Institute for Economic Studies, Keio University

Keio-IES Discussion Paper Series

From Battlefield to Marketplace: Legacies, Highways and Polycentric Development in Vietnam

能勢 学、澤田 康幸

2026 年 7 月 9 日
DP2026-014
https://ies.keio.ac.jp/publications/27888/

Institute for Economic Studies, Keio University
2-15-45 Mita, Minato-ku, Tokyo 108-8345, Japan
ies-office@adst.keio.ac.jp
9 July, 2026

※ **DP2025-010 の改訂版**

From Battlefield to Marketplace: Legacies, Highways and Polycentric Development in Vietnam

能勢 学、澤田 康幸

IES Keio DP2026-014
2026 年 7 月 9 日
JEL Classification: H54, O14, O18, R11, R12, R42
キーワード: spatial structural transformation, market access, agglomeration, input-output linkages, core-periphery

【要旨】
Why do transport investments induce polycentric industrial growth in specific instances, rather than reinforcing existing cores? By investigating large-scale Greater Mekong Subregion highway projects in Northern Vietnam, where postwar state-led reconstruction restored monocentric urban structure following wartime devastation, this paper shows that later corridor investments generated nonlinear, polycentric industrial growth. Improved market access induced large manufacturing entry in agrarian peripheries, with effects amplified by local supply-chain depth. A spatial general equilibrium model with input-output linkages replicates this mechanism. The corridor raised income but widened spatial inequality, a trade-off mitigated by feeder-road complementarity. Infrastructure acts as a spatial big push only when connectivity and production networks jointly trigger new clusters. Overall, a structural shift from monocentric battlefield resilience to polycentric market expansion is catalyzed by infrastructure investments.

能勢 学
慶應義塾大学 経済学部
mnose@keio.jp

澤田 康幸
東京大学 経済学部、経済学研究科
sawada@e.u-tokyo.ac.jp

謝辞：The authors thank Gilles Duranton, Vernon Henderson, Tomoya Mori, Simon Fuchs, Yuhei Miyauchi, Donald Davis, Rajeev Dehejia, Yasuyuki Todo, Toshihiro Okubo, Hisaki Kono, Yuki Higuchi, Katsuya Takii, Shuhei Kitamura, Hirokazu Ishise, and Takuma Kamada for their insightful comments and suggestions. We also thank seminar participants at 2026World Bank LEADS, 2025 UEA North American meeting, ADBUTokyo Urban Development Conference, IEA, IIPF, CSAE, JEA, JSIE, JADE, Osaka University, and Kyoto University for their useful feedback. We are grateful to Hiroyuki Miyazaki, Dan Tran, and Tung Nguyen for excellent data analytical support. The project was supported by the Asian Development Bank, 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 remaining errors are our own.

From Battlefield to Marketplace: Legacies, Highways and Polycentric Development in Vietnam*

Manabu Nose†

Yasuyuki Sawada‡

July 9, 2026

Abstract

Why do transport investments induce polycentric industrial growth in specific instances, rather than reinforcing existing cores? By investigating large-scale Greater Mekong Subregion highway projects in Northern Vietnam, where postwar state-led reconstruction restored monocentric urban structure following wartime devastation, this paper shows that later corridor investments generated nonlinear, polycentric industrial growth. Improved market access induced large manufacturing entry in agrarian peripheries, with effects amplified by local supply-chain depth. A spatial general equilibrium model with input-output linkages replicates this mechanism. The corridor raised income but widened spatial inequality, a trade-off mitigated by feeder-road complementarity. Infrastructure acts as a spatial big push only when connectivity and production networks jointly trigger new clusters. Overall, a structural shift from monocentric battlefield resilience to polycentric market expansion is catalyzed by infrastructure investments.

Keywords: spatial structural transformation, market access, agglomeration, input-output linkages, core-periphery

JEL Codes: H54, O14, O18, R11, R12, R42

*The authors thank Gilles Duranton, Vernon Henderson, Tomoya Mori, Simon Fuchs, Yuhei Miyauchi, Donald Davis, Rajeev Dehejia, Yasuyuki Todo, Toshihiro Okubo, Hisaki Kono, Yuki Higuchi, Katsuya Takii, Shuhei Kitamura, Hirokazu Ishise, and Takuma Kamada for their insightful comments and suggestions. We also thank seminar participants at 2026 World Bank LEADS, 2025 UEA North American meeting, ADB-UTokyo Urban Development Conference, IEA, IIPF, CSAE, JEA, JSIE, JADE, Osaka University, and Kyoto University for their useful feedback. We are grateful to Hiroyuki Miyazaki, Dan Tran, and Tung Nguyen for excellent data analytical support. The project was supported by the Asian Development Bank, 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 remaining errors are our own.

†Keio University, Faculty of Economics; mnose@keio.jp (Corresponding author)

‡The University of Tokyo, Faculty of Economics; sawada@e.u-tokyo.ac.jp

1 Introduction

Why are urban systems often resilient to physical destruction yet vulnerable to changes in market access? Cities have faced a range of shocks throughout history, such as natural hazards, pandemics, and wars, that ruin buildings, infrastructure, and dense urban cores. In principle, such destruction could trigger substantial and persistent reallocation of economic activities. Yet recent evidence suggests a more nuanced pattern: urban structure is often resilient to physical destruction but reshaped by economic and political shocks (Glaeser, 2022; Lin and Rauch, 2022; Takeda and Yamagishi, 2026). This paper empirically studies this resilience-versus-vulnerability hypothesis in postwar Northern Vietnam, where the same urban system underwent a rare two-stage transformation. Following wartime destruction, state-led reconstruction restored a largely monocentric economy centered on re-established urban cores, whereas later market integration through corridor investments generated a nonlinear shift toward polycentric industrialization.

The central question is why subsequent transport investments have catalyzed polycentric industrial growth rather than reinforcing existing cores. While quantitative spatial models (QSM) provide a flexible framework for analyzing trade cost reductions (Allen and Arkolakis, 2014; Redding, 2022), their application to developing and post-conflict economies where historical legacies, market frictions, and production networks differ from those in advanced economies remains an open empirical question. In rapidly transitioning economies characterized by pre-existing institutional and regulatory rigidities, infrastructure does not merely facilitate marginal trade; it can trigger a spatial take-off by integrating fragmented production networks and inducing labor reallocation across sectors. Capturing such non-linear transformations still requires a framework tailored to the structural features of developing economies.

By investigating large-scale Greater Mekong Subregion (GMS) highway projects in Northern Vietnam, this paper extends the canonical QSM to endogenously explain the mechanisms behind the emergence of new industrial centers in the periphery. With possible sectoral reallocation and input-output (IO) amplification, our framework aims to address the *fenced city* challenge, the difficulty of modeling the transition from an empty periphery to a thriving hub (Duranton and Mori, 2025). Northern Vietnam offers an ideal laboratory to test this framework: heavy aerial bombardment during the Vietnam War destroyed transportation networks in core cities (Miguel and Roland, 2011), and subsequent state-led reconstruction entrenched the economic dominance of Hanoi and Haiphong, while peripheral districts remained largely agrarian and isolated. Against this polarized baseline, the GMS program launched in 1992, aimed at transforming former battlefields into integrated economic corridors, invested in highway upgrading during the 2010s. Despite being widely regarded as a successful model of corridor-led development (Asian Development Bank, 2020), the long-run spatial effects of the GMS highways and the mechanisms behind their success remain poorly understood in the literature.

Combining geocoded firm census data, commune-level statistics, and satellite imageries, we document a pattern that lies outside the ordinal comparative statics of standard models. The GMS upgrades did not simply reinforce agglomeration in urban cores (*centripetal* forces) nor induce monotonic dispersion (*centrifugal* forces). Instead, they triggered the nonlinear spatial structural transformation: the emergence of multiple new manufacturing hubs in previously agrarian peripheral districts. This shift from a *monocentric* system centered on Hanoi to a *polycentric* industrialization offers a novel contribution to the literature, as the role of infrastructure in driving such spatial structural shifts has remained underexplored.[1][2]

We establish these findings in four steps. First, Section 2 demonstrates that despite the catastrophic damage inflicted by the Vietnam War, post-war reconstruction enabled the historical core to regain its economic dominance, ultimately reinforcing a stark core–periphery configuration. Second, in Section 3, using a geography-based instrumental variable (IV) constructed from a least-cost path (LCP) spanning-tree network, we show that improved market access causally increased manufacturing firm entry. The estimated agglomeration dynamics are disproportionately stronger in the periphery relative to the saturated Red River Delta (RRD) core, a convex pattern less commonly documented in advanced-economy settings (Baum-Snow, 2007; Duranton and Turner, 2012; Faber, 2014). Third, difference-in-differences (DID) estimates in Section 4 reveal a nonlinear corridor-centric pattern: peripheral communes within 20–50 km of the highway accumulated 0.39 log points more manufacturing firms and 0.85 log points more jobs over a decade. These effects are strongly heterogeneous: communes with deeper upstream supply-chain linkages experience stronger agglomeration response near the highway corridor. Finally, Section 5 develops a spatial general equilibrium model augmented with IO linkages and validates it against the data using the Adão *et al.* (2024)'s goodness-of-fit test. While the IV and DID estimates establish the key empirical patterns, they cannot by themselves determine whether the nonlinear agglomeration responses arise from input-output amplification, competitive catch-up in thin peripheral markets, or their interaction. Nor can they capture the general equilibrium adjustments, such as wage responses and sectoral reallocation, that shape the aggregate spatial equilibrium. A structural framework is therefore needed to decipher the reduced-form patterns, quantify the underlying mechanisms, and evaluate counterfactual infrastructure designs. The estimated model, featuring IO amplification and endogenous wage-driven dispersion, quantitatively accounts for the

[1] While a few recent papers have begun to study structural transformation in space (e.g., Trew, 2020; Fajgelbaum and Redding, 2022; Nagy, 2023; Eckert and Peters, 2022), empirical evidence on the role of transport infrastructure in driving structural shifts remains scarce.

[2] Theoretically, Akamatsu *et al.* (2025) show that multimodal agglomeration can emerge under a global dispersion force, which induces the establishment of firms distant from urban centers to avoid local competition. At the same time, the flattening of existing agglomeration may occur as a result of local dispersion forces driven by land scarcity.

corridor-centric development observed in northern Vietnam.

Counterfactual analysis in Section 6 reveals a fundamental efficiency–equity trade-off in corridor-led development and shows how infrastructure design can mitigate the tension. The GMS corridor generated strong IO amplification but spatially concentrated gains, as inland districts lacking *last-mile* road access could not activate local IO multipliers. A budget-neutral reallocation that adds feeder roads reduces spatial inequality while preserving most efficiency gains by extending market access to inland areas and activating supply-chain linkages across a broader geography, consistent with evidence from rural road programs in developing economies (Asher and Novosad, 2020). In contrast, policies focused solely on trunk highways or manufacturing upgrading raise aggregate income but intensify spatial concentration. The key lesson is that combining trunk connectivity with feeder-road investment, rather than corridor expansion alone, is essential for achieving both overall efficiency and spatial equity.

We believe that this paper makes four contributions. First, it provides new reduced-form causal evidence on the two-stage postwar transformation of urban structure from monocentric recovery after physical destruction to polycentric industrialization following market integration in a post-conflict economy, contributing to the literature on urban resilience and vulnerability (Glaeser, 2022; Lin and Rauch, 2022). Second, it establishes that IO linkages and market access jointly produce spatially heterogeneous firm agglomeration, extending the big push logic of pecuniary externalities (Murphy *et al.*, 1989; Krugman, 1992) to a spatial model.[3] Third, it applies the Adão *et al.* (2024)'s model validation framework to infrastructure evaluation in a developing economy, showing that IO amplification is quantitatively essential to account for the observed corridor-centric development pattern (Matsuyama, 1995). Fourth, it provides policy diagnostics, demonstrating that achieving inclusive regional growth requires a *triple-trigger* mechanism: (i) sufficiently large market accessibility, (ii) deep local supply-chains to sustain agglomeration, and (iii) complementary feeder-roads to diffuse gains beyond established nodes. These findings reveal that an interregional highway alone is insufficient for spatial inclusion without the local capacity to activate industrial multipliers.

The paper proceeds as follows. After we describe the initial setting and data in Section 2, Section 3 presents IV-based reduced-form evidence on the causal effect of market access. Section 4 examines the spatial and supply-chain mechanisms. Section 5 develops, calibrates, and validates the quantitative model. Section 6 conducts counterfactual analysis. Section 7 concludes.

[3]Whereas spatial-equilibrium studies of transport infrastructure (Tsivanidis, 2026; Morten and Oliveira, 2024; Allen and Arkolakis, 2022) mainly consider migration and commuting choices, we focus on producers' entry decisions in the presence of input-output linkages and sectoral labor reallocation.

2 Setting and Data

In this section, with satellite-based and firm-level data, we demonstrate a two-stage evolution of Northern Vietnam's economic geography. The post-war period is characterized by a monocentric structure centered on Hanoi. With the GMS corridor upgrades, however, the region exhibits a marked shift toward polycentric industrialization. The evidence points to a transition from *battlefield* resilience to *marketplace* expansion through infrastructure investments.

2.1 Historical Background, Core-Periphery Baseline, and GMS Highway Upgrades

Following Vietnam's reunification in 1975 after the devastating war, economic recovery in the north was heavily skewed toward the RRD (Miguel and Roland, 2011). The government prioritized reconstruction of war-damaged infrastructure in Hanoi and the port city of Haiphong (Tan, 2014), while northern border provinces, historically a battlefield zone due to the Sino-Vietnamese conflict of the late 1970s, remained isolated and underdeveloped.

Figure 1 documents this asymmetric legacy by plotting wartime bombing intensity against districts' market potential (calculated by eq. (A3) in Appendix D). In the north, heavily bombed locations had regained *higher* market potential, consistent with their prioritization in postwar reconstruction.[4] Even after large wartime shocks, reconstructed cores re-established their pre-war prominence, a pattern of strong path dependence (Davis and Weinstein, 2002) that shaped the baseline for later highway investments.

By the end of the 1990s, the result was stark core-periphery divergence. Core districts in the RRD region were far more densely populated than the national average (Table 1a). Peripheral areas in the Northeast (hereafter, NE) and Northwest (hereafter, NW) remained largely agricultural through the 1990s, though population in the NW began to grow from a lower base. About one-third of all registered firms in Vietnam were located in the RRD, where services dominated the local economy (Table 1b). After the start of GMS highway construction, the increase in the number of firms was particularly pronounced in the northern regions. Based on regional averages of district-level enterprise growth rates, the Northwest recorded a 76% increase between 2011 and 2016, exceeding the national average, while the Northeast also experienced a large increase between 2016 and 2020. Manufacturing employment in northern provinces also grew at 7.4% per year, outpacing the national average of 5.1%, yet gains were disproportionately concentrated in and around

[4] In the south, by contrast, bombing intensity correlates negatively with initial market potential, suggesting a different wartime legacy and reconstruction trajectory. Our resilience interpretation therefore focuses on northern Vietnam, where heavily bombed locations were more likely to be reconstructed as the postwar core, consistent with the spatial resilience to physical shocks (Davis and Weinstein, 2002; Glaeser, 2022).

the core. In the mountainous border regions, poverty rates exceeded 50% and electrification was below 40% (Figure 2). This polarized baseline, an economically dynamic core surrounded by agrarian and infrastructure-poor peripheries, is the starting point against later GMS highway investments.

Northern Vietnam thus provides a distinctive setting because wartime destruction and its aftermath created a stark core–periphery structure. Heavy bombing damaged transport networks, and subsequent reconstruction reinforced the primacy of Hanoi and Haiphong, while many peripheral districts remained agrarian and relatively disconnected. In the paper's framing, this persistence of a core-centered structure captures the city system's resilience to physical shocks, a robust recovery from the *battlefield* legacy that sets up the later quasi-experimental test.

Accordingly, we focus on the North-South Economic Corridor, which upgraded connectivity from Kunming and Lao Cai (the border city with Yunnan province in China) to Hanoi, with extended links to Haiphong and Lang Son (a border city with Guangxi province). The centerpiece was the Noi Bai–Lao Cai Expressway and upgrades to national highways connecting to the Chinese border (Figure 3). These investments replaced slow, winding mountain roads with modern multi-lane expressways, reducing travel time between the border (Lao Cai) and Hanoi from over 10 hours to about 3.5 hours, with daily traffic rising from 4,000 to 18,000-19,000 passenger car units (PCU) by 2017 (GMS Secretariat, 2018).

2.2 Data

Our analysis draws upon a rich set of data sources combining granular ground-level information and remote sensing imageries (see Appendix A).[5] The comprehensive spatial panel data covering Northern Vietnam from 2000 to 2020 are constructed from the Vietnam Enterprise Survey (VES) conducted by the General Statistics Office (GSO), which provides annual information on firm location, employment, and ownership for the universe of registered firms. We aggregate our analysis to the district and commune levels to measure local industrial agglomeration. Our study area covers approximately 4,350 communes within 270 districts across three northern regions. We define "**core cities**" as districts in the RRD region, including Hanoi and Haiphong, two central provinces in the north.[6] Districts in the NW and NE regions are defined as "**peripheral cities**" (Figure 4).

Additionally, we employ the harmonized Input-Output Database compiled by the Organisation for Economic Co-operation and Development (OECD) and the 2007 Viet-

[5] Donaldson and Storeygard (2016) reviews recent applications of satellite data in economics. Our paper closely relates to Burchfield *et al.* (2006) which uses remote-sensing data to track the evolution of land use and urban sprawl at finer grid level over decades in the U.S.

[6] In Vietnam, cities directly under central government include Hanoi, Haiphong, Danang, Can Tho, and Ho Chi Minh City. These constitute the socioeconomic cores in each region.

namese Input-Output Tables, published by the General Statistics Office (GSO), to construct the supply chain network data and related parameters. To incorporate spatial and geographical dimensions, we integrate the Japan Aerospace Exploration Agency (JAXA) Advanced Land Observing Satellite (ALOS) Land Cover Map alongside the National Aeronautics and Space Administration (NASA) Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Global Digital Elevation Model.

Due to inconsistencies in tracking state-owned enterprises (SOEs; including collectives) in the annual survey, we restrict the analysis to private registered firms to avoid confounding effects from the ongoing SOE reforms.

2.3 Spatial Patterns: From Monocentric to Polycentric

Before turning to econometric analyses, we document the raw spatial transformation using high-resolution satellite imageries over decades, the motivating fact that drives our empirical investigation. 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 to track built-up area expansion at the commune level from 1990 to 2019.

Figure 5 shows the expansion of built-up areas along the GMS corridor from the onset of the *Doi Moi* reform in 1990 until 2019. The pattern is striking in its three-phase structure. In 1990–2000, economic activity was tightly concentrated in the Hanoi–Haiphong axis; outside this core, virtually no district exceeded 10% built-up coverage. The mountainous periphery faced high trade costs, with transport connections to the capital consisting of degraded colonial-era roads. In 2000–2010, as the economy liberalized, built-up areas began spreading toward peripheral cities along the coast (Haiphong, Ha Long) and toward the Chinese border (Lao Cai, Yen Bai). After 2010, improved connectivity facilitated by the GMS investments further accelerated peri-urban expansion, particularly along the corridor. Urban fringes also sprawled further down into remote peripheral areas, such as the Laos-border provinces of Hoa Binh and Son La, facilitated by the National Highway 6 (locally known as Quốc Lộ 6 (QL6)) corridor. Taken together, this satellite evidence documents a clear transition from a *monocentric* structure centered on Hanoi toward a more *polycentric* configuration. We interpret the earlier monocentricity as reflecting the city's resilience to physical shocks, i.e., robust recovery from the *battlefield*, while the subsequent move toward polycentricity appears driven by the evolution of the diversified *marketplaces* through infrastructure investments and the resulting expansion of market access.

The satellite-based evidence is consistent with highway effects but cannot rule out confounders: the same period saw broad Vietnamese economic growth, ongoing place-based industrial zone policies, and growing foreign direct investments. The next two sections use quasi-experimental variation to isolate the causal contribution of the GMS investments.

3 Does Market Access Drive Industrial Clusters?

This section asks whether GMS highway upgrades increased manufacturing firm entry, and whether the effect differs between core and periphery. We use an IV approach that exploits exogenous geographic variation in connectivity improvements.

3.1 Empirical Strategy

Firm Market Access. We follow Redding (2022) and measure market access using a theoretically grounded Firm Market Access (FMA) index:

$$FMA_{s,i} = \sum_{j=1}^{J} \tau_{ij}^{1-\sigma} \cdot \frac{E_{s,j}}{P_j^{1-\sigma}} \quad (1)$$

where τ_{ij} is the iceberg trade cost between districts, $E_{s,j}$ is total expenditure in sector s at destination j, and P_j is the destination price index.[7] The FMA index is defined as the sum of destination markets' real expenditure, discounted by bilateral trade costs. We compute trade costs directly from freight trade values and volumes in the VES. The distance elasticity of trade is set at 4 based on Simonovska and Waugh (2014). Figure 6 illustrates the resulting changes in trade costs and FMA over 2011–2020: large decreases in costs and large increases in market access are concentrated along the GMS corridor, particularly in previously isolated peripheral areas. An exception is the area around Yên Bái province, a mountainous region where geographical constraints limit the construction of connecting roads to the highways.

Identification challenge. A key hurdle in estimating the causal impacts of highways is their non-random placement. Since roads are often built to connect growing commercial hubs (reverse causality) or to stimulate struggling regions (selection bias), the Ordinary Least Squares (OLS) estimates are likely systematically biased.[8] In fact, the routes in the GMS program were driven primarily by the geopolitical goal of regional integra-

[7] Because sector-level local price indices cannot be constructed from the VES, we replace $P_{s,j}$ of the structural model (Appendix G) with a district-level price index P_j in the reduced-form estimation.

[8] Estimating the causal effect of highways is challenging because randomized experiments are inherently infeasible. Existing literature has used various instruments for infrastructure, such as (a) historical transport network (Baum-Snow *et al.*, 2017; Duranton and Turner, 2012; Michaels, 2008), (b) counterfactual least-cost-path network (Faber, 2014; Morten and Oliveira, 2024), (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 (Dinkelman, 2011; Duflo and Pande, 2007).

tion, connecting Vietnam to the broader Mekong sub-region and facilitating trans-border commerce with China, rather than by anticipation of local district-level demand (Asian Development Bank, 2020). ADB Board documents and government decrees confirm that route selection followed a corridor logic, determined at the sub-regional level, with specific alignments shaped by topography rather than local economic conditions. This exogeneity argument is formalized in our instrumental variable strategy, which exploits sharp topographic contrasts between the mountainous border regions and the flat RRD to isolate quasi-experimental variation in connectivity improvements. More specifically, to address this econometric estimation issue, we combine the *inconsequential unit* approach with least-cost path (LCP) and other geography-based instrumental variables (Neubert and Nikolić, 2026; Gibbons *et al.*, 2019; Ahlfeldt and Feddersen, 2018). We restrict the sample to districts within 100 km of the project site and exploit variation in the magnitude of cross-regional market access changes among otherwise similar local locations.

MST and Broad Geography-Based Instruments. Following the LCP network literature (Faber, 2014; Morten and Oliveira, 2024), we construct a hypothetical Minimum Spanning Tree (MST) highway network derived from LCP analysis (see Appendix B). The instrument exploits northern Vietnam's unique topological conditions: sharp contrasts between the flat hinterland of the RRD and the mountainous border regions which are inhabited by ethnic minority groups, create a natural cost surface that determines engineering-optimal routes independent of local economic shocks.

Formally, we generate the LCP spanning tree network connecting provincial capitals and major nodes using terrain and land-cover data (slope, urban development, water bodies, wetlands). We adopt the following cost function from the transport engineering literature (Jha *et al.*, 2001; Jong and Schonfeld, 2003):

$$c_i = 1 + Slope_i + 25 \times Developed_i + 25 \times Water_i + 25 \times Wetland_i$$

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 and NASA-ASTER global digital elevation model (both in 30 meter spatial resolution, based on satellite imagery from 1990) as inputs and apply Kruskal's algorithm to identify the minimum spanning tree (Figure 7). The resulting network is independent of local economic shocks and satisfies the exclusion restriction: it affects FMA growth through terrain-driven route placement, not local demand conditions. Accordingly, we use district-level measures of distance to the MST network constructed from 1990 satellite imagery as an IV.

