Kent Academic Repository

Full text document (pdf)

Citation for published version

Ngulube, Mumba (2021) Essays on the Economic Impacts of Upgraded Highways in Zambia. Doctor of Philosophy (PhD) thesis, University of Kent,.

DOI

Link to record in KAR

https://kar.kent.ac.uk/91575/

Document Version

UNSPECIFIED

Copyright & reuse
Content in the Kent Academic Repository is made available for research purposes. Unless otherwise stated all content is protected by copyright and in the absence of an open licence (eg Creative Commons), permissions for further reuse of content should be sought from the publisher, author or other copyright holder.

Versions of research
The version in the Kent Academic Repository may differ from the final published version.
Users are advised to check **http://kar.kent.ac.uk** for the status of the paper. **Users should always cite the published version of record.**

Enquiries
For any further enquiries regarding the licence status of this document, please contact:
researchsupport@kent.ac.uk

If you believe this document infringes copyright then please contact the KAR admin team with the take-down information provided at **http://kar.kent.ac.uk/contact.html**

School of Economics
University of Kent

Essays on the Economic Impacts of Upgraded Highways in Zambia

Mumba Ngulube

Thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy in the School of Economics at the University of Kent - October, 2021.

Abstract

This thesis aims to contribute to the literature on the economic impacts of upgraded highways in developing economies by using firm-level panel data and global nighttime light data from remote sensing technology. The thesis consists of three self-contained essays as follows.

The first essay, entitled "***Improved Roads and Firm Performance: Evidence from Zambia***," investigates the impact of improving a section of the Nacala Road Corridor in Zambia on the performance of firms across 64 sectors. The analysis in this essay is twofold. First, I investigate the impact of a fall in transport costs due to the improved highway on firms located in bands at different distances from the highway. Second, I investigate the impact on firms of a change in accessibility to Port Nacala resulting from the improved highway. In both cases, a number of firm outcomes are analysed, including sales, labour costs, assets and transportation costs. A difference-in-differences estimation technique is applied to compare outcomes in firms close to the improved road or in close proximity to the seaport with those of firms far away. I use a new district-referenced firm panel dataset for the period 2013 to 2017 and show that after the road is improved, compared to firms far away, firms in the peripheral area and close to the road experience positive growth in fuel and lubricant costs. These firms also experience negative growth in assets. However, effects on firm output, labour costs, freight transport costs and raw material inventory costs are inconclusive. Results on the effect on firms of a change in accessibility to Port Nacala yields inconclusive results. This essay contributes to the literature by using a new district-referenced, disaggregated firm panel dataset from a country in Southern Africa to study the effects of improved highways. The impact on a number of firm outcomes of the improved highway is inconclusive, suggesting that it is not easy to observe spatial effects of improved highways on firm economic activity across many sectors.

The second essay, entitled *"**Improved Highways, Transportation Costs and Trade: Firm Panel Data Evidence from Zambia**,"* examines the effects on the performance of firms engaged in international trade in a landlocked country of reduced transport costs arising from improving an international highway leading to a seaport. This essay considers whether the spatial proximity to an improved highway may have distinct differential impacts on firms in the tradable sector compared to firms in the non-tradable sector over time. Motivating this essay is the debate that Africa trades very little with itself and the rest of the world because of poor transport infrastructure among other factors. This chapter explores how the removal of a trade obstacle of poor-quality transport infrastructure impacts firms engaged in international trade. Relying on a triple difference strategy and using firm panel data for 2013 to 2017, I estimate

the effect of the upgraded highway on the output, cost of sales, labour costs and transportation costs of firms in the tradable sector compared to those in the non-tradable sector. Results show that the improved international highway brings gains in trade of increased exportable output and reduced inventory costs. Relative to firms in the non-tradable sector, firms in the tradable sector in the peripheral area close to the road experience a large positive growth in gross profits, sales, fuel costs and labour costs after the highway is improved. The upgraded highway has insignificant effects in the centre which, before the intervention, is already star connected to other efficient transport links. Results in this essay contribute by providing a better understanding of how reduced transport costs to a seaport impact international trade activities in a landlocked country. Results also suggest that investing in international highways can be an effective policy to promote increased economic activity in the tradable sector outside metropolitan cities.

The last essay entitled *"**Estimating the Impact of Upgraded Highways Using Global Nighttime Light Data in Zambia**,"* aims at measuring the impact of transport cost reductions arising from upgraded highways on economic activity measured at a local level of grid cells of 0.1 decimal degrees (approximately $11\,km^2$). In the existing literature, nighttime light data reflect real economic activity and are correlated with the true GDP. Based on this, I use the harmonised nighttime light remote sensing data to proxy economic growth in the absence of economic data at the grid cell level. Relying on a difference-in-differences strategy, I compare outcomes in grid cells close to the upgraded highway with grid cells equidistant to highways not upgraded over time. The event study method is also used to estimate the temporal effects of the treatment during and after construction. Results from both the difference-in-differences methodology and the event study indicate that grid cells surrounding the upgraded highway experience positive growth in nighttime lights after the highway is upgraded. I translate the growth in nighttime lights associated with the upgrade of the highway to infer growth of approximately 21 per cent in GDP at the local grid cell level after the highway is upgraded. This essay provides evidence of the impact of reduced transport costs on GDP measured at a local level using harmonised nighttime lights data in a landlocked developing country.

Acknowledgments

I start by thanking the University of Kent, School of Economics for awarding me the Doctoral Research Scholarship to pursue this degree. Further, I wish to thank everyone that supported and influenced me during my PhD journey.

To Dr. Amrit Amirapu my first supervisor, I appreciate the guidance and patience shown during my gradual understanding of concepts and methodologies. To my second supervisor Dr. Alex Klein, thank you for the invaluable feedback. I am also grateful to the Faculty of Economics for the brilliant suggestions and to the Development Economics Research Centre team and the Kent/Sussex Virtual PhD Seminar members for the constructive feedback that enriched my work.

Special gratitude also goes to the officers at the Zambia Revenue Authority, the Road Development Agency, Ministry of Finance and the African Development Bank for providing the data for this research.

To all my friends, thank you for politely pretending to be interested in my work. To my friends in Canterbury, you made my stay memorable. Oluwakemi and Ayo, you were great co-parents to my kids and helpful in my low moments. To Goher and family, thank you for always being kind and loving. To my Zambian brother in Canterbury, Dr. John, God could not have placed you in our lives at a better time. To Sevgi my office mate, your advice was priceless. To Lucas, thank you for always checking up on me and sparing time to discuss research. To my friends in Zambia (Ruwezi, Nancy, Mr. Banda and Mr. Mwape) and Andrea in South Africa, you brought a healthy balance twist to my life in our countless virtual interactions, thank you.

To my family, I owe you special thanks. Thus, I dedicate this thesis to my husband Eustace Zulu for his unwavering support and love, whilst I and my kids stayed away from home for four years. To my two children, Ethan Tawanda Zulu and Elsa Sarah Zulu, you gave me the strength to complete my studies. To my mother-in-law, Mrs. Angela Zulu, you played a critical part by traveling to provide emotional and family support, thank you. To my cousin Ngabwa and my brother-in-law Ashley, you were fantastic co-parents during my conference trips outside Europe. To my dad Mr. John Ngulube thank you for showing me unconditional love and support. To my brothers and sister, I love you. Finally, to the memory of a great woman, my beloved mother Mrs. Alice Ngulube (1960 to 2008), I know you are proud of me and smiling in heaven.

Contents

List of Figures

List of Tables

1 Introduction

Globally, transport infrastructure is recognised as crucial for economic development.[1] It promotes sustainable and inclusive development by opening up areas to increased economic activity.[2] It is therefore, no wonder that building reliable, quality and resilient infrastructure is one of the seventeen United Nations Sustainable Development Goals. In Africa, development efforts are reportedly stifled by poor transport infrastructure, which increases transportation costs. Further, compared to coastal countries, landlocked countries are said to experience higher transport costs, which account for 60 per cent of the cost of goods (Lmao and Venables, 2001). Half of the world's poor resides in Africa, and the continent lags behind others for every kind of infrastructure (World Bank, 2018; Yepes, Pierce and Foster, 2008). In addition, in the past two decades, governments in Africa have invested large amounts of public resources in transport infrastructure to promote trade, economic growth and access to social services (Jedwab and Storeygard, 2017).[3] However, despite significant investment in road infrastructure in the region, there is scant research investigating the impacts of such investments on economic activity at a microeconomic level. Many of the studies on this topic are macroeconomic studies on developed countries like the United States (US) (Duranton, 2015; Micheals, 2008). Few studies are conducted on Africa because of a lack of economic data, especially at the subnational level. Another reason to conduct this study on Africa is that the continent is distinct from others. It includes a large number of landlocked countries (15) and small economies with a human population of less than 20 million and a GDP of US$ 10 billion or less (Vivien and Briceño-Garmendia, 2010). Therefore, Africa faces different initial conditions such that results may vary from those found elsewhere.

Based on the above observations, the three essays that follow attempt to bridge the gap in the literature by using firm-level panel data and nighttime light remote sensing data to study the impact of upgrading a highway in a landlocked country in Africa.

In the first essay, I investigate the impact of improving a section of the Nacala Road Corridor – an international highway terminating at a trade seaport – on the performance of firms in sixty-four sectors in Zambia. I identify two changes that occur when a highway is improved and thus conduct two investigations. First, I analyse the impact of a fall in transport costs due to the highway improvements on firms located in different distance bands from the highway.

[1] See Gurara et al (2018) and Ivanova and Masarova (2013).

[2] (Holl, 2016).

[3] Roads receive 50 to 60 percent of total government spending on infrastructure in Africa (Vivien and Briceño-Garmendia, 2010).

Second, I study the impact on firms of a fall in costs of transport to a trade seaport. I build a simple theoretical model to provide a clear understanding of how improved roads affect economic activity through firms. This model takes into consideration sectoral heterogeneity in production technologies and shows how changes in the price of transportation services affect firm output, demand for labour and other inputs in production. I estimate results using the difference-in-differences (DID) methodology. Using firm-level district-referenced panel data for 2013 and 2017, I compare outcomes between firms within 400 km of the upgraded highway with firms outside this radius over time. The distance bandwidth of 400 km is used because there are few firms in districts close to the upgraded highway in distance bandwidths less than this.[4] Road development projects are prone to endogeneity concerns; there is a concern that areas that receive upgraded roads may systematically be different from areas that do not. Roads are usually developed in a particular area because either it evidences trends of high economic growth and roads are developed to deal with the increasing economic activity. Alternatively, the area shows low economic growth and roads are developed to attract economic activity to the area. I deal with the endogeneity concern of firm self-selectivity by incorporating firm fixed effects in all estimations, while unobserved factors affecting all locations equally over time are controlled using time fixed effects. Endogeneity concerns associated with reverse causality are addressed by running estimations on a sample that excludes the endogenously chosen capital city, Lusaka.[5] Results indicate that after the road is improved, compared to those far away, firms in the peripheral area within 400 km of the road, experience positive growth in fuel costs. The effect of the upgraded highway on most firm outcomes is inconclusive given the available data. In the final empirical specification, I assess how a fall in transport costs to the Port of Nacala affects firm outcomes. A change in the accessibility of the port is captured by a rank variable that measures how close a firm is to the Port of Nacala in comparison to other seaports. Results on the effect of a change in accessibility to the Port of Nacala are also inconclusive.

The second essay investigates the effect of improving a section of the Nacala Road Corridor on firms engaged in international trade in Zambia. Existing literature points out that Africa trades too little with itself and the rest of the world because of high transport costs arising from poor infrastructure (Amjadi and Yeats, 1995; Sequeira, 2011).It is important to know how and to what extent upgraded highways impact participation in international trade in order to increase the participation of African countries. Using firm-level panel data and a triple difference estimation strategy with location and time fixed effects, I estimate the impact over time in the tradable sector of reduced transport costs arising from the upgraded highway

[4] There are 16 firms within 180 km of the improved road, 14 firms in 2013 and 2 firms in 2017. Using firm fixed effects in regression analysis drops all singletons in the estimation. To increase the number of firms in the treatment group appearing in both years, the driving distance buffer is extended to 400 km from the upgraded highway. In the distance band of 400 km there are 94 firms in 2013 and 73 firms in 2017.

[5] The Nacala road corridor was built and upgraded to connect the capital city of Zambia and the Capital city of Malawi to the port of Nacala at the shortest driving distance. Thus, locations laying in between are included in the treatment by virtual of being located along this area and not reasons connected to the outcomes of interest in this study. These areas in-between the two capital city are exogenously treated by the upgraded highway. Chandra and Thompson (2000) use this same strategy in studying the economic effects of the interstate highways in the United States.

on gross profit, sales and various firm costs. The triple difference methodology compares the double difference in the tradable sector to the double difference in the non-tradable sector. It thus uses an additional within-state control group (the non-tradable sector) to remove location-specific time-varying confounding factors. Results indicate that in the sample excluding the endogenously chosen city of Lusaka, relative to firms in the non-tradable sector, firms in the tradable sector near the upgraded highway experience a large positive growth in gross profits, sales, fuel costs and labour costs after the upgrade.

The third and final essay investigates the impact of upgrading a section of the Nacala Road Corridor in Zambia on gross domestic product (GDP) measured at a local level of grid cells of 0.1 decimal degrees (approximately $11\,km^2$). Due to the absence of GDP data at this fine-grained level, nighttime light remote sensing data is used as a proxy;[6] harmonised global nighttime light data covering the period 2013 to 2018 is used.[7] Using the DID methodology, I compare the growth in nighttime lights over time in grid cells located near the upgraded highway to that in grids an equal distance from highways not upgraded. In the analysis, variation in exposure to treatment within the grid cells is captured by calculating the straight-line distance from each grid centroid to its nearest highway [8]In another DID estimation, the effect on grid lights of the improved highway is estimated using the minimum travel cost data. Using an event study methodology, I estimate the temporal effects of the treatment during and after construction. Estimates from the DID and the event study indicate that grid cells surrounding the intervened highway experience positive growth in lights after the highway is upgraded. Lastly, I translate the growth in lights at the grid level to determine what it means at the local level in terms of GDP growth.[9] Results indicate that after the highway is upgraded, grid cells in close proximity to the intervened highway experience GDP growth that is 21 per cent higher than that of grids equidistant to highways not upgraded.

The three essays in this thesis provide a better understanding of the impact of upgraded highways in landlocked countries in Africa. The essays inform policymakers and researchers of the effects of upgraded highways on economic activity measured at a subnational level using firm data for the region. Knowledge is also built on the impact of upgraded highways on the tradable sector and on GDP measured at a fine local level using nighttime light remote sensing data.

[6] A number of studies show that nighttime light data is a good predictor of GDP especially in developing countries where development is infrastructure driven (Henderson et al, 2012).

[7] There are two types of nighttime light data, the VIIRS data and the DMSP data. The DMSP data is available from 1992 to 2013 and the VIIRS data from 2012 to date. The road project analysed was upgraded in 2014 and thus, the VIIRS data coverage coincides with the road construction project. To estimate road effects, I use the DID methodology where I need to check whether the outcome variables show parallel trends in the pre-treatment period in the control and treatment groups which is reassuring for the interpretation of estimates as causal. The VIIRS data cannot be used to check for common pre-trends as it has data for only one pre-treatment period. Thus, I use the harmonized global nighttime lights dataset made available by Li et al (2020). This dataset combines the DMSP data from 1992 to 2013 with the VIIRS data from 2013 to 2018. In the harmonized dataset, the VIIRS data is simulated to be like the DMSP data for easy comparability.

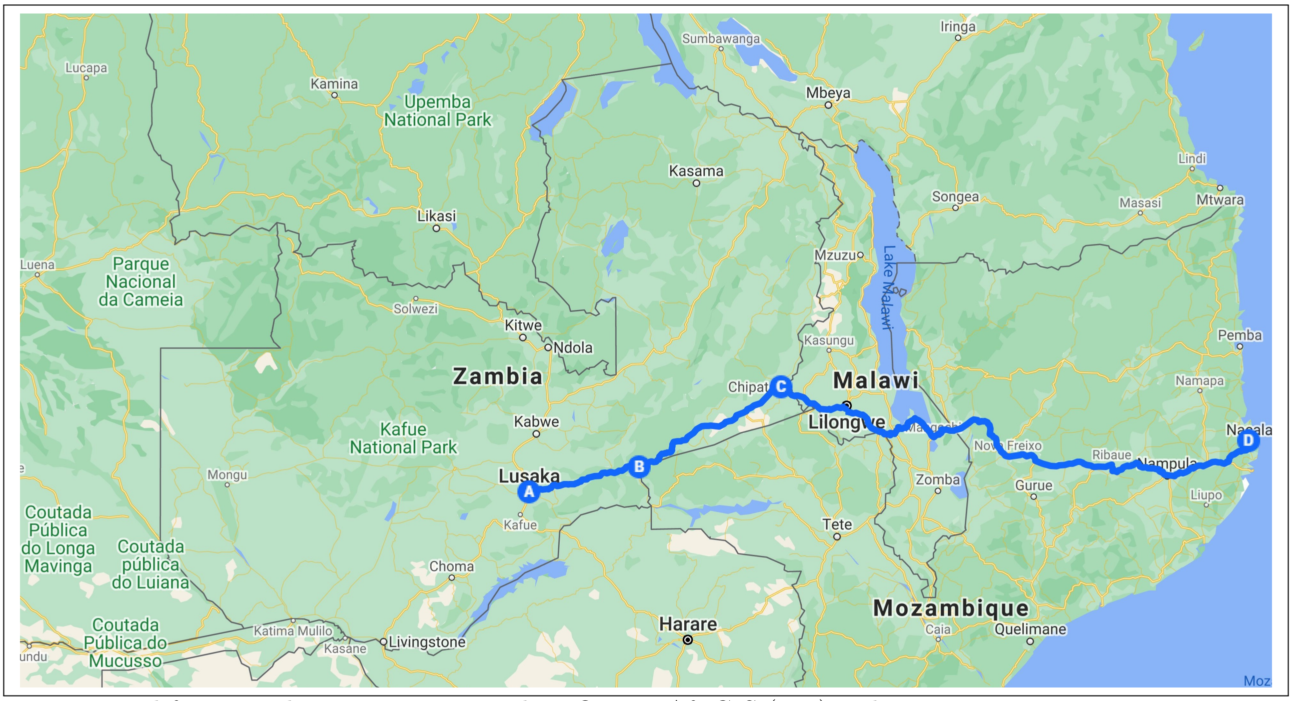
[8] Grids are grouped into distance bins using the straight-line distance information.

