Centre for Economic Performance

Discussion Paper ISSN 2042-2695

No. 2062
December 2024

Roads to development? Urbanization without growth in Zambia

Cong Peng
Yao Wang
Wenfan Chen

LSE THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

UKRI Economic and Social Research Council

Abstract
This study explores the impacts of road improvements in a country characterized by "urbanization without growth." Our analysis reveals that, although road upgrades increase population growth, they do not significantly advance economic development and tend to worsen living conditions. Utilizing a combination of empirical evidence and a spatial equilibrium model, we identify that constrained industrial capacities and congestion from high population density limit the efficacy of road development policies in enhancing GDP and overall welfare. Our results also indicate that strategically targeting road placement in regions with higher economic productivity could yield better economic outcomes.

Key words: road improvements, urbanization, industrialization, quantitative spatial model, satellite imagery, Africa
JEL: O1; R1; R4

This paper was produced as part of the Centre's Urban Programme. The Centre for Economic Performance is financed by the Economic and Social Research Council.

We gratefully acknowledge funding from the International Growth Centre (IGC) at LSE. We also appreciate the IGC Zambia country office for facilitating coordination with local entities. Special thanks to IGC Country Economist Benjamin Shawa for spearheading strong engagements and interviews with Zambian ministries to enrich understanding of the Zambian context and Miljan Sladoje for his assistance in developing the first research proposal and collecting data from the Road Development Agency that launched this research. We are thankful to the Road Development Agency and Zambia's Central Statistical Office for their data and support. We would like to thank Vernon Henderson, Gordon Hanson, Gordon McCord, Matthew Turner, Frédéric Robert-Nicoud, Qinghua Zhang, Zhi Wang, Jianfeng Wu, Cheng-keat Tang, Alex Rothen-berg, and Margaret Jodlowski for their insightful comments and suggestions. We also appreciate the valuable feedback from seminar participants at the Urban Economic Association, COMPIE, Harvard University, Université de Genève, Nanyang Technological University, Fudan University and Peking University. Additionally, we gratefully acknowledge Yulu Tang for her important contributions to this research. This paper was previously circulated under the title: Roads to Development? Examining the Zambian Context Using AI-Sat.

Cong Peng, Peking University and Centre for Economic Performance at London School of Economics. Yao Wang, Ohio State University. Wenfan Chen, J.P. Morgan.

Published by
Centre for Economic Performance
London School of Economic and Political Science
Houghton Street, London WC2A 2AE

1 Introduction

Policymakers often view investment in transportation infrastructure as a critical catalyst for economic growth. Research across various contexts demonstrates that road construction enhances welfare through mechanisms such as the multiplier effect, reducing trade and migration frictions, and channeling resources to more productive urban areas (Duranton and Turner, 2012; Allen and Arkolakis, 2014; Redding and Turner, 2015; Donaldson, 2018; Fajgelbaum and Schaal, 2020). However, in many least-developed regions, key drivers of such benefits may be absent. In many African countries, road constructions are frequently outsourced to foreign firms (United Nations Economic Commission for Africa, 2016) and often prioritizes political motives over strategic economic development, limiting local spending and weakening the multiplier effect (Burgess et al., 2015; McCartney, 2023). Moreover, many African urban centers function primarily as consumption hubs lacking a robust manufacturing base, with a significant portion of the workforce engaged in non-tradable sectors, which constrains their production capacity and hampers formal employment growth (Glaeser, 2014; Gollin et al., 2016; Venables, 2017). This phenomenon has been extensively documented in the literature as "urbanization without growth" (Fay and Opal, 2000; Henderson and Kriticos, 2018; Henderson and Turner, 2020). In such contexts, road improvements may yield less economic growth than in more industrialized countries. Therefore, road improvements alone may not spur substantial economic growth. Instead, priority should be given to developing tradable sectors, human capital, and other factors that can counteract the mechanisms behind "urbanization without growth."

This paper uses high-resolution satellite imagery enhanced by deep learning techniques, combined with a quantitative spatial model, to assess the economic and welfare impacts of a large-scale road infrastructure project in Zambia. Over the last decade, Zambia has invested heavily in road construction and maintenance, committing billions of dollars to major initiatives like the *Link Zambia 8000* program. This ambitious project aimed to upgrade and expand 8,000 kilometers of roads nationwide to improve trade routes, boost productivity, and attract foreign investment. Between 2011 to 2017, road infrastructure accounted for 42% of Zambia's expenditure on non-financial assets. However, this massive investment does not seem to be achieving significant economic success. In 2022, the IMF criticized Zambia for "years of economic mismanagement," highlighting the road-building program as an example of a public investment drive that failed to deliver substantial growth (IMF, 2022). This calls for a more timely evaluation of the effects of road projects and a deeper understanding of the role of road infrastructure in economic development in low-income contexts.

The first challenge in evaluating the economic impacts of road construction in the setting is the lack of high-resolution, localized data on road quality and economic outcomes. Such data is crucial since road impacts are often highly localized, redistributing economic activity at a micro level. While areas with direct road access may gain benefits like better product access, new logistics services (Duranton and Turner, 2012), and easier school commutes (Adukia et al., 2020), locations just 20 km away may still face last-mile access issues. Yet, economic outcomes are seldom monitored at such fine spatial resolutions, and panel data on road quality over long periods is especially scarce.[1] To address this challenge, we employ satellite imagery and recent AI advancements to identify and classify road surfaces (paved or

[1]Notable exceptions include the work of Gertler et al. (2024) on road maintenance in Indonesia and the more recent study by Currier et al. (2023), which utilizes Uber data in a developed country context.

unpaved) and detect buildings at near-building levels across Zambia, dating back to 2009. Specifically, using 6,000 proprietary high-resolution satellite images from Planet Lab, we trained deep learning models built on Facebook's to classify billions of pixels at a 5-meter resolution, producing highly granular datasets on road conditions and economic development for 2009, 2014, and 2019.

The second challenge is addressing endogeneity concerns due to unobserved location-specific factors influencing both roads and economic growth, as well as reverse causality, where roads are built in anticipation of future growth. Moreover, following the literature on road construction evaluation, we utilize market access—a measure of accessibility between locations weighted by economic mass—as a summary indicator of road development (e.g., Donaldson, 2018; Faber, 2014; Michaels, 2008). However, changes in market access are influenced not only by road improvements but also by shifts in economic activity in neighboring areas, which are themselves outcomes of interest, potentially reversing the causal direction. To address these issues, we adopt the newly developed identification strategy of Jedwab and Storeygard (2021), using variations in market access induced by distant road improvements as an instrumental variable (IV). [2] Specifically, we construct the IV by calculating the exogenous component in the changes in market access based on road upgrades at least 200 km away from the target area, and excluding economic growth in the neighboring locations. This approach alleviates reverse causality concerns and minimizes the influence of local factors that could introduce endogeneity.[3]

Leveraging the innovative panel dataset and the instrumental variable strategy based on remote road variations, we find that a 1% increase in the growth of market access from road improvements leads to a 3.8% increase in the growth of built-up areas. However, this is accompanied by a negative, though statistically insignificant, effect on the nighttime light growth. Interpreting built-up areas as a proxy for population and nighttime light as a proxy for income, we infer that improved connectivity through road development stimulates local population growth without necessarily raising income growth.[4]

We further explore the impact of increased market access on living standards and environmental outcomes. Using Demographic and Health Survey (DHS) data from 2007 and 2018/19, we analyze how road upgrades affect living conditions. Our findings reveal that higher market access growth leads to increased crowding in housing and worsens household health, suggesting that road development may exacerbate urban living conditions. To assess environmental quality, we use Aerosol Optical Depth (AOD) as a measure of air pollution (Gendron-Carrier et al., 2018) and find that a 1% increase in market access growth increases air pollution by 1.1%, consistent with findings from East Africa (Singh et al., 2020). Our

[2] Recent literature has addressed endogeneity concerns through various identification strategies, including (a) using historical road construction plans as instruments, under the assumption that these plans are unrelated to current conditions (Baum-Snow et al., 2020, 2017; Baum-Snow, 2007; Duranton and Turner, 2012); (b) using least-cost road networks derived from physical geography as an instrument, assuming that physical geography is unrelated to present-day conditions (Faber, 2014); (c) exploiting regions that are incidental beneficiaries of new infrastructure links (Gibbons et al., 2019; Chandra and Thompson, 2000; Ghani et al., 2015); and (d) calculating market access measures using variations in market access induced by roads built far away (Jedwab and Storeygard, 2021; Jaworski and Kitchens, 2019), inspired by the approach of using "friends of friends" to estimate peer effects in social network literature.

[3] We choose this method due to the unique political context in the country, where road construction has been reported to prioritize political motives over strategic economic development, resulting in projects that are broadly dispersed rather than concentrated in key corridors (McCartney, 2023). Newly paved or rehabilitated improvements identified in our study are spread across the country, supporting anecdotal evidence that road investments are influenced by election-driven considerations.

[4] Although night lights may not serve as a reliable proxy for growth, especially in non-urban areas (Gibson et al., 2021), our analysis centers on urban areas and formal income. Our primary focus is on assessing whether road infrastructure can support the expansion of the tradable sector, a critical driver for sustaining long-term economic growth.

results align with the phenomenon of "urbanization without growth," where cities expand in population and physical size but lack corresponding economic development, income growth, or improvements in living standards, often accompanied by environmental degradation.

Why are roads less effective in countries experiencing "urbanization without growth"? The literature identifies two main drivers behind the this phenomenon: (1) limited manufacturing development—few productive sectors exist to absorb additional labor supply and create job opportunities; and (2) urban congestion—population growth often reduces living standards by overwhelming inadequate infrastructure and placing excessive strain on public goods and services. We examined whether these factors could shed light on the limited impact of Zambia's road-building initiative. Our findings suggest that improved market access through road development does not significantly stimulate economic growth, even in regions with comparatively stronger manufacturing bases. This indicates that the roads may not be effectively supporting the movement of tradable goods, as manufacturing in these areas remains in the early stages of development. Additionally, in more congested provincial capitals, higher growth of market access significantly reduces the growth of nighttime light intensity, a proxy for economic activity. As roads encourage population concentration in these cities, these results suggest that congestion effects may indeed dampen the economic benefits of road investments.

Our reduced-form results compare changes in outcomes between locations with larger versus smaller increases in market access, which essentially eliminates any effect that is common to both groups. As a result, these reduced-form results alone cannot capture the aggregate and distributional effects of the road upgrades. To quantify its overall welfare and growth effects, better understand the mechanisms, and explore potential improvements, we apply a quantitative spatial model based on Allen and Arkolakis (2018), Fan (2019), and Tombe and Zhu (2019) to depict the economic geography of Zambia. The model encompasses both agricultural and non-agricultural sectors across multiple locations linked by costly trade and migration. It incorporates productivity and amenity variations across locations and sectors, influenced by congestion and agglomeration effects. Using sectoral wage and population data from the World Bank Living Conditions Monitoring Survey and Census data, along with parameters calibrated from existing literature, we estimate key model parameters and infer local productivity and amenity levels. We then employ the calibrated model to conduct policy simulations.

Our findings reveal that road improvement policies in Zambia yield only modest aggregate gains, increasing national GDP by 0.07% and welfare by 0.2%, while notably reducing the share of non-agricultural GDP and employment. These limited outcomes largely result from inefficient targeting of road investments: the current approach favors more populous areas over regions with higher productivity or amenity levels. In a country with a limited industrial base, this strategy is unlikely to drive structural transformation or meaningful economic growth. Consistent with our reduced-form mechanism analysis, simulations indicate that strong congestion effects and weak agglomeration benefits due to lack of manufacturing activities are key factors constraining the policy's effectiveness.

Finally, we explore potential improvements to the current policy by conducting simulations with our model. The results indicate that targeting road improvements in the most productive locations, as identified through model inversion, produces significantly better outcomes—yielding more than three times the GDP growth and over twice the welfare gains compared to the current policy. Better targeting also modestly increases the non-agricultural share of both national employment and GDP, signaling

progress toward structural transformation. Additionally, when simulating the effects of sequentially reducing travel times across different locations, we find that policies focused on areas with higher productivity, larger populations, and greater initial employment in the non-agricultural sector deliver the most substantial improvements in GDP, welfare, and structural change.

The paper is organized as follows: Section 2 summarizes the related literature. Section 3 provides an overview of Zambia's context and transportation network. Section 4 describes the data sources, and Section 5 outlines the empirical approach. Section 6 presents causal findings based on the innovative panel data constructed for this study. Section 7 explores the underlying mechanisms using factors behind the "urbanization without growth" phenomenon. Section 8 introduces a spatial equilibrium model that clarifies these mechanisms and allows for counterfactual analysis. Finally, Section 10 concludes with remarks.

2 Related Literature

Our paper contributes to the extensive literature evaluating the effects of road construction on economic growth. While roads are generally viewed as critical for economic development, evidence on their impacts varies. Some studies find that road development brings significant benefits, such as increased industrial output (Gibbons et al., 2019; Ghani et al., 2015), urban population growth (Jedwab and Storeygard, 2021), reduced transaction costs (Storeygard, 2016), and higher school enrollment rates (Adukia et al., 2020). However, other research suggests more complex outcomes. For instance, Asher and Novosad (2020) and Banerjee et al. (2020) report limited effects on economic growth in rural India and China, respectively, while Rothenberg (2013) finds that road improvements in Indonesia significantly increase the spatial dispersion of manufacturing activity, limiting agglomeration benefits. Additionally, road-building can have uneven impacts across regions and populations. Baum-Snow et al. (2020) and Faber (2014) show that highway construction in China led to economic declines in hinterland cities, while Fretz et al. (2021) finds that highways in Switzerland exacerbate income segregation. These diverse findings highlight the complex, context-dependent nature of road infrastructure's effects on economic development.

Our paper extends this literature by focusing on the contexts of the least developed countries, which are characterized by a weak industrial base and severe constraints on public resources. We employ a combination of reduced-form analysis and a quantitative spatial model to examine the factors driving the limited effectiveness of road improvement policies in such settings and to explore potential avenues for improving their design and implementation. The closest study to ours is Gertler et al. (2024), which examines how road quality influences welfare outcomes in Indonesia, a lower-middle-income country with more industrial activity than Zambia.

Our paper also contributes to the growing literature that leverages high-resolution satellite imagery (Henderson et al., 2012; Baragwanath et al., 2019) and deep learning (Khachiyan et al., 2021; Burke et al., 2020) to analyze transportation infrastructure. Satellite data provide valuable, consistent, and accessible information across both time and space, even in regions with limited traditional data sources (Donaldson and Storeygard, 2016; Henderson et al., 2012). These tools have been applied to identifying urban markets (Baragwanath et al., 2019), tracking air pollution effects (Jayachandran, 2009), monitoring de-

forestation (Burgess et al., 2012), slum investments (Marx et al., Forthcoming), and climate impacts on crops (Costinot et al., 2016). Our study uses AI to construct Zambia's road network and detect road improvements from 2009 to 2019, introducing scalable methods for transportation data in developing countries. While most studies emphasize road construction, relatively few focus on the effects of road maintenance aside from Gertler et al. (2024). By leveraging AI and satellite data, we can consistently measure road surface conditions across the country over time, tracking changes in road quality due to both new construction and maintenance activities. Furthermore, we use AI to predict the expansion of built-up areas—serving as a proxy for population growth—at an unprecedented 5-meter resolution, providing new insights into economic dynamics at a near-building level.

3 Background

Zambia gained independence from the United Kingdom in 1964 and has maintained political and socioeconomic stability since. The founding president's slogan, "One Zambia, One Nation," fostered unity among the country's 70+ ethnic groups. As one of the earliest African nations to gain independence, Zambia also supported its neighbors' liberation movements, contributing to regional stability. Strategically located at the crossroads of Southern and Central Africa, Zambia is Africa's second-largest copper producer and is rich in resources like gold, cobalt, manganese, and nickel. In response to the global energy transition, the government aims to increase copper production from 800,000 to 3 million metric tons annually by 2031 (Vandome, 2023; Chifunda, 2024). Positioned as a key transit hub for trade corridors within Common Market for Eastern and Southern Africa (COMESA) and Southern African Development Community (SADC), Zambia relies on its eight neighboring countries for port access, making robust transportation infrastructure crucial for enhancing its trade competitiveness.

History of Road Building Over three-quarters of traded goods in Zambia are transported by road or rail. Historically, railways were the main mode for copper transport, but since the 1990s, road freight has taken precedence, now carrying over 70 percent of Zambia's goods trade (Raballand and Whitworth, 2012; UNCTAD, 2011).[5] Annually, around 4 million metric tons of copper pass through Zambia, primarily (3 million metric tons) from the Democratic Republic of Congo (DRC). The heavy loads from copper shipments and large mining equipment put substantial strain on roads, compounded by poor-quality pavement and seasonal flooding, accelerating infrastructure deterioration (Raballand and Whitworth, 2012). According to data from the Road Development Agency, roughly 20 percent of the Core Road Network deteriorated between 2008 and 2015, as evidenced by rising International Roughness Index (IRI) scores—a standard measure of road surface roughness, where higher values indicate rougher conditions (see Appendix Figure A.1).[6]

[5]Local reports emphasize the urgent need for road improvements to support economic growth and highlight transportation's critical role in fostering development (European Commission, 2020).

[6]The International Roughness Index (IRI) is a standard metric developed by the World Bank to evaluate road surface smoothness or roughness by measuring the vertical movement of a vehicle traveling over a road segment. Expressed in meters per kilometer (m/km), lower IRI values indicate smoother roads, while higher values denote rougher surfaces. This index is widely used in road maintenance and infrastructure management to assess road quality and prioritize repairs or improvements.

In the mid-2000s, a combination of debt relief and rising copper prices positioned Zambia as a fiscally resourced country with an improving economic outlook.[7] Building on this fiscal space, in 2011 the newly elected Patriotic Front (PF) government, led by President Michael Sata, adopted an expansionary fiscal policy focused on infrastructure development. In 2012, the *Link Zambia 8000* project launched to improve connectivity by constructing and upgrading 8,200 km of roads at an estimated $5.46 billion (Plan, 2017). Other major initiatives included the *Lusaka Urban Road (L400)* project to upgrade 402 km of urban roads in Lusaka (AidData, 2024) and *Pave Zambia 2000*, which aimed to rehabilitate 2,000 km of urban roads using interlocking paving technology (Development Minerals, 2024). However, this surge in spending was not matched by revenue growth; between 2011 and 2014, average revenues remained at 17% of GDP while expenditures rose from 20% to 23% of GDP (Mbewe et al., 2024).

In its early years, the *Link Zambia 8000* program surfaced an average of 260 km of roads annually, reaching 666 km by 2015. However, this pace slowed sharply from 2016 as Zambia's economic conditions deteriorated due to a severe drought impacting hydropower-dependent electricity generation and falling copper prices that slashed export earnings (Mbewe et al., 2024). By 2019, only 830 km of roads had been completed (Road Development Agency, 2019). Despite limited progress, interviews with stakeholders suggest most allocated funds were spent. Similarly, the *Pave Zambia 2000* initiative struggled, surfacing just 8.5 km of roads by 2017 (Road Development Agency, 2017). These road expenditures significantly increased Zambia's debt, which reached 140% of GDP in 2020, leading it to become one of the first African nations to declare bankruptcy during the COVID-19 pandemic (Statista, 2024; Mbewe et al., 2024). Since taking office in 2021, President Hakainde Hichilema has focused on debt restructuring and austerity measures, including scaling back road projects (MHID, 2020; Vandome, 2023).

Misallocation Concerns Critics argue that political motivations, rather than cost-effectiveness or evidence-based prioritization, drive road development and spending in Zambia. The lack of reliable, geo-referenced data on road usage, deterioration, and improvements challenges effective planning. Roads, as highly visible public services, often serve as tools for incumbents seeking re-election. Marx (2018) shows that incumbents in Sub-Saharan Africa frequently gain electoral rewards for completing visible infrastructure projects, creating incentives to prioritize these projects near election times. Harding and Stasavage (2014) finds in Ghana that a one-standard-deviation improvement in road conditions increases the incumbent party's vote share by one-quarter of a standard deviation. Burgess et al. (2015) find that in Kenya, road-building investments increase significantly in the constituencies of political leaders. Additionally, traditional leaders or Chiefs influence road placement in rural areas. While the Road Development Agency manages trunk roads, the Ministry of Local Government oversees intraprovincial roads, with Chiefs playing key roles in rural project decisions. Research suggests that Members of Parliament (MPs) engage in clientelism, working with Chiefs to secure local development support, as Chiefs hold sway over land use, voluntary labor, and project monitoring (Baldwin, 2013).

Despite these challenges, Zambia's road network expanded by over 2,900 km between 2012 and

[7]Debt relief programs reduced Zambia's debt from a peak of US$7.1 billion (approximately 130 percent of GDP) in 2004 to US$500 million (25 percent of GDP) by 2006. From 2006 to 2013, the debt-to-GDP ratio remained low, fluctuating between 19 and 26 percent (IMF, 2024). Simultaneously, rising copper prices fueled a boom in the mining and quarrying sector, with its GDP contributions increasing from about 10 percent in 2000 to 15 percent in 2010 (Mbewe et al., 2024).

2019, warranting a closer analysis of its economic impact.[8] Urban areas also saw significant expansion, with over 850 km of new roads added across Lusaka, Copperbelt, and Central Provinces.[9]

4 Data

4.1 Spatial Units

The basic spatial unit in this study is the hexagon, based on a global spatial indexing system that divides the Earth's surface into hexagons at multiple resolutions.[10] This indexing system provides a consistent framework, enabling uniform spatial boundaries globally. For our analysis, each hexagon has a side length of approximately 10.8 km. Panel (a) in Appendix Figure A.2 shows Zambia's administrative boundaries at the provincial and district levels, while Panel (b) overlays the hexagons, demonstrating their relative sizes compared to these administrative boundaries and highlighting the gain in spatial granularity. Using hexagons as the spatial unit offers two key advantages. First, survey data, such as from the Demographic and Health Surveys (DHS), are spatially displaced up to 2 km in urban areas and 5 km in rural areas for confidentiality. Aggregating data to the hexagon level reduces the impact of this displacement, stabilizing the spatial referencing of survey clusters. Second, nighttime light data, often used as a proxy for economic activity, loses predictive accuracy when aggregated into smaller units (Gibson et al., 2021). Aggregating to the hexagon level enhances its reliability as an economic indicator.

For urban-specific analysis, we define urban boundaries using delineation methods similar to those in Baragwanath et al. (2019). These boundaries serve as spatial masks for urban-focused outcomes like built-up areas, air pollution, and green space, which we aggregate within urban boundaries before further aggregating to the hexagon level. We use this urban mask to exclude very small rural settlements because the quality of satellite data for these areas is often poor. This approach ensures more accurate and reliable analysis by focusing on regions where satellite imagery is sufficiently detailed and clear. [11]

4.2 Remote Sensing-based Outcomes

We use satellite data to measure key outcome variables: built-up area, nighttime light, Aerosol Optical Depth (AOD) for air quality, and the Normalized Difference Vegetation Index (NDVI) for green cover across three time periods—2009, 2014, and 2019. Built-up area data is further refined using AI techniques. These variables vary in temporal frequency, with some available monthly and others annually. To maintain consistency in our regression analysis, we aggregate all variables to the hexagon-year level.