MST distance captures a single source of variation, whether a district locates close to the engineering-optimal network, but does not span the full distribution of market access conditions needed to identify nonlinear responses. Hence, our instrument set Z_{ct_0} incorporates broader geography-based variables to capture northern Vietnam's unique topological conditions, i.e., land elevation, sunshine (both measured in 1999), and U.S. bombing intensity during the Vietnam War (Miguel and Roland, 2011). These geographic variables add independent variation in the effective cost of market integration for districts that face natural barriers even when proximate to the road network. We use U.S. bombing intensity as a historically predetermined source of variation in subsequent infrastructure development, based on prior evidence that bombing exposure had little persistent direct effects on long-run Vietnamese economic outcomes (Miguel and Roland, 2011; Glaeser, 2022). Together, these instruments create natural barriers to market expansion and are uncorrelated with post-2011 productivity shocks conditional on regional fixed effects. In this way, we employ an overidentified specification because identifying nonlinear effects of market access requires at least two instruments, while additionally permitting Hansen's overidentifying restriction test.

Regression. We estimate a first-difference two-stage least squares (2SLS) specification over 2011–2020:

$$\Delta \ln FMA_{ct} = \delta_0 + Z_{ct_0}\delta_1 + X_{ct_0}\delta_2 + \Delta\nu_{ct} \quad (2)$$

$$\Delta \ln y_{ct} = \beta_0 + \beta_1 \Delta \ln FMA_{ct} + X_{ct_0}\gamma + \Delta\varepsilon_{ct} \quad (3)$$

where y_{ct} is the number of enterprises in commune c. Controls X_{ct_0} include the initial share of female workers, social security coverage, proximity to industrial zones (within 5 km), and distance to national roads. We estimate separately for the RRD core and the NE/NW periphery to capture the non-linearities, where centrifugal and centripetal forces operate differently across regions. Standard errors are clustered at the commune level.

3.1.1 IV Results: Large, Convex Effects Concentrated in the Periphery

Table 2 reports the first-stage estimates. Log distance to the MST network is strongly negatively associated with FMA growth in both regions (Kleibergen-Paap F-statistics exceed the rule-of-thumb threshold of 10 in all specifications). Land elevation and sunshine significantly influence FMA growth in the periphery, whereas historical bombing is associated with lower subsequent market access growth in core areas, consistent with the path-dependent reconstruction pattern documented in Section 2.1. Heavily bombed core cities recovered resiliently after the war, which substantially raised their initial level of market access. As a result, these cities had less scope for further FMA growth in the subsequent period, generating a pattern of mean reversion.

Table 3 presents the second-stage estimates.[9] We report OLS alongside 2SLS for comparison. In the NE and NW regions (Panel B), the observed downward bias of OLS is consistent with highways being directed toward initially lagging areas where improved market access yields stronger impacts (Redding and Turner, 2015; Jedwab and Storeygard, 2020). While the Hansen J-statistics broadly support that overidentifying restrictions are valid, the estimates reveal a striking core-periphery asymmetry. In the NE and NW periphery regions (Panel B), a 1% increase in FMA raises manufacturing firm entry by approximately 0.14%, with similarly strong effects in service sectors. By contrast, in the RRD core (Panel A), elasticities are relatively moderate, consistent with the market saturation and heightened competition in already-dense markets during the 2010s.

Crucially, the squared-term specifications (columns 3, 6, 9) reveal accelerating (convex) effects in the periphery: the positive and significant coefficients on the squared firm market access term, $(\Delta \ln FMA)^2$, indicate that agglomeration accelerates once market access crosses a critical level. In the core, the squared terms are generally insignificant (except manufacturing), consistent with diminishing returns in mature agglomerations that have already exhausted most within-region external economies.

This asymmetry suggests a threshold-type transition between centrifugal and centripetal regimes, consistent with core-periphery models in which regions below a critical level of trade integration remain at a low-agglomeration equilibrium while those above it undergo rapid catch-up (Krugman, 1991; Krugman and Venables, 1995). Section 4 examines the empirical sources of this nonlinearity, and Section 5 formalizes the mechanism in a quantitative spatial model.

4 Sources of Nonlinear and Polycentric Development

The IV estimation results in the previous section establish two facts: (i) improved market access causally increases manufacturing firm entry, and (ii) this effect is larger and accelerates with connectivity in the periphery compared to the core. However, these results do not unpack why the effects are nonlinear and spatially asymmetric. This section examines two potential complementary mechanisms: the spatial configuration of agglomeration by distance zone, and the role of supply-chain depth in amplifying corridor effects.

[9] Table 2 and 3 report results from the preferred overidentified specification. Appendix C presents first- and second-stage results under both just-identified (MST IV only) and overidentified specifications. The overidentified model yields stronger first-stage relevance and more precisely identifies non-linear market access effects, particularly in the NE and NW periphery.

4.1 Spatial Shape of Effects: Distance Decay and Exposure Bands

To examine *where* the agglomeration occurs, we employ a flexible two-way fixed-effect difference-in-differences (DID) specification that allows the highway effect to vary by distance zone and to evolve over time:

$$
\begin{aligned}
\ln y_{ct} = &\beta_1 \mathbb{1}[Dist_c \leq 20\text{km}] \times POST + \beta_2 \mathbb{1}[20\text{km} < Dist_c \leq 50\text{km}] \times POST \\
&+ \beta_3 \mathbb{1}[Dist_c \leq 20\text{km}] \times POST \times t + \beta_4 \mathbb{1}[20\text{km} < Dist_c \leq 50\text{km}] \times POST \times t \\
&+ \gamma X_{ct} + \mu_c + \tau_t + \lambda_{dt} + \varepsilon_{ct},
\end{aligned} \tag{4}
$$

where y_{ct} denotes firm counts or employment in commune c for year t. $Dist$ is the commune's distance to the GMS highway. $POST$ indicates post-construction years ($t \geq 2011$), and μ_c, τ_t, λ_{dt} denote commune, year, and district-specific trend fixed effects. X_{ct} controls for the effect of industrial zones. The 20 km and 50 km distance cutoffs are chosen based on semi-parametric estimates in Appendix D.2, which show rapid market potential growth within 20–50 km of the highway with effects declining beyond 50 km distance buffer. Following Greenstone *et al.* (2010), we allow both a *mean shift* (β_1, β_2) and a *trend break* (β_3, β_4) in the outcome to distinguish immediate from cumulative effects. In the specification with trend break parameters (columns 2, 4, and 6 of Tables 4 and 5), we report the combined 10-year effect $\beta_1 + 10\beta_3$ (for communes within 20km) and $\beta_2 + 10\beta_4$ (for communes in the 20-50km zone) as our main estimand.

Our control group consists of communes outside the 50 km exposure band, where the direct accessibility gains from the highway are substantially weaker. Conditional on commune and year fixed effects, district-specific trends, and baseline controls, they provide a valid counterfactual for treated communes in the absence of the highway upgrade. This assumption is supported by the event-study evidence in Figures 8a–b using the pooled data from the core and periphery regions, which shows parallel pre-construction trends in both firm counts and employment. Importantly, the parallel trend holds when defining treatment as either 20 km or 50 km, reinforcing the robustness of our design. After 2011, outcomes in treated areas diverge upward, with marked acceleration after 2015 coinciding with the completion of major highway sections, a timing consistent with the infrastructure channel rather than contemporaneous demand shocks.

Periphery results. In the NE and NW periphery (Table 4, Panel B), we find a distinctly polycentric pattern that contrasts with a standard monocentric prediction of corridor-led agglomeration. The estimated effect on manufacturing firm entry within 20 km is 0.47 log points, corresponding to 59% increase over 10 years. Importantly, the effects are not confined to areas immediately adjacent to the corridor. In the 20–50 km zone, the estimates indicate 0.39 log points more manufacturing firms, 0.13 log points more transport firms, and 0.25 log points more trade firms, corresponding to increases of 48%, 14%, and 28%, respectively. This wider spatial diffusion in the periphery contrasts with the more localized pattern observed in the core, and is not easily reconciled with a purely local agglomeration mechanism operating in the immediate corridor vicinity.

Employment effects (Table 5) are even more pronounced. Manufacturing employment increases by 0.78 log points within 20 km and by 0.85 log points in the 20–50 km zone, corresponding to 117% and 133% increases over 10 years, respectively. Second-tier cities such as Hoa Binh, Son La, and Thai Nguyen, connected via national highways (QL3 and QL6) (Figure 3), recorded strong expansions in manufacturing (textiles, food processing) and trade. The diffused spatial development is consistent with satellite-based evidence of peri-urban expansion (Figure 5) and the FMA improvements (Figure 6).

Core results. In the RRD core (Table 4, Panel A), the effects are concentrated within 20 km of the corridor and become smaller in the 20–50 km band. The estimated effect on manufacturing firms within 20 km is 0.35 log points, corresponding to 41% increase over 10 years, while the effect further away are statistically insignificant, except for trade. This spatial concentration is consistent with the reinforcement of existing urban and industrial hubs near the highway corridor. Trade activities, including retail, food services, and accommodation, show a somewhat more diffuse pattern, with positive effects extending to 20–50 km zone. This likely reflects demand spillovers from improved connectivity, which encouraged the formation of small consumer-oriented businesses in surrounding areas. Similar spatial patterns are found for employment in Table 5.

These results document that highway access does not affect all locations and industries uniformly. To ensure that these patterns reflect the intrinsic impact of infrastructure rather than confounding factors, and to provide a preliminary look at the industrial drivers behind this spatial clustering, we examine the robustness of our results against concurrent place-based policies and explore how industry-specific characteristics—particularly production linkages—matter in shaping the observed agglomeration patterns.

Robustness checks. In Appendix E, we address the concern that place-based policies, specifically industrial zones (IZs) which often co-locate with highways (Figure A5), possibly confound the observed effects. Using a geocoded database of IZs in northern Vietnam (Kono *et al.*, 2025), we check the stability of the DID estimates when communes within 1, 5, and 10km from IZs are removed. Excluding these communes generally reduces point estimates in peripheral areas, reflecting that place-based policies contribute to attracting firms in the periphery. Nonetheless, the main conclusion holds: highways exert an independent effect on agglomeration beyond industrial policy (Table A3).

Industry heterogeneity and production linkages. Table A4 in Appendix F disaggregates the DID estimates by investigating spatial clustering across manufacturing industries. Within 20 km of the highway, both core and periphery experienced entry across a broader set of manufacturing industries. In the 20–50 km zone, the effects are concentrated in lighter, labor-intensive sectors (such as textiles, garments, wood, paper, and leather products), while heavy and intermediate industries (e.g., metals, machinery, chemicals) display weaker or insignificant responses. We also find emerging IT clusters within 50 km in both core and periphery regions. This pattern is consistent with the hypothesis that industries with stronger production linkages prioritize proximity to transport hubs where dense buyer-supplier networks amplify spillovers, while lighter industries with weaker supply chain linkages spread more diffusely once connectivity improves.

The findings suggest that the complementarity between transport infrastructure and supply-chain intensity plays a pivotal role in shaping the industrial landscape, a mechanism we quantify formally in the next section.

4.2 The Role of Supply Chains: Amplification and Second-Tier Cities

Building on the New Economic Geography (NEG) model, this section extends the DID framework to quantify the importance of supply chain spillovers (Ellison *et al.*, 2010) in reshaping the core-periphery pattern.[10]

Measuring supply chain linkages. We measure intersectoral supply chain linkages using the OECD Harmonized IO Tables (45 sectors). Following Antràs *et al.* (2012), we construct a commune-level upstreamness index M_{ct} as the expenditure-weighted average of industry-level upstreamness L_{st}, using the initial share of industry s's sales in each commune c as weights:

$$M_{ct} = \sum_{s=1}^{45} L_{st} \frac{Sales_{cs,0}}{\sum_s Sales_{c,0}} \tag{5}$$

A higher M_{ct} indicates specialization in industries that are further upstream (more distant from final consumers) and thus likely to benefit disproportionately more when trade costs fall (as intermediate trade expands). In 2006, local intersectoral production linkages were weak, reflecting Vietnam's reliance on imported inputs from China.[11] Over the following decade, supply chains deepened along highway corridors, reinforced by industrial policies promoting mechanical engineering, electronics, textiles and technology industries (Figure A4).[12]

[10]The NEG model (Krugman and Venables, 1995; Puga and Venables, 1996; Venables, 1996; Redding and Venables, 2004) emphasizes that reductions in trade costs from highway investments interact with inter-industry linkages to generate spatial externalities and shape agglomeration patterns.

[11]See Okumura (2021)

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

Generalized DID. To estimate how highway effects vary with supply-chain depth, we employ a generalized DID (Lindo *et al.*, 2020; Callaway *et al.*, 2024) that estimates the average causal response (ACR) to a unit increase in M_{ct}, comparing communes categorized by the equal-sized decile of each commune's distance from the GMS corridor:

$$\ln y_{ct} = \beta\, \mathbb{1}[Q(Dist_c) = q] \times POST \times M_{ct} + \gamma M_{ct} + \delta X_{ct} + \mu_c + \tau_t + \lambda_{dt} + \varepsilon_{ct} \qquad (6)$$

where y_{ct} denotes firm counts and $\mathbb{1}[Q(Dist_c) = q]$ is a dummy for the q-th distance decile. In this model, we compare the heterogeneity in gains from the treatment, i.e., the average treatment effect on the treated (ATT) for treatment group g and a group just below it $g-1$ in the supply chain (Callaway *et al.*, 2024). Identification assumes equal ATT across supply-chain groups at each distance level. In reality, communes differ in their expected returns from the highway upgrade depending on initial industrial composition and stage of local industrial development. If firms locate in areas where they anticipate larger gains, this may potentially bias the estimate. We mitigate bias from initial industrial composition by controlling for pre-GMS industrial level differences.

Results. Figure 9 plots the ACR coefficient β with respect to firm counts by distance decile for the RRD core (left) and NE & NW periphery (right). Three patterns stand out. First, the marginal effect of the highway is significantly larger in communes with stronger upstream linkages: deeper supply chains amplify the agglomeration response near the highway corridor. Second, in the periphery, input-output multiplier effects decline with distance, but recover at 30–50 km, precisely where secondary cities anchor local production networks. Removing communes along major national highways flattens this secondary peak (dashed line), confirming that second-tier cities mediate agglomeration diffusion. Third, in the core, multiplier effects are elevated close to the corridor but show less secondary recovery, consistent with the core's pre-existing, denser supply chains providing agglomeration benefits throughout the treatment zone.

These results provide direct evidence for the IO amplification channel: once trade costs fall, locations with deeper supply chains experience amplified agglomeration. The secondary-city relay effect further shows that supply-chain linkages propagate through the regional production network rather than dying out beyond the immediate corridor.

4.3 From Reduced-Form Evidence to Structural Mechanism

The reduced-form evidence in Sections 3 and 4 points to nonlinear and spatially heterogeneous responses to market access improvements, while the supply-chain results suggest that input-output linkages magnify these responses once connectivity improves sufficiently. We formalize this interpretation using the quantitative spatial model in Section 5, where endogenous labor reallocation between agricultural and non-agricultural sectors, agglomeration externalities, and input-output linkages jointly generate threshold-type behavior. Together, the framework is designed to capture the key mechanisms identified above and to provide a basis for welfare-relevant counterfactual analysis.

5 Quantitative Spatial Model

The framework builds on the new economic geography model of Krugman and Venables (1995), Puga and Venables (1996), and Redding and Venables (2004), augmented by adding an input-output production structure and agglomeration externalities. The model is designed to be theoretically tractable and empirically disciplined, via calibration to the Vietnamese data and the Adão *et al.* (2024)'s causal validation framework.

5.1 Model

Environment. Consider an economy with J locations (districts) indexed by $i, j \in \{1, \ldots, J\}$. The economy is populated by consumers with total income $w_i L_i$, where w_i is the wage in region i and L_i is the regional labor endowment. There are S sectors, broadly categorized into Agriculture (A) and Nonagriculture (NA), which comprises manufacturing and service industries.

Preferences. Consumers have Cobb-Douglas preferences with expenditure share α_A on agriculture and $(1-\alpha_A)$ on a CES composite of nonagricultural varieties. The composite $C_{NA,i}$ is defined over S sectors and the continuum of varieties ω in sector s.

$$U_i = C_{A,i}^{\alpha_A} C_{NA,i}^{1-\alpha_A}, \qquad C_{NA,i} = \prod_{s=1}^{S} Q_{s,i}^{\theta_s}, \qquad Q_{s,i} = \left(\int_{\omega \in \Omega_s} q_{s,\omega,i}^{\frac{\sigma-1}{\sigma}} d\omega \right)^{\frac{\sigma}{\sigma-1}} \tag{7}$$

where $\sigma > 1$ is the elasticity of substitution across varieties and θ_s is the expenditure share on sector s within nonagriculture, with $\sum_{s=1}^{S} \theta_s = 1$.

Agriculture. The agricultural sector operates under perfect competition with constant-returns-to-scale technology, $Y_{A,i} = A_{A,i} L_{A,i}^{\beta_L} K_i^{1-\beta_L}$, where $A_{A,i}$ is agricultural productivity, $L_{A,i}$ is agricultural labor, and β_L is the labor share. Agricultural land K_i is location-specific and inelastically supplied. As nonagriculture expands at location i, agricultural employment $L_{A,i}$ falls, raising wages w_i. This process acts as the engine of structural transformation: highway improvements that raise nonagricultural returns draw labor out of agriculture, consistent with the broad pattern of structural transformation in developing Asia (Herrendorf *et al.*, 2014).

Nonagriculture with IO linkages. Nonagriculture operates under monopolistic competition with increasing returns to scale. Each firm in sector s at location i produces goods and services using labor and intermediate inputs from all sectors, following a Cobb-Douglas production function:

$$Y_{NA,s,i} = A_{NA,s,i} \, L_{NA,s,i}^{\eta_s} \prod_{r=1}^{S} X_{r,s,i}^{\mu_{rs}} \tag{8}$$

where $A_{NA,s,i}$ is nonagricultural sector productivity, $L_{NA,s,i}$ is nonagricultural sector labor, and $X_{r,s,i}$ is the bundle of sector r inputs used by sector s in region i. η_s is the labor share in sector s. μ_{rs} is the input-output coefficient (sector r's expenditure share in sector s production) and $\eta_s + \sum_r \mu_{rs} = 1$. The IO structure generates two interlocking linkages. A *backward linkage* occurs when higher nonagricultural revenue at location i increases demand for upstream intermediate inputs, raising revenues and supporting more firms. A *forward linkage* occurs when a larger local firm base reduces the price index $P_{s,j} = \left(\sum_i n_{s,i}(p_{s,i}\tau_{ij})^{1-\sigma}\right)^{1/(1-\sigma)}$ for downstream producers, lowering production costs and encouraging further entry. Mediated through the Firm Market Access index, these linkages represent the IO amplification mechanism documented empirically in Section 4.2.

Agglomeration externality. We incorporate a Marshallian agglomeration externality: total factor productivity in nonagriculture increases with local firm density.

$$A_{NA,s,i} = \bar{A}_{NA,s,i} \, (n_{s,i})^{\lambda} \tag{9}$$

where $\lambda > 0$ is the agglomeration elasticity. This reduced-form specification parameterizes localization economies through three channels: labor market pooling, input sharing, and knowledge spillovers (Rosenthal and Strange, 2004; Duranton and Puga, 2004). Both IO linkages and the agglomeration externality reinforce each other in the agglomeration force, while wage adjustments create the dispersion force.

Free entry and equilibrium. We assume free entry in the nonagriculture sector. Firms enter location i until operating profits equal a fixed labor cost $w_i F$. Free entry drives per-firm profits to zero:

$$\pi_{s,i} = \tfrac{1}{\sigma} \, p_{s,i}^{1-\sigma} \, FMA_{s,i} - w_i F = 0$$

This zero-profit condition fixes per-firm revenue at the break-even level, $p_{s,i}^{1-\sigma} \, FMA_{s,i} = \sigma w_i F$, and the equilibrium number of firms then follows from total sectoral revenue $R_{s,i}$ in each region:

$$n_{s,i}^{*} = \frac{R_{s,i}}{\sigma w_i F} \tag{10}$$

A highway upgrade reduces τ_{ij}, raising $FMA_{s,i}$ and revenue above break-even level, which induces firm entry. Through the input-output channel, higher nonagricultural revenue also raises demand for local intermediates, further reducing input costs and attracting additional entry (the IO amplification loop). Endogenous wages clear local labor markets and constitute the dispersion force that bounds this loop.

Proposition 1. [13] *Suppose the agglomeration elasticity $\lambda > 0$ and IO coefficients $\mu_{rs} > 0$. Then there exists a threshold $\overline{FMA}(\lambda, \mu_{rs}, F)$ such that:*

1. *For $FMA_{s,i} < \overline{FMA}$, the dispersion force dominates, $\frac{dn^*_{s,i}}{d(FMA_{s,i})}$ is small, and the location remains in a low-activity agrarian equilibrium.*
2. *For $FMA_{s,i} > \overline{FMA}$, the combined agglomeration and IO multiplier forces dominate, $\frac{dn^*_{s,i}}{d(FMA_{s,i})}$ is large and increasing, and rapid industrialization follows.*
3. *The threshold $\overline{FMA}$ is* increasing *in μ_{rs}: stronger IO linkages raise the threshold but simultaneously amplify post-threshold agglomeration.*

Claim #3 merits economic interpretation. Before sufficient connectivity is established, stronger IO linkages raise the threshold because intermediate trade flows are thin. Once the threshold is crossed, those same IO coefficients generate larger multiplier effects: each additional firm expands the demand for intermediates proportionally more, and each reduction in the price index feeds back more strongly into entry. The result is a steeper, more discontinuous transition at the threshold for high-μ_{rs} locations. This mechanism provides one structural rationalization for explaining the core-periphery asymmetry: the RRD core, with dense pre-existing supply chains (high μ_{rs}) and market access already above $\overline{FMA}$, experiences diminishing returns, whereas the periphery, with nascent linkages, faces a higher threshold but exhibits self-reinforcing agglomeration once it is crossed.

Figure A7 illustrates this mechanism: equilibrium firm counts simulated across trade costs $\tau \in [1.0, 2.0]$ for a two-region calibration show a U-shaped (periphery) versus concave (core) relationship between firm counts and market access.

5.2 Calibration

We calibrate the model for Northern Vietnam around the highway upgrade period. Table 6 summarizes fixed parameters and data inputs.

Fixed structural parameters. We set the elasticity of substitution $\sigma = 7$, within the range of estimates in the trade literature (Anderson and van Wincoop, 2004; Head and

[13] See Appendix G for the proof.

Ries, 2001). The input-output share parameters, μ_{rs}, are calibrated using the 2007 Vietnamese Input-Output Tables published by the General Statistics Office (GSO). Specifically, we aggregate the original 138 sectors into the three sectors of the theoretical model and compute the corresponding intersectoral input shares. These calibrated shares are then cross-checked against the OECD Input-Output Table. The agricultural consumption share $\alpha_A = 0.492$ is taken from GSO household expenditure surveys.

Estimated parameters. Two parameters are estimated from the spatial distribution of manufacturing firms.[14] First, the agglomeration elasticity is estimated by simulated method of moments (SMM) such that $\hat{\lambda}$ minimizes the discrepancy between model-implied and observed firm shares across districts:

$$\hat{\lambda} = \arg \min_{\lambda \in [0.05, 0.40]} \sum_{j=1}^{J} \left(\frac{n_j^{\text{model}}(\lambda)}{\sum_k n_k^{\text{model}}(\lambda)} - \frac{n_j^{\text{data}}}{\sum_k n_k^{\text{data}}} \right)^2 \tag{11}$$

The objective is minimized at $\hat{\lambda} = 0.10$. While existing studies (Duranton and Puga, 2020; Ahlfeldt and Pietrostefani, 2019) find elasticities of productivity with respect to density around 0.04 in developed countries, our estimate aligns with recent evidence from emerging economies where agglomeration externalities are often more pronounced (Grover *et al.*, 2023; Henderson and Turner, 2020). Specifically, studies on Vietnamese manufacturing document significant productivity gains from localization, supporting an elasticity in this range (Howard *et al.*, 2014). This is consistent with the view that in rapidly industrializing economies, benefits from labor pooling and supplier networks are substantial.