[9] I use the geospatial Demographic Health Survey data for the years 2013 and 2018 for Zambia to calculate the GDP-lights correlation coefficient.

2 Background on the Highway Project and Country of Study

The three essays presented in this thesis seek to investigate from a specific context the economic impact of upgrading a section of the Nacala Road Corridor in Zambia. The Nacala Road Corridor is one of the large transport infrastructure development projects implemented to improve the core road network in the Southern Africa Development Community (SADC) at an estimated cost of USD 841.51 million (UA 614.24 million).[1] Works under this project involve the reconstruction of more than 1,165 km of selected sections along the Nacala Road Corridor and the construction of two one-stop border posts. The Nacala Road Corridor stretches from Lusaka in Zambia, passing through Malawi and northern Mozambique to the Port of Nacala.

Figure 2.1: Map of Nacala Road Corridor



Map accessed from google maps using Map data @ 2021 AfriGIS (Pty) Ltd.
Note: Location A is Lusaka city, B is Luangwa Bridge, C is Mwami Border and D is Port Nacala.

This corridor is part of the Trans-Africa Highway System and is one of the core transport corridors in Southern Africa identified under the SADC Program for Infrastructure Development in Africa. The SADC Regional Infrastructure Development Master Plan (2012–2020) also

[1] UA is Africa Development Bank (AfDB) currency.

emphasises the centrality of Zambia in the quest to accelerate and facilitate the development of regional transit and transport infrastructure (Ministry of transport and communication, 2019). Source documents from the Africa Development Bank (AfDB) indicate that the transport corridor was established to provide the shortest driving distance from north-eastern Zambia, Malawi and northern Mozambique to Port Nacala. The upgrade of the corridor aims to catalyse regional economic integration and trade competitiveness of the SADC region. Other objectives of the project include increasing the traffic volume on the corridor from 14 per cent in 2014 to 25 per cent by 2025, reducing the travel time in rehabilitated road sections by 30 per cent and reducing vehicle-operation costs by 30 per cent. The transport project was initially set to be executed in three phases, but this was later revised to five. Table 2.1 sets out the details of the five project phases.

Table 2.1: Road Project Phases

Phase	Section Description	Project Approval Date	Project Completion Date	Estimated Cost USD	Funding Institutions	Status of Project
I	Reconstruction of 486 km Nampula-Cuamba-Muita road in Mozambique and 14 km of a new road bypass in Lilongwe city in Malawi.	24/06/2009	December 2021	$246.64 Million	AfDB, JICA, Korean Exim Bank and Government of Mozambique	Progress is 75%
II	Reconstruction of 360 km of road sections in Zambia, from Luangwa bridge to Mwami, border with Malawi.	27/09/2010	December 2021	$326.57 Million	AfDB, EIB, EU and Government of Zambia	Completed
III	Reconstruction of 175 km Muita-Massangulo-Lichinga road in Mozambique completed in 2019.	23/10/2013	December 2021	$136.53 Million	AfDB, JICA and Government of Mozambique	Progress 65%
IV	Reconstruction of 75km Liwonde-Mangochi road in Malawi, and construction of a One-Stop-Border-Post (OSBP) between Malawi and Zambia at Mchinji.	12/2013	December 2021	$69.71 Million	AfDB, Government of Malawi and Government of Zambia	Progress is 70%
V	Reconstruction of a 55 km Nsipe-Liwonde road in Malawi, and construction of a One-Stop-Border-Post (OSBP) between Malawi and Mozambique at Chiponde	06/2019	December 2024	$62.05 Million	EU and Government of Malawi	In progress

Source: African Development Bank.

The three essays in this thesis focus on phase II of the Nacala Road Corridor, which is located in Zambia. During the time of research this was the only section of the transport corridor fully completed.[2] Phase II of the corridor is the section between B and C in Figure 2.1.

[2]One caveat to note in my research is that the full impact of Phase II of the construction project may not be felt until some of the other phases are complete.

At a national level, this large transport infrastructure development project responds to Zambia's needs to improve its poor road network. Zambia is a landlocked lower-middle-income country located in Southern Africa. It is one of the largest and most sparsely populated countries in the world, and, not surprisingly, its road network is one of the longest, per person and by GDP, in Africa (Raballand and Whitworth, 2011). The population of Zambia is approximately 18 million, spread over a land area of 752,614 square kilometres. In Zambia, roads are the dominant mode of transportation carrying over 90 per cent of passengers and close to 97 per cent of goods (World Bank, 2014). A large part of the road network was constructed during the first decade after independence in 1964, when Zambia was one of Africa's most prosperous economies (Raballand and Whitworth, 2011).[3] However, as of 1990, when a road assessment was conducted by the World Bank, only 20 per cent of the total road network of 67,671 km was in good condition (World Bank, 1997). Zambia is a copper-dependent country, and collapsing revenue from the copper mines immediately after nationalisation, combined with general economic mismanagement, led to large fiscal deficits and debt accumulation, resulting in discretionary fiscal resources drying up; after wages and debt servicing, virtually nothing remained (Raballand and Whitworth, 2011). Road maintenance was an inevitable casualty of the economic collapse (Raballand and Whitworth, 2011).

Transport costs in Zambia are among the highest in the region because of poor road infrastructure combined with long distances between locations (Limao and Venables, 2001). Good-quality, reliable transport systems are needed to connect its widely dispersed rural population countrywide and for international trade. In 2012, the Zambian government made a policy decision to increase its investment in road infrastructure development. The objective was to construct, rehabilitate and upgrade Zambia's core road network of 40,454 km (out of a total network of 67,671 km) deemed the road stock necessary to spur economic development. In the same year, the government launched the Link Zambia 8000 Project with the aim of transforming Zambia into a highly road-linked landlocked country by 2017. However, as of 2018, only 9 per cent of road construction progress was achieved under the project. This is mainly because the 42 road construction projects launched in 2012 as a goal of the Link Zambia 8000 Project put considerable financial pressure on the country's treasury. Construction on the Nacala Road Corridor was completed in 2017 due to joint funding by the Zambian government, African Development Bank, European Union, European Investment Bank and Agence Francaise de Développement.

In Zambia, the total length of the Nacala Road Corridor is 592 km. Phase II of the highway project analysed in this thesis upgrades the 360 km length of road that runs from Luangwa Bridge to Mwami border. The old road rehabilitated under phase II of the Nacala Road Corridor Project was constructed between 1967 and 1972. Source documents from the Road Development Agency in Zambia show the old road as a 6.1 m wide single carriageway with 1.8 m shoulders built using crushed stone on a natural gravel sub-base with 40 mm asphalt and

[3] In 1964 Zambia had the fourth highest GDP per capita in Sub-Saharan Africa after South Africa, Gabon and (South) Rhodesia (NationMaster.com) (Raballand and Whitworth, 2011).

concrete surfacing. The condition of the road was poor; it had been extensively and roughly patched in several emergency maintenance interventions in the 1990s. The old road design did not meet the requirements of a strategic multinational road corridor, and this negatively affected the service level of the corridor. The geometric design of the old road raised safety concerns for users of the road due to the extreme narrowness of certain cross sections and the lack of paved shoulders. The upgrade and rehabilitation widened the road to 2 x 3.4 m wide paved traffic lanes with 2 x 1.5 m wide paved shoulders. All works were done within the confines of the existing road width, and there was no resettlement of residents. Improved sections of the road were opened to traffic as works progressed; for instance, the Nyimba–Petauke section was opened in December 2015. Other works on the corridor involved the rehabilitation of three bridges (B5, B6 and B7) located between Nyimba and Sinda. Bridges B6 and B7 were replaced completely, while Bridge B5 required remedial works. Construction works on Phase II of the corridor were executed through four separate civil works contracts of UA 69.471 million or USD 2.5m (1 USD = 28.1800 UA).[4] The contract began in June 2013 and had a target completion date of June 2015. Physical construction on site started in late December 2013, and road construction diversion was established in January 2014. Earthworks began in April 2014, and asphalt surfacing in August 2015, more than two years from the original commencement date of the contract. Commencement was delayed for various administrative and financial reasons. The project was completed on 2 August 2017, and as of 2018, the project was in a 12-month defects-notification period.

[4] UA is the official currency for the African Development Bank projects.

3 Improved Roads and Firm Performance: Evidence from Zambia

3.1 Introduction

Roads are a prerequisite for economic development (Ivanova and Masarova, 2013). Good quality roads decrease transportation costs, improve access to markets, foster economic integration, stimulate competition and generate other 'wider' economic benefits (Gibbons et al, 2019).[1] At the same time, road development projects are capital-intensive and are mainly financed using public resources. For development planning purposes, it is important to understand the full extent of the effects of reduced transport costs arising from road infrastructure development. However, the literature is not clear about the size of effects and the channels through which transport investments impact economic activity (Holls, 2016). In Africa, for instance, little is known about this relationship or the extent of effects, and evidence of direct causal effects are rare.

This essay attempts to bridge this gap in the literature by investigating how improved roads impact economic activity through firms in a country in Southern Africa. Two research questions are addressed in this essay. The first is, what are the effects on firm profits, costs and size of reduced transport costs arising from improved highways? The second question is, do reduced transport costs to trade seaports affect firm performance? In analysing the above, I study the effect of improving 360 km of the Nacala Road Corridor on firms in Zambia, a country in Southern Africa with a poor road network.[2] I use a new disaggregated district-referenced firm-level panel dataset covering the period 2013 to 2017. I estimate the effects of the improved road using the difference-in-differences (DID) methodology. In the first empirical specification, I use a dummy treatment variable that equals one if a firm is located within 400 km of the road corridor and zero otherwise.[3] In the second empirical estimation, I explore the non-linear

[1] Roads attract firms into the surrounding vicinity and increase economic activity (Holl, 2016). Reduced transport costs contribute to the creation of a conducive business environment that favours firm development and growth. See Adenikinju (2005); Chissokho and Seck (2013) on the literature of how the growth potential of SMEs and even large-scale businesses are impaired by a lack of functional infrastructure.

[2] According to a roads assessment undertaken by the World Bank in Zambia in 1991, 80 percent of the roads were in poor condition (World Bank, 2005).

[3] Inspection of the dataset displays that firms are sparsely distributed in Zambia and a cut off of 400 km is used in order to have a sufficient number of districts and firms in the treatment group. See appendix figure 3.A.1 in for more information. However, estimations are also run on the 200 km bandwidth for robustness checks.

effects of the improved road on firms located in incremental driving-distance bands of 200 km. The third empirical estimation captures changes in firm performance induced by variation in accessibility to Port Nacala as a result of the road improvements. I group firms into ranked bins based on the shortest driving distance in kilometres from a district's central business area to Port Nacala in comparison to the distance to seven other trade seaports.

I deal with the potential endogeneity problem of roads being established or improved in districts with trends of high or low economic activity by treating the improved road as a natural experiment in the intermediate districts through which it passes. The aim of improving the Nacala Road Corridor was to connect the capital city of Malawi (Lilongwe) and the capital city of Zambia (Lusaka) to Port Nacala in Mozambique at the shortest driving distance for increased sub-regional trade, competitiveness and economic growth. I treat the placement and improvement of the road in areas outside Lusaka as exogenous because this geographical space received treatment only because of its location along the shortest driving path between the two capitals and not because of its baseline economic outcomes.[4] Thus, I exclude firms located in Lusaka in one sample. The firm-level panel dataset allows me to control for time-invariant unobserved heterogeneity across firms and time-varying unobserved confounders that affect all groups. Hence, I include firm and time fixed effects in the DID estimations to capture causal effects on firms of the improved road.[5] I include these controls and run regressions on a second sample that includes Lusaka.

Results indicate that after the road is improved, compared with those far away, firms in the peripheral region within 400 km of the road experience positive growth in fuel costs suggesting increased road utilisation. The same firms also experience negative growth in total assets, indicating increased competition.[6] However, the effect of the improved road on most firm outcomes (output and labour costs) are inconclusive given the available data. In addition, results on the effects on firm outcomes of changes in accessibility to Port Nacala are inconclusive.

This essay is motivated by debates in the literature on the economic impacts of roads. Some economists view investments in road infrastructure as akin to "bridges to nowhere" – spending yielding few economic benefits with large cost overruns. Others view such investments as an effective way to boost economic activity both in the short and long run (Leduc and Wilson, 2013). There is currently limited empirical evidence, especially from developing countries, to support conceptual frameworks that treat road infrastructure investments as boosting economic activity and reducing poverty. This essay focuses on Africa because, according to the World Bank, (2018), half of the global poor reside on this continent. One of the reasons African

[4] This is similar to the work of Chandra and Thompson (2000), Michaels (2008) and Datta (2012) who exploit the quasi-experimental nature of roads by demonstrating that roads are exogenous to the non-metropolitan counties they run through provided that, there was no manipulation to include or exclude some intermediate areas on factors correlated with the are an important tool tooutcomes of interest during road development.

[5] I also control for observed time varying factors that affect the outcome variable by including various covariates in the estimation.

[6] The negative growth in total assets is attributed to a decrease in current assets. Increased competition decreases sales, cash and book balances, and account receivables of firms. The assumption in this paper is that the analysis is in the short-run, a period when firm fixed assets do not vary.

countries (including some that are resource-rich) have a lower economic growth rate than those in the Western and newly industrialised countries is their poor quality infrastructure (Vivien and Briceño-Garmendia, 2010). Based on this fact, many governments and donor organisations worldwide have been channelling resources to infrastructure development. In 2007, the World Bank allocated 20 per cent of its total allocation to transport infrastructure projects, a larger share than that to education, health and social services combined (World Bank, 2007). In Africa, spending on roads is the single-largest infrastructure item ranging from 50 to 60 per cent of general government spending in middle-income countries and low-income countries, respectively (Vivien and Briceño-Garmendia, 2010). Given the limited resources faced by governments and the capital-intensive nature of road projects, excessive investment that yield little economic benefits can be detrimental to growth in the long run, and thus the need for further research on the topic.

This essay makes three contributions to the literature. It is the first to bring to the analysis a district-referenced panel dataset of firms across multiple (64) sectors in an African country where retail trade is the dominant sector in the dataset.[7] Further, the data used in this essay is new and not available to the public; authorisation to access the data from the Zambia Revenue Authority was obtained from the Ministry of Finance in 2017. In addition, most studies on developing countries focus on single sectors such as manufacturing (Ghani et al., 2016; in India), agriculture (Jedwab and Moradi, 2011; in Africa) and exporting firms (Martincus et al., 2017; in Peru). There are a few studies analysing effects across multiple sectors (Datta, 2012; using the World Bank Enterprise Survey for India).[8] However, none use a dataset dominated by retail firms. Second, I use a modern econometric methodology to provide plausible causal estimates of the impact of improved roads on firms in Africa. I use a DID estimation, a robust methodology not previously applied to similar research in the region. I use a quasi-experimental design to analyse the effects on various outcomes of firms across multiple sectors in Africa of improving a road. Like other studies on developing economies, studies on Africa focus on a single sector and mainly analyse the effects of railway systems on firm exports, agricultural activity or the effect of roads on population growth and the manufacturing sector.[9] A study in the region that is close to this one is that of Casaburi et al. (2013) who analyse the effect of improved rural roads on crop prices in rural markets in Sierra-lone. Firm-level empirical studies are rare in Africa because of data availability challenges. Lastly, this essay provides empirical evidence of the economic impacts on firms of an improved highway in a continent with a different economic geography than that in existing studies. I argue that initial conditions in Africa are distinct, and thus results may differ from those reported in the literature. Africa is sparsely populated. It has a low population density of thirty-six people per $100km^2$ of arable land, a low urbanisation rate of 35 per cent, rapid rates of urban growth of 3.6 per cent a year, a relatively large number of landlocked countries (15), and numerous small economies with

[7] Retail trade constitutes 41 percent of the firms in the dataset.

[8] In developed countries, Gibbons et al. (2019) investigates economic effects of new road additions using the universe of businesses in Britain.

[9] See Sequeira (2011), Jedwab and Moradi (2011), Jedwab and Storeyguard (2017) and Shiferaw et al (2012).

populations less than 20 million and gross domestic product (GDP) of USD 10 billion or less (Vivien and Briceño-Garmendia, 2010). Road conditions are worse than in other continents (Yepes, Pierce and Foster 2008). As compared to elsewhere, Africa faces a high backlog of road maintenance of about 40 per cent (Vivien and Briceño-Garmendia, 2010). This is despite the continent having started with road stocks not very different from those in South or East Asia in the 1960s (Vivien and Briceño-Garmendia, 2010). Infrastructure services are reported to be twice as expensive in Africa than on other continents (Schmidt et al., 2010). Infrastructure also constrains business in Africa and depresses firm productivity by 40 per cent (Escribano et al., 2008).

Findings in this essay are of interest to researchers and policy makers – investment in roads is a common tool to boost economic growth. Yet, empirical research on the effects of improved roads in developing countries, especially in Africa, remains scarce. The remainder of this essay is structured as follows. Section 3.2 details the related literature. Section 3.3 discusses the data and variables. Section 3.4 presents the theoretical framework. Section 3.5 sets out the methodology. Section 3.6 presents the results. Finally, Section 3.7 concludes.