AI-predicted Built-up We identify built-up areas at an unprecedented 5-meter resolution for 2009, 2014, and 2019 using Planet RapidEye, a high-resolution satellite imagery product from Planet Labs. The RapidEye constellation, comprising five satellites, captures imagery focused on agriculture, forestry,

[8] Based on RDA reports from 2013–2019, about 2,900 km of roads were rehabilitated or upgraded. See Road Development Agency (2023) for the reports.

[9] The *Lusaka Urban Roads (L400)* project added over 662 km, while the Zambia Township Roads project contributed 145 km in Copperbelt and 43 km in Central Province by 2019 (Road Development Agency, 2019).

[10] This system, widely used by companies like Uber, supports 16 levels of resolution, from large hexagons covering 1,107 km² to smaller ones of 0.5 m².

[11] Please see Appendix B for details. Examples of these urban masks are shown in Appendix Figure B.17.

and environmental monitoring. Each image includes five spectral bands: Blue (440-510 nm) for water and soil-vegetation distinction, Green (520-590 nm) for chlorophyll sensitivity, Red (630-685 nm) for vegetation health, Red Edge (690-730 nm) for detecting vegetation stress, and Near-Infrared (760-850 nm) for healthy vegetation and NDVI calculations (see Appendix Figure A.3 for a 25 km $\times$ 25 km example). The images are orthorectified and corrected for radiometric and sensor distortions, making them suitable for deep learning applications.

To identify built-up areas, we trained a deep learning model using Facebook's 30-meter population density map as labeled data.[12] The model was trained to identify pixel patterns that suggest the presence of buildings, allowing it to classify each pixel in satellite images from earlier years as either built-up or not. Appendix Figure A.4 demonstrates the model's sensitivity in detecting built-up structures in Lusaka, Zambia's largest city. The right panel shows AI-predicted built-up areas, with red indicating structures from 2009, blue representing new developments from 2009-2014, and green marking growth from 2014-2019. The left panel compares Google Earth images from 2009 and 2019 in the area circled in yellow, validating the model's accuracy. Notably, existing built-up area datasets that cover multiple years, such as the Global Human Settlement Layer (GHSL) developed by the European Commission, are derived from 10-meter Sentinel or 30-meter Landsat data (Pesaresi et al., 2023). In contrast, our AI-predicted built-up data achieves a consistent 5-meter resolution across all years, significantly enhancing spatial detail.

Nighttime Light We measure economic activity using NASA's recently released Black Marble nighttime light data, which is specifically processed for monitoring human activities, offering enhanced comparability across space and time relative to the widely used VIIRS data from NOAA (Román et al., 2018).[13] Appendix Figure A.5 compares Zambia's national GDP with trends from both VIIRS and Black Marble data. Unlike VIIRS, which shows a consistent upward trend regardless of GDP fluctuations, Black Marble closely mirrors GDP movements, including the decline starting in 2015. This suggests that Black Marble data is more effective at capturing time series variations in economic activities than VIIRS. To our knowledge, this study is the first to incorporate Black Marble data in economic research. We use data from May to August (Zambia's cool, dry season) for the years 2012-2019 to reduce cloud cover interference. While nighttime light data is unavailable for 2009, we utilize 2012 data as a proxy.

The Economic Meanings of Built-up and Nighttime Light Data Satellite data, including measures of built-up areas and nighttime lights, have become widely used proxies for economic growth in economics and urban planning research (Henderson et al., 2012; Chen and Nordhaus, 2011; Pinkovskiy and Sala-i Martin, 2016; Storeygard, 2016; Donaldson and Storeygard, 2016). To better interpret our results and understand the economic meanings of our built-up and nighttime light data, we assess their correlation

[12] Facebook's map designates 30-meter grid cells as built-up if buildings are detected (Tiecke et al., 2017). However, this map is only available for around 2019. Our approach extends this classification to earlier years and significantly enhances spatial resolution. The data is accessible here: https://dataforgood.facebook.com/dfg/tools/high-resolution-population-density-maps.

[13] NASA's Black Marble product suite enables global mapping of nighttime lights (NTL) linked to human-driven patterns, such as electrification, economic conditions, and disruptions caused by disasters. It also provides insights into seasonal variations in human activity and urbanization. While both Black Marble and NOAA's VIIRS data originate from the same instrument, they differ in processing methods.

with key economic indicators: firm employment and revenue. We use firm-level data from the 2011 National Business Establishments Register, which provides detailed information such as firm locations (coordinates), year of establishment, primary economic activities, and annual turnover. Appendix Figure A.8 illustrates that nighttime light data from 2012 significantly correlates with firms' aggregate revenue, while AI-predicted built-up areas are closely associated with aggregate employment at the hexagon level. Building on these findings, we interpret nighttime lights as a proxy for economic output and built-up areas as a proxy for population, consistent with established practices in the economics literature (Chen and Nordhaus, 2011; Henderson et al., 2012). It is worth noting, however, that employment data may not perfectly represent population levels due to the high prevalence of informal employment in Africa.[14] We also examine changes in built-up areas and nighttime light over 2009–2019. Panel (a) of Figure 1 shows a rightward shift in the distribution of built-up areas, indicating an expansion in development over time. Panel (b) of Figure 1 illustrates changes in nighttime light distribution, revealing a modest increases in economic activity.

Air Quality and Green Cover We measure air quality using Aerosol Optical Depth (AOD) data, following Gendron-Carrier et al. (2018). AOD data is sourced from the Moderate Resolution Imaging Spectroradiometers (MODIS) aboard the Terra and Aqua satellites, providing daily measurements of atmospheric aerosol concentration at a 3 km resolution, making it a valuable proxy for urban air pollution.[15] Over the 2009–2019 period, air pollution levels exhibit fluctuations, with a general increase from 2009 to 2013 followed by a decline from 2013 to 2019, as shown in Panel (c) of Figure 1.

We use the Normalized Difference Vegetation Index (NDVI) to monitor changes in forest and crop cover since 1999. NDVI measures the difference between visible and near-infrared reflectance, providing an estimate of vegetation density. To distinguish between forests and croplands, we supplement the NDVI data with 300-meter resolution Global Land Cover data from 2009, provided by the European Space Agency. Panel (d) of Figure 1 reveals a slight increase in green space over the study period.

4.3 Road Network Data

Digitized Road Network We begin by digitizing and mapping Zambia's Core Road Network (CRN), the road network essential for the country's socioeconomic development, based on planned construction and rehabilitation projects from the *Link Zambia 8000* initiative provided by the Road Development Agency (RDA).[16] Existing road segment locations were primarily sourced from OpenStreetMap (OSM), while planned roads not yet on OSM were verified using Google Earth satellite imagery. Appendix Figure A.6 displays the resulting road map, with different colors indicating road types based on function and importance. The CRN comprises Trunk (T), Main (M), District (D), Primary Feeder (PF), and Urban (U) roads. Trunk roads connect major cities and borders, supporting long-distance trade; Main roads link provincial capitals and towns for regional transport; District roads connect rural areas to nearby

[14] According to the 2015 Living Conditions Monitoring Survey, 88.7% of Zambia's employed population works informally. Nearly all workers in agriculture, fishing, and forestry are informally employed, with informality especially high in rural areas (97% of workers) and still significant in urban areas, where fewer than one in four jobs are formal (Tassot et al., 2019).

[15] AOD quantifies the extent to which aerosol particles prevent direct sunlight from reaching the ground, offering a comprehensive measure of overall air pollution.

[16] Maintaining the quality of this network is a primary responsibility of the RDA.

towns; Primary Feeder roads support movement from agricultural areas to main networks; and Urban roads serve local traffic within towns and cities. In total, the road network covers a total of 40,113 km as of 2015. Our analysis focuses on the Trunk, Main, and District roads most relevant for inter-city and cross-border transport.

AI-predicted road surface condition We processed approximately 100 billion pixels from 6,000 high-resolution Planet RapidEye satellite images, provided by Planet Labs, to identify and classify road segments in Zambia across three key time periods. These are the same images used for identifying built-up areas. Using an AI-based deep learning model built upon Facebook's (now Meta) approach (Bonafilia et al., 2019), we classified roads into three categories: paved roads, dirt (or earth) roads, and no roads. The model was trained using 2019 OpenStreetMap (OSM) data and satellite imagery from 2018 to 2020. After training, we applied the model to satellite images from 2009-2011 and 2014-2016, aligning these periods with the 2009 and 2014 data used in our study. This method allows us to track changes in road surface conditions over time and construct a historical road network, with a continuous score, ranging from zero to one, indicating the likelihood that a surface is paved. Appendix Figure A.7 shows AI-predicted road scores for 2009 and 2019 in panels (a) and (b), with panel (c) illustrating changes between these years.

To validate the model's performance, we cross-checked the results against the 2015 Road Condition Survey (RCS 2015) conducted by Zambia's Road Development Agency (RDA). The RCS 2015 provides detailed, geo-referenced data on road conditions by segment, which we used to map road conditions across the digitized Core Road Network (CRN). Out of 930 roads surveyed, 436 were officially classified as paved by RCS. We were able to match 429 roads surveyed by RCS with our AI-predicted road segments. Among them, 58 are earth roads, 73 are gravel roads, and 298 are paved roads according to the RCS definition.[17] Appendix Figure A.9 compares the AI-predicted road condition scores from 2014/16 to the actual surface types recorded in the RCS 2015. The results indicate that AI-predicted scores for earth and gravel roads, as defined by the RCS, cluster near zero, with median values of 0.01 and 0.02, respectively. In contrast, paved roads have a significantly higher score distribution, with a median of 0.52. Some discrepancies, such as outliers in the earth and gravel categories, may stem from differences in time periods; for example, our AI scores rely on 2016 imagery, while the RCS data is from 2015, suggesting possible road upgrades between 2015 and 2016. Additionally, the wider range of AI-predicted scores for RCS-defined paved roads suggests that certain segments classified as "paved" by the RCS may, in fact, contain unpaved sections.

4.4 Household Surveys

We use the Demographic and Health Surveys (DHS) for data on household measures, primarily the survey years 2007 and 2018/19.[18] All waves of the surveys are geo-referenced, making them well-suited for integration with our fine-grained spatial market access data. The DHS data has two limitations for

[17] Some roads could not be matched due to errors in geocoded nodes and links.

[18] The 2018/19 survey includes successful interviews with a nationally representative sample of 13,683 women aged 15-49 and 12,132 men aged 15-59, across 12,831 households. The 2007 survey covers a smaller sample, with interviews of 7,146 women aged 15-49 and 6,500 men aged 15-59, also nationally representative. Additionally, we briefly use data from the 2013-14 survey for certain cross-sectional analyses.

causal inference. First, it is a cross-sectional dataset of households rather than a panel. Second, for confidentiality, the survey locations, or DHS clusters, are randomly displaced up to 2 km in urban areas and up to 5 km in rural areas. This can introduce inaccuracies when aggregating DHS clusters into hexagons, as some clusters may shift across hexagonal boundaries. To address this, we developed an innovative approach to convert the data into a location-based panel. Given that DHS survey clusters are displaced by no more than 5 km, and our hexagons have a side length of 10.8 km, a displaced cluster could only move as far as a neighboring hexagon. Thus, we group adjacent hexagons that were surveyed at least once in both survey years into what we call Stable Spatial Panel Units (SPUs). Within each SPU, we aggregate key outcomes, independent variables, and covariates by calculating their mean values, allowing us to track changes in these variables at the SPU level over time. For DHS outcomes, we also compute the total number of respondents in each SPU and use these totals as weights in regression analyses when these outcomes serve as dependent variables. One limitation of this approach is that it may reduce sample representativeness, potentially introducing bias toward larger locations that are more frequently surveyed.

5 Empirical Strategy

5.1 Construction of Market Access

To measure road improvement, we construct market access, M_{it}, using AI-predicted road networks and built-up area data, following the methodology of Donaldson and Hornbeck (2016) and Jedwab and Storeygard (2021). The unit of analysis is a hexagon, each fully contiguous and non-overlapping, covering the entire country. This market access measure reflects an economic-mass-weighted average of accessibility between locations. We define market access for hexagon i in period t as follows:

$$M_{it} = \sum_{j \neq i} N_{jt}(1 + t_{ijt})^{-\theta} \tag{1}$$

where N_{jt} represents the built-up area of hexagon j in period t, and t_{ijt} denotes the travel time between hexagons i and j in period t, which we discuss in the next paragraph. The parameter θ captures the elasticity of trade flows with respect to travel time, typically ranging from -0.64 to -1.39 based on existing literature.[19] We report results using $\theta = -1$ and examine the effects across this range for sensitivity analysis.

We construct travel time, t_{ijt}, in an innovative way that accounts for road quality. Specifically, travel time is calculated as:

$$t_{ijt} = \sum_{k \in ij} (D/v(s_{kt}, c_{kt})) \tag{2}$$

where k represents the hexagons along the route from i to j, and D is the distance between the centroids

[19] In Donaldson (2018), θ is derived from the product of the elasticity of trade flows to trade costs and the elasticity of trade costs to travel time. The former is estimated at 3.8 for India and 8.22 for historical USA in Donaldson and Hornbeck (2016), while the latter is 0.169 in Donaldson (2018). Consequently, θ usually ranges from -0.64 to -1.39, though adjustments are needed for the African context. We initially set θ to -1 but explore values from -0.5 to -2.5 in increments of 0.5 for robustness checks.

of two hexagons, set as a constant 18.7 km based on their size. The travel speed $v(s_{kt}, c_{ijtk})$ depends on the maximum allowable speed for the highest-ranked road segment, s_{kt}, between i and j in period t, and the road condition score, c_{kt}, derived from the deep learning model. This score represents the probability that a road segment is paved. Specifically, travel speed is calculated as $v_{kt} = (s_{kt} - 20)c_{kt} + 20$, where a score of 0 yields a speed of 20 km/h (the cap for dirt roads), and a score of 1 reaches the maximum speed cap s_{kt}. Maximum speeds are capped at 80 km/h for trunk roads and 60 km/h for district and main roads. If no road exists between two hexagons, a default speed of 10 km/h is applied. This method enables us to generate a detailed (historical) travel time matrix for Zambia in 2009, 2014, and 2019, which is not accessible through Google Maps.[20]

Panel (e) of Figure 1 shows the distribution of market access across Zambia during the study periods, revealing a substantial improvement over the decade from 2009 to 2019, with an average increase of approximately 0.8 log points. Figure 2 maps the spatial distribution of these changes. Panel (a) illustrates absolute changes in market access between 2009 and 2019, with gains concentrated around established economic hubs, particularly along the north-south corridor connecting Lusaka, the capital, to Ndola, the country's second-largest city. In contrast, Panel (b) displays relative growth in market access, which is more widely distributed beyond these core areas. Notably, the northwestern corridor linking Ndola to the Angolan border experiences the most significant relative improvement.

5.2 Main Specification

We examine how increased market access to other locations, driven by road upgrades and construction, influences a range of outcomes for each location. For outcomes observed through remote sensing at the hexagon level, we estimate the following change-to-change specification:

$$\Delta \ln Y_i = \alpha_0 + \beta_m \Delta \ln M_i + X_i \beta_x + \Delta W_i \beta_w + \epsilon_i \tag{3}$$

where i indexes hexagons. Y_i represents the outcome of interest, including AI-predicted built-up areas, yearly average nighttime light, annual AOD (air quality), and yearly average NDVI (green space). Δ indicates the long difference between 2009 and 2019.[21] M_i measures market access, as defined in Section 5.1. The coefficient β_m captures the elasticity of an outcome with respect to changes in market access. X_i includes control variables such as distances to rivers and lakes, proximity to the national border. It also includes the log initial outcome level in 2009 to control for historical trends or mean reversion as in Duranton and Turner (2012). ΔW_i includes changes in annual average precipitation and temperatures. α_0 is a constant term.[22]

For DHS panel outcomes based on Stable Spatial Panel Units (SPUs), we estimate a modified specification:

$$\Delta Y_i = \alpha_0 + \beta_m \Delta \ln M_i + X_i \beta_x + \Delta W_i \beta_w + \varepsilon_i \tag{4}$$

[20] Google Maps provides only real-time and recent traffic data, typically limited to recent years and without comprehensive archives for public access.

[21] Nighttime light data from 2012 is used to approximate outcomes around 2009.

[22] Although data around 2014 is available, we do not use five-year intervals in our analysis. Since road scores and market access are derived from images taken one year before and after each reporting year, a five-year interval would provide insufficient time for outcomes to adjust.

where i indexes each SPU. Since many DHS outcomes are discrete, we examine changes in outcomes, ΔY_i, which are first aggregated at the SPU-year level by calculating averages. The remaining components are consistent with our main specification in equation (3). In this linear-log model, a one percent increase in market access corresponds to an average change in outcomes of $\beta_m/100$.

Our market access measure is a function of built-up areas, which introduces the concern of reverse causality. Additionally, unobserved local factors, such as economic potential or regional policies that affect both market access and economic outcomes, may be omitted, leading to biased estimates. To address the endogeneity of market access, we employ an instrumental variable based on changes in remote roads following Jedwab and Storeygard (2021). Specifically, we construct the IV, M_{it}^{masked}, for hexagon i at time t as follows:

$$M_{it}^{\text{masked}} = \sum_{j \neq i} N_{j0} (1 + t_{ijt}^{\text{masked}})^{-\theta}, \tag{5}$$

where N_{j0} represents the initial 2009 built-up area of neighboring hexagon j relative to hexagon i, excluding the built-up area in hexagon i itself. This helps mitigating reverse causality. The term t_{ijt}^{masked} denotes travel time calculated from a modified road network, where road conditions within a 200 km radius of hexagon i are held constant at their 2009 levels, while conditions outside this radius are updated to their 2019 levels. Thus, any differences between the modified network and the initial 2009 network reflect changes solely in distant, non-local roads, effectively isolating the impact of these distant changes and filtering out local conditions that might influence the outcome through both road access and unobserved factors.

Figure 3 illustrates our implementation of the instrumental variable. Panels (a) and (b) display the road networks and road scores for a focal hexagon (indicated by the white dot) in 2009 and 2019, respectively, which we use to calculate market access for each year. The difference in market access between these years serves as our key independent variable. Comparing Panels (a) and (b), we observe road improvements indicated by increased road scores; for instance, the horizontal road below the focal hexagon changes from pink and purple to orange. These road upgrades contributes to higher market access for the focal hexagon. However, these upgrades may be driven by unobserved factors or trends in the outcome variable, potentially introducing endogeneity.

To address this issue, we create a synthetic 2019 road map, shown in Panel (c), to construct our IV. In this map, we draw a masked area with a 200 km boundary around the focal hexagon (represented by the green hexagon). Within this masked area, the road network is reverted to its initial 2009 conditions (the road network within the green hexagon in Panel (a)). Outside the masked region, the road network remains as it was in 2019 (the road network outside the green hexagon in Panel (b)). We then use this synthetic road map to calculate a synthetic market access measure for 2019 and, by differencing it with the 2009 market access, obtain our IV. In calculating this synthetic market access, we also hold neighboring hexagons' built-up levels constant at their 2009 values. This approach effectively removes the influence of local road changes (inside the masked hexagon), isolating market access variation driven by distant road upgrades, which are assumed to be exogenous to local conditions.

6 Results

Built-up Table 1 presents the estimates of the effect of changes in market access on built-up areas. The dependent variable is the change in log built-up between 2009 and 2019. Market access is calculated under the assumption that the decay parameter $\theta = 1$, with results for alternative decay factors discussed later in this section. We cluster standard errors at the district level.[23] Columns 1 to 3 show the OLS estimates, progressively adding control variables. Column 1 reveals a significant positive correlation between growth in built-up areas and increased market access. To account for historical trends or mean reversion, Column 2 includes the initial log of built-up, and the coefficient's magnitude slightly decreases. Column 3 further incorporates geographic controls, including distances to rivers and lakes to account for the influence of water transport and distance to the national border to address potential mechanical reductions in market access near borders, given that our data is limited to Zambia and excludes cross-border roads. Additionally, we control for changes in yearly average precipitation and temperature to account for the role of climate change in agricultural productivity and subsequent rural-to-urban migration (Henderson et al., 2017). The correlation remains significantly positive.

Columns 4 to 6 replicate the specifications from Columns 1 to 3, using instrumental variables (IV) for market access.[24] The IV estimates are slightly larger than the OLS estimates, suggesting a bias in road development toward less developed areas. This observation aligns with the literature indicating that roads were placed more often for political reasons than for economic ones (Burgess et al., 2015; McCartney, 2023). In our preferred specification in Column 6, a 1 percent increase in market access is associated with a 3.8 percent increase in built-up areas.

Ideally, we would estimate θ directly; however, this would require detailed data on trade flows between locations, which is scarce in developing countries. Instead, we select $\theta = 1$ based on the Bayesian Information Criterion (BIC) after testing the specification in Column 6 of Table 1 with various values of θ. A lower BIC value indicates a better fit according to Bayesian statistics. As shown in Appendix Figure A.10, the BIC remains stable until θ reaches a value of 3. We set $\theta = 1$ and provide key results across different values of θ as a robustness check.

Nighttime Light Table 2 presents the results for nighttime light. Column 1 reports OLS estimates, and Column 2 shows IV estimates. In both, we observe that increases in market access are associated with declines in nighttime light, though the effect is not statistically significant. If we interpret nighttime light as a proxy for income and built-up area as a proxy for population (as discussed in Section 4.2), these findings suggest that improved connectivity may encourage population growth without a corresponding increase in income or GDP growth. We also examine the ratio of nighttime light to built-up area as a proxy for average income, with results displayed in Columns 3 and 4. The IV estimates in Column 4 indicate a statistically significant decline in average income.[25] Although this result may seem counterin-

[23] Zambia is divided into 10 provinces and 116 districts.

[24] See Panel (b) of Appendix Table A.1 for the first stage results.

[25] Jedwab and Storeygard (2021) finds that market access increases nighttime light over a 30-year period, using the DMSP-OLS Nighttime Lights Time Series. While DMSP-OLS data cover a longer period (1992–2013), they have lower spatial resolution and less sensitive sensors than VIIRS nighttime lights, which we use here for higher resolution and more advanced sensor capabilities (Gibson et al., 2021). Additionally, DMSP-OLS data are top-coded in city centers where light intensity often exceeds the sensor's dynamic range.

tuitive, it aligns with the broader literature on "urbanization without growth" observed in many African countries (Fay and Opal, 2000). In such contexts, urbanization can strain the provision of basic infrastructure and services, especially when not accompanied by structural transformation (Henderson and Turner, 2020). The challenges and costs of urbanization without industrialization can outweigh the agglomeration benefits, as limited manufacturing activity fails to absorb the expanding urban labor force, thereby constraining both productivity and income growth.

Livings Conditions, Urban Infrastructure, and Amenity A key aspect of "urbanization without growth" is that as population move to densely populated areas, particularly cities, increased pressure on local public goods leads to deteriorating infrastructure, amenities, and overall living standards. We examine whether improved road connectivity—which drives expansion in built-up areas and potentially fuels population growth—also contributes to a decline in urban amenities and living conditions, using data from the Demographic and Health Surveys (DHS). Utilizing a panel from 69 stable spatial units, we analyze DHS outcomes observed in 2007 and 2018/19 to assess the impact of changes in market access by estimating equation (4). Panel A of Table 3 presents demographic outcomes. In Column 1, we find that the number of males per household increases with better market access, though the effect is muted for females, indicating increasingly crowded living conditions. This is further supported by Column 3, which shows that the number of bedrooms per household increases at a similar rate. Columns 6 and 7 reveal reductions in the size of agricultural land and the number of animals owned as market access improves. These findings are consistent with findings of increased growth in built-up areas.

