geospatial data sets constructed from Vietnamese firm census, micro surveys as well as high-resolution satellite imageries.[2] We first describe the non-parametric relationship between highway network improvements and each district's market access, and estimate how the enhanced market access, in turn, promoted industrial development differently between core and periphery areas. We employ a firm market access (FMA) approach based on an economic geography model (Redding (2022)), besides a reduced-form model of market potential, i.e., the Harris index (Donaldson and Hornbeck, 2016; Combes and Gobillon, 2015). The growth of local markets is often attributed to exogenous locational advantages. In our study, we construct least cost path spanning tree networks along with large variations in geographic conditions between the China-Vietnam border (mountainous) and the Mekong Delta as instrumental variables (IV) to estimate the aggregate causal impact running from better highway network to enhanced market access, and ultimately, the agglomeration of firms and workers over space.

[2] Donaldson and Storeygard (2016) reviews recent applications of satellite data in Economics. Our satellite-based analysis closely relates to Burchfield *et al.* (2006) and Gibbons *et al.* (2019).

The polycentric core-periphery development could emerge when transport cost reduction promotes agglomeration of firms near the highway, while also creating diffusion of business to suburban areas through production chains (Krugman and Venables, 1995).[3] We additionally examine the local production network as another source of polycentric spatial development by estimating the heterogeneity of GMS highways' treatment effect by supply chain linkages. We measure the supply chain linkage based on input-output tables and apply the generalized difference-in-difference (DID) design with continuous treatment (Lindo *et al.* (2020)). The flexible DID specification quantifies the medium-term impact of the GMS highway construction on firm agglomeration and employment, accounting for the size of the impact multiplier through local supply chains.

To preview our empirical results, two main findings have emerged. First, our IV regression shows that improved firm market access from the GMS highway upgrade generated stronger agglomeration effects than competition among manufacturing, transport, and trade firms. The effects are especially pronounced in peripheral regions, where the agglomeration elasticity is particularly large. In the core region, agglomeration was more spatially concentrated – largely driven by foreign firm entry and clustering of suppliers and services within 20km from the GMS highway.

By contrast, peripheral areas in the northeast and northwest – initially characterized by underdeveloped markets– experienced substantial manufacturing firm entry and job creation when improved connectivity raised market access above a critical threshold level. This shift enabled the entry of small firms and the expansion of formal employment near the highway. The manufacturing sector grew by 47% within 20km and 39% within 50km radius of the highway over a decade, and service industries expanded across a wider spatial scale, extending to 50km from the GMS highway. As a result of new formal job creation, we also find an increase in the share of manufacturing workers. While the urban system had been largely *monocentric* – centered on the capital city (Hanoi) 20-30 years ago, our findings indicate that the expansion of interregional connectivity contributed to a transition toward a more *polycentric* structure, with manufacturing activity spreading to the periphery.

Second, we estimate heterogeneous treatment effects of the GMS highway, focusing on input-output linkages in local production networks. These linkages have deepened over

[3] Akamatsu *et al.* (2019) shows that multimodal agglomeration can emerge under a global dispersion force (such as local competition effect) which induces the establishment of firms distant from urban centers. At the same time, the flattening of existing agglomeration could emerge due to a local dispersion force (such as the local scarcity of land).

the past decade, supported by market liberalization and industrial policy. Our generalized DID results indicate that transport cost reductions led to a more pronounced agglomeration of manufacturing firms in communes with strong upstream linkages, particularly those located near the highway. However, these supply chain effects are more evident in core regions than in the periphery. This suggests that while production linkages amplify agglomeration where they are already dense, supply chain connections alone cannot fully explain the emergence of a polycentric development pattern. Instead, this spatial transformation in peripheral areas is more plausibly attributed to catch-up dynamics and weaker competitive pressure that facilitated firm entry once market access conditions improved.

Finally, we demonstrate the robustness of our main findings against alternative mechanisms that could possibly influence firms' agglomeration and job creation. Specifically, we account for (a) the impact of industrial zones, (b) spatial spillovers via secondary road linkages, and (c) knowledge spillover through co-agglomeration. While these factors may have played a role, our results consistently point to the central role of interregional transport improvements in driving structural change and spatial reorganization. This focus on internal transportation infrastructure as a driver of urban and industrial transformation offers a novel contribution to the literature, which has thus far received less attention in the literature. [4]

The rest of this paper is organized as follows. Sections 2 introduces our data and background. Section 3 provides a spatial general equilibrium framework to guide the empirical analyses. In section 4, we construct least-cost path spanning tree networks as our main IV and estimate the effect of firm market access on agglomeration and employment growth. Sections 5 and 6 estimate the generalized DID regressions to examine the GMS highway and supply chain linkages as sources of polycentric spatial development. Finally, Section 7 provides a conclusion.

[4] Except a few recent works such as Fajgelbaum and Redding (2021), Nagy (2020), Eckert and Peters (2022), and Trew (2020). Fajgelbaum and Redding (2021) found that Argentina's integration into the world economy in the late-19th century triggered structural change due to a spatial Balassa-Samuelson effect.

2 Data and Context

2.1 Data

Our analysis draws upon a rich set of data sources combining micro-level information and remote sensing imagery. The main dataset is constructed from three traditional micro datasets on registered firms, households, and workers (see Appendix A for details). In addition, we incorporate satellite imageries on historical land cover and night-time light (NTL) to capture long-term spatial changes in development.

For land cover, we use Landsat satellite imagery since 1972 and focus on five benchmark years: 1990, 2000, 2010, 2015, and 2019. To track local economic development, we utilize the monthly average Day-Night Band (DNB) radiance values from the NASA Visible and Infrared Imaging Radiometer Suite (VIIRS), available since April 2012. VIIRS offers higher resolution (15 arc-seconds, approximately 500m grid) than the older Defense Meteorological Satellite Program (DMSP), providing more precise proxy indicators of local economic activity in both urban and rural areas (Gibson *et al.* (2021)).[5]

Among 696 districts nationwide, our study area includes about 4,350 communes within 270 districts across three northern regions of Vietnam. We define "core cities" as districts in the Red River Delta (RRD) region, including Hanoi and Haiphong– two central provinces in the north.[6] Districts in the the northwestern (NW) and northeastern (NE) regions are defined as "peripheral cities" (see Figure 1).

Due to inconsistencies in tracking state-owned enterprises (SOEs; including collectives) in the annual enterprise survey, we limit our analysis to private registered firms and exclude SOEs. This exclusion helps avoid confounding effects from the ongoing SOE reforms since 2011 and enhances the identification of the impact of infrastructure improvements.

2.2 Background

Our focus is on understanding how recent infrastructure investments have shaped the spatial reallocation of economic activities in northern Vietnam. As our baseline condition, Figure 2 illustrates large socio-economic disparities as of 1999. Peripheral regions in

[5] Gibson *et al.* (2021) discusses the superiority of using VIIRS as a proxy for GDP in both urban and rural cities. Spatial inequality tends to be understated when the DMSP is used.

[6] In Vietnam, central provinces, or cities directly under central government, includes Hanoi, Haiphong, Danang, Can Tho, and Ho Chi Minh City. These cities are the socio-economic centers in the region.

the northwestern (mountainous region) and northeastern Vietnam exhibited initially high poverty and inequality with limited access to basic infrastructure. In contrast, metropolitan areas around Hanoi and Haiphong displayed significantly higher living standards.

Core cities had higher population density, with the RRD region's average population size about 1.3 times larger than the national average in 2019 (Table 1a). In contrast, peripheral areas, particularly near national borders, remained primarily agricultural. However, over the past decade, the population size has also grown by 19% in the northwestern peripheral cities.

Similarly, Enterprise Census demonstrates a high concentration of industrial activities in the RRD region, home to roughly one-third of the nation's registered firms (Table 1b). Since 2000, all northern districts have experienced substantial growth in firm numbers, driven by market entry of foreign firms and the creation of small- andmedium-sized enterprises (SMEs). As a result, manufacturing employment grew at an annual rate of 7.4% in northern regions, surpassing the national average of 5.1%.[7] The average labor income in northern regions, initially below the national average in 1999, grew faster over the past two decades, particularly in manufacturing and transportation. The number of registered transportation companies increased dramatically from 850 in 2001 to 12,835 in 2020.

2.3 GMS Highway Construction Program

Transport infrastructure development has played a crucial role in transforming northern Vietnam. Figure 3 displays the geo-coded location of the GMS highways connecting Kunming, Lao Cai, Hanoi, Lang Son, and Haiphong. Prior to construction, international container cargo from Kunming (upper left of the figure) were often sent by rail or road to ports in Guangzhou province in China (1,053km away from Kunming) rather than being sent via the Kunming - Lao Cai - Hanoi/Haiphong route due to long travel times (up to six days). The construction of the GMS expressway would improve connectivity to Haiphong port, which is only 813km away from Kunming, with better road pavement. In this context, the construction was expected to significantly shorten the cargo travel time between Kunming and Haiphong from more than 2 days to 1 day.

For example, the Hanoi – Lao Cai highway (CT.05), started in 2008 and completed in 2014, was estimated to reduce travel time from over 10 hours to about 3.5 hours. Daily

[7] Note that the Enterprise Surveys (VES) underreport the country's overall employment as a large share of micro and small businesses do not register, or do not need to register, to the authorities. Additionally, a large share of employment is within household businesses, which are not recorded in the VES.

traffic surged to 18,000-19,000 passenger car unit (PCU) in 2017, far exceeding the pre-2015 volume of 4,000 PCU (GMS Secretariat, 2018). The Hanoi-Haiphong highway (CT.04), started in 2009 and completed in 2015 with a US$2 billion investment under the Built-Operate-Transfer (BOT) scheme, features six travel lanes and two emergency lanes. This modern highway, running parallel to the old national road (QL5), allows travel speeds of up to 120 km per hour, reducing inter-city travel time from 3 hours to under 2 hours.[8][9]

As part of the GMS eastern corridor, the Hanoi-Lang Son (HN-LS) expressway enhances access between northeastern Vietnam and China's Guangxi region. The expressway consists of three sections: Hanoi - Bac Giang (46km), Bac Giang – Chi Lang (63.1km), and Chi Lang - Huu Nghi (Lang Son, 43.3km). Of these, the first two sections were built under BOT contracts and were finished in 2016 and 2019, respectively.

As shown in appendix C, interprovincial freight traffic has grown substantially over a decade, particularly in urban centers, as represented by denser traffic volume in 2020. This suggests that GMS highways have made a meaningful contribution to local economic development, especially in core cities where transport firms are concentrated.

2.4 Agglomeration Patterns of Firms: Core vs. Periphery

Table 1b highlights contrasts in industrial composition between core and peripheral cities. Within the RRD region, 70% of firms operate in services and one-third of employment is in service sector. However, peripheral cities have experienced an increase in manufacturing employment share from 56% in 2011 to 62% in 2016. Construction dominates remote border districts (such as ones along the border with China), accounting for about half of total employment.

The regional trends suggest a structural shift in peripheral areas, from agriculture and construction towards manufacturing and services. Manufacturing employment grew especially fast in the northeast (66%) and northwest (51%) peripheral regions. In contrast, core cities experienced a shifts in workers from manufacturing toward services.

[8]The old national road was built during the French colonial period and allows speed limits of 80km (50km) per hour outside (inside) residential areas.

[9]After the completion of CT.05, the construction of Haiphong– Ha Long – Van Don – Mong Cai expressway (CT.06) started. This expressway is part of the North-South Economic Corridor (NSEC)-4, connecting Nanning (PRC) and Haiphong through the coastline. The construction of the first two sections was completed earlier in 2018, while the last section, between Van Don and Mong Cai, was completed in 2022.

2.5 City Development from High-Resolution Satellite Images

To motivate our empirical analysis, we document spatial development trends – the agglomeration and the decentralization of economic activities – using high-resolution satellite imagery from 1990, 2000, 2010, and 2019.

The satellite image archives provided on the Google Earth Engine enable us to track the growth and changes around the target areas over the past decades (Miyazaki *et al.*, 2019). We obtain cloud-free images from Landsat 5 and 8 and classify land cover using a supervised neural network algorithm. We define four land cover classes – waterbody, vegetation, bare land, and built-up area – and compute the percentage of built-up areas by commune using ArcGIS zonal statistics.

Figure 4 shows the long-term expansion of urban areas along the GMS corridor in the northern Vietnam since the outset of the *Doi Moi* reform in 1990 until 2019. Built-up areas were minimal (less than 10%) outside Hanoi until 2000, but grew rapidly after 2000, resulting in an increase in built-ups and the diffusion of settlements to peripheral cities near the coast (Haiphong and Ha Long) and toward the border with China (Lao Cai, Yen Bai). Improved transport connectivity after 2010 further accelerated this trend, especially along the GMS highway corridor. We also observe that the urban fringe has sprawled further down to the remote peripheral areas like the border with Laos (Hoa Binh, Son La) through an existing Asian Highway 13 (AH-13) corridor (Figure 3).

Figure 5 illustrates growth in NTL intensity, showing a consistent picture. Urban core areas near Hanoi and industrial hubs in northeastern Vietnam saw the most rapid gains, while peripheral areas along highways also lit up over years. The expansion of city fringe towards peripheral areas may reflect the impact of the GMS transport investment.

Landsat imageries suggest a transition from *monocentric* to *polycentric* urban forms. Initially, economic activities were concentrated in Hanoi, suggesting strong agglomeration economies due to Marshall's local externalities (Henderson, 1974). With higher wages and land prices, the land development tended to be compact. On the other hand, the monocentric city model also predicts that improved transport lowers commuting costs within a city and will induce suburban growth. As Figures 4 and 5 show, like the U.S. in 1970s (Burchfield *et al.*, 2006), the suburban areas outside Hanoi, particularly to the east and southeast, have become more polycentric with decentralization of economic activities to peripheries. The extent of decentralization varies across regions and industries depending on the strength of agglomeration economies (Glaeser and Kahn, 2001, 2004).

3 Theoretical Framework

3.1 Model Overview

To analyze the spatial reorganization of economic activities in response to infrastructure upgrades, we develop a multi-region, multi-sector spatial general equilibrium model (see Appendix B for details of model). The framework extends the core-periphery tradition in economic geography (Krugman and Venables, 1995; Venables, 1996; Puga and Venables, 1996; Redding and Venables, 2004), incorporating the following key features:

Regions and Sectors. The model features multiple regions ($j = 1, \ldots, J$), each capable of producing both agricultural and manufacturing goods. Agriculture operates under perfect competition and uses a constant-returns-to-scale technology with land and labor as inputs. In contrast, manufacturing is monopolistically competitive and consists of S differentiated sectors, where each firm produces a unique variety of goods.

Input-Output Linkages. Each manufacturing sector utilizes intermediate inputs sourced from all other sectors. These relationships are captured by the input-output coefficient matrix μ_{rs}, which quantifies the share of inputs from sector r used in the production of sector s. This structure allows for downstream and upstream dependencies across sectors.

Trade Costs and CES Demand. Goods are traded across regions with iceberg trade costs $\tau_{ji} \geq 1$, implying that only a fraction of shipped goods arrives at the destination. Both consumers and firms exhibit constant elasticity of substitution (CES) preferences over varieties and sectors, shaping demand patterns based on relative prices and trade frictions.

Firm Entry and Productivity. Manufacturing sectors are characterized by free entry, ensuring a zero-profit equilibrium condition. Sectoral productivity is not fixed but increases endogenously with the number of firms in the same region due to agglomeration externalities. Specifically, manufacturing productivity takes the form $A_{M,s,i} = A_{M,\text{base},s,i} \cdot n_{s,i}^{\alpha}$, where $A_{M,\text{base},s,i}$ represents base productivity, $n_{s,i}$ is the number of firms in sector s in region i, and $\alpha > 0$ reflects the strength of scale economies from local firm concentration.

Market Access. Firms' effective revenues depend on their ability to reach destination markets. Market access of a firm in region i, sector s, selling its products to destination markets j is determined by total potential demand in each destination, trade costs, and prices in the destination region:

$$MA_{s,i} = \sum_j \tau_{ij}^{1-\sigma} \cdot \frac{n_{s,j}}{P_{s,j}^{1-\sigma}}$$

where $n_{s,j}$ is the number of firms in destination market j and τ_{ij} is iceberg trade costs.

The model jointly determines wages, number of firms, sectoral prices, and trade flows, balancing agglomeration and competition effects across space. In equilibrium, the number of firms in sector s in region i adjusts to satisfy the zero-profit condition:

$$n_{s,i} = \frac{\sum_j \pi_{s,i} \cdot E_{s,i}}{F \cdot w_i}.$$

where $\pi_{s,i}$ is the firm's profit, $E_{s,i}$ is the regional demand for sector s, w_i is the wage in region i, and F is a fixed entry cost (hiring initial labors).

3.2 Equilibrium Calibration and Simulation Design

To operationalize the model, we calibrate it to two regions—Periphery and Core—and three sectors representing Primary, Secondary, and Tertiary (Services) industries. Table below summarizes the parameter values used.

Calibrated Parameters

Parameter	**Description**	**Value**
σ	Elasticity of substitution	5
θ	Agri. share in consumption	0.5
F	Fixed cost of firm entry	1.0
α	Agglomeration elasticity	0.5
γ	Competition sensitivity	0.1
S	Number of sectors	3
J	Number of regions	2
η_s	Labor share in sector s	[0.2, 0.3, 0.4]
$A_{M,\text{base}}$	Manufacturing productivity, by region	[1.0, 1.2]
A_A	Agricultural productivity, by region	[1.2, 1.0]
K_A	Land (Agri.), by region	[1.2, 1.0]
L_j	Labor endowment, by region	[1.0, 1.0]

Two scenarios for input-output linkage matrices are considered: *Weak* and *Strong*, capturing variation in inter-sectoral dependence across regions.

Input-Output Linkage Coefficients (μ_{rs})

Source r	Weak			Strong		
	$s=1$	$s=2$	$s=3$	$s=1$	$s=2$	$s=3$
$r=1$ (Primary)	0.05	0.03	0.02	0.10	0.05	0.05
$r=2$ (Secondary)	0.10	0.10	0.05	0.20	0.25	0.15
$r=3$ (Service)	0.02	0.03	0.05	0.10	0.10	0.20

3.3 Theoretical Hypotheses: Market Access and Firm Distribution

We simulate the model under varying trade cost levels $\tau \in [1.0, 2.0]$ to trace the equilibrium number of firms in each region. Figure 6 demonstrates the simulation results to rationalize the spatial heterogeneity in firm agglomeration in responses to infrastructure upgrades as observed in data, highlighting the nonlinear effects of trade cost reductions and the importance of sectoral linkages.

Our theoretical analysis provides several predictions on how market access improvements affect regional firm dynamics:

(a) Peripheral Regions: At low market access levels, the market is still small, and increasing competition with more varieties lowers firms' effective revenues with the sensitivity γ (Equation (10) in Appendix B). As a result, firm entry is suppressed. Once access improves beyond a certain threshold, the agglomeration effect encourages entry. Thus, a *V-shaped* relationship is expected between market access and the number of firms in secondary and tertiary sectors. With stronger supply chain linkages, firms face higher entry costs and thus need a higher level of market access before realizing strong agglomeration benefits.

(b) Core Regions: Core regions already enjoy agglomeration benefits. Market access improvements produce strong initial gains in firm numbers, but exhibit diminishing returns as saturation sets in.

(c) Input-Output Linkages: Stronger μ_{rs} magnify the gains from agglomeration but also increase the threshold needed in periphery regions. Stronger input-output linkages as it shifts the agglomeration threshold rightward for the periphery, and steepen the early gains for the core. Linkages amplify both the benefits and frictions of regional integration.

These hypotheses guide the empirical investigation that follows, particularly focusing on how transport improvements (e.g., highways) affect firm distribution across regions with heterogeneous economic structures.

4 Empirical Analysis

Given the theoretical predictions, we quantify the impact of the GMS transport infrastructure investments on local market access and examine how it explains the observed changes in agglomeration patterns between core and peripheral cities. A community not only benefits from the direct access to highway, but also from improved accessibility to neighboring markets through the road network indirectly. We take advantage of geo-coded GMS highway network data and follow the "market access" approach (Donaldson and Hornbeck (2016)) to summarizes overall market accessibility improvement through lower transport cost due to the highway upgrades.

4.1 Measurement and Extent of Market Potential

We start from characterizing the market access using the reduced-form "market potential" (MP) index that captures the size of own market as well as surrounding markets where firms trade with. Assuming the iceberg cost of trade, the Harris MP index expresses effective density of workers as the distance weighted sum of the employment of all districts within an accessible distance (Combes and Gobillon, 2015):

$$MP_{ct} = \sum_{l \neq c} \frac{den_{lt}}{d_{cl}}, \tag{1}$$

where den_{lt} is employment density in commune l at year t and d_{cl} is the shortest distance between the centroids of communes c and l. This index calculates the shortest distance for all pairwise combinations of communes and keep neighboring districts within 50 km for each district.