3.2 Related Literature

Research reiterating the importance of investing in infrastructure for faster economic growth and development dates back to Rostow (1959) who emphasized investments in transportation as one of the key sectors for investment in an economy under the preconditions for take-off stage. A later seminal paper by Aschauer (1989a) influenced the literature by showing that the contribution of basic services to economic performance was large. Aschauer (1989a, 1989b), used a production function approach and related output to employment, private capital and public capital investment in infrastructure. He obtained output elasticities with respect to public capital investment of 0.34 and 0.39 in the US economy. Aschauer's work led to an explosion of research on the effects of public infrastructure investment on economic performance. Despite applying the same methodology to international, regional and sector-specific data, subsequent analyses failed to replicate similar large effects and at times failed to find significant positive effects (Pereira and Andraz, 2013). There are mixed results in the older studies as a result of many factors, including their use of different types of public capital. The World Bank reports that earlier studies fell short on econometric grounds, and this made the results highly questionable; they were not only implausibly large but also lacked robustness in the use of appropriate econometric techniques (Fedderke and Bogetic, 2006). For instance, Aschauer (1989a, 1989b), Munnell (1990) and Finn (1993) used single equation static production functions while Shah (1992), Dalenburg (1987) and Morrison (1988) used dual cost functions and derived input demand equations. These methodologies failed to control for simultaneity, the effect of confounding factors, spurious correlation of time series data and model misspecification, which left the matter of causality unresolved (Pereira and Andraz, 2013).

Recent studies use robust econometric methodologies to attempt to provide plausible causal estimates of the impact of infrastructure on various economic outcomes.[10] Scholars have attempted to address the problem of endogeneity associated with road investment projects and the issue of estimating infrastructure stock from public capital investment. Studies resort to using binary variables of the regressor of interest or other plausible instruments in addition to advanced econometric techniques of identification, specification and estimation to infer the causal effect of roads. Common econometric methodologies used include DID estimation, panel fixed effects, long-differenced estimation and instrumental variable approach.[11] Sudies using difference-in-differences estimation include Martincus et al. (2017) and Datta (2012). Martincus et al. (2017) investigate the impact of new domestic road infrastructure on firm exports in Peru and use a binary indicator as the regressor of interest; this takes the value one if the firm is located in a region where there are shorter distance due to a new road and zero otherwise. The authors further apply an instrumental variable approach using the historic pre-Columbian Inca-road network in addition to the DID methodology as an alternative means of addressing endogeneity concerns. Datta (2012) analyses India's Golden Quadrilateral Program and a major road's impact on firm inventory costs and supplier relationships using two treatment variables. Their first treatment variable is a binary variable that takes the value of one if the firm was located on an improved road and zero otherwise. The second treatment variable is the distance of a firm from the improved road. Ghani et al (2016), use long-differenced estimation and straight-line instrumental variables estimation to study the impact of transportation infrastructure on the organisation and efficiency of manufacturing activity in India. Gibbons et al. (2017)) use firm panel fixed effects to investigate the impact of new road infrastructure on employment and labour productivity using plant-level data in the United Kingdom.

A number of identification strategies are employed in the literature to identify the causal effects of roads. Studies like Chandra and Thompson (2000), Michaels (2008) and Datta (2012) treat the construction of roads as a natural experiment by considering as randomly selected into the treatment the intermediate areas falling along the roads connecting two major cities. Other studies use instrumental variables to address the endogeneity problem of road investment projects. These include Sequeira (2011), Faber (2014), Ghani et al. (2016), Wang (2016), Martincus et al. (2017) and Jedwab and Storeygard (2017).

Investing in road infrastructure has direct and indirect effects in an economy. Direct effects include increased accessibility to new markets and technology, lower transportation costs, easy access to social services, and job creation. The indirect effects involve 'wider economic benefits' of total factor productivity effects arising from agglomeration economies (Graham, 2007; and Sequeira, 2011).[12] Based on the hypothesised benefits to economies of investing in infrastructure, investments in transport infrastructure rank highly on the development agendas of

[10] Donaldson (2018), Duranton et al (2014), Faber (2014) and Jedwab and Storeygard (2017) to mention a few.

[11] Long differenced estimation is the difference-in-differences methodology with post period observations averaged over some number of years. See Ghani et al, (2016) for a detailed explanation on long-differenced estimation.

[12] The nature and availability of infrastructure influences the marginal productivity of private capital and therefore, public investment in infrastructure crowds-in private investment (Kessides, 1993).

many countries and garner huge public-resource allocations. In Peru, for instance, to finance the expansion of the road network, public resources allocated to road infrastructure increased by 610 per cent (in nominal US dollars) to reach 1.3 per cent of the country's GDP in 2010 from its initial level of 0.5 per cent in 2003 (Martincus et al., 2017). In China, the National Trunk Road Development Program was constructed at an estimated cost of USD 120 billion over a 15-year period (Faber, 2014). In India, Phase 1 of the National Roads Development Project was constructed at an estimated cost of USD 7 billion (Ghani et al., 2016). The above examples highlight the emphasis placed by policy makers on investment in infrastructure for faster economic growth and development. In fact, some scholars and policy makers link the failure of some countries to achieve adequate growth to a lack of adequate infrastructure. For instance, the slowdown in productivity growth in many OECD countries during the 1970s and 1980s is often attributed to infrastructure deterioration due to fiscal consolidation policies or benign neglect (Pereira and Andraz, 2013).

Empirical studies, as mentioned earlier, show mixed results regarding the magnitude and direction of the net effect on economies of infrastructure development. This also emerges in recent studies. Despite using modern econometric methodologies, recent empirical work continues to report that infrastructure has positive effects on economic growth. However, recent studies also provide countervailing evidence of either ambiguous, insignificant or negative impacts of public capital on development prospects (Fedderke and Bogetic, 2006). Studies reporting positive impacts include Datta (2012), which finds that in India, improved roads caused firms to reduce their input inventory holding days and to change suppliers to optimise operations. Martincus et al. (2017) and Sequeira (2011) investigate trade activities in Peru, and Mozambique and South Africa, respectively, and find that transport infrastructure positively impacts firm exports. Gibbons et al. (2019) find that new investment in infrastructure produces economic benefits even in mature economies like the UK. Studies with mixed results include Bonaglia et al. (2000) and Pedroni and Canning (2004). Bonaglia and et al. (2000) .investigate the impact of infrastructure on the economic performance of Italy as a whole and of its macro-regions. Canning and Pedroni (2004) analyse the long-run impact of infrastructure provision on per capita income in various countries. According to Fedderke and Bogetic (2006) of the World Bank, mixed empirical results in the recent literature may be the result of various factors:

1. There is a potential endogeneity problem of road placement; roads may be allocated to certain areas and not others due to an interplay of political, economic, social and environmental factors (Robinson and Torvik, 2005). Therefore, areas that receive roads may be completely different to areas that do not. Failure to address this concern results in biased and inconsistent estimates.
2. The type of dependent variables used in the model affects the results. The effect of infrastructure on labour, output and total factor productivity differ in size and direction.
3. 3. Aggregate measures of infrastructure hide the productivity impact of infrastructure at a disaggregated level. Shioji (2001)), studying the US and Japan, obtains positive results only when public capital is disaggregated. Conversely, some economists argue that studies

> at a small geographic scale fail to capture network and spill-over effects that aggregate studies capture. Pereira and Andraz (2004) argue that only 20 per cent of the aggregate effects of public investment in roads in the US are captured as direct effects in the state itself; the remaining 80 per cent correspond to spill-over effects from public investments in roads in other states. Hence, studies at different geographic scales are important as they provide useful insights into the economic impacts of roads.

Despite the mixed empirical results, there are a few stylised facts evident from the literature. First, there is little consensus on the magnitude and direction of the effect of infrastructure on economic performance. Second, infrastructure effects are higher in less developed than in developed economies due to a high level of infrastructure inadequacy in developing economies. Lastly, microeconomic channels of households and firms, through which infrastructure influences economic activity, remain largely unexplored, especially in developing countries (Sequeira, 2011).

3.3 Data

3.3.1 Firm Data

The firm-level district-referenced data used in this essay comes from the Zambia Revenue Authority (ZRA) online tax system. The system captures yearly balance sheet information of firms that have turnover of or over approximately USD 67,300.00 (ZMW800, 000.00). As required by the Income Tax Act of 2012, firms submit income tax returns and financial statements electronically to the ZRA on 21 June each year for the preceding year's transactions. The ZRA online tax system was launched in 2013. [13] This essay uses the data for the year 2013 (the period before the highway upgrade) and the year 2017 (the period in which the highway improvement project was completed). In this dataset, firm data is available for only one pre-treatment period and one post-treatment period. The full sample database has 8,757 observations in fifty-seven districts.[14]

Table 3.1 shows that there are 1,952 firms in 2013 and 1,560 in 2017 in the sample including Lusaka, and 761 firms in 2013 and 561 in 2017 in the sample that excludes Lusaka. Firms are identified by a unique taxpayer number, which is important for identifying the location of a firm with branches in several districts. Further, the district information allows for the computation of the shortest driving distance of a district in which a firm is located to the nearest point on the improved road, using google maps. The period covered in the firm dataset allows for the study of the impact on firm outcomes of a road improvement project on which construction began in January 2014 and finished in August 2017. I take 2013 as the pre-treatment period

[13] The year 2013 is the earliest data year in the database.

[14] When observations with missing data are dropped, 3512 observations remain in the full sample. The number of districts in my dataset is nearly half the total number of districts in Zambia reported to be 117 as at 2018. As shown in figure 3.1, firm information is not available for all the districts.

and 2017 as the post-treatment period. I do not compute labour productivity factors, which are important indicators of firm performance, because the data does not contain information on the number of workers employed. The data is converted from nominal to real terms using the World Bank national GDP deflator (base year 2010) to ensure the comparison of firm data over time. The national GDP deflator is used in the absence of data on district-level GDP deflators for Zambia.

Table 3.1 shows that the number of firms in both samples is higher in 2013 than in 2017.[15] This is attributed to the fact that some firms may not have filed their tax returns for the post-construction period while others merged or merely shut down operations gradually as firms only de-register if operations discontinue and not for the purposes of downgrading to lower tax bands.

Table 3.1: Summary Statistics

						Change (2017-13)	
Sample	Ave. Driving Km to improved road	No. of firms 2013	Ave. Population 2013	No. of firms 2017	Ave. Population 2017	No. of firms	Ave. Population
Sample with Lusaka	339	1,952	1,459,284	1,560	1,717,177	-392	257,893***
(S.E)	(1.716)		(12013.86)		(14742.44)		
Sample without Lusaka	521	761	420,963	561	472,087	-200	51,124 ***
(S.E)	(2.439)		(6699.70)		(9314.93)		

* denotes significance at 90%, ** at 95% and *** at 99%

Source: The author

The distribution of firms by industry at baseline in the two samples is shown in Table 3.2. The dataset has sixty-four sectors based on the two-digit division of the International Standards of Industrial Classification (ISIC) of all economic activities of the United Nations, Revision 4. The baseline results show that in both samples, the highest number of firms are engaged in retail and agricultural activities. This is a typical economic structure in African countries where the manufacturing sector is still underdeveloped and economic activity is driven by numerous small and medium enterprises engaged in retail and rain-fed agricultural businesses that are easily established from a resource perspective.

[15] See Appendix table 3.B.1 for more details on the summary statistics by sample and treatment.

Table 3.2: Baseline Distribution of Firms by Industry

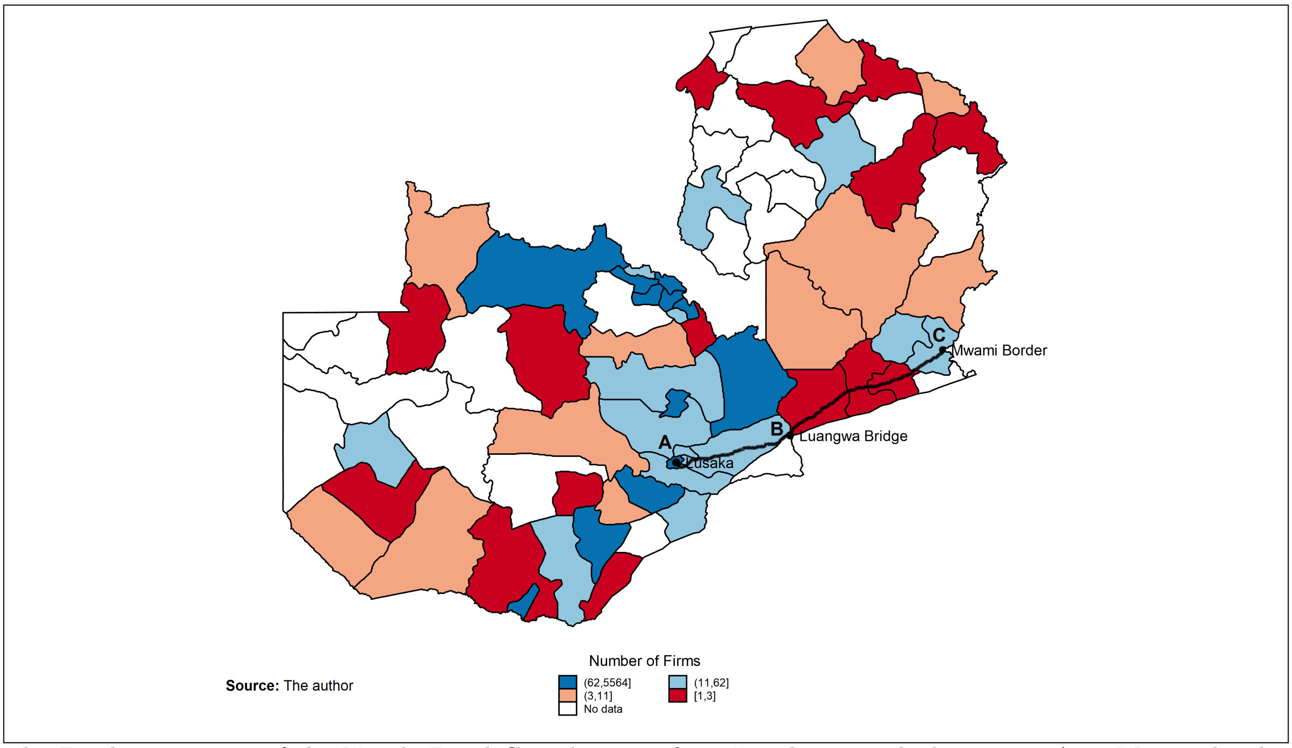
		Sample Excludes Lusaka		Full Sample	
Industrial Class	2 Digit ISIC	No. of firms	Percent Share	No. of firms	Percent Share
Retail trade, except of motor vehicles and motorcycles	47	315	41.39	783	40.11
Crop and animal production, hunting and related service activities	1	43	5.65	88	4.51
Other professional, scientific and technical activities	74	42	5.52	70	3.59
Land transport and transport via pipelines	49	31	4.07	56	2.87
Wholesale trade, except of motor vehicles and motorcycles	46	27	3.55	2	0.1
Manufacture of food products	10	23	3.02	41	2.1
Accommodation	55	19	2.5	49	2.51
Other mining and quarrying	8	19	2.5	52	2.66
Other manufacturing	32	18	2.37	35	1.79
Warehousing and support activities for transportation	52	17	2.23	41	2.1
Wholesale and retail trade and repair of motor vehicles and motorcycles	45	16	2.1	43	2.2
Architectural and engineering activities; technical testing and analysis	71	15	1.97	36	1.84
Office administrative, office support and other business support activities	82	14	1.84	83	4.25
Construction of buildings	41	13	1.71	63	3.23
Fishing and aquaculture	3	12	1.58	29	1.49
Education	85	10	1.31	32	1.64
Mining of metal ores	7	10	1.31	16	0.82
Food and beverage service activities	56	7	0.92	22	1.13
Human health activities	86	7	0.92	22	1.13
Manufacture of fabricated metal products, except machinery and equipment	25	7	0.92	11	0.56
Specialized construction activities	43	7	0.92	11	0.56
Activities of membership organizations	94	6	0.79	16	0.82
Manufacture of beverages	11	6	0.79	13	0.67
Other personal service activities	96	6	0.79	19	0.97
Rental and leasing activities	77	6	0.79	10	0.51
Security and investigation activities	80	6	0.79	10	0.51
Others		59	7.72	299	15.29
Total		761	100.0	1952	100.0

3.3.2 Road Data

Information on the highway project comes from the Road Development Agency (RDA) in Zambia. The Public Roads Act of 2002 establishes the RDA as responsible for the care, maintenance and construction of public roads in Zambia. Source documents used for information on the Nacala Road Corridor Project include pre- and post-implementation project reports. The reports set out project work sections, financing partners, types of works done, project phases and duration, challenges faced and resources employed.

The section of the Nacala Road Corridor upgraded in Zambia is a 360 km length of road running from Luangwa bridge to the Mwami border post. Figure 2.1 is a map showing the Nacala Road Corridor in Zambia and the distribution of firms by district.[16]

Figure 3.1: Map of Nacala Road Corridor in Zambia



The Zambian portion of the Nacala Road Corridor runs from Lusaka city which is point A to Mwami border at Point C. The upgraded section is from Luangwa bridge Point B to Mwami border point C.

The road analysed in this essay is an international highway and it could be argued that the outcomes analysed should be limited to those of firms engaged in international trade. However, the improvement of an international highway may also have broad spatial effects on economic activity. Therefore, it is also important to know the effect of the improved highway on all firms (whether engaged in international trade or not) located in different distance bands from the improved highway.

[16]Appendix figure 3.A.2 is a map showing the Nacala road corridor and districts in the two treatment groups. This paper studies only phase 2 of the upgrade of the Nacala road corridor. Phase 1 and 3 road improvements are done in Mozambique and Malawi. The data for Mozambique and Malawi were not available at the time of data collection to allow a within and cross country analysis of effects of upgrading the road corridor on firms in these three countries.