Second, we calibrate the agricultural labor share β_L via a grid search over the range established by Gollin (2002), who shows that once self-employment and unpaid family labor, both prevalent in Vietnamese agriculture, are properly accounted for, the labor share in developing-country agriculture lies between 0.65 and 0.85. Minimizing the Root-Mean-Squared Error (RMSE) between model-implied and observed spatial cumulative distribution functions (CDFs) of manufacturing firms (Appendix Figure A8) yields $\hat{\beta}_L = 0.85$, the upper bound of this range. This high labor intensity is consistent with the structural transformation literature, which documents that agricultural production in low-income economies is substantially more labor-intensive than in other sectors (Herrendorf *et al.*, 2014), and reflects the reality of Vietnamese agriculture where smallholder farming and family labor remain dominant.

Trade costs and remaining inputs. Baseline trade costs τ_{ij} are constructed from GIS-based minimum-travel-time paths on the pre-upgrade road network. Initial district wages

[14] We calibrate and validate the model against manufacturing firm entry rather than nonagricultural firms more broadly because manufacturing entry is the primary object of interest in this paper, given its central role in the polycentric industrial growth we document.

w_j and labor endowments L_j are from the VES and Population Census. The size of agricultural land $K_{A,j}$ (for rice paddies and crops) at district-level is computed based on the JAXA ALOS high-resolution land-use and land-cover maps.

5.3 Model Validation

A central challenge in structural modeling is establishing that the framework captures the correct economic mechanisms rather than merely fitting the data through flexible functional forms. To address this challenge, we perform two complementary validation exercises before using the structural model for counterfactual policy analysis.

5.3.1 Horse Race Across Specifications

First, we evaluate the necessity of our model's core ingredients, IO linkages and endogenous wages, by comparing three nested specifications against the observed CDF of manufacturing firms by distance from the GMS highway network (Figure 10).

Model with fixed wages (no dispersion). Shutting down wage adjustment isolates the combined effect of agglomeration and IO amplification. The resulting CDF exhibits a "winner-takes-all" pattern: nearly 100% of firms concentrate within 10 km of the highway, far exceeding the 75% observed. Without the dispersion force of rising labor costs, the IO-driven amplification runs unchecked, leading to excessive spatial concentration. In an economy where structural transformation is ongoing, the countervailing pressure through labor reallocation from agriculture is a critical determinant of the spatial equilibrium.

Model without IO linkages (endogenous wages, no amplification). When intermediate goods trades ($\mu_{rs} \equiv 0$) are removed, market access improvement feeds directly into firm profit, with no multiplicative demand-linkage effect to sustain clustering. As a result, the CDF substantially undershoots: only 60% of manufacturing firms are predicted within 25 km of the corridor versus 87% observed. This illustrates why trade models that abstract from supply chains systematically underpredict agglomeration in settings where intermediate goods linkage is salient (Bartelme and Gorodnichenko, 2015).

Preferred model (IO linkages + endogenous wages). The full specification with $\hat{\lambda} = 0.10$ closely matches the smooth shape of the CDF over space, strong near-corridor clustering without complete concentration. IO amplification generates the pronounced corridor clustering, while endogenous wages flatten the distribution beyond 10 km by raising costs for firms that might otherwise concentrate further along the corridor.

The horse race suggests that the combination of agglomeration externalities, IO amplification, and dispersion forces is quantitatively essential to replicate the observed polycentric pattern.

5.3.2 IV-based Goodness-of-Fit Test

Second, we apply the validation test of Adão *et al.* (2024) to assess whether the model correctly specifies the causal mechanism to predict the spatial developments induced by the highway upgrades.

Inference. We decompose the observed changes in district-level firm density $\Delta \ln n_j^{\text{data}}$ into the model-implied response to the highway upgrade and a residual term:

$$\underbrace{\Delta \ln n_j^{\text{data}}}_{\text{Total change}} = \underbrace{\Delta \ln n_j^{\text{model}}(\tau_{\text{pre}}, \tau_{\text{post}})}_{\text{Trade cost shock}} + \underbrace{u_j}_{\text{Residual local shocks}} \tag{12}$$

where $\Delta \ln n_j^{\text{model}}$ is the model-implied general equilibrium response to the trade cost change $(\tau_{\text{pre}} \to \tau_{\text{post}})$, holding other structural primitives fixed. u_j absorbs all other local shocks (e.g., local demand and productivity shocks, industrial policies, FDIs) which are unrelated to the highway.

Test statistic. Let $y_j \equiv \Delta \ln n_j^{\text{data}}$ and $x_j(\hat{\lambda}) \equiv \Delta \ln n_j^{\text{model}}$. Adao-Costinot-Donaldson (ACD)'s goodness-of-fit test statistic evaluates whether the model's prediction error is orthogonal to the exogenous MST-based IV ($\tilde{z}_j$) used in our reduced-form analysis:

$$\hat{\beta}^z = \frac{\sum_j \tilde{z}_j (y_j - x_j(\hat{\lambda}))}{\sum_j \tilde{z}_j \, x_j(\hat{\lambda})} \tag{13}$$

where $\tilde{z}_j = z_j - \bar{z}$ is the demeaned MST-based shift-share instrument. The null hypothesis $H_0 : \beta^z = 0$ is equivalent to $\hat{\beta}^{\text{IV}} = 1$: the 2SLS slope between observed and predicted changes being unity, the prediction of a correctly specified causal model.[15] We report two specifications following Adão *et al.* (2024): a naive IV that ignores estimation uncertainty in $\hat{\lambda}$, and a conservative IV that corrects for the estimation error in $\hat{\lambda}$.

Results. As shown in Table 7, neither IV specification rejects H_0. Under standard errors that account for estimation uncertainty in $\hat{\lambda}$ (Panel B), both the naive and preferred IV estimates remain insignificant: $\hat{\beta}^z_{\text{naive}} = 0.030$ (p-value $= 0.284$) and $\hat{\beta}^z_{\text{pref}} = 0.030$ (p-value $= 0.514$). A first-stage F-statistic of 14.12 confirms instrument relevance. The positive

[15]Therefore, non-rejection of the null hypothesis implies that the model generates the correct direction and magnitude of agglomeration response for each unit of policy-induced market access change.

correlation between model-implied and observed changes ($\rho(y_j, x_j(\hat{\lambda})) = 0.155$) provides directional confirmation.[16] This non-rejection under both IV specifications indicates that our structural framework generates an agglomeration response that aligns with the causal variation identified in our empirical strategy.

Taken together, by passing both the horse race and ACD causal test, the preferred model, with IO amplification and endogenous wage adjustments, demonstrates its ability to map highway-induced market access improvements to spatial agglomeration outcomes.

6 Counterfactual Policy Analysis

Having established the validity of the model, we use it to evaluate three counterfactual scenarios addressing two policy questions. The first, *does the spatial design of infrastructure matter?*, is examined through budget-neutral reallocation of the GMS investment envelope (CF1 and CF2). The second, *how does infrastructure interact with industrial structure?*, is examined through a complementary industrial policy scenario (CF3). These counterfactuals exploit the validated model structure to quantify general equilibrium welfare effects of infrastructure design choices and their interaction with local production networks.

6.1 Design of Counterfactuals

CF1: GMS Downgrade + Trunk Highway Reallocation. We partially downgrade the GMS corridor (target speed from 100 to 80 km/h) and reallocate the saved budget, equated to the engineering cost premium of the GMS speed upgrade, to upgrading the national highway and major roads, which provides extended connectivity to the NE and NW periphery (Figure 11). This budget-neutral exercise compares the marginal value of concentrating transport capacity on the main GMS corridor with an alternative design that redistributes trunk connectivity toward a broader set of peripheral regions. It parallels the logic of Morten and Oliveira (2024), who evaluate Brazil's actual Brasilia-centered highway network against a hypothetical Rio-centered network. However, our counterfactual reallocates the marginal engineering budget from a corridor-focused upgrade toward a more spatially dispersed trunk-road improvement, rather than relocating the spatial center of a national highway system.

[16] The correlation of 0.155 is a lower bound on the model's explanatory power. By construction, the residual u_j absorbs all local shocks unrelated to the highway, so the highway-attributable share of spatial variation is only a part of total observed variation.

CF2: GMS Downgrade + Trunk Highway and Feeder Road Reallocation. This scenario similarly downgrades the GMS corridor and upgrades local road networks. However, we split the saved budget 50:50 between upgrading national highways/major roads and upgrading feeder roads (secondary and tertiary roads connecting rural communes to the national network; Figure 11). Feeder roads provide *local market access*, the last-mile connectivity that allows inland districts to participate in the supply chains along the GMS corridor (Asher and Novosad, 2020). In our setting, feeder roads allow local intermediate producers to participate in the supply chains activated by the GMS corridor, generating multiplier gains that spread broadly through the regional production system.

CF3: Manufacturing Technology Upgrading. Holding the transport network fixed, we simulate a deepened IO matrix that strengthens manufacturing value-added by (i) raising manufacturing-to-manufacturing IO shares, capturing deeper domestic supply chains, and (ii) reducing reliance on imported intermediates.[17] This counterfactual captures industrial policies complementary to infrastructure that deepen local production networks, such as industrial zone development, technology transfer programs, and local supplier upgrading. By holding trade costs fixed, CF3 isolates the IO amplification channel from the market access channel, providing a comparison with CF1 and CF2.

6.2 Endogenous Trade Costs and Welfare Evaluation

These counterfactuals fundamentally alter the configuration of the road network by changing relative travel speed and time across highways and feeder roads. This prompts firms and transporters to re-optimize their routes. Following the route-choice framework (Fan *et al.*, 2023), trade costs under CF1 and CF2 are endogenized based on the optimal path between any two districts:

$$\tau_{ij} = \exp(\kappa \cdot d_{ij}(\mathcal{G})) \tag{14}$$

where $d_{ij}(\mathcal{G})$ is minimum travel time across the counterfactual network graph $\mathcal{G}$. The elasticity of trade costs with respect to travel time, κ, governs the spatial sensitivity of economic activity in the counterfactual exercise. We select $\hat{\kappa} = 0.101$ based on empirical consistency, spatial data fit, and numerical stability. First, this value aligns with the range of 0.05–0.15 found by Donaldson (2018) and is close to the recent estimate of 0.134 by Elhan-Kayalar *et al.* (2025), who study the GMS highway expansion in neighboring Yunnan, China. Our choice is consistent with the evidence that developing economies face substantially higher internal trade barriers than developed markets (Atkin and Donaldson, 2015; Sotelo, 2020). Second, $\hat{\kappa}$ is identified through a two-stage grid search (fine

[17] CF3 changes only manufacturing's input-sourcing shares, raising its reliance on domestically produced manufacturing inputs while reducing its reliance on agricultural and service inputs. Agriculture's and services' own input-sourcing structures are held fixed.

step size of 0.001) that minimizes the Gaussian-weighted RMSE between observed and model-implied spatial CDFs of manufacturing firms (see Appendix H for details). Third, we employ a continuation method to trace the equilibrium path, selecting a baseline value within a stability zone where counterfactual responses are numerically continuous.

We evaluate the welfare consequences of these scenarios using two key metrics. First, we measure aggregate welfare as the total real income in the economy, defined as $W = \sum_j (\frac{w_j L_j}{P_j})$, which captures the overall efficiency gains from the policy. Second, to assess the distributional impact, we calculate spatial inequality using the population-weighted Gini coefficient of district-level real income. All counterfactuals are evaluated relative to the baseline model scenario, normalized to 100. These metrics allow us to quantify the trade-off between maximizing aggregate growth and reducing regional disparities—a central tension in the design of large-scale infrastructure projects.

6.3 Results

Table 8 reports welfare outcomes across counterfactual scenarios, and Figure 12 compares the corresponding spatial CDFs of manufacturing firms relative to the baseline.

Infrastructure design: corridor vs. dispersed investment (CF1 and CF2). Both infrastructure reallocation scenarios yield higher aggregate real income than the baseline (CF1: 137.19%; CF2: 116.29%), confirming that dispersing road investment activates IO amplification across a broader spatial base. As more districts' market access crosses the threshold of Proposition 1, the production network mobilizes a larger share of the economy and contributes to stronger industrial growth and welfare improvement.

Under CF1 and CF2 in Figure 12, the cumulative share of firms within 5–10 km of the GMS corridor falls sharply compared with the baseline, indicating that budget reallocation toward national highways or feeder roads relocates manufacturing activity away from the corridor and toward newly connected interior nodes.

The comparison between CF1 and CF2 highlights the critical role of *local market access*. CF1 increases the Gini index by 16% because reallocating funds solely to national highways benefits already-connected nodes, leaving behind peripheral districts that lack last-mile feeder road access. In contrast, CF2 reduces the Gini index to 93% of the baseline level. By investing half of the saved budget in feeder roads, CF2 provides the last-mile connectivity, enabling inland districts to be integrated into supply-chain linkages.

Infrastructure vs. industrial policy (CF3). The manufacturing upgrading scenario delivers the largest aggregate income gain (159.8% of the baseline). Deepening IO linkages and reducing reliance on imported intermediates amplifies agglomeration economies where supply chains are already dense. Crucially, Figure 12 reveals that the spatial firm distribution of CF3 moves in the opposite direction from the reallocation scenarios: the cumulative share of firms within 5 km of the GMS corridor rises to approximately 72%, well above the baseline. IO deepening pulls firms toward the corridor rather than dispersing them, because the densest supply-chain hubs reap the largest gains from richer intermediate-input linkages. This spatial concentration explains why CF3 simultaneously achieves the highest efficiency and an increase in the Gini (107.69%), because intensified agglomeration at corridor-adjacent locations outpaces the periphery.

The comparison between CF2 and CF3 reveals the policy complementarity formalized in Proposition 1. While IO deepening (CF3) raises the *post-threshold* growth potential of the production network, feeder road investment (CF2) effectively pushes peripheral districts beyond the agglomeration threshold by lowering market access costs. Therefore, a welfare-maximizing strategy involves a coordinated investment in trunk connectivity, feeder road extension, and IO deepening to amplify the welfare gains.[18]

6.4 Implications for Policy Design

Our counterfactual analysis provides three insights into the spatial design of infrastructure and industrial policy. First, our findings suggest that the GMS corridor program, while transformative, does not lie on the efficiency-equity frontier. CF2 shows that a budget-neutral reallocation toward feeder road connectivity can improve both aggregate real income and spatial equity relative to the actual program. Second, the threshold nonlinearity in industrialization—whereby certain connectivity gains are required to trigger self-sustaining agglomeration—suggests a role for the strategic sequencing of investments. Reallocating resources to bridge the *last-mile* gap may allow peripheral districts to cross the agglomeration threshold, potentially generating higher marginal social returns than incremental upgrades to already-adequate trunk infrastructure. Finally, we document a complementarity between infrastructure and industrial policy. This strategic complementarity suggests that coordinating transport investments with supplier development programs is critical for fostering inclusive and sustained regional growth.

7 Conclusion

This paper has examined the spatial transformation of Northern Vietnam from a battlefield into a dynamic marketplace as a two-stage process, using the GMS highway upgrades as a quasi-experiment in large-scale transport infrastructure investment. Satellite

[18]The large welfare gains in CF3 should be interpreted with caution, as the scenario abstracts from the political economy constraints and implementation costs of industrial policy. We interpret the estimate as an upper-bound illustration of the complementarity between infrastructure and industrial policy.

imagery documents a clear evolution from a monocentric urban structure centered on Hanoi after the Vietnam War to a polycentric configuration extending into the hinterland. Instrumental variable and difference-in-differences estimates show that this shift was driven by the emergence of new manufacturing clusters in previously agrarian peripheries, not by simple reinforcement of the industrial core.

Our central empirical finding is that polycentric development reflects nonlinear market access responses amplified by supply-chain linkages. IV estimates reveal that firm entry accelerates as FMA grows in the periphery. DID estimates show that the spatial extent of these effects scales with supply-chain depth and that second-tier cities mediate diffusion through production networks.

The quantitative model rationalizes the observed polycentric development pattern through a threshold-type mechanism arising from the interaction of endogenous labor reallocation, agglomeration externalities, and input-output linkages. While IO linkages foster corridor-led clustering, endogenous wage responses appear to mitigate "winner-takes-all" concentration, allowing for the emergence of secondary centers. The threshold mechanism provides a framework for understanding the conditions under which a *spatial big push* can occur. While large-scale trunk infrastructure may be necessary to generate sufficiently large connectivity gains, its impact on inclusive growth potentially depends on complementary local market access and the depth of regional production networks. These findings point to the potential importance of coordinating infrastructure design with supplier development programs to sustain growth across a broader range of locations.

Future research could further refine these insights in several directions. First, while our static model captures the equilibrium shift, exploring the dynamic adjustment paths and the long-run stability of newly emerged clusters would be valuable. Second, our finding of strategic complementarity between different types of infrastructure suggests that the synergistic effects between road transport networks and other trade-facilitating assets—such as ports—merit further investigation (Fuchs and Wong, 2026; Elhan-Kayalar *et al.*, 2024). Third, evaluating the interaction between place-based policies (e.g., industrial zones) and production networks could clarify how targeted interventions complement broad-based market access improvements (Kono *et al.*, 2025). The Vietnamese experience provides a compelling illustration of how strategic infrastructure, when combined with local production networks and complementary market access improvements, can reshape the economic geography of post-conflict nations.

References

ADÃO, R., COSTINOT, A. and DONALDSON, D. (2024). Putting quantitative models to the test: An application to the U.S.-China trade war. *The Quarterly Journal of Economics*, **140** (2), 1471–1524.

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.

— and PIETROSTEFANI, E. (2019). The economic effects of density: A synthesis. *Journal of Urban Economics*, **111**, 93–107.

AKAMATSU, T., MORI, T., OSAWA, M. and TAKAYAMA, Y. (2025). *Spatial scale of agglomeration and dispersion: Number, spacing, and the spatial extent of cities*. Report, arXiv.org, papers.

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

— and — (2022). The welfare effects of transportation infrastructure improvements. *Review of Economic Studies*, **89** (6), 2911–2957.

ANDERSON, J. E. and VAN WINCOOP, E. (2004). Trade costs. *Journal of Economic Literature*, **42** (3), 691–751.

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.

ASHER, S. and NOVOSAD, P. (2020). Rural roads and local economic development. *American Economic Review*, **110** (3), 797–823.

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

ATKIN, D. and DONALDSON, D. (2015). Who's getting globalized?: The size and implications of intra-national trade costs. *NBER Working Paper No. 21439*.

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.

BARTELME, D. and GORODNICHENKO, Y. (2015). Linkages and economic development. *NBER Working Paper No. 21251*.

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. (2024). Difference-in-differences with a continuous treatment. *NBER Working Paper No. 32117*.

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

DAVIS, D. R. and WEINSTEIN, D. E. (2002). Bones, bombs, and break points: The geography of economic activity. *The American Economic Review*, **92** (5), 1269–1289.

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

DONALDSON, D. (2018). Railroads of the Raj: Estimating the impact of transportation infrastructure. *American Economic Review*, **108** (4-5), 899–934.

— 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 MORI, T. (2025). How Masa Fujita shaped the present of spatial economics and how he will inspire its future. *The Japanese Economic Review*.

— and PUGA, D. (2004). *Chapter 48 - Micro-Foundations of Urban Agglomeration Economies*, Elsevier, vol. 4, p. 2063–2117.

— and — (2020). The economics of urban density. *Journal of Economic Perspectives*, **34** (3), 3–26.

— 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. and SHANGGUAN, R. (2025). Connectivity gains from regional integration in the Greater Mekong Subregion. *Asian Economic Policy Review, forthcoming*.

—, —, —, —, — and TUNG, N. T. (2024). Foreign Direct Investment location and trade dynamics in Viet Nam after the US–PRC trade war. *ADB Economics Working Paper No. 761*.

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. (2022). Trade, structural transformation, and development: Evidence from Argentina 1869–1914. *Journal of Political Economy*, **130** (5), 1249–1318.

FAN, J., LU, Y. and LUO, W. (2023). Valuing domestic transport infrastructure: A view from the route choice of exporters. *Review of Economics and Statistics*, **105** (6), 1562–1579.

FUCHS, S. and WONG, W. F. (2026). Multimodal transport networks. *NBER Working Paper No. 35065*.

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.

GLAESER, E. L. (2022). Urban resilience. *Urban Studies*, **59** (1), 3–35.

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

GOLLIN, D. (2002). Getting income shares right. *Journal of Political Economy*, **110** (2), 458–474.

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.

GROVER, A., LALL, S. and TIMMIS, J. (2023). Agglomeration economies in developing countries: A meta-analysis. *Regional Science and Urban Economics*, **101**, 103901.

HEAD, K. and RIES, J. (2001). Increasing returns versus national product differentiation as an explanation for the pattern of U.S.-Canada trade. *The American Economic Review*, **91** (4), 858–876.

HENDERSON, J. V. and TURNER, M. A. (2020). Urbanization in the developing world: Too early or too slow? *Journal of Economic Perspectives*, **34** (3), 150–173.

HERRENDORF, B., ROGERSON, R. and VALENTINYI, (2014). *Chapter 6 - Growth and Structural Transformation*, Elsevier, vol. 2, p. 855–941.

HOWARD, E., NEWMAN, C., RAND, J. and TARP, F. (2014). Productivity-enhancing manufacturing clusters: Evidence from Vietnam. *UNU-WIDER Working Paper No. 14*.

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., LE, H.-M., NOSE, M. and SAWADA, Y. (2025). Industrial zones and production networks: Evidence from Vietnam, DP No. E-25-005, Kyoto University.

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

— (1992). Toward a counter-counterrevolution in development theory. *The World Bank Economic Review*, **6** (suppl_1), 15–38.

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

LIN, J. and RAUCH, F. (2022). What future for history dependence in spatial economics? *Regional Science and Urban Economics*, **94**, 103628.

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.

MATSUYAMA, K. (1995). Complementarities and cumulative processes in models of monopolistic competition. *Journal of Economic Literature*, **33** (2), 701–729.

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.

MURPHY, K. M., SHLEIFER, A. and VISHNY, R. W. (1989). Industrialization and the big push. *Journal of Political Economy*, **97** (5), 1003–1026.

NAGY, D. K. (2023). Hinterlands, city formation and growth: Evidence from the U.S. westward expansion. *The Review of Economic Studies*, **90** (6), 3238–3281.

NEUBERT, M. and NIKOLIĆ, S. (2026). Why railways fail: Colonial railways and economic development in Habsburg Bosnia-Herzegovina. *Journal of Development Economics*, p. 103747.

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.

ROSENTHAL, S. S. and STRANGE, W. C. (2004). *Chapter 49 - Evidence on the Nature and Sources of Agglomeration Economies*, Elsevier, vol. 4, p. 2119–2171.

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

SOTELO, S. (2020). Domestic trade frictions and agriculture. *Journal of Political Economy*, **128** (7), 2690–2738.

TAKEDA, K. and YAMAGISHI, A. (2026). The economic dynamics of city structure: Evidence from Hiroshima's recovery. *Forthcoming, Journal of Political Economy*.

TAN, D. (2014). The Greater Mekong Subregion programme: reflections for a renewed paradigm of regionalism. *Asia Europe Journal*, **12** (4), 383–399.

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

TSIVANIDIS, N. (2026). Evaluating the impact of urban transit infrastructure: Evidence from Bogotá's transmilenio. *American Economic Review*, **116** (2), 418–463.

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

TABLE 1: Descriptive Statistics

(a) District Population, 1999–2019

	Number of districts	District Population (average)			Population Growth (%)	
		1999	2009	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) Number of Enterprises, 2011–2020

	All Firms			Manufacturing			Services			Enterprise Growth (%)	
	2011	2016	2020	2011	2016	2020	2011	2016	2020	2011-2016	2016-2020
Red River Delta	104,827	161,876	258,018	15,917	24,078	38,830	73,020	114,677	181,378	49.1	80.7
North East	10,294	13,875	24,729	1,783	2,434	4,891	5,382	7,889	13,608	34.3	92.8
North West	5,518	7,872	12,376	736	904	1,489	2,706	4,255	6,580	76.4	66.3
Total (National)	329,121	516,803	822,618	52,596	76,312	121,948	220,581	361,047	566,038	55.3	77.2

NOTE: Panel (a) reports regional averages across districts for district population and district-level population growth over 1999-2009 and 2009-2019. Panel (b) reports each region's total number of registered enterprises in 2011, 2016, and 2020, separately for all firms, manufacturing firms, and service firms. For each region, average enterprise growth across districts over 2011-2016 and 2016-2020 is also reported. Population data are from the 1999, 2009, and 2019 Population Censuses, while enterprise data are from the 2011, 2016, and 2020 Enterprise Censuses. Total (National) row covers all eight regions in Vietnam.