In panel B of Table 3, we analyze key income and infrastructure indicators from the DHS, including access to bank accounts, electricity, radios, televisions, refrigerators, mobile phones, and mosquito nets. The results show that these indicators do not improve with increased market access; in fact, most coefficients are negative, albeit not statistically significant. Finally, Panel C of Table 3 shows that improved market access is associated with adverse health outcomes, including a reduction in birth weight, a higher probability of diarrhea, and a decreased likelihood of receiving vitamin A in the past six months.

These results suggest that road upgrades lead to increased crowding and declining household living standards, potentially accelerating the spread of slums given the high prevalence of informal housing in Zambia. Admittedly, slums can provide the rural poor with an initial foothold for accessing better opportunities (Marx et al., 2013; Glaeser and Henderson, 2017). However, the economic benefits of such transitions often take generations to fully materialize, which lies beyond the decade-long time frame of our study.

Air Pollution and Green Space We also examine the impact of road upgrades on environmental outcomes. Columns 1 and 4 of Table 4 show results for air pollution, measured by Aerosol Optical Depth (AOD), and green space, measured by Normalized Difference Vegetation Index (NDVI), respectively, using the preferred specification from Column 6 of Table 1. The findings suggest that a 1 percent increase in market access is associated with an approximately 1.1 percent increase in air pollution (Column 1) and a 0.77 percent reduction in green space (Column 4). These results are consistent with those in Table 1 and align with established literature in environmental science and economics, which shows that population growth typically lead to higher air pollution and deforestation (Southgate et al., 1991; Cropper

et al., 1999; Deininger and Minten, 2002; Foster and Rosenzweig, 2003; Brock and Taylor, 2005; Glaeser and Kahn, 2010; Kleinschroth et al., 2019).

We further examine the differential effects between urban and rural regions in the remaining columns of Table 4 with Columns 2 and 5 focusing on rural areas and Columns 3 and 6 on urban areas.[26] The results indicate that road upgrades lead to a greater increase in air pollution (comparing Columns 2 and 3) and more deforestation (comparing Columns 5 and 6) in rural areas than in urban areas. This likely reflects the lack of industrial activities in urban areas and environmentally harmful agricultural practices in rural areas.

Robustness Check Our results are robust across different values of the decay factor, θ, as shown in Appendix Table A.2. Each column presents the results corresponding to a specific value of θ. A smaller θ implies less discounting of access to distant markets. The direction of the estimates remains consistent, although the magnitude of the effects increases with smaller values of θ. [27] As highlighted by Jedwab and Storeygard (2021), the coefficient of market access is sensitive to the parameter θ, complicating the generalization of results despite θ being widely used in the literature. However, they demonstrate that estimating the effect of a one-standard-deviation increase in log market access yields robust results, unaffected by variations in θ. Following their approach, we present our estimation results in Appendix Figure A.11. Our findings confirm that the impact of a one-standard-deviation increase in log market access on key outcomes remains consistent across a range of θ values from 0.5 to 2.5, mirroring the stability observed by Jedwab and Storeygard (2021).

Currently, all of our results use changes in market access, constructed based on both changes in built-up areas and road upgrades. To what extent are these findings directly influenced by road improvements alone? To address this, we regress the outcome variables directly on the instrumental variable, following the main specification in Column 6 of Table 2. Panel A of Appendix Table A.1 presents these reduced-form results, where the variation in IV market access is driven exclusively by changes in travel time, as built-up areas are held constant at their initial levels. Thus, these reduced-form regressions effectively isolate the impact of road improvements on outcomes. We find that the results remain consistent across all four outcomes of interest, with magnitudes even larger than those in our baseline IV specification.

7 Mechanism: Urbanization without Growth

Our findings suggest that road construction in Zambia spurs population growth without corresponding income growth, while contributing to declining living standards and environmental quality. This pattern aligns with extensive evidence on the "urbanization without growth" phenomenon, where rapid urban population growth in many African countries has not been accompanied by structural transformation

[26] Urbanized areas are identified using the urban mask introduced in Section 4.1.

[27] Our estimates are higher than those reported in existing studies, likely due to the finer spatial granularity of our data. Instead of focusing solely on distant primary cities, we construct market access using adjacent, granular hexagons that encompass a range of settlements, from cities to villages. This granular approach appears to increase the elasticity of economic activity in response to market access. The economic intuition is that changes in market access have a greater impact on the relative growth of small places. For example, a new road passing through a village can create business opportunities that might double its size. However, such a dramatic transformation is unlikely to occur in a city, where the baseline size and existing infrastructure dampen the relative impact of additional market access.

(Fay and Opal, 2000). Drawing on the literature about the drivers of "urbanization without growth," we can better understand the mechanisms that may explain why road upgrades in Zambia have yielded limited positive impact over the past decade.

A key factor behind this phenomenon is the insufficient manufacturing activity to absorb the expanding urban labor force, which limits the potential for productivity and income growth. In many cases, rural-to-urban migration in Africa is not driven by better job opportunities but rather by a lack of infrastructure and services in rural areas, as seen in the significant presence of subsistence farmers in cities (Henderson and Kriticos, 2018). Another important factor is severe congestion in populous cities: as people move into urban areas, local governments face enormous challenges in providing adequate infrastructure and public services (Henderson and Turner, 2020). This strain on public resources often leads to a decline in average living standards.[28]

In this section, we investigate whether (1) a lack of an industrial base and (2) urban congestion help explain the limited success of road upgrade projects in Zambia.

7.1 Lack of Industrial Foundations

We investigate the limited industrial base as a mechanism by examining the differential effects in cities with a stronger manufacturing base. Using the 2011 Business Establishment Register, we aggregate sectoral revenue for agriculture, mining, manufacturing, and other sectors within each hexagon.[29] A hexagon is classified as manufacturing-intensive if it ranks in the top 20% for aggregate manufacturing revenue. Appendix Figure A.12 highlights these manufacturing-dominant hexagons. We then estimate the following regression:

$$\Delta lnY_i = \alpha_0 + \beta_m \Delta lnM_i + \beta_{mz} \Delta lnM_i \times Z_i + \beta_z Z_i + X_i \beta_x + \Delta W_i \beta_w + \varepsilon_i \tag{6}$$

The specification is similar to equation (3), but adds an interaction terms between ΔM_i and a dummy variable, Z_i, indicating whether a hexagon is manufacturing-intensive (ranks in the top 20% for aggregate manufacturing revenue). Additionally, we instrument $\Delta \ln M_i$ and $\Delta \ln M_i \times Z_i$ using $\Delta M_i^{\text{masked}}$ and $\Delta \ln M_i^{\text{masked}} \times Z_i$.

Table 5 reports results for five key outcomes: built-up areas, nighttime light, the nighttime light-to-built-up ratio, air pollution, and green space. In Column 1, the interaction term's coefficient (-0.456) is small relative to its standard errors, indicating no significant increase in built-up areas within manufacturing sectors. This suggests that manufacturing does not meaningfully interact with road infrastructure in influencing urban expansion. Columns 2 and 3 reveal positive interaction effects for nighttime light and the nighttime light-to-built-up ratio, suggesting that the impact of market access on urban growth and light intensity is more pronounced in manufacturing-intensive regions. These findings highlight the potential role of manufacturing capacity in amplifying the benefits of road infrastructure. However, the effects are not statistically significant, likely due to the absence of robust instruments for manufacturing activities, as evidenced by the low F-statistics in the first stage. Despite this limitation, the insignif-

[28] This is consistent with Au and Henderson (2006)'s theory of optimal city size, which suggests that improved market access, without accompanying structural change, can lead to population growth that reduces income growth rates.

[29] The manufacturing sector is classified according to the ISIC Rev 4 standard under code C. Agriculture is classified as ISIC code A, mining as ISIC code B, and all other activities fall under the remaining ISIC codes.

icance of both β_m and β_{mz} suggestively implies that improved market access does not substantially drive growth, even in regions with relatively developed manufacturing sectors. This aligns with the observation that Zambia's manufacturing sector predominantly consists of small-scale production and processing activities rather than large-scale industrial operations. Columns 4 and 5 provide additional insights, showing that improved market access is significantly associated with reduced air pollution and increased green space in areas with high manufacturing intensity. Typically, roads facilitating industrial activities would be expected to increase air pollution and decrease green space due to expanded industrial operations. The observed opposite effects suggest that the lack of extensive industrialization in these regions limits the ability of road infrastructure to stimulate significant economic development. This further underscores the constrained role of Zambia's manufacturing sector in leveraging road improvements for broader economic growth.

We also investigate how other aspects of structural transformation, including international trade and mining, influence our results. Proximity to ports can significantly influence local economic growth by enhancing trade and access to markets (Storeygard, 2016). In column 1 of Appendix Table A.3, we replace changes in market access with changes in access to two key land-based trading hubs. The first is Chipata, located in the east of the country, which connects Zambia to Mozambique's main seaport via Malawi. The second is Kazungula, situated in the southwest, which serves as a transit route to major seaports in South Africa and Namibia. Following Jedwab and Storeygard (2021), the growth of built-up areas in these border towns can be seen as a proxy for trade volume. We find no significant effects of improved market access to trading hubs on economic growth. This may be because imported tradables do not pass through these border towns but are instead distributed from Lusaka, especially goods transported via railway. Given this, and to explore how access to the national political center influences outcomes, we examine how changes in market access to Lusaka, the national capital, affect the results. Column 2 shows that better connectivity to Lusaka is associated with a decline in income, consistent with emerging literature on urban growth shadows from large economic centers (Chen et al., 2017; Cuberes et al., 2021). Additionally, enhanced connectivity to Lusaka is linked to increased air pollution, likely due to spillovers from large cities. In Column 3, we turn our focus to the mining sector, which accounts for a significant portion of Zambia's GDP, and assess how better connectivity to the mining areas in the Copperbelt province impacts outcomes. We find that while better access to mining areas does not drive economic growth, it does lead to increased pollution.

To sum up, our analysis shows that road upgrades in Zambia have not substantially driven economic growth, even in regions with a stronger manufacturing base. Additionally, we find that improved connectivity to the national capital and mining regions results in environmental degradation without boosting economic growth, reinforcing the notion that Zambia's current structural landscape limits the potential benefits of infrastructure investments.

7.2 Congestion

Another key driver of "urbanization without growth" in Zambia is severe congestion in rapidly expanding major cities. To explore this, we examine whether the effects of market access differ in more populous areas, focusing on provincial capitals. In Table 6, we extend our baseline regression by additionally in-

teracting changes in market access with a dummy for provincial capital cities.[30] Both market access and its interaction with the provincial capital indicator are instrumented using changes in IV market access and the interaction term. The results show that market access has a larger effect on built-up areas in provincial capitals compared to other regions, though the interaction term is noisy. Notably, we find a significant negative effect of market access on nighttime lights and the nighttime light-to-built-up ratio in these cities, indicating that strong congestion forces in larger cities may drive population growth with minimal income gains or even declines.[31] While high population density does not necessarily imply congestion, it is often an indicator, as supported by the literature. Given these nuances and our interest in understanding general equilibrium effects, we will further examine the role of congestion using a quantitative equilibrium model in Sections 8 and 9.

8 Quantitative Spatial Model

Locations in Zambia are interconnected through trade and migration, implying that changes in one area can affect others. Our reduced-form results compare changes in outcomes between locations with larger versus smaller increases in market access, which essentially eliminates any effect that is common to both groups. As a result, these reduced-form results alone cannot capture the aggregate and distributional effects of the road upgrades (see Redding and Rossi-Hansberg (2017) for discussion). To quantify the broader impacts of Zambia's road construction, analyze the economic forces behind its challenges, and explore potential policy improvements, we apply a quantitative spatial model mainly based on Allen and Arkolakis (2018) to describe the region's economic geography. This model enables us to evaluate both the aggregate and distributional welfare impacts of current road-building policies, while also allowing for policy simulations that explore potential optimizations. Additionally, we extend the model to incorporate multiple sectors, following recent advancements in economic geography research (Fan, 2019; Tombe and Zhu, 2019; Zárate, 2022; Rothenberg et al., 2024), to analyze implications for structural transformation. The model features multiple interconnected locations with costly trade and migration and multiple sectors, capturing endogenous agglomeration and congestion effects. In this section, we describe the model's structure in detail.

8.1 Setup

The model considers a country with N locations, with each location indexed by i for the origin and j for the destination. Each location has two sectors: $s \in U, R$, where U representing the non-agricultural sector and R denotes the agricultural sector. The sectors in each location is heterogeneous in terms of amenities and productivity. Locations are interconnected through costly trade and migration. In each location, representative firms produce distinct varieties of goods using labor, following the Armington assumption (Armington, 1969). In equilibrium, local wages and prices adjust to balance both the goods and labor markets.

[30] There are ten provincial capitals. We assign a value of 1 to hexagons within a 10 km buffer around each city's centroid and 0 otherwise.

[31] This finding aligns with literature on Africa's unique urbanization trajectory (Henderson and Kriticos, 2018). For comparison, Baum-Snow et al. (2020) find that in China, highway connections benefit regional primate cities, the largest and often disproportionately large cities in a region, while causing GDP declines and population losses in surrounding areas.

8.2 Migration Flows

The total population is fixed at $\bar{L}$.[32] Workers are initially located in their birthplace i (the origin) and choose a destination j and sector s to maximize utility. Each location has an initial, exogenous labor stock L_i^0. Workers value local consumption goods and amenities, and they select a sector and location based on the most attractive combination of wages, prices, and amenities, adjusted for the migration costs associated with relocating from their birthplace. The indirect utility of worker ω, born in district i and moving to district j, is given by:

$$V_{ijs}(\omega) = \frac{w_{js} u_{js} \epsilon_{ijs}(\omega)}{P_j D_{ij}} \tag{7}$$

where w_{js} denotes the nominal wage in sector s in location i and P_j represents the price index in location j. The term u_{js} indicates the local amenities in location j for sector s, which can be further broken down into:

$$u_{js} = \bar{u}_{js} L_j^{\beta} \tag{8}$$

where $\bar{u}_{js}$ represents the exogenous component of amenities (e.g., soil quality, proximity to the coast), while Lj^{β} reflects the endogenous component influenced by the local population L_j with the elasticity β, which captures congestion effects, such as increased traffic, overcrowding, and declining living standards caused by insufficient infrastructure and inadequate provision of public goods.

The term D_{ij} represents the migration cost associated with moving from location i to j. The variable $\epsilon_{ijs}(i)$ denotes the idiosyncratic preference of worker i, which follows a Frechet distribution given by $F(\overrightarrow{\epsilon}) = \exp\left(-\sum_{j\in N}\left(\sum_{s\in S}\epsilon_{ijs}^{-\nu}\right)^{\frac{\eta}{\nu}}\right)$. Here, η and ν are parameters that influence the dispersion of preferences across locations and sectors. As η decreases, workers' preferences become more dispersed spatially, indicating a greater variation in attachment to specific locations due to preference heterogeneity and a reduced sensitivity to changes in other welfare components such as wages, prices, and amenities.

Each worker selects the location and sector that maximize their indirect utility. According to the properties of the Frechet distribution, the probability that a worker born in location i will relocate to location j is given by:

$$\pi_{ijs} = \frac{\left(\frac{\Phi_j}{P_j D_{ij}}\right)^{\eta}}{\sum_{j'}\left(\frac{\Phi_{j'}}{P_{j'} D_{ij'}}\right)^{\eta}} \frac{(u_{js} w_{js})^{\nu}}{\sum_{s'}(u_{js'} w_{js'})^{\nu}} \tag{9}$$

where $\Phi_j^{\nu} = \sum_{s'}(u_{js'} w_{js'})^{\nu}$ captures the overall attractiveness of location j to residents. Specifically, we define the market access as follows,

$$W_i = \sum_{j'}\left(\frac{\Phi_{j'}}{P_{j'} D_{ij'}}\right)^{\eta} \tag{10}$$

which captures residents' access to well-paid jobs, amenities with lower cost of living from location i.

The *labor supply* in location j for sector s can be expressed as:

[32] We abstract from population growth.

$$\tilde{L}_{js} = \sum_{i \in N} \pi_{ijs} L_i^0 \tag{11}$$

By the property of Frechet distribution, the average welfare of workers living in location i is,

$$E\left[\max_{j,s} V_{ijz}(i)\right] = \Gamma\left(1 - \frac{1}{\eta}\right) \left\{ \sum_{j'} \left(\frac{W_{j'}}{P_{j'} D_{ij'}} \right)^{\eta} \right\}^{\frac{1}{\eta}} \tag{12}$$

where $\Gamma(\cdot)$ is the Gamma function, and $\Gamma(1 - \frac{1}{\eta})$ is a constant.

8.3 Demand for goods

Workers have uniform nested constant elasticity of substitution (CES) preferences over the varieties of consumption goods produced in each location and sector. At the upper level, they allocate their budget between agricultural and non-agricultural goods, while at the lower level, they choose among varieties within each sector. We use ξ to represent the elasticity of substitution between non-agricultural and agricultural goods, and let σ_s denote the elasticity of substitution among varieties within sector s. We assume that trade flows in the agricultural sector are more sensitive to trade costs, leading to $\sigma_U < \sigma_R$. The total consumer expenditure for an individual in location i, denoted by q_i, is given by:

$$q_i = \left(\sum_{s \in S} q_{is}^{\frac{\xi - 1}{\xi}} \right)^{\frac{\xi}{\xi - 1}} \qquad \text{where} \qquad q_{is} = \left(\sum_{j \in N} \int_{\omega} q_{jis}(\omega)^{\frac{\sigma_s - 1}{\sigma_s}} \, d\omega \right)^{\frac{\sigma_s}{\sigma_s - 1}}, \tag{13}$$

where $q_{jis}(\omega)$ represents the quantity of variety ω produced in location j and sector s that is consumed in location i. The price index for consumption goods in location i, denoted by P_i, is given by:

$$P_i = \left(\sum_{s \in S} P_{is}^{1 - \xi} \right)^{\frac{1}{1 - \xi}} \qquad \text{where} \qquad P_{is} = \left(\sum_{j \in N} \int_{\omega} p_{jis}(\omega)^{1 - \sigma_s} \, d\omega \right)^{\frac{1}{1 - \sigma_s}}, \tag{14}$$

where $p_{jis}(\omega)$ denotes the unit price of variety ω produced in location j and sector s but consumed in location i. This price is the factory-door price, $p_{js}(\omega)$, adjusted for the iceberg trade cost, expressed as $p_{jis}(\omega) = \tau_{ji} p_{js}(\omega)$.

8.4 Production

Each location i and sector s is assumed to produce a differentiated good y_{is} using labor L_{is}. Many firms produce homogeneous goods within a perfectly competitive market. The production function for a representative firm in location i and sector s is:

$$y_{is} = A_{is} L_{is} \tag{15}$$

where A_{is} is the overall productivity in location i and sector s,

$$A_{is} = \bar{A}_{is} L_i^{\alpha}, \tag{16}$$

which comprises an intrinsic component, $\bar{A}_{is}$, determined by exogenous factors such as soil quality and proximity to the coast, and an endogenous component influenced by the local population with elasticity α, capturing agglomeration externalities like better labor market matching, sharing infrastructure, and knowledge spillovers.

By perfect competition, the unit price of y_{is} is,

$$p_{is} = \frac{w_{is}}{A_{is}} \tag{17}$$

Given the product demand specified in equation (13), the share of expenditure in location j on goods produced in location i and sector s can be expressed as:

$$\lambda_{ijs} = \underbrace{\left(\frac{{P_{js}}^{1-\xi}}{\sum_{s' \in S} {P_{js'}}^{1-\xi}} \right)}_{\lambda_{js}} \underbrace{\left(\frac{p_{ijs}^{1-\sigma_s}}{\sum_{i' \in N} p_{i'js}^{1-\sigma_s}} \right)}_{\lambda_{ijs|s}}, \tag{18}$$

where $P_{js} = \left(\sum_{i' \in N} p_{i'js}^{1-\sigma_s} \right)^{\frac{1}{1-\sigma_s}}$.

8.5 Equilibrium

An equilibrium is defined over the endogenous variables $\{w_{is}, L_{is}, \omega_i, \phi_i, p_{is}, P_i\}$, subject to labor market clearing, goods market clearing, and the condition of a closed economy where total population is constant. The formal details are provided in the Model Appendix D.

8.6 Model Estimation

We use data and parameter estimates from the literature to calibrate the key parameters of our model. The population and migration data for this analysis is derived from a 10% sample of the 2000 and 2010 Zambian censuses provided by IPUMS. Wage data comes from the 2015 Zambia Living Conditions Monitoring Survey (LCMS), based on household income. To ensure representativeness, we calculated district-level averages and medians for wages, adjusting for survey weights. Wages were further disaggregated by sector (agricultural vs. non-agricultural) to offer more detailed insights at the district level. Employment data was used to estimate the non-agricultural employment share, which, when combined with population figures from the 2010 census, enabled us to estimate sectoral employment for each district in 2010.

First, we estimate Equation (9) to obtain the migration elasticity. Aggregating across both sectors for each origin-destination pair ($\pi_{ij} = \sum_s \pi_{ijs}$), and incorporating parameterization of the migration costs as follows,

$$D_{ij} = exp\left\{\beta_t travelTime_{ij}\right\} \tag{19}$$

Equation (9) can be expressed as follows,

$$\pi_{ij} = \exp\left\{ \underbrace{\ln(\eta\Phi_j/P_j)}_{\text{destination FE}} + \underbrace{\ln\sum_{j'}(\Phi_{j'}/P_{j'})^{\eta}}_{\text{origin FE}} - \eta\beta_t travelTime_{ij} + \epsilon_{ij} \right\} . \tag{20}$$

We estimate this equation using Poisson Pseudo Maximum Likelihood (PPML), controlling for both origin and destination fixed effects. The results, shown in Appendix Table A.4, report two-way clustered standard errors by origin and destination districts. Column 2 demonstrates a migration elasticity with respect to travel time of -0.95, which aligns with estimates from Indonesia, China, and India (Bryan and Morten, 2019; Tombe and Zhu, 2019; Ghose, 2021). In Column 1, we find similar effects using physical distance as a measure of travel cost. [33]

Following established literature, we calibrate additional parameters for our baseline model as outlined in Appendix Table A.5. Combining our estimated migration elasticity with other calibrated parameters and district-sector-level data on wages and population, we use the model framework (detailed in Model Appendix D) to recover local amenities, $\bar{U}_{is}$, and productivity, $\bar{A}_{is}$. The inverted fundamentals $\bar{U}_{is}$ and $\bar{A}_{is}$, shown in Figure A.14, provide insight into Zambia's economic geography. We observe higher productivity and amenity levels in Lusaka, the capital, and Copperbelt Province, known for copper mining, underscoring the validity of these inverted fundamentals.

9 Counterfactuals

We apply the quantitative spatial model developed in the previous section to assess the regional and aggregate effects of the current road upgrades in Zambia and conduct alternative policy experiments to explore potential mechanisms behind the limited impact of the existing policy and identify ways to enhance its effectiveness.

The Model-based Market Access we use the model structure to invert the market access, as defined in equation (10), to align with the market access defined in our reduced-form results. While both measures capture accessibility to "better" locations, their approaches differ: model-based market access directly quantifies access to desirable features such as higher wages, lower living costs, and greater amenities, whereas reduced-form market access reflects this accessibility indirectly through a revealed preference framework, based on the assumption that more desirable locations attract larger populations. To validate the quantitative spatial model, we compare the simulated changes in market access with those derived from the reduced-form approach. Panel (a) of Figure 4 demonstrates a correlation of 0.48 between the reduced-form and model-simulated changes in market access, supporting the use of the model for further welfare analysis and policy simulations. Panel (b) of Figure 4 shows a distribution of log changes in model-inverted market access that closely mirrors the patterns observed in the data, as seen in Panel (b) of Figure 2.