3.3.3 Population Data

In some regressions, I interact the population of each district with the post variable to control for differences in trends of districts of different sizes. The population data comes from the Central Statistical Office (CSO) database of Zambia. The census in Zambia is conducted every ten years, with the most recent in 2010. The 2013 and 2017 population data used in the analysis are projection figures collated by CSO with the technical input of the US Census Bureau for the purposes of national planning needs assessments. The projections are calculated at an average growth rate of 2.7 per cent per annum (CSO, 2013). The limitation of using population data is that the figures are not exact and are subject to measurement errors (CSO,2013). Further, the projection data excludes the demographic impact of the HIV/AIDS epidemic and takes immigration and net international migration for Zambia to be zero or negligible from 2010 and throughout the projection period up to 2035 (CSO,2013).

As shown in Table 3.1, in the dataset, the average population in districts increases over time in the range of 12.14 per cent to 17.67 per cent in the sample excluding Lusaka and the full sample, respectively.

3.3.4 Dependent Variables

As allowed by the available data, the dependent variables in this study are annual gross profits, sales, wage costs, total employment costs, cost of raw materials inventory, transport costs, fuel lubricant costs and total assets. The dependent variables in this study are defined as follows. Sales are the gross receipts of the firm from output sold. Gross profits is the difference between sales and the cost of goods sold. Direct wage costs refer to the costs of labour directly engaged in the production process. Total employment expenses constitute direct wage costs, salaries of labour not directly engaged in production, including all company employee costs of subscriptions, pension contributions, expatriate staff costs and any other employment-related costs. The cost of raw materials inventory is the cost of closing inventory, which comprises raw materials and work in progress. Transport costs include freight expense by land, air or sea and costs pertaining to handling, packaging and warehousing goods, insurance costs, and seaport and border-clearance costs. Fuel and lubricant costs are expenses for fuel and other oils utilised in the transit vehicles. Total assets include total current assets and fixed assets. Current assets constitute cash and bank balances, inventory and accounts receivable, while fixed assets include land, buildings and machinery.[17]

Conceptually, bad roads depress firm activity and thus, removing the constraint (dilapidated road infrastructure) should result in better outcomes. This is especially the case for firms in districts closer to the improved road post-treatment as they will experience a greater reduction in transport costs. Table 3.3 shows the average change in the outcome variables over time

[17] For more information on variables in the dataset see https://www.zra.org.zm/documentUpload.htm?actionCode=eForms and open Income tax annual company ITF-2-A eform (ITAC-2).

for firms located within 400 km (treatment group) and outside 400 km (control group) of the improved road in both samples. [18] The dependent variables are transformed into inverse hyperbolic sine (ihs) to normalise the skewed distribution of the data. Another advantage of using inverse hyperbolic sine transformation is that zeros and negative numbers are retained in the dataset, and coefficients can be interpreted as semi-elastic. The full sample has 8,757 observations, but I drop observations with missing data, leaving 3,512 observations in the full sample and 1,322 observations in the sample that excludes Lusaka. Of the missing values in both samples, approximately 50 per cent are for costs of raw materials inventory and 40 per cent for transport costs.

Hypothetically, firms in close proximity to the improved road should experience greater changes in outcomes than firms located far away as improved roads cause a fall in the price of transportation services that is inversely proportional to distance. The results set out in Table 3.3 are, for the most part, not statistically different from zero based on the z-test. This suggests firms in the treated and control groups do not experience significant changes in the majority of the dependent variables over time. Further, the direction of change in the mean values of the dependent variables in the two samples and treatment groups is mixed.[19] It should be noted that the results in Table 3.3 are simply unconditional changes that describe the behaviour of variables in the data set over time.

[18] I calculate the within-group change in means for each outcome variable over time using the z test. The z test examines the significance of the within-group differences or conducts a matched pair test of equality of means before and after the intervention in each group. The control group is identical in both samples.

[19] Out of the eight dependent variables, the average change in four variables (gross profit, sales, total employment costs and assets) are consistently negative in both samples and treatment groups.

Table 3.3: Dependent Variables

	Treatment Group		Control Group
	Mean change (2017-13)		Mean change (2017-13)
	Full Sample	Excludes Lusaka	Both Samples
ihs_Gross Profit	-0.139	-0.280	-0.531
	(0.213)	(0.743)	(0.294)
ihs_Sales	-0.017	-0.291	-0.259
	(0.187)	(0.602)	(0.252)
ihs_Direct Wages	-0.112	0.811	-0.153
	(0.130)	(0.637)	(0.234)
ihs_Freight & Transport	0.003	-0.062	0.030
	(0.117)	(0.555)	(0.165)
his_Fuel_Lubricant	-0.202	0.618	-0.377*
	(0.148)	(0.590)	(0.211)
ihs_Tot closing Raw maerials	-0.135	0.105	-0.005
	(0.116)	(0.536)	(0.163)
ihs _Tot Employment Expense	-0.027	-0.926*	-0.029
	(0.163)	(0.558)	(0.220)
ihs_Total Assets	-0.431*	-0.555	-0.736**
	(0.241)	(0.673)	(0.297)

* Denotes significance at 90%, ** at 95% and *** at 99%. Standard errors are in parentheses.
Results in this table are calculated using the z test. The z test examines the significance of the within group differences or conducts the matched pair test of equality of means before and after the intervention in each group.

3.4 Theoretical Framework

Roads affect firm-level productivity by yielding savings in transport costs which consequently result in reduced input and output costs (Holl, 2016). Transportation services are an intermediate good; that is, it is both an input and output in the production process. As an input to the production process, firms in various sectors of the economy combine transportation services with other factor inputs to produce some output. As an output in the production process, transportation services are the core business of a firm and are produced from a combination of inputs. I build a simple theoretical model to provide a clear understanding of how improved roads affect economic activity through firms. The model takes into consideration sectoral heterogeneity in production technologies and shows how changes in the price of transportation services affect firm output, demand for labour and other inputs in production. In this framework, how firms substitute factor inputs in production in response to a change in the price of transportation services depends on the value of the elasticity of substitution.

Theoretical studies close to my theoretical framework include Holl (2016) on the impact of highways on productivity in Spain and Moro (2007) on the influence on total factor productivity growth in the in the US of a change in price and the share of intermediate goods. The theoretical framework in this essay differs from these studies as follows. The above studies use the constant elasticity of substitution (CES) production function and the Cobb–Douglas production function to explain the influence of changes in factor prices and shares on total factor productivity. By contrast, I focus on explaining the influence of a change in the price of transportation services on firm output, demand for labour and raw materials in the short run.

3.4.1 Sectoral Value Added Production Function

$$Y_{tk} = A' \left[\beta L_{tk}^{\frac{\sigma-1}{\sigma}} + \theta T_{tk}^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\psi}{\sigma-1}} \tag{3.1}$$

I use a CES production function to explain the relationship of improved roads and firm performance where Y_{tk} is the value added output of firms at time t in sector k. L_{tk} represents labour input and T_{tk} other inputs in production such as transportation services and other raw materials. The assumption is that in the short-run capital is fixed and therefore, firms react to road improvements by adjusting labour and other inputs in production. Let $A^{'} = A\bar{K_{tk}}$ be the total factor productivity term that includes constant capital $\bar{K_{tk}}$. $\theta = 1 - \beta$ and β, $\theta > 0$ with $\beta + \theta = 1$. ψ is the returns to scale parameter . $\psi = 1$ is constant returns to scale, $\psi > 1$ is increasing returns to scale and $\psi < 1$ is decreasing returns to scale. σ = Elasticity of substitution of inputs and lies between $0 < \sigma < \infty$. If $\sigma \to \infty$ this indicates perfect substitutability. If $\sigma \to 0$, this indicates no substitutability and if $\sigma \to 1$ this indicates some substitutability and you have the Cobb Douglas production function. The Cobb Douglas production function is continuous, increasing and quasi-concave.[20]

3.4.1.1 Firm's Problem when $\sigma = 1$

Profit maximising firms will choose the least cost production plan for every level of output. They will choose a level of output for production and how much of which factors to use to produce it. Firms seek to solve:

$$max_{\{L,T\}} \qquad P_t A' L_{tk}^{\beta} T_{tk}^{\theta} - wL_{tk} - \gamma T_{tk} \tag{3.2}$$

From the equation above, I derive the supply function and two input demand functions for other raw materials or transportation services and labour.[21]

[20]It is continuous to ensure that small changes in the vector inputs leads to small changes in the amount of value-added output produced; strictly increasing to ensure that employing strictly more of every input results in strictly more output and is strictly quasi concave implying that inputs are complementary to an extent as very little production will take place if one of the inputs is in low quantities while the other is in high abundance. All inputs are important in production and input quantity utilisation is equally important.

[21]See appendix 3.C for detailed derivations of all the equations in this section.

$$T_{tk}^{*}\left(P_{t},w,\gamma\right)=\left(P_{t}A'\right)^{\frac{1}{1-\beta-\theta}}\left(\frac{\beta}{w}\right)^{\frac{\beta}{1-\beta-\theta}}\left(\frac{\theta}{\gamma}\right)^{\frac{1-\beta}{1-\beta-\theta}} \tag{3.3}$$

$$L_{tk}^{*}\left(P_{t,},w,\gamma\right)=\left(P_{t}A'\right)^{\frac{1}{1-\beta-\theta}}\left(\frac{\beta}{w}\right)^{\frac{1-\theta}{1-\beta-\theta}}\left(\frac{\theta}{\gamma}\right)^{\frac{\theta}{1-\beta-\theta}} \tag{3.4}$$

$$Y_{tk}^{*}\left(P_{t},w,\gamma\right)=P_{t}^{\frac{\beta+\theta}{1-\beta-\theta}}A'^{\frac{1}{1-\beta-\theta}}\left(\frac{\beta}{w}\right)^{\frac{\beta}{1-\beta-\theta}}\left(\frac{\theta}{\gamma}\right)^{\frac{\theta}{1-\beta-\theta}} \tag{3.5}$$

Proposition 1:

A fall in (γ) ***the price of transportation services has a positive effect on firm valued added output, the demand for labour, transportation services and other inputs in production.***

Equation 3.3 is the transportation services or raw materials input demand function and it indicates that a fall in γ the price of transportation services means a high $T_{tk}^{*}\left(P_{t},w,\gamma\right)$. This indicates that improved roads cause firms to demand more raw materials and to increase utilisation of the upgraded road. A fall in the price of transportation services allows firms to access wider markets and to experience efficiency gains by purchasing cheap raw materials. Access to cheap input markets allows firms to purchase more raw materials for the same cost and to produce more output for the same cost. On the other hand, accessibility to wider output markets increases the demand for output and firms respond by increasing production capacities, which triggers increased demand for raw materials and labour inputs.

Equation 3.4 is the labour input demand function and indicates the relationship between prices of transportation services and a firms' demand for labour. In this equation, a fall in γ infers a high $L_{tk}^{*}\left(P_{t},w,\gamma\right)$ which suggests that in the short-run firms react to a road improvement project by employing more labour in production. Similarly, the value added output supply function as shown by equation 3.5 displays the relationship between prices of transportation services and output. In this equation, a fall in γ implies a higher $Y_{tk}^{*}\left(P_{t},w,\gamma\right)$ suggesting that road improvement projects cause firms to increase current production capacities.

The key supposition under this subsection is that, road improvements introduce positive shocks in firm production processes by reducing the price of transportation services.

3.4.1.2 Predictions on Empirical data based on the Cobb Douglas production function:

Firms located near the improved road compared to those far away incur a higher fall in transport costs and thus are expected to experience:

1. A high growth in sales, gross profits and labour expenses after the road is improved. Firms near the improved road face high transport cost reductions and start to access previously inaccessible cheap input markets and competitively priced output markets. Road improvement projects increase firm output capacities, firms experience an increase in demand originating from new markets and since in the short run capital is fixed, firms respond by hiring more labour and other raw materials.
2. A high growth in freight transport and fuel costs after the road is improved. Firms near the road experience a larger fall in transport costs compared to firms far away after the road is improved and thus, these firms will optimise utilisation of the road transport system more.
3. Reduced growth in raw materials inventory costs. After the road is improved increased accessibility to cheaper input markets may result in firms holding a larger amount of raw materials inventory for the same cost price or hold the same quantity of raw materials inventory at a lower cost. Firms may also deplete raw materials inventory because of increased production driven by increased demand from new output markets.
4. A high growth in assets after the road is improved because of a positive growth in sales. Increased output demand causes firms to make positive output and labour adjustments that result in an expansion of total assets. In the short run, only current assets change (cash, bank balances and trade receivables).

3.4.1.3 Firm's problem when $\sigma \neq 1$

When $\sigma \neq 1$, we have the standard CES production function. To study the substitution effect of a change in price of transportation services γ on the demand for labour and other raw materials, and on value-added output quantity, I differentiate the unconditional input demand functions and supply function with respect to the price of transportation services to get the equations below.[22]

$$\frac{\partial L^*_{tk}\left(P_t, w, \gamma\right)}{\partial \gamma} = \frac{\sigma - 1 - \psi\sigma}{(\psi - 1)} \left(P_t A' \psi\right)^{-\frac{1}{\psi - 1}} \left(\frac{w}{\beta}\right)^{-\sigma} \left(\frac{\gamma}{\theta}\right)^{-\sigma} \left[\beta \left(\frac{w}{\beta}\right)^{1-\sigma} + \theta \left(\frac{\gamma}{\theta}\right)^{1-\sigma}\right]^{-\frac{\psi}{(\sigma-1)(\psi-1)}} \tag{3.6}$$

$$\frac{\partial T^*_{tk}\left(P_t, w, \gamma\right)}{\partial \gamma} = \left(P_t A' \psi\right)^{-\frac{1}{\psi - 1}} \left\{ \begin{array}{c} -\sigma \left(\frac{\gamma}{\theta}\right)^{\sigma - 1} \frac{1}{\theta} \left[\beta \left(\frac{w}{\beta}\right)^{1-\sigma} + \theta \left(\frac{\gamma}{\theta}\right)^{1-\sigma}\right]^{\frac{\sigma - 1 - \psi\sigma}{(\sigma-1)(\psi-1)}} + (1-\sigma) \frac{\sigma - 1 - \psi\sigma}{(\sigma - 1)(\psi - 1)} \left(\frac{\gamma}{\theta}\right)^{-2\sigma} \\ \left[\beta \left(\frac{w}{\beta}\right)^{1-\sigma} + \theta \left(\frac{\gamma}{\theta}\right)^{1-\sigma}\right]^{\frac{2(\sigma-1) - \psi(2\sigma - 1)}{(\sigma-1)(\psi-1)}} \end{array} \right\} \tag{3.7}$$

[22] See appendix 3.C for detailed derivations of all the equations in this subsection.

$$\frac{\partial Y_{tk}^{*}\left(P_t,w,\gamma\right)}{\partial\gamma}=-\frac{\psi\left(1-\sigma\right)}{\left(\sigma-1\right)\left(\psi-1\right)}\left(P_tA'\psi\right)^{-\frac{\psi}{\psi-1}}\left(\frac{\gamma}{\theta}\right)^{-\sigma}\left[\beta\left(\frac{w}{\beta}\right)^{1-\sigma}+\theta\left(\frac{\gamma}{\theta}\right)^{1-\sigma}\right]^{\frac{-\psi\sigma-\sigma-1}{(\sigma-1)(\psi-1)}} \tag{3.8}$$

Proposition 2:

1. ***An increase in the price of transportation services γ increases the utilisation of labour in production. However, the extent to which firms substitute labour for transportation services or raw materials depends on σ the elasticity of substitution.***
2. ***An increase in the price of transportation services decreases the utilisation of transportation services and raw materials in production and, reduces value added output.***

If $\sigma > 1$, there is easy substitutability of factor inputs, and firms respond to a change in the price of transportation costs by changing the input mix of their least least-cost bundle.

Equation 3.6 shows that firms intensify the use of labour in production when γ, the price of transportation services, increases. To minimise costs in production, firms respond by changing the input mix in their least least-cost bundles to favour the now relatively cheaper labour input. This reduces the transportation services to labour ratio, the marginal product of labour and the marginal rate of technical substitution of transportation services by labour. In equilibrium, firms will intensify labour in production to a point where the marginal product of labour equals its wage rate w. Equation 3.7 shows that firms respond to an increase in the price of transportation services γ by reducing the quantities of transportation services and raw materials $T_{tk}^{*}\left(P_t,w,\gamma\right)$ used in production. As raw materials and transportation services are relatively more expensive than labour, they are easily substituted with labour. Equation 3.8 shows that an increase in the price of transportation services reduces value-added output. High transport costs result in firms failing to access wider output and input markets, which in turn limits production output.

If $\sigma < 1$, there is no substitutability between factor inputs and firms are not able to change the input mix of their least-cost bundle to minimise costs. Firms require fixed proportions of capital, labour, transportation services and other raw materials to produce the optimal output quantity. Therefore, an increase in the price of transportation services produces an increase in the cost per unit of output and a loss for firms producing the same output volume. By contrast, a fall in the price of transportation services will decrease the cost per unit of output in production and increase firm profits, given the unchanged optimal output volume. Firms are generally only able to adjust production capacities downwards or upwards in the long run where all inputs of production are variable.

3.4.1.4 Predictions on Empirical data based on the CES production function.

1. If the CES of inputs is greater than 1 ($\sigma > 1$) , a fall in γ will cause firms near the improved road compared to those far away to experience:
   a) a) Increased growth in gross profits and sales. Gross profits increases as firms experience reduced costs per unit of output as they substitute labour with transportation services. There is increased use of transportation services and other raw material inputs in production since they become relatively cheaper than labour. Sales increase due to increased demand for output arising from new markets that were previously inaccessible.
   b) Increased growth in firm assets due to increased sales resulting from demand in new markets.
   c) Increased growth in labour costs. Demand from new markets causes firms to increase production quantities and demand more labour and raw material inputs.
2. If the elasticity of substitution inputs $\sigma < 1$, a fall in γ will cause firms near the improved road compared to those far away to experience:
   a) Increased growth in gross profits due to reduced cost per unit of output.
   b) No significant growth in sales because fixed input quantities are used to produce a fixed output level. This is because, in the short run, some inputs like capital are fixed, so firms continue to use the same fixed input quantities and produce the same output quantity
   c) No significant growth in labour costs and raw materials inventory costs due to a fixed proportion of inputs being used to produce a fixed output.
   d) Increased growth in current assets due to a growth in gross profits.