TABLE 2: Determinants of Firm Market Access Growth, 2011–2020 (First Stage)

	(1) Δ Log FMA	(2) Δ Log FMA
Log distance to MST	-0.106***	-0.054***
	(0.016)	(0.018)
Land elevation (z-score)	0.125	-0.796***
	(0.233)	(0.094)
Sunshine (z-score)	-0.223***	0.570***
	(0.056)	(0.078)
U.S. bomb, missiles and rockets (log+1)	-0.446***	0.012
	(0.024)	(0.035)
Observations	2,387	1,437
Region FE	YES	YES
Sample	RRD region	NE & NW regions

NOTE: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors clustered at the commune level.

TABLE 3: Firm Market Access Growth and Industrial Clustering (Second Stage)

	Panel A. RRD regions								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Manufacturing			Service: transport			Service: trade		
Δ Log FMA	0.046***	-0.016	0.111***	0.039***	0.028	0.049***	0.047***	0.040	0.072***
	(0.006)	(0.044)	(0.020)	(0.005)	(0.040)	(0.017)	(0.006)	(0.044)	(0.019)
Δ Log FMA squared			0.023***			0.011			0.012
			(0.008)			(0.007)			(0.008)
Industrial zone ($\leq 5km$) in 2011	0.139***	0.139***	0.146***	0.213***	0.213***	0.217***	0.171***	0.171***	0.175***
	(0.038)	(0.039)	(0.039)	(0.038)	(0.038)	(0.038)	(0.037)	(0.037)	(0.037)
Observations	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		33.52	54.02		33.52	54.02		33.52	54.02
Hansen J-statistics (p-value)		0.162	0.603		0.022	0.120		0.039	0.212
	Panel B. NE & NW regions								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Manufacturing			Service: transport			Service: trade		
Δ Log FMA	0.024***	0.140***	0.260***	0.006	0.021	0.151***	0.013**	0.067***	0.179***
	(0.006)	(0.029)	(0.067)	(0.005)	(0.013)	(0.041)	(0.006)	(0.021)	(0.063)
Δ Log FMA squared			0.015*			0.016***			0.014*
			(0.009)			(0.005)			(0.008)
Industrial zone ($\leq 5km$) in 2011	0.306***	0.235**	0.218**	0.163***	0.142**	0.123*	0.174**	0.108	0.091
	(0.085)	(0.092)	(0.093)	(0.063)	(0.064)	(0.064)	(0.087)	(0.088)	(0.090)
Observations	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		40.760	45.531		40.760	45.531		40.760	45.531
Hansen J-statistics (p-value)		0.250	0.363		0.282	0.167		0.286	0.212

NOTE: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors clustered at the commune level. First-difference regressions take differences from 2011 to 2020. Distance to MST network, combined with land elevation, sunshine, and U.S. bomb intensity, are used as IVs in columns 2, 3, 5, 6, 8 and 9. Controls include initial employment, share of formal employment, female workers, and distance to national roads.

TABLE 4: Difference-in-Differences: 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						
After 10 years (Within 20km)		0.347***		0.304***		0.607
After 10 years (20-50km)		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						
After 10 years (Within 20km)		0.465***		0.257***		0.324***
After 10 years (20-50km)		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. The combined effects, $\beta_1 + 10\beta_3$ for communes within 20km and $\beta_2 + 10\beta_4$ for communes in the 20-50km zone, estimate the cumulative treatment effect of the GMS highway over a 10-year horizon.

TABLE 5: Difference-in-Differences: 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						
After 10 years (Within 20km)		0.625***		0.460***		0.834***
After 10 years (20-50km)		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						
After 10 years (Within 20km)		0.776***		0.381***		0.381***
After 10 years (20-50km)		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. The combined effects, $\beta_1 + 10\beta_3$ for communes within 20km and $\beta_2 + 10\beta_4$ for communes in the 20-50km zone, estimate the cumulative treatment effect over a 10-year horizon.

TABLE 6: Fixed Parameters and Data Inputs

Description	**Parameter**	**Value**	**Source/Description**
A. Structural Parameters			
Elasticity of substitution	σ	7.0	Anderson and van Wincoop (2004)
Agglomeration elasticity	λ	0.10	Grover *et al.* (2023), Howard *et al.* (2014)
Agricultural consumption share	α_A	0.492	GSO
Agricultural labor share	β_L	0.85	Gollin (2002)
Fixed entry cost	F	35.0	Normalized units
B. Data Inputs (Baseline for Calibration)			
Trade costs	τ_{ij}	GIS analysis	Travel time in pre/post road networks
Initial district wages	w_j	Enterprise Survey	Average wage in district j
Labor endowment	L_j	Population Census	Workforce in district j
Agricultural land	$K_{A,j}$	JAXA ALOS	Farmland size (Ha) in district j
Input-Output (IO) share	μ_{rs}	GSO IO Tables	Aggregated to 3-sector structure

NOTE: IO shares are aggregated to three sectors (Agriculture, Manufacturing, Services). λ and β_L are in the range of values estimated in the literature. Within the range, both parameters are estimated from the spatial distribution of manufacturing firms (Section 5.2).

TABLE 7: IV-based Goodness-of-Fit Test: Impact of the GMS Highway Upgrade on Firm Entry

Goodness-of-fit measure	(1) Correlation $\rho(y_j, x_j(\hat{\lambda}))$	(2) Naive IV $\hat{\beta}^z_{\text{naive}}$	(3) Preferred IV $\hat{\beta}^z_{\text{pref}}$
Point estimate	0.155	0.032	0.019
Panel A. Inference ignoring estimation uncertainty			
Std. err.		0.030	0.037
p-value (H_0: $\beta^z = 0$)		0.292	0.435
Panel B. Inference accounting for estimation uncertainty			
Std. err.		0.030	0.030
p-value (H_0: $\beta^z = 0$)		0.284	0.514
First-stage F (preferred IV)		—	14.12

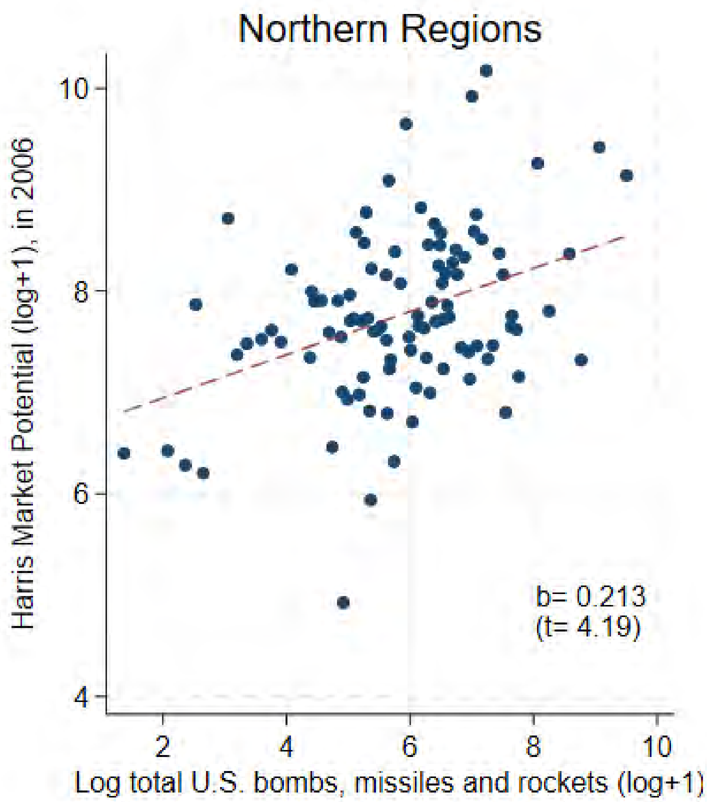
NOTE: $N = 255$ districts. $y_j = \Delta \ln n_j^{\text{data}}$; $x_j = \Delta \ln n_j^{\text{model}}$ under calibrated $\hat{\lambda} = 0.10$. All instruments are demeaned prior to estimation. Inference in column (3) accounts for estimation uncertainty following Adão *et al.* (2024).

TABLE 8: Welfare Impact of Counterfactual Scenarios

Scenario	**Total Real Income** (Baseline = 100)	**Gini Index** (Baseline = 100)
Model (Baseline)	100.00	100.00
CF1: GMS Downgrade + Highway only	137.19	115.79
CF2: GMS Downgrade + Highway & Feeder roads	116.29	93.36
CF3: MFG Upgrading	159.80	107.69

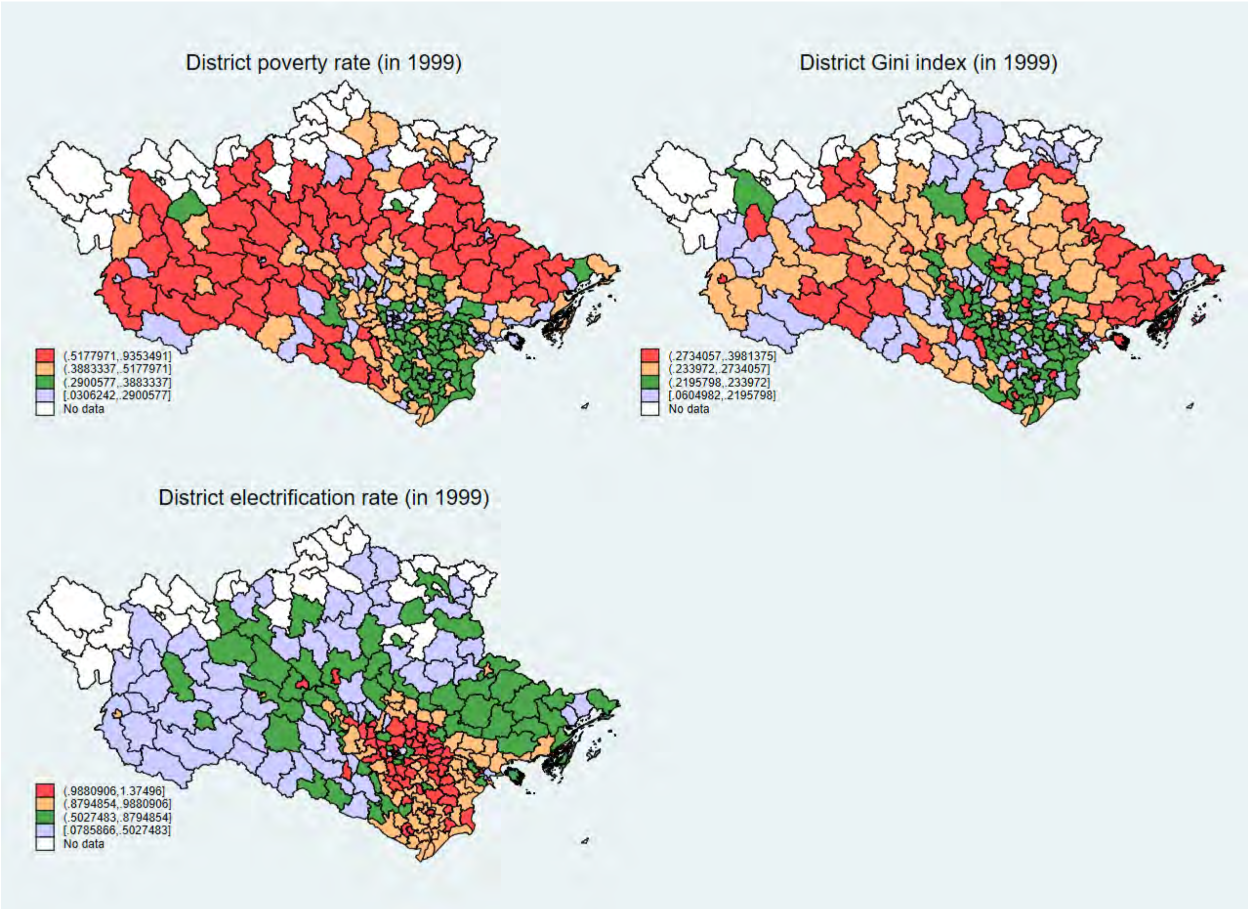
NOTE: Total Real Income and Gini Index as percentages of Model (Baseline) value (normalized to 100). Gini Index is computed over $\frac{w_j L_j}{P_j}$ across all Northern Vietnamese districts, weighted by L_j.

FIGURE 1: Legacy of U.S. Bombing and Initial Market Condition



NOTE: Conditional correlation with provincial-level fixed effects. Each dot is the bin average of the Harris market potential index in 2006. Bins defined by 50 equally-sized groups of log total U.S. bombs, missiles, and rockets from Miguel and Roland (2011).

FIGURE 2: Baseline Socio-Economic Conditions in 1999



SOURCE: Miguel and Roland (2011).

FIGURE 3: GMS Highway and National Highways to Peripheries

NOTE: QL denotes "Quốc Lộ", the Vietnamese term for national highway. QL3 and QL6 correspond to National Highway No. 3 and No. 6. The section of QL6—Dien Bien Phu–Son La, Son La–Hoa Binh, and Hoa Binh–Hanoi—is designated as part of Asian Highway 13 (AH13) corridor within Vietnam.

FIGURE 4: Sample: Core vs. Periphery

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

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

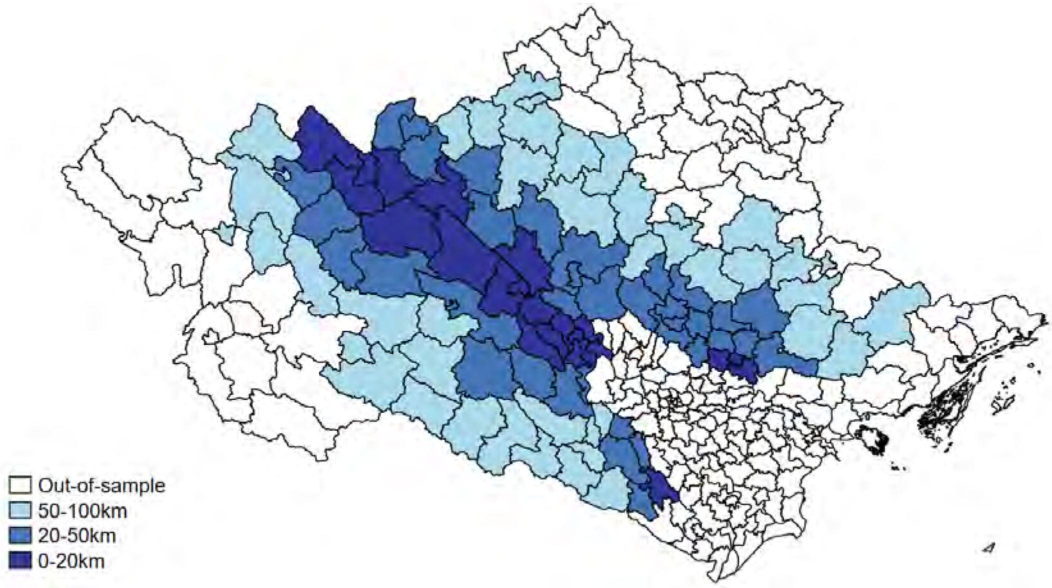
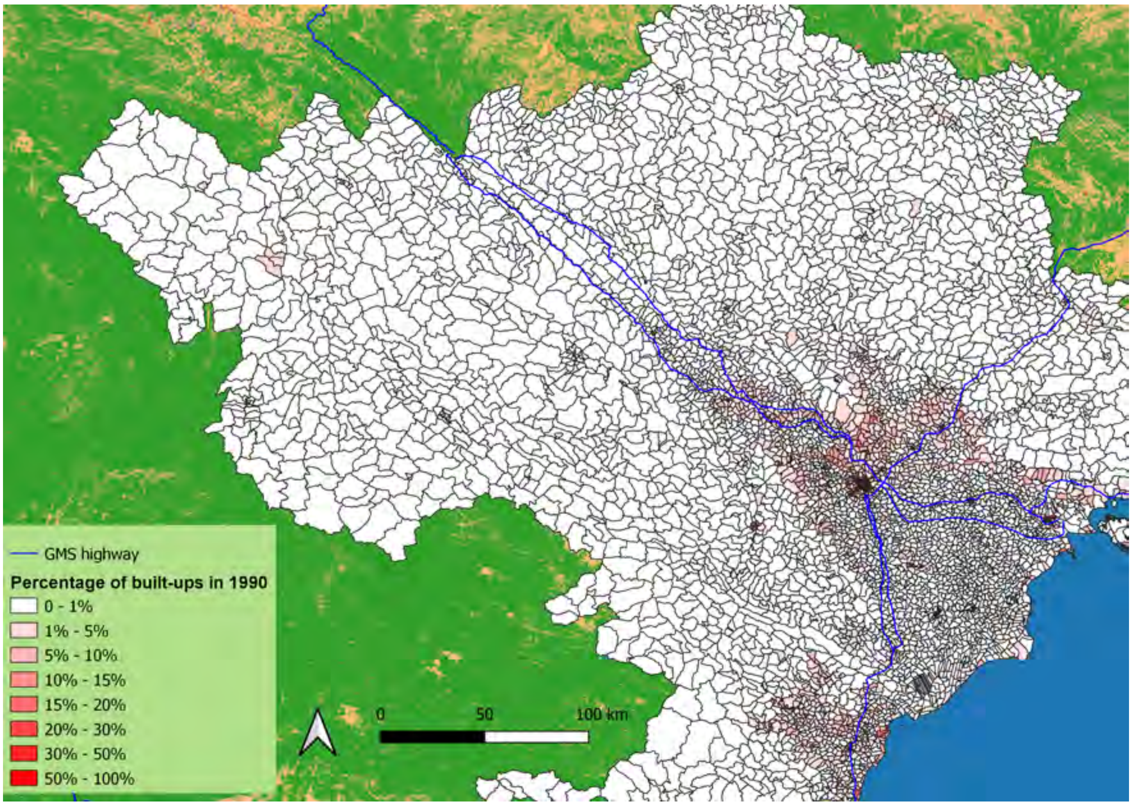


FIGURE 5: Percentage of Built-ups within Communes: 1990, 2000, 2010, and 2019

(a) In 1990



(b) In 2000

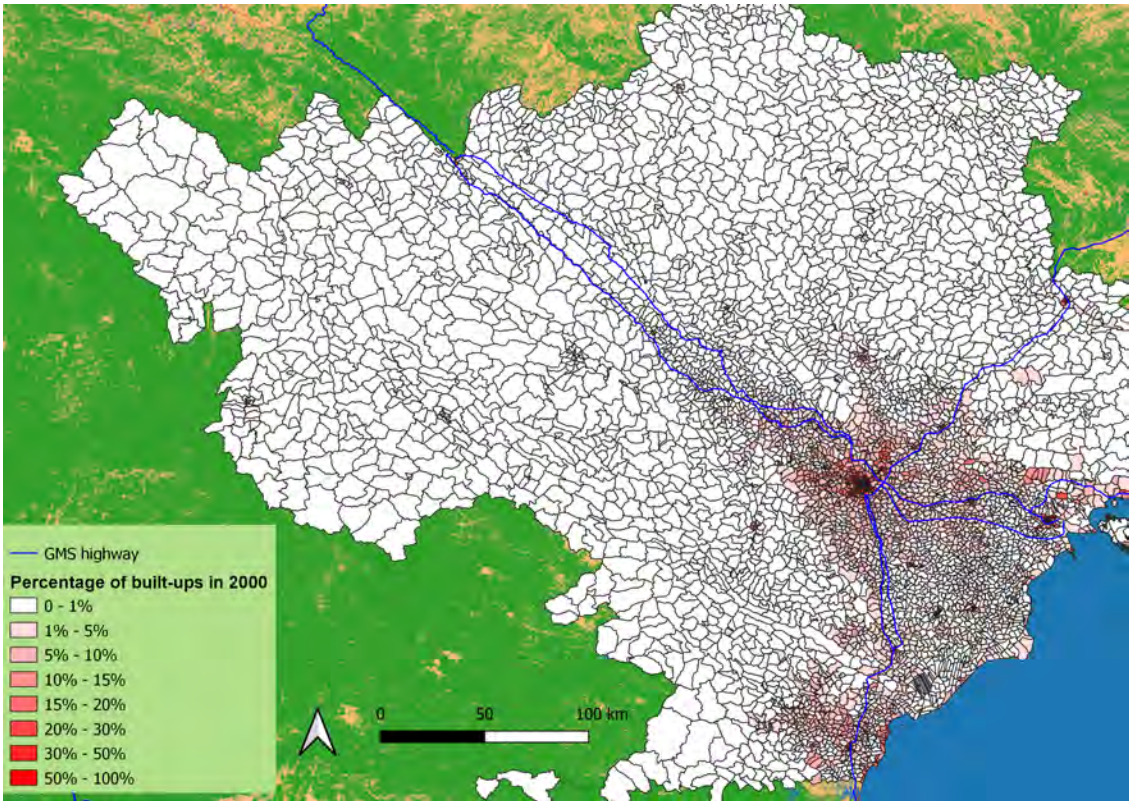
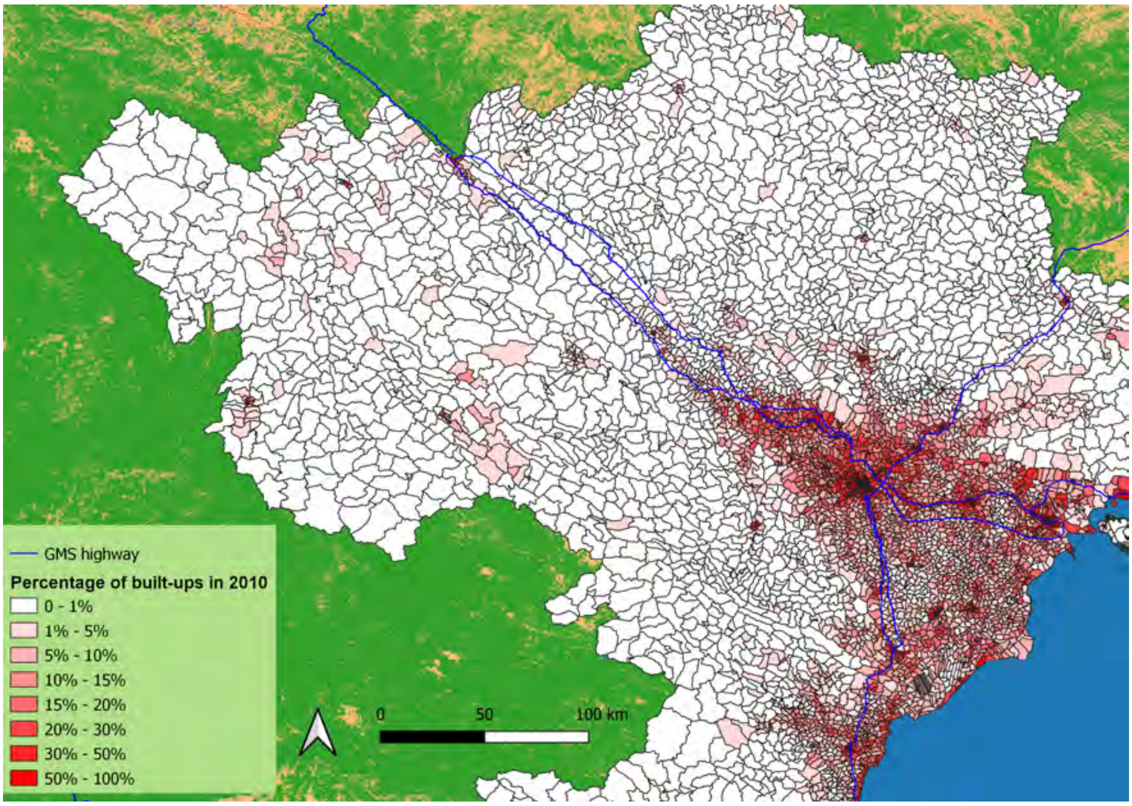
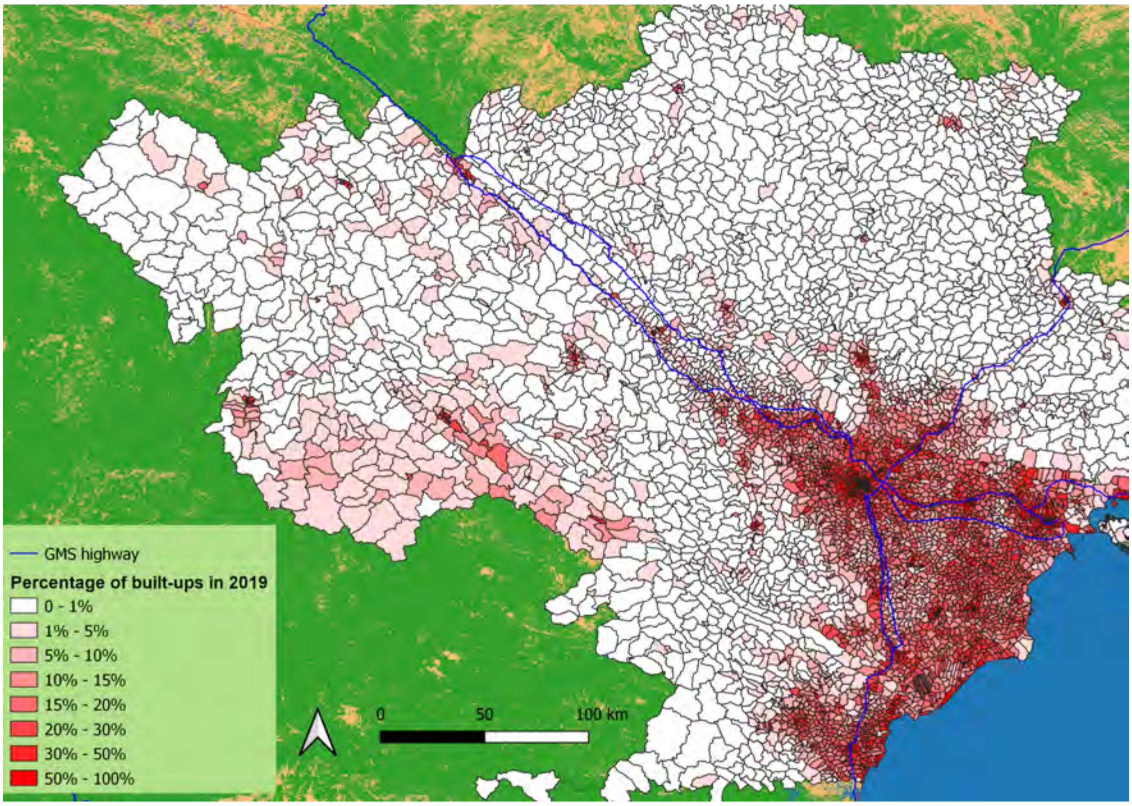


FIGURE 5: Percentage of Built-ups within Communes (cont.)