[33] The high correlation between pairwise physical distance and travel time is illustrated in Appendix Figure A.13.

The Overall and Distributional Impacts of the Current Road Construction We use the model to quantify the overall growth and welfare impacts of the Zambia's road-building policy by simulating the effects of reducing travel times between locations in Zambia from their 2009 levels to their 2019 level. Column 1 of Table 7 reports the resulting welfare and output changes due to Zambia's road upgrades. We find that this policy increases national GDP by only 0.065%, less than one-tenth of a percent, and leads to a modest welfare improvement of 0.2%. Notably, much of the GDP growth is driven by the agricultural sector. In fact, the policy negatively affects both the employment and GDP shares of the non-agricultural sector, suggesting it fails to promote and may even hinder structural transformation from agriculture to non-agriculture. This aligns with the reduced-form analysis, which finds no significant effects of road-building on economic growth as measured by nightlights.

Inefficient Targeting One of the key findings of our reduced-form analysis is that, while improved market access fosters population growth, it does not drive structural transformation in the affected locations. Our model suggests that one potential driver behind this is that new roads may not be directing population growth to areas with the highest economic potential. In Table 8, we analyze the correlation between changes in market access and initial location characteristics. The results indicate that while the current policy targets more populous locations (Column 1), there is no significant relationship between changes in market access and non-agricultural productivity (column 2) or amenities (column 3). This aligns with our earlier discussion in Section 3, indicating that road construction decisions are not primarily influenced by economic incentives. By attracting populations to locations without a solid industrial base or the potential for structural transformation, the program's positive economic impact is likely to be limited.

To investigate this further, we analyze the correlation between population growth in each district (measured by changes in the log of total population) and structural change (measured by changes in the log of GDP share in non-agricultural sectors), controlling for local changes in market access to isolate the policy's direct effects. Figure 5 illustrates a non-positive relationship between structural change and population growth, represented by the red line, and a negative relationship in larger cities (provincial capitals), denoted by the blue line. This indicates that districts with higher population growth do not experience corresponding increases in structural transformation from agriculture to non-agriculture under the current policy. This finding supports our earlier discussion in Section 7 and aligns with the literature on "urbanization without growth".

The Role of Congestion and Agglomeration As detailed in Section 7, a key factor limiting the success of the policy is the substantial congestion resulting from population growth in Zambia. Additionally, the lack of a robust industrial base may contribute to insufficient agglomeration, undermining the roads' contribution to economic growth.[34] We investigate the roles of congestion and agglomeration forces in evaluating the policy by assessing the sensitivity of our results to changes in key parameters. In Figure 6, each panel presents the outcomes of various simulations where the road-building policy (which reduces travel time) is rerun with one parameter altered at a time, while holding all others at their baseline values.

[34] It is well established that manufacturing provides higher agglomeration benefits than agricultureKrugman (1991); Fujita et al. (1999).

We then calculate the impact of the road construction on GDP and the non-agricultural GDP share. The y-axis shows the percentage changes in these outcomes, and the x-axis shows the value (magnitude) of the parameter that is changed. The dashed lines indicate the baseline parameter values and simulation results.

Panels A and C illustrate that when congestion forces are stronger (i.e., a larger magnitude of β), the road-building policy in Zambia has even weaker positive effects on GDP and more detrimental effects on structural change (measured by the non-agricultural GDP share). Conversely, Panels B and D demonstrate that with stronger agglomeration forces (i.e., a larger magnitude of α), the road-building policy is more effective at enhancing GDP growth and structural change. These results highlight that strong congestion and weak agglomeration forces are critical factors in the limited success of the current road-building policy.

Alternative Policies Where should policymakers target for optimal outcomes? To better understand the mechanisms behind the limited impact of the current road policy and to inform potential improvements, we use the model to conduct additional policy experiments. The average reduction in travel time under the current policy across all origin-destination pairs is around 7%. In the first experiment, we apply this 7% reduction in travel time to the five most productive locations, based on our inverted local productivity measure. Column 2 of Table 7 presents the aggregate results of this targeted approach. Compared to the current policy (Column 1), targeting more productive locations generates over three times the GDP growth and more than double the welfare gains. Additionally, this targeted approach shows a modest increase in both employment and GDP share in the non-agricultural sector, signaling structural change.

Next, we examine other local characteristics that policymakers could consider to achieve better outcomes. In a second set of experiments, we sequentially reduced travel time by 10% for each location (i.e., the travel time from each treated location to all other locations) and used the model to simulate the aggregate impacts. We then analyzed how these impacts correlated with the characteristics of the treated areas. The results, presented in Table 9, indicate that the policy delivers greater improvements in GDP, welfare, and industrialization when it targets more productive locations with higher initial employment shares in the non-agricultural sector and larger population sizes.

10 Conclusion

While transportation infrastructure is often seen as a catalyst for economic development, our results suggest that these benefits may not readily materialize in least-developed countries, where industrial activity lags and cities face significant overcrowding. These findings, which align with the mechanisms underlying the phenomenon of "urbanization without growth," had not been empirically tested in this context before due to the lack of granular data on roads and economic outcomes. By leveraging AI to analyze a vast trove of high-resolution satellite data, we assembled the necessary dataset to conduct this analysis. As anticipated, Zambia's road improvements have primarily driven population growth without corresponding increases in income. Additionally, we observe declining living standards and heightened environmental degradation. These results underscore that road paving alone may not be an

effective development strategy for countries experiencing "urbanization without growth."

Our study also suggests two likely overlooked aspects of the failure of road-building efforts: one economic and the other political. From an economic perspective, road construction alone may reinforce reliance on agriculture rather than fostering industrialization without complementary policies. Lower transportation costs often allow sectors to leverage their comparative advantages, which, in developing countries, tends to be agriculture rather than industrial activities. From a political perspective, road allocation is frequently influenced by political incentives, as noted by Acemoglu and Robinson (2012). This can lead to roads being constructed in locations that serve political interests rather than areas with the highest economic potential, thereby diminishing their effectiveness. Furthermore, the challenge of optimal road placement reflects broader issues of limited state capacity. While these aspects extend beyond the current scope of this study, they merit further research to deepen our understanding of the role of road improvements in economic development in the least developed contexts.

Declaration "During the preparation of this work the authors used ChatGPT in order to proofread, and applied deep learning models to predict built-up area and road surface conditions as described in the paper. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication."

References

ACEMOGLU, D. AND J. A. ROBINSON (2012): *Why Nations Fail: The Origins of Power, Prosperity, and Poverty*, New York: Crown Business.

ADUKIA, A., S. ASHER, AND P. NOVOSAD (2020): "Educational investment responses to economic opportunity: evidence from Indian road construction," *American Economic Journal: Applied Economics*, 12, 348–76.

AIDDATA (2024): "Project ID 31329: Rehabilitation of Zambia's Road Infrastructure," Accessed: 2024-10-09.

ALLEN, T. AND C. ARKOLAKIS (2014): "Trade and the Topography of the Spatial Economy," *Quarterly Journal of Economics*, 129, 1085–1140.

——— (2018): "13 Modern spatial economics: a primer," *World Trade Evolution*, 435.

ARMINGTON, P. S. (1969): "A theory of demand for products distinguished by place of production," *Staff Papers*, 16, 159–178.

ASHER, S. AND P. NOVOSAD (2020): "Rural roads and local economic development," *American economic review*, 110, 797–823.

AU, C. C. AND J. V. HENDERSON (2006): "Are Chinese cities too small?" *Review of Economic Studies*, 73, 549–576.

BALDWIN, K. (2013): "Why Vote with the Chief? Political Connections and Public Goods Provision in Zambia," *American Journal of Political Science*, 57, 794–809.

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

BARAGWANATH, K., R. GOLDBLATT, G. HANSON, AND A. K. KHANDELWAL (2019): "Detecting urban markets with satellite imagery: An application to India," *Journal of Urban Economics*, 103173.

BAUM-SNOW, N. (2007): "Did Highways Cause Suburbanization?" *The Quarterly Journal of Economics*, 122, 775–805.

BAUM-SNOW, N., L. BRANDT, J. V. HENDERSON, M. A. TURNER, AND Q. ZHANG (2017): "Roads, Railroads, and Decentralization of Chinese Cities," *The Review of Economics and Statistics*, 99, 435–448.

BAUM-SNOW, N., J. V. HENDERSON, M. A. TURNER, Q. ZHANG, AND L. BRANDT (2020): "Does investment in national highways help or hurt hinterland city growth?" *Journal of Urban Economics*, 115, 103124, cities in China.

BONAFILIA, D., D. YANG, J. GILL, AND S. BASU (2019): "Building High Resolution Maps for Humanitarian Aid and Development with Weakly-and Semi-Supervised Learning," *The IEEE Conference on Computer Vision and Pattern Recognition - Workshops*.

BROCK, W. A. AND M. S. TAYLOR (2005): "Economic Growth and the Environment: A Review of Theory and Empirics," in *Handbook of Economic Growth*, vol. 1B, 1749–1821.

BRYAN, G. AND M. MORTEN (2019): "The aggregate productivity effects of internal migration: Evidence from Indonesia," *Journal of Political Economy*, 127, 2229–2268.

BURGESS, R., M. HANSEN, B. A. OLKEN, P. POTAPOV, AND S. SIEBER (2012): "The Political Economy of Deforestation in the Tropics," *The Quarterly Journal of Economics*, 127, 1707–1754.

BURGESS, R., R. JEDWAB, E. MIGUEL, A. MORJARIA, AND G. PADRÓ I MIQUEL (2015): "The value of democracy: evidence from road building in Kenya," *American Economic Review*, 105, 1817–1851.

BURKE, M., A. DRISCOLL, D. B. LOBELL, AND S. ERMON (2020): "Using satellite imagery to understand and promote sustainable development," .

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

CHEN, X. AND W. D. NORDHAUS (2011): "Using luminosity data as a proxy for economic statistics," *Proceedings of the National Academy of Sciences*, 108, 8589–8594.

CHEN, Y., J. V. HENDERSON, AND W. CAI (2017): "Political favoritism in China's capital markets and its effect on city sizes," *Journal of Urban Economics*, 98, 69–87, urbanization in Developing Countries: Past and Present.

CHIFUNDA, E. (2024): "National Three(3) Million Tones Copper Production Strategy By 2031," Tech. rep., Zambia Ministry of Mines.

COSTINOT, A., D. DONALDSON, AND C. SMITH (2016): "Evolving Comparative Advantage and the Impact of Climate Change in Agricultural Markets: Evidence from 1.7 Million Fields around the World," *Journal of Political Economy*, 124, 205–248.

CROPPER, M., C. GRIFFITHS, AND M. MANI (1999): "Roads, Population Pressures, and Deforestation in Thailand, 1976-1989," *Land Economics*, 75, 58–73.

CUBERES, D., K. DESMET, AND J. RAPPAPORT (2021): "Urban growth shadows," *Journal of Urban Economics*, 123, 103334.

DEININGER, K. AND B. MINTEN (2002): "Determinants of Deforestation and the Economics of Protection: An Application to Mexico," *American Journal of Agricultural Economics*, 84, 943–960.

DEVELOPMENT MINERALS (2024): "Zambia: Paving the Road to Success," Accessed: 2024-10-09.

DONALDSON, D. (2018): "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure," *American Economic Review*, 108, 899–934.

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

DONALDSON, D. AND A. STOREYGARD (2016): "The View from Above: Applications of Satellite Data in Economics," *Journal of Economic Perspectives*, 30, 171–198.

DURANTON, G. AND M. A. TURNER (2012): "Urban Growth and Transportation," *The Review of Economic Studies*, 79, 1407–1440.

EDMOND, C., V. MIDRIGAN, AND D. Y. XU (2015): "Competition, markups, and the gains from international trade," *American Economic Review*, 105, 3183–3221.

EUROPEAN COMMISSION (2020): "A highway to success for Emmanuel and Peter in Zambia," Accessed: 2024-09-10.

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

FAJGELBAUM, P. D. AND E. SCHAAL (2020): "Optimal Transport Networks in Spatial Equilibrium," *Econometrica*, 88, 1411–1452.

FAN, J. (2019): "Internal geography, labor mobility, and the distributional impacts of trade," *American Economic Journal: Macroeconomics*, 11, 252–88.

FAY, M. AND C. OPAL (2000): "Urbanization Without Growth : A Not-So-Uncommon Phenomenon," Tech. rep.

FEENSTRA, R. C., P. LUCK, M. OBSTFELD, AND K. N. RUSS (2018): "In search of the Armington elasticity," *Review of Economics and Statistics*, 100, 135–150.

FOSTER, A. D. AND M. R. ROSENZWEIG (2003): "Economic Growth and the Rise of Forests," *The Quarterly Journal of Economics*, 118, 601–637.

FRETZ, S., R. PARCHET, AND F. ROBERT-NICOUD (2021): "Highways, Market Access, and Spatial Sorting*," *The Economic Journal*, ueab070.

FUJITA, M., P. KRUGMAN, AND A. J. VENABLES (1999): *The Spatial Economy: Cities, Regions, and International Trade*, Cambridge, MA: MIT Press.

GENDRON-CARRIER, N., M. GONZALEZ-NAVARRO, S. POLLONI, AND M. A. TURNER (2018): "Subways and Urban Air Pollution," Working Paper 24183, National Bureau of Economic Research.

GERTLER, P. J., M. GONZALEZ-NAVARRO, T. GRAČNER, AND A. D. ROTHENBERG (2024): "Road maintenance and local economic development: Evidence from Indonesia's highways," *Journal of Urban Economics*, 143, 103687.

GHANI, E., A. G. GOSWAMI, AND W. R. KERR (2015): "Highway to Success: The Impact of the Golden Quadrilateral Project for the Location and Performance of Indian Manufacturing," *The Economic Journal*, 126, 317–357.

GHOSE, D. (2021): *Trade, Internal Migration, and Human Capital: Who Gains from India's IT Boom?*, World Bank.

GIBBONS, S., T. LYYTIÄINEN, H. G. OVERMAN, AND R. SANCHIS-GUARNER (2019): "New road infrastructure: the effect on firms," *Journal of Urban Economics*, 110, 35–50.

GIBSON, J., S. OLIVIA, G. BOE-GIBSON, AND C. LI (2021): "Which night lights data should we use in economics, and where?" *Journal of Development Economics*, 149, 102602.

GLAESER, E. L. (2014): "A world of cities: The causes and consequences of urbanization in poorer countries," *Journal of the European Economic Association*, 12, 1154–1199.

GLAESER, E. L. AND J. V. HENDERSON (2017): "Urban Economics for the Developing World: An Introduction," *Journal of Urban Economics*, 98, 1–5.

GLAESER, E. L. AND M. E. KAHN (2010): "The Greenness of Cities: Carbon Dioxide Emissions and Urban Development," *Journal of Urban Economics*, 67, 404–418.

GOLLIN, D., R. JEDWAB, AND D. VOLLRATH (2016): "Urbanization with and without industrialization," *Journal of Economic Growth*, 21, 35–70.

HARDING, R. AND D. STASAVAGE (2014): "What Democracy Does (and Doesn't Do) for Basic Services: School Fees, School Inputs, and African Elections," *The Journal of Politics*, 76, 229–245.

HENDERSON, J. AND S. KRITICOS (2018): "The Development of the African System of Cities," *Annual Review of Economics*, 10, 287–314.

HENDERSON, J. V., A. STOREYGARD, AND U. DEICHMANN (2017): "Has climate change driven urbanization in Africa?" *Journal of Development Economics*, 124, 60–82.

HENDERSON, J. V., A. STOREYGARD, AND D. N. WEIL (2012): "Measuring economic growth from outer space," *American Economic Review*, 102, 994–1028.

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

IMF (2022): "Zambia: Request for an Arrangement Under the Extended Credit Facility-Press Release; Staff Report; and Statement by the Executive Director for Zambia," Accessed: 2024-09-10.

IMF (2024): "Government Gross Debt as a Percentage of GDP - Sub-Saharan Africa," Accessed: 2024-10-09.

JAWORSKI, T. AND C. T. KITCHENS (2019): "National Policy for Regional Development: Historical Evidence from Appalachian Highways," *The Review of Economics and Statistics*, 101, 777–790.

JAYACHANDRAN, S. (2009): "Air Quality and Early-Life Mortality: Evidence from Indonesia's Wildfires," *Journal of Human Resources*, 44, 916–954.

JEDWAB, R. AND A. STOREYGARD (2021): "The Average and Heterogeneous Effects of Transportation Investments: Evidence from Sub-Saharan Africa 1960-2010," *Journal of the European Economic Association*, jvab027.

KHACHIYAN, A., A. THOMAS, H. ZHOU, AND G. HANSON (2021): "Using Neural Networks to Predict Micro-Spatial Economic Growth," *American Economic Review: Insights*.

KHANNA, G., W. LIANG, A. M. MOBARAK, AND R. SONG (2021): "The productivity consequences of pollution-induced migration in China," *NBER Working Paper*.

KLEINSCHROTH, F., N. LAPORTE, W. F. LAURANCE, S. J. GOETZ, AND J. GHAZOUL (2019): "Road expansion and persistence in forests of the Congo Basin," *Nature Sustainability*, 2, 628–634.

KRUGMAN, P. (1991): *Geography and Trade*, Cambridge, MA: MIT Press.

MARX, B. (2018): "Elections as Incentives: Project Completion and Visibility in African Politics," Tech. rep., SciencesPo.

MARX, B., T. STOKER, AND T. SURI (2013): "The Economics of Slums in the Developing World," *Journal of Economic Perspectives*, 27, 187–210.

MARX, B., T. N. STOKER, AND T. SURI (Forthcoming): "There is no free house: Ethnic patronage in a Kenyan slum," *American Economic Journal*.

MBEWE, M., I. MASILOKWA, M. K. THÉODORE HUMANN, M. KESSLER, AND S. MWAMBA1 (2024): "The road to Zambia's 2020 sovereign debt default," *Zambian Institute for Policy Analysis and Research (ZIPAR)*.

MCCARTNEY, M. (2023): "A Zambian Road to Nowhere? The History of Big (Road) Infrastructure and the 2011-2022 Zambian Road-Building Boom," Unpublished report.

MHID (2020): "Link Zambia 8000," Tech. rep., Zambian Ministry of Housing, Infrastructure.

MICHAELS, G. (2008): "The Effect of Trade on the Demand for Skill: Evidence from the Interstate Highway System," *The Review of Economics and Statistics*, 90, 683–701.

OEHMCKE, S., C. THRYSOE, A. BORGSTAD, M. A. V. SALLES, M. BRANDT, AND F. GIESEKE (2019): "Detecting Hardly Visible Roads in Low-Resolution Satellite Time Series Data," *Proceedings - 2019 IEEE International Conference on Big Data, Big Data 2019*, 2403–2412.

PESARESI, M., L. MAFFENINI, S. FREIRE, P. POLITIS, AND M. SCHIAVINA (2023): "GHS-SDATA R2023A - GHS supporting data," Dataset.

PINKOVSKIY, M. AND X. SALA-I MARTIN (2016): "Lights, Camera … Income! Estimating Poverty Using National Accounts, Survey Means, and Lights," *Quarterly Journal of Economics*, 131, 579–631.

PLAN, N. D. (2017): "Seventh National Development Plan 2017-2021," Zambia Ministry of Planning.

RABALLAND, G. AND A. WHITWORTH (2012): "Should the Zambian Government Invest in Railways?" Working paper no. 3, Zambia Institute for Policy Analysis and Research ZIPAR.

REDDING, S. J. AND E. ROSSI-HANSBERG (2017): "Quantitative spatial economics," *Annual Review of Economics*, 9, 21–58.

REDDING, S. J. AND M. A. TURNER (2015): "Transportation Costs and the Spatial Organization of Economic Activity," in *Handbook of Regional and Urban Economics*, Elsevier, vol. 5, 1339–1398.

ROAD DEVELOPMENT AGENCY (2017): "Annual Report 2017," Accessed: 2024-10-09.

——— (2019): "Annual Report 2019," Accessed: 2024-10-09.

——— (2023): "Annual Report 2023," .

ROMÁN, M. O., Z. WANG, Q. SUN, V. KALB, S. D. MILLER, A. MOLTHAN, L. SCHULTZ, J. BELL, E. C. STOKES, B. PANDEY, K. C. SETO, D. HALL, T. ODA, R. E. WOLFE, G. LIN, N. GOLPAYEGANI, S. DEVADIGA, C. DAVIDSON, S. SARKAR, C. PRADERAS, J. SCHMALTZ, R. BOLLER, J. STEVENS, O. M. RAMOS GONZÁLEZ, E. PADILLA, J. ALONSO, Y. DETRÉS, R. ARMSTRONG, I. MIRANDA, Y. CONTE, N. MARRERO, K. MACMANUS, T. ESCH, AND E. J. MASUOKA (2018): "NASA's Black Marble nighttime lights product suite," *Remote Sensing of Environment*, 210, 113–143.

ROTHENBERG, A., Y. WANG, AND A. CHARI (2024): "When Regional Policies Fail: An Evaluation of Indonesia's Integrated Economic Development Zones," *Available at SSRN 4717284*.

ROTHENBERG, A. D. (2013): "Transport infrastructure and firm location choice in equilibrium: evidence from Indonesia's highways," *Unpublished Manuscript, Department of Economics, University of California, Berkeley*.

SIMONOVSKA, I. AND M. E. WAUGH (2014): "The elasticity of trade: Estimates and evidence," *Journal of international Economics*, 92, 34–50.

SINGH, A., W. R. AVIS, AND F. D. POPE (2020): "Visibility as a proxy for air quality in East Africa," *Environmental Research Letters*, 15.

SOUTHGATE, D., R. SIERRA, AND L. BROWN (1991): "The Causes of Tropical Deforestation in Ecuador: A Statistical Analysis," *World Development*, 19, 1145–1151.

STATISTA (2024): "National Debt of Zambia in Relation to Gross Domestic Product (GDP)," Accessed: 2024-10-09.

STOREYGARD, A. (2016): "Farther on down the Road: Transport Costs, Trade and Urban Growth in Sub-Saharan Africa," *The Review of Economic Studies*, 83, 1263–1295.

TASSOT, C., L. PELLERANO, AND J. LA (2019): "Informality and Poverty in Zambia: Findings from the 2015 Living Conditions and Monitoring Survey," Tech. rep., International Labour Organisation.

TIECKE, T. G., X. LIU, A. ZHANG, A. GROS, N. LI, G. YETMAN, T. KILIC, S. MURRAY, B. BLANKESPOOR, E. B. PRYDZ, AND H.-A. H. DANG (2017): "Mapping the world population one building at a time," .

TOMBE, T. AND X. ZHU (2019): "Trade, migration, and productivity: A quantitative analysis of china," *American Economic Review*, 109, 1843–72.

UNCTAD (2011): "An investment guide to Zambia: Opportunities and conditions," Tech. rep., The United Nations.

UNITED NATIONS ECONOMIC COMMISSION FOR AFRICA (2016): "Africa's Construction and Energy Sectors through Modern Industrial Policy," Policy Brief 16/102, United Nations Economic Commission for Africa (UNECA), accessed: 2024-11-16.

VANDOME, C. (2023): "Zambia's developing international relations," *Chatham House Research Papers*.