3.5 Methodology

The objective of this essay is to estimate the causal effect of an improved road on firm output and the costs of transport, labour and raw materials inventory. I analyse how firms located within 400 km of the road behave in comparison to firms in districts outside that radius to capture the effect of reduced transport costs that arise after the highway is improved on firms at different proximity to the improved road. I measure the distance to the improved road from the central business area of each district using Google Maps.[23] The average driving distance from the improved road is 521 km for firms in the sample that excludes Lusaka and 339 km for firms in the full sample (including Lusaka). There are only sixteen firms within 180 km of the

[23]The central business area in Zambia is the area where most business offices are located. District council offices and the post office are located in this area and thus, I take these two locations as my reference starting points in a district.

improved road, fourteen in 2013 and two in 2017 and singletons are dropped in the estimation using firm fixed effects. The geographical spread of firms in the dataset indicates that there are few firms in distance bandwidths within 180 km. To capture more firms, I investigate the effects over time of the improved highway on firms within 400 km of the improved road.[24] In another specification, I explore how firms clustered in 200 km incremental driving-distance bands from the improved road behave compared to those located in the farthest band – more than 600 km from the improved road.

Further, I investigate how reduced transport costs to Port Nacala induced by the improved highway affect firm performance. In this analysis, I calculate the driving distance from each district to each of the eight seaports that Zambia can use for international trade that are accessible through its neighbouring countries. Thereafter, I create a rank variable that indicates, for each district, how near the Port of Nacala is compared to the seven other trading seaports.[25] The data shows that Port Beira is the nearest trading seaport to most of the districts in Zambia. The ranking of Port Nacala for each district, therefore, starts from Rank 2 and ends at Rank 8[26] Rank 2 indicates that Port Nacala is the trading port second nearest to a district and Rank 8 indicates it is the farthest trading port in comparison to the other seven.[27] Using the rank variable, I measure the impact on firms of a change in accessibility of Port Nacala produced by the improved road.[28] Changes in outcomes of firms in districts in Ranks 2 and 3 and those in Ranks 4 and 5 are compared to changes in outcomes of firms in districts with Ranks 6 to 8. Appendix Table 3.B.2 indicates the rank of Port Nacala and the shortest driving km distance for each district to the eight seaports used by Zambian firms to access international markets. Appendix Figure 3.A.3 shows the map of Zambia and the eight seaports.

In all the above estimations, a DID methodology is applied to a panel database of firms in forty-seven districts observed over five years.[29] In all estimations, I control for observed and unobserved factors by including other covariates, time and firm fixed effects. With firm fixed effects, highway effects are identified using variation within firms over time after accounting for factors that affect all groups. Firm fixed effects also address the issue of firms self-selecting into districts located near the road over time.

[24] Appendix figure 3.A.1 shows the spatial distribution of firms.

[25] Zambia has eight seaports it can use to access international markets through its neighbouring countries namely: Beira, Maputo and Nacala in Mozambique; Dar-es-salam in Tanzania; Durban, East London and Port Elizabeth in South Africa and Walvis Bay in Namibia. Zambia is landlocked.

[26] Port Beira in Mozambique is the nearest port to most of the districts in the data, however, due to port handling capacity challenges and route accessibility challenges firms may not always use the nearest port to access international markets.

[27] The rank dummy variable is used to capture a change in accessibility to Port Nacala because it groups together firms in districts with the same rank of Port Nacala as their nearest trading port in comparison to other sea ports. Grouping districts based on the shortest driving distance to port Nacala may result in estimating effects of a change in accessibility to Port Nacala on districts identified as close to Port Nacala, but whose firms do not use the Port Nacala for business because other seaports are much closer based on the rank variable.

[28] Due to few observations in some ranks, I pull the rank groups together.

[29] The distribution of districts in the sample after dropping missing values is 47. 12 districts are within 400 km; 35 districts are more than 400 km; 4 districts are within 200 km, 8 districts are between 201 to 400 km; 17 districts are between 401 to 600 km and the reference group has 18 districts laying between 601 to 1041 km.

3.5.1 Empirical Estimation Model 1

$$Y_{ijt} = \alpha_0 + \beta_1 Post_t + \beta_2 d400_j + \beta_3 d400_j * Post_t + \gamma X_{ijt} + \tau_i + \epsilon_{ijt} \tag{3.9}$$

In the model above, Y_{ijt} is the outcome variable of the ith firm in district j at time t. $Post_t$ is the year fixed effects.[30] $d400_j$ is a binary variable taking the value one if j indexes a district within 400 km of the improved road and zero otherwise.[31] X_{ijt} indicates other observable factors affecting the outcome variable. τ_i are the firm fixed effects and ϵ_{ijt} is the idiosyncratic error term. In this model, the parameter of interest is β_3 which measures the average change in outcomes of firms located within 400 km of the improved road relative to the average change in outcomes after the road is improved of firms outside this radius.

3.5.2 Empirical Estimation Model 2

$$\begin{aligned} Y_{ijt} = \alpha_0 + \beta_1 Post_t + \theta_1 Dist0\_200km_j * Post_t + \theta_2 Dist201\_400km_j * Post_t + \\ \theta_3 Dist401\_600km_j * Post_t + \gamma X_{ijt} + \tau_i + \epsilon_{ijt} \end{aligned} \tag{3.10}$$

In Model 2, all other terms remain the same as in Model 1 except for the treatment variables. Here, $Dist0\_200km_j$ is a binary variable taking the value of one if j indexes a district laying within 200 km of the improved road and zero otherwise. $Dist201\_400km_j$ is a binary variable taking the value of 1 if j indexes a district laying within 201 km to 400 km of the improved road and zero otherwise. $Dist401\_600km_j$ is a binary variable taking the value of one if j indexes a district laying within 401 km to 600 km of the improved road and zero otherwise. In Model 2, the parameters of interest are θ_1,θ_2 and θ_3. θ_1 gives the estimate of the average change in outcomes for firms located within 200 km of the improved road in comparison with the average change in outcomes of firms laying more than 600 km from the road. θ_2 measures the average change in outcomes of firms laying within 201 km to 400 km compared to the average change in outcomes of firms more than 600 km from the improved road. θ_3 measures the average change in outcomes of firms located within 401 km to 600 km compared to the average change in outcomes of firms more than 600 km from the improved road.

3.5.3 Empirical Estimation Model 3

$$Y_{ijt} = \alpha_t + \delta_1 Rank2\_3_j * Post_t + \delta_2 Rank4\_5_j * Post_t + \gamma X_{ijt} + \tau_i + \epsilon_{ijt} \tag{3.11}$$

Here, $Rank2\_3_j$ is a binary variable taking the value of one if district j ranks Port Nacala as its $2nd$or $3rd$ nearest trading seaport and zero otherwise. $Rank4\_5_j$ is a binary variable taking the value of one if district j ranks Port Nacala as its $4th$or $5th$ nearest trading seaport and

[30] $Post_t$is a dummy variable that equals 1 if the year is 2017. This dummy has no group subscript because time period does not vary across groups

[31] $d400_j$ dummy has no time subscript because group membership is time invariant.

zero otherwise. In this model, the parameters of interests are δ_1 and δ_2; δ_1 gives the average change in outcomes of firms located in districts with Rank2_3 compared to the average change in districts with Rank6_8 while δ_2 gives the differential change in outcomes of firms located in Rank4_5 in comparison to firms in Rank6_8.

3.5.4 Identification Strategy

Estimating the causal effect of road projects is difficult because of the endogeneity problem associated with road development projects. Roads are either placed or improved in areas experiencing high economic activity or in places lagging behind economically in order to stimulate economic activity. In this essay, to identify the effect of the improved road on firm performance, I use two identification strategies. First, I assume common trends in outcome variables over time in the absence of treatment in the control and treated groups.[32] I am not able to check for pre-trends in the outcome variable in the control and treatment groups as the relevant data is only available for one pre-treatment period. However, I check for pre-trends using district population values for 1990 and 2010. Appendix Figure 3.A.4 shows population trends by treatment group. The data indicates that districts in the treatment and control groups have pre-trends in covariates. This signals the possibility that districts in the two groups may not be on divergent paths in relation to the dependent variables in the pre-treatment period. Secondly, DID estimation is prone to endogeneity problems of simultaneity or reverse causality.[33] Therefore, I treat the improvement of this international highway as a natural experiment to strengthen the identification that the intervention is unrelated to the baseline outcome. This is because the placement and improvement of the Nacala Road Corridor were driven by the high economic activity in the capital cities of Zambia and Malawi, and, thus, areas between these two capital cities were randomly treated. Areas in-between the two capital cities were included in treatment because they happen to be located along the route between the two cities with the shortest driving distance and not because of existing baseline economic activity. Thus, the development of the road is exogenous in this area. Therefore, I run estimations on two samples, one in which Lusaka (the endogenously connected city) is dropped, and I emphasise the results of this sample. The technique of treating the highway placement as exogenous to the non-metropolitan areas it runs through is one used in Chandra and Thompson (2000), Michaels (2008) and Datta (2012). These argue that if the precise route of road location or improvement is not manipulated to include or exclude some intermediate areas because of factors correlated with the outcomes of interest, then the road construction can be treated as exogenous to the areas through which the road runs.

[32] In the absence of treatment, the unobserved differences of firms in both groups remains constant overtime. I cannot directly test the common trends assumption but I can check for pre-trends in the outcome variable in the control and treatment groups to check whether the groups were trending on different paths.

[33] Was the allocation of the intervention influenced by the outcome variable at baseline.

3.6 Results

This section proceeds as follows. For each firm outcome, results are presented based on the three model specifications. The first table shows the differential effect of the improved road on outcomes of firms within 400 km compared to firms outside this radius. The second table shows the effect of proximity to the improved highway on firm outcomes over time. The last table indicates the impact on firms of a change in accessibility to Port Nacala induced by the improved road.

3.6.1 Firm output

Table 3.4 shows the effects of the improved road on firm gross profits and sales estimated under empirical Model 1. The group dummy variable $d400km_j$ captures unobserved time invariant differences between groups. The term of interest is $d400km_j^{*}Post_t$, an interaction term of firms in districts laying within 400 km of the improved road and the post-construction period. I include time and firm fixed effects in all estimations. I control for differences in district trends in Columns 2 and 4 by including an interaction term of the total population of each district with the post-period variable. All estimations use robust standard errors clustered at the district level. Coefficients on the term of interest differ in the two samples, and, therefore, I emphasise results for the sample excluding Lusaka (over those for the full sample) as it is controlled for the endogeneity problem of reverse causality.

Results in Table 3.4 indicate that after the road is improved, relative to those further out, firms within 400 km experience growth in gross profits and sales that are statistically insignificant. A major concern with these results is that the coefficients are large and, therefore, noisy. [34] This indicates that the road did have some effect on firm output, but given the available dataset, results are inconclusive. By contrast, coefficients on gross profits are positive while those on sales are negative. Negative differential effects in sales after the improvements suggest areas located within 400 km of the road experience increased competition from new products entering the market owing to high transport-cost reductions in the area. Table 3.4 further shows that the negative coefficients on sales are large in the sample excluding Lusaka (-0.86 and -0.787) but reduce in size (-0.004 and -0.271) when Lusaka is included. This implies that after the road is improved, the growth in sales in Lusaka is likely positive. As for firms in locations outside Lusaka and within 400 km of the improved road, Column 4 shows that they experience a growth in sales that is 54.48 per cent less (-0.787) after the road is improved than for locations more than 400km from the improved road.[35] Results in the two samples suggest that

[34]The standard errors are large and so results are statistically insignificant. Standard errors are large in both samples but they are much larger in the sample that excludes Lusaka because the sample size is smaller.

[35]When regression models have log transformed outcomes the impact of a one-unit change in a covariate (X) is calculated by exponentiating the coefficient, and subtracting one from the number (Mitnik et al, 2018). In this case $(exp\,(\beta_j) - 1) = exp\,(-0.788) - 1 = -0.54$. To give it a percentage interpretation multiply this result by 100. When the estimated coefficient is less than 0.10 the interpretation that a unit increase in X is associated with an average of $100 * \beta_j$ percent increase in Y works well (Mitnik et al, 2018). I refer to the

firms in the centre supply products to the peripheral areas; consumers in the periphery find it convenient to buy cheap products directly from the centre because of reduced transport costs. [36]My findings are similar to Faini (1983), who argues that the reduction in transport costs between Northern and Southern Italy in the 1950s deprived firms in the Mezzogiorno of the protection they previously enjoyed and accelerated the de-industrialisation process in Southern Italy. Puga (2002) also reports that a better connection between regions at different levels of development not only gives firms in the less developed region better access to the inputs and markets of the more developed region but makes it easier for firms in richer regions to supply poorer regions at a distance; this can harm the industrialisation prospects of less developed areas. Further, firms outside Lusaka positioned within 400 km of the road experience positive growth in gross profits after improvements of 268 per cent (1.302) compared to those outside this radius. This implies that these firms experience a substantial fall in production costs that generate a positive gross profit despite experiencing negative growth in sales. Results in Table 3.4 also show that estimates on gross profits and sales do not change much after controlling for differences in district trends in both samples. This indicates that firms within 400 km of the improved road may not be on a different growth trajectory than firms outside this radius.[37]

Table 3.4: Effects of Improved Road on Firm Output - Model 1

	(1) Gross Profit	(2) Gross Profit	(3) Sales	(4) Sales
Sample Excluding Lusaka				
Post	-0.747	-0.375	0.269	-0.342
	(0.611)	(4.842)	(0.503)	(1.407)
d400km*Post	1.332	1.302	-0.836	-0.787
	(0.905)	(1.020)	(0.724)	(0.733)
Total Population*Post	No	Yes	No	Yes
Observations	1322	1322	1322	1322
R^2	0.0637	0.0637	0.2126	0.2129
Full Sample				
Post	-0.016	1.591	-0.597	-2.352
	(0.583)	(3.024)	(0.371)	(1.438)
d400km*Post	0.552	0.795	-0.004	-0.271
	(0.381)	(0.488)	(0.216)	(0.363)
Total Population*Post	No	Yes	No	Yes
Observations	3512	3512	3512	3512
R^2	0.0538	0.0547	0.2045	0.2067

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: Standard errors in parentheses and clustered at location (district) level. The full sample has 48 clusters while the sample excluding Lusaka has 47 clusters. The full sample includes Lusaka city, which is located within 400km of the improved road. All variables are in inverse hyperbolic sine. Inverse hyperbolic sine transformation is similar to a logarithm transformation except that this transformation allows for retaining zero and negative valued observations. Post is a dummy variable that equals one when the year is 2017 and zero otherwise. d400km is a dummy variable that takes the value one if a firm is located in a district within 400km of the improved road or zero otherwise. To control for differences in district growth trends, regressions in column 2 and 4 include the interaction term Total Population*Post.

exponentiated coefficients throughout the paper unless indicated otherwise.

[36]Products flowing to the periphery are either high quality or cheaply priced identical products or substitutes.

[37]In the small sample, coefficients on gross profit fall from 1.332 to 1.302 while that of sales fall from -0.836 to -0.787.

Table 3.5 displays results estimated under empirical Model 2, where the terms of interest are $Dist0\_200km_j^{*}Post_t$, $Dist201\_400_j^{*}Post_t$ and $Dist401\_600_j * Post_t$. The coefficients on the terms of interest are statistically insignificant in both samples. As in the previous model, I interpret these results as inconclusive because the coefficients are large. However, based on the size of coefficients and direction of effect, the following is deduced. In the sample excluding Lusaka, compared to firms more than 600 km from the improved road, firms within 200 km experience differential changes in gross profits and sales that are negative. Columns 2 and 4 indicate that after the road is improved, firms within 200 km experience negative growth in gross profits and sales of 99.32 per cent (−4.989) and 99.23 per cent (−4.866) respectively, compared to firms located more than 600 km from the improved road. This shows that firms at this distance experience high competition effects because of increased area accessibility. Firms in the distance bin of 201 km and 400 km, as shown in Column 2, after the improvements experience differential changes in gross profits of 242 per cent higher (1.230) than those of firms located more than 600 km from the road. Positive growth in gross profits implies gains from reduced costs in production in this distance bin. As Column 4 indicates, these firms experience a differential change in sales of 19 per cent less (−0.214) after the road is improved. After improvements, firms in the third distance bin located between 401 km and 600 km experience a differential change in gross profit of 59.59 per cent less (−0.906) and a differential change in sales of 12.41 per cent higher (0.117) in comparison to firms located more than 600 km from the road. This shows that due to their distance from the improved road, firms in this area experience fewer competition effects and, because of minimal reductions in transport costs, do not enjoy reduced costs in production induced by the improved road. Further, in Columns 2 and 4, where I control for differences in district trends, in both samples, the size of the estimates in absolute terms reduces as the distance from the improved road increases. This shows the road effects are less pronounced the farther a firm is located from the improved road.