(c) In 2010



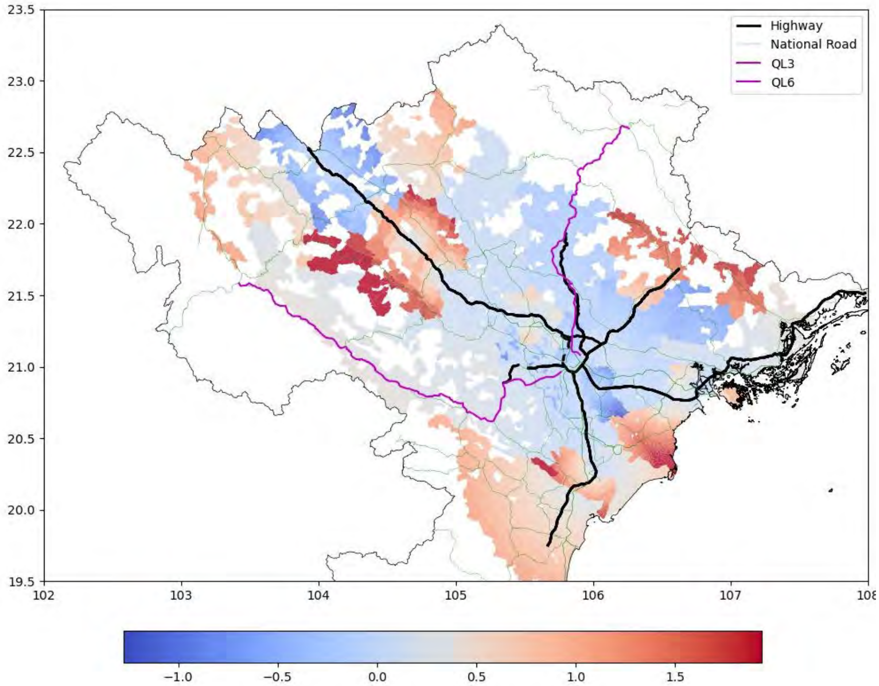
(d) In 2019



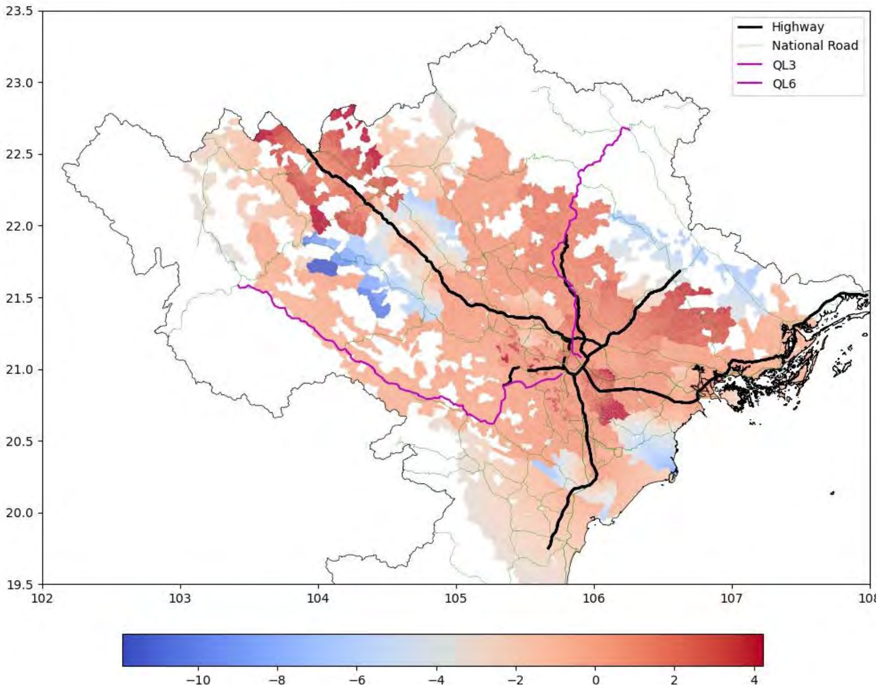
SOURCE: NASA Landsat 5 and 8.
NOTE: The proportion of built-up area is computed for each commune. Supervised classification using the neural network algorithm is applied to training data from Landsat 5 and 8 images.

FIGURE 6: Spatial Distribution of Trade Costs and Firm Market Access

(a) Change in Interregional Trade Costs, 2011–2020



(b) Improvement in Firm Market Access, 2011–2020



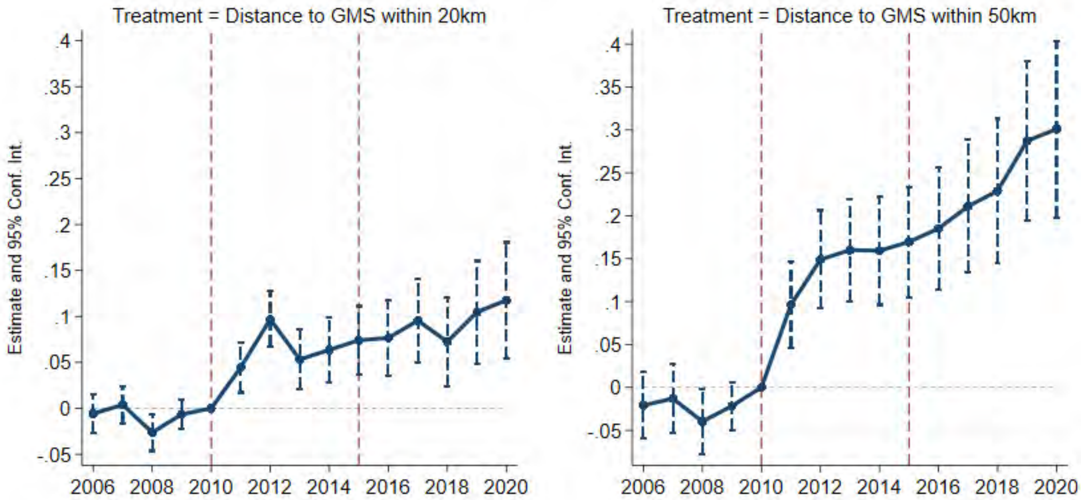
SOURCE: VES 2011, 2020. GSO.
NOTE: Trade cost changes computed from real freight price changes in the VES. Firm Market Access is computed by equation (1).

FIGURE 7: Maps of Minimum Spanning Tree Network

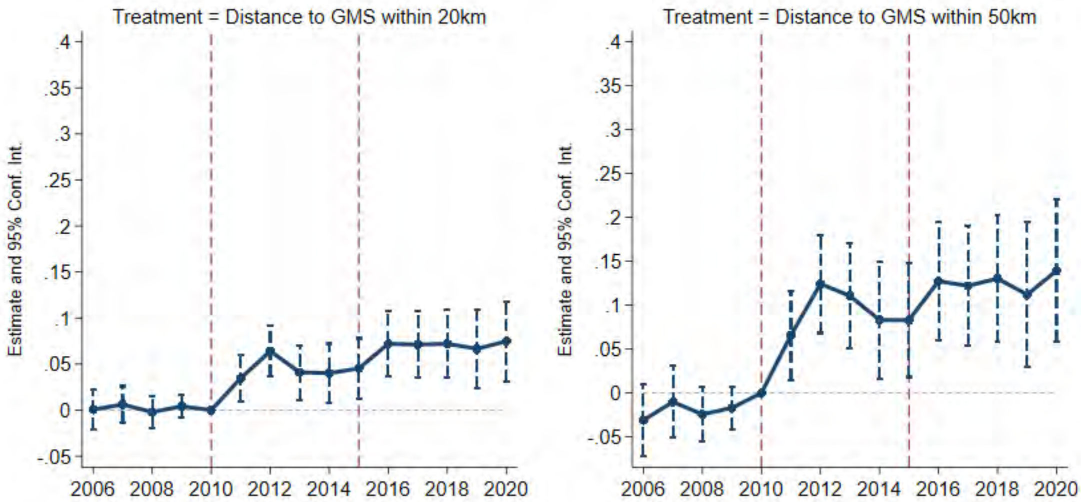
SOURCE: JAXA-ALOS, NASA-ASTER Global Digital Elevation Model.
NOTE: The blue routes are the result of a least cost path and minimum spanning tree algorithms. In the first step, Dijkstra's optimal route algorithm is applied to land cover and elevation data to construct least costly paths between each bilateral pair of the targeted destination. In the second step, the bilateral cost parameters are fed into Kruskal's minimum spanning tree algorithm to identify the subset of routes that connect all targeted nodes on a single continuous network. Euclidean spanning tree is available upon request.

FIGURE 8: Parallel Trends: Clustering of Firms and Employment

(a) Treatment effect on number of firms (log)



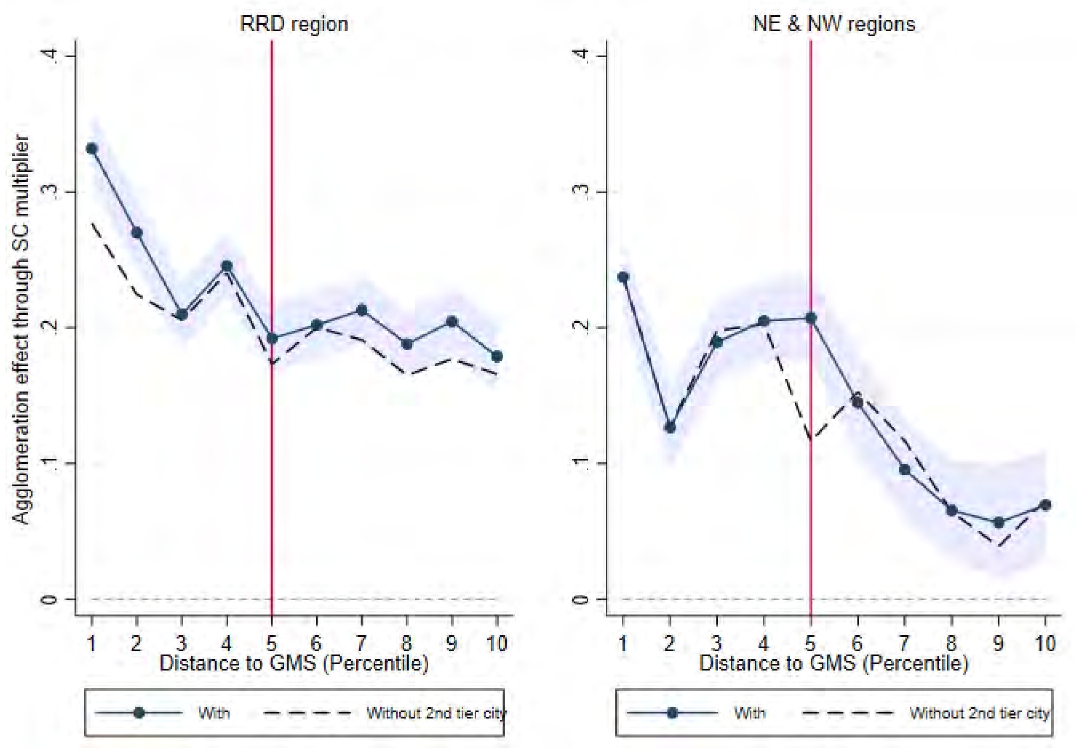
(b) Treatment effect on employment (log)



NOTE: Coefficient plots for β_t from two-way FE model: $\frac{y_{ct}-y_{c2010}}{y_{c2010}} = \alpha_t + \beta_t \mathbb{1}[Dist_c \leq d] + \gamma_t X_{it} + \mu_c + \tau_t + \varepsilon_{ct}$ where $d = 20$ or 50km. Dashed vertical lines at 2010 and 2015 mark start and completion of main GMS investment. 95% confidence intervals are based on Huber-White robust standard errors.

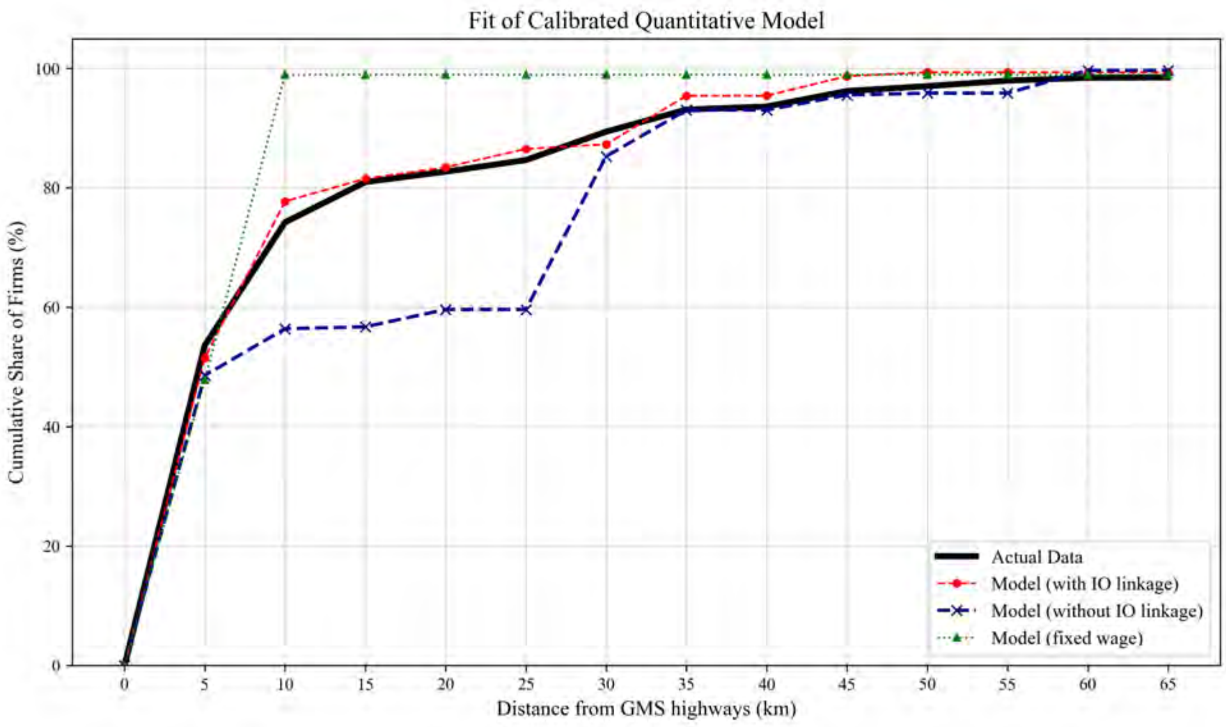
FIGURE 9: Input-Output Multiplier Effect on Number of Manufacturing Firms (in log)

Coefficient of β (Distance $\times$ Upstreamness)



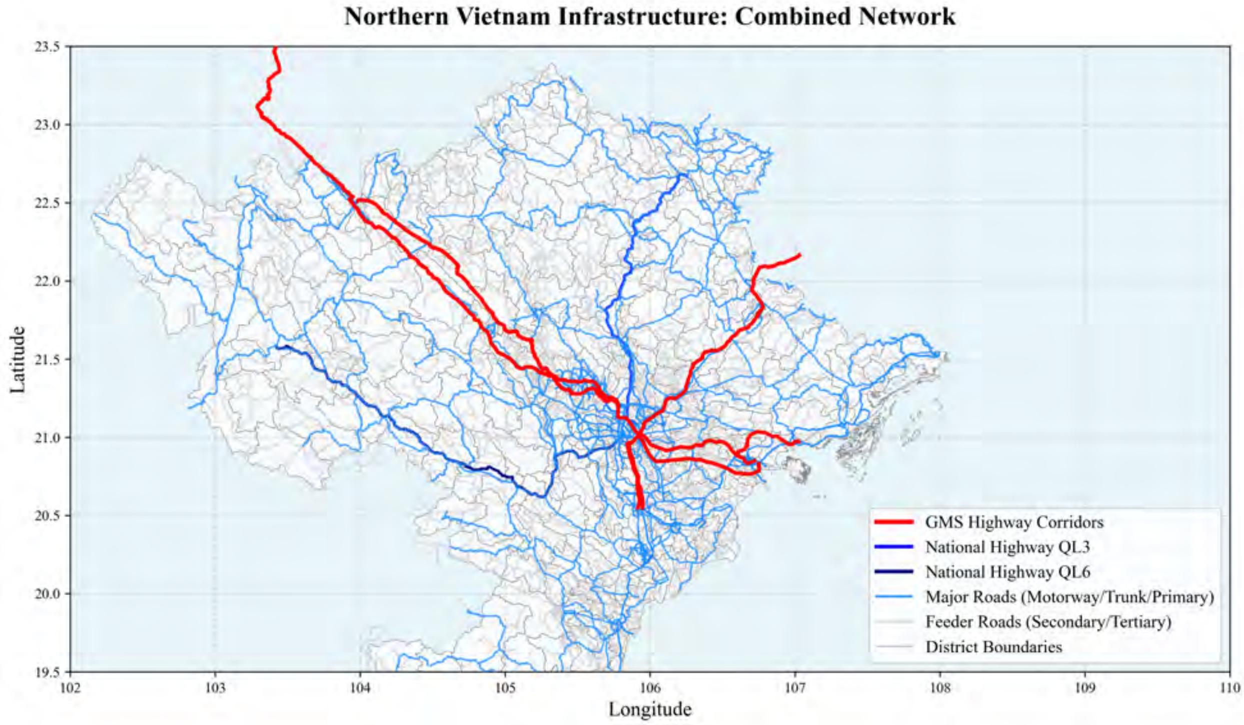
SOURCE: VES 2006–20, OECD Harmonized Input-Output Table, 2021 ed. Shaded area: 95% confidence intervals. Dashed line excludes communes along QL3 and QL6.

FIGURE 10: Model Fit: CDF of Firm Shares by Distance from GMS Highways



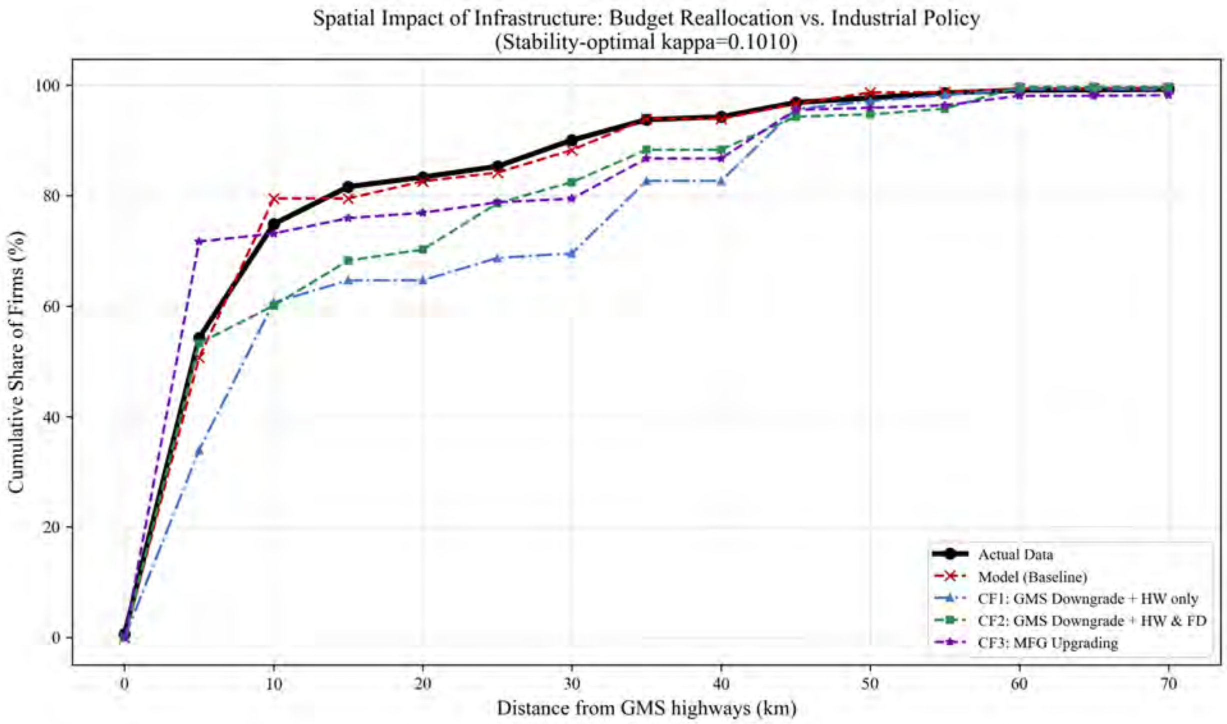
NOTE: Cumulative distribution of manufacturing firms by distance from GMS corridors across model specifications. "Actual Data" in black solid; "Model (with IO linkage)" in red dashed; "Model (without IO linkage)" in dark-blue dashed; and "Model (fixed wage)" in green dotted lines.

FIGURE 11: Northern Vietnam Road Network: GMS Corridors and Baseline Infrastructure



NOTE: Red lines: GMS highway corridors (upgraded during study period). Dark blue: two major national highways (QL3, QL6). Light blue: major roads on the Open Street Map (OSM). Grey: feeder roads (secondary or tertiary rural roads).

FIGURE 12: Counterfactual Spatial CDFs: Infrastructure vs. Industrial Policy ($\kappa = 0.101$)



NOTE: CDFs of manufacturing firms by distance from GMS corridors under the model baseline and three counterfactuals. “Actual Data” (black solid) and “Model (Baseline)” (red dashed) confirm model fit. CF1 (blue): highway (HW) reallocation only. CF2 (green): highway + feeder (FD) roads. CF3 (purple): manufacturing IO upgrading.