Two panels in Figure A4 illustrate the spatial distribution of the market potential in northern Vietnam. The market potential has concentrated in core cities near Hanoi and expanded over the decade. From 2011 to 2016, we found faster increase in MP index in districts along the GMS highways. Mountainous districts near Chinese border also experienced a moderate expansion of market potential. In Thai Nguyen and Bac Giang provinces, the MP index has increased greater than 50% (red color) due to their locational advantage. While Bac Giang is an important midpoint in the HN-LS expressway, Thai

Nguyen is close to both HN-LS and HN-LC highways.

Built on the findings from the satellite images, this subsection estimates the spatial scale of the GMS project impacts on local market development in terms of market accessibility. Given the non-linearity of the project impact, we apply the Robinson (1988)'s semi-parametric regression model:

$$\ln MP_{drt} = m(GMSdist_{dr}) + x_{drt}\beta + \epsilon_{drt}, \tag{2}$$

where $m(\bullet)$ is a non-parametric component as a function of $GMSdist$, which is the Euclidian distance between the People Committee (PC) office of district d in region r (the economic center of each district) and the nearest interchange point of the GMS highway. The parametric part is composed of variables x_{drt}, including provincial dummies and the distance to the capital city, Hanoi. Figure 7 shows the relationship between the change in MP index and the distance to the GMS highways with the 95% confidence interval. For the non-parametric estimates of the MP index, we report results separately for the short-term impact (from 2011 to 2016, left figure) and the medium-term impact (from 2011 to 2020, around five years after the project completion, right figure).

The left chart in Figure 7 shows that market grew equally around 20-40% within 50km from the GMS highways. Remote districts beyond 70km away from the highways experienced almost no change in the market access. Over the period 2011-2016, the market potential expanded by around 30% to 40% for places within 60km and the positive effect gets smaller as the location moves away from the highway. The market potential expanded even faster in the NE and NW regions at nearly 50% within 20km from the highway. This shows that the GMS significantly affected the market potential of places within 70km at different magnitude with potentially large agglomeration effects.

Based on these estimation results, we define treatment exposure as highway access within 20km and 50km radii, which appear to mark the thresholds beyond which the effects largely dissipate.

4.2 Instrumental Variable Regression

We leverage the GMS highway upgrade as a quasi-experiment that facilitated transport connectivity for cities closer to the highway network. However, estimating the effect of market access on firm agglomeration by OLS necessitates a strong assumption that GMS networks between nodal cities were randomly assigned. Planners might have possibly targeted politically important and economically prosperous districts in building the GMS highway network. To uncover the relationship between local market expansion and agglomeration patterns across communes, we apply the instrumental variable (IV) regression to estimate the arguably causal effect of market access (MA) growth on firm growth over nine years in the first-difference (FD) specification. In other words, we empirically implement the theoretical relationship illustrated in Figure 6.

The challenge to estimate the causal effect of highways is that, first, randomized experiments are almost impossible to design and implement by nature; and, second, actual roads are built in growing places or areas with development prospects to meet the traffic demand, or lagging areas to accelerate inclusive development, which suggests possible existence of reversed causality. We minimize this potential bias in estimation due to non-random allocation of highways and the induced market expansion by adopting the inconsequential unit approach (Gibbons *et al.*, 2019; Ahlfeldt and Feddersen, 2018; Redding and Turner, 2015; Chandra and Thompson, 2000), which serves as a complement to our instrumental variable identification strategy.[10] We restrict our sample to districts within 100km from the GMS transport network. In this way, we estimate effects from variations in the intensity of treatment (the change in market potential induced by the GMS highway) within the buffer zone of 100km from the project site. The identification comes from comparison of units experiencing larger and smaller changes in market potential, amongst the sub-sample that are all close to the road.

Previous literature has proposed various IV strategies to tackle the endogenous placement of infrastructure based on: (a) historical transport network (Baum-Snow *et al.*, 2017; Duranton and Turner, 2012; Michaels, 2008), (b) counterfactual least-cost-path network (Faber, 2014), (c) straight-line connections (Banerjee *et al.*, 2020; Ghani *et al.*, 2016), (d) non-local changes to the road network (Jedwab and Storeygard, 2020), and (e) geographic barriers like land gradient and elevation (Dinkelman, 2011; Duflo and Pande, 2007). These topological instruments have not been applied to transportation infrastructure, but certainly predicts the region's market potential growth by affecting the cost of firms' market entry and workers' occupational mobility.

[10] In this approach, treatment, i.e., exposure to infrastructure, can be considered exogenous at the local level, conditional on controlling for regional-level factors, because local characteristics are inconsequential to national-level decision-making in large-scale infrastructure investments.

4.2.1 The Minimum Spanning Tree (MST) Highway Networks

Our IV strategy follows the second line of related literature, while exploiting exogenous variations in northern Vietnam's unique topological conditions (related to the last strand of literature). This section explains how we construct hypothetical minimum spanning tree highway networks as an instrumental variable for actual route placement following Faber (2014) and Morten and Oliveira (2024). We assume that the government and policymakers chose the shortest path to connect core coastal cities to peripheral boarder cities with PRC. The least cost path serves as a plausible instrument because it reflects an exogenous, engineering-based logic of network design and satisfies key econometric identification criteria, i.e., relevance and exclusion restriction.

4.2.2 Preparing a Construction Cost Surface

To construct the least-cost path (LCP) spanning tree network, we adapt the following construction cost function from the transport engineering literature (Jha *et al.* (2001); Jong and Schonfeld (2003)).

$$c_i = 1 + slope_i + 25 * Developed_i + 25 * Water_i + 25 * Wetland_i,$$

where c_i is the cost of crossing a pixel of land i and $Slope_i$ is the average slope gradient of land i. $Developed_i$, $Water_i$, and $Wetland_i$ are dummies indicating whether land i is covered by built structures, water, or wetland. We use JAXA-ALOS landcover map[11] and NASA-ASTER global digital elevation model[12] (both in 30 meter spatial resolution).

For computational purpose, we resampled original cost layer to 100 meter resolution and calculated hypothetical LCP for provincial capitals and 13 targeted highway intersections in the northern Vietnam using the Dijkstra's optimal route algorithm. We calculate the LCPs for 1990, 2005, 2010, 2016, and 2020 (Figure 9).[13] LCPs between the centroid of 270 district capitals are also calculated with a 500 meter resolution. As the final step, the bilateral cost parameters are fed into Kruskal's MST algorithm to identify the routes that connect all targeted nodes on a single continuous network subject to global construction

[11] https://www.eorc.jaxa.jp/ALOS/en/dataset/lulc/lulc$_{v}nm_{v}$2109$_{e}.htm$

[12] https://asterweb.jpl.nasa.gov/gdem.asp

[13] The hypothetical LCPs cover 13 target sections of the following 8 highway network in Northern Vietnam: (i) Hanoi - Lao Cai; (ii) Hanoi - Hai Phong; (iii) Hai Phong - Ha Long - Mong Cai - Van Don; (iv) Hanoi - Bac Giang - Lang Son; (v) Hanoi - Thai Nguyen - Bac Kan - Cao Bang; (vi) Lang - Hoa Lac; (vii) Hoa Lac - Hoa Binh; (viii) Cau Gie - Ninh Binh.

cost minimization (Figure 10).

4.3 Firm Market Access (FMA)

The reduced-form MP index defines the accessible market within geographically proximate neighborhoods (communes within a 50km radius). However, this may inadvertently underestimate the true impact of substantial GMS investments that establish connections between urban core cities and remote mountainous periphery areas across considerably long distances. Moreover, the Harris index lacks micro-foundation based on a spatial equilibrium model.

As suggested by Redding (2022), our preferred specification involves utilizing a theoretically-consistent Firm Market Access (FMA) index, defined below in Equation (3). The FMA index explicitly integrates (a) the elasticity of interregional trade with respect to distance and (b) the impacts of local wage and price indices in determining the market demand for each location, in addition to employment density.

$$FMA_i = \sum_j \tau_{ij}^{1-\sigma} \frac{(w_j L_j)}{P_j^{1-\sigma}}, \tag{3}$$

where τ is the direct measurement of freight transport cost for commune i. The distance elasticity of trade, denoted as σ, is set at 4 based on Simonovska and Waugh (2014). The transport cost parameter is estimated as a unit cost of transportation services using the module in the VES that records trade value and volume for transport companies.

Figure 8 illustrates a notable decrease in interregional real transport costs over a decade (from 2011 to 2020), particularly along the highway in peripheral areas, except for the area around Yên Bái province, which is a mountainous region with geographical constraints that limit the construction of connecting roads to the highways. The estimated FMA consistently indicates a significant enhancement in firm market access, seemingly attributed to improved transport connectivity.

4.4 Instrumental Variable Regression

We exploit the exogenous initial variations in the distance to MST network from each commune's centroid, as well as geographic conditions in the northern Vietnam as IVs for ΔFMA in the first-stage. Geographic conditions significantly differ between flat hinterland of the Red River (core districts) and mountain plateaus in NE and NW regions that are inhabited by tribal groups (peripheral districts). The mountain plateaus are irregular in elevation and very rugged. Temperature and weather within RRD region are also diversified according to terrains and seasons. The initial geographic conditions would affect the amenity and create natural barriers for expanding local markets.

Accordingly, a set of possible IVs, Z_{ct_0}, comprises 1999 district-level variations in historical geographic conditions (as a proxy for city amenities), including the elevation level, average temperature, and the amount of sunshine, all taken from Miguel and Roland (2011). We estimate the FD model with the distance to the hypothetical MST network and its interaction with selected set of geographic variables that exhibit the strongest association with the growth in FMA between 2011 and 2020.

In the second-stage regression, we estimate the effect of firm market access on the number of manufacturing, transport, and trade firms. We will present the following first-difference two-stage least squares (2SLS) regression:

$$\begin{aligned} \Delta \ln FMA_{ct} &= \delta_0 + Z_{ct_0}\delta_1 + X_{ct_0}\delta_2 + \Delta\nu_{ct} \\ \Delta \ln y_{ct} &= \beta_0 + \beta_1 \Delta \ln FMA_{ct} + X_{ct_0}\gamma + \Delta\varepsilon_{ct}, \end{aligned} \tag{4}$$

where we use two census rounds in 2011 and 2020 to compute the first-difference; y is the number of enterprise at the commune-level; and X_{ct_0} is control variables including the initial share of female workers, workers with social security, and the employment share in industrial zones before the highway construction in 2011. We also fix the 2011 employment size in each commune to better identify the market potential growth due to the highway construction. Given that our identifying variation in the MST network is at the commune-level, we cluster standard errors at the commune level.

We estimate Equation (5) separately for the core and periphery subsamples. By doing so, we believe we can address non-linearity arising from differentiated centrifugal and centripetal forces.

4.4.1 IV Results

Table 2 exhibits the IV results in which Kleibergen-Paap F statistics exceeds the rule-of-thumb threshold of 10, showing that our IVs are valid. The results also affirm the positive association between enhanced firm market access and agglomeration, rather than competition, of manufacturing, transport, and trade firms. Notably, the IV estimation results in columns (2), (5), and (8) indicates significantly large elasticity of the agglomeration effect for manufacturing and service firms in the peripheral area (Panel B). This likely reflects disproportionately larger gains of interregional trade for peripheral cities, once market access reaches a certain level, by linking them with Hanoi and Haiphong port, an event more accurately captured in the FMA index. In contrast, The estimated elasticity is relatively muted for core cities, reflecting that the positive effect of market access is more moderate as market access approaches saturation which is also offset by the competition effect, as predicted by our theoretical model.

In columns (3), (6), and (9), we add the square term of FMA growth to capture potential nonlinearity, specifically, whether an increase in firm numbers accelerates or diminishes once market access reaches a certain level. While the estimated coefficient on the squared term is insignificant for core cities, we find strongly positive nonlinear effect, significant at 5% level for transport firms and marginally significant for others in the peripheral cities. These results show an increasing marginal effect as market access improves in these cities. This appears broadly consistent with our model prediction of the *V-shaped* relationship that the agglomeration effect starts to outweigh the negative competition effects as market grows in peripheral cities.

Similar to the empirical pattern found in the literature (Redding and Turner (2015); Jedwab and Storeygard (2020)), larger estimated coefficients from IV than those from OLS imply that the highway network was more likely to be built toward lagging areas in the northern Vietnam where the changes in market access have larger impacts. This significant downward bias may also be the result of measurement error in the market access.

In sum, our empirical findings show that improved firm market access promotes stronger agglomeration economies than competition among manufacturing, transport, and trade firms. The agglomeration elasticity is particularly large in peripheral regions, likely reflecting greater interregional trade gains once these areas are connected to key hubs such as Hanoi and Haiphong. In contrast, core regions exhibit smaller elasticities, consistent with the effects of market access saturation and increased inter-firm competition, as predicted by theory. Overall, these results point to the emergence of polycentric patterns in spatial development and structural transformation.

5 GMS Highways as Source of Polycentric Development

The market integration through lower transport costs could facilitate industrial concentration to central regions with larger market at the expense of peripheral regions. The polycenetric development after the upgrade in inter-regional GMS highways shows the co-existence of agglomeration of firms to core areas as well as diffusion of production to peripheral areas. Since this would be a reflection of alternative mechanisms at play beyond the market size (home market effect), we follow Krugman (1991) and Fujita *et al.* (1999) model and examine the role of transport cost in the following analysis.

5.1 The Difference-in-Differences Regression

The identification of the impact of GMS highway on firms and employments, uncovering the role of transport cost, is challenging as we do not know the treatment group exactly. This is simply because the exposure to upgraded roads are continuous depending on geographic distance from the highway. As the baseline model, we apply a flexible two-way fixed effect DID specification for heterogeneous treatment effect analysis in which physical distance is discretized into a set of indicator variables for being 0-20 km and 20-50km from the GMS highway.[14]

$$
\begin{aligned}
y_{ct} = {} & \beta_1 \mathbb{1}[Dist_c \leq d_1] \times POST + \beta_2 \mathbb{1}[d_1 < Dist_c \leq d_2] \times POST \\
& + \beta_3 \mathbb{1}[Dist_c \leq d_1] \times POST \times t + \beta_4 \mathbb{1}[d_1 < Dist_c \leq d_2] \times POST \times t \\
& + \gamma X_{ct} + \mu_c + \tau_t + \lambda_{dt} + \varepsilon_{ct},
\end{aligned}
\tag{5}
$$

where y_{ct} is the number of firms and employment in commune c for year t. $Dist$ is the distance from the GMS highway, measuring accessibility to the highway for firms locating in each commune. The distance thresholds are set at d_1=20km and d_2=50km. The variable, $POST$, takes one for the post-GMS construction years ($t \geq 2011$). μ_c, τ_t, and λ_{dt} are commune and year fixed effects, and district-level trend prior to the start year of the GMS highway upgrades. X_{ct} controls for the effect of industrial zones. t is a simple time trend.

Similar to Greenstone *et al.* (2010), we first test for a mean shift in the outcome (β_1) among treated communes within 20km or 20-50km from the GMS highway after 2011. In

[14] In terms of the distance, semi-parametric estimates in subsection 4.1 reveal that the districts within 20km distance buffer from the GMS highway experienced rapid expansion in the market potential, while the market expansion decreases to zero above 50km from the highway. Based on this, we use 20km and 50km as the boundary of the highway's treatment zone.

another specification, we allow both a mean shift and a trend break in the outcome (β_1 and β_2) among the treated communes to investigate whether the effect occurs immediately or evolves over the time.

The key identification assumption in the DID regression is that changes at communes far away from the highway (more than 50km away from the highway) could serve as a valid counterfactual for communes near the highway, i.e., they would follow a parallel trend without the highway construction. Figures 11a and b plot the differential trend in the number of firms and employments among the areas within and above 50km from the highway (2010 as the reference year). Reassuringly, communes in each distance group exhibit parallel trend before the highway construction started, which supports the validity of our identifying assumption. After construction began in 2011, the estimated coefficient exhibits an upward trend, which accelerated further after 2015 upon the completion of the main GMS highway construction.

5.1.1 Baseline Results

Agglomeration of Firms. Results in columns 1, 3, and 5 of Table 3 confirms the visual impression in Figure 11 that the starting of highway construction is associated with significantly stronger firm agglomeration in communes near the highway relative to communes located in remote areas (more than 50km away from the highway as the reference group). In core regions, the effect is the largest in the immediate neighborhood of the highway. Within 20km, there was 16%, 15%, and 36% increase in the number of manufacturing, transport, and trade firms respectively on average. For trade business (e.g., retail, food and accommodation) which serves consumers, the agglomeration effect remains strong for 20-50km distance buffer. This may indicate some formation of small trade start-ups in surrounding areas due to the positive demand spillover from the central business areas.

Similarly, the positive effect is evident in peripheral regions where its significance extends to communes located up to 50km from the highway. This suggest some spatial spillover of the GMS impact to the outskirt places which are connected to the GMS highway through secondary road networks. Several second-tier cities (such as Hoa Binh, Son La, and Thai Nguyen; see Figure 3), being connected to the GMS highway through main national roads, have experienced the expansion of manufacturing production (e.g., food processing, textile) and trade services. The wider geographic scale of agglomeration in the peripheral region seems consistent with the satellite-based evidence in Figure 4 as well as an improvement in FMA in the peripheral areas further away from the GMS highway.

Columns 2, 4, and 6 show both the mean shift and trend break coefficients. Using two parameters, we calibrate the estimated change in the number of firms over 10 years as $\beta_1 + 10\beta_2$ (combined effect "after 10 years" as reported at the bottom of the table). In core RRD region, the estimates in three columns suggest that the GMS highway construction is associated with about 35% and 30% increase in manufacturing and transport firms for communes within 20km over 10 years. The combined effect is insignificant for communes within 20-50km distance buffer, except trading firms. In the peripheral NE and NW regions, communes within 20km experience even stronger agglomeration effect, in particular for manufacturing firms that grew by 47% over 10 years. Even outside 20km, the positive agglomeration effect remains significant at about 39%, 13%, and 25% for manufacturing, transport, and trading firms in the 20-50km distance buffer areas.

In summary, the results confirm that highway construction under the GMS program significantly enhanced firm agglomeration, particularly in communes located within 20km of the highway. The effects were especially pronounced for manufacturing firms in peripheral regions and trade services in core regions, with the latter driven by strong positive demand spillovers from central economic hubs. In short, the highway produced a sizable but geographically concentrated boost in the core – a pattern of incremental agglomeration that mirrors the model's predicted plateauing effect in consistent with competition effect near already-developed regions. The initially peripheral regions, in turn, reaped the larger proportional gains once they overcame the entry threshold. While we do not observe an immediate boom in firms in the remote communes, once connectivity surpasses a critical threshold, economically lagged districts started to benefit the most from connectivity-driven agglomeration.

Job Creation. Results in Table 4 similarly show that the beginning of highway construction is associated with significant increase in sector employments near the highway relative to communes outside 50km distance buffer. Using the estimates in columns 2, 4, and 6, the immediate vicinity of the highway in the core area (within 20km) has experienced about 62%, 46%, and 83% increase in manufacturing, transport, and trade jobs over 10 years. In the peripheral area, communes within a 20km radius experienced a substantial increase in manufacturing jobs (78%), while communes in the 20-50km distance also witnessed significant growth in both manufacturing and service job opportunities.

5.2 Proximity to Industrial Zones

As Figure 12 shows, industrial zones are typically located along highways or national roads in northern Vietnam. Therefore, communes near the highway likely benefit both from improved transport connectivity and the effect of industrial zones. Inclusion of such communes in the treatment group could overestimate the treatment effect of the GMS.

To remove the confounding effect of industrial zones, we first geocode the location of all industrial zones in northern Vietnam using the Google Earth Pro as well as their establishment year. Then, we examine the stability of the DID estimates in tables 3 and 4 for the sub-sample when communes within a certain distance (1, 5, and 10km) from the centroid of each industrial zones are removed.

As Table 5 shows, dropping communes within close proximity to the industrial zones generally reduces the magnitude of the mean shift and trend break estimates. This reflects that the earlier estimate also includes the positive agglomeration effect due to the industrial zones. In Table 5(a), when affected communes within 5km or 10km from the industrial zones are dropped from the sample, the magnitude of the mean shift estimate (β_1) gets smaller in both RRD and NE & NW regions.

The 10-year agglomeration effect ($\beta_1 + 10\beta_2$) also tends to decrease in both core and peripheral cities due to smaller trend break estimates. The reduction in the combined effect is pronounced for communes located within a 20km radius in peripheral areas, where the combined effect is almost halved when communes located within 10km from industrial zones are dropped. This reflects the critical role of place-based policy in attracting firms in the peripheral cities in northern Vietnam. Table 5(b) similarly shows the robustness of the GMS impact on job creation. Although the magnitude of treatment effect decreases, our main findings remain unchanged.

5.3 Technology Spillover

In urban centers, the clustering of firms may be partially attributed to positive production externality resulting from the exchange of knowledge and technology through formal or informal interaction (Greenstone *et al.* (2010)). Existing literature suggests that such spillover effects play an important role in stimulating innovation, particularly in certain high-tech industries. This dynamics could be pertinent to the small neighborhood surrounding the center of Hanoi, where foreign firms and high-tech companies concentrate.