Table 3.5: Effects of Improved Road on Firm Output - Model 2

	(1) Gross Profit	(2) Gross Profit	(3) Sales	(4) Sales
Sample Excluding Lusaka				
Post	-0.306	-1.828	-0.100	-0.820
	(1.986)	(4.220)	(0.545)	(1.634)
Dist0_200km*Post	-4.790	-4.989	-4.771	-4.866
	(4.408)	(4.391)	(4.098)	(4.063)
Dist201_400km*Post	1.331	1.230	-0.166	-0.214
	(1.962)	(2.030)	(0.544)	(0.552)
Dist401_600km*Post	-0.633	-0.906	0.246	0.117
	(1.852)	(2.108)	(0.315)	(0.474)
Total Population*Post	No	Yes	No	Yes
Observations	1322	1322	1322	1322
R^2	0.0848	0.0855	0.2464	0.2469
Full Sample				
Post	0.832	2.363	-0.622	-1.685^{*}
	(1.946)	(2.861)	(0.396)	(0.981)
Dist0_200km*Post	-4.978	-4.638	-4.719	-4.955
	(4.327)	(4.508)	(4.118)	(4.023)
Dist201_400km*Post	0.050	0.693	0.327	-0.119
	(1.845)	(2.167)	(0.289)	(0.500)
Dist401_600km*Post	-0.689	-0.301	0.237	-0.033
	(1.855)	(2.030)	(0.295)	(0.399)
Total Population*Post	No	Yes	No	Yes
Observations	3512	3512	3512	3512
R^2	0.0595	0.0605	0.2158	0.2168

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: Standard errors in parentheses and clustered at location (district) level. The full sample has 48 clusters while the sample excluding Lusaka has 47 clusters. The full sample includes Lusaka city, which is located within 400km of the improved road. All variables are in inverse hyperbolic sine. Inverse hyperbolic sine transformation is similar to a logarithm transformation except that this transformation allows for retaining zero and negative valued observations. Post is a dummy variable that equals one when the year is 2017 and zero otherwise. Dist0_200km is a dummy variable that takes the value one if a firm is located in a district within 200km of the improved road or zero otherwise. Dist201_400km is a dummy variable that equals one if a firm is located in a district laying within 201 to 400km from the improved road. Dist401_600km is a dummy variable that equals one if a firm is located in a district laying within 401 to 600km from the improved road. To control for differences in district growth trends, regressions in column 2 and 4 include the interaction term Total Population*Post. The control group includes firms in districts laying within 601 to 1100km, which is the furthest point in the dataset.

Table 3.6 shows results generated from Model 3. The terms of interest are $Rank2\_3_j * Post_t$ and $Rank4\_5_j * Post_t$. These capture growth in firm output resulting from changes in the accessibility of Port Nacala. Similar to results generated in previous model specifications, Table 3.6 shows that results are inconclusive. In both samples, firms in districts with Ranks 2 to 3 and Ranks 4 to 5 experience changes in sales and gross profit that are negative compared to firms in districts with Ranks 6 to 8 after the road is improved. This signals that firms with increased accessibility to the seaport may be facing increased competition.

The general conclusion under firm output is that the impact of the improved road on firm output in the short run is inconclusive given the available dataset. Results are insignificant with large coefficients, and, therefore, it cannot be stated that the road did not have any effects on firm output. These findings are robust to the different model specifications used in this essay, and the following is deduced from the coefficient size and direction of effect. Coefficients on gross profits and sales are negative for firms within 200 km of the improved road, which suggest increased competition in this distance bin. Firms in the distance bin of 200 km to 400 km and those within 400 km of the improved road, coefficients on gross profits are positive while those on sales remain negative. Negative differential changes in sales are associated with increased competition effects, while the growth in GDP is attributed to advantages in production as a result of reduced transport costs. Similarly, firms with increased access to Port Nacala experience negative differential changes in gross profits and sales after the road is improved.

Table 3.6: Effects of Improved Road on Firm Output - Model 3

	(1) Gross Profit	(2) Gross Profit	(3) Sales	(4) Sales
Sample Excluding Lusaka				
Post	1.354	-0.323	0.226	-1.641
	(1.870)	(4.773)	(0.443)	(1.730)
Rank2_3*Post	-2.484	-2.747	-0.289	-0.582
	(1.718)	(1.868)	(0.319)	(0.501)
Rank4_5*Post	-1.140	-1.226	-0.398	-0.494
	(1.825)	(1.846)	(0.403)	(0.433)
Total Population*Post	No	Yes	No	Yes
Observations	1322	1322	1322	1322
R^2	0.0740	0.0750	0.2027	0.2063
Full Sample				
Post	1.100	0.541	-0.502	-1.425*
	(1.746)	(2.471)	(0.421)	(0.746)
Rank2_3*Post	-2.411	-2.645	-0.138	-0.524
	(1.700)	(1.798)	(0.329)	(0.402)
Rank4_5*Post	-0.999	-1.277	-0.076	-0.536
	(1.795)	(1.835)	(0.455)	(0.468)
Total Population*Post	No	Yes	No	Yes
Observations	3512	3512	3512	3512
R^2	0.0589	0.0592	0.2045	0.2064

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: Standard errors in parentheses and clustered at location (district) level. The full sample has 48 clusters while the sample excluding Lusaka has 47 clusters. All variables are in inverse hyperbolic sine. Inverse hyperbolic sine transformation is similar to a logarithm but unlike logarithm, transformation allows for retaining zero and negative valued observations. Post is a dummy variable that takes the value one when the year is 2017 and zero otherwise. Rank2_3 is a dummy variable that takes the value one when a firm is located in a district whose distance from Port Nacala is ranked 2 or 3 but zero otherwise. Rank4_5 is a dummy variable that takes the value one when a firm is located in a district whose distance from Port Nacala is ranked 4 or 5 or zero otherwise. Rank2_3*Post is an interaction term which gives the differential change of firms in Rank2_3 compared to firms in Rank6_8 (the control group). Rank4_5*Post shows the differential changes of firms in Rank4_5 compared to firms in Rank6_8.

3.6.2 Firm Labour Costs

Table 3.7 displays results of the impact of the improved road on firm labour costs generated under Model 1. Two types of labour costs are considered, direct wage costs and total employment costs. Direct wage costs reflect firm labour expenses on employees directly engaged in the production process while total employment expenses include direct wage costs and all other employee-related expenses, such as salaries of administrative support staff and pension contributions. In the sample excluding Lusaka, the results of the impact of the road are insignificant. However, the coefficients are not small enough to confidently claim that the improved road has no effect on firm labour costs. Based on the size of the coefficients, Column 2 shows that after the road is improved, firms outside Lusaka within 400 km of the road experience a differential change in direct wages that is 54.30 per cent less (−0.783) than for firms located outside this radius. The reduction in direct labour costs for firms in this area is not unusual given that after the road is improved, these firms experience a differential change in sales that is 54.48 per cent less than that for firms outside this radius. Results on direct wages in the sample that includes Lusaka reveal similar effects, other than the much smaller coefficients (−0.062). This suggests the differential change in direct wage costs in Lusaka is positive. Further, Column 4 indicates that after the road is improved, firms outside Lusaka located within 400 km of the improved road experience small and insignificant effects on employment expenses (coefficient of 0.005). When Lusaka is added to the sample, these firms experience differential changes in employment expenses of 39.10 per cent less than firms outside this radius. The results for labour costs show that after the road is improved, as sales fall because of increased competition in the area, firms outside Lusaka close to the improved road cut labour in production. Firms in the centre, by contrast, experience positive growth in direct wage costs either because they hire more workers or increase working hours to increase the supply of products to the peripheral areas. However, these firms cut down employment expenses for labour not directly engaged in production.

Table 3.7: Effect of Improved Road on Firm Labour Costs - Model 1

	(1) Direct Wages	(2) Direct Wages	(3) Employment Expense	(4) Employment Expense
Sample Excluding Lusaka				
Post	0.234	0.130	-0.435	-2.820**
	(0.876)	(1.977)	(0.341)	(1.260)
d400km*Post	-0.792	-0.783	-0.186	0.005
	(0.884)	(0.883)	(0.336)	(0.366)
Total Population*Post	No	Yes	No	Yes
Observations	1322	1322	1322	1322
R^2	0.0757	0.0758	0.1031	0.1077
Full Sample				
Post	-0.976*	-2.763	-0.186	-0.408
	(0.500)	(1.925)	(0.291)	(1.078)
d400km*Post	0.209	-0.062	-0.462**	-0.496*
	(0.229)	(0.418)	(0.220)	(0.275)
Total Population*Post	No	Yes	No	Yes
Observations	3512	3512	3512	3512
R^2	0.0114	0.0135	0.0761	0.0762

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: Standard errors in parentheses and clustered at location (district) level. The full sample has 48 clusters while the sample excluding Lusaka has 47 clusters. The full sample includes Lusaka city, which is located within 400km of the improved road. All variables are in inverse hyperbolic sine. Inverse hyperbolic sine transformation is similar to a logarithm transformation except that this transformation allows for retaining zero and negative valued observations. Post is a dummy variable that equals one when the year is 2017 and zero otherwise. d400km is a dummy variable that takes the value one if a firm is located in a district within 400km of the improved road or zero otherwise. To control for differences in district growth trends, regressions in column 2 and 4 include the interaction term Total Population*Post.

Table 3.8 displays the results generated using empirical Model 2. Results in this table are inconclusive. Coefficients on the variables of interest are not small but are statistically insignificant in both samples. Columns 2 and 4 show that firms outside Lusaka within 200 km of the improved road experience a differential change in direct wages that is 23.74 per cent higher (0.213) and a differential change in total expenses that is 84.92 per cent lower (−1.892) than for firms located more than 600 km from the improved road. The negative growth in employment expenses indicates firms in this area cut down expenses on labour not engaged directly in production. Additionally, firms located between 200 and 600 km away from the improved road experience a differential change in direct wages that is less than for firms located more than 600 km from the improved road. The growth in direct wages is positive only for firms within 200 km. This signals a movement of direct wage labour from areas located more than 200 km to areas near the improved road. The negative growth in direct wages is highest in the distance bin adjacent to the 200 km distance bin. This finding supports the theory that improved roads increase economic activity in surrounding areas.[38]

[38] New firms enter the area surrounding the improved road or existing firms expand operations due to increased accessibility to markets. See Gibbons et al (2019).

Table 3.8: Effect of Improved Road on Firm Labour Costs - Model 2

	(1) Direct Wages	(2) Direct Wages	(3) Employment Expense	(4) Employment Expense
Sample Excluding Lusaka				
Post	0.302	-1.072	-0.375	-1.455
	(1.159)	(1.912)	(0.471)	(1.446)
Dist0_200km*Post	0.396	0.213	-1.748	-1.892
	(0.913)	(1.003)	(2.280)	(2.227)
Dist201_400km*Post	-1.246	-1.342	0.095	0.020
	(1.297)	(1.314)	(0.375)	(0.396)
Dist401_600km*Post	-0.450	-0.702	0.156	-0.042
	(0.873)	(1.055)	(0.354)	(0.446)
Total Population*Post	No	Yes	No	Yes
Observations	1322	1322	1322	1322
R^2	0.0292	0.0300	0.0171	0.0180
Full Sample				
Post	-0.645	-2.875	-0.076	0.081
	(1.039)	(1.871)	(0.422)	(0.872)
Dist0_200km*Post	0.561	0.063	-1.800	-1.765
	(0.939)	(0.985)	(2.257)	(2.279)
Dist201_400km*Post	-0.171	-1.110	-0.244	-0.178
	(0.891)	(1.239)	(0.342)	(0.456)
Dist401_600km*Post	-0.390	-0.960	0.137	0.177
	(0.888)	(1.017)	(0.361)	(0.414)
Total Population*Post	No	Yes	No	Yes
Observations	3512	3512	3512	3512
R^2	0.0104	0.0145	0.0061	0.0061

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: Standard errors in parentheses and clustered at location (district) level. The full sample has 48 clusters while the sample excluding Lusaka has 47 clusters. The full sample includes Lusaka city, which is located within 400km of the improved road. All variables are in inverse hyperbolic sine. Inverse hyperbolic sine transformation is similar to a logarithm transformation except that this transformation allows for retaining zero and negative valued observations. Post is a dummy variable that equals one when the year is 2017 and zero otherwise. Dist0_200km is a dummy variable that takes the value one if a firm is located in a district within 200km of the improved road or zero otherwise. Dist201_400km is a dummy variable that equals one if a firm is located in a district laying within 201 to 400km from the improved road. Dist401_600km is a dummy variable that equals one if a firm is located in a district laying within 401 to 600km from the improved road. To control for differences in district growth trends, regressions in column 2 and 4 include the interaction term Total Population*Post. The control group includes firms in districts laying within 601 to 1100km, which is the furthest point in the dataset.

Table 3.9 shows the results of empirical Model 3. Results in this table are largely statistically insignificant and, based on the size of the estimates, are also inconclusive. Column 2 shows that after the road is improved, firms outside Lusaka, with Port Nacala as their nearest seaport, experience a growth in direct wages that is 57 per cent lower (−0.844) and a growth in employment expenses that is 34. 30 per cent lower (−0.420) than that of firms in districts with Port Nacala as their 6th (or 7th or 8th) farthest seaport. The negative growth in labour costs occurs as these firms experience negative growth in sales and gross profits due to increased competition that arises after the road is improved. The effect of the road is similar in the sample, including Lusaka, except the coefficient on direct wages is now weakly statistically significant.

The summary in relation to labour costs is that the results are inconclusive since the estimates are large and statistically insignificant. Coefficients signal that firms outside Lusaka within 400 km of the improved road and those near to the Port of Nacala compared to those far away, experience negative growth in labour costs. The analysis under the 200 km distance bin further indicates that firms in this area experience positive growth in direct wage costs while firms in distance bins greater than 200 km experience negative growth. This signals labour moving to locations near the road after it is improved.

Table 3.9: Effect of Improved Road on Firm Labour Costs - Model 3

	(1) Direct Wages	(2) Direct Wages	(3) Employment Expense	(4) Employment Expense
Sample Excluding Lusaka				
Post	-0.166	-2.265	-0.356	-2.100
	(0.822)	(2.609)	(0.591)	(1.307)
Rank2_3*Post	-0.504**	-0.844	-0.137	-0.420
	(0.189)	(0.558)	(0.591)	(0.685)
Rank4_5*Post	-0.318	-0.433	0.108	0.012
	(0.624)	(0.619)	(0.561)	(0.608)
Total Population*Post	No	Yes	No	Yes
Observations	1322	1322	1322	1322
R^2	0.0241	0.0262	0.0123	0.0151
Full Sample				
Post	-1.106**	-2.366**	0.557	0.977
	(0.474)	(0.951)	(0.595)	(0.890)
Rank2_3*Post	-0.326	-0.853*	-0.309	-0.133
	(0.251)	(0.432)	(0.532)	(0.593)
Rank4_5*Post	0.202	-0.427	-0.396	-0.187
	(0.521)	(0.610)	(0.537)	(0.548)
Total Population*Post	No	Yes	No	Yes
Observations	3512	3512	3512	3512
R^2	0.0103	0.0135	0.0020	0.0025

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: Standard errors in parentheses and clustered at location (district) level. The full sample has 48 clusters while the sample excluding Lusaka has 47 clusters. All variables are in inverse hyperbolic sine. Inverse hyperbolic sine transformation is similar to a logarithm but unlike logarithm, transformation allows for retaining zero and negative valued observations. Post is a dummy variable that takes the value one when the year is 2017 and zero otherwise. Rank2_3 is a dummy variable that takes the value one when a firm is located in a district whose distance from Port Nacala is ranked 2 or 3 but zero otherwise. Rank4_5 is a dummy variable that takes the value one when a firm is located in a district whose distance from Port Nacala is ranked 4 or 5 or zero otherwise. Rank2_3*Post is an interaction term which gives the differential change of firms in Rank2_3 compared to firms in Rank6_8 (the control group). Rank4_5*Post shows the differential changes of firms in Rank4_5 compared to firms in Rank6_8.

3.6.3 Firm Transportation Costs

Table 3.10 shows the effects of the improved road on firm transportation costs of namely, freight transport expense and fuel and lubricant costs. Freight transport expenses are logistical costs arising from transporting inputs and finished goods and services by road, rail, air and sea. These logistical costs include distribution costs pertaining to handling, packaging and storing goods as well as insurance costs, seaport and border-clearance costs. Fuel and Lubricant costs are expenses related to fuel and oil use by the firm vehicle fleet. Column 2 in the sample excluding

Lusaka indicates that firms within 400 km of the improved road experience a differential change in freight transport costs that is statistically insignificant but 139.65 per cent higher (0.874) than firms more than 400 km from the improved road. Column 4 shows that the same firms experience a differential change in fuel and lubricant costs that is 414.49 per cent higher (1.638) compared to firms located more than 400 km from the improved road. The differential change in fuel and lubricant costs is statistically significant at the 1 per cent level in both samples. The results for fuel and lubricants costs signal that after the road is improved, firms within 400 km of the road increase the utilisation of road transport services through increasing the number of fleet trips to markets (this could be more trips to old markets or trips to new markets previously inaccessible because of poor roads). This results in increases in fuel and lubricant costs under operating expenses in a firm's income statement. This result is in line with the theoretical predictions in this essay that improved roads cause a fall in the price of firm transportation services, which causes firms to intensify their use of road transport services in production (the now cheaper input) provided the elasticity of substitution between inputs is greater than one. The growth in firm fuel and lubricant costs is associated with increased utilisation of the transport infrastructure and not fuel-pump-price increases – the DID strategy used to estimate the impact of the improved road on fuel and lubricant costs controls for time fixed effects. An increase in fuel pump prices happens countrywide, and thus factors that affect the whole dataset over time, such as fuel price increases, are controlled for under the DID estimation strategy.

Table 3.10: Effect of Improved Road on Firm Transportation Costs - Model 1

	(1) Transport	(2) Transport	(3) Fuel & Lubricants	(4) Fuel & Lubricants
Sample Excluding Lusaka				
Post	-1.142**	-0.603	-0.185	3.063
	(0.472)	(1.538)	(0.461)	(2.621)
d400km*Post	0.917*	0.874	1.900***	1.638***
	(0.510)	(0.545)	(0.564)	(0.589)
Total Population*Post	No	Yes	No	Yes
Observations	1322	1322	1322	1322
R^2	0.0276	0.0279	0.0483	0.0545
Full Sample				
Post	-0.548	0.891	0.604	3.598**
	(0.425)	(1.194)	(0.483)	(1.603)
d400km*Post	0.276	0.493*	1.049***	1.502***
	(0.210)	(0.250)	(0.282)	(0.244)
Total Population*Post	No	Yes	No	Yes
Observations	3512	3512	3512	3512
R^2	0.0060	0.0079	0.0140	0.0194

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: Standard errors in parentheses and clustered at location (district) level. The full sample has 48 clusters while the sample excluding Lusaka has 47 clusters. The full sample includes Lusaka city, which is located within 400km of the improved road. All variables are in inverse hyperbolic sine. Inverse hyperbolic sine transformation is similar to a logarithm transformation except that this transformation allows for retaining zero and negative valued observations. Post is a dummy variable that equals one when the year is 2017 and zero otherwise. d400km is a dummy variable that takes the value one if a firm is located in a district within 400km of the improved road or zero otherwise. To control for differences in district growth trends, regressions in column 2 and 4 include the interaction term Total Population*Post.