VENABLES, A. J. (2017): "Breaking into tradables: Urban form and urban function in a developing city," *Journal of Urban Economics*, 98, 88–97.

YOU, Y., J. LI, S. REDDI, J. HSEU, S. KUMAR, S. BHOJANAPALLI, X. SONG, J. DEMMEL, K. KEUTZER, AND C.-J. HSIEH (2019): "Large Batch Optimization for Deep Learning: Training BERT in 76 minutes," .

ZÁRATE, R. D. (2022): "Spatial misallocation, informality, and transit improvements," *Development Research*.

ZHOU, L., C. ZHANG, AND M. WU (2018): "D-linknet: Linknet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2018-June, 192–196.

Table 1: THE EFFECT OF MARKET ACCESS: BUILT-UP

	OLS			IV		
	(1)	(2)	(3)	(4)	(5)	(6)
Δ Log MA	3.510***	3.103***	3.163***	4.596**	3.484*	3.798*
	(1.100)	(1.092)	(1.151)	(2.099)	(2.059)	(2.134)
Initial Log outcome		-0.045***	-0.048***		-0.045***	-0.047***
		(0.017)	(0.016)		(0.017)	(0.016)
Log Distances to River			-0.041			-0.039
			(0.041)			(0.042)
Log Distances to Lake			-0.024			-0.033
			(0.030)			(0.037)
Log Distances to National Border			-0.038			-0.038
			(0.037)			(0.037)
Changes in Precipitation			0.210			0.189
			(0.256)			(0.259)
Changes in Temperature			1.395			1.466
			(1.511)			(1.548)
KPW F-Stats				46.68	44.97	47.05
R^2	0.01502	0.02732	0.03084	0.01358	0.02714	0.03043
N	2488	2488	2488	2488	2488	2488

Notes: This table presents our estimates of the impact of market access and the dependent variable is the change in log built-up. We assume $\theta = 1$, with a mask area of approximately 200 km. Columns 1 to 3 report the OLS estimates, while columns 4 to 6 display the IV estimates. Standard errors are clustered at the district level. */**/*** denotes significant at the 10% / 5% / 1% levels.

Table 2: THE EFFECT OF MARKET ACCESS: NIGHTTIME LIGHT

	NTL Black Marble		NTL/Built-up	
	(1) OLS	(2) IV	(3) OLS	(4) IV
Δ Log MA	-0.877	-4.123	-2.727	-8.385*
	(2.074)	(3.778)	(2.170)	(4.384)
Initial Log outcome	-0.120**	-0.135***	-0.281***	-0.315***
	(0.047)	(0.048)	(0.053)	(0.065)
Log Distances to River	-0.225***	-0.228***	-0.249***	-0.260***
	(0.083)	(0.081)	(0.071)	(0.072)
Log Distances to Lake	0.044	0.046	0.087**	0.091**
	(0.038)	(0.038)	(0.041)	(0.043)
Log Distances to National Border	0.112**	0.125**	0.071	0.092
	(0.051)	(0.052)	(0.057)	(0.063)
Changes in Precipitation	0.940**	1.061**	0.840*	1.056**
	(0.460)	(0.480)	(0.458)	(0.510)
Changes in Temperature	-1.188	-1.754	0.675	-0.088
	(2.547)	(2.722)	(1.932)	(2.120)
KPW F-Stats		34.70		32.20
R^2	0.11703	0.10267	0.22214	0.17771
N	322	322	322	322

Notes: This table presents our estimates of the impact of market access, with the dependent variable being the change in log nighttime light. We assume $\theta = 1$ and a mask area of approximately 200 km. Columns 1 and 3 report the OLS estimates, while columns 2 and 4 show the IV estimates. Standard errors are clustered at the district level. */**/*** denotes significant at the 10% / 5% / 1% levels.

Table 3: THE EFFECT OF MARKET ACCESS: LIVING STANDARDS (DHS)

Panel A:	(1) Men	(2) Women	(3) Bedrooms	(4) Education	(5) Age	(6) Arg Land	(7) Own animals
Δ Log MA	2.033*	-0.024	2.030*	4.350	-3.502	-356.570**	-1.148*
	(1.109)	(0.851)	(1.134)	(5.635)	(9.637)	(160.834)	(0.642)
KPW F-Stats	18.58	18.58	18.58	18.58	18.58	18.58	18.58
R^2	0.14	-0.00	0.19	-0.06	0.04	-0.01	-0.26
N	69	69	69	69	69	69	69
Panel B:	(1) Bank	(2) Electricity	(3) Radio	(4) Television	(5) Fridge	(6) Mobile	(7) Mos. Net
Δ Log MA	-0.665	-0.542	0.131	-0.085	-0.894	0.547	-0.209
	(0.715)	(0.793)	(0.457)	(0.812)	(0.623)	(0.788)	(1.854)
KPW F-Stats	18.58	18.58	18.58	18.58	18.58	18.58	18.58
R^2	0.03	0.01	-0.02	-0.00	0.07	0.04	-0.00
N	69	69	69	69	69	69	69
Panel C:	(1) Weight	(2) BMI	(3) BCG	(4) DPT	(5) Vaccine	(6) Diarrhea	(7) Vitamin
Δ Log MA	-1874.264*	85.431	0.631	-0.657	1.180	1.041**	-1.553*
	(1115.293)	(210.364)	(0.564)	(0.528)	(1.287)	(0.418)	(0.913)
R^2	-0.13042	0.02888	0.06542	-0.08155	0.05978	0.09814	-0.12761
N	66	69	69	69	58	69	68

Notes: This table presents IV estimates for three groups of DHS outcomes. The dependent variables are the changes in the mean of the outcomes listed in the column titles of each panel. Regressions are weighted by the total number of households in each stable panel unit (SPU). We assume $\theta = 1$ and a mask area of approximately 200 km, with changes in market access instrumented. Robust standard errors are applied and clustered at the district level. BMI (Body Mass Index) is calculated by dividing a person's weight in kilograms by the square of their height in meters. BCG (Bacillus Calmette-Guérin) is a vaccine against tuberculosis and related mycobacterial infections. The DPT vaccine provides protection against diphtheria, tetanus, and pertussis (whooping cough). Standard errors are clustered at the district level. */**/*** denotes significant at the 10% / 5% / 1% levels.

Table 4: THE EFFECT OF MARKET ACCESS: ENVIRONMENTAL OUTCOMES

	AOD			NDVI		
	(1) All	(2) Rural	(3) Urban	(4) All	(5) Rural	(6) Urban
Δ Log MA	1.070***	1.080***	0.843***	-0.767**	-0.774**	-0.330
	(0.294)	(0.297)	(0.194)	(0.341)	(0.349)	(0.362)
Initial Log outcome	-0.381***	-0.384***	-0.444***	-0.350***	-0.339***	-0.233*
	(0.078)	(0.079)	(0.061)	(0.063)	(0.065)	(0.126)
Log Distances to River	0.003	0.003	0.004	0.002	0.002	0.001
	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.006)
Log Distances to Lake	0.004	0.004	0.004	0.041***	0.042***	0.017***
	(0.007)	(0.007)	(0.006)	(0.012)	(0.013)	(0.006)
Log Distances to National Border	-0.011**	-0.011**	-0.012***	0.009	0.009	0.009
	(0.005)	(0.005)	(0.005)	(0.006)	(0.007)	(0.008)
Changes in Precipitation	0.006	0.005	-0.000	0.097**	0.094**	0.057
	(0.030)	(0.030)	(0.027)	(0.040)	(0.039)	(0.052)
Changes in Temperature	-0.065	-0.070	0.034	-1.074***	-1.090***	-1.061***
	(0.180)	(0.182)	(0.169)	(0.230)	(0.230)	(0.254)
KPW F-Stats	41.84	41.79	31.39	43.83	43.12	37.82
R^2	0.15767	0.15324	0.30213	0.26568	0.25327	0.12377
N	2488	2488	1347	2461	2460	1273

Notes: This table presents our estimates of the impact of market access, with the dependent variable being the change in log AOD in columns 1–3 and log NDVI in columns 4–6. We assume $\theta = 1$ and a mask area of approximately 200 km. Column 1 and 4 show results from all pixels within a hexagon, columns 2 and 5 display results from pixels in rural (non-urbanized) areas only, while columns 3 and 6 present results for urban areas. Standard errors are clustered at the district level. */**/*** denotes significant at the 10% / 5% / 1% levels.

Table 5: MECHANISM: MANUFACTURING-INTENSIVE LOCATIONS

	(1) Built-up	(2) NTL	(3) NTL/Built	(4) AOD	(5) NDVI
Changes in Log Market Access	3.985*	-9.569	-15.879*	1.144***	-0.885**
	(2.364)	(8.818)	(8.654)	(0.320)	(0.357)
Δ Log Market Access × Manufacture	-0.456	8.194	12.328	-0.534**	0.716***
	(3.503)	(9.671)	(8.707)	(0.233)	(0.253)
Manufacture Dummy	0.722	-5.768	-9.189	0.445**	-0.619***
	(2.782)	(7.603)	(6.887)	(0.190)	(0.207)
Initial Log outcome	-0.054***	-0.231***	-0.377***	-0.383***	-0.352***
	(0.017)	(0.060)	(0.077)	(0.080)	(0.063)
KPW F-Stats	22.98	8.09	7.54	20.08	21.27
R^2	0.03301	0.12309	0.15997	0.15029	0.26114
N	2488	322	322	2488	2461

Notes: This table presents the results on the impact of market access between manufacturing-intensive locations and other areas. The dependent variables are changes in log built-up, log nighttime light, log (nighttime light / built-up), log AOD, and log NDVI, respectively. We assume $\theta = 1$ and a mask area of approximately 200 km. Both changes in market access and its interaction with the manufacturing dummy are instrumented. Standard errors are clustered at the district level. */**/*** denotes significant at the 10% / 5% / 1% levels.

Table 6: MECHANISM: PROVINCIAL CAPITALS

	(1) Built-up	(2) NTL	(3) NTL/Built	(4) AOD	(5) NDVI
Changes in Log Market Access	3.790*	0.195	-3.373	1.063***	-0.788**
	(2.176)	(2.868)	(3.454)	(0.300)	(0.348)
Δ Log Market Access $\times$ Regional Capital	4.840	-24.747***	-27.444***	0.444	1.152*
	(3.142)	(8.493)	(10.490)	(0.668)	(0.693)
Regional Capital Dummy	-3.550	20.080***	22.270***	-0.356	-0.949*
	(2.563)	(6.724)	(8.294)	(0.532)	(0.562)
Initial Log outcome	-0.048***	-0.142***	-0.303***	-0.381***	-0.350***
	(0.017)	(0.050)	(0.058)	(0.078)	(0.063)
KPW F-Stats	23.24	14.30	13.36	20.55	21.48
R^2	0.03083	0.11563	0.20877	0.15597	0.26312
N	2488	322	322	2488	2461

Notes: This table presents the results on the impact of market access between provincial capital cities and other locations. The dependent variables are changes in log built-up, log nighttime light, log (nighttime light / built-up), log AOD, and log NDVI, respectively. We assume $\theta = 1$ and a mask area of approximately 200 km. Both changes in market access and its interaction term with the regional capital dummy are instrumented. Standard errors are clustered at the district level. */**/*** denotes significant at the 10% / 5% / 1% levels.

Table 7: POLICY SIMULATIONS

	Baseline **(1)**	**+Reducing Travel Time for Productive Cities** **(2)**
Change in GDP (%)	0.065	0.205
Change in Non-agricultural GDP (%)	0.051	0.191
Change in Agricultural GDP (%)	0.397	0.521
Change in Welfare (%)	0.225	0.554
Change in Non-Agricultural GDP Share (%)	-0.007	0.001
Change in Non-Agricultural Emp. Share (%)	-0.017	0.002

Notes: Authors' calculations. This table shows percentage changes in outcomes by comparing actual variable values with their counterfactual counterparts, under a scenario where travel time between locations is reduced by a specific road policy. Column 1 presents results reflecting changes based on the current (actual) travel times. Column 2 illustrates a scenario in which travel times to all other locations are reduced by 7% for the 10 most productive cities, as identified by the local fundamentals estimated from our model.

Table 8: CORELATION BETWEEN MARKET ACCESS AND LOCAL CHARACTERISTICS

	Initial Population **(1)**	**Non-ag Productivity** **(2)**	**Non-ag Amenities** **(3)**
Log Changes in Market Access	0.180**	0.004	34.268
	(0.080)	(0.008)	(20.502)

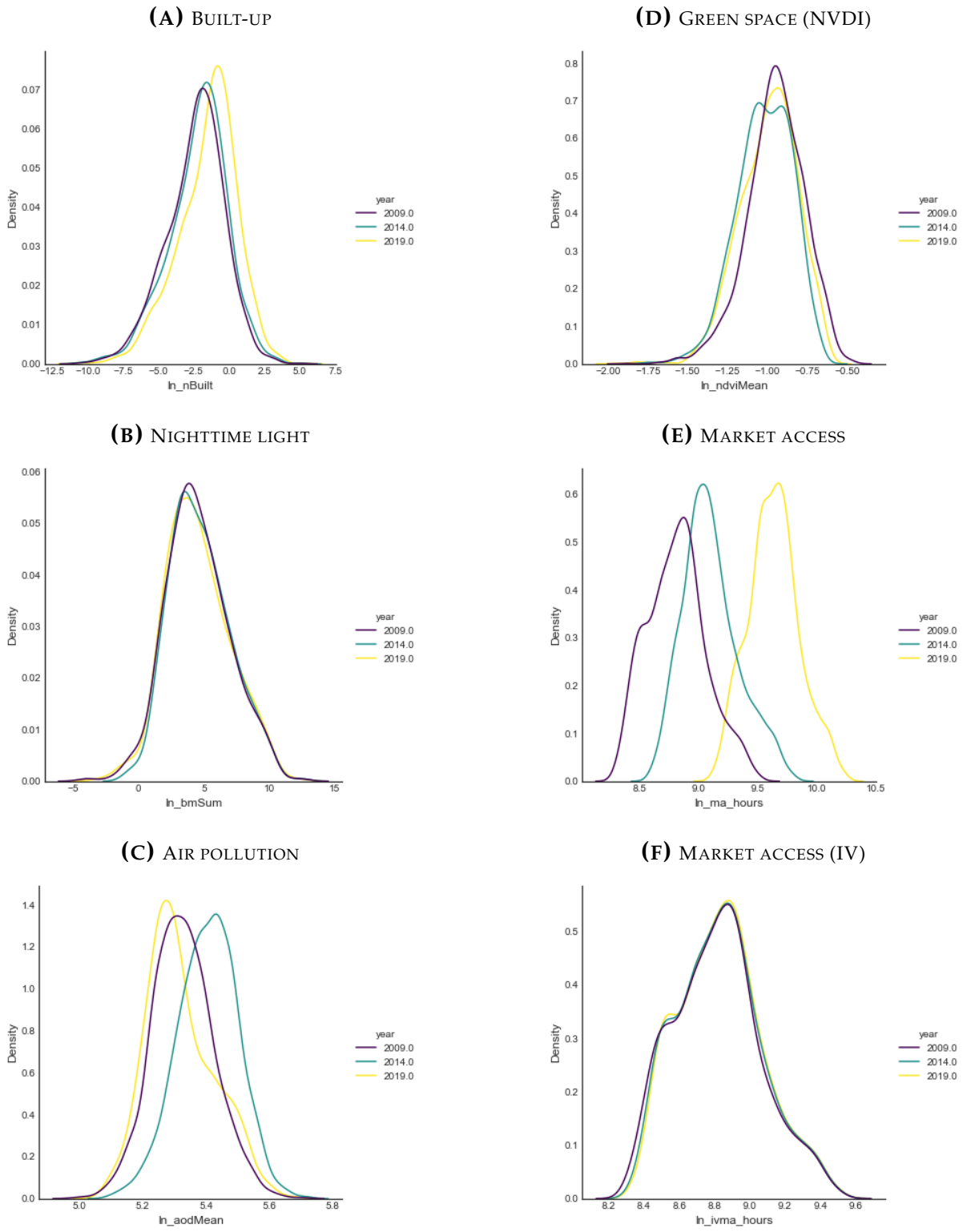
Notes: This table presents the correlation between the log changes in model-inverted market access, driven by travel time reductions under the current road policy, and the observed or inverted characteristics of each location, as specified in the column headings. Robust standard errors are shown in parentheses. */**/*** denotes significant at the 10% / 5% / 1% levels.

Table 9: THE DETERMINANTS OF IMPACTS OF INDIVIDUALLY REDUCING TRAVEL TIME

	Non-ag Productivity (1)	Ag Productivity (2)	Non-ag Amenities (3)	Ag Amenities (4)	Initial Non-ag Population Share (5)	Initial Population (6)
Change in Real GDP (%)	0.087***	0.095***	200.056***	-9.833*	4.904***	15.864***
	(0.010)	(0.019)	(52.441)	(5.713)	(0.691)	(2.090)
Change in Real Urban GDP (%)	0.088***	0.091***	200.511***	-9.230	4.890***	16.399***
	(0.010)	(0.020)	(53.879)	(5.865)	(0.722)	(2.109)
Change in Real Rural GDP (%)	0.050**	0.187***	189.094**	-23.893***	5.225***	3.348
	(0.021)	(0.020)	(80.021)	(7.670)	(1.168)	(4.266)
Change in Welfare (%)	0.179***	0.245***	569.154***	-19.455	11.751***	43.876***
	(0.027)	(0.040)	(112.376)	(13.309)	(1.539)	(3.523)
Change in Industrialization (Pop) (%)	0.016***	0.018***	43.906***	-1.884	0.866***	4.449***
	(0.003)	(0.004)	(11.109)	(1.226)	(0.168)	(0.205)
Change in Industrialization (GDP) (%)	0.006***	0.003	13.874***	0.063	0.243***	1.700***
	(0.001)	(0.002)	(4.980)	(0.529)	(0.080)	(0.139)

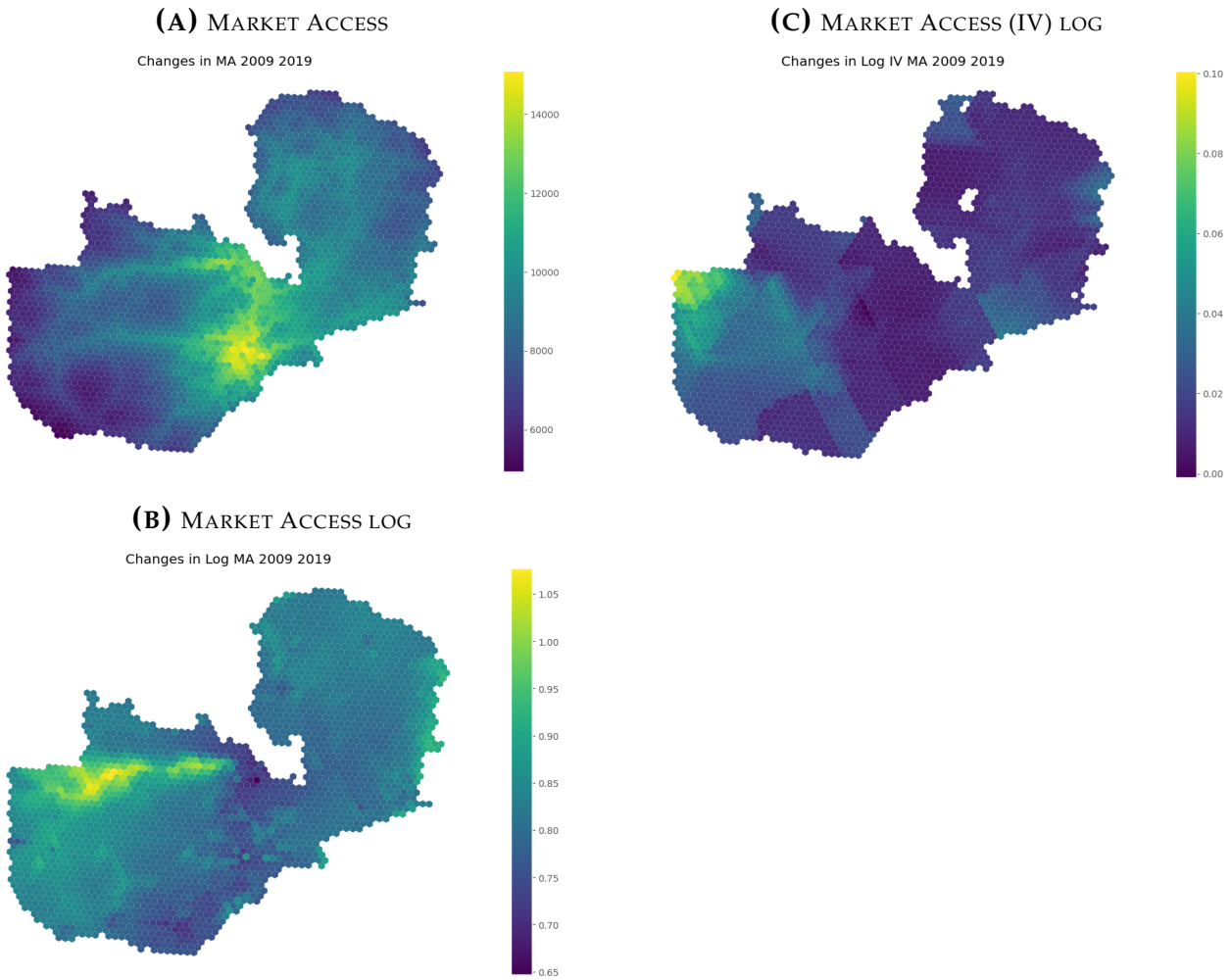
Notes: This table presents the correlation between aggregate welfare changes (in percentage) and the observed or inverted characteristics of each treated location when travel time is reduced by 10% for each location individually. For ease of visualization, the dependent variables are rescaled by multiplying by 100. Robust standard errors are shown in parentheses. */**/*** denotes significant at the 10% / 5% / 1% levels.

Figure 1: Distribution of Key Variables



Notes: This figure displays AI-predicted road quality across the core road network. In the top two panels, roads shown in brighter colors have a higher predicted probability of having a paved surface. In the bottom panel, brighter colors indicate more significant changes in road surface conditions.

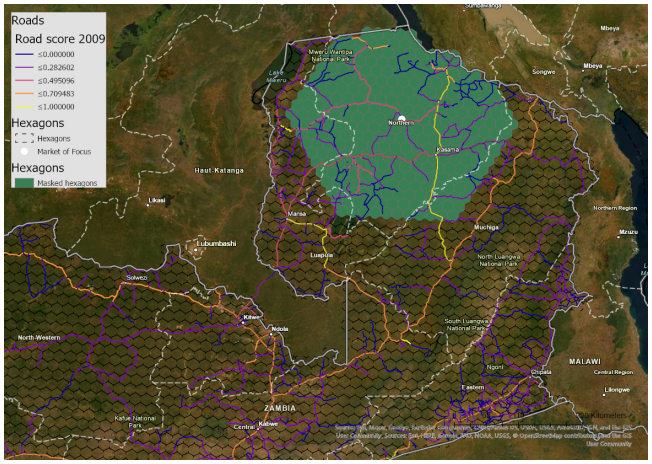
Figure 2: Spatial Distribution of the Changes in Market Access



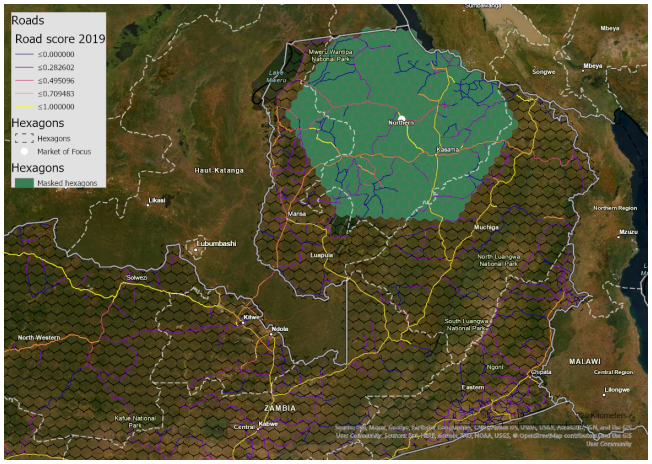
Notes: Panel (a) shows the spatial distribution of changes in market access between 2009 and 2019. Panel (b) displays the spatial distribution of changes in log market access over the same period. Panel (c) presents the spatial distribution of the instrumental variable (IV), capturing improvements in roads more than 200km away, weighted by economic mass in 2009, and excluding the growth in built-up areas.