However, an earlier VES survey reports firm-level R&D investment, which record only 3% of communes report positive R&D spending nationwide. At the intensive margin, the R&D spending is in very small scale (approximately 0.1% of each commune's gross regional output on average).[15] This evidence implies that innovation-related spending (which serves as a proxy for the extent of technology spillovers/knowledge sharing among firms) is likely be limited in Vietnam. At most, innovation spending has been limited only to small urban core neighborhoods. In peripheral areas, the possibility of local technology spillover creating substantial externalities and explaining a significant agglomeration effect is improbable. Therefore, we believe that this factor does not pose a major concern for our main findings.

6 Heterogeneous Treatment by Supply Chain Linkages

We also consider supply chain linkages as a key driver of polycentric spatial development. Our model (Appendix B) based on New Economic Geography (NEG) literature (Krugman and Venables (1995); Puga and Venables (1996); Venables (1996); Redding and Venables (2004)) considers the interaction between transport costs and trade in differentiated intermediate goods, which creates externalities by linking firms and consumers. The development of supply chains network makes peripheral region, with higher transport costs but cheaper initial wages, more attractive. As backward linkage with upstream industries and forward linkage with downstream customers tighten, positive intermediate production spillovers would reshape the core-periphery pattern. As transport costs drop further, the model predicts that the industrial location could diffuse to peripheral regions, which results in convergence in core-periphery income disparity.

Besides physical locational distance, the impact of GMS highway construction would differ by each commune's particular production structure in local supply chains. The target causal parameter β (GMS highway impact) would be larger if firms in the commune are strongly interconnected to firms located in the same or different communes within accessible markets. An exposure to transport cost reduction would differ by the extent of intermediate goods transactions and transport service usage. The identification of the impact of transport investment also requires the specification that accounts for indirect spillovers through local intermediate production chains.

[15] Gross regional output at the commune level is measured by the sum of total labor income and firm's gross operating profits, which are available in the VES.

In this regard, we extend the event study to deal with input-output linkages facilitated by the transport network (Ellison *et al.* (2010)). Given the absence of interfirm trade data for Vietnam, we adopt the approach similar to Li *et al.* (2017) and assess intersectoral supply chain linkage using the OECD's harmonized input-output (IO) table, encompassing 45 sectors, available for all years up to 2018. This data enables us to gauge the overall strength of IO linkages, typically measured by the Leontief inverse of the IO matrix. To explicitly account for production chains, we examine alternative IO statistics such as *upstreamness* (the distance of each industry from consumers in Vietnam's supply chain, *a la* Antràs *et al.* (2012)). Subsequently, we define local upstreamness (M_{ct}) as the average of industry-level index L_{st} for each commune c in year t using the initial share of each industry s's sales as the weighting factor:

$$M_{ct} = \sum_s L_{st} \frac{Sales_{cst}}{\sum Sales_{ct}}. \tag{6}$$

Figure 13 illustrates the spatial distribution of local production multipliers in northern Vietnam. Given high import dependency of raw materials and intermediate goods from China, local industries initially exhibit weak interlinkages in 2006.[16] Over the course of a decade, there is a clear and significant deepening of supply chain strength, particularly evident along the road network in the northwestern part of Vietnam. The deepened production linkage would also be justified by Vietnam government's industrial policy to foster domestic suppliers and small-sized support industries in target areas (e.g., mechanical engineering, electronics, assembly of machines, textiles, and high-techs).[17] The VES indicates increasing concentration of upstream manufacturing industries (including food processing, wood and paper production, rubber and plastic products, electric products) in peripheral areas. These industries purchase materials and parts from local suppliers and supply processed goods to downstream transporters, distributers, as well as wholesale or retail businesses. Additionally, the diffusion of textile industries is observed along the southeastern coast of Hanoi, extending towards the South China Sea.

Using the constructed IO metric and generalized DID design (Lindo *et al.* (2020); Callaway *et al.* (2021)), we aim to estimate the average causal response (ACR) to a unit change in local upstreamness M. We estimate heterogeneous treatment at finer 10km distance

[16]While empirical studies on Vietnam's production network are scarce, several studies, such as Okumura (2021), point immature local supporting industries with low value creations as weakness for Vietnam's industrial development.

[17]See the Prime Minister's Resolution No. 12/2011/QD-TTg and No. 1483/QD-TTg.

bins, and for each distance group, estimate ATT for communes with treatment level $j \in T$ relative to adjacent communes with treatment $j-1 \in T$. Like Lindo *et al.* (2020), one way to recover the ACR for treatment group j is to estimate ATT for a given treatment level j (i.e., the interaction term β):

$$y_{ct} = \beta \mathbb{1}[Q(Dist_c) = q] \times POST \times M_{ct} + \gamma M_{ct} + \delta X_{ct} + \mu_c + \tau_t + \lambda_{dt} + \varepsilon_{ct}. \tag{7}$$

Instead of estimating β by 10km bins, $\mathbb{1}[Q(Dist_c) = q]$ is a dummy for the equal-sized decile of the GMS distance from each commune's centroid. In the continuous treatment DID design, we compare the heterogeneity in gains from the treatment, i.e., the ATT for treatment group j and a group just below it $j-1$ in the supply chain (Callaway *et al.*, 2021). The identification needs equal ATT, i.e., communes at the different location of the supply chain should be similarly treated by the GMS investment.

In reality, each commune would have heterogeneous expected returns from the GMS by moving up or down the industrial layer of Vietnam's supply chain due to several factors, such as an initial industrial composition and the level of industrial development in each locality. If firms choose locations based on the difference in an expected treatment effect, this could potentially bias our estimates for a particular ACR group. To minimize the selection bias, our approach is to control for pre-trend term (the growth of number of firms or employment before the GMS investment) in all our regressions.

6.1 Robustness Check to Spillovers Through Local Roads

To improve the identification of the impact of GMS highway upgrade, we also run an alternative specification that accounts for spatial spillovers to neighborhood control areas through local road networks. Communes that are not in the neighborhood of the highway but located on local roads could be indirectly affected by the highway construction via the secondary local transport links. Inclusion of such communes within control group could possibly attenuate the direct impact of the highway upgrade. To isolate the indirect effect due to the treatment spillovers, we re-estimate Equation (7) after removing the communes where two main national roads (NH3 and NH13) path through. This robustness check is separately reported below in Figure 14 as dash lines.

6.2 Results of Heterogeneous Treatment Effect

We plot the coefficient β of the ACR to explore heterogeneous effects of the GMS highway through input-output linkages, separately for core (shown in the left chart) and periphery regions (right chart). In the case of manufacturing firms, we examine heterogeneous effects stemming from upward demand linkages - the channel through input use in the production process (Figure 14). The finding reveals a more pronounced input-output production multiplier, as reflected by larger ACR estimates when the proximity to the GMS highway is closer.

In Figure 14, manufacturing firm agglomeration manifests due to upward demand linkage, with equally strong magnitude in both RRD and NE & NW regions. This result underscores the presence of upstream industries actively sourcing local materials and products. The improvement in transport connectivity near the GMS highway has facilitated intermediate trade, thereby promoting the co-agglomeration of manufacturing firms in both regions. Nevertheless, in remote peripheral areas located beyond 50km from the GMS highway, the positive multiplier effect seems to decline at a faster rate. This trend is likely attributable to weaker production network linkages in rural area.

We detect non-linearity in multiplier effects in regard to physical distance from the GMS highway. For example, we observe a sharp dip at the second percentile in the peripheral regions. Also, we observe some picks in the multiplier effect in locations around 40km or 70km away from the GMS highway in the RRD region. In the NE & NW regions, the multiplier effect also gets stronger in communes located beyond 80km away from the GMS highway. The non-linearity in β coefficient in regard to distance would be driven by positive externality spatial spillovers occurred in several large second-tier cities. When second-tier cities are excluded, the effect gets flattened out (dash line).

Generally speaking, large ACR near the highway indicates positive externality resulting from production spillovers. The locations near the GMS highway benefit directly from reduced transport costs, which gets amplified when local industries are upstream and tightly interconnected. While the direct impact of the GMS highways is consistently large on average, the production externality is relatively weaker in peripheral areas except for large secondary cities located along the highways.[18]

The overall findings from the generalized DID estimation supports that the GMS project

[18]Similar results could be obtained for the GMS effect on commune's total employment. We find significant positive treatment externality on manufacturing employments in both core and periphery regions near the GMS highway. Results are available upon request.

facilitated significant economy-wide industrialization. In core region, production externality on manufacturing firms and employments concentrates in communes near the highway, which shapes typical core-periphery pattern. For service trade, positive externality diffuses more equality over space possibly through consumers' demand linkages. In the peripheral regions, the spillovers appear to cover wider geographic scale due to production linkages via local transport networks.

However, the linkage effects appear to be stronger in core regions than in peripheral ones, suggesting that supply chain linkages alone cannot account for the emergence of polycentric development. Instead, this spatial pattern is more plausibly driven by catch-up effects from a low baseline level and lower competitive pressure in peripheral areas.

7 Conclusions

This paper examined how interregional transport investments reshaped the spatial economic structure of northern Vietnam, a region historically affected by conflict and under-development. We found that the upgrading of the GMS highway promoted broad-based industrialization by connecting peripheral rural areas to core urban centers. Improved connectivity expanded market potential within a 50-kilometer radius of the GMS corridor, reduced trade costs, and fostered strong agglomeration economies among manufacturing firms in both core and peripheral districts.

Satellite imageries revealed a transition from a monocentric urban structure – previously centered around Hanoi– to a more polycentric spatial configuration. Our results suggested that this shift was largely driven by the GMS interregional transport investment. Notably, we found significant heterogeneity in its impact across regions, with disproportionately larger agglomeration benefits found for peripheral areas due to improved firm market access. The asymmetric impact reflects underlying differences in industrial linkages: core areas benefited from dense input-output networks, whereas peripheral districts – initially lagging – experienced greater market access gains that enabled catch-up dynamics. Once market access surpassed a certain threshold, agglomeration forces began to dominate competition effects, supporting manufacturing firm growth and employment expansion along the highway. Together, these factors have contributed to the emergence of a polycentric spatial distribution of economic activity.

Difference-in-difference estimates confirmed substantial firm agglomeration along the

GMS corridor, especially in manufacturing. Urban core districts within 20km radius of the highway saw a 35% increase in manufacturing firms over a decade. The surrounding peripheral rural areas recorded larger increase of 47% within 20km and and 39% within 50km radius of the highway, which diffused over wider spatial scale. Employment rose in parallel, particularly in rural areas, indicating spatially broader industrialization. Within the core region, we also observed the diffusion of textile and service industries from Hanoi to adjacent coastal areas towards the South China Sea. These patterns remained robust after accounting for the proximity to industrial zones and technology spillovers. This indicates that transport infrastructure not only enhanced market access, but it itself was instrumental in shaping the spatial distribution of economic activity.

Finally, generalized difference-in-difference analysis showed significant heterogeneity in the treatment effect of the GMS highway. Specifically, manufacturing firms with stronger upstream linkages were more likely to cluster near the highway, generating significant input-output multiplier effects. This finding remains robust when accounting for the spatial spillovers to second-tier peripheral cities through national roads. However, since these effects were more pronounced in core regions, supply chain linkages alone could not explain the emergence of polycentric development. Rather, the transformation also reflected the activation of agglomeration dynamics in peripheral areas once market access reached critical mass.

Taken together, our findings demonstrated the potential of well-designed infrastructure investments to catalyze broad development payoffs through spatial integration, industrial clustering, and inclusive regional development, even in regions shaped by historical conflict and marginalization.

As a next step, calibrating a spatial structural general equilibrium model will allow for deeper insights into the mechanisms driving these transformations, building on insights from our reduced-form empirical analysis. This approach enables a structural interpretation of the dynamic interplay between transport investment, market access, and spatial economic transformation. More specifically, there are also two important directions for extending our current study. First, incorporating open-economy features into the model would allow us to capture the critical synergistic effects between port infrastructure and a well-maintained road network in facilitating trade (Elhan-Kayalar *et al.*, 2024). Second, regarding the role of industrial zones, it is essential to evaluate the aggregate impact of place-based policies through the lens of production networks in Vietnam (Kono *et al.*, 2025). We believe these represent crucial avenues for future research.

References

AHLFELDT, G. M. and FEDDERSEN, A. (2018). From periphery to core: measuring agglomeration effects using high-speed rail. *Journal of Economic Geography*, **18** (2), 355–390.

AKAMATSU, T., MORI, T., OSAWA, M. and TAKAYAMA, Y. (2019). Multimodal agglomeration in economic geography. *arXiv preprint arXiv:1912.05113*.

ALLEN, T. and ARKOLAKIS, C. (2014). Trade and the topography of the spatial economy. *Quarterly Journal of Economics*, **129** (3), 1085–1140.

ANTRÀS, P., CHOR, D., FALLY, T. and HILLBERRY, R. (2012). Measuring the upstreamness of production and trade flows. *American Economic Review*, **102** (3), 412–416.

ASIAN DEVELOPMENT BANK (2020). *Asia's Journey to Prosperity: Policy, Market, and Technology Over 50 Years*. Manila, Philippines: Asian Development Bank.

BANERJEE, A., DUFLO, E. and QIAN, N. (2020). On the road: Access to transportation infrastructure and economic growth in China. *Journal of Development Economics*, **145**, 102442.

BAUM-SNOW, N. (2007). Did highways cause suburbanization? *Quarterly Journal of Economics*, **122** (2), 775–805.

—, BRANDT, L., HENDERSON, J. V., TURNER, M. A. and ZHANG, Q. (2017). Roads, railroads, and decentralization of Chinese cities. *Review of Economics and Statistics*, **99** (3), 435–448.

BURCHFIELD, M., OVERMAN, H. G., PUGA, D. and TURNER, M. A. (2006). Causes of sprawl: A portrait from space. *Quarterly Journal of Economics*, **121** (2), 587–633.

CALLAWAY, B., GOODMAN-BACON, A. and SANT'ANNA, P. H. (2021). Difference-in-differences with a continuous treatment. *arXiv preprint arXiv:2107.02637*.

CHANDRA, A. and THOMPSON, E. (2000). Does public infrastructure affect economic activity?: Evidence from the rural interstate highway system. *Regional Science and Urban Economics*, **30** (4), 457–490.

COMBES, P.-P. and GOBILLON, L. (2015). The empirics of agglomeration economies. *Handbook of Regional and Urban Economics*, **5**, 247–348.

DINKELMAN, T. (2011). The effects of rural electrification on employment: New evidence from South Africa. *American Economic Review*, **101** (7), 3078–3108.

DONALDSON, D. and HORNBECK, R. (2016). Railroads and American economic growth: A "market access" approach. *Quarterly Journal of Economics*, **131** (2), 799–858.

— and STOREYGARD, A. (2016). The view from above: Applications of satellite data in economics. *Journal of Economic Perspectives*, **30** (4), 171–98.

DUFLO, E. and PANDE, R. (2007). Dams. *Quarterly Journal of Economics*, **122** (2), 601–646.

DURANTON, G. and TURNER, M. A. (2012). Urban growth and transportation. *Review of Economic Studies*, **79** (4), 1407–1440.

ECKERT, F. and PETERS, M. (2022). Spatial structural change, NBER Working Paper No. 30489.

ELHAN-KAYALAR, Y., KUCHERYAVYY, K., NOSE, M., SAWADA, Y., SHANGGUAN, R. and TUNG, N. T. (2024). Foreign direct investment location and trade dynamics in Viet Nam after the US–PRC trade war, memograph.

ELLISON, G., GLAESER, E. L. and KERR, W. R. (2010). What causes industry agglomeration? evidence from coagglomeration patterns. *American Economic Review*, **100** (3), 1195–1213.

FABER, B. (2014). Trade integration, market size, and industrialization: evidence from China's national trunk highway system. *Review of Economic Studies*, **81** (3), 1046–1070.

FAJGELBAUM, P. and REDDING, S. J. (2021). Trade, structural transformation and development: Evidence from Argentina 1869-1914.

FORSLID, R. and OTTAVIANO, G. I. (2003). An analytically solvable core-periphery model. *Journal of Economic Geography*, **3** (3), 229–240.

FUJITA, M., KRUGMAN, P. R. and VENABLES, A. (1999). *The spatial economy: Cities, regions, and international trade*. MIT press.

GHANI, E., GOSWAMI, A. G. and KERR, W. R. (2016). Highway to success: The impact of the Golden Quadrilateral project for the location and performance of Indian manufacturing. *Economic Journal*, **126** (591), 317–357.

GIBBONS, S., LYYTIKÄINEN, T., OVERMAN, H. G. and SANCHIS-GUARNER, R. (2019). New road infrastructure: The effects on firms. *Journal of Urban Economics*, **110**, 35–50.

GIBSON, J., OLIVIA, S., BOE-GIBSON, G. and LI, C. (2021). Which night lights data should we use in economics, and where? *Journal of Development Economics*, **149**.

GLAESER, E. L. and KAHN, M. E. (2001). Decentralized employment and the transformation of the American city.

— and — (2004). Sprawl and urban growth. *Handbook of Regional and Urban Economics*, **4**, 2481–2527.

GMS SECRETARIAT (2018). *Assessment of GMS Economic Corridors - Viet Nam*. Report.

GREENSTONE, M., HORNBECK, R. and MORETTI, E. (2010). Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings. *Journal of Political Economy*, **118** (3), 536–598.

HENDERSON, J. V. (1974). The sizes and types of cities. *American Economic Review*, pp. 640–656.

JEDWAB, R. and STOREYGARD, A. (2020). The average and heterogeneous effects of transportation investments: Evidence from sub-Saharan Africa 1960-2010. *Journal of the European Economic Association*.

JHA, M. K., MCCALL, C. and SCHONFELD, P. (2001). Using gis, genetic algorithms, and visualization in highway development. *Computer-Aided Civil and Infrastructure Engineering*, **16** (6), 399–414.

JONG, J.-C. and SCHONFELD, P. (2003). An evolutionary model for simultaneously optimizing three-dimensional highway alignments. *Transportation Research Part B: Methodological*, **37** (2), 107–128.

KONO, H., NOSE, M. and SAWADA, Y. (2025). Industrial zones and production networks: Evidence from Vietnam, mimeograph.

KRUGMAN, P. (1991). Increasing returns and economic geography. *Journal of Political Economy*, **99** (3), 483–499.

— and VENABLES, A. J. (1995). Globalization and the inequality of nations. *Quarterly Journal of Economics*, **110** (4), 857–880.

LI, Z., WU, M. and CHEN, B. R. (2017). Is road infrastructure investment in China excessive? evidence from productivity of firms. *Regional Science and Urban Economics*, **65**, 116–126.

LINDO, J. M., MYERS, C. K., SCHLOSSER, A. and CUNNINGHAM, S. (2020). How far is too far? new evidence on abortion clinic closures, access, and abortions. *Journal of Human Resources*, **55** (4), 1137–1160.

LU, P.-H. and TSAI, J.-F. (2025). A footloose entrepreneur model with productivity heterogeneity in two regions. *Australian Economic Papers*, **64** (1), 121–138.

MICHAELS, G. (2008). The effect of trade on the demand for skill: Evidence from the interstate highway system. *Review of Economics and Statistics*, **90** (4), 683–701.

MIGUEL, E. and ROLAND, G. (2011). The long-run impact of bombing Vietnam. *Journal of Development Economics*, **96** (1), 1–15.

MIYAZAKI, H., BHUSHAN, H. and WAKIYA, K. (2019). Urban growth modeling using historical landsat satellite data archive on google earth engine.

MORTEN, M. and OLIVEIRA, J. (2024). The effects of roads on trade and migration: Evidence from a planned capital city. *American Economic Journal: Applied Economics*, **16** (2), 389–421.

NAGY, D. K. (2020). Hinterlands, city formation and growth: Evidence from the U.S. westward expansion, university of Pompeu Fabra.

OKUBO, T. (2010). Firm heterogeneity and location choice. *Kobe University, DP2010-11*.

PUGA, D. and VENABLES, A. J. (1996). The spread of industry: spatial agglomeration in economic development. *Journal of the Japanese and International Economies*, **10** (4), 440–464.

REDDING, S. and VENABLES, A. J. (2004). Economic geography and international inequality. *Journal of International Economics*, **62** (1), 53–82.

REDDING, S. J. (2022). Trade and geography. *Handbook of International Economics*, **5**, 147–217.

— and TURNER, M. A. (2015). Transportation costs and the spatial organization of economic activity. *Handbook of Regional and Urban Economics*, **5**, 1339–1398.

ROBINSON, P. M. (1988). Root-n-consistent semiparametric regression. *Econometrica*, **56**, 931–954.

SIMONOVSKA, I. and WAUGH, M. E. (2014). The elasticity of trade: estimates and evidence. *Journal of International Economics*, **92** (1), 34–50.

TREW, A. (2020). Endogenous infrastructure development and spatial takeoff in the first industrial revolution. *American Economic Journal: Macroeconomics*, **12** (2), 44–93.

VENABLES, A. J. (1996). Equilibrium locations of vertically linked industries. *International Economic Review*, pp. 341–359.