Table 3.11 shows results generated from empirical Model 2. In both samples, differential changes in freight transport costs for firms located in the various distance bands compared to firms more than 600 km from the improved road are inconclusive though the sign of the effect looks largely positive. Further, the size of estimates on freight transport falls with increased distance from the improved road. As regards fuel and lubricant costs, Column 4 shows that compared to firms located more than 600 km from the improved road, the differential change in fuel and lubricant costs for firms located between 201 km and 400 km is positive. The coefficients are 2.787 and 2.674 in the small sample and full sample, respectively, and are strongly statistically significant. These results indicate that after the road is improved, firms outside Lusaka between 201 and 400 km from the road experience growth in fuel and lubricants of 1,523 per cent higher (2.787) than that of firms located more than 600 km from the road. Firms within 200 km experience negative differential changes that are statistically insignificant compared to firms more than 600 km from the improved road. A plausible explanation for this result is that firms in this distance band experience high competition effects and fail to increase their utilisation of the road infrastructure.

Table 3.11: Effect of Improved Road on Firm Transportation Costs - Model 2

	(1) Transport	(2) Transport	(3) Fuel & Lubricants	(4) Fuel & Lubricants
Sample Excluding Lusaka				
Post	-1.042	-0.235	-0.495	3.190
	(0.705)	(1.413)	(0.582)	(2.749)
Dist0_200km*Post	3.000	3.108	-2.288	-1.796
	(1.984)	(2.035)	(2.141)	(1.975)
Dist201_400km*Post	0.660	0.716	2.530***	2.787***
	(0.755)	(0.744)	(0.474)	(0.515)
Dist401_600km*Post	-0.036	0.112	0.182	0.858
	(0.554)	(0.594)	(0.371)	(0.708)
Total Population*Post	No	Yes	No	Yes
Observations	1322	1322	1322	1322
R^2	0.0366	0.0373	0.0703	0.0783
Full Sample				
Post	-0.596	0.513	0.647	4.067***
	(0.685)	(1.261)	(0.521)	(1.327)
Dist0_200km*Post	2.922	3.170	-2.487	-1.723
	(1.954)	(2.060)	(2.218)	(1.920)
Dist201_400km*Post	0.153	0.620	1.234***	2.674***
	(0.582)	(0.700)	(0.428)	(0.618)
Dist401_600km*Post	-0.065	0.219	0.109	0.983*
	(0.551)	(0.586)	(0.389)	(0.542)
Total Population*Post	No	Yes	No	Yes
Observations	3512	3512	3512	3512
R^2	0.0103	0.0117	0.0190	0.0277

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: Standard errors in parentheses and clustered at location (district) level. The full sample has 48 clusters while the sample excluding Lusaka has 47 clusters. The full sample includes Lusaka city, which is located within 400km of the improved road. All variables are in inverse hyperbolic sine. Inverse hyperbolic sine transformation is similar to a logarithm transformation except that this transformation allows for retaining zero and negative valued observations. Post is a dummy variable that equals one when the year is 2017 and zero otherwise. Dist0_200km is a dummy variable that takes the value one if a firm is located in a district within 200km of the improved road or zero otherwise. Dist201_400km is a dummy variable that equals one if a firm is located in a district laying within 201 to 400km from the improved road. Dist401_600km is a dummy variable that equals one if a firm is located in a district laying within 401 to 600km from the improved road. To control for differences in district growth trends, regressions in column 2 and 4 include the interaction term Total Population*Post. The control group includes firms in districts laying within 601 to 1100km, which is the furthest point in the dataset.

Table 3.12 displays the results for empirical Model 3. In both samples, firms in districts with Ranks 2–3 experience growth in freight transport costs and fuel and lubricant costs that are inconclusive. By contrast, as indicated in Column 4, firms in Ranks 4–5 from Port Nacala experience a differential change in fuel and lubricant costs that is 938 per cent higher (0.234) than that of firms in Ranks 6–8 from Port Nacala; this result is statistically significant at the 1 per cent level.

The conclusion under this subsection is that the effect of the improved road on a firm's freight transport costs are inconclusive. This result resonates to an extent with studies in Africa that find the effects on transport costs of road quality are insignificant. For instance, Raballand and Macchi (2008) argue that the high transport costs faced by firms in Africa are not because of poor quality road infrastructure but because of inefficiencies in the structure of transport markets. They argue that a large portion of transportation costs in developing countries are driven by factors that are not necessarily related to infrastructure quality, such as inefficient processes and corruption at seaports.[39] Macchi and Sequeira (2009) also argue that the poor function of soft infrastructures at ports due to corruption directly affects port services, which in turn affect the returns to investments made in the hard infrastructure of ports.[40]

The conclusion under fuel and lubricant costs is that firms near the improved highway and those with Port Nacala as the nearest accessible seaport experience positive growth in fuel and lubricant costs compared to firms far away. These results are strongly statistically significant and show that firms close to the road benefit from its improvement and leverage those benefits by increasing road utilisation. Improving the road decreases travel time, which is a transportation cost, and increases accessibility. Therefore, firms optimise by increasing the number of trips or travelled distance, which results in positive growth in firm fuel and lubricant costs.

[39] Raballand and Macchi (2008) in their study of the costs, rates and performance of the trucking industry across Western and Central Africa find that transport prices charged to businesses are high, while the actual costs incurred by trucking companies in Western Africa to move cargo do not differ greatly from those faced by companies in the developed world.

[40] Sequeira et al 2011 define hard infrastructures as including quality of roads, railways and ports, while soft infrastructures are regulations of transport markets, port policies and rules regulating the movement of goods across borders.

Table 3.12: Effect of Improved Road on Firm Transportation Costs - Model 3

	(1) Transport	(2) Transport	(3) Fuel & Lubricants	(4) Fuel & Lubricants
Sample Excluding Lusaka				
Post	-0.561	0.858	-1.058	2.958
	(1.214)	(1.703)	(0.988)	(2.540)
Rank2_3*Post	-0.206	0.024	0.242	0.892
	(0.769)	(0.824)	(0.570)	(0.818)
Rank4_5*Post	-0.093	-0.015	2.118**	2.340**
	(1.064)	(1.064)	(0.817)	(0.902)
Total Population*Post	No	Yes	No	Yes
Observations	1322	1322	1322	1322
R^2	0.0147	0.0169	0.0393	0.0496
Full Sample				
Post	-0.580	-0.296	-1.532**	-1.092
	(0.861)	(0.958)	(0.598)	(1.421)
Rank2_3*Post	-0.202	-0.083	0.331	0.516
	(0.730)	(0.831)	(0.515)	(0.787)
Rank4_5*Post	-0.082	0.060	2.381***	2.601***
	(0.922)	(1.041)	(0.683)	(0.914)
Total Population*Post	No	Yes	No	Yes
Observations	3512	3512	3512	3512
R^2	0.0047	0.0050	0.0136	0.0139

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: Standard errors in parentheses and clustered at location (district) level. The full sample has 48 clusters while the sample excluding Lusaka has 47 clusters. All variables are in inverse hyperbolic sine. Inverse hyperbolic sine transformation is similar to a logarithm but unlike logarithm, transformation allows for retaining zero and negative valued observations. Post is a dummy variable that takes the value one when the year is 2017 and zero otherwise. Rank2_3 is a dummy variable that takes the value one when a firm is located in a district whose distance from Port Nacala is ranked 2 or 3 but zero otherwise. Rank4_5 is a dummy variable that takes the value one when a firm is located in a district whose distance from Port Nacala is ranked 4 or 5 or zero otherwise. Rank2_3*Post is an interaction term which gives the differential change of firms in Rank2_3 compared to firms in Rank6_8 (the control group). Rank4_5*Post shows the differential changes of firms in Rank4_5 compared to firms in Rank6_8.

3.6.4 Firm Inventory and Size

Table 3.13 depicts the effect of the improved road on inventory costs and total assets for firms located within 400 km of the improved road relative to firms outside this radius. After the road is improved, firms outside Lusaka within 400 km of the road experience a differential change in raw materials inventory that is 48.59 per cent higher (0.396) compared to firms outside this radius. This result is inconclusive as it is insignificant, but the estimate is not sufficiently small to strongly support the conclusion that the road had no effect on raw materials inventory costs. However, when firms in Lusaka are included in the sample, the effect is statistically significant, as shown in Column 2.[41] A disclaimer on results in the sample including Lusaka (the endogenously chosen city) is that results are expected to be positively high. Column 4

[41] In the full sample, firms within 400 km experience a growth in raw material inventory of 180.95 percent higher (1.033) compared to firms outside this radius.This indicates firms in the centre experience a growth in raw material inventory driven by an increase in sales after the road is improved.

shows that firms outside Lusaka within 400 km of the 400 km of the improved road, experience growth in assets that is 95 per cent less (−0.2.934) than for firms located outside this radius.[42] This negative growth in assets is attributed to increased competition, which negatively affects current assets through a fall in sales. Positive growth in the costs of raw materials inventory, in conjunction with negative growth in sales, shows firms react to increased competition by slowing down production and holding more inventory after the road is improved.

Table 3.13: Effects of Improved Road on Firm Inventory and Size - Model 1

	(1) Raw Materials Inventory	(2) Raw Materials Inventory	(3) Assets	(4) Assets
Sample Excluding Lusaka				
Post	-0.161	6.523^{*}	-0.104	8.189
	(0.911)	(3.869)	(1.159)	(5.843)
d400km*Post	0.937	0.396	-2.263	-2.934^{*}
	(0.859)	(0.810)	(1.466)	(1.564)
Total Population*Post	No	Yes	No	Yes
Observations	1322	1322	1322	1322
R^2	0.0074	0.0283	0.0317	0.0499
Full Sample				
Post	0.275	4.019^{*}	-0.492	2.268
	(0.440)	(2.109)	(0.971)	(3.860)
d400km*Post	0.467	1.033^{**}	-1.845^{***}	-1.427^{**}
	(0.381)	(0.460)	(0.513)	(0.703)
Total Population*Post	No	Yes	No	Yes
Observations	3512	3512	3512	3512
R^2	0.0033	0.0143	0.0216	0.0235

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: Standard errors in parentheses and clustered at location (district) level. The full sample has 48 clusters while the sample excluding Lusaka has 47 clusters. The full sample includes Lusaka city, which is located within 400km of the improved road. All variables are in inverse hyperbolic sine. Inverse hyperbolic sine transformation is similar to a logarithm transformation except that this transformation allows for retaining zero and negative valued observations. Post is a dummy variable that equals one when the year is 2017 and zero otherwise. d400km is a dummy variable that takes the value one if a firm is located in a district within 400km of the improved road or zero otherwise. To control for differences in district growth trends, regressions in column 2 and 4 include the interaction term Total Population*Post.

Columns 2 and 4 In Table 3.14 show that firms located within different distance bins from the improved road compared to firms in the last bin (those located more than 600 km from the improved road) experience negative differential changes in raw materials inventory costs and assets that are statistically insignificant. The estimates are large, and these results are thus considered inclusive.

[42]This result is significant at the 10 percent level.

Table 3.14: Effects of Improved Road on Firm Inventory and Size - Model 2

	(1) Raw Materials	(2) Raw Materials	(3) Assets	(4) Assets
Sample Excluding Lusaka				
Post	1.943	5.224	1.852	5.916
	(1.603)	(3.706)	(2.124)	(5.171)
Dist0_200km*Post	-1.848	-1.410	-12.976*	-12.433
	(1.247)	(1.500)	(7.438)	(7.659)
Dist201_400km*Post	-1.018	-0.789	-3.476*	-3.192
	(1.582)	(1.657)	(2.035)	(2.025)
Dist401_600km*Post	-2.183*	-1.581	-2.277	-1.531
	(1.252)	(1.631)	(1.808)	(1.787)
Total Population*Post	No	Yes	No	Yes
Observations	1322	1322	1322	1322
R^2	0.0330	0.0380	0.0854	0.0897
Full Sample				
Post	2.378*	4.358*	2.035	3.838
	(1.321)	(2.269)	(1.987)	(3.347)
Dist0_200km*Post	-1.923	-1.481	-13.008*	-12.605
	(1.230)	(1.420)	(7.412)	(7.580)
Dist201_400km*Post	-1.511	-0.677	-3.683**	-2.924
	(1.220)	(1.654)	(1.795)	(1.912)
Dist401_600km*Post	-2.211*	-1.705	-2.289	-1.828
	(1.238)	(1.470)	(1.802)	(1.824)
Total Population*Post	No	Yes	No	Yes
Observations	3512	3512	3512	3512
R^2	0.0183	0.0221	0.0392	0.0402

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: Standard errors in parentheses and clustered at location (district) level. The full sample has 48 clusters while the sample excluding Lusaka has 47 clusters. The full sample includes Lusaka city, which is located within 400km of the improved road. All variables are in inverse hyperbolic sine. Inverse hyperbolic sine transformation is similar to a logarithm transformation except that this transformation allows for retaining zero and negative valued observations. Post is a dummy variable that equals one when the year is 2017 and zero otherwise. Dist0_200km is a dummy variable that takes the value one if a firm is located in a district within 200km of the improved road or zero otherwise. Dist201_400km is a dummy variable that equals one if a firm is located in a district laying within 201 to 400km from the improved road. Dist401_600km is a dummy variable that equals one if a firm is located in a district laying within 401 to 600km from the improved road. To control for differences in district growth trends, regressions in column 2 and 4 include the interaction term Total Population*Post. The control group includes firms in districts laying within 601 to 1100km, which is the furthest point in the dataset.

Columns 2 and 4 in Table 3.15 show that firms in districts with Ranks 2–3 and Ranks 4–5 experience a positive differential change in raw materials inventory costs and a negative differential change in assets when compared to firms in Ranks 6–8. These results are also inconclusive.

The general conclusion under this subsection is that after the road is improved, firms close to it or firms with increased accessibility to Port Nacala experience growth in inventory costs that are inconclusive given the available data. Differential changes in assets are negative in all model specifications but are significant only in Model 1. Negative asset growth is not unusual for firms experiencing negative changes in sales and labour costs due to increased competition in the area immediately surrounding the improved road. Negative effects in sales will affect current assets and total assets negatively. This signals that the firms close to the improved road are shrinking after the improvements. This is attributed to increased competition effects.

Table 3.15: Effects of Improved Road on Firm Inventory and Size - Model 3

	(1) Raw Materials Inventory	(2) Raw Materials Inventory	(3) Assets	(4) Assets
Sample Excluding Lusaka				
Post	-0.082	7.243	1.916	4.206
	(1.293)	(4.501)	(3.430)	(6.124)
Rank2_3*Post	-0.118	1.069	-3.249	-2.878
	(0.411)	(0.939)	(3.168)	(3.083)
Rank4_5*Post	0.493	0.896	-3.076	-2.950
	(0.883)	(0.782)	(3.276)	(3.216)
Total Population*Post	No	Yes	No	Yes
Observations	1322	1322	1322	1322
R^2	0.0030	0.0303	0.0294	0.0309
Full Sample				
Post	-0.348	0.975	4.516	7.253*
	(0.528)	(1.691)	(3.236)	(3.790)
Rank2_3*Post	-0.068	0.486	-3.740	-2.595
	(0.237)	(0.782)	(3.077)	(3.094)
Rank4_5*Post	0.640	1.300	-4.513	-3.146
	(0.545)	(0.802)	(3.163)	(3.125)
Total Population*Post	No	Yes	No	Yes
Observations	3512	3512	3512	3512
R^2	0.0019	0.0061	0.0151	0.0208

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: Standard errors in parentheses and clustered at location (district) level. The full sample has 48 clusters while the sample excluding Lusaka has 47 clusters. All variables are in inverse hyperbolic sine. Inverse hyperbolic sine transformation is similar to a logarithm but unlike logarithm, transformation allows for retaining zero and negative valued observations. Post is a dummy variable that takes the value one when the year is 2017 and zero otherwise. Rank2_3 is a dummy variable that takes the value one when a firm is located in a district whose distance from Port Nacala is ranked 2 or 3 but zero otherwise. Rank4_5 is a dummy variable that takes the value one when a firm is located in a district whose distance from Port Nacala is ranked 4 or 5 or zero otherwise. Rank2_3*Post is an interaction term which gives the differential change of firms in Rank2_3 compared to firms in Rank6_8 (the control group). Rank4_5*Post shows the differential changes of firms in Rank4_5 compared to firms in Rank6_8.

3.7 Conclusion

The aim of this essay was to investigate the impact of the improvement of an international highway connecting two capital cities in Southern Africa to a trading seaport on firm outcomes in a landlocked country. The essay investigated how firms located within different distance bins from the improved highway perform in comparison to firms located far away. The essay also investigated how changes in accessibility to Port Nacala induced by the improved highway affects outcomes of firms with increased accessibility to the seaport of interest in comparison to firms with low accessibility to the seaport of interest. I addressed potential endogeneity problems by treating the districts the highway passes through as exogenous to treatment since the intention of the international highway was to connect the two big cities to a specific port of trade at the shortest driving distance. The district-referenced panel data for the period 2013 to 2017 allowed for the use of firm fixed effects in a DID estimation strategy to control for unobserved time-invariant factors. I further added other covariates in the estimation to control for observed factors.

The essay presented results from three estimation models run on two samples, one excluding Lusaka, the endogenously chosen city. The results show that after the road is improved, firms close to the improved road experienced growth in fuel and lubricant costs that is 414 per cent higher relative to firms far away. Further, the impact of the road on many of the firm outcomes are inconclusive. Results are statistically insignificant while the coefficients are large, making it impossible to definitively conclude that the road had no impact. The coefficients under the various empirical models indicate that the improved highway produces negative differential changes on sales, labour costs and assets of firms close to the improved road or with increased accessibility to Port Nacala. These results suggest increased competition affects locations nearer the road as firms shrink in terms of sales, labour costs and assets.