Figure 3: IV Constructed by Masking out local roads

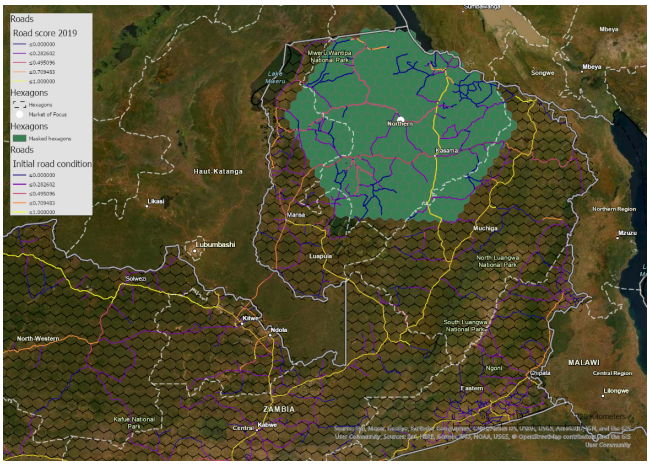
(A) Road Maps Used for Constructing MA, 2009



(B) Road Maps Used for Constructing MA, 2019

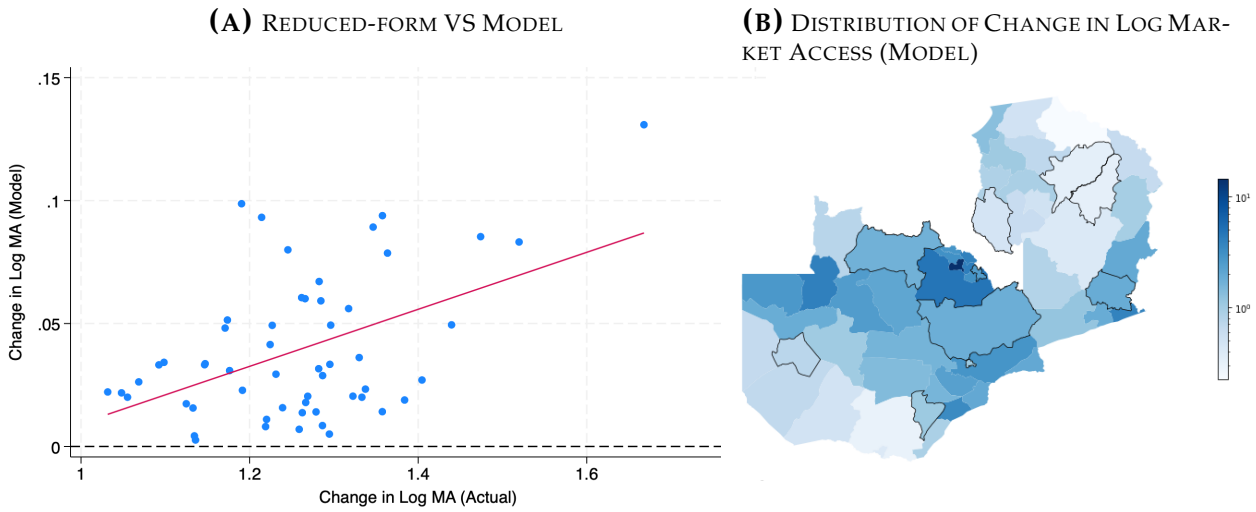


(C) Synthetic Road Maps Used for Constructing IV



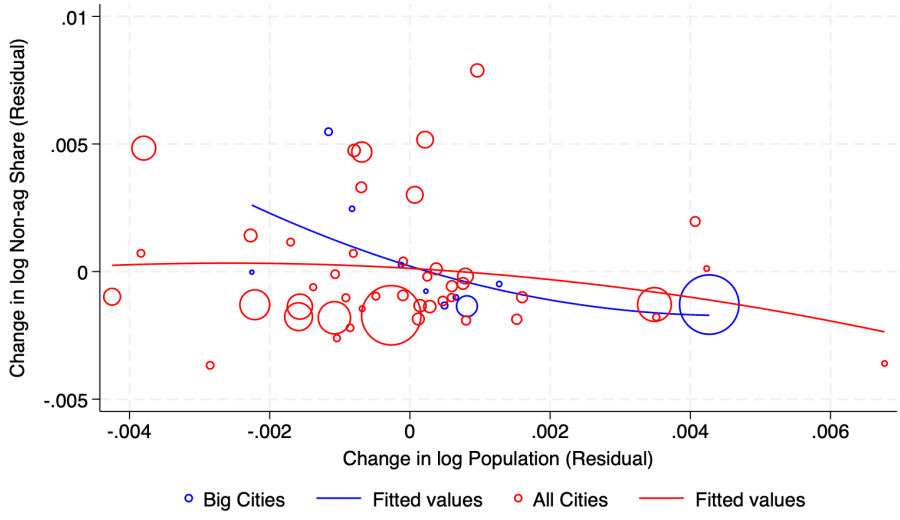
Notes: This figure illustrates the instrumental variable strategy following Jedwab and Storeygard (2021). Panel (a) presents the market access for the focal market, calculated using road scores from 2009, while Panel (b) shows the market access based on 2019 road scores. In Panel (c), we instrument the changes in market access for the focal market (white-filled circle) using changes in market access driven by distant road improvements (areas outside the mask, indicated by large green hexagons), with neighboring markets' built-up levels held constant at their initial values.

Figure 4: Changes in Market Access from Road Building, Model Simulation



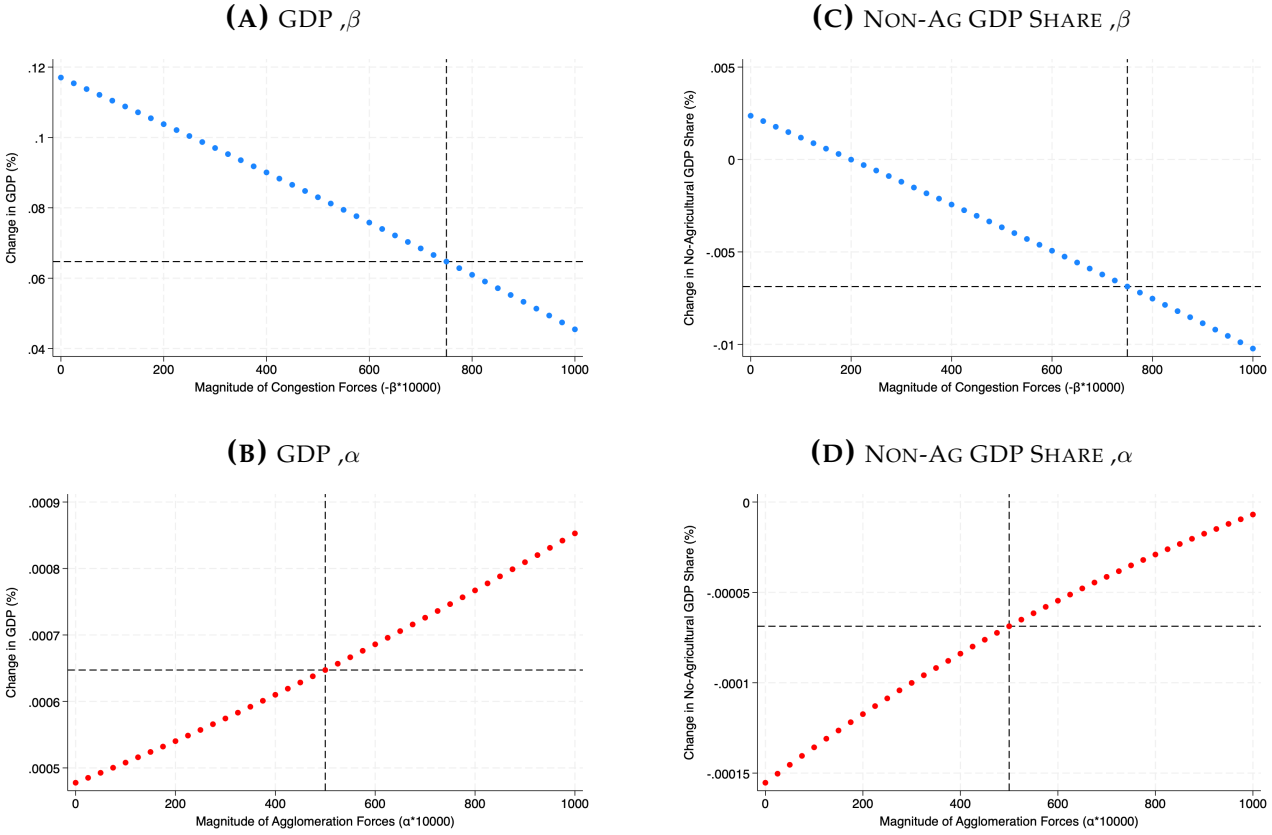
Notes: This figure shows the features of the changes in market access from the road building policy. Specifically, we calculate the percentage change in the endogenous variables, including the model-based market access, ($\hat{X} = (X_t - X_{t-1})/X_{t-1}$) and plot them against the percentage change in reduced-form market access ($\hat{MA} = (MA_t - MA_{t-1})/MA_{t-1}$). Panel A displays the correlation between the changes in log model-inverted market access and the changes in log reduced-form market access, with each blue dot representing a specific district. Panel B presents the spatial distribution of changes in log market access, where darker shades indicate areas with greater increases.

Figure 5: Local Structural Changes and Population Growth from Road Building, Model Simulation



Notes: This figure illustrates the correlation between changes in log population and changes in the log share of non-agricultural GDP, after purging out changes in market access. Each blue (or red) dot represents a district outside of a provincial city, with the size of each dot proportional to the district's population size. The blue (or red) line shows a quadratic fit of this relationship. Similar negative relationship is found using linear fit.

Figure 6: GDP and Non-ag GDP Shares Changes from Road Building, Changes in β and α



Notes: This figure presents the sensitivity of our primary conclusions on the impact of Zambia's road-building policy across different parameter values. In each panel, each dot represents the outcome of a separate simulation in which we rerun the model, varying one parameter while keeping all others fixed at their baseline values. We calculate the policy's effects on national GDP (Panels A and B) and on the share of non-agricultural GDP (Panels C and D), both shown on the y-axis. The dotted lines on the y-axis represent the baseline results, while the dotted line on the x-axis marks the baseline parameter value.

A Appendix Tables and Figures

Table A.1: FIRST STAGE AND REDUCED-FORM RESULTS

	(1) Built-up	(2) NTL	(3) NTL/Built	(4) AOD	(5) NDVI
Panel A: *Reduced-form*					
Changes in Log Market Access (IV)	6.982*	-10.688	-21.683*	1.862***	-1.396**
	(4.003)	(10.051)	(11.538)	(0.526)	(0.611)
Panel B: *First stage*					
Changes in Log Market Access (IV)	1.838***	2.592***	2.586***	1.740***	1.820***
	(0.268)	(0.440)	(0.456)	(0.269)	(0.275)
N	2488	322	322	2488	2461

Notes: The top panel of the table presents the first-stage results, while the bottom panel displays the reduced-form results. We assume $\theta = 1$ and a mask area of approximately 200 km. Standard errors are clustered at the district level. */**/*** denotes significant at the 10% / 5% / 1% levels.

Table A.2: ROBUSTNESS: THE EFFECT OF MARKET ACCESS UNDER DIFFERENT θ

	(1) 0.5	(2) 1	(3) 1.5	(4) 2	(5) 2.5
Built-up					
Changes in Log Market Access	7.986*	3.798*	2.356*	1.577	0.964
	(4.607)	(2.134)	(1.363)	(1.060)	(1.010)
KPW F-Stats	50.55	47.05	38.08	26.67	16.48
R^2	0.03	0.03	0.03	0.04	0.04
N	2488	2488	2488	2488	2488
NTL					
Changes in Log Market Access	-11.588	-4.123	-1.878	-0.827	-0.190
	(9.522)	(3.778)	(2.121)	(1.431)	(1.118)
KPW F-Stats	31.45	34.70	34.22	29.89	23.54
R^2	0.09	0.10	0.11	0.12	0.12
N	322	322	322	322	322
NTL/Built					
Changes in Log Market Access	-22.263*	-8.385*	-4.243*	-2.371	-1.325
	(11.390)	(4.384)	(2.391)	(1.575)	(1.217)
KPW F-Stats	28.61	32.20	32.40	28.99	23.61
R^2	0.14	0.18	0.20	0.21	0.22
N	322	322	322	322	322
AOD					
Changes in Log Market Access	2.296***	1.070***	0.689***	0.531***	0.478***
	(0.661)	(0.294)	(0.181)	(0.136)	(0.130)
KPW F-Stats	46.00	41.84	33.32	23.07	14.12
R^2	0.17	0.16	0.13	0.05	-0.13
N	2488	2488	2488	2488	2488
NDVI					
Changes in Log Market Access	-1.611**	-0.767**	-0.504**	-0.396**	-0.364**
	(0.737)	(0.341)	(0.217)	(0.166)	(0.152)
KPW F-Stats	49.58	43.83	33.55	22.13	12.68
R^2	0.27	0.27	0.26	0.24	0.19
N	2461	2461	2461	2461	2461

Notes: This table presents IV estimates under varying assumptions of θ, ranging from 0.5 to 2.5. The dependent variables are changes in log built-up, log nighttime light, log (nighttime light / built-up), log AOD, and log NDVI, as indicated in the column titles across the four panels. Standard errors are clustered at the district level. */**/*** denotes significant at the 10% / 5% / 1% levels.

Table A.3: INTERNATIONAL TRADE, NATIONAL CAPITAL, AND MINING CENTERS

	(1) Ports	(2) Capital	(3) Mining
Built-up			
Δ Log MA	3.487	-6.852	2.026
	(2.586)	(4.672)	(1.239)
KPW F-Stats	12.55	32.83	33.85
R^2	0.01	-0.00	0.02
N	2488	2488	2488
NTL			
Δ Log MA	2.304	-6.897	-5.022
	(6.070)	(4.228)	(24.417)
KPW F-Stats	7.24	20.05	0.31
R^2	0.11	0.06	0.06
N	322	322	322
NTL/Built			
Δ Log MA	0.181	-9.786**	-32.680
	(5.611)	(4.934)	(68.348)
KPW F-Stats	6.91	19.70	0.27
R^2	0.21	0.09	-1.65
N	322	322	322
AOD			
Δ Log MA	0.025	0.516**	0.393**
	(0.292)	(0.243)	(0.183)
KPW F-Stats	7.46	37.97	31.51
R^2	0.22	0.19	0.22
N	2488	2488	2488
NDVI			
Δ Log MA	-1.630***	-0.374	-0.290
	(0.498)	(0.385)	(0.187)
KPW F-Stats	12.47	32.10	38.12
R^2	0.14	0.30	0.29
N	2461	2461	2461

Notes:This table examines the effects of changes in access to trade ports, the national capital, and mining centers—driven by road construction—on key outcomes. The dependent variables include changes in log built-up, log nighttime light, log (nighttime light / built-up), log AOD, and log NDVI, as shown across the four panels. The primary regressors of interest are changes in market access to these critical locations, as specified in the column titles. These regressors are instrumented using hypothetical market access measures to these locations, constructed similarly to the main IV market access measure. The regression specifications follow Column 6 in Table 1. Standard errors are clustered at the district level. */**/*** denotes significant at the 10% / 5% / 1% levels.

Table A.4: MIGRATION GRAVITY ESTIMATION

	(1)	(2)
Ln(Physical Distance)	-1.040***	
	(0.082)	
Log (Travel Time), hr		-0.946***
		(0.071)
N of region pairs	9,075	9,075
N Clusters	55	55
Pseudo R-squared	0.561	0.562

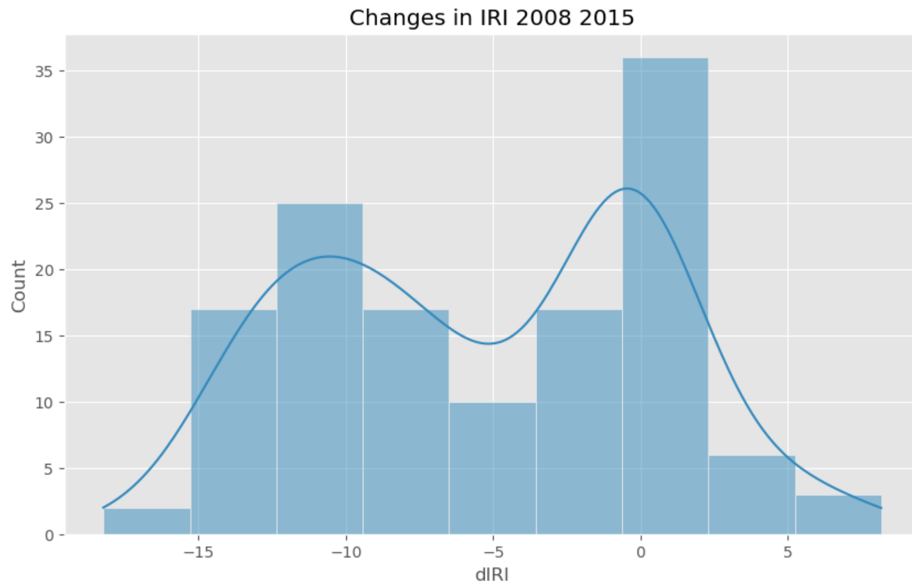
Notes: This table shows the results of estimating the migration gravity equation, equation (20). We estimate this equation using Poisson Pseudo Maximum Likelihood (PPML) with destination-, origin-, and year-fixed effects. Robust standard errors, clustered two-way by origin and destination district, are reported in parentheses. */**/*** denotes significant at the 10% / 5% / 1% levels.

Table A.5: Parameterization of the Model

Parameter	Value	Literature
Agglomeration Parameters α	0.05	Bryan and Morten 2019
Congestion Parameters ν	-0.075	Bryan and Morten 2019
Armington Elasticity of Substitution σ	5	Simonovska and Waugh 2014; Feenstra et al. 2018
Sector Elasticity of Substitution between Goods ξ	2	Edmond et al. 2015
Sector Dispersion Parameter ν	1.95	Zárate 2022
Sector Dispersion Parameter Dispersion Parameter η	1.6	Tombe and Zhu 2019; Khanna et al. 2021 Fan 2019
Trade cost elasticity τ_{ij}	-0.63	Donaldson and Hornbeck 2016

Notes: This table presents the baseline parameter values used to calibrate the model, along with their corresponding sources.

Figure A.1: The distribution of the changes in IRI from 2008 to 2015

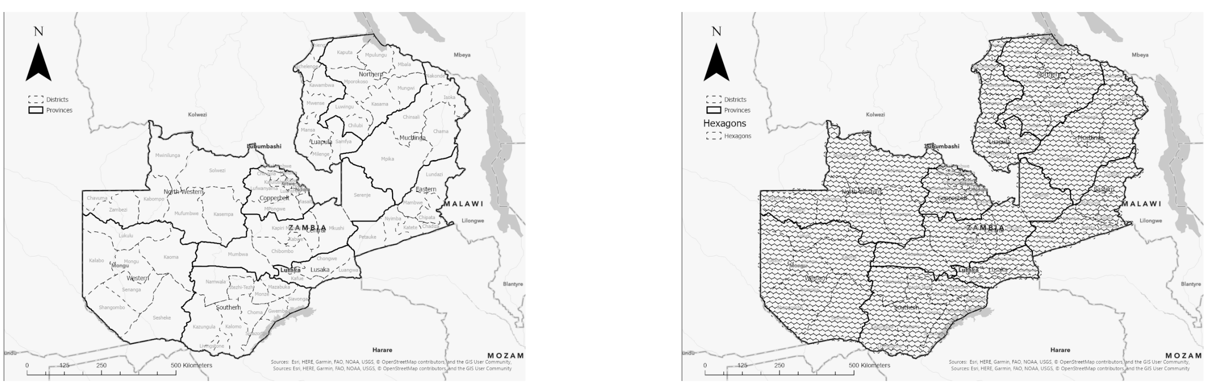


Notes: This graph presents a histogram of changes in the International Roughness Index (IRI) for road segments reported in the Road Development Agency's report, comparing data from 2008 to 2015.

Figure A.2: Geography of Zambia and Hexagons

(A) Provinces and Districts

(B) Hexagons

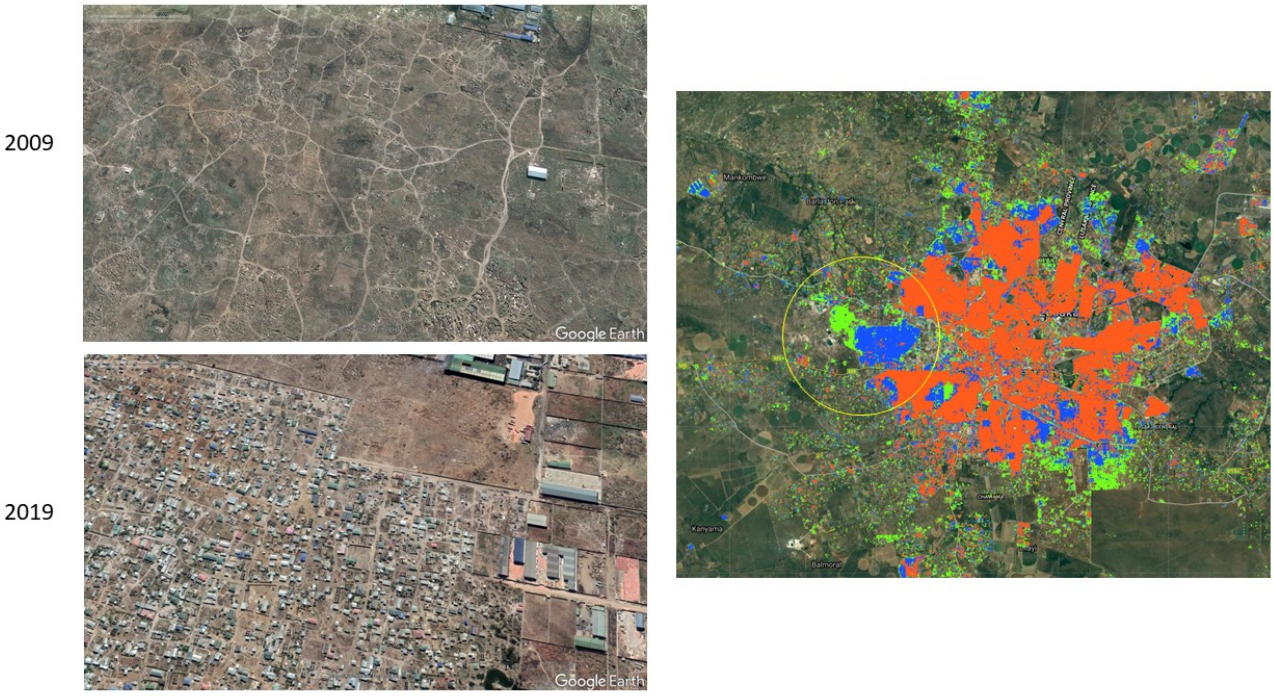


Notes: Panel (A) displays the boundaries of provinces and districts in Zambia, while Panel (B) illustrates the hexagonal grid cells used as the basic units for spatial analysis.