TABLE 1: Descriptive Statistics

(a) Changes in District Population, 1999–2019

	Number of districts	Population (people)			Population Growth (%)	
		In 1999	In 2009	In 2019	1999–2009	2009–2019
Red River Delta	129	138,282	151,809	174,752	13.1	15.0
North East	82	84,018	85,079	95,364	5.5	10.0
North West	59	59,368	69,103	79,882	22.8	18.7
Total (National)	696	108,907	121,415	135,874	17.2	11.4

(b) Changes in Number of Enterprises

	All firms			Manufacturing			Services		
	2011	2016	2020	2011	2016	2020	2011	2016	2020
Red River Delta	104,871	161,930	258,109	15,926	24,091	38,855	73,036	114,702	181,433
North East	10,294	13,875	24,729	1,783	2,434	4,891	5,382	7,889	13,608
North West	5,476	7,818	12,285	727	891	1,464	2,690	4,230	6,525
Total (National)	329,121	516,803	822,677	52,596	76,312	121,951	220,581	361,047	566,085

NOTE.—Data in Panel (a) are at the district level and district boundaries are adjusted based on the GADM version 3.6 (as of 2017). Data are from Population Census 1999, 2009, and 2019. Panel (b) includes all types of registered firms. Data are from Enterprises Census 2011, 2016, and 2020. Total covers all 8 regions in Vietnam.

TABLE 2: Agglomeration Effect using Firm Market Access

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A. RRD regions									
	Manufacturing			Service: transport			Service: trade		
Δ Log FMA	0.046***	0.035	0.047*	0.039***	0.045**	0.034	0.047***	0.076***	0.091***
	(0.006)	(0.023)	(0.026)	(0.005)	(0.021)	(0.024)	(0.006)	(0.021)	(0.024)
Δ Log FMA squared			0.009			-0.007			0.010
			(0.008)			(0.008)			(0.008)
Industrial zone ($\leq 5km$) in 2011	0.139***	0.139***	0.142***	0.213***	0.213***	0.211***	0.171***	0.170***	0.174***
	(0.038)	(0.038)	(0.038)	(0.038)	(0.038)	(0.038)	(0.037)	(0.037)	(0.037)
N	2,387	2,387	2,387	2,387	2,387	2,387	2,387	2,387	2,387
Region FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Model	OLS	2SLS	2SLS	OLS	2SLS	2SLS	OLS	2SLS	2SLS
Kleibergen-Paap F statistics		118.40	52.74		118.40	52.74		118.40	52.74
Panel B. NE & NW regions									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Manufacturing			Service: transport			Service: trade		
Δ Log FMA	0.024***	0.240***	0.309***	0.006	0.114***	0.177***	0.013**	0.203***	0.270***
	(0.006)	(0.041)	(0.058)	(0.005)	(0.026)	(0.037)	(0.006)	(0.036)	(0.055)
Δ Log FMA squared			0.017			0.016**			0.017
			(0.011)			(0.007)			(0.010)
Industrial zone ($\leq 5km$) in 2011	0.306***	0.172*	0.195**	0.163***	0.083	0.104	0.174**	0.022	0.043
	(0.085)	(0.098)	(0.098)	(0.063)	(0.066)	(0.066)	(0.087)	(0.093)	(0.095)
N	1,576	1,437	1,437	1,576	1,437	1,437	1,576	1,437	1,437
Region FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Model	OLS	2SLS	2SLS	OLS	2SLS	2SLS	OLS	2SLS	2SLS
Kleibergen-Paap F statistics		45.06	29.80		45.06	29.80		45.06	29.80

NOTE.—*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors, clustered at the commune level, are reported in the bracket. First-difference regressions take differences from 2011 to 2020 (the long-term effect). Distance to MST network and its interaction with land elevation, temperature, and sunshine are used as the IV in columns 2, 3, 5, 6, 8 and 9.

TABLE 3: Difference-in-Differences Specification: Agglomeration Effect of GMS Highway

	Log (number of registered firms)					
	Manufacturing		Service: transport		Service: trade	
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: RRD region					
β_1: Within 20km	0.161***	-0.066	0.149**	-0.041	0.360***	0.058
	(0.060)	(0.058)	(0.058)	(0.050)	(0.057)	(0.054)
β_2: 20-50km	0.028	-0.101*	-0.055	-0.055	0.165***	-0.060
	(0.061)	(0.060)	(0.059)	(0.051)	(0.058)	(0.055)
β_3: Post-trend (within 20km)		0.041***		0.035***		0.055***
		(0.009)		(0.007)		(0.007)
β_4: Post-trend (20-50km)		0.023***		-0.000		0.041***
		(0.009)		(0.007)		(0.008)
Observations	33,429	33,429	33,429	33,429	33,429	33,429
Combined effect of β_1 and β_2						
After 10 years (20km)		0.347***		0.304***		0.607
After 10 years (50km)		0.133		-0.056		0.349***
	Panel B: NE & NW regions					
β_1: Within 20km	0.260***	0.011	0.130***	-0.024	0.180***	0.005
	(0.032)	(0.028)	(0.024)	(0.022)	(0.030)	(0.029)
β_2: 20-50km	0.216***	0.003	0.069***	-0.008	0.134***	-0.006
	(0.026)	(0.025)	(0.019)	(0.019)	(0.026)	(0.025)
β_3: Post-trend (within 20km)		0.045***		0.028***		0.032***
		(0.005)		(0.004)		(0.005)
β_4: Post-trend (20-50km)		0.039***		0.014***		0.025***
		(0.004)		(0.003)		(0.004)
Observations	19,429	19,429	19,429	19,429	19,429	19,429
Combined effect of β_1 and β_2						
After 10 years (20km)		0.465***		0.257***		0.324***
After 10 years (50km)		0.391***		0.133***		0.249***
Commune FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Pre-trend controlled	YES	YES	YES	YES	YES	YES

NOTE.—*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors, clustered at the commune level, are reported in the bracket. In columns 2, 4, and 6 of each panel, the combined effect of β_1 and β_2 reports the linear combination of the mean shift and trend break parameters ($\beta_1 + 10\beta_2$). It estimates the cumulative effect of the GMS highway a decade after the initiation of main GMS highway upgrades in northern Vietnam.

TABLE 4: Difference-in-Differences Specification: Job Creation Effect of GMS Highway

	Log (employment in registered firms)					
	Manufacturing		Service: transport		Service: trade	
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: RRD region					
β_1: Within 20km	0.396**	0.115	0.226**	-0.060	0.514***	0.124
	(0.162)	(0.182)	(0.101)	(0.101)	(0.121)	(0.136)
β_2: 20-50km	0.342**	0.015	-0.061	-0.116	0.330***	-0.012
	(0.169)	(0.189)	(0.105)	(0.106)	(0.125)	(0.140)
β_3: Post-trend (within 20km)		0.051*		0.052***		0.071***
		(0.028)		(0.014)		(0.016)
β_4: Post-trend (20-50km)		0.059**		0.010		0.062***
		(0.029)		(0.014)		(0.017)
Observations	33,429	33,429	33,429	33,429	33,429	33,429
Combined effect of β_1 and β_2						
After 10 years (20km)		0.625***		0.460***		0.834***
After 10 years (50km)		0.610**		-0.017		0.611***
	Panel B. NE & NW regions					
β_1: Within 20km	0.524***	0.217**	0.183***	-0.056	0.261***	0.116
	(0.105)	(0.099)	(0.062)	(0.065)	(0.067)	(0.074)
β_2: Within 20-50km	0.576***	0.244***	0.142***	0.008	0.223***	0.060
	(0.093)	(0.092)	(0.051)	(0.055)	(0.060)	(0.065)
β_3: Post-trend (within 20km)		0.056***		0.044***		0.027**
		(0.015)		(0.009)		(0.010)
β_4: Post-trend (20-50km)		0.060***		0.024***		0.030***
		(0.013)		(0.007)		(0.009)
Observations	19,429	19,429	19,429	19,429	19,429	19,429
Combined effect of β_1 and β_2						
After 10 years (20km)		0.776***		0.381***		0.381***
After 10 years (50km)		0.848***		0.252***		0.357***
Commune FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Pre-trend controlled	YES	YES	YES	YES	YES	YES

NOTE.—*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors, clustered at the commune level, are reported in the bracket. In columns 2, 4, and 6 of each panel, the combined effect of β_1 and β_2 reports the linear combination of the mean shift and trend break parameters ($\beta_1 + 10\beta_2$). It estimates the cumulative effect of the GMS highway a decade after the initiation of main GMS highway upgrades in northern Vietnam.

TABLE 5: Robustness Check: Dropping Communes near Industrial Zones

(a) Agglomeration Effect

	Log(number of firms)								
	Panel A. RRD region								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Manufacturing			Service: transport			Service: trade		
Drop distance from IZ	$\leq 1km$	$\leq 5km$	$\leq 10km$	$\leq 1km$	$\leq 5km$	$\leq 10km$	$\leq 1km$	$\leq 5km$	$\leq 10km$
Mean shift									
β_1: Within 20km	0.157***	0.137**	0.130**	0.142**	0.095	0.015	0.355***	0.326***	0.289***
	(0.060)	(0.060)	(0.063)	(0.058)	(0.058)	(0.059)	(0.057)	(0.057)	(0.059)
β_2: 20-50km	0.028	0.016	-0.026	-0.058	-0.093	-0.121**	0.164***	0.141**	0.092
	(0.061)	(0.061)	(0.063)	(0.059)	(0.058)	(0.059)	(0.058)	(0.059)	(0.059)
Combined effect of β_1 and β_2									
After 10 years (20km)	0.338***	0.299***	0.318***	0.292***	0.214***	0.104	0.601***	0.559***	0.532***
After 10 years (50km)	0.134	0.128	0.105	-0.060	-0.120	-0.192**	0.347***	0.312***	0.264***
Observations	33,065	25,657	14,208	33,065	25,657	14,208	33,065	25,657	14,208
	Panel B. NE & NW region								
Mean shift									
β_1: Within 20km	0.256***	0.205***	0.133***	0.124***	0.077***	0.041*	0.174***	0.106***	0.036
	(0.032)	(0.032)	(0.034)	(0.024)	(0.023)	(0.024)	(0.030)	(0.031)	(0.032)
β_2: 20-50km	0.217***	0.207***	0.189***	0.068***	0.053***	0.049***	0.133***	0.118***	0.114***
	(0.026)	(0.026)	(0.027)	(0.019)	(0.018)	(0.018)	(0.026)	(0.027)	(0.028)
Combined effect of β_1 and β_2									
After 10 years (20km)	0.462***	0.375***	0.230***	0.247***	0.179***	0.086**	0.314***	0.218***	0.093*
After 10 years (50km)	0.392***	0.380***	0.355***	0.133***	0.110***	0.096***	0.247***	0.222***	0.218***
Observations	19,324	17,604	15,003	19,324	17,604	15,003	19,324	17,604	15,003
Commune & Year FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES
Pre-trend controlled	YES	YES	YES	YES	YES	YES	YES	YES	YES

NOTE.—*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors, clustered at the commune level, are reported in the bracket. Columns 1, 4, and 7 present the estimates for a sub-sample that excludes communes located within 1km from the centroid of the closest industrial zones (IZs). Similarly, columns 2, 5, 8 as well as columns 3, 6, and 9 implement the same robustness checks by excluding communes within 5km and 10km from the IZs, respectively. The combined effect of β_1 and β_2 reports the linear combination of the mean shift and trend break parameters ($\beta_1 + 10\beta_2$). It estimates the cumulative effect of the GMS highway a decade after the initiation of main GMS highway upgrades in northern Vietnam.

(b) Job Creation

	Log(employment in registered firms) Panel A. RRD region								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
		Manufacturing			Service: transport			Service: trade	
Drop distance from IZ	$\leq 1km$	$\leq 5km$	$\leq 10km$	$\leq 1km$	$\leq 5km$	$\leq 10km$	$\leq 1km$	$\leq 5km$	$\leq 10km$
Mean shift									
β_1: Within 20km	0.390**	0.432***	0.572***	0.214**	0.142	0.094	0.509***	0.502***	0.508***
	(0.162)	(0.164)	(0.173)	(0.101)	(0.102)	(0.106)	(0.121)	(0.123)	(0.127)
β_2: 20-50km	0.345**	0.366**	0.320*	-0.066	-0.109	-0.136	0.331***	0.319**	0.245*
	(0.169)	(0.171)	(0.176)	(0.105)	(0.106)	(0.107)	(0.125)	(0.127)	(0.129)
Combined effect of β_1 and β_2									
After 10 years (20km)	0.610***	0.656***	1.019***	0.442***	0.335**	0.305**	0.829***	0.822***	0.881***
After 10 years (50km)	0.615**	0.712***	0.815***	-0.023	-0.086	-0.159	0.612***	0.614***	0.574***
Observations	33,065	25,657	14,208	33,065	25,657	14,208	33,065	25,657	14,208
	Panel B. NE & NW region								
Mean shift									
β_1: Within 20km	0.517***	0.417***	0.247**	0.176***	0.071	0.043	0.259***	0.169**	0.113
	(0.106)	(0.111)	(0.120)	(0.062)	(0.063)	(0.065)	(0.068)	(0.070)	(0.077)
β_2: Within 20km	0.579***	0.535***	0.482***	0.141***	0.111**	0.097**	0.223***	0.191***	0.192***
	(0.093)	(0.093)	(0.097)	(0.050)	(0.048)	(0.048)	(0.060)	(0.060)	(0.062)
Combined effect of β_1 and β_2									
After 10 years (20km)	0.767***	0.620***	0.300*	0.375***	0.250***	0.150	0.377***	0.286***	0.218*
After 10 years (50km)	0.850***	0.819***	0.750***	0.255***	0.206***	0.165**	0.355***	0.314***	0.321***
Observations	19,324	17,604	15,003	19,324	17,604	15,003	19,324	17,604	15,003
Commune & Year FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES
Pre-trend controlled	YES	YES	YES	YES	YES	YES	YES	YES	YES

NOTE.—*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors, clustered at the commune level, are reported in the bracket. Columns 1, 4, and 7 present the estimates for a sub-sample that excludes communes located within 1km from the centroid of the closest industrial zones (IZs). Similarly, columns 2, 5, 8 as well as columns 3, 6, and 9 implement the same robustness checks by excluding communes within 5km and 10km from the IZs, respectively. The combined effect of β_1 and β_2 reports the linear combination of the mean shift and trend break parameters ($\beta_1 + 10\beta_2$). It estimates the cumulative effect of the GMS highway a decade after the initiation of main GMS highway upgrades in northern Vietnam.

FIGURE 1: Sample: Core vs. Periphery

((a)) Core Cities: Red River Delta (RRD) region

((b)) Peripheral Cities: Northeast (NE) and Northwestern (NW) regions

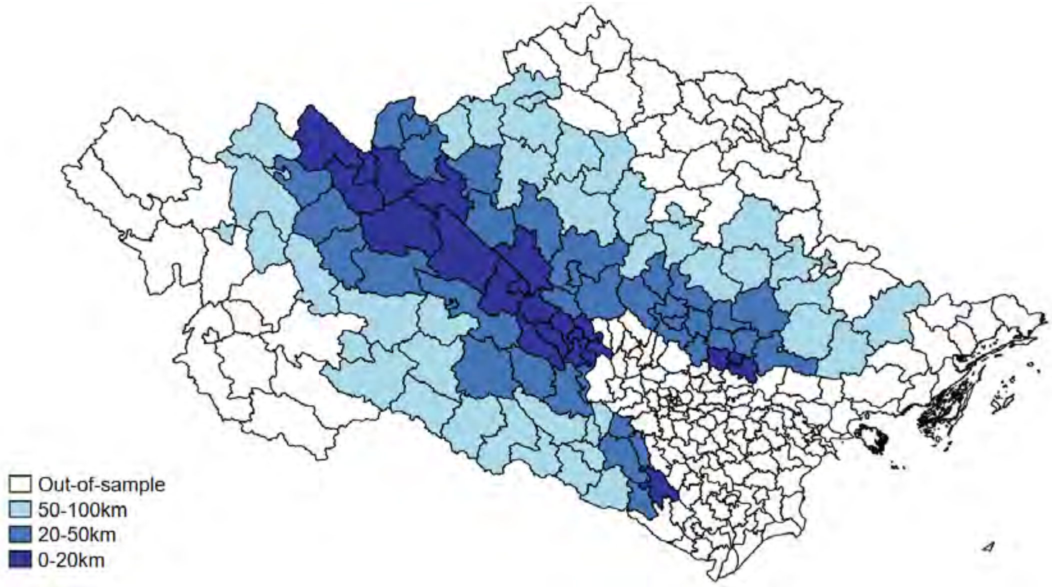
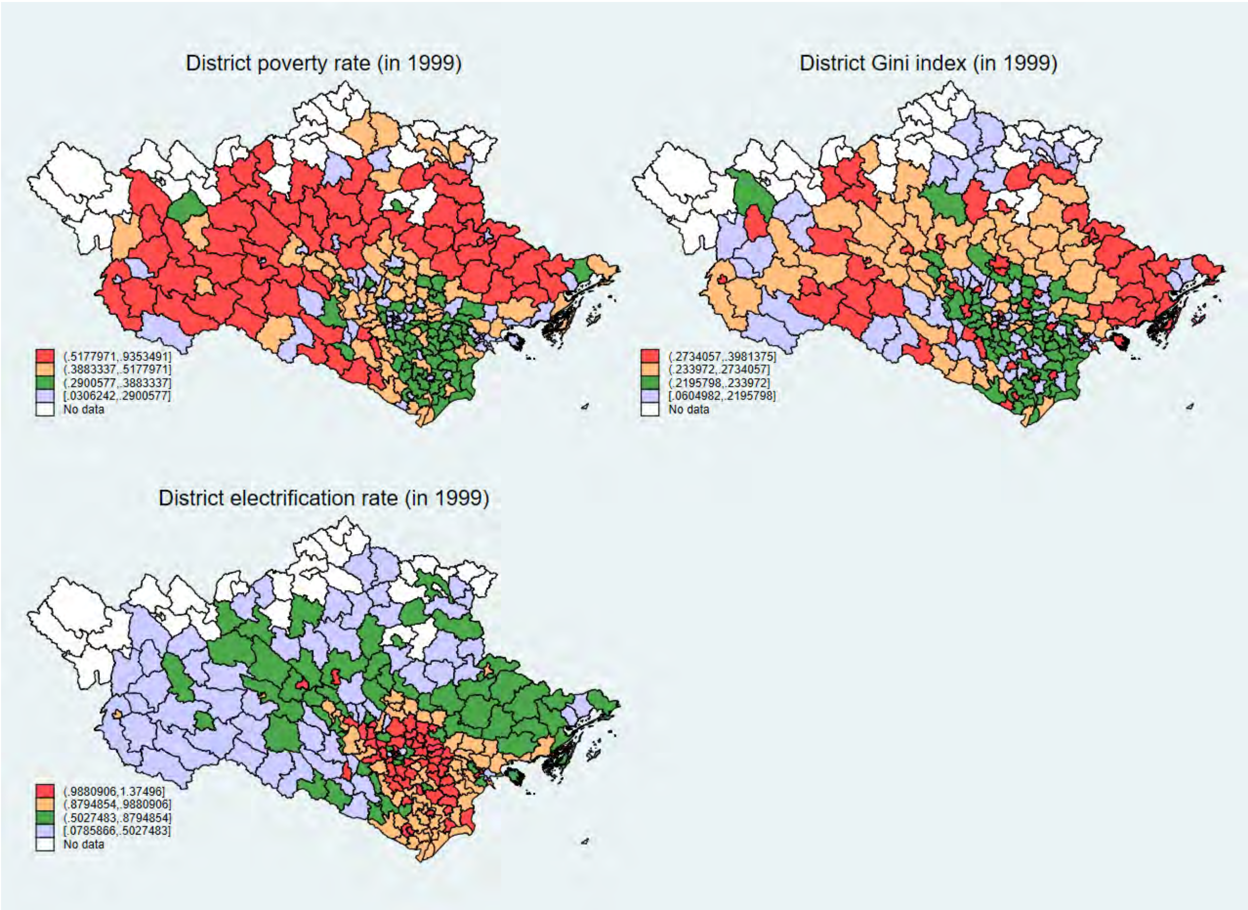


FIGURE 2: Baseline Socio-Economic Conditions: Poverty Rate, Inequality, and Access to Electricity in 1999

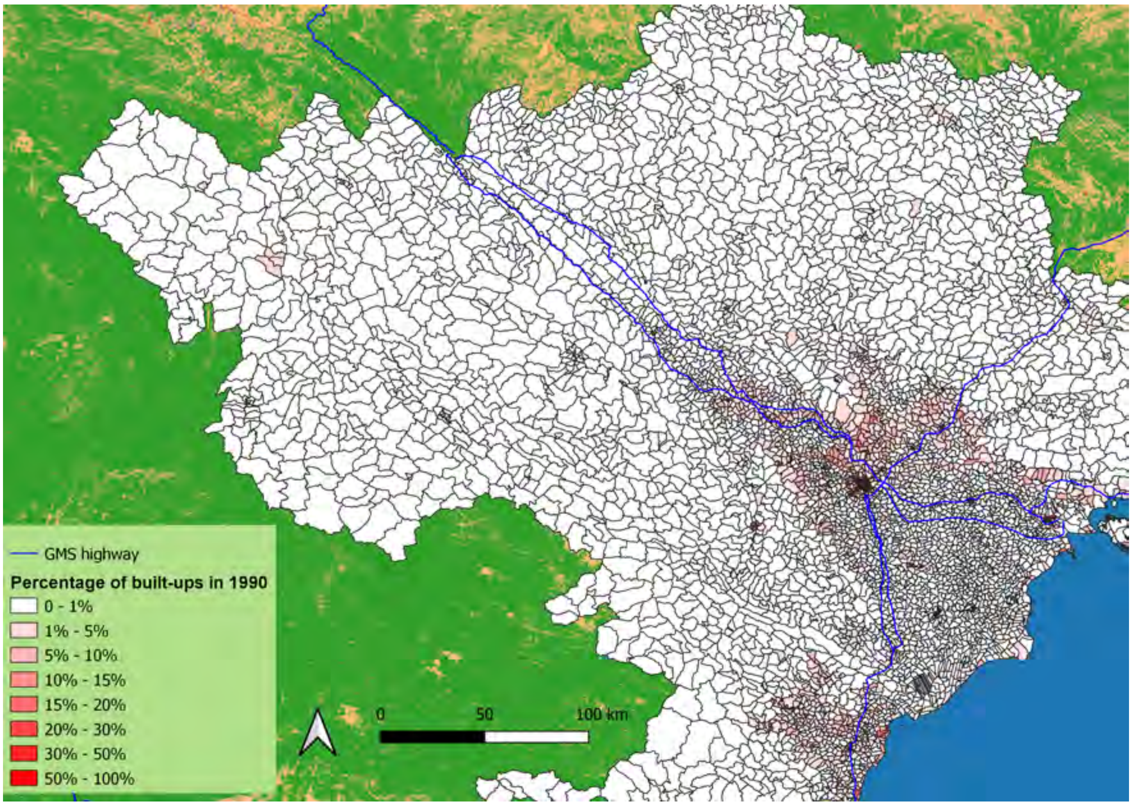


Source.—Miguel and Roland (2011).