Online Appendix

Table of Contents

List of Appendix Tables

List of Appendix Figures

Appendix A 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 Standards Survey (VHLSS). All the data allow us to track the geographical location of respondents at the commune level, the finest administrative unit in Vietnam.[19]

The **VES** is a large-scale establishment dataset conducted annually by the General Statistics Office (GSO) since 2000. The rounds in 2001, 2006, 2011, 2016, and 2020 are census rounds and thus cover the universe of registered firms. For other years, the VES covers all registered firms above a certain employment threshold, which varies across provinces and generally increases over time. The VES comprises a main module for all firms and specific modules for firms in different industries. We focus on the Main and Transportation and Industrial Production (Manufacturing) modules. We geocoded each firm's location at the commune and district levels for the spatial analysis.

The **LFS** has been conducted annually by the GSO since 2007. Each round contains approximately 750,000–830,000 observations representing the whole population. Geographical information at the district level cannot be identified in the rounds between 2008 and 2010. We employ the LFS to estimate district-level population in 2007 and during 2010–2020. The 2009 and 2019 Population Censuses, as well as the 2014 Mid-term Population Survey, are used to construct total population at the commune and district levels.

The **VHLSS** is a biennial survey covering around 45,000 households each round. It contains information on population, housing value, and human capital stock as captured by average years of schooling. The VHLSS covers both formal and informal workers, but location information is based on living place rather than working place. An additional sub-sample of around 9,000 households provides detailed expenditure data covering monthly and annual household consumption on food, drink, non-food, and durable items.

Table 1 in the main text provides descriptive statistics for population changes at the district level and enterprise-level descriptive statistics.

[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 a fourth-tier administrative unit called "village", but it is rather a historical unit and not a current governmental working unit. Throughout this paper we use the Global Administrative Areas (GADM) for our spatial analysis.

Appendix B MST/LCP Construction Details

To address the endogeneity of highway placement, we construct a hypothetical Minimum Spanning Tree (MST) highway network derived from Least Cost Paths (LCP). The MST network reflects least-cost road network engineering logic and provides exogenous variation in route placement that satisfies relevance and exclusion restriction.

B.1 Cost Surface and LCP Algorithm

We generate the least-cost path network between provincial capitals and major intersections (Figure A1). Using terrain and land-cover data (slope, urban development, water bodies, wetlands), we create a raster grid where the cost of traversing a cell depends solely on exogenous geographic features. We adopt the following cost function from the transport engineering literature (Jha *et al.*, 2001; Jong and Schonfeld, 2003):

$$c_i = 1 + Slope_i + 25 \times Developed_i + 25 \times Water_i + 25 \times Wetland_i \qquad \text{(A1)}$$

where c_i is the cost of crossing pixel i; $Slope_i$ is the average slope gradient from NASA-ASTER global digital elevation models (30m resolution); and $Developed_i$, $Water_i$, $Wetland_i$ are dummies indicating land covered by built structures, water, or wetland, respectively, from the JAXA-ALOS land-cover map.[20] The factor of 25 reflects the high relative construction cost of elevated roads, bridges, or embankments over water and wetland.

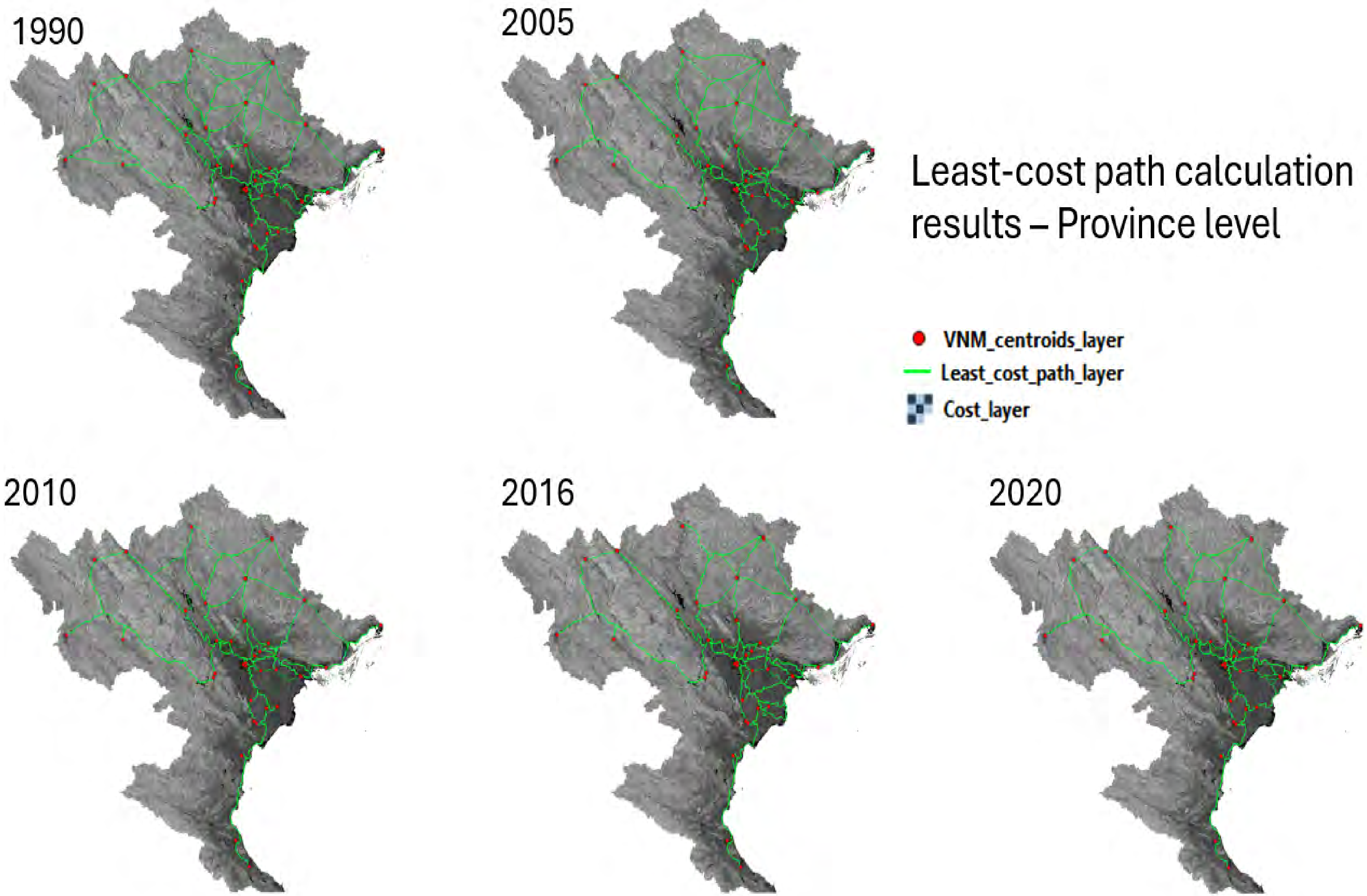
For computational efficiency we resample the cost layer to 100m resolution and calculate hypothetical LCPs for provincial capitals and 13 targeted highway intersections in northern Vietnam using Dijkstra's optimal route algorithm. LCPs are computed for 1990, 2005, 2010, 2016, and 2020 (Figure A1).

B.2 Minimum Spanning Tree (MST)

As the final step, the bilateral cost parameters derived from the LCP analysis are fed into Kruskal's MST algorithm. This identifies the subset of routes that connects all targeted nodes on a single continuous network subject to global construction cost minimization (Figure 7). The resulting network serves as our counterfactual engineering-optimal network: it is independent of local economic shocks or political targeting, since route selection is determined entirely by terrain. The distance from each commune's centroid to the nearest MST segment constitutes our primary instrument for FMA growth.

[20] JAXA-ALOS: https://www.eorc.jaxa.jp/ALOS/en/dataset/lulc/lulc_vnm_v2109_e.htm; NASA-ASTER DEM: https://asterweb.jpl.nasa.gov/gdem.asp (both at 30m resolution).

FIGURE A1: Maps of Hypothetical Least Cost Path (at Province Level)



Appendix C Market Access Regression with Alternative IV Specification

TABLE A1: First Stage Regression: Just-identified vs. Over-identified Model Specification

	(1)	(2)	(3)	(4)
	Δ Log FMA	Δ Log FMA	Δ Log FMA	Δ Log FMA
Log distance to MST	-0.138***	-0.106***	-0.105***	-0.054***
	(0.017)	(0.016)	(0.020)	(0.018)
Land elevation (z-score)		0.125		-0.796***
		(0.233)		(0.094)
Sunshine (z-score)		-0.223***		0.570***
		(0.056)		(0.078)
U.S. bomb, missiles and rockets (log+1)		-0.446***		0.012
		(0.024)		(0.035)
Observations	2,387	2,387	1,437	1,437
Region FE	YES	YES	YES	YES
Sample	RRD region	RRD region	NE & NW regions	NE & NW regions

NOTE: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors clustered at the commune level.

TABLE A2: Second Stage Regression: Just-identified vs. Over-identified Model Specification

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A. RRD regions									
	Manufacturing			Service: transport			Service: trade		
Δ Log FMA	-0.005	-0.016	0.111***	0.044	0.028	0.049***	0.055	0.040	0.072***
	(0.044)	(0.044)	(0.020)	(0.041)	(0.040)	(0.017)	(0.045)	(0.044)	(0.019)
Δ Log FMA squared			0.023***			0.011			0.012
			(0.008)			(0.007)			(0.008)
Industrial zone ($\leq 5km$) in 2011	0.139***	0.139***	0.146***	0.213***	0.213***	0.217***	0.170***	0.171***	0.175***
	(0.038)	(0.039)	(0.039)	(0.038)	(0.038)	(0.038)	(0.037)	(0.037)	(0.037)
Observations	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	Just-ID	Over-ID	Over-ID	Just-ID	Over-ID	Over-ID	Just-ID	Over-ID	Over-ID
Kleibergen-Paap F statistics	64.61	33.52	54.02	64.61	33.52	54.02	64.61	33.52	54.02
Hansen J-statistics (p-value)		0.162	0.603		0.022	0.120		0.039	0.212
Panel B. NE & NW regions									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Manufacturing			Service: transport			Service: trade		
Δ Log FMA	0.060	0.140***	0.260***	0.070	0.021	0.151***	0.137*	0.067***	0.179***
	(0.072)	(0.029)	(0.067)	(0.051)	(0.013)	(0.041)	(0.073)	(0.021)	(0.063)
Δ Log FMA squared			0.015*			0.016***			0.014*
			(0.009)			(0.005)			(0.008)
Industrial zone ($\leq 5km$) in 2011	0.286***	0.235**	0.218**	0.111	0.142**	0.123*	0.064	0.108	0.091
	(0.102)	(0.092)	(0.093)	(0.073)	(0.064)	(0.064)	(0.102)	(0.088)	(0.090)
Observations	1,437	1,437	1,437	1,437	1,437	1,437	1,437	1,437	1,437
Region FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Model	Just-ID	Over-ID	Over-ID	Just-ID	Over-ID	Over-ID	Just-ID	Over-ID	Over-ID
Kleibergen-Paap F statistics	28.843	40.760	45.531	28.843	40.760	45.531	28.843	40.760	45.531
Hansen J-statistics (p-value)		0.250	0.363		0.282	0.167		0.286	0.212

NOTE: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors clustered at the commune level. First-difference regressions take differences from 2011 to 2020. Columns 1, 4, and 7 are just-identified (ID) model using distance to MST network as the IV. Columns 2, 3, 5, 6, 8, and 9 are over-identified model as in Table 3. Controls include initial employment, share of formal employment, female workers, and distance to national roads.

Appendix D Market Access and Supply-Chain Measurement

D.1 Firm Market Access (FMA) Construction

While the Harris-type market potential (MP) index captures local accessibility within a fixed radius, it may underestimate the impact of long-distance connections created by GMS corridors linking urban core cities to remote mountainous peripheries. Moreover, the Harris index lacks theoretical micro-foundations. We therefore employ a theoretically consistent Firm Market Access (FMA) index (Redding, 2022) as the primary measure of connectivity improvements.

The FMA is defined as the sum of destination markets' real expenditure, discounted by bilateral trade costs:

$$FMA_i = \sum_{j}^{J} \tau_{ij}^{1-\sigma} \frac{E_{s,j}}{P_j^{1-\sigma}} \tag{A2}$$

where τ_{ij} is the freight trade cost between districts i and j, and the interregional trade elasticity is set at 4 following Simonovska and Waugh (2014). $E_{s,j}$ is total expenditure at destination, measured by local wage bills, and P_j is the destination price index. Trade cost parameter τ_{ij} is estimated as a unit cost of transportation services using the VES freight module, which records trade values and volumes for transport companies.

Figure 6 illustrates a notable decrease in interregional trade costs over 2011–2020, particularly along the highway in peripheral areas. An exception is the area around Yên Bái province, a mountainous region where geographical constraints limit the construction of connecting roads. The estimated FMA consistently indicates a significant enhancement in firm market access, primarily attributable to improved transport connectivity.

D.2 Harris Market Potential Index and Semi-Parametric Estimates

As a complement to the FMA, we compute a reduced-form Harris market potential (MP) index, which captures both local market size and accessibility of surrounding markets. Assuming iceberg trade costs, the Harris index measures effective worker density as a distance-weighted sum of employment in neighboring districts (Combes and Gobillon, 2015):

$$MP_{ct} = \sum_{l \neq c} \frac{den_{lt}}{d_{cl}} \tag{A3}$$

where den_{lt} is employment density in commune l at year t and d_{cl} is the shortest distance between centroids of communes c and l. The index aggregates across all communes within a 50-kilometer radius.

Figure A2 shows the spatial distribution of market potential in northern Vietnam. Market potential has historically concentrated in core cities around Hanoi but expanded markedly during the 2010s. Between 2011 and 2016, districts located along the GMS highways experienced the fastest increases in their MP index. Mountainous districts near the Chinese border also experienced a moderate expansion. These patterns underscore how the GMS corridor reshaped the geography of market potential—reinforcing advantages of well-connected intermediate areas while extending accessibility improvements into previously remote districts.

To quantify the spatial scale of the GMS impact on local market accessibility, and to motivate the 20km and 50km treatment thresholds used in the difference-in-differences specification, we apply a semi-parametric regression following Robinson (1988):

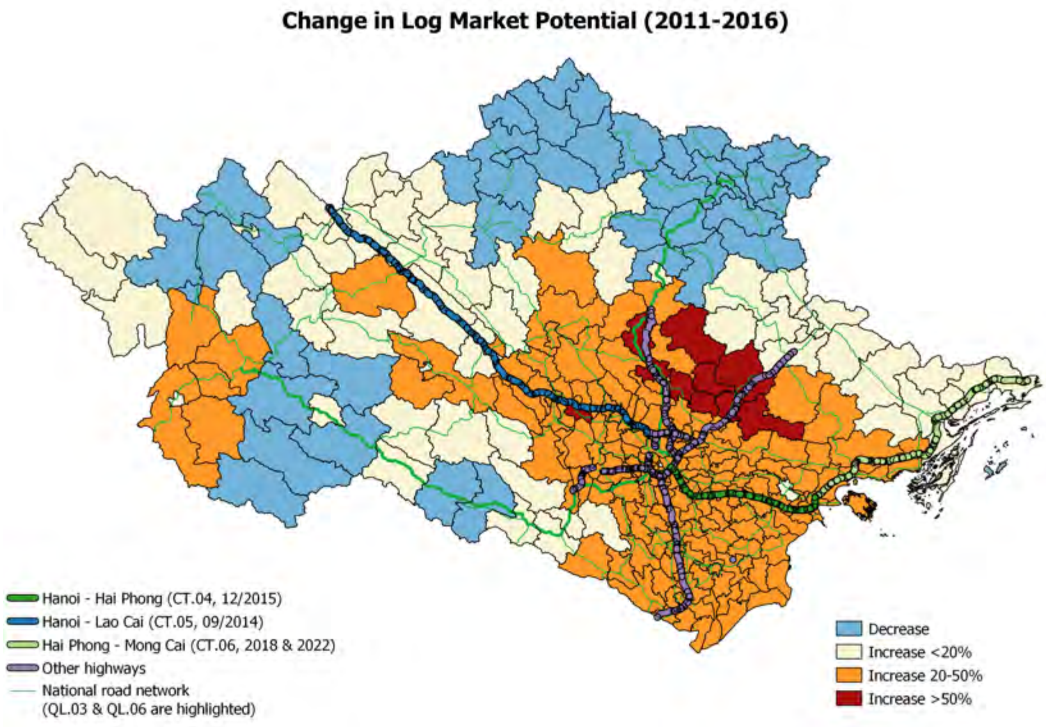
$$\ln MP_{drt} = m(GMSdist_{dr}) + x_{drt}\beta + \epsilon_{drt} \tag{A4}$$

where $m(\cdot)$ is a non-parametric function of $GMSdist$—the Euclidean distance between the People's Committee office of district d in region r and the nearest GMS highway interchange. The parametric part x_{drt} includes provincial dummies and distance to Hanoi. Figure A3 plots the estimated relationship between distance to the GMS highways and changes in the MP index, with 95% confidence intervals, reported separately for the short run (2011–2016) and medium run (2011–2020).

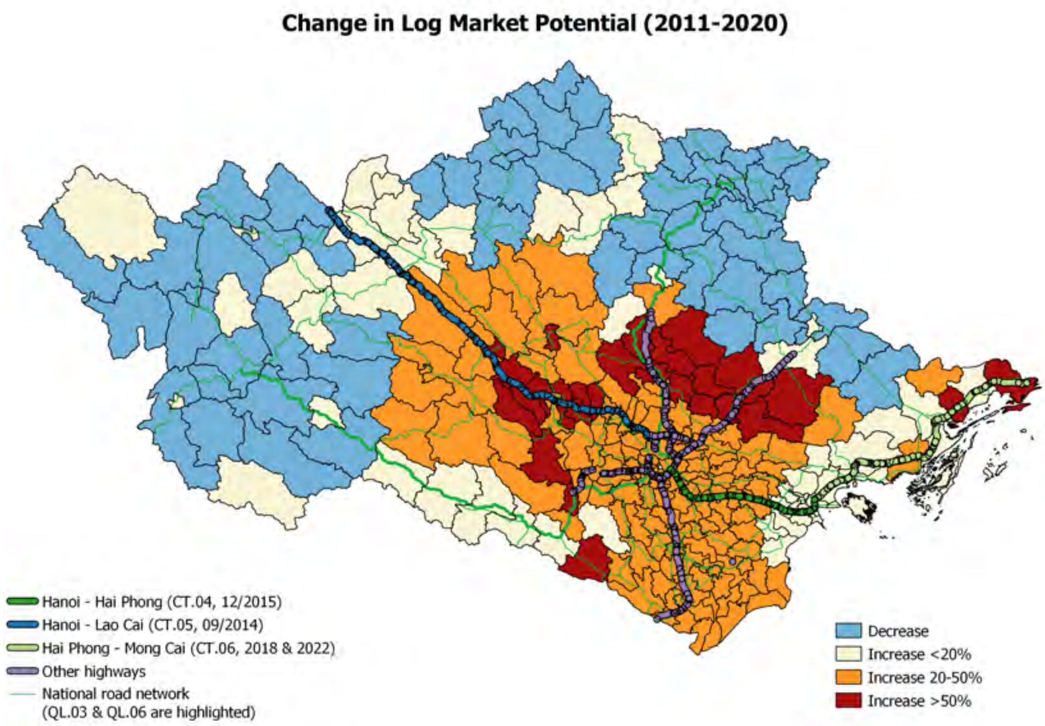
The short-run estimates show that districts within 50km of the highways experienced a 20–40% increase in market potential, while remote areas beyond 70km recorded virtually no change. Within 20km in the NE and NW regions, market potential expanded at nearly 50%, highlighting particularly strong accessibility improvements in those areas. Based on this evidence, we define treatment exposure as highway access within 20km and 50km—the thresholds beyond which the effects largely dissipate.

FIGURE A2: Spatial Distribution of the MP Index

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

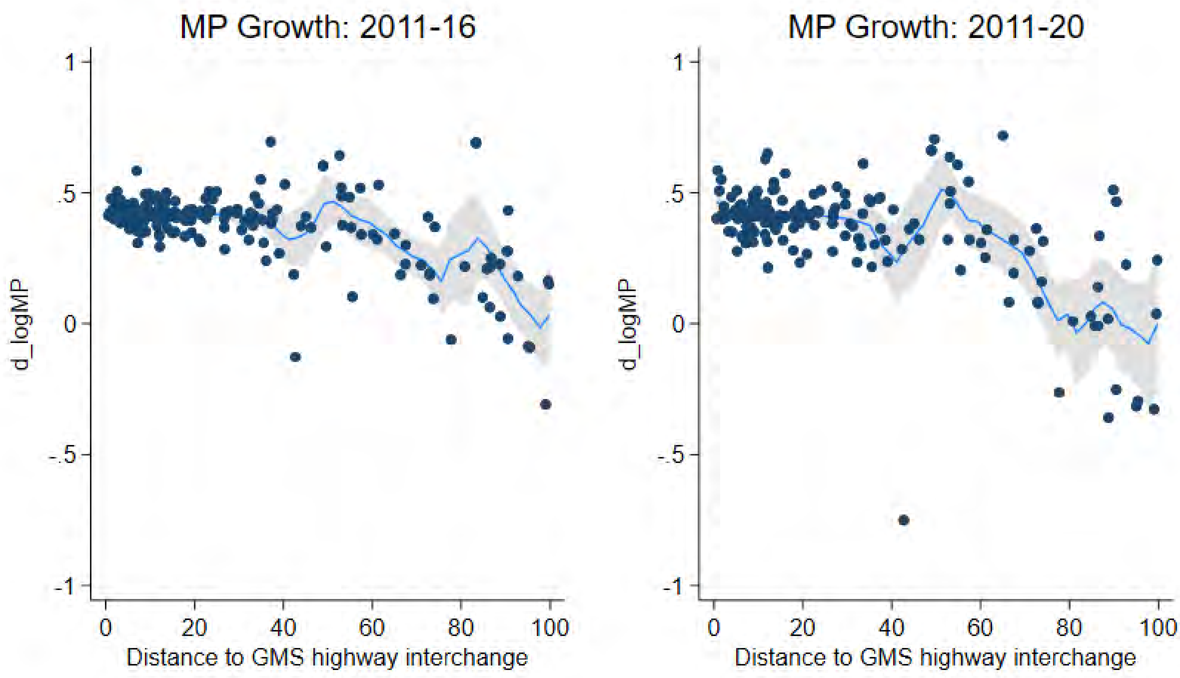


(b) Change in the MP Index from 2011 to 2020



NOTE: MP index computed using total district employment of registered firms within a 50km radius.

FIGURE A3: Change in the MP Index and the Distance from the GMS Corridor



NOTE: Sample covers all districts in northern Vietnam. x-axis is the Euclidean distance between each district's People's Committee office and the nearest GMS interchange. Provincial dummies and distance to Hanoi are controlled as parametric factors.

D.3 Supply Chain Linkage Measurement

Besides physical distance, the effect of GMS highways is expected to vary by each commune's production structure. Places with stronger integration into input-output networks should experience larger treatment effects, as trade cost savings magnify intermediate transactions.