Figure A.3: An Example of the Planet RapidEye Product

Notes: This is a typical satellite image, covering an area of 25 km × 25 km with a resolution of 5000 × 5000 pixels across 4 bands. Radiometric and sensor corrections have been applied, and the imagery is orthorectified using RPCs and an elevation model.

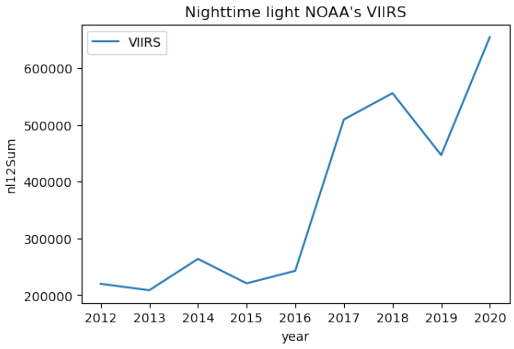
Figure A.4: Built-up results: Lusaka



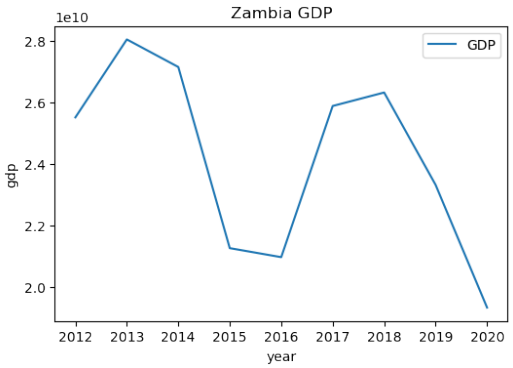
Notes: This figure illustrates how our AT-predicted built-up closely align with actual built-up growth. The image on the right shows the AI-predicted built-up for Lusaka, the largest city in Zambia. Areas overlaid with red indicates the state of built-up in 2009, while blue indicates new built-up developed between 2009 and 2014, and green indicates additional built-up developed between 2014 and 2019. On the left, we zoom in on the area circled in yellow for the years 2009 (top) and 2019 (bottom), taken from Google Earth. Visually, we see that the model prediction aligns well with actual built-up growth over the years.

Figure A.5: Comparing GDP with Different Sources of Nighttime light data

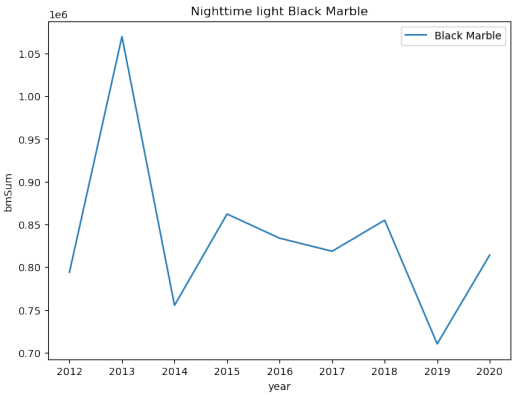
(A) NOAA's VIIRS



(C) GDP

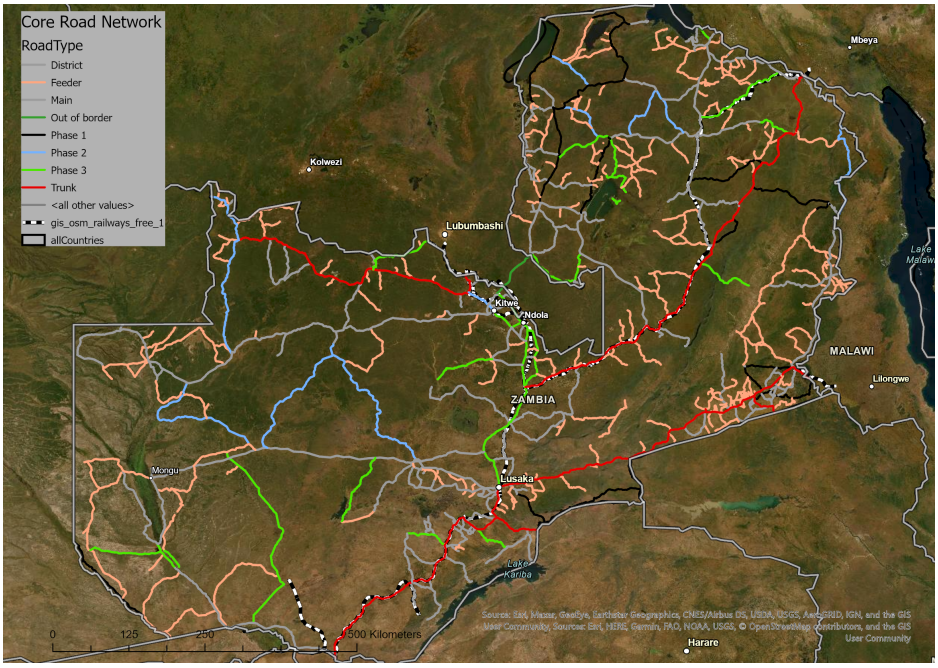


(B) Black Marble



Notes: The figure displays a comparison of time series data for nighttime lights from NOAA's VIIRS and Black Marble, plotted against GDP data from 2012 to 2019.

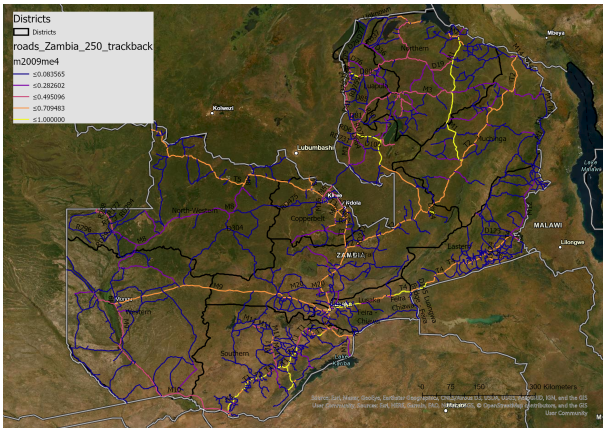
Figure A.6: Core Road Network in Zambia



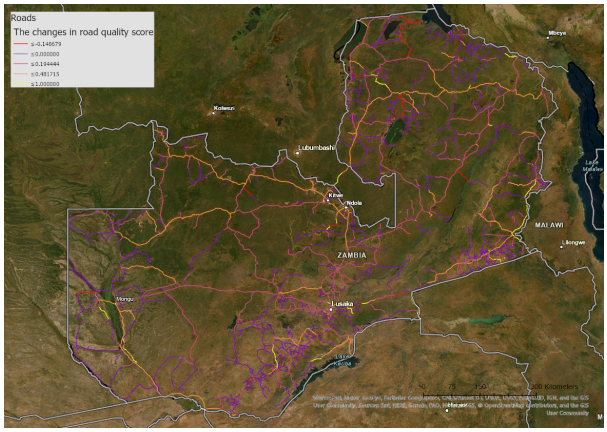
Notes: This figure displays the core road network digitized from the Link Zambia 8000 construction plans, provided by the Road Development Agency of Zambia. Road segment locations are primarily derived from OpenStreetMap data, with satellite imagery from Google Earth used for additional verification.

Figure A.7: Predicted Paved Roads in Zambia

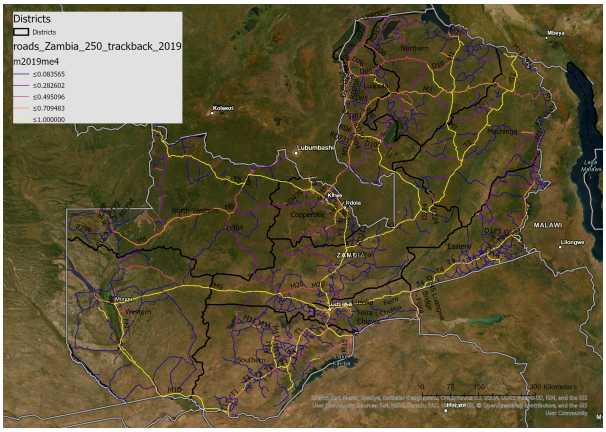
(A) ROAD CONDITION SCORE 2009



(C) CHANGES BETWEEN 2009 AND 2019



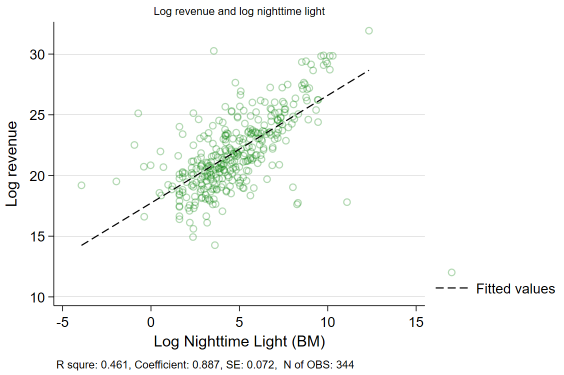
(B) ROAD CONDITION SCORE 2019



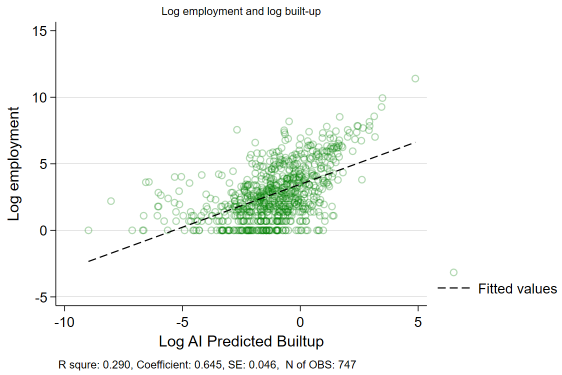
Notes: This figure illustrates AI-predicted road quality within the core road network. In the top two panels, brighter colors indicate a higher predicted probability of paved surfaces. In the bottom panel, brighter colors represent more substantial changes in road surface conditions.

Figure A.8: Correlation of Nighttime Light and Built-up and Economic Outcomes

(A) AGGREGATE REVENUE VS NIGHTTIME LIGHT

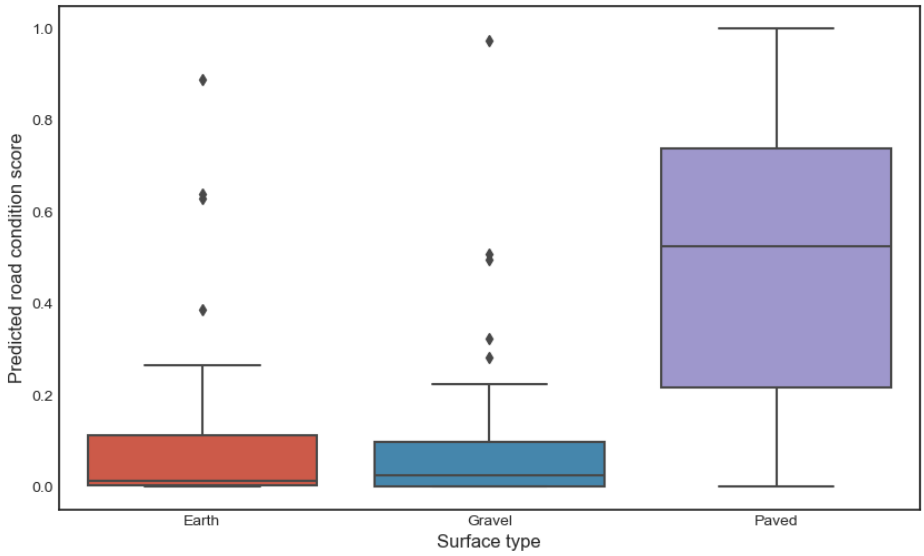


(B) AGGREGATE EMPLOYMENT VS BUILT-UP



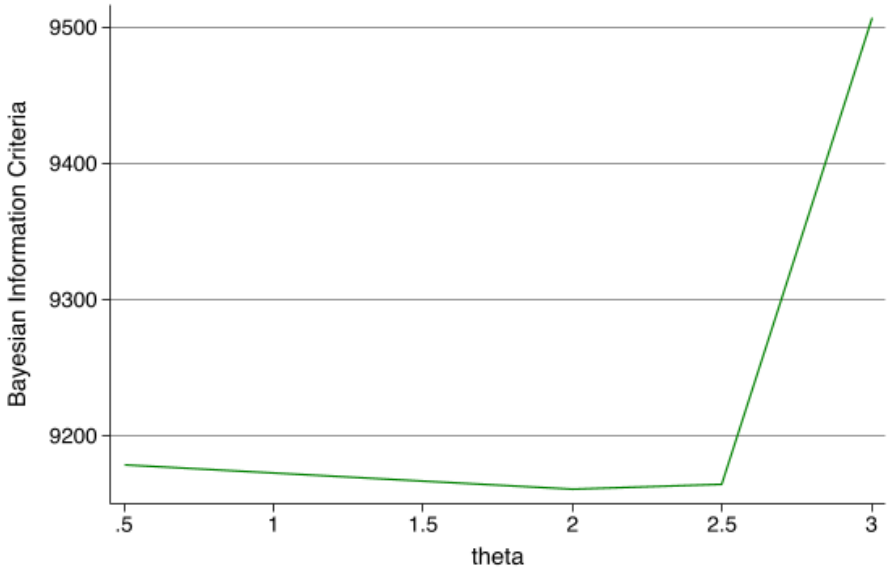
Notes: This figure displays the correlation between aggregate revenue, employment, and nighttime light. The left panel plots aggregate revenue against aggregate nighttime light, while the right panel plots aggregate employment against aggregate built-up area. Data are aggregated at the hexagon level, with revenue and employment data sourced from Zambia's geo-coded Business Establishments Register for 2011.

Figure A.9: Validation of Road Scores



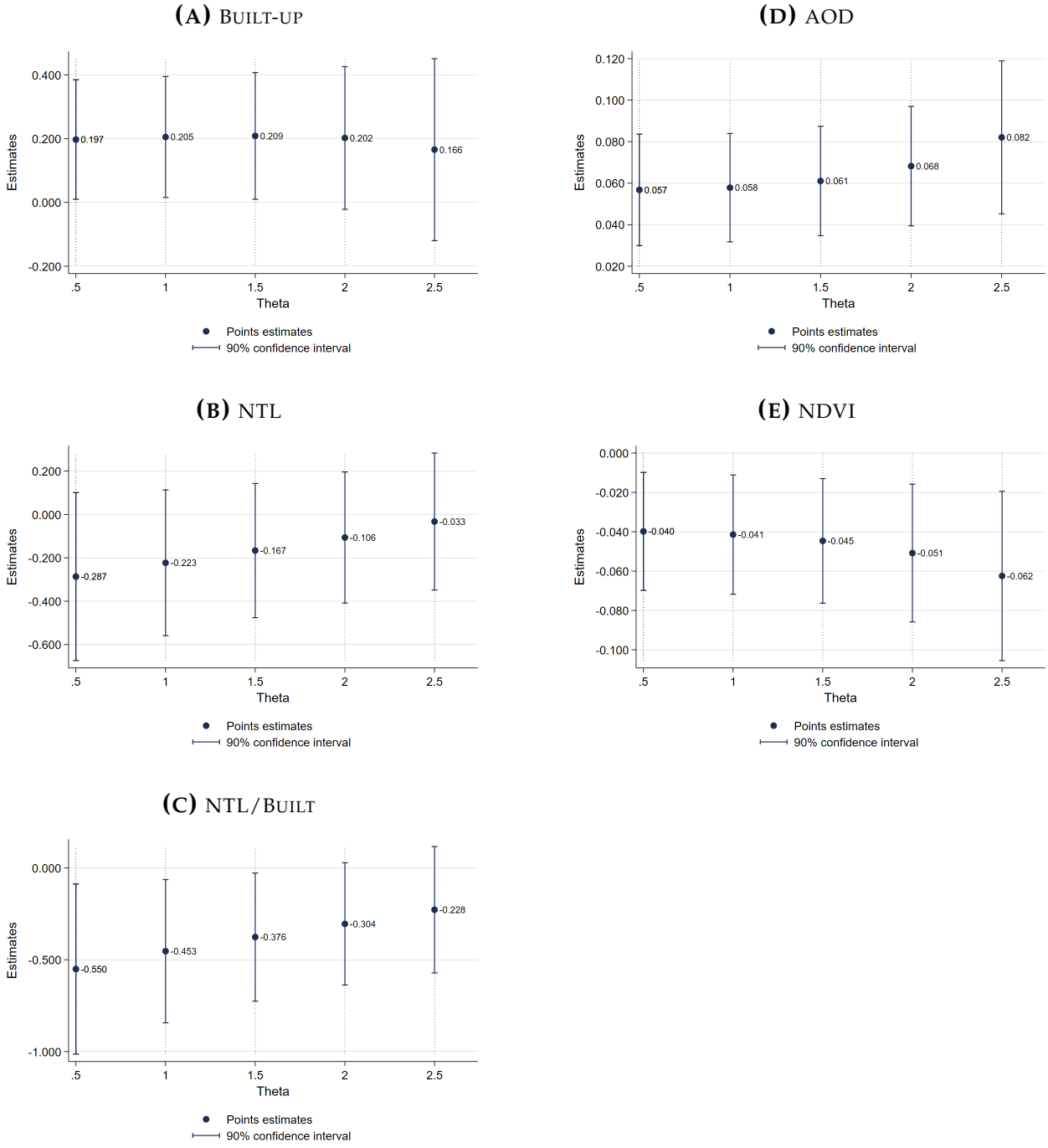
Notes: This figure compares AI-predicted road condition scores in 2014/16 across three surface types recorded in the government's 2015 Road Condition Survey. Each box displays the distribution of AI-predicted road condition scores for each type of ground truth surface condition. The sample includes 58 dirt or earth roads, 73 gravel roads, and 298 paved roads.

Figure A.10: Parameter Choice: BIC under Different θ



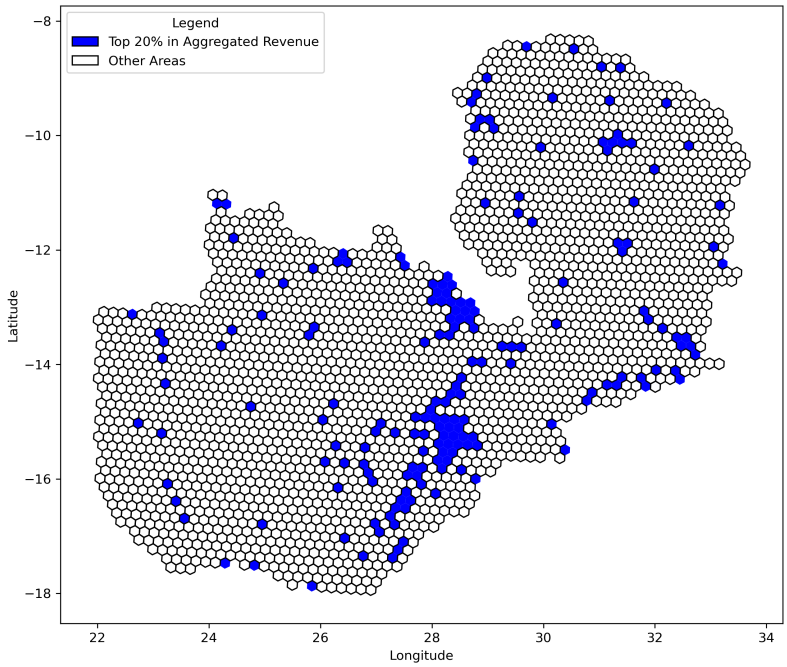
Notes: The figure shows the Bayesian Information Criterion (BIC) value under different θ for built-up. Regression specification follows column 6 in Table 1.

Figure A.11: Robustness: IV Estimates under Different θ, One-standard-deviation Increase in log Market Access



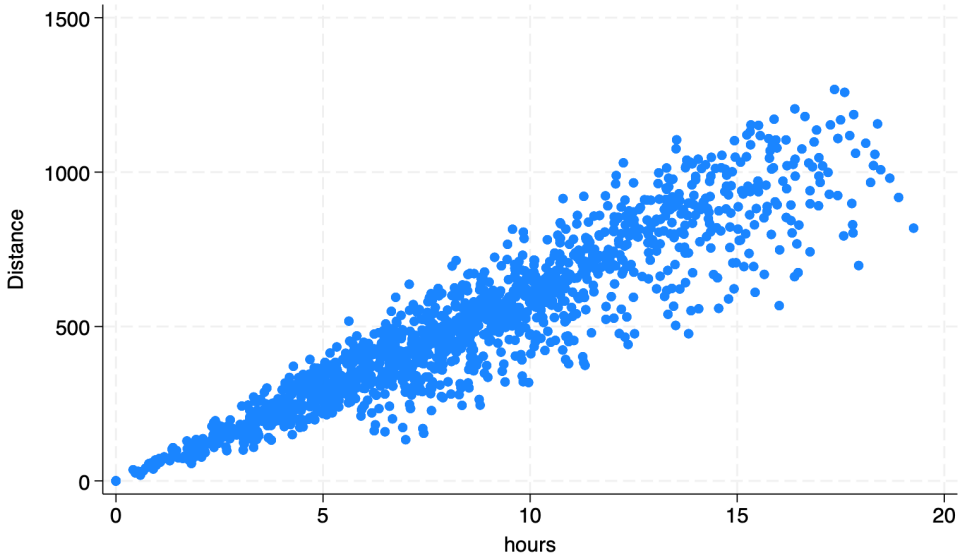
Notes: The figure shows the IV estimates under different θ for the five key outcomes in the paper. Each dot shows the IV estimates of the effect of a one standard deviation change in log market access constructed under θ. Regression specification follows column 6 in Table 1.