FIGURE 3: GMS Highway and National Highways to Peripheries

FIGURE 4: Percentage of Built-ups within Communes: in 1990, 2000, 2010, and 2019

((a)) In 1990



((b)) In 2000

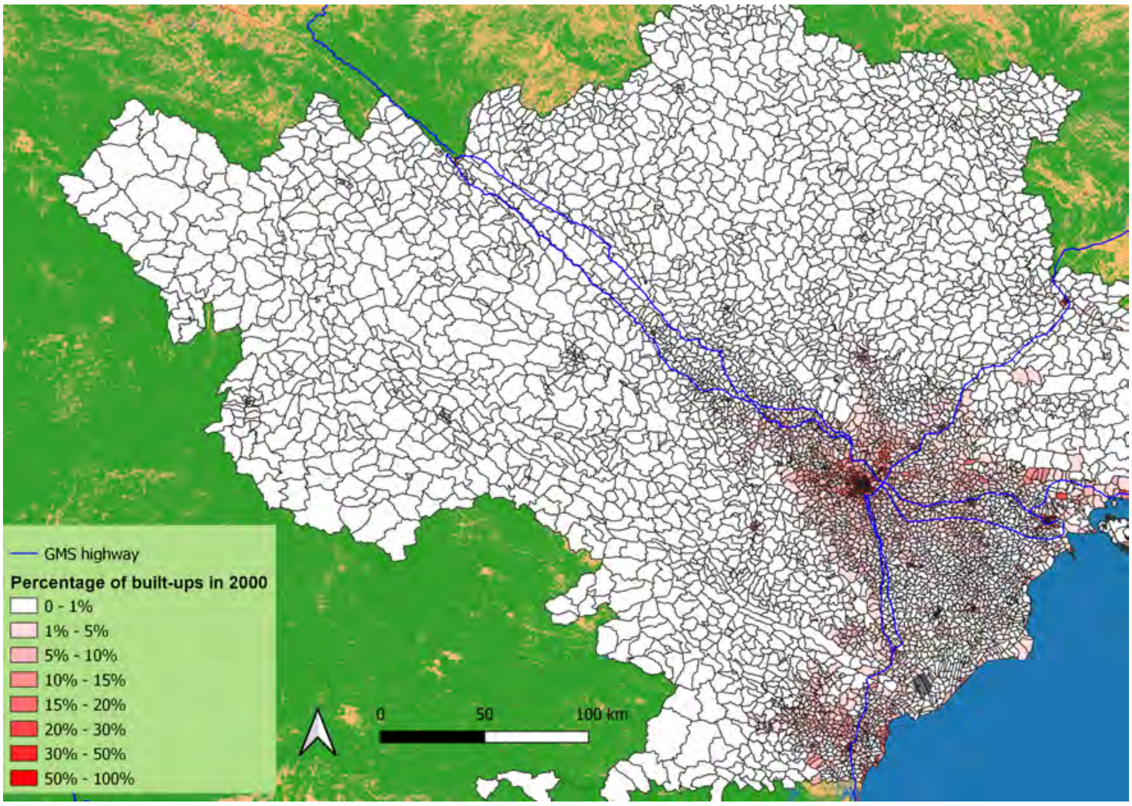
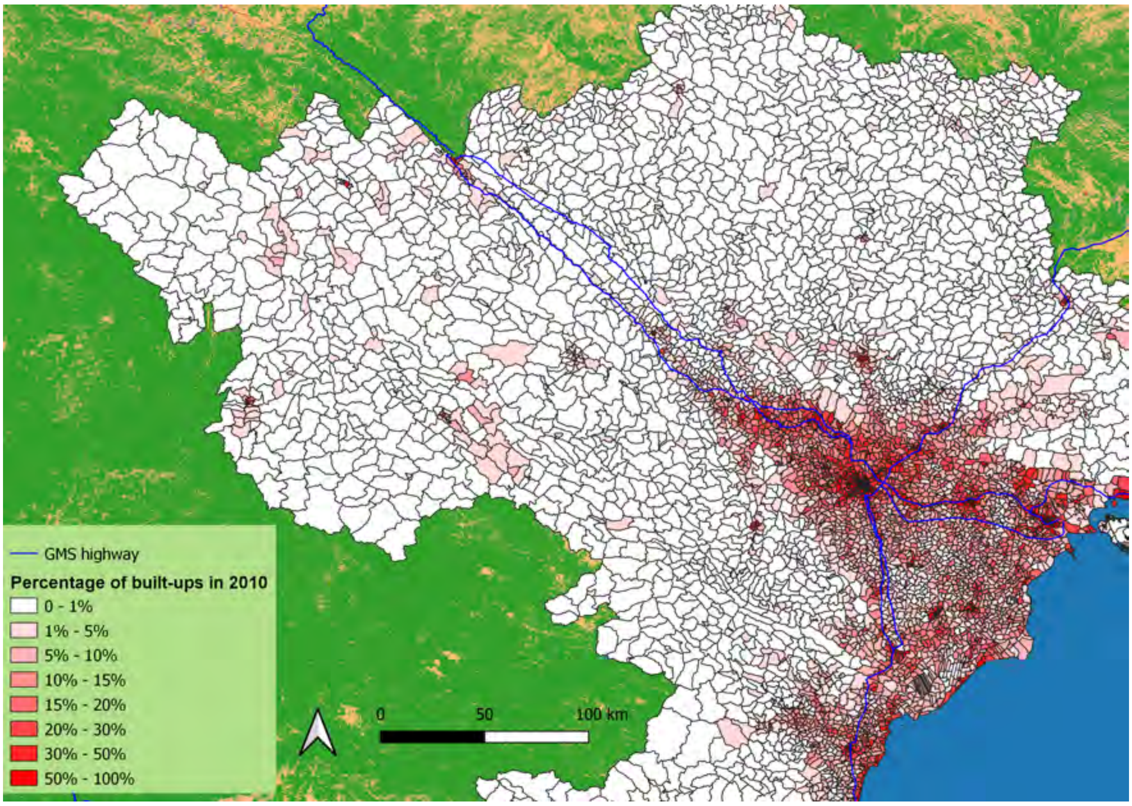


FIGURE 4: Percentage of Built-ups within Communes: in 1990, 2000, 2010, and 2019 (cont.)

((c)) In 2010



((d)) In 2019

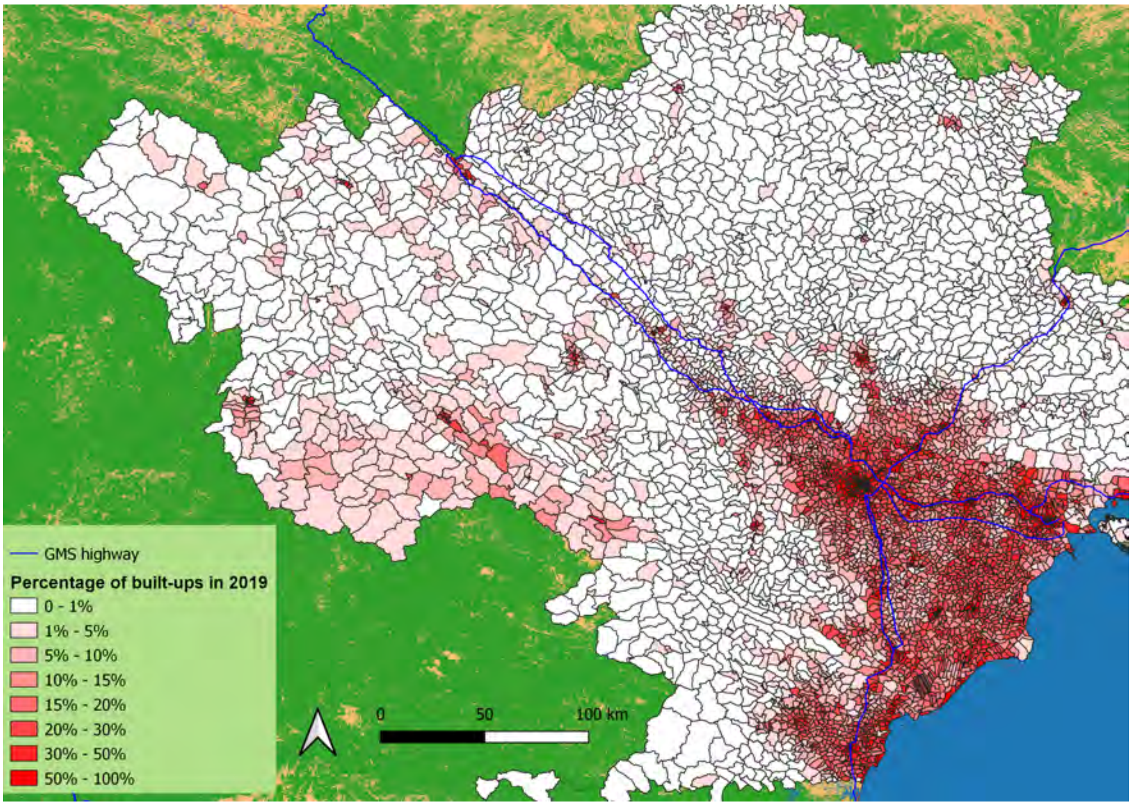


FIGURE 5: Evolution of Change in the VIIRS Nighttime Light, 2012-2019

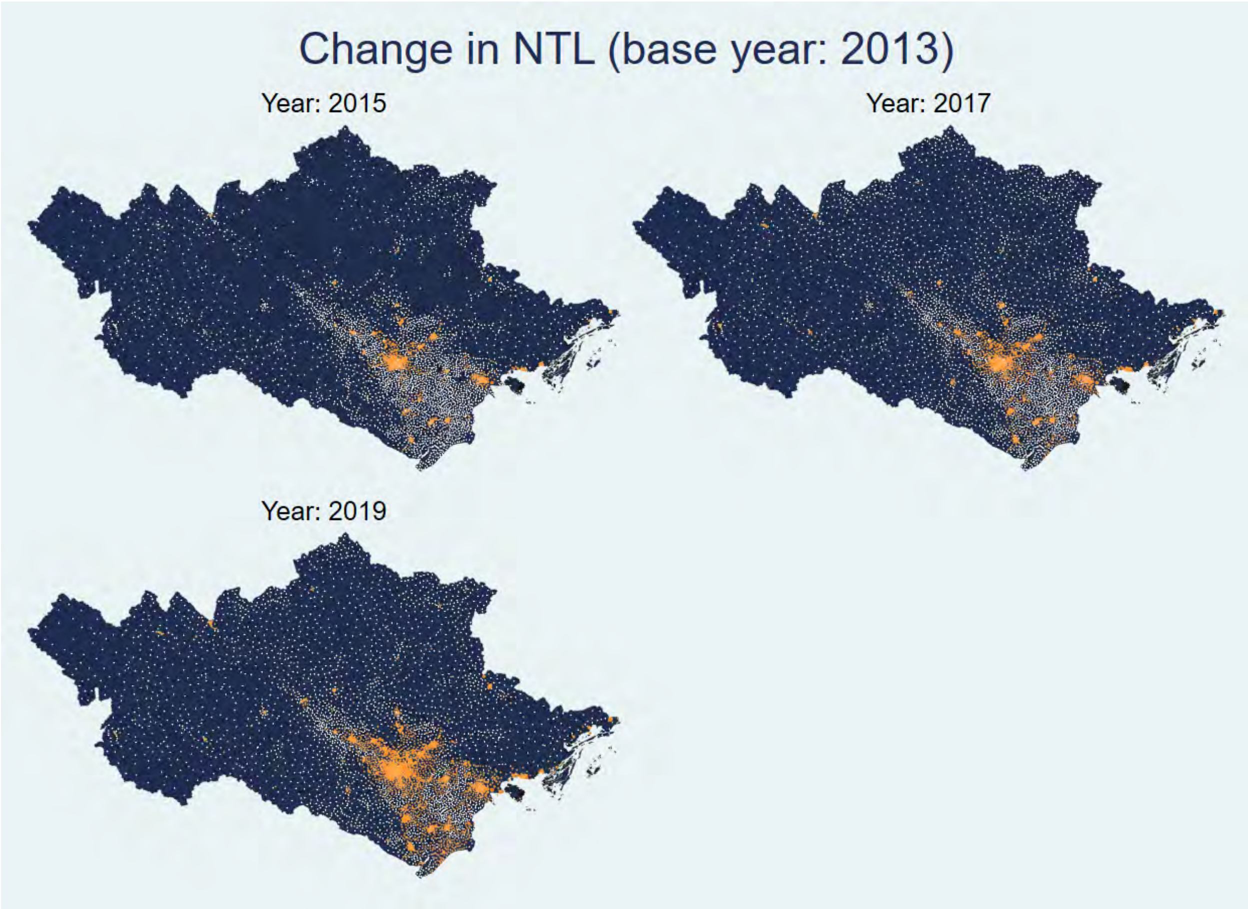
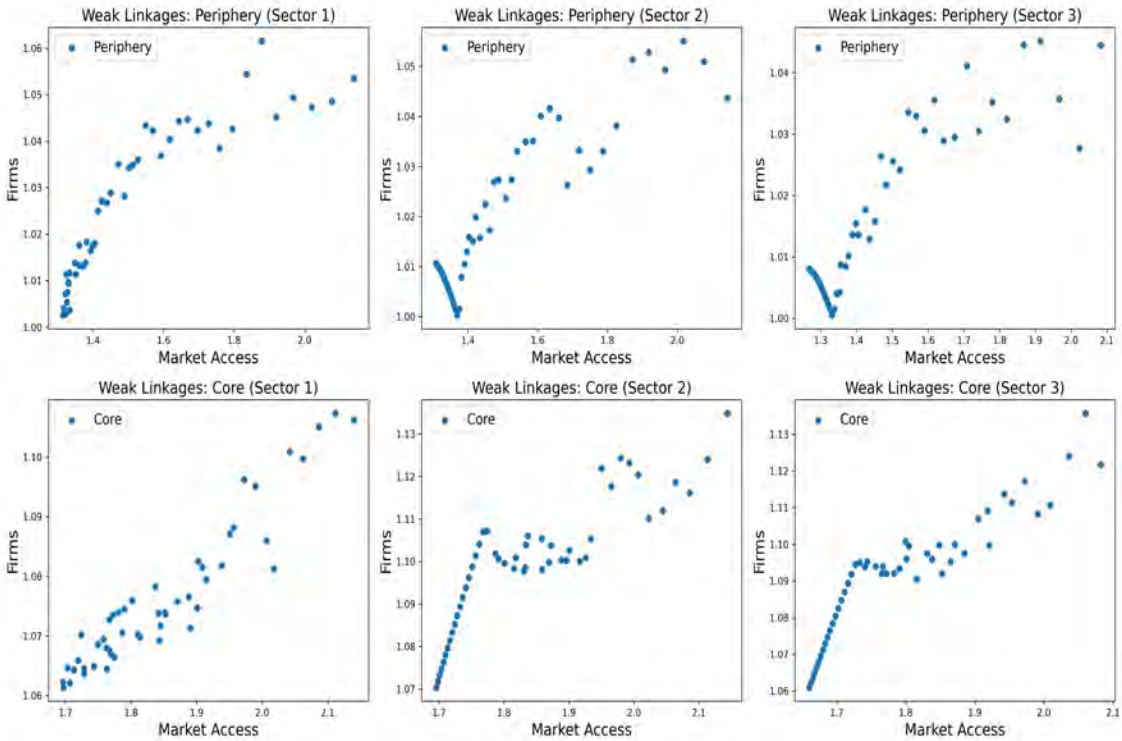


FIGURE 6: Simulation Results: Market Access vs. Number of Manufacturing Firms

((a)) Weak Input-Output Linkage Case: Market Access vs. Number of Manufacturing Firms



((b)) Strong Input-Output Linkage Case: Market Access vs. Number of Manufacturing Firms

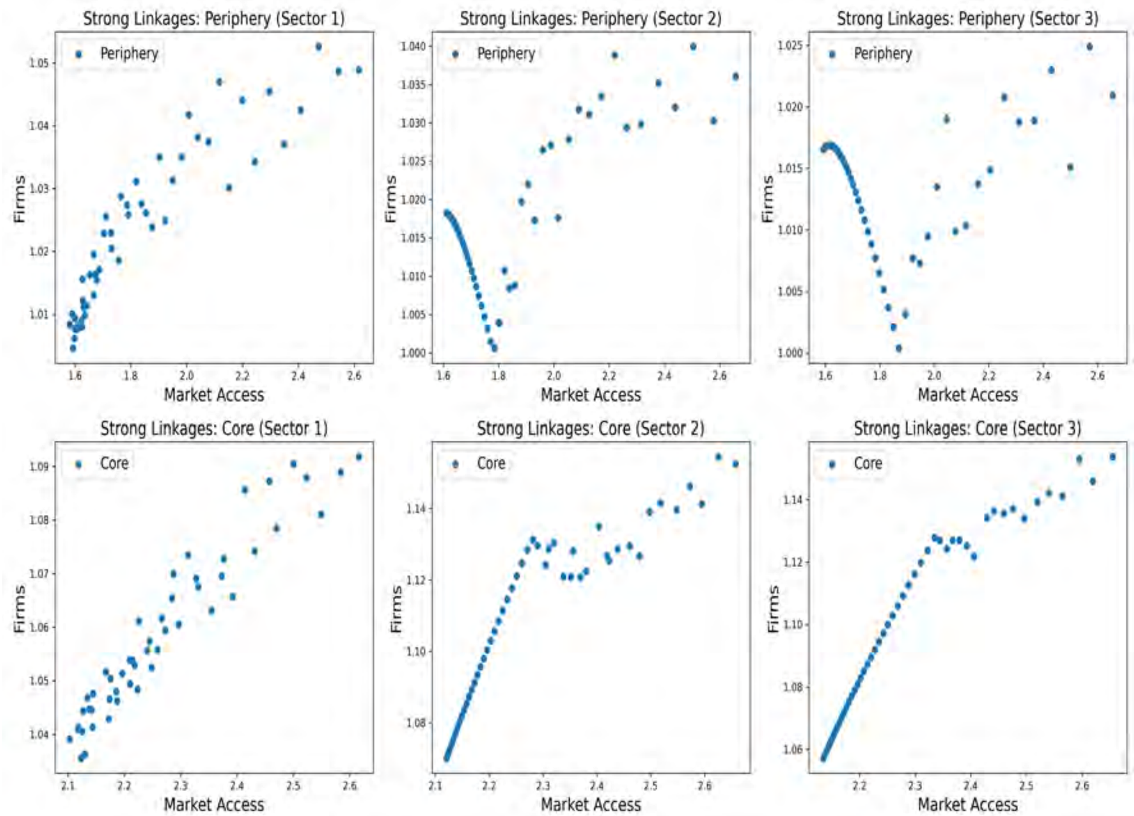
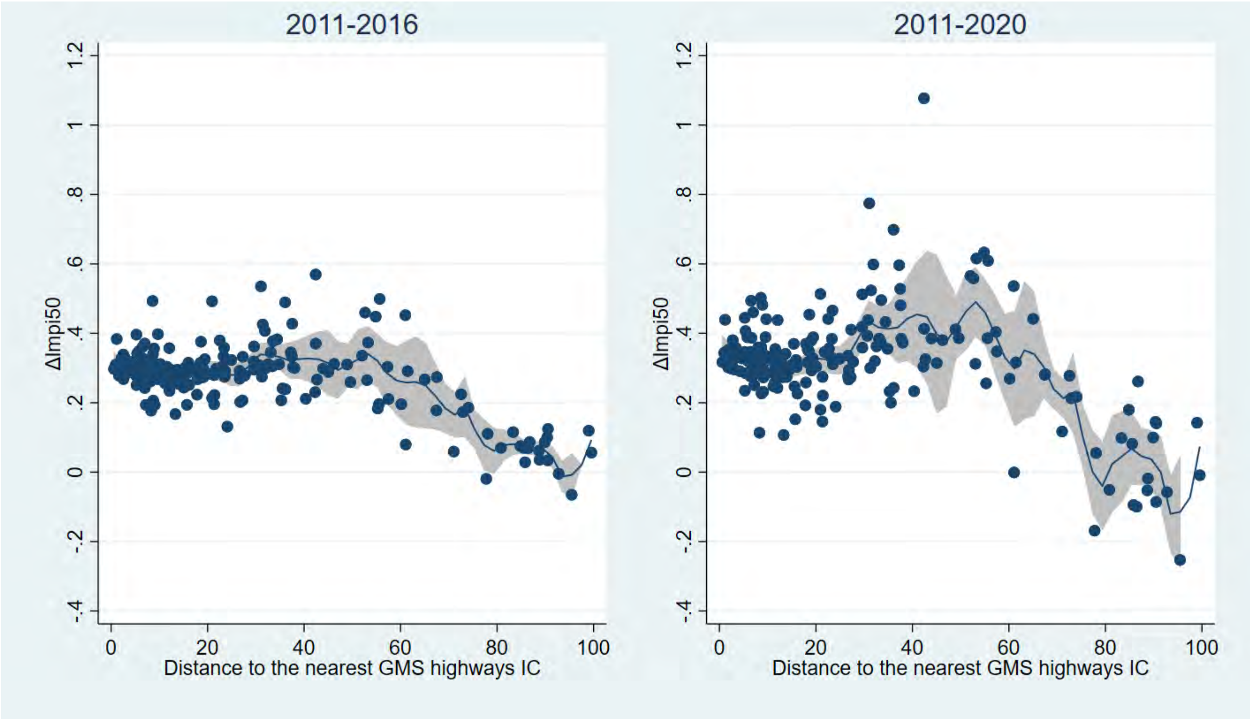