We measure intersectoral supply chain linkages using the OECD Harmonized Input-Output (IO) Tables (45 sectors). The overall strength of local input-output linkages is commonly measured by the Leontief inverse of the IO matrix. To better reflect production chains, we construct an index of local *upstreamness*—the distance of each industry from consumers, *à la* Antràs *et al.* (2012). Commune-level upstreamness M_{ct} is defined as the expenditure-weighted average of industry-level upstreamness L_{st}, using the initial share of industry's sales for each commune as weights:

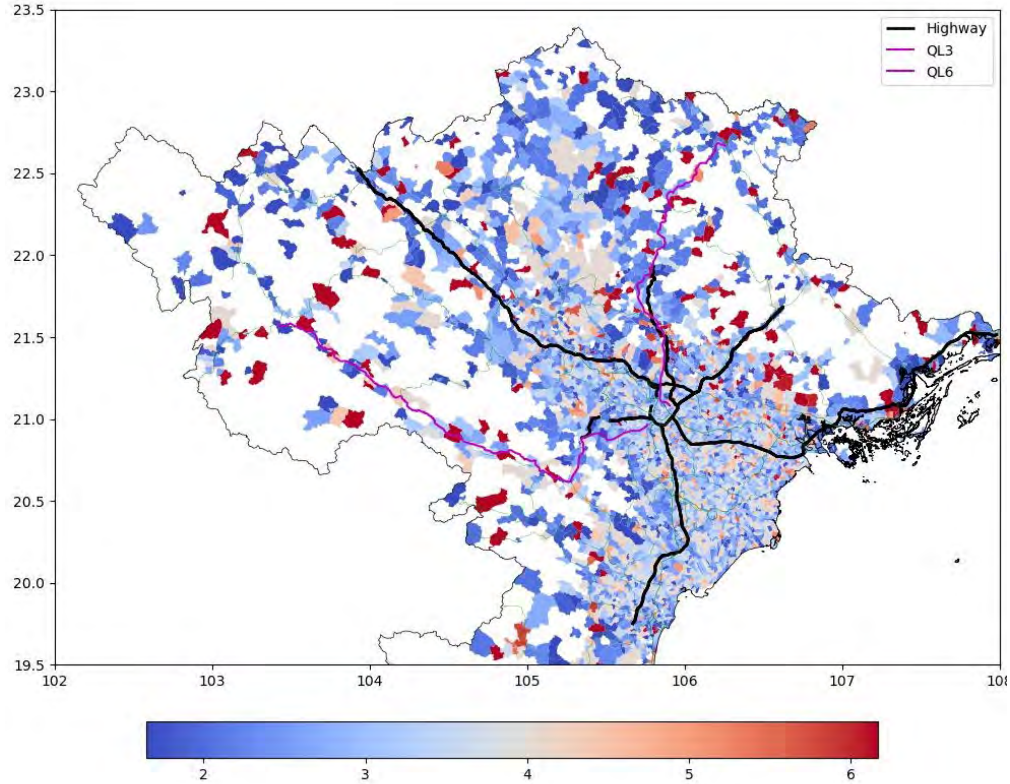
$$M_{ct} = \sum_{s=1}^{45} L_{st} \frac{Sales_{cs,0}}{\sum_s Sales_{c,0}} \tag{A5}$$

A higher value of M_{ct} indicates that the commune specializes in industries that are further upstream (more distant from final consumers) and thus likely to benefit from forward linkages when trade costs fall.

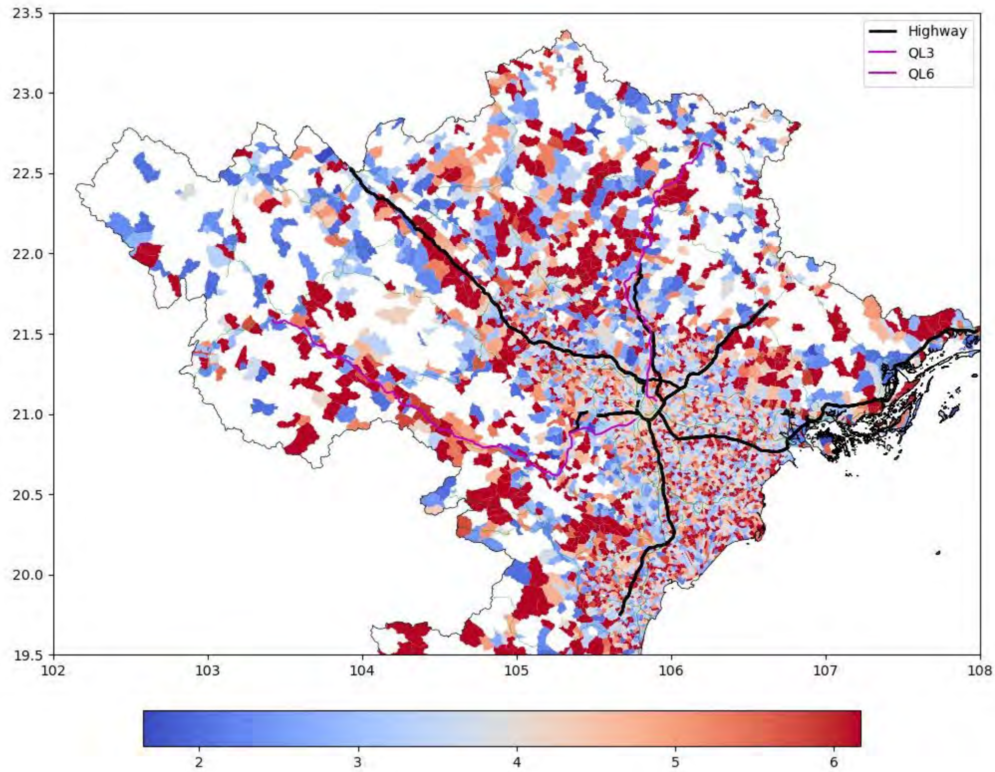
Figure A4 illustrates the evolution of production multipliers (Leontief inverse) across space. In 2006, local linkages were weak, reflecting Vietnam's reliance on imported inputs, particularly from China. Over the course of a decade, supply chains deepened, especially along highway corridors in NW Vietnam. This trend was reinforced by government industrial policies promoting mechanical engineering, electronics and assembly, textiles, and related technology industries. Consistent with this, VES data point to growing concentrations of upstream industries (including food processing, wood and paper production, rubber and plastic products, electric products) in peripheral areas. These industries increasingly source inputs locally and supply processed goods to transport, wholesale, and retail firms, indicating stronger vertical linkages. Textile industries also spread from Hanoi's southeastern coast toward the South China Sea, further illustrating the spatial diffusion of supply chains.

FIGURE A4: Spatial Distribution of Input-output Linkages: Leontief Inverse

(a) In 2011



(b) In 2017



SOURCE: VES 2006–20, OECD Harmonized Input-Output Table, 2021 ed.

Appendix E Robustness Checks

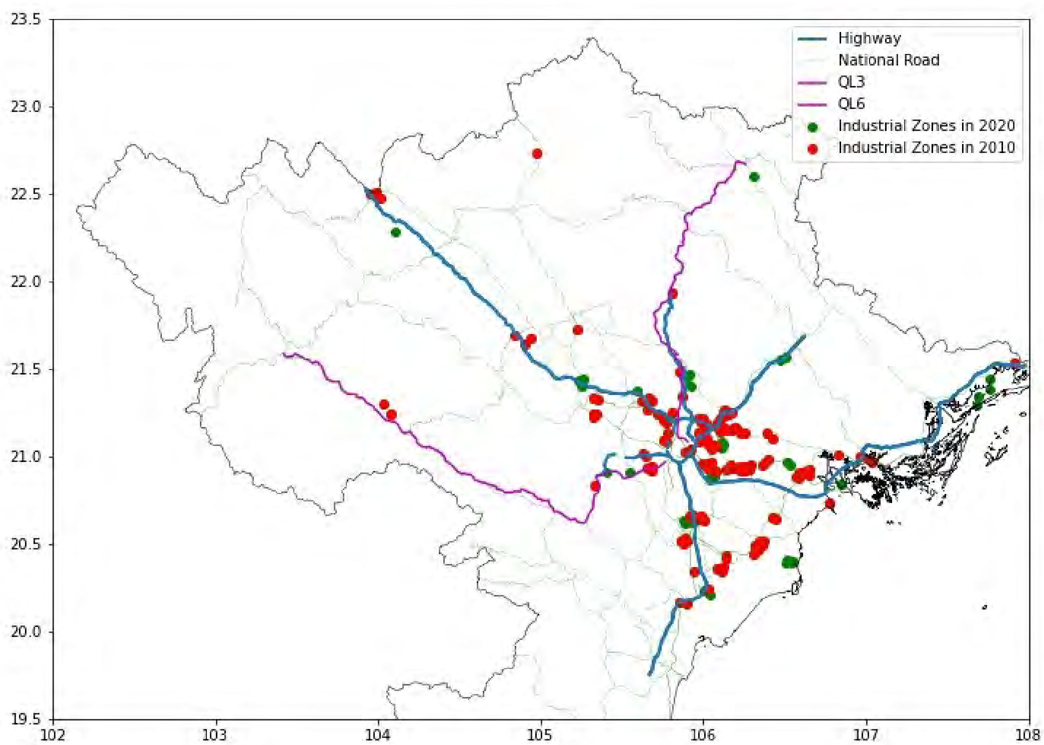
This section presents robustness checks for the reduced-form results in Section 4.1.

E.1 Industrial Zone Exclusion

Industrial zones (IZs) are often co-located along highways (Figure A5), raising concerns that part of the observed effect may reflect place-based IZ policies rather than connectivity itself. Including such communes in the treatment group could overestimate the highway treatment effect. To address this, we geocode all industrial zones in northern Vietnam using Google Earth Pro together with their establishment years. We then re-estimate the DID specification excluding communes within 1, 5, and 10km of any IZ.

As shown in Table A3(a), effect sizes generally decline once these communes are dropped—especially in peripheral areas within 20km, where combined 10-year effects are nearly halved when communes within 10km of industrial zones are excluded. This reflects the critical role of place-based policy in attracting firms to peripheral cities in northern Vietnam. Nonetheless, the main conclusions hold: highways exert an independent effect on agglomeration beyond industrial zone policy. Table A3(b) similarly shows the robustness of the GMS impact on job creation.

FIGURE A5: Location of Industrial Zones in Northern Vietnam



SOURCE: Vietnamese official documents and decrees.

TABLE A3: 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									
After 10 years (Within 20km)	0.338***	0.299***	0.318***	0.292***	0.214***	0.104	0.601***	0.559***	0.532***
After 10 years (20-50km)	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									
After 10 years (Within 20km)	0.462***	0.375***	0.230***	0.247***	0.179***	0.086**	0.314***	0.218***	0.093*
After 10 years (20-50km)	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. Columns 1, 4, 7 exclude communes within 1km of nearest Industrial Zone center; Columns 2, 5, 8 exclude communes within 5km; and Columns 3, 6, 9 exclude communes within 10km. The combined effects, $\beta_1 + 10\beta_3$ for communes within 20km and $\beta_2 + 10\beta_4$ for communes in the 20-50km zone, estimate the cumulative treatment effect over a 10-year horizon.

(b) Job Creation

	Log(employment in registered firms) Panel A. RRD region								
	(1)	(2) Manufacturing	(3)	(4)	(5) Service: transport	(6)	(7)	(8) Service: trade	(9)
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									
After 10 years (Within 20km)	0.610***	0.656***	1.019***	0.442***	0.335**	0.305**	0.829***	0.822***	0.881***
After 10 years (20-50km)	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: 20-50km	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									
After 10 years (Within 20km)	0.767***	0.620***	0.300*	0.375***	0.250***	0.150	0.377***	0.286***	0.218*
After 10 years (20-50km)	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. The combined effects, $\beta_1 + 10\beta_3$ for communes within 20km and $\beta_2 + 10\beta_4$ for communes in the 20-50km zone, estimate the cumulative treatment effect over a 10-year horizon.

E.2 Technology Spillovers

In urban centers, knowledge spillovers through formal or informal interaction across firms could partially explain industrial clustering (Greenstone *et al.*, 2010). Literature suggests that such spillover effects play an important role in stimulating innovation, particularly in high-tech industries—a dynamic potentially relevant to the small neighborhood surrounding central Hanoi, where foreign firms and high-tech companies concentrate.

To rule out the possibility that results are driven by pure technology spillovers rather than input-output linkages, we examined R&D expenditure data from the VES. Innovation spending is negligible ($<$ 3% of communes) and concentrated solely in the Hanoi core. This is consistent with the industry-level DID results in Appendix F, which show that IT clusters in the periphery emerged only after highway connectivity improved sufficiently. The absence of R&D in the periphery supports our interpretation that polycentric development is driven by cost linkages and market access improvements rather than knowledge spillovers.

Appendix F Industry-Specific DID Estimates

Table A4 disaggregates the DID estimates by the type of manufacturing industries. The results show clear heterogeneity.

Within 20km of the highways, both core and peripheral regions experienced significant entry across a broad set of manufacturing industries. In the 20–50km zone, however, effects are concentrated in light, labor-intensive sectors such as textiles, garments, wood, paper, and leather products, while heavy and intermediate industries (e.g., metals, machinery, chemicals) display weaker or insignificant responses.

The patterns also reveal an important role for industrial zones in shaping manufacturing agglomeration. Zones located along the highways reinforced entry across sectors and were effective in attracting foreign affiliates to peripheral regions. The IT and computer industry shows notable clustering within 50km in both core and periphery regions, suggesting that infrastructure upgrades created opportunities for technology-related industries to expand beyond the core once industrial zones were established.

This heterogeneity underscores that industries with stronger production linkages and scale economies tend to cluster near highway hubs, where dense buyer-supplier networks amplify spillovers. By contrast, entry further from the highways is dominated by lighter industries with weaker supply chain linkages and limited complementarities.

TABLE A4: Difference-in-Differences Regression: By Industry Type

Panel A: RRD region

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
	Dependent variable: Log number of firms														
	FDI	IT	PC	Electric	Machine	Textile	Agri-related	Food/Bevarage	Mining	Petroleum	Chemical	Rubber/Plastic	Wood/Paper	Metal	Construction
IZs dummy (<=5km)	-0.006	-0.104***	0.005	-0.016	-0.026**	-0.064***	0.005	-0.027	-0.054***	-0.004	-0.039**	0.003	-0.021	-0.027	-0.090***
	(0.028)	(0.027)	(0.021)	(0.013)	(0.012)	(0.024)	(0.024)	(0.022)	(0.013)	(0.004)	(0.015)	(0.026)	(0.025)	(0.029)	(0.032)
Combined effect of β_1 and β_2															
After 10 years (Within 20km)	0.292***	0.539***	0.205***	0.124***	0.120***	0.322***	-0.552***	0.069	-0.186***	0.007	0.178***	0.113	0.181***	0.207***	0.156
After 10 years (20-50km)	0.014	0.175***	0.020***	0.019***	0.024	0.289***	-0.428***	-0.052	-0.215***	-0.007	0.001	-0.102	-0.027	-0.072	-0.114
Observations	33,429	30,395	30,395	30,395	30,395	30,395	30,395	30,395	30,395	30,395	30,395	30,395	30,395	30,395	30,395
Commune FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Pre-trend	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Post-trend	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES

Panel B: NE & NW regions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
	Dependent variable: Log number of firms														
	FDI	IT	PC	Electric	Machine	Textile	Agri-related	Food/Bevarage	Mining	Petroleum	Chemical	Rubber/Plastic	Wood/Paper	Metal	Construction
IZs dummy (<=5km)	0.184***	0.130**	0.085*	0.027	-0.011	0.002	-0.070**	0.032	0.071**	0.008	0.054*	0.065**	0.025	0.145***	0.194***
	(0.065)	(0.052)	(0.047)	(0.020)	(0.019)	(0.034)	(0.035)	(0.031)	(0.033)	(0.006)	(0.029)	(0.031)	(0.035)	(0.048)	(0.050)
Combined effect of β_1 and β_2															
After 10 years (Within 20km)	0.120***	0.241***	0.167***	0.045***	0.020*	0.205***	0.064	0.018	0.094***	0.005	0.063***	0.255***	0.228***	0.187***	0.217***
After 10 years (20-50km)	0.068***	0.111***	0.064***	0.018**	0.024**	0.109***	0.002	0.039	0.139***	0.005	0.005	0.173***	0.262***	0.125***	0.120***
Observations	19,429	18,262	18,262	18,262	18,262	18,262	18,262	18,262	18,262	18,262	18,262	18,262	18,262	18,262	18,262
Commune FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Pre-trend	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Post-trend	YES	YES	YES	YES	YES	YES	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.

Appendix G Derivation of Spatial Equilibrium Model

G.1 Full Model Equations

The model consists of J locations indexed by i, j, with S sectors categorized into Agriculture (A) and Nonagriculture (NA), including manufacturing and service industries.

Preferences. Consumers have Cobb-Douglas preferences with expenditure share α_A on agriculture and $(1 - \alpha_A)$ on a CES composite of nonagricultural varieties. The composite $C_{NA,i}$ is defined over S sectors and the continuum of varieties ω in sector s.

$$U_i = C_{A,i}^{\alpha_A} C_{NA,i}^{1-\alpha_A}, \qquad C_{NA,i} = \prod_{s=1}^{S} Q_{s,i}^{\theta_s}, \qquad Q_{s,i} = \left(\int_{\omega \in \Omega_s} q_{s,\omega,i}^{\frac{\sigma-1}{\sigma}} d\omega \right)^{\frac{\sigma}{\sigma-1}} \tag{A6}$$

where $\sigma > 1$ is the elasticity of substitution across varieties and θ_s is the expenditure share of sector s within nonagriculture, with $\sum_{s=1}^{S} \theta_s = 1$. Consumers allocate income according to Cobb-Douglas weights:

$$e_{A,j} = \alpha_A w_j L_j, \qquad \sum_s e_{s,j} = (1 - \alpha_A) w_j L_j, \quad e_{s,j} = \theta_s (1 - \alpha_A) w_j L_j$$

Agriculture. Agricultural production is characterized by perfect competition and constant returns to scale under a Cobb-Douglas production technology:

$$Y_{A,i} = A_{A,i} L_{A,i}^{\beta_L} K_i^{1-\beta_L} \tag{A7}$$

where $A_{A,i}$ is agricultural productivity and β_L is the labor share. The agricultural wage equals the marginal product of labor:

$$w_i = \beta_L A_{A,i} \left(\frac{L_{A,i}}{K_i} \right)^{\beta_L - 1} \tag{A8}$$

As nonagriculture expands, $L_{A,i}$ falls, raising w_i through this equation—the key dispersion force preventing excessive concentration.

Nonagriculture. Nonagricultural sector consists of firms operating under monopolistic competition and increasing returns to scale, under a Cobb-Douglas production that integrates intermediate inputs from all sectors of the economy:

$$Y_{NA,s,i} = A_{NA,s,i} L_{NA,s,i}^{\eta_s} \prod_{r=1}^{S} X_{r,s,i}^{\mu_{rs}} \tag{A9}$$

where $A_{NA,s,i}$ is nonagricultural sector productivity, $L_{NA,s,i}$ is nonagricultural sector labor, and $X_{r,s,i}$ is the bundle of sector r inputs used by sector s in region i. η_s is the labor share in sector s. μ_{rs} is the input-output coefficient (sector r's expenditure share in sector s

production) and $\eta_s + \sum_r \mu_{rs} = 1$. Marginal cost for a firm in sector s, region i is:

$$c_{s,i} = \frac{1}{A_{NA,s,i}} w_i^{\eta_s} \prod_{r=1}^{S} P_{r,i}^{\mu_{rs}}, \qquad p_{s,i} = \frac{\sigma}{\sigma - 1} c_{s,i} \tag{A10}$$

Firms set price with a constant markup. We incorporate a Marshallian agglomeration externality as follows:

$$A_{NA,s,i} = \bar{A}_{NA,s,i} \left(n_{s,i}\right)^{\lambda} \tag{A11}$$

where $\lambda > 0$ is the agglomeration elasticity.

Trade Costs and Firm Market Access. The sectoral CES price index is defined over products supplied from all regions ($i \in J$) taking the following CES form:

$$P_{s,j} = \left(\sum_{i=1}^{J} n_{s,i} \left(p_{s,i} \tau_{ij}\right)^{1-\sigma} \right)^{\frac{1}{1-\sigma}} \tag{A12}$$

where τ_{ij} is the iceberg trade cost between regions i and j, and $n_{s,i}$ is the number of firms producing good s in region i. A fall in trade costs τ_{ij} reduces $P_{s,j}$, making goods cheaper in region j and stimulating demand.

Following Redding (2022), Firm Market Access (FMA) 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 iceberg trade costs $\tau_{ij} \geq 1$ and prices in the destination region. We model the GMS highway upgrades as a reduction in τ_{ij} between connected districts.

$$FMA_{s,i} = \sum_{j=1}^{J} \tau_{ij}^{1-\sigma} \cdot \frac{E_{s,j}}{P_{s,j}^{1-\sigma}} \tag{A13}$$

where $E_{s,j}$ is total expenditure on sector s goods in destination j, which includes both final consumption and intermediate input demand for sector s goods from all other sectors:

$$E_{s,j} = \theta_s (1 - \alpha_A) w_j L_j + \sum_{r=1}^{S} \mu_{sr} R_{r,j} \tag{A14}$$

An increase in $FMA_{s,i}$ directly raises the potential revenue of firms in location i.

Free Entry and Equilibrium Conditions.

Firm Entry We assume free entry in the nonagricultural sector. Firms enter location i until operating profits equal the fixed cost of entry. Each firm in sector s and region i incurs a fixed labor cost $w_i F$. Because of constant markups, a constant fraction $\frac{1}{\sigma}$ of a firm's revenue $r_{s,i}$ is operating profit. From CES demand, revenue is given by:

$$r_{s,i} \;=\; p_{s,i}^{1-\sigma}\, FMA_{s,i}$$

The profit of a firm is defined as:

$$\pi_{s,i} \;=\; \frac{1}{\sigma} r_{s,i} \;-\; w_i F$$

Free entry drives profits to zero in equilibrium. The resulting *zero-profit condition* (ZPC), $p_{s,i}^{1-\sigma}\, FMA_{s,i} = \sigma w_i F$, requires each firm to cover its fixed cost. The equilibrium number of firms $n_{s,i}$ then follows from total sectoral revenues $R_{s,i}$ in each region:

$$n^*_{s,i} \;=\; \frac{R_{s,i}}{\sigma\, w_i F} \tag{A15}$$

This pins down how many firms a region can support: a rise in firm market access $FMA_{s,i}$ raises firm revenue above break-even, inducing entry. Entry in turn lowers the local price index $P_{s,i}$ and, through the competition channel, pushes firm revenue back to the break-even level, so that $n_{s,i}$ adjusts upward until zero profits are restored.

It is through this channel that improved connectivity reshapes the spatial distribution of activity. The highway upgrade reduces trade costs τ_{ij}, which raises $FMA_{s,i}$ and induces entry in the newly connected regions. Because expenditure $E_{s,j}$ embeds input-output linkages across sectors, this entry feeds back into the market access of upstream and downstream firms, amplifying the initial shock and triggering self-reinforcing industrial clustering—the Big Push mechanism.

Labor market. Wages w_i are determined endogenously to clear local labor markets. Labor is mobile across sectors within a district. We use observed district-level wages from the Enterprise Census to calibrate baseline local productivity fundamentals. While the observed data serve as the initial values for calibration, the equilibrium wage will be adjusted endogenously in response to changes in labor demand driven by market access shocks to equate local labor supply and demand in equilibrium.

G.2 Derivation of Proposition 1

Setup. Let us consider the equilibrium of a nonagricultural sector in location i. Let n denote firm count, $w(n)$ the market-clearing wage, and $P_{NA}(n)$ the CES price index. The ZPC (eq. A15), combined with agglomeration $A(n) = \bar{A} n^{\lambda}$ and CES marginal cost (eq. A10), yields a fixed-point equation:

$$n = G(n; FMA_0, \mu) \equiv FMA_0 \cdot \hat{G}(n; \mu) \tag{A16}$$

where

$$\hat{G}(n;\mu) = \frac{(\sigma-1)^{\sigma-1}}{\sigma^\sigma F} \cdot \underbrace{\bar{A}^{\sigma-1} n^{\lambda(\sigma-1)}}_{\text{agglomeration (+)}} \cdot \underbrace{P_{NA}(n)^{\mu(1-\sigma)}}_{\text{IO linkage (+)}} \cdot \underbrace{\frac{E(n;\mu)}{w(n)^\sigma}}_{\text{wage dispersion (-)}}$$

and $FMA_0 = \sum_{j=1}^{J} \tau_{ij}^{1-\sigma} \frac{E_j}{P_{NA,j}^{1-\sigma}}$ is the exogenous market access component.

Proof of Proposition 1 The fixed-point condition (eq. A16) is equivalent to finding where the curve $G(n)$ crosses the 45-degree line in the (n, G) plane. Whether that crossing is unique or multiple depends on the slope of G relative to unity. Define the log-elasticity:

$$\varepsilon_G(n;\mu) \equiv \frac{\partial \ln G}{\partial \ln n} = \underbrace{\lambda(\sigma-1)}_{\substack{\text{agglomeration}\\(+)}} - \underbrace{\sigma\,\varepsilon_w(n)}_{\substack{\text{wage}\\\text{competition }(-)}} + \underbrace{\mu(\sigma-1)|\varepsilon_P(n)|}_{\substack{\text{IO cost}\\\text{saving }(+)}} + \underbrace{\frac{\mu\, n\sigma F w(n)[1+\varepsilon_w(n)]}{E(n;\mu)}}_{\substack{\text{IO demand}\\(+)}} \quad \text{(A17)}$$

where $\varepsilon_w \equiv \frac{\partial \ln w}{\partial \ln n} > 0$ and $\varepsilon_P \equiv \frac{\partial \ln P_{NA}}{\partial \ln n} < 0$. This elasticity measures whether G is *steeper* than the 45-degree line ($\varepsilon_G > 1$) or *flatter* ($\varepsilon_G < 1$). The threshold $\overline{FMA}$ and the multiplicity of equilibria are both direct consequences of how ε_G varies with n.