Figure A.12: High Manufacturing Revenue Share Locations in Zambia



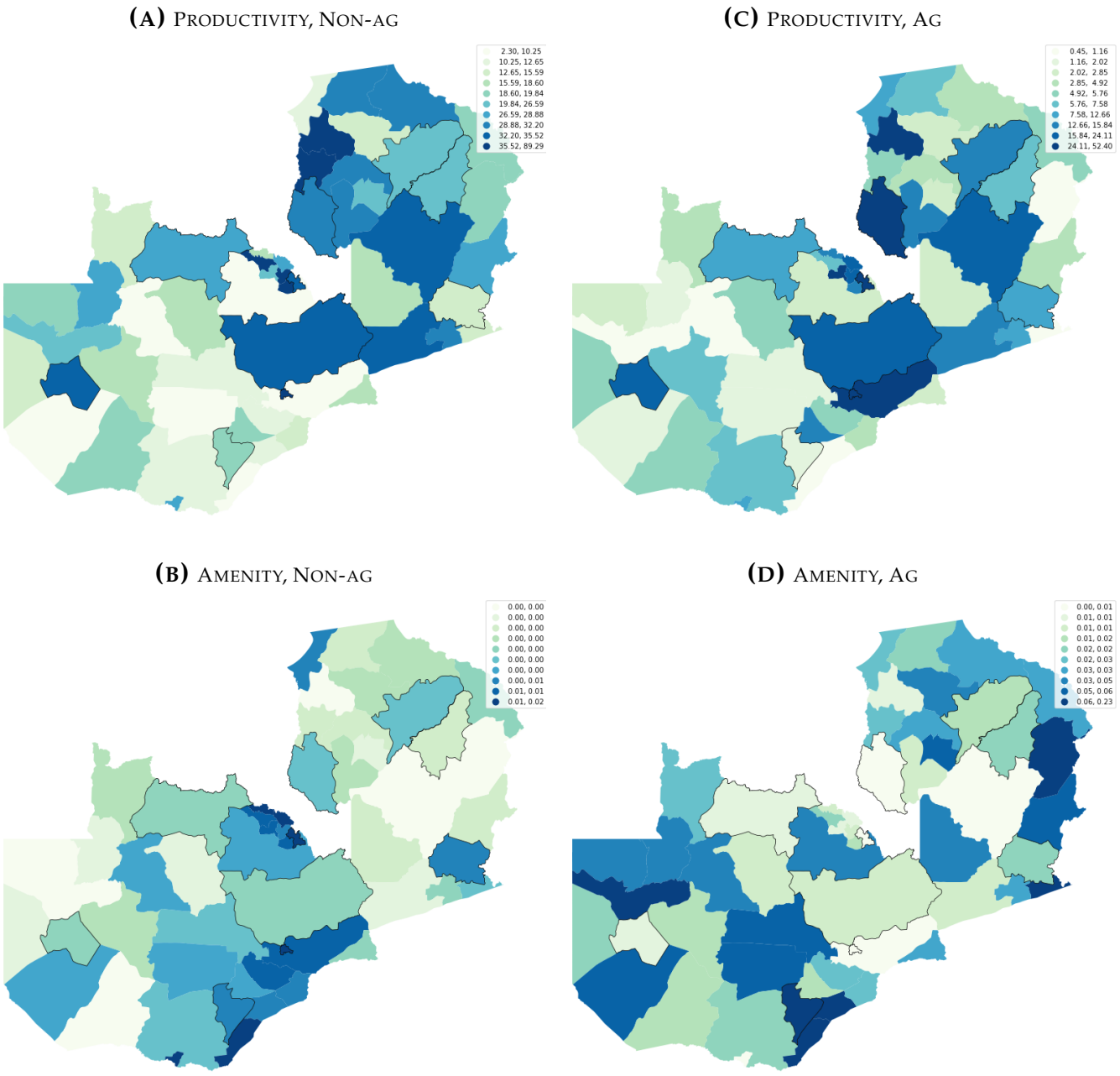
Notes: This figure maps hexagons with the top 20% highest manufacturing revenue share. Firm revenues by sector are first aggregated at the hexagon level. Hexagons shaded in blue represent those in the top 20% for aggregate revenue from manufacturing.

Figure A.13: Correlation between Physical Distance and Travel Time



Notes: This figure shows the correlation between physical distance and travel time. Each dot represents an origin-destination pair.

Figure A.14: Inverted Local Fundamentals, Exogeneous



Notes: This figure maps the inverted local fundamentals, with each panel title specifying the type, using the equilibrium conditions from equations (21) to (29).

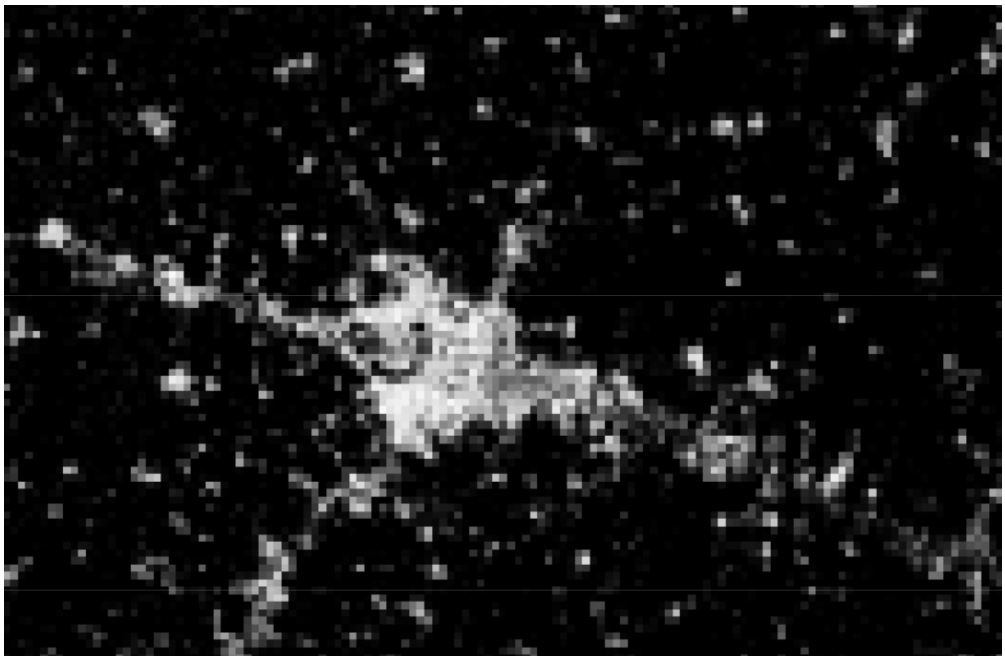
B Methods to create urban boundaries

We create urban boundaries based on built-up data following the six steps below:

- Step 1: Using an inverse distance weighted kernel, aggregate pixels from 5m resolution up to 250m resolution to obtain a smoothed built-up density (Figure B.15). This step smooths out gaps between discrete built-up pixels and reduces computational cost.

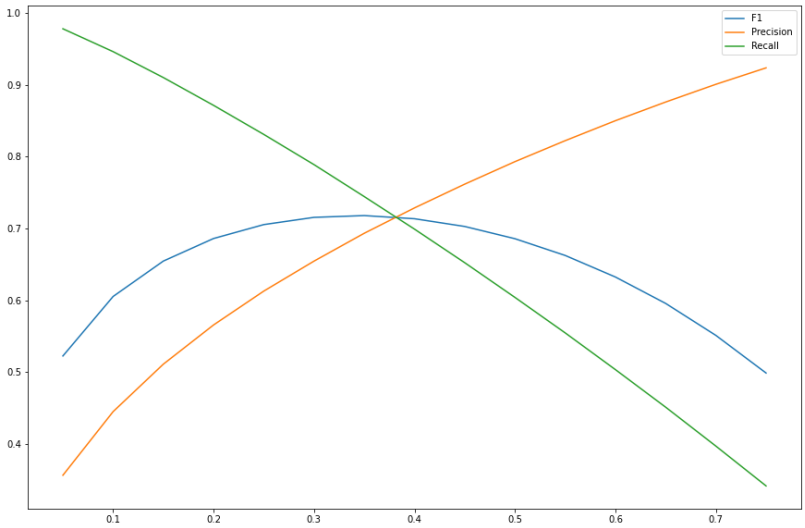
- Step 2: Pick a threshold for the smoothed built-up density p, and define urban boundaries as the pixels with smoothed built-up density above p.
- Step 3: Measure the amount of built-up or relevant information captured in the markets based on threshold p. The measure we use is the F1 score, which is accepted in the field of Information Retrieval. Specifically, $F1 = \frac{\rho * \gamma}{\rho + \gamma}$, where:
  - ρ represents precision: Given urban boundaries B defined under p, the share of built-up area within B relative to the total area of B.
  - γ represents recall: Given urban boundaries B defined under p, the share of built-up area within B relative to all built-up areas (both inside and outside of B).
- Step 4: Iterate steps 2 and 3 to find $\hat{p}$ that maximizes the F1 score, as shown in Figure B.16.
- Step 5: Generate boundaries based on $\hat{p}$. These boundaries tend to capture the core built-up areas of markets.
- Step 6: Create a 500m buffer around the boundaries and cluster the boundaries that are within 1km of each other.

Figure B.15: Smoothed built-up in 250m pixels



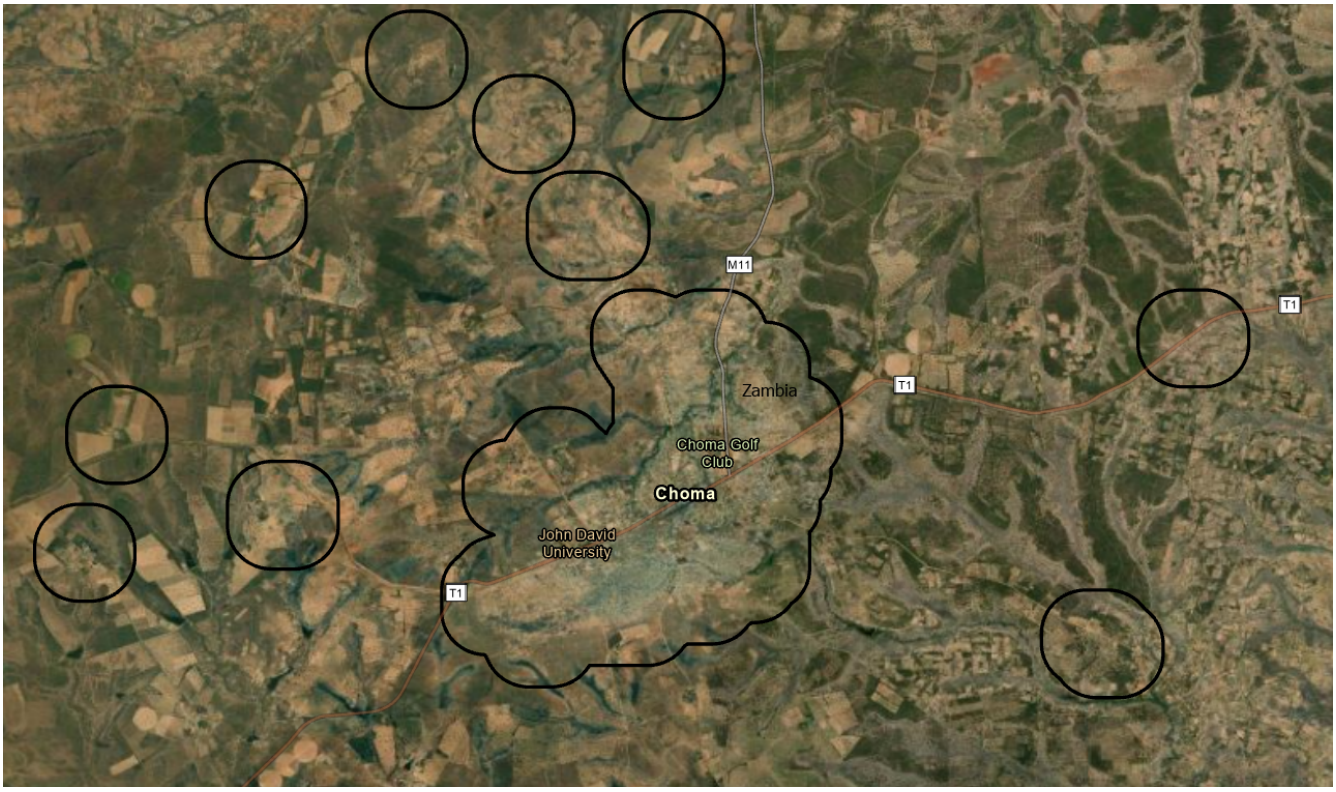
Notes: This figure illustrates the first step in defining urban boundaries. We applied an inverse distance-weighted kernel on built-up areas predicted by the deep learning model, aggregating pixels from a 5m resolution up to a 250m resolution to produce a smoothed built-up density. This process bridges gaps between discrete built-up pixels and lowers computational costs.

Figure B.16: Search for p that maximize $F1$ score



Notes: This figure illustrates how precision, recall, and the $F1$ score vary with different values of p. Given urban boundaries B defined under p, precision (ρ) measures the proportion of built-up area within B relative to the total area of B. Recall (γ) measures the proportion of built-up area within B relative to all built-up areas (both inside and outside B). The $F1$ score, calculated as $F1 = \frac{2\rho\gamma}{\rho+\gamma}$, reflects the overall accuracy of the urban boundary B. A higher $F1$ score indicates that the boundaries more effectively capture dense built-up areas.

Figure B.17: Examples of urban masks



Notes: This figure provide an example of the identification of the urban masks. Black lines denote the identified urban boundaries, which encompass cities, towns, and villages. More details are provided in Appendix Section B.

C Road surface condition classification using deep learning

C.1 Dataset Creation

The raw data consists of scenes obtained from the Planet Lab RapidEye constellation. Each scene has a resolution of 5 meters and a size of 5000 by 5000 pixels, covering an area of 25 km by 25 km. We use a rolling window technique to crop these scenes into smaller tiles, each measuring 250 by 250 pixels. These tiles are then upsampled to 256 by 256 pixels. Each tile includes five spectral bands: RGB, Red Edge, and Near Infrared. Roads from OpenStreetMap (OSM) serve as the labels, with a 5-meter buffer created around the road centerlines. Roads in OSM are generally classified into four categories:

- **Class 1 (Small Roads):** 'service', 'residential', 'track', 'living street', 'pedestrian', 'footway', 'unknown', 'path', 'steps', 'track grade1', 'track grade2', 'cycleway', 'track grade3', 'track grade4', 'track grade5', 'bridleway'
- **Class 2:** 'unclassified' (with specific definitions in OSM)
- **Class 3 (Big Roads):** 'primary', 'primary link', 'secondary', 'secondary link', 'tertiary', 'tertiary link'
- **Class 4 (Trunk/Motorway):** 'trunk', 'trunk link', 'motorway', 'motorway link' (high visibility and most likely paved)

Since smaller roads might not be easily visible at a 5-meter resolution and are more prone to mislabeling, we follow the approach of Oehmcke et al. (2019) and create four binary classifiers to detect roads from each class. First,

the tiles are ranked based on the highest classification of road present. For example, a rank 4 tile must contain a trunk or motorway road, but it may also include roads from other classes. In contrast, rank 3 tiles will not contain trunk or motorway roads.

The classifiers are designed as follows:

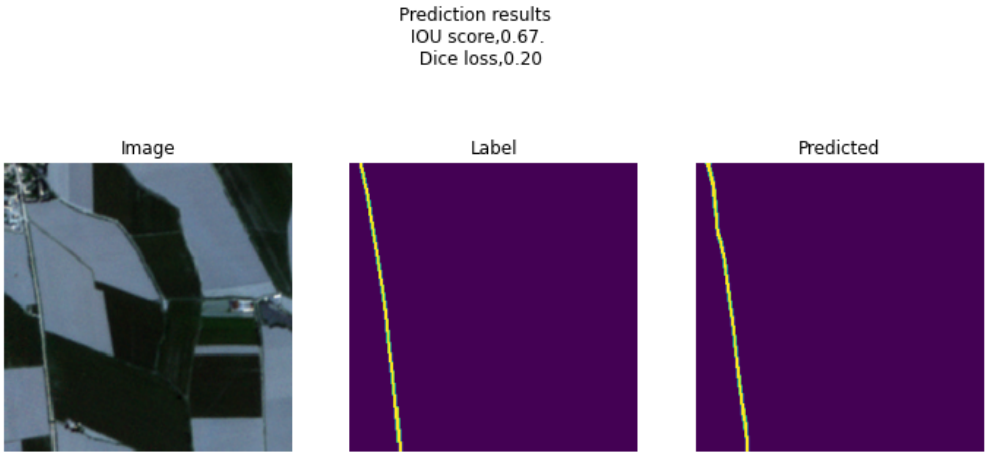
- **Classifier 1:** Detects all road classes using samples from tiles containing any type of road.
- **Classifier 2:** Targets classes 2, 3, and 4, using samples from tiles with roads of class 2 or higher (rank 2+). Smaller roads are labeled as background.
- **Classifier 3:** Targets classes 3 and 4, using samples from tiles with roads of class 3 or higher (rank 3+). Smaller and unclassified roads are labeled as background.
- **Classifier 4:** Targets trunk/motorways (class 4), using samples from tiles containing only trunk/motorway roads (rank 4). Other roads are labeled as background.

To ensure a balanced dataset for the four road classes, we limit the number of tiles in lower-ranked categories to match the number of tiles in the highest rank (26,819 labeled tiles per rank). As a result, the training dataset for Classifier 1 is four times larger than that for Classifier 4. For sample splitting, 10% of the data is reserved for validation, 10% for testing, and the remainder for training. To minimize spatial correlation in the error terms, data randomization is performed at the scene level, ensuring that tiles in the validation and test sets are, on average, 25 km away from the training tiles.

C.2 Training Details

We use random crops of size 256×256 from the tiles, followed by mean subtraction. The DLinkNet34 model (Zhou et al., 2018) is trained on the four datasets with a batch size of 128. We use the Layer-wise Adaptive Minibatch (LAMB) optimizer (You et al., 2019) with a momentum of 0.9. The learning rate follows a cosine annealing schedule, starting from 1e-4, increasing to 1e-3, and then reducing back to 1e-4. The loss function used is $1 -$ dice score. Appendix Figure C.18 presents an example of a tile cropped from the satellite images used for training, along with the corresponding labeled tiles and the predicted results.

Figure C.18: Demonstration of labeled tiles and prediction made by the deep learning model



Notes: The figure displays sample tiles used for training in the deep learning model. The left tile shows the raw input image, the middle tile represents the ground truth label, and the right tile depicts the road predicted by the model.

D Quantitative Spatial Model Appendix

Given trade costs τ_{ij}, migration costs M_{ij}, local fundamental productivity $\bar{A}_{is}$ and amenities $\bar{u}_{is}$, initial populations L_i^{0e}, preference parameters $\{\sigma, \eta\}$, and production technology parameters $\{\theta_i, \rho\}$, an equilibrium is defined over a series of endogenous variables $\{w_{is}, L_{is}, \omega_i, \phi_i, p_{is}, P_i\}$ such that,

- **Labor Market Clearing:** Labor demand stated in Eq.(17) and Eq.(17) equals labor supply stated in Eq.(11) for each location i and each sector s.
- **Goods Market Clearing:** For each location i and each sector s, the total payment to labor equals total sales, and the total expenditure equals to total payments to goods.
- **Closed Model:** The model is closed in the sense that the total population is constant.[35]

The equilibrium can be formally expressed as follows,

$$w_{is}L_{is} = \sum_{j\in N}\left(\frac{{P_{js}}^{1-\xi}}{\sum_{s'\in S}{P_{js'}}^{1-\xi}}\right)\left(\frac{p_{ijs}^{1-\sigma_s}}{\sum_{i'\in N}p_{i'js}^{1-\sigma_s}}\right)w_{js}L_{js} \tag{21}$$

$$P_{js} = \left(\sum_{i'\in N}p_{i'js}^{1-\sigma_s}\right)^{\frac{1}{1-\sigma_s}} \tag{22}$$

$$L_j = \sum_{i\in N}\frac{\left(\frac{\Phi_j}{P_jD_{ij}}\right)^{\eta}}{\sum_{j'}\left(\frac{\Phi_{j'}}{P_{j'}D_{ij'}}\right)^{\eta}}L_i^0 \tag{23}$$

$$W_i = \sum_{j'}\left(\frac{\Phi_{j'}}{P_{j'}D_{ij'}}\right)^{\eta} \tag{24}$$

$$\tag{25}$$

We denote $\omega_j = \left(\frac{\Phi_j}{P_j}\right)^{\eta}$, $W_i = \sum_{j'}\left(\frac{\Phi_{j'}}{P_{j'}D_{ij'}}\right)^{\eta}$, $T_{ij} = {\tau_{ij}}^{1-\sigma_s}$ and $M_{ij} = {D_{ij}}^{-\eta}$, then we have the following system that can be used to invert the price and welfare composites.

$${p_{is}}^{\sigma_s-1} = \sum_{j\in N}\left(\frac{{P_{js}}^{\sigma_s-\xi}}{\sum_{s'\in S}{P_{js'}}^{1-\xi}}\right)T_{ij}\left(\frac{Y_{js}}{Y_{is}}\right) \tag{26}$$

$$\left({P_{js}}^{\sigma_s-1}\right)^{-1} = \sum_{i\in N}\left({p_{i's}}^{\sigma-1}\right)^{-1}T_{i'j} \tag{27}$$

$${\omega_j}^{-1} = \sum_{i'\in N}M_{i'j}\left(\frac{L_i^0}{L_j}\right){W_i}^{-1} \tag{28}$$

$$W_i = \sum_{j'\in N}M_{ij'}\omega_{j'} \tag{29}$$

Since $\omega_j = \left(\frac{\Phi_j}{P_j}\right)^{\eta}$, we can use the inverted ω_j and wage data to recover Φ_j. Then we can combine Φ_j and the sectoral composition of each location, $\pi_{is|j} = \frac{(u_{js}w_{js})^{\nu}}{\sum_{s'}(u_{js'}w_{js'})^{\nu}}$, to back out the amenity level of each location.

[35] The initial population of a location should equal to the total out-flows of population to all location (include stayers), and the total population should equal to the in-migration from all locations (include stayers).

CENTRE FOR ECONOMIC PERFORMANCE
Recent Discussion Papers

2061	Gabriel M. Ahlfeldt Fabian Bald Duncan Roth Tobias Seidel	Measuring quality of life under spatial frictions
2060	Tessa Hall Alan Manning	Only human? Immigration and firm productivity in Britain
2059	Shadi Farahzadi	The integration penalty: Impact of 9/11 on the Muslim marriage market
2058	Italo Colantone Gianmarco I.P. Ottaviano Kohei Takeda	Trade and intergenerational income mobility: Theory and evidence from the US
2057	Maria Guadalupe Veronica Rappoport Bernard Salanié Catherine Thomas	The perfect match: Assortative matching in mergers and acquisitions
2056	Fabrizio Leone	Global robots
2055	Luca Fontanelli Flavio Calvino Chiara Criscuolo Lionel Nesta Elena Verdolini	The role of human capital for AI adoption: Evidence from French firms
2054	Saul Estrin Andrea Herrmann Moren Lévesque Tomasz Mickiewicz Mark Sanders	New venture creation: Innovativeness, speed-to-breakeven and revenue tradeoffs
2053	Stephen J. Redding	Quantitative urban economics

2052	Esteban M. Aucejo Spencer Perry Basit Zafar	Assessing the cost of balancing college and work activities: The gig economy meets online education
2051	Jonathan Colmer Suvy Qin John Voorheis Reed Walker	Income, wealth and environmental inequality in the United States
2050	Debopam Bhattacharya Ekaterina Oparina Qianya Xu	Empirical welfare analysis with hedonic budget constraints
2049	Jonathan Colmer Eleanor Krause Eva Lyubich John Voorheis	Transitional costs and the decline in coal: Worker-level evidence
2048	Ekaterina Oparina Andrew E. Clark Richard Layard	The Easterlin paradox at 50
2047	Stephen J. Redding	Spatial economics
2046	Stephen Machin Matteo Sandi	Crime and education
2045	Hanno Foerster Tim Obermeier Bastian Schulz	Job displacement, remarriage and marital sorting
2044	Randi Hjalmarsson Stephen Machin Paolo Pinotti	Crime and the labor market
2043	Emanuel Ornelas	Political competition and the strategic adoption of free trade agreements