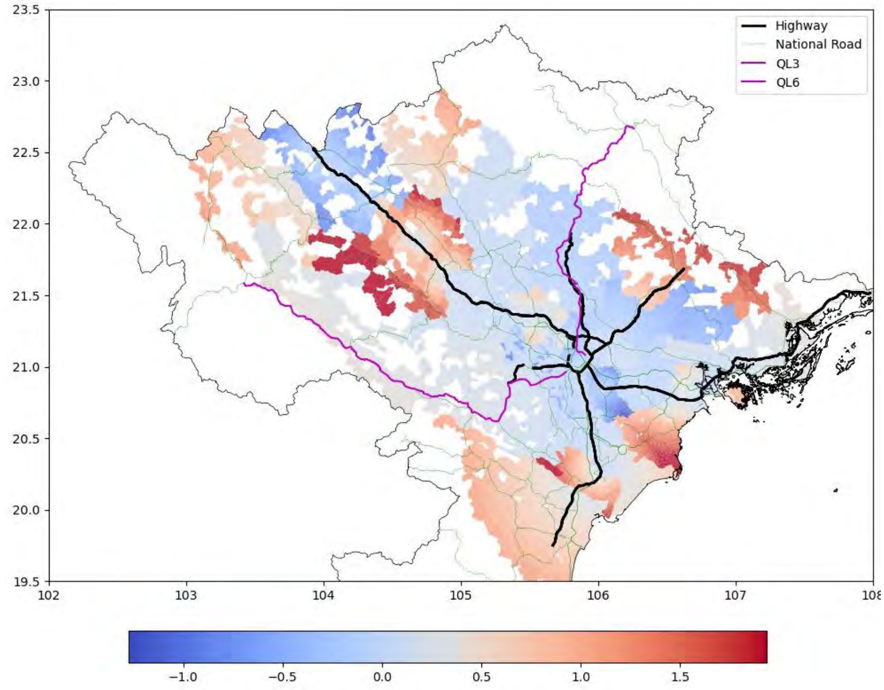
FIGURE 7: Change in the MP Index and the Distance from the GMS Corridor



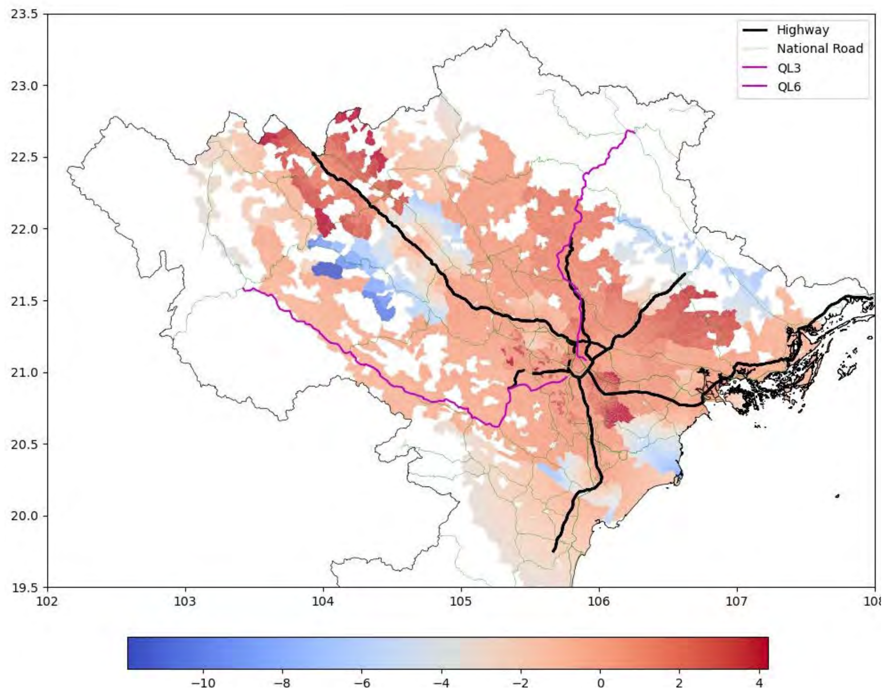
NOTE.—The sample covers all districts in northern Vietnam. x-axis is the Euclidian distance between each district's PC office to the nearest interchange point of the GMS highway. Provincial dummies and the distance to Hanoi are controlled as the parametric factor.

FIGURE 8: Spatial Distribution of Transport Costs and the Firm Market Access

((a)) Change in Interregional Transport Costs, from 2011 to 2020



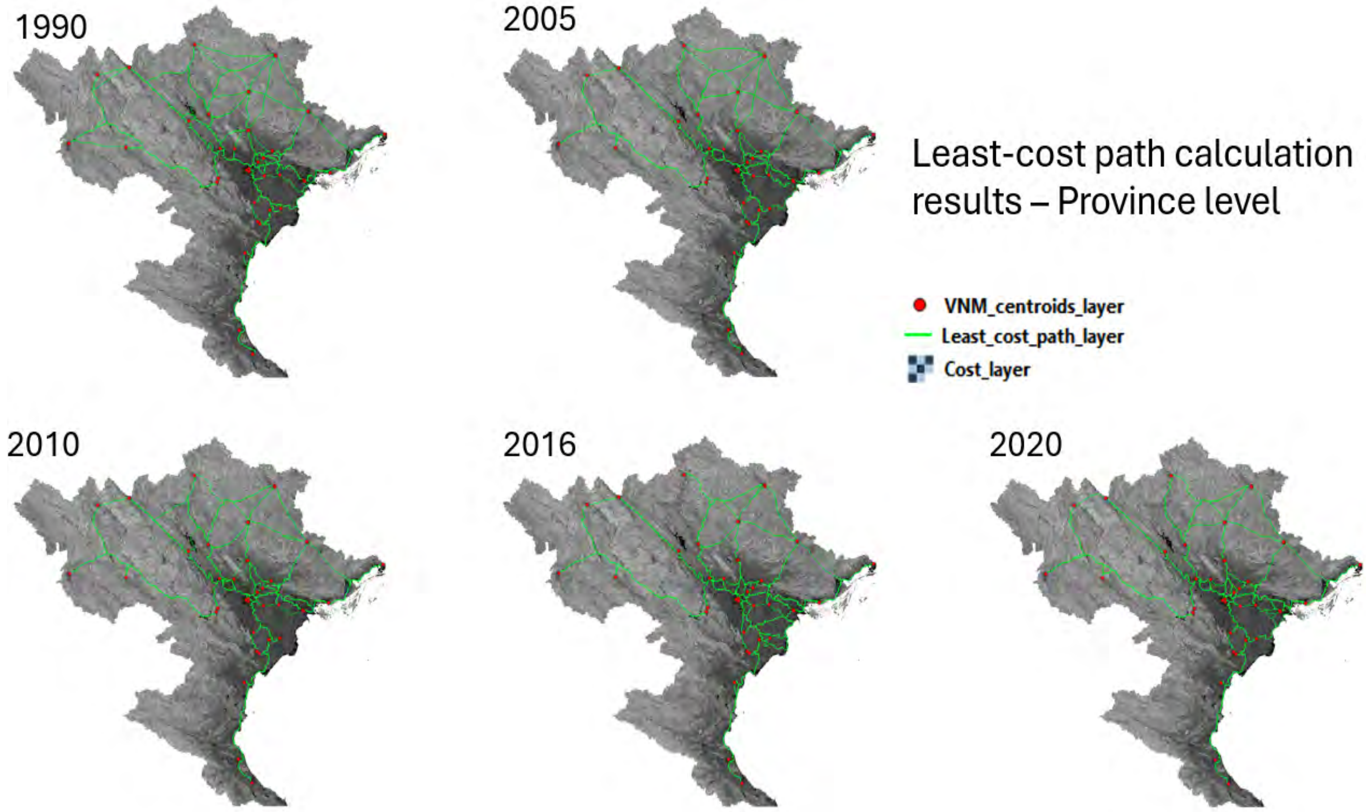
((b)) Improvement in Firm Market Access, from 2011 to 2020



SOURCE.—VES 2011, 2020. GSO.
NOTE.—The change in transport costs from 2011 to 2020 is characterized by the real price changes in transportation services. The real price is computed based on freight trade values and volumes of transport companies (deflated by CPI inflation) and is further adjusted for price variations across provinces and years. The computation of Firm Market Access (FMA) uses Equation (3) using the calculated real transport costs, real wages and employment data.

FIGURE 9: Maps of Hypothetical Least Cost Path

((a)) Least-cost path at province level



((b)) Least-cost path at district level

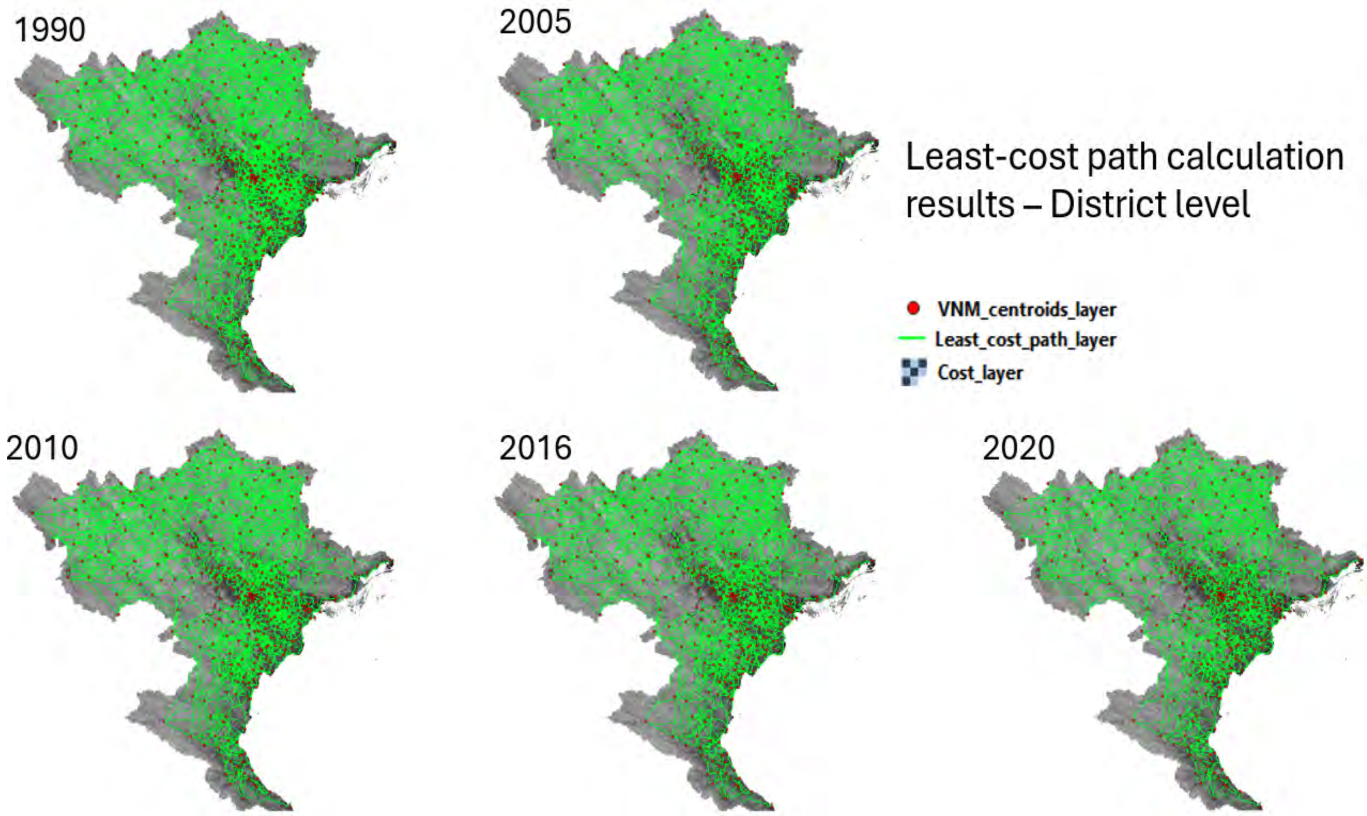


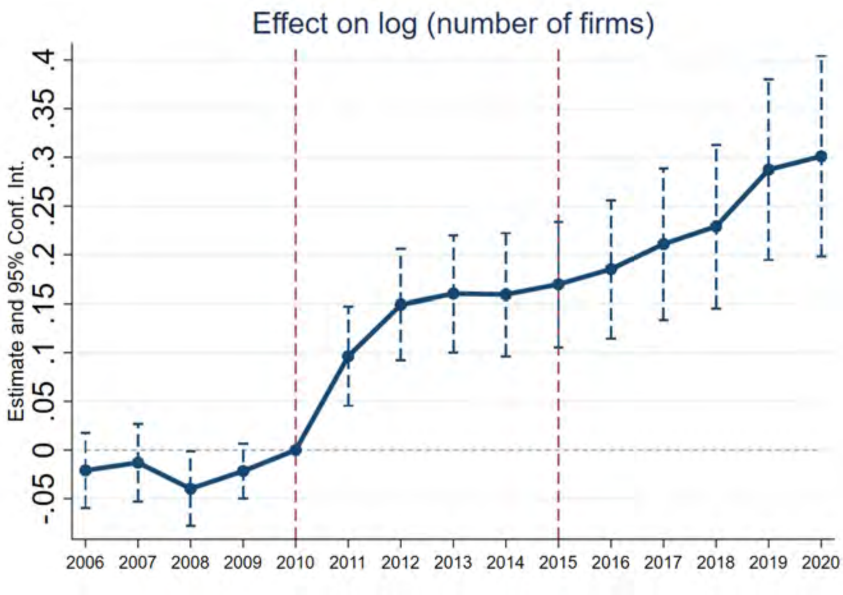
FIGURE 10: Maps of Minimum Spanning Tree Network

((a)) Minimum Spanning Tree Network (Actual)

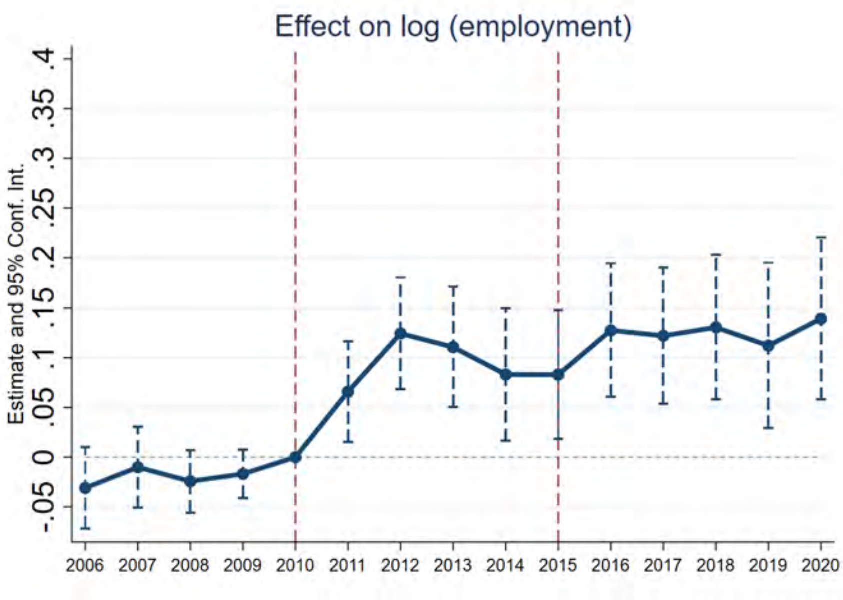
((b)) Minimum Spanning Tree Network (Euclidean)

FIGURE 11: Parallel Trends: Agglomeration and Employment

((a)) Log number of firms

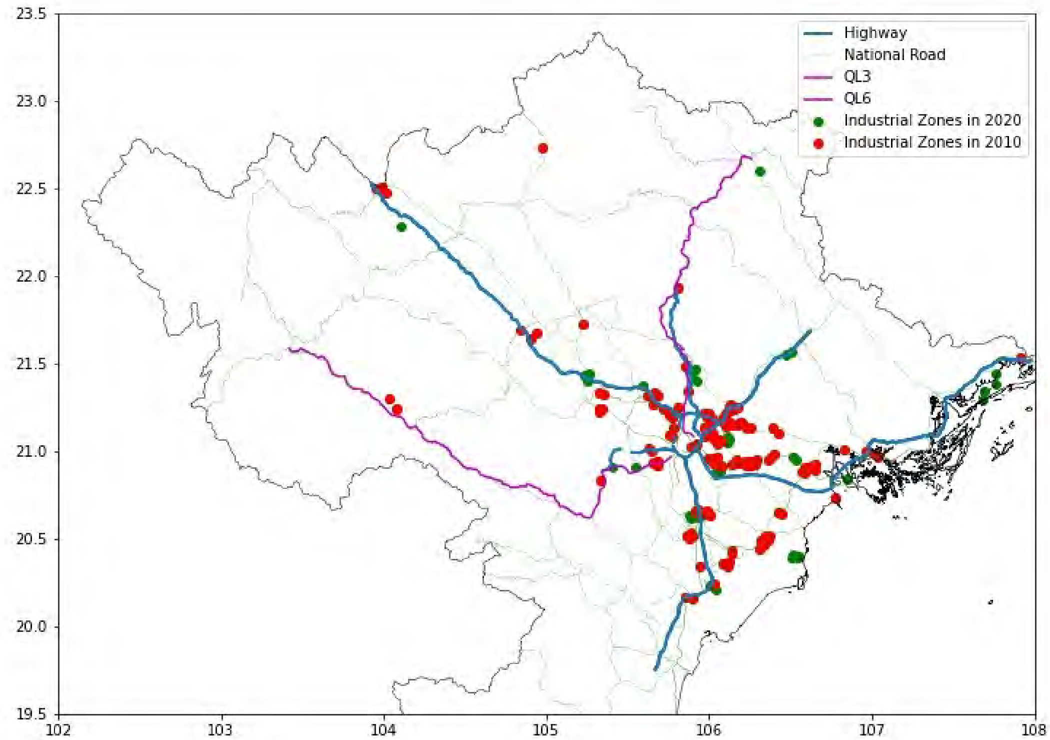


((b)) Log employment



NOTE.—The coefficient plots show the differential trend of each outcome variable β_t for communes near the GMS highway (within 50km) using two-way fixed effect model: $\frac{y_{ct}-y_{c2010}}{y_{c2010}} = \alpha_t + \beta_t \mathbb{1}[Distance_c \leq 50km] + \gamma_t X_{it} + \mu_c + \tau_t + \varepsilon_{ct}$. X_{it} controls for the effect of industrial zones. The dashed vertical lines in 2010 and 2015 indicate the timing when the construction of main GMS highway investment in the northern Vietnam started and completed, respectively. 95% confidence intervals, base on Huber-White robust standard errors, are reported as error bars.