Unique equilibrium ($0 \leq \mu \leq \mu^*$). Without IO linkages ($\mu = 0$), only agglomeration and wage competition remain. Wage competition dominates as n rises, so $\varepsilon_G < 1$ throughout and G crosses the 45-degree line exactly once, yielding a unique low-activity equilibrium (Panel (a), Figure A6).

In the intermediate case ($0 < \mu \leq \mu^*$), IO linkages are present but insufficiently strong to push ε_G above unity. The G-curve remains everywhere flatter than the 45-degree line, so the equilibrium is still unique. However, as IO demand raises total expenditure, G shifts upward and yields a higher unique equilibrium n^* than under $\mu = 0$.

Multiple equilibria ($\mu > \mu^*$). With IO linkages, the IO demand term—which measures intermediate purchases as a share of total expenditure—grows disproportionately at intermediate firm counts: each additional firm both buys inputs and generates revenue that rival firms spend on inputs. When μ exceeds a critical value μ^*, this feedback pushes ε_G above unity on an interval (n_1, n_2), causing the G-curve to bulge above the 45-degree line. The curve then crosses the line three times rather than once—from above at n^* (stable), from below at n_{crit} (unstable), and from above again at n^{**} (stable).

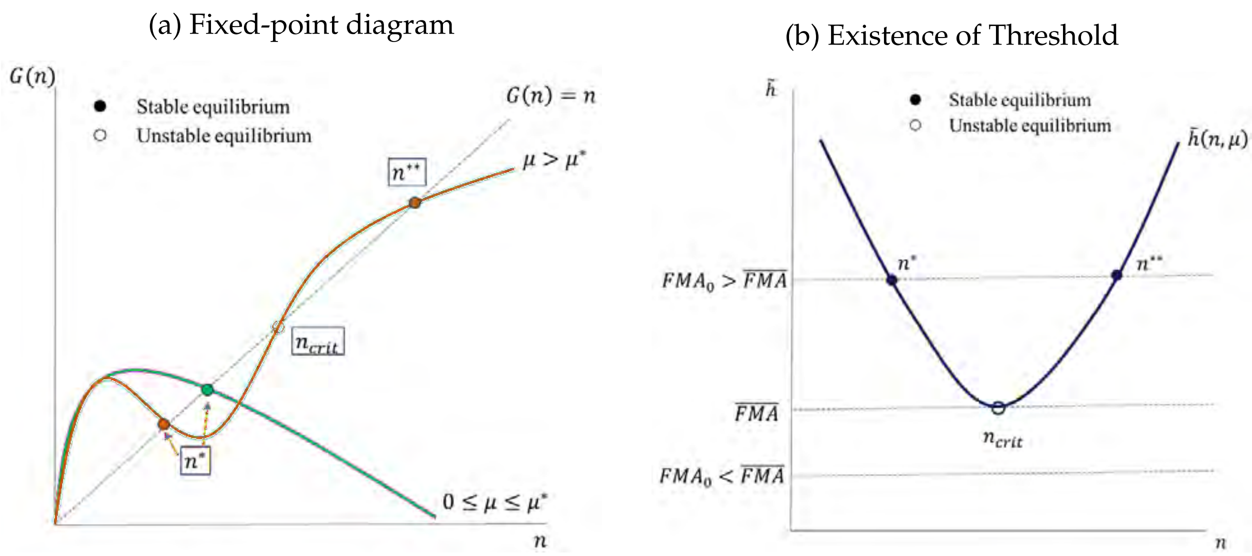
From the G-diagram to the threshold (Panel (b), Figure A6). To characterize how equilibria depend on FMA_0, define

$$\tilde{h}(n;\mu) \equiv \frac{n}{\hat{G}(n;\mu)}, \qquad \overline{FMA}(\mu) \equiv \min_{n\in(0,\bar{n})} \tilde{h}(n;\mu)$$

Solving eq. (A16) is equivalent to finding where the horizontal line FMA_0 intersects $\tilde{h}(n;\mu)$. Since $\frac{\partial \ln \tilde{h}}{\partial \ln n} = 1 - \varepsilon_G$, the region where $\varepsilon_G > 1$ on (n_1, n_2) makes $\tilde{h}$ locally *decreasing*, creating the valley visible in Panel (b). The valley floor at n_{crit} (where $\varepsilon_G = 1$) is exactly $\overline{FMA}(\mu)$. The three cases follow directly from Panel (b):

- $FMA_0 < \overline{FMA}$: the horizontal line lies below n_{crit}. The only equilibrium is the trivial $n = 0$: no firm can break even, and nonagriculture does not emerge.
- $FMA_0 > \overline{FMA}$: the line clears the valley and intersects $\tilde{h}$ at three points $n^* < n_{\text{crit}} < n^{**}$. The outer two are stable ($G' < 1$); the middle is an unstable saddle ($G' > 1$). The high equilibrium n^{**} is self-sustaining via IO revenue circulation.
- $FMA_0 = \overline{FMA}$: the line is tangent to the valley floor, merging the low and unstable equilibria in a saddle-node bifurcation at n_{crit}.

FIGURE A6: Fixed Point Equilibrium Solution



Why do stronger IO linkages raise the threshold and amplify post-threshold growth? Higher μ has *opposite* effects at n_{crit} and n^{**}, because the intermediate market is thin at the former and thick at the latter.

To see this, differentiate $\ln \hat{G}$ with respect to μ:

$$\left.\frac{\partial \ln \hat{G}}{\partial \mu}\right|_n = \underbrace{(1-\sigma)\ln P_{NA}(n)}_{\text{input cost channel} < 0} + \underbrace{\frac{n\sigma F w(n) - (1-\alpha_A) w(n) L}{E(n;\mu)}}_{\text{net demand channel}}$$

The input-cost channel is always negative: higher μ raises the importance of intermediate inputs, but $P_{NA} > 1$, so input costs increase. The net demand channel reflects whether IO demand ($n\sigma F w$) or final consumption ($(1-\alpha_A) w L$) dominates total expenditure.

At n_{crit} (thin markets): activity is low, so $n_{\text{crit}}\sigma F w \ll (1-\alpha_A) w L$ and both channels are negative: $\partial\hat{G}/\partial\mu|_{n_{\text{crit}}} < 0$. By the envelope theorem,

$$\frac{d\overline{FMA}}{d\mu} = \left.\frac{\partial\tilde{h}}{\partial\mu}\right|_{n_{\text{crit}}} = \frac{-n_{\text{crit}}}{\hat{G}^2}\cdot\underbrace{\left.\frac{\partial\hat{G}}{\partial\mu}\right|_{n_{\text{crit}}}}_{<0} > 0$$

Stronger IO linkages *raise* the threshold: before connectivity is sufficient, higher μ amplifies cost exposure on thin intermediate markets, making it harder for the first firms to break even.

At n^{**} (thick markets): IO demand is large, so the net demand channel turns positive and—for sufficiently large μ—dominates the cost channel: $\left.\frac{\partial\hat{G}}{\partial\mu}\right|_{n^{**}} > 0$. The fixed-point map G shifts upward, raising n^{**}. Stronger IO linkages *amplify* post-threshold industrialization: each additional firm expands intermediate demand proportionally more, reinforcing further entry.

Together, IO depth simultaneously raises the entry barrier and steepens the post-threshold takeoff, generating a *threshold-plus-amplification* mechanism consistent with the convex reduced-form responses observed in peripheral districts.

G.3 Core-Periphery Simulation

The equilibrium is determined by the interaction of wages, firm counts, and market access. We solve iteratively for equilibrium firm counts, wages, and prices using trade costs τ, firm counts, and IO linkages μ_{rs} for two regions (Periphery and Core) separately.

Table A5 summarizes key parameter values. The IO linkage matrices for the weak and strong linkage cases are given in Table A6. Simulations explore the effect of varying trade costs $\tau \in [1.0, 2.0]$ and IO linkage strength on equilibrium firm counts, illustrating the model-implied threshold-type responses in the periphery and the concave, saturating pattern in the core.

TABLE A5: Parameters for the Illustrative Two-region Simulation

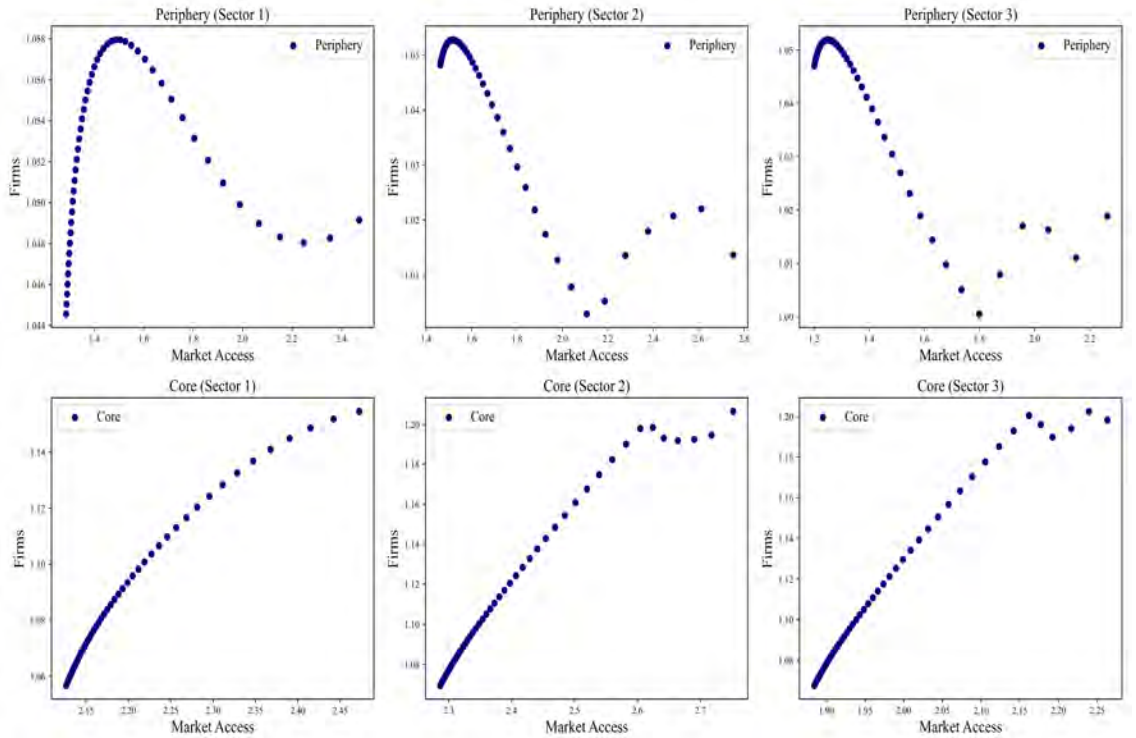
Parameter	**Description**	**Value**
σ	Elasticity of substitution	7
λ	Agglomeration elasticity	0.10
α_A	Agricultural consumption share	0.492
β_L	Agricultural labor share	0.750
F	Fixed entry cost	1.0
S	Number of sectors	3
J	Number of regions	2
η_s	Labor share in sector s	[0.52, 0.20, 0.28]
$A_{NA,\text{base}}$	Nonagriculture productivity, by region	[1.0, 1.2]
A_A	Agricultural productivity, by region	[1.2, 1.0]
K_A	Agricultural land	[1.2, 1.0]
L_j	Labor endowment	[1.0, 1.0]

TABLE A6: Input-Output Coefficients (μ_{rs}) for the Illustrative Simulation

Source r	**Sector**		
	$s = 1$	$s = 2$	$s = 3$
$r = 1$ (Primary)	0.149	0.079	0.014
$r = 2$ (Secondary)	0.139	0.241	0.097
$r = 3$ (Service)	0.074	0.117	0.157

NOTE: 2007 Vietnamese Input-Output Table (GSO).

FIGURE A7: Simulation Results: Market Access vs. Number of Manufacturing Firms



NOTE: Equilibrium manufacturing firm counts simulated across trade costs $\tau \in [1.0, 2.0]$ for the two-region (Core/Periphery), three-sector calibration described above. The model-implied U-shaped (threshold) relationship in the periphery and the concave (saturating) relationship in the core illustrate Proposition 1 in the main text.

G.4 Calibration Details

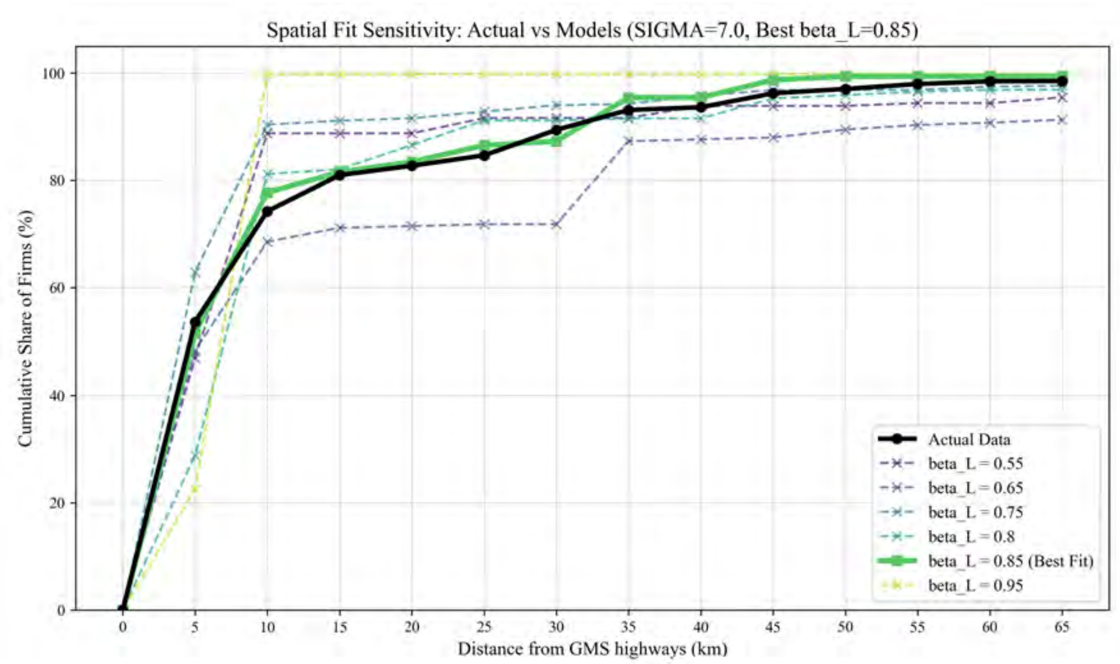
Agglomeration elasticity λ. We estimate λ by simulated method of moments, choosing the value that minimizes the discrepancy between model-implied and observed firm shares across districts:

$$\hat{\lambda} = \arg \min_{\lambda \in [0.05, 0.40]} \sum_{j=1}^{J} \left(\frac{n_j^{\text{model}}(\lambda)}{\sum_k n_k^{\text{model}}(\lambda)} - \frac{n_j^{\text{data}}}{\sum_k n_k^{\text{data}}} \right)^2 \tag{A18}$$

The model is solved under the post-upgrade transport network τ_{post} with endogenous wages. The objective is minimized at $\hat{\lambda} = 0.10$, implying that a district doubling its firm count experiences approximately $100 \times \hat{\lambda} \times \ln(2) \approx 6.9\%$ reduction in unit production costs—economically meaningful scale effects tempered by endogenous wage adjustment.

Agricultural labor share β_L. We calibrate β_L by minimizing the RMSE between the model-implied and observed cumulative distribution of manufacturing firms by distance from the GMS corridor. Figure A8 reports the RMSE surface over a grid of β_L values. The spatial fit is sensitive to β_L: values below 0.75 generate distributions that are too flat (insufficient corridor concentration), while $\beta_L = 0.95$ produces an implausible kink at 10km. The best fit across all distance bands is achieved at $\beta_L = 0.85$.

FIGURE A8: Sensitivity of Spatial Fit to Agricultural Labor Share



NOTE: CDFs of manufacturing firms by distance under alternative β_L values. "Actual Data" (black solid). Best-fit: $\beta_L = 0.85$ (green solid). Lower values ($\beta_L \leq 0.75$): insufficient corridor concentration. $\beta_L = 0.95$: implausible kink at 10km. All other parameters: $\hat{\lambda} = 0.10$, $\sigma = 7.0$.

Appendix H Counterfactual Details

H.1 Route Choice Model and Endogenous Trade Cost

We endogenize trade costs under counterfactual transport networks (CF1 and CF2) using the iceberg formulation following Fan *et al.* (2023):

$$\tau_{ij} = \exp(\kappa \cdot d_{ij}(\mathcal{G})) \tag{A19}$$

where $d_{ij}(\mathcal{G})$ is the minimum travel time between districts i and j, and κ is the trade-cost elasticity. The minimum travel time is determined endogenously by firms' optimal route choice given prevailing road speeds.

Route choice. Figure A9 illustrates the mechanism for a representative origin-destination pair (O, D) located off the GMS corridor. Firms can choose among three routes:

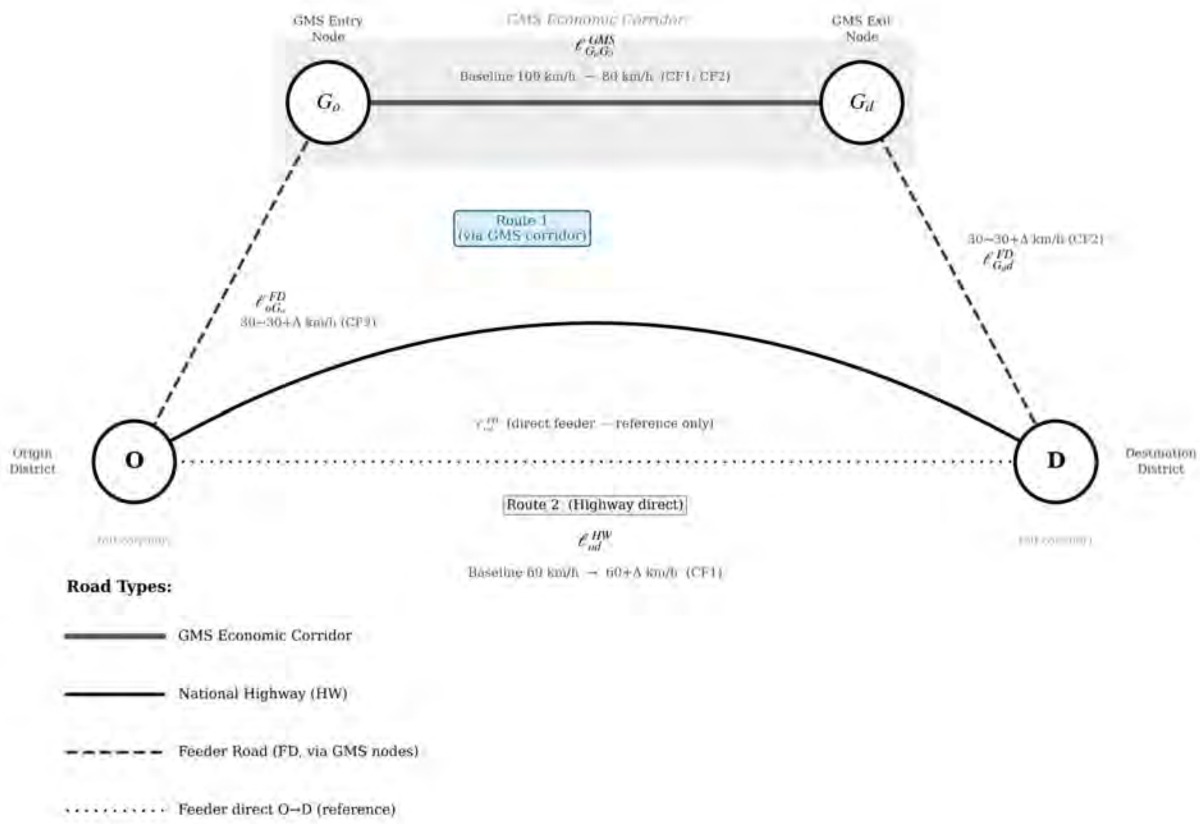
- **Route 1 (via GMS):** $O \xrightarrow{\text{feeder}} G_o \xrightarrow{\text{GMS}} G_d \xrightarrow{\text{feeder}} D$, with travel time $t^{(1)}_{OD} = \ell^{FD}_{oG_o}/v^{FD} + \ell^{GMS}_{G_oG_d}/v^{GMS} + \ell^{FD}_{G_dD}/v^{FD}$
- **Route 2 (highway direct):** $O \xrightarrow{\text{HW}} D$, with $t^{(2)}_{OD} = \ell^{HW}_{OD}/v^{HW}$
- **Route 3 (feeder direct):** $O \xrightarrow{\text{FD}} D$, with $t^{(3)}_{OD} = \ell^{FD}_{OD}/v^{FD}$

Under the actual program, the GMS corridor operates at 100km/h, making Route 1 attractive for districts close to the corridor's entry and exit nodes (G_o, G_d) despite the feeder-road detour. Route 2 is slower (60km/h) but direct, dominating for districts far from the corridor.

Under CF1, the GMS corridor is degraded to 80km/h while the budget savings are reallocated to upgrade the national highway network (v^{HW} increases). For a district pair marginally using Route 1 under the baseline, this raises $t^{(1)}_{OD}$ and simultaneously lowers $t^{(2)}_{OD}$, potentially reversing the route ranking. Through eq. (A19), districts that successfully switch to Route 2 experience lower trade costs, while those that relied on the GMS corridor but cannot switch lose locational advantage. This is the primary spatial reallocation mechanism in CF1.

CF2 additionally upgrades feeder roads (v^{FD} increases), lowering the feeder-road penalty embedded in Route 1 and benefiting inland districts that are off both the GMS corridor and the main highway network. The two-channel investment in CF2 generates broader spatial dispersion relative to CF1.

FIGURE A9: Route Choice and Trade Cost Determination



NOTE: Route choice for a representative firm shipping goods between origin O and destination D, both located off the GMS corridor. Nodes G_o and G_d are the nearest GMS entry and exit points. Trade costs: $\tau_{OD} = \exp(\kappa \cdot t^*_{OD})$ where t^*_{OD} is the minimum travel time across routes. Under CF1, GMS speed falls from 100 to 80km/h and the budget is reallocated to national highways. Under CF2, feeder roads are additionally upgraded, broadening access to inland districts.

H.2 Kappa Calibration

Objective Function and Identification. We estimate the trade cost elasticity κ by minimizing the RMSE between the model-implied and observed spatial CDF of manufacturing firms. The estimation objective $\mathcal{L}(\kappa)$ is defined as follows:

$$\mathcal{L}(\kappa) = \sqrt{\sum_d \omega_d \left(\text{CDF}_d^{\text{data}} - \text{CDF}_d^{\text{model}}(\kappa)\right)^2} + \zeta \cdot \mathbb{I}(\text{bias} > 0.5) \tag{A20}$$

where ω_d represents weights derived from a Gaussian kernel ($h = 20$km) that prioritize districts within 0–30km of the GMS corridor, where the identifying variation for κ is highest. The second term is a "sanity check" penalty ($\zeta = 3.0$) that activates if the model-implied spatial centroid (firm-count-weighted mean distance) deviates from the observed value by more than 50%, thereby excluding implausible concentration or dispersion patterns.

Two-Stage Search and Model Averaging. The optimization followed a two-stage procedure: an initial coarse evaluation over $\kappa \in [0.03, 0.15]$ with a step size of 0.01, followed by a refinement stage with a step size of 0.001 around the global minimum. To account for identification uncertainty, we also calculate a model-averaged estimate using inverse-loss weights for all candidate values within 130% of the minimum loss, which yielded a value close to our point estimate of $\hat{\kappa} = 0.101$. This value implies that a one-hour increase in travel time raises trade costs by approximately $e^{0.101} - 1 \approx 10.6\%$, consistent with gravity-model estimates in the region. Specifically, Elhan-Kayalar *et al.* (2025) estimate a trade cost elasticity of 0.134 for the GMS expressway network in Yunnan, China—a setting comparable to ours in terms of geography and policy.