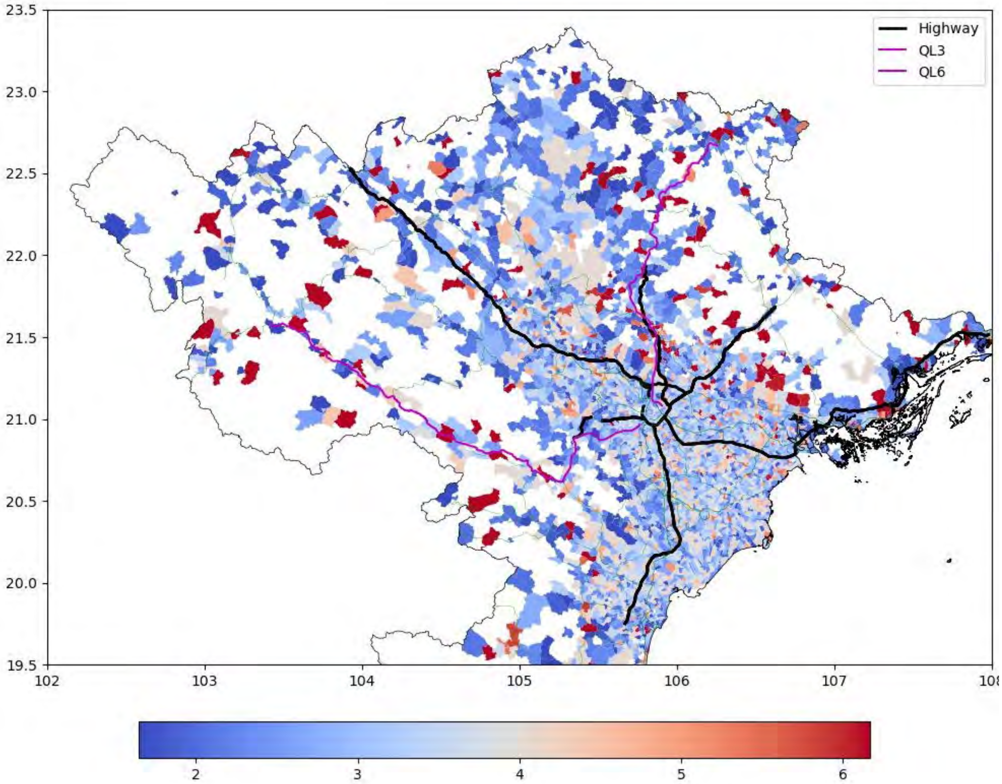
FIGURE 12: Location of Industrial Zones in Northern Vietnam



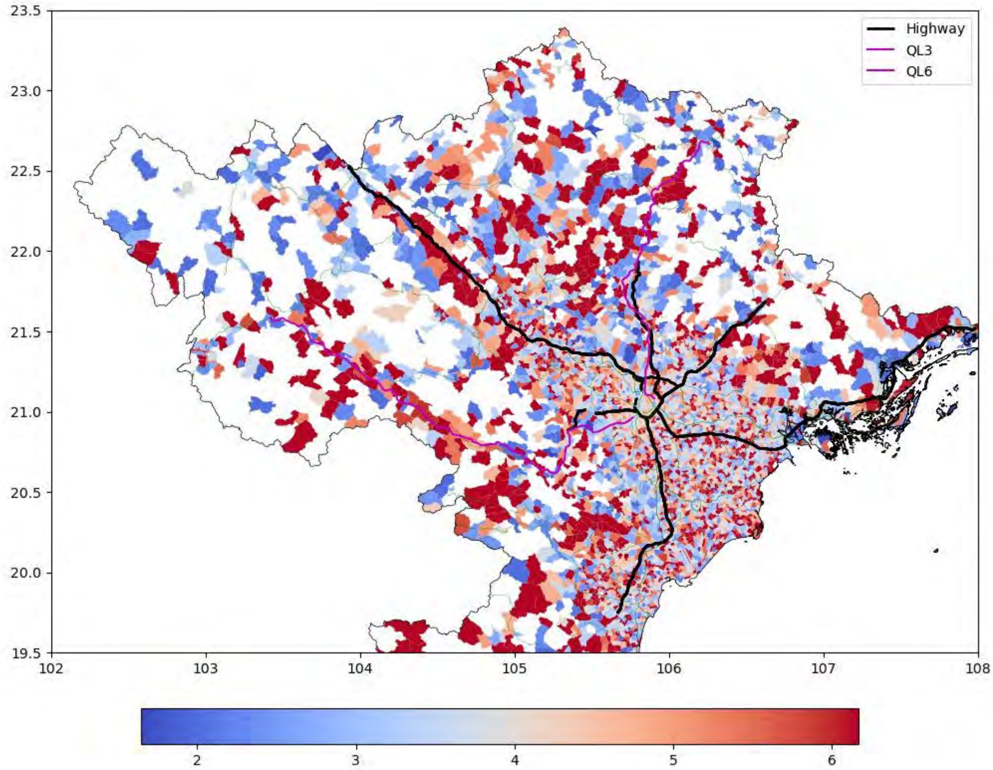
Source.—Vietnamese official documents and decrees.

FIGURE 13: Spatial Distribution of Input-output Linkages: Leontieff Inverse

((a)) In 2011

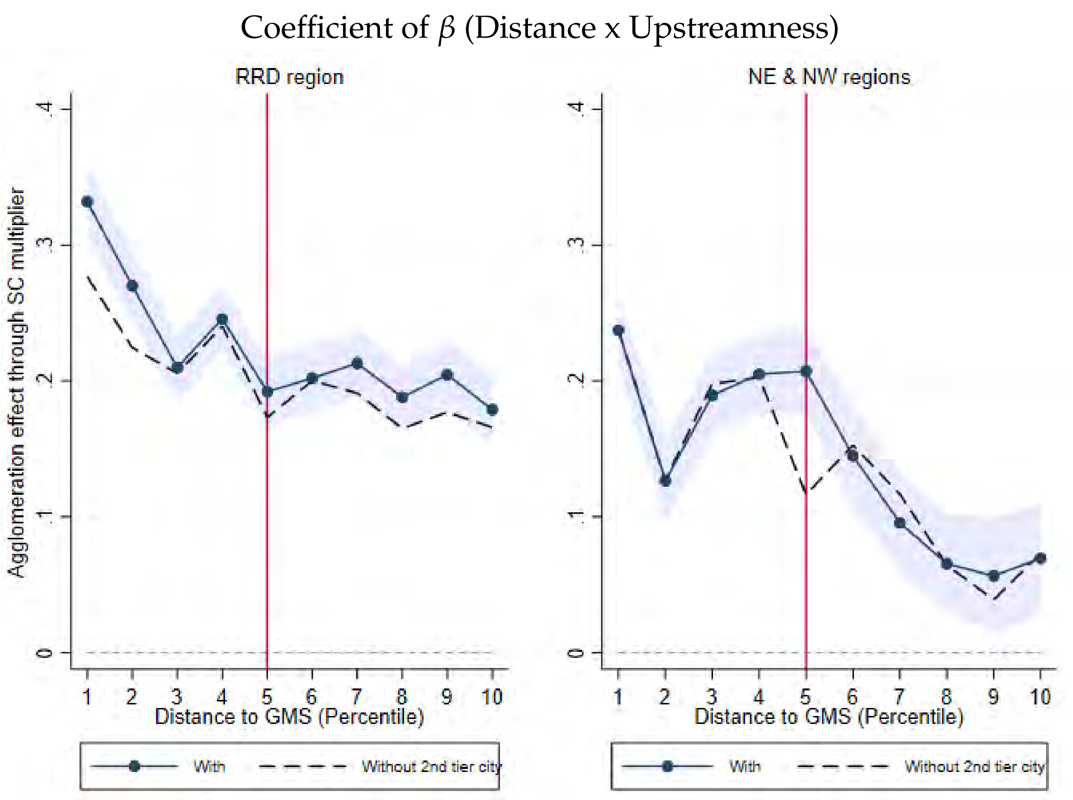


((b)) In 2017



Source.—VES2006-20, OECD Harmonized Input-output Table, 2021 ed.

FIGURE 14: Input-output Multiplier Effect on Manufacturing Firm Agglomeration



Source.—VES2006-20, OECD Harmonized Input-output Table, 2021 ed.
Note: shaded area indicates 95% confidence intervals.

Online Appendix

A Appendix 1: Micro Data

We collected data from four major surveys in Vietnam: the Vietnam Enterprise Survey (VES), Labor Force Survey (LFS), Population Census and Vietnam Household Living Standard Survey (VHLSS). All the data allow us to track geographical location of the respondent at the commune level, which is the finest administrative unit in Vietnam.[19]

The **VES** is a large-scale establishment data, which has been conducted annually by the General Statistics Office (GSO) since 2000. The rounds in 2001, 2006, 2011, 2016, and 2020 are census, and thus, cover the universe of registered firms. For other years, the VES covers all registered firms with employment size above a certain threshold, which varies across provinces where firms locate and generally increases over time. For smaller enterprises, a sample of 10–20% of firms were chosen for the survey. For the rest of the registered firms, the GSO creates a list and collects basic information such as legal type, total employment, and industry. The VES comprises of a main module for all firms and specific modules for firms in different industries. In this Component, we focus on the Main and following specific modules: Transportation, and Industrial Production (Manufacturing). We geocoded each firm's location at the commune and district levels for the spatial analysis. We compute most of our main variables, such as employment, sales per worker, and profit margins, for registered firms using the VES data.

The **LFS** has been conducted annually by the GSO since 2007. Each LFS round contains approximately 750–830 thousand observations, representing the whole population. However, geographical information at the district level cannot be identified in the rounds between 2008 and 2010. We employ the LFS to estimate the district-level population in 2007 and during the 2010-2020 period. The 2009 and 2019 Population Censuses, as well as the 2014 Mid-term Population Survey,, are utilized to construct the total population at the commune and district levels.

The **VHLSS** is a biennial survey that covers around 45,000 households each round. The data constructed from the VHLSS contain information on population, housing value, and human capital stock, as captured by the average years of schooling. The VHLSS cov-

[19] In Vietnam, administrative units are divided into three tiers: (i) municipality and province; (ii) urban district, provincial city, town, and rural district; and (iii) ward, township, and commune. There is 4-level administrative unit called "village", but it is rather a historical unit, not a current governmental working unit. Throughout this Project Component, we use the The Global Administrative Areas (GADM) version 3.6 for our spatial analysis.

ers both formal and informal workers, but the location information is based on the living place rather than the working place. The VHLSS data also collect additional information on household expenditures including expenditures on housing and education for the smaller set of respondents, covering about 9,000 households. The expenditure data also cover monthly and annual household consumption on food, drink, non-food, and durable items.

B Appendix 2. The Model

What were the drivers of polycentric city development in the northern Vietnam? We consider a multi-region, multi-sector spatial general equilibrium model with input-output linkages and iceberg trade costs to guide our empirical analysis. The new economic geography model (Redding and Venables (2004); Puga and Venables (1996); Venables (1996); Krugman and Venables (1995)) provides a relevant framework to derive theoretical predictions on a firm's location decision in response to a highway upgrade program.

Let us consider a general model with multiple regions $j = (1, ..., J)$, and each region can produce both agricultural and manufacturing outputs. When we calibrate the model in Section B.7, we consider a Core-Periphery model. Manufacturing is composed of s different sectors, each of which is assumed to be monopolistically competitive. The model considers endogenous labor allocation between agriculture and manufacturing.

B.1 Agricultural Production

Agriculture is perfectly competitive. The agricultural production uses Cobb-Douglas technology with labor L_A and land K:

$$Y_{A,i} = A_{A,i} L_{A,i}^{\theta} K_i^{1-\theta}, \tag{8}$$

where A_A is agricultural productivity (exogenously endowed in each region). Agricultural wages are determined by marginal productivity:

$$w_{A,i} = \theta A_{A,i} \left(\frac{L_{A,i}}{K_i} \right)^{\theta - 1}$$

B.2 Manufacturing Production and Input-Output Linkages

Each sector s in region i produces goods using labor $L_{M,s,i}$ and intermediate inputs $X_{r,s,i}$ sourced from other sector r. The production function in sector s follows a Cobb-Douglas specification:

$$Y_{M,s,i} = A_{M,s,i} \cdot L_{M,s,i}^{\eta_s} \cdot \prod_{r=1}^{S} X_{r,s,i}^{\mu_{rs}}, \tag{9}$$

where $A_{M,s,i}$ is the productivity in sector s in region i, and η_s is the labor share in sector s.[20]

[20]We introduce heterogeneity in manufacturing base productivity across sectors and regions with pro-

Manufacturing sectors use intermediate inputs, modeled by an input-output coefficient matrix μ_{rs}, which represents the share of inputs from sector r used in sector s's production.

In each region i, each sector s's price index is defined over products supplied from all regions ($i \in J$) taking the following CES form:

$$P_{s,i} = \left(\sum_{j}^{J} n_{s,j} (p_{s,j} \tau_{ji})^{1-\sigma} \right)^{\frac{1}{1-\sigma}},$$

where τ_{ji} is the trade cost between regions j and i, and $n_{s,j}$ is the number of firms producing good s in region j. σ is the elasticity of substitution across varieties within sector r. $p_{s,j}$ is the producer price of good s in region j, which is set with a markup over marginal cost:

$$p_{s,j} = \frac{\sigma}{\sigma - 1} \cdot \underbrace{w_j^{1-\eta_s - \sum_{r=1}^{S} \mu_{rs}} \cdot \prod_{r=1}^{S} P_{r,j}^{\mu_{rs}}}_{\text{Marginal Cost}}$$

Firm in sector s demands a fixed share μ_{rs} of its revenue to purchase inputs from sector r for intermediate input $X_{r,s,i}$:

$$X_{r,s,i} = \mu_{rs} \cdot R_{s,i}(\gamma),$$

where $R_{s,i}(\gamma)$ is the effective revenues of a firm in sector s located in region i derive from selling to consumers in destination region r. The parameter, μ_{rs}, captures backward linkages. The effective revenue, $R_{s,i}(\gamma)$, is given by:

$$R_{s,i}(\gamma) = \sum_{r=1}^{S} E_{s,r} \cdot \left(\frac{n_{s,i}}{n_{r,i}} \right)^{\gamma} \cdot \left(\frac{n_{s,i}}{n_{s,j}} \right)^{\mu_{rs}} \cdot T_{ji}^{r}. \tag{10}$$

The term $(\frac{n_{s,i}}{n_{r,i}})^{\gamma}$ adjusts for intersectoral competition effect based on the relative number of firms between sectors in region i. Higher the competition (more firms in sector r), the lower the effective revenues with the sensitivity $\gamma \in [0, 1]$. Higher γ increases the negative impact of facing many competing varieties in destination r. This formulation captures the idea that market access is diluted when competitors are abundant, and helps regulate the incentives for firm entry across space. Intuitively, the middle part, $(\frac{n_{s,i}}{n_{s,j}})^{\mu_{rs}}$, represents the size of intermediate goods demand through input-output linkages.

The bilateral trade share T_{ji}^{r} is defined as the fraction of sector r's expenditure in region

ductivity enhancement through agglomeration (see Equation (11)). A footloose entrepreneur model (Forslid and Ottaviano (2003); Okubo (2010); Lu and Tsai (2025)) provides an alternative framework to consider the locational choice of firms with heterogeneous productivity.

i that is spent on goods imported from region j. Under CES price index and iceberg trade costs $\tau_{ji}^r (\geq 1)$ for shipping one unit of goods from region j to region i in sector r:

$$T_{ji}^r = \frac{\left(p_{r,j} \cdot \tau_{ji}^r\right)^{1-\sigma} \cdot n_{r,j}}{\sum_{k=1}^{J} \left(p_{r,k} \cdot \tau_{ki}^r\right)^{1-\sigma} \cdot n_{r,k}}$$

Agglomeration of Firms The model incorporates the agglomeration economy in the manufacturing sectors. Specifically, the productivity in sector s and region i, denoted $A_{M,s,i}$, is increasing in the number of firms $n_{s,i}$ operating locally:

$$A_{M,s,i} = A_{M,\text{base},s,i} \cdot n_{s,i}^{\alpha}, \tag{11}$$

where $A_{M,\text{base},s,i}$ is the baseline productivity level in sector s, region i, $\alpha > 0$ is the elasticity of productivity with respect to the number of firms, capturing the strength of the agglomeration externality. This functional form reflects increasing returns to scale arising from agglomeration forces such as shared supplier networks, knowledge spillovers, and access to specialized labor.

B.3 Consumer Preferences

Consumers in both regions have Cobb-Douglas preferences, spending a fraction θ of income on agriculture and $(1 - \theta)$ on a composite of differentiated manufacturing goods. The demand for sector s goods follows a CES utility function:

$$U_i = C_{A,i}^{\theta} \cdot C_{M,i}^{1-\theta}, \quad \theta \in (0, 1),$$

where $C_{A,i}$ is consumption of agricultural goods in region i, $C_{M,i}$ is composite consumption of manufacturing varieties in region i, and θ: expenditure share on agriculture.

The manufacturing composite is itself a CES aggregate over sectors and varieties:

$$C_{M,i} = \prod_{s=1}^{S} \left(\int_{\omega \in \Omega_s} q_{s,\omega,i}^{\frac{\sigma-1}{\sigma}} d\omega \right)^{\frac{\beta_s}{\sigma-1}}$$

where $q_{s,\omega,i}$ is consumption of variety $\omega \in \Omega_s$, Ω_s is set of varieties in sector s, β_s is the share of sector s in the manufacturing composite, with $\sum_s \beta_s = 1$.

Consumers allocate total income $w_i L_i$ across sectors according to Cobb-Douglas shares.

Manufacturing expenditure is further allocated across sectors and varieties. If varieties in sector s, region i, are symmetric and trade incurs iceberg costs $\tau_{ji} \geq 1$, the optimal demand for a single variety ω produced in sector s and region i is:

$$q_{s,\omega,i} = \left(\frac{p_{s,\omega,j} \cdot \tau_{ji}}{P_{s,i}} \right)^{-\sigma} \cdot \frac{E_{s,i}}{P_{s,i}} \tag{12}$$

This equation implies that an increase in the delivered price of the variety relative to the price index of competing varieties leads to higher demand. The higher the real expenditure on sector s goods in region i, the more is spent on each variety on average.

Aggregate Regional Demand

We can also characterize total expenditure on sector s in region i includes both final consumption and intermediate input demand from other sectors:

$$E_{s,i} = \beta_s(1-\theta)w_i L_i + \sum_{r=1}^{S} \mu_{rs} \cdot X_{r,i} \tag{13}$$

where β_s is manufacturing sector share of spending on sector s, μ_{rs} is input-output coefficient from sector s to sector r, and $X_{r,i}$ is total sales of sector r in region i.

B.4 Firm Entry and Market Structure

The model incorporates monopolistic competition, with free entry of firms into manufacturing sectors. Each firm faces a fixed entry cost, F (hiring initial labors). The number of firms in sector s in region i adjusts to satisfy the zero-profit condition:

$$\pi_{s,i} = \frac{(p_{s,i}\tau_{ik})^{1-\sigma}}{P_{s,i}^{1-\sigma}} \cdot n_{s,i} \cdot E_{s,i} - F \cdot w_i, \tag{14}$$

where $\pi_{s,i}$ is the firm's profit, $E_{s,i}$ is the regional demand for sector s, and w_i is the wage in region i. The number of firms, $n_{s,i}$, is determined by the equilibrium between firm entry and exit.

B.5 Labor and Wages Determination in Manufacturing

The allocation of labor between agriculture and manufacturing is endogenous in the model. Total labor in each region is divided between the manufacturing sector L_M and agriculture L_A:

$$L_M = \frac{L_i}{1 + \frac{(1-\theta)K}{\theta}}, \qquad L_A = L_i - L_M.$$

The wage rate in each region is determined by the marginal productivity of labor in both sectors. In the manufacturing sectors, wages are set through the free-entry condition and depend on the number of firms and sectoral productivity. The equilibrium wage is solved iteratively, accounting for both competition and agglomeration effects.

B.6 Market Access

In this model, *market access* measures the ability of a firm in region i, sector s, to sell its products across all destination markets j. It aggregates the potential demand a firm can access, accounting for trade costs and prices in the destination region.

We define the market access of a firm in region i, sector s, as:

$$MA_{s,i} = \sum_j \tau_{ij}^{1-\sigma} \cdot \frac{n_{s,j}}{P_{s,j}^{1-\sigma}} \tag{15}$$

This expression is derived from summing over the optimal demand for a single variety (see previous section) across all destination regions j. It captures the total size of accessible markets for region i.

The Market Access Index reflects how *attractive* region i is as a production location from a sales perspective.

- **Increasing in destination number of firms** $n_{s,j}$: larger, richer markets offer more sales opportunities.
- **Decreasing in trade costs** τ_{ij}: higher trade frictions reduce the competitiveness of firms in distant markets.

Thus, firms located in regions with better access to large and close markets will experience stronger demand and higher revenues. This concept plays a central role in explaining spatial patterns of industrial concentration and wage variation in the model.

B.7 Equilibrium and Core-Periphery Model Calibration

The equilibrium in this model is determined by the interaction of wages, the number of firms, and market access. The model uses iterative procedures to solve for the equilibrium number of firms, wages, and market prices, based on the trade costs τ, the number of firms, and the input-output linkages μ_{rs} for region 1 (Periphery) and region 2 (Core region) separately.

Simulations in the model explore the effects of varying market access (trade costs) and input-output linkages on the number of firms. Table below summarizes key parameter values in the calibration. Table below compares the weak vs. strong input-output linkage matrices.

Calibrated Parameters

Parameter	**Description**	**Value**
σ	Elasticity of substitution	5
θ	Agri. share in consumption	0.5
F	Fixed cost of firm entry	1.0
α	Agglomeration elasticity	0.5
γ	Competition sensitivity	0.1
S	Number of sectors	3
J	Number of regions	2
η_s	Labor share in sector s	[0.2, 0.3, 0.4]
$A_{M,\text{base}}$	Manufacturing productivity, by region	[1.0, 1.2]
A_A	Agricultural productivity, by region	[1.2, 1.0]
K_A	Land (Agri.), by region	[1.2, 1.0]
L_j	Labor endowment, by region	[1.0, 1.0]

Input-Output Linkage Coefficients (μ_{rs})

Source r	**Weak**			**Strong**		
	$s=1$	$s=2$	$s=3$	$s=1$	$s=2$	$s=3$
$r=1$ (Primary)	0.05	0.03	0.02	0.10	0.05	0.05
$r=2$ (Secondary)	0.10	0.10	0.05	0.20	0.25	0.15
$r=3$ (Service)	0.02	0.03	0.05	0.10	0.10	0.20

As the above table shows, we consider three sectors: the "Primary" sector with low intersectoral linkages, the "Secondary" sector with strong upstream and downstream linkages, and the "Service" sector with strong downstream linkages.

B.8 Two Competing Forces in the Equilibrium

Competition Effect

The competition effect in the model arises from the relationship between the number of firms and market competition. As the number of firms in a region increases, individual firms face higher competition, lowering their market share.

An increase in the number of firms, $n_{s,i}$, lowers the price index $P_{s,i}$, which in turn decreases individual firm revenues. The number of firms is determined by the zero-profit condition, where revenue equals the cost of entry:

$$n_{s,i} = \frac{\sum_j \pi_{s,i} \cdot E_{s,i}}{F \cdot w_i}.$$

Thus, competition limits firm entry in a region.

Agglomeration Effect

Agglomeration effect arising from an increase in $n_{s,i}$ raises productivity $A_{M,s,i}$, potentially lowering costs and encouraging further firm entry. This positive feedback loop can lead to the emergence of industrial concentration in certain regions (typically the core), especially when trade costs are low and input-output linkages are strong. These effects operate via:

1. **Market Access:** More firms in a region increase demand for goods and services, which reduces the cost of production and increases consumer welfare.
2. **Input Linkages:** Agglomerated regions benefit from stronger input-output linkages. Firms in the region benefit from lower costs of intermediate goods, leading to higher overall productivity.

In the model, agglomeration benefits are captured by the input-output matrix μ_{rs}, which represents the interdependence between sectors. A higher concentration of firms in a region leads to stronger demand for intermediate goods from other sectors, raising overall regional income and boosting firm profitability. The agglomeration effect leads to self-reinforcing growth: as more firms enter a region, the regional price index $P_{s,j}$ falls, which lowers costs and further attracts firms.

B.9 Simulation Results

In the analysis of the relationship between market access and the number of firms, we observe different patterns for the **periphery** and **core** regions. Simulation results for different input-output linkage scenarios (weak (Figure A1) vs. strong (Figure A2) highlight significantly positive but nonlinear relationship between market access improvement (driven by lower trade costs τ) and the number of firms in each region. These results underscore the importance of trade costs and regional linkages in shaping the spatial distribution of economic activities.

To rationalize this result, we focus on the interplay of two key mechanisms in the model: the **competition effect** and the **agglomeration effect** in driving the nonlinearity in the effect of market access.

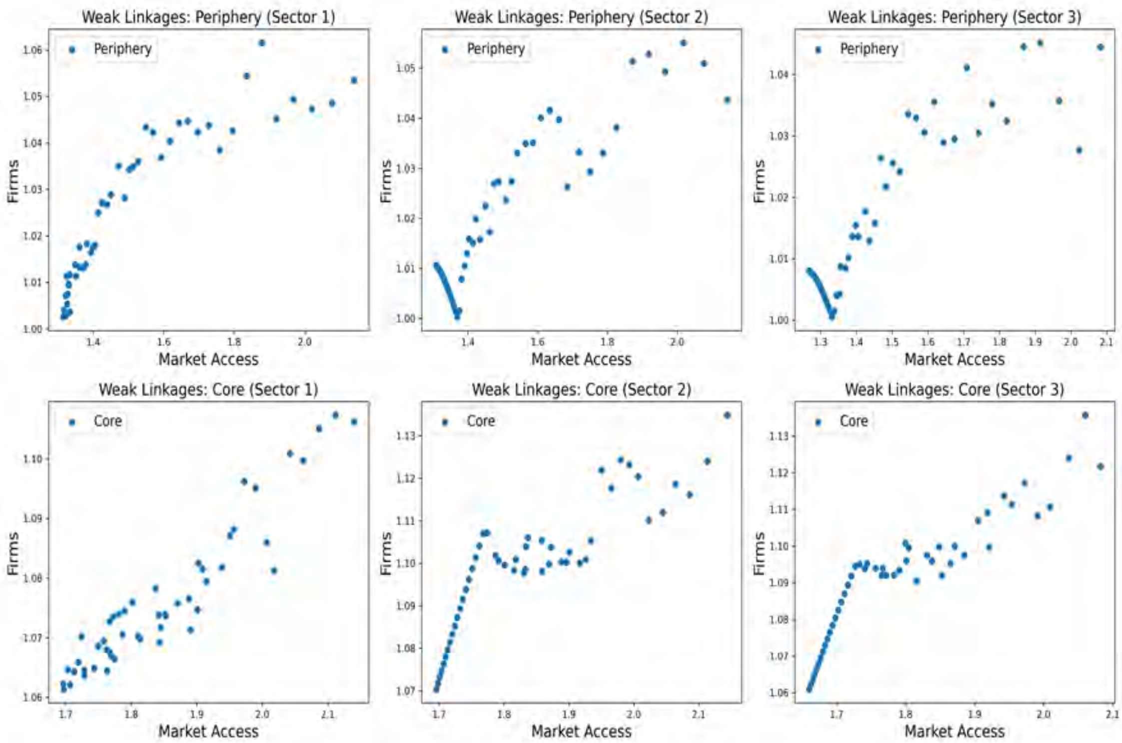


FIGURE A1: Weak Input-Output Linkage Case: Market Access vs. Number of Manufacturing Firms

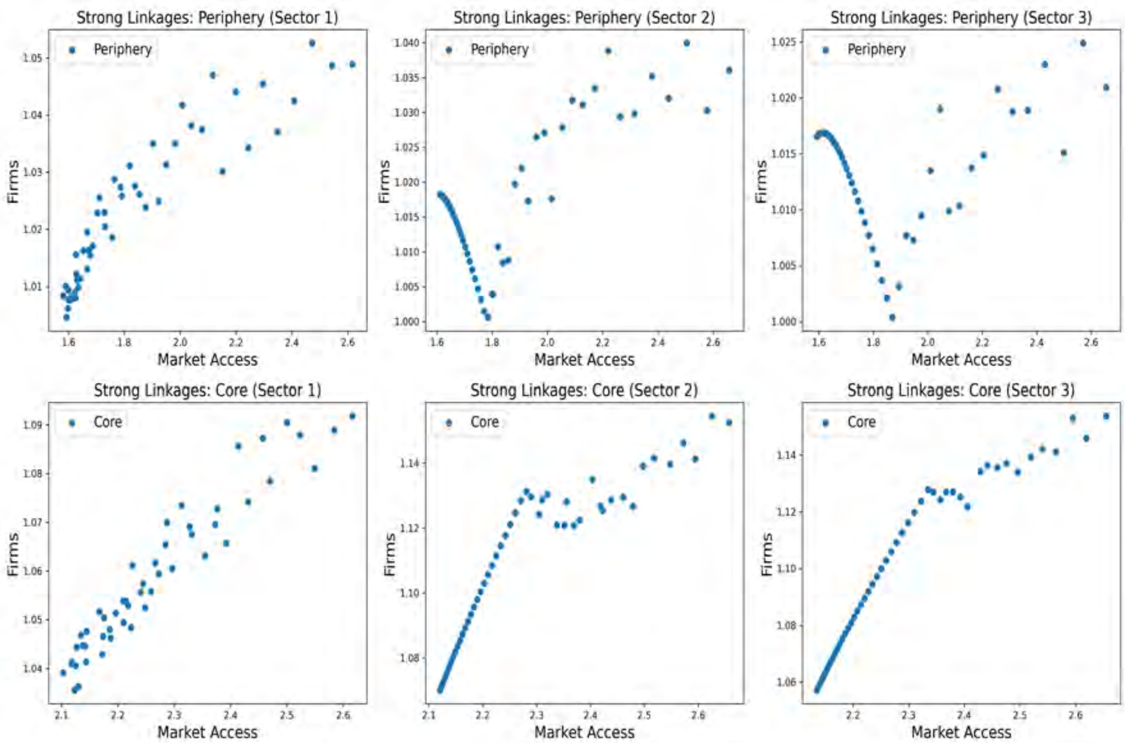


FIGURE A2: Strong Input-Output Linkage Case: Market Access vs. Number of Manufacturing Firms

Periphery Region

In the periphery region, the effect of market access on the number of firms follows a V-shape, except sector 1 (primary industry with limited intermediate linkage) where the effect is always positive.

At low levels of market access, the competition effect dominates. As more firms enter the market, they face increasing competition, but the market is still small, and the infrastructure for firms is not as developed. In this situation, market access may not yet be sufficient to compensate for the higher costs and the increased pressure from competition. The result is a decline in the number of firms at lower levels of market access, as firms struggle to survive in a market with limited opportunities.

However, as market access improves and reaches a certain threshold, the agglomeration effect starts to outweigh the adverse competition effects. With better access to markets, firms begin to benefit from lower transportation costs, improved input-output linkages, and a larger consumer base. This creates an environment where firms can achieve economies of scale, reduce costs, and increase profitability, which results in a positive effect on the number of firms. The agglomeration effect becomes especially pronounced as firms cluster together, sharing resources and benefiting from stronger economic ties.

The threshold level of market access required for the positive agglomeration effect to kick in is higher when input-output linkages are strong. This suggests that in peripheral regions with stronger linkages, firms face more initial challenges—such as higher entry costs or more complex supply chains—before they can fully benefit from the improved market access. Therefore, regions with stronger linkages need a higher level of market access to overcome the initial inefficiencies and reach a tipping point where the benefits of scale and agglomeration can be realized.

Core Region

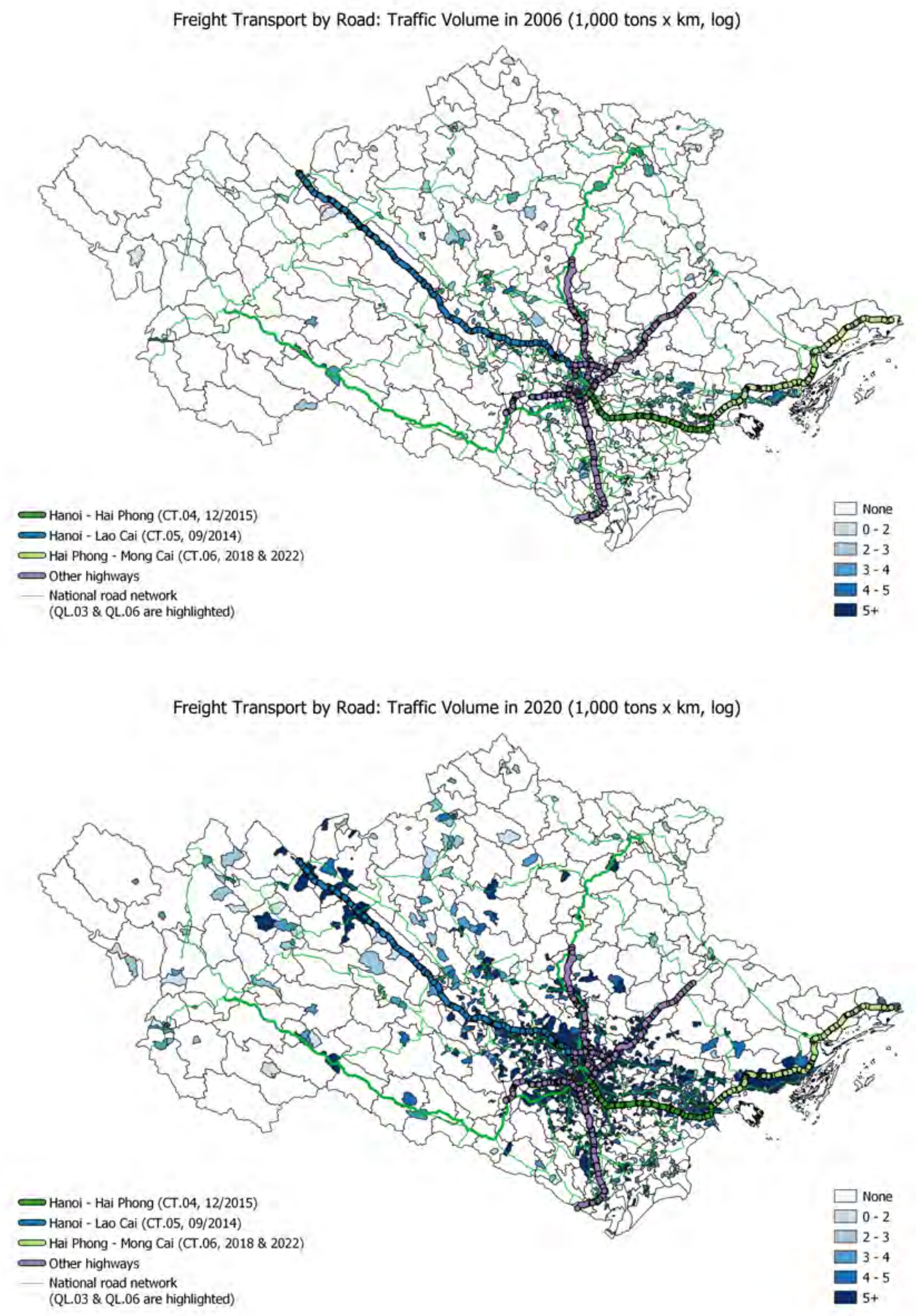
In contrast, the relationship in the core region is more straightforward and gradually positive. At low levels of market access, the agglomeration effect is already quite strong. Core regions tend to have better infrastructure, larger markets, and a higher concentration of firms. As a result, even at low levels of market access, firms in core regions can quickly benefit from reduced transport costs and the advantages of clustering. This leads to a sharp increase in the number of firms as market access improves.

As market access continues to improve, the competition effect becomes more significant. However, the core region has a larger capacity to absorb new firms and a better environment for sustaining firm growth. Therefore, the positive effect of market access becomes more moderate as market access increases beyond a certain point. This indicates that while market access continues to bring benefits, the incremental gains become smaller as the market approaches saturation. At this stage, firms may experience diminishing returns to market access, and the curve flattens.

The relationship is even more pronounced in sectors with stronger input-output linkages. In these sectors, firms benefit more from agglomeration in the early stages of market access improvement because the stronger linkages create more opportunities for firms to reduce costs and improve productivity. However, at higher levels of market access, the core region reaches a point of saturation, where the competition effect moderates further increases in the number of firms. Essentially, the agglomeration effect starts to diminish as the region becomes more densely populated, with firms and resources becoming scarcer.

C Appendix 3: Development in Transportation Sector

FIGURE A3: Freight Transport Recorded by Registered Enterprises: 2006 vs. 2020

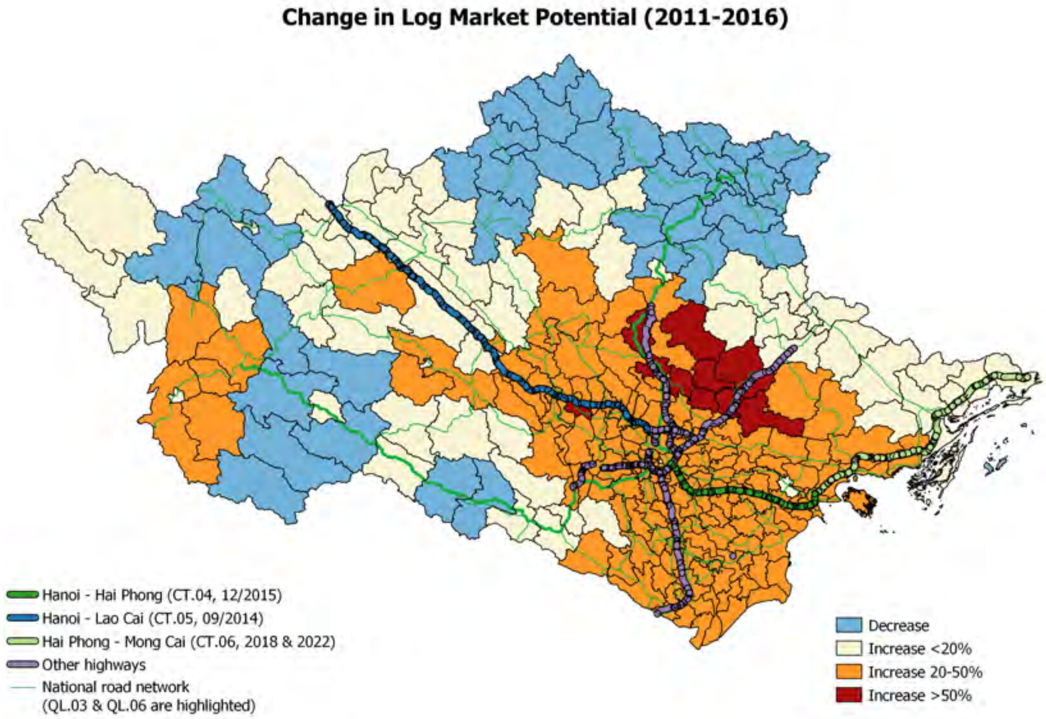


NOTE.—This figure plots the traffic volume of freight transportation by road, which was recorded by registered firms in the VES. Data is aggregated at the commune level.

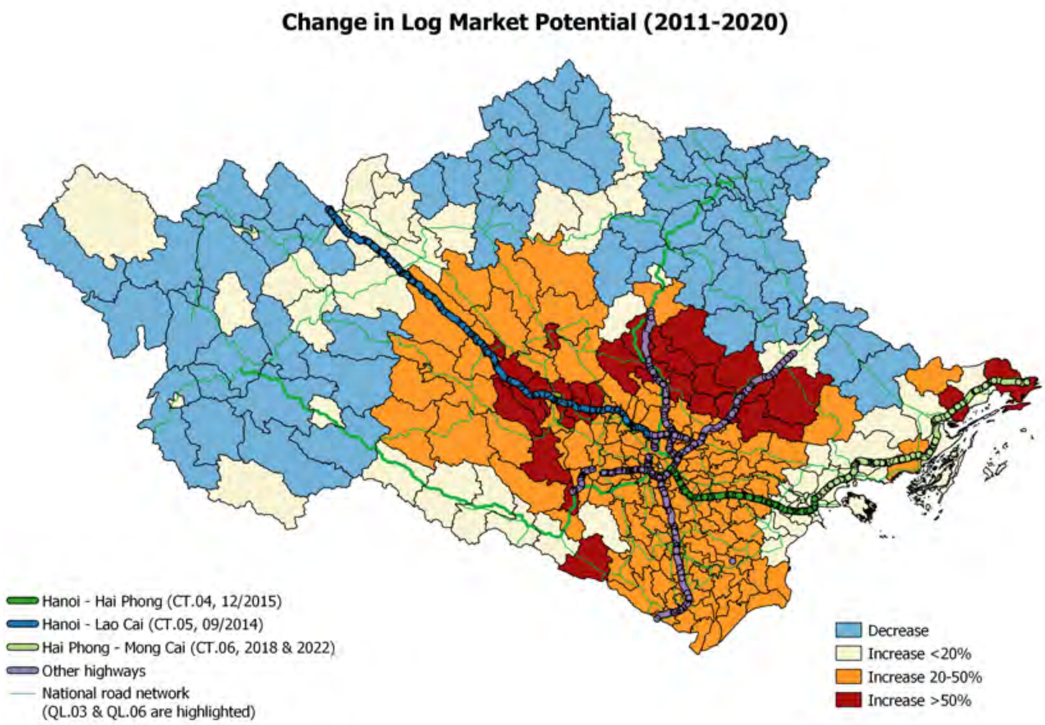
D Appendix 4: Descriptive Evidence using Harris Index

FIGURE A4: Spatial Distribution of the MP Index

((a)) Change in the MP Index from 2011 to 2016



((b)) Change in the MP Index from 2011 to 2020



NOTE.—MP index is computed using total district employment of registered firms. A distance of 50km is used for MP index calculation.