The findings in this essay have important policy implications. Existing studies in the literature recommend investments in road transport infrastructure to promote increased economic growth. These studies report that part of the reason African countries lag behind those elsewhere in terms of development is their insufficient and poor level of infrastructure that impedes economic activity.[43] This essay informs policy makers and researchers as to how investments in a highway infrastructure impact economic activity measured using firm panel data in a landlocked country in Africa. The study shows that the use of road infrastructure investment policy as a tool to accelerate economic growth at a microeconomic level in the short run has inconclusive impacts on most firm outcomes. This suggests it is not easy to observe the economic impacts of improved roads on firm economic activity across many sectors. However, in the short run, firms experience a reduction in the price of transportation services and resort to optimising operations by utilising more transport services. Therefore, this policy favours the promotion of regional integration through increased accessibility to wider markets. Lastly, this essay provides short-run empirical evidence of the impact of upgraded highways in Africa. The literature is still not clear on the long-term impacts on firm activity of upgraded highways in the region or what the impacts of new highways are in Africa.

[43] See Limao and Venables, 2001.

3.A Appendix: Figures

Figure 3.A.1: Spatial Distribution of Firms

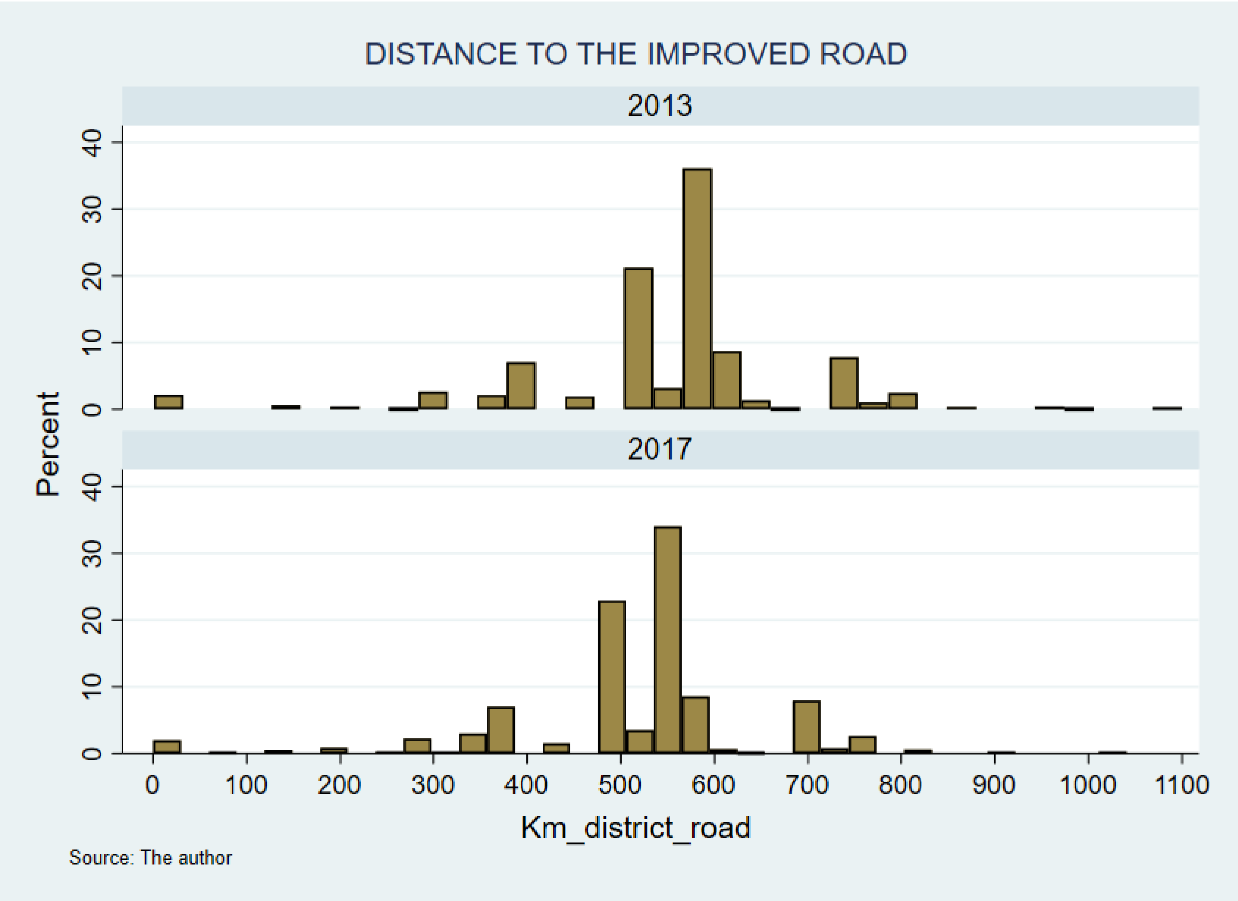
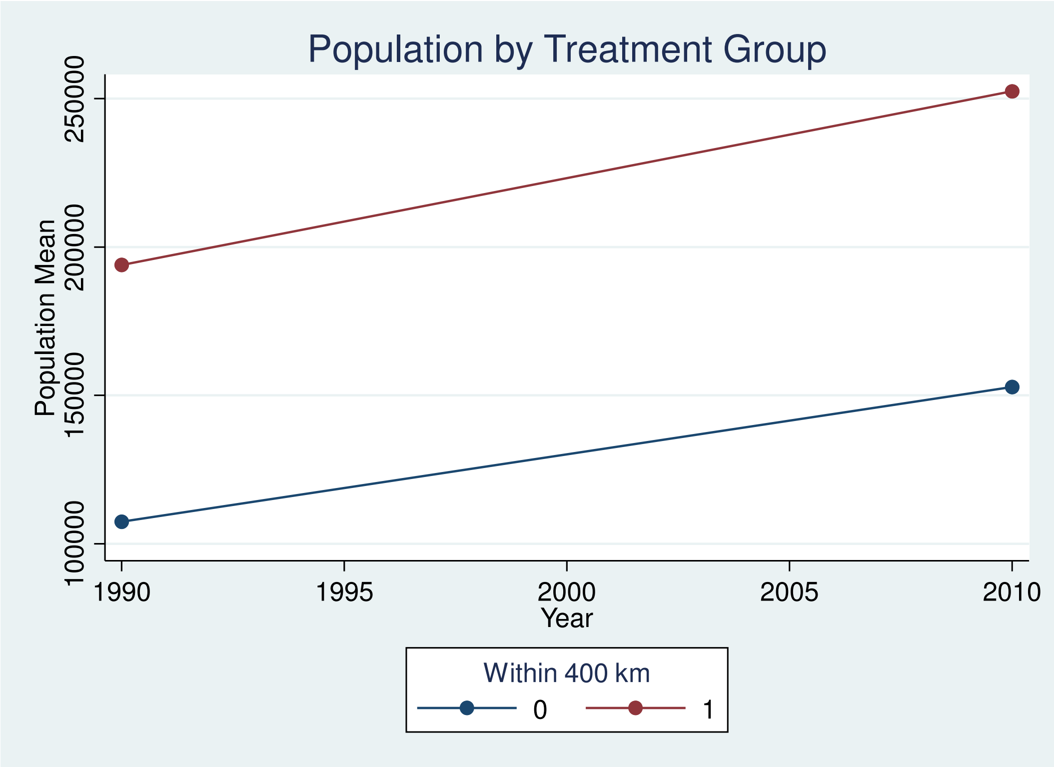


Figure 3.A.2: Map of Nacala Road Corridor Project and Districts by Treatment

Figure 3.A.3: Map of Zambia and the Eight Seaports

Figure 3.A.4: Population Trends by treatment group



3.B Appendix: Tables

Table 3.B.1: Summary Statistics by Sample and Treatment

District	Ave. Driving Km to improved road or District on Road 2013	2013 No. of firms 2013	2013 Ave. ihs_Population 2013	2017 No. of firms 2017	2017 Ave. ihs_Population 2017	Change (2017-13) No of firms	Change (2017-13) Ave. ihs_Population
Full Sample)							
Model 1							
Treatment Group (d400km=1)	234	1,285	8.165	1,072	8.326	-213	0.16***
(S.E)	(1.108)		(0.016)		(0.018)		
Control Group	550	667	6.663	488	6.789	-179	0.125***
(S.E)	(3.261)		(0.022)		(0.026)		
Model 2							
Treatment Group (Km_distG1)	34	17	6.569	4	6.577	-13	0.008
(S.E)	(17.936)		(0.141)		(0.225)		
Treatment Group (Km_distG2)	237	1,268	8.186	1,068	8.332	-200	0.146***
(S.E)	(0.881)		(0.015)		(0.018)		
Treatment Group (Km_distG3)	524	584	6.781	427	6.911	-157	0.13***
(S.E)	(1.644)		(0.02)		(0.024)		
Control Group (601km plus)	755	83	5.836	61	5.935	-22	0.099*
(S.E)	(9.295)		(0.04)		(0.044)		
Model 3							
Treatment Group (Rank2_3)	330	1,812	7.780	1,445	7.989	-367	0.209***
(S.E)	(3.617)		(0.019)		(0.021)		
Treatment Group 2 (Rank4_5)	426	103	6.063	75	6.126	-28	0.06*
(S.E)	(12.169)		(0.023)		(0.023)		
Control Group (Rank6_8)	704	37	5.815	40	5.878	3	0.06
(S.E)	(16.406)		(0.035)		(0.053)		
SAMPLE (Excludes Lusaka)							
Model 1							
Treatment Group (d400km=1)	299	94	6.177	73	6.112	-21	-0.065
(S.E)	(13.520)		(0.034)		(0.03)		
Control Group	550	667	6.663	488	6.789	-179	0.125***
(S.E)	(3.258)		(0.022)		(0.026)		
Model 2							
Treatment Group (Km_distG1)	33	17	6.569	4	6.577	-13	0.01
(S.E)	(17.981)		(0.141)		(0.225)		
Treatment Group (Km_distG2)	358	77	6.090	69	6.085	-8	-0.006
(S.E)	(2.616)		(0.015)		(0.026)		
Treatment Group (Km_distG3)	524	584	6.781	427	6.911	-157	0.13***
(S.E)	(1.644)		(0.02)		(0.024)		
Control Group (601km plus)	734	83	5.836	61	5.935	-22	0.099*
(S.E)	(9.307)		(0.04)		(0.044)		
Model 3							
Treatment Group (Rank2_3)	524	621	6.740	446	6.871	-175	0.131***
(S.E)	(4.391)		(0.021)		(0.025)		
Treatment Group 2 (Rank4_5)	426	103	6.063	75	6.126	-28	0.06*
(S.E)	(12.169)		(0.023)		(0.023)		
Control Group (Rank6_8)	703	37	5.815	40	5.878	3	0.06
(S.E)	(16.402)		(0.035)		(0.053)		

* denotes significance at 90%, ** at 95% and *** at 99%

Table 3.B.2: Distance of districts from the seaports

Location	Ranking of Nacala	Durban	Port Elizabeth	East London	Walvis Bay	Nacala	Beira	Dar es salam	Maputo
Rank2_3									
Chipata	2	2,431	2,916	2,834	2,640	1,143	973	1,565	1,820
Isoka	2	3,038	3,522	3,441	2,853	1,587	1,595	1,033	2,539
Lundazi	2	2,621	3,106	3,024	2,821	1,240	1,149	1,332	2,093
Mfuwe	2	2,540	3,025	2,944	2,740	1,236	1,068	1,572	2,013
Nakonde	2	3,148	3,633	3,552	2,964	1,665	1,706	919	2,599
Nyimba	2	2,510	2,995	2,913	2,476	1,350	1,052	1,771	1,899
Petauke	2	2,482	2,967	2,885	2,474	1,322	1,024	1,743	1,871
Rufunsa	2	2,205	2,690	2,609	2,230	1,542	1,126	1,862	1,656
Sinda	2	2,426	2,911	2,829	2,516	1,266	968	1,687	1,815
Chambishi	3	2,537	3,022	2,940	2,353	1,971	1,457	1,882	1,987
Chibombo	3	2,228	2,713	2,632	2,022	1,798	1,150	1,854	1,679
chililabombwe	3	2,584	3,068	2,987	2,486	2,006	1,504	1,929	2,034
Chingola	3	2,561	3,046	2,964	2,471	1,984	1,482	1,906	2,012
Chirundu	3	1,995	2,480	2,398	2,055	1,938	916	2,075	1,446
Chisamba	3	2,199	2,684	2,602	2,058	1,737	1,119	1,913	1,650
Chongwe	3	2,108	2,592	2,511	2,112	1,658	1,028	1,979	1,558
Kalulushi	3	2,523	3,007	2,926	2,422	1,946	1,443	1,868	1,973
Kasama	3	2,984	3,469	3,388	2,800	1,660	1,493	1,160	2,435
Kitwe	3	2,507	2,992	2,911	2,323	1,931	1,428	1,853	1,958
Lusaka	3	2,133	2,633	2,536	2,078	1,889	1,071	1,937	1,583
Mansa	3	2,894	3,379	3,297	2,709	1,840	1,815	1,500	2,345
Masaiti	3	2,414	2,899	2,818	2,230	1,831	1,334	1,754	1,865
Mbala	3	3,153	3,638	3,557	2,969	1,856	1,662	1,112	2,604
Mkushi	3	2,429	2,913	2,832	2,244	1,724	1,350	1,646	1,880
Mpika	3	2,775	3,260	3,179	2,591	1,451	1,284	1,294	2,226
Mporokoso	3	3,161	3,646	3,565	2,977	1,950	1,670	1,297	2,612
Mpulungu	3	3,188	3,673	3,591	3,003	1,864	1,696	1,147	2,639
Mufulira	3	2,515	3,000	2,919	2,331	1,938	1,436	1,860	1,966
Nchelenge	3	3,132	3,617	3,535	2,905	2,033	1,866	1,527	2,583
Ndola	3	2,446	2,931	2,850	2,262	1,869	1,367	1,791	1,897
Serenje	3	2,537	3,022	2,941	2,353	1,615	1,458	1,538	1,988
Shiwangandu	3	2,869	3,354	3,273	2,686	1,582	1,504	1,218	2,320
Siavonga	3	2,018	2,503	2,422	2,089	1,961	939	2,125	1,469
Solwezi	3	2,594	2,995	2,921	2,297	2,159	1,657	2,082	2,188
Rank4_5									
Kabwe	4	2,271	2,756	2,675	2,087	1,841	1,191	1,799	1,722
Kapiri-Mposhi	4	2,330	2,815	2,733	2,146	1,817	1,250	1,739	1,780
Luanshya	4	2,447	2,932	2,850	2,263	1,921	1,367	1,843	1,897
Mpongwe	4	2,427	2,912	2,830	2,242	1,901	1,348	1,823	1,878
Choma	5	2,000	2,410	2,327	1,787	2,204	1,182	2,217	1,663
Chilanga	5	2,112	2,597	2,516	2,048	2,055	1,033	1,957	1,563
Kafue	5	2,086	2,571	2,489	2,022	2,029	1,007	1,982	1,537
kasempa	5	2,414	2,824	2,741	2,117	2,252	1,668	2,177	2,041
Maamba	5	2,059	2,468	2,385	1,846	2,223	1,200	2,274	1,721
Mazabuka	5	2,158	2,568	2,485	1,945	2,082	1,060	2,061	1,821
Mwinilunga	5	2,631	3,041	2,958	2,209	2,430	1,928	2,353	2,356
Senanga	5	1718	1996	1934	1193	1871	1287	2017	1456
Rank6_8									
Kalomo	6	1,934	2,344	2,261	1,721	2,270	1,247	2,283	1,597
Monze	6	2,096	2,506	2,423	1,884	2,143	1,120	2,121	1,759
Mumbwa	6	2,185	2,595	2,512	1,882	2,279	1,257	2,051	1,787
Itezhi-Tezhi	7	2,022	2,432	2,349	1,719	2,390	1,368	2,166	1,783
Kabompo	7	2,387	2,797	2,714	1,965	2,734	1,712	2,351	2,112
Kazungula	7	1,745	2,155	2,072	1,534	2,344	1,318	2,473	1,469
Livingstone	7	1,806	2,216	2,133	1,593	2,282	1,256	2,411	1,475
Sesheke	7	1,879	2,289	2,206	1,402	2,476	1,450	2,485	1,604
Zimba	7	1,889	2,299	2,216	1,676	2,320	1,298	2,333	1,552
Mongu	8	2,191	2,601	2,518	1,702	2,662	1,639	2,433	1,952
Shangombo	8	2,187	2,597	2,514	1,698	2,784	1,758	2,699	1,948

3.C Appendix: Solutions of Theoretical Model

Sector Specific Value Added Production Function

I consider a Cobb-Douglas production function as the specification allows for the control of shares of inputs in value added output.

$$Y_{tk} = AL_{tk}^{\beta}K_{tk}^{\alpha}T_{tk}^{1-\alpha-\beta} \tag{3.C.1}$$

Where:

Y_{tk} = Output of firms in sector k at time t, A = Total factor productivity, L_{tk}= Labour in firms in sector k at time t. K_{tk} = Capital of firms in sector k at time t and T_{kt} = Other inputs either transportation services or raw materials utilised at time t in sector k. The output elasticities α,β, $1-\alpha-\beta$ are of capital, labour and other inputs respectively, and $1-\alpha-\beta < 1$ but greater than zero.

In the short run, capital is fixed and firm reactions to a road improvement project are through adjustments of labour and other raw materials in production. Let $A^{'} = \bar{K}_{tk}^{\alpha}$ which gives the reduced form production function of two inputs.

$$Y_{tk} = A^{'}L_{tk}^{\beta}T_{tk}^{\theta} \tag{3.C.2}$$

Where θ =$1-\beta$ and β,$\theta > 0$ with $\beta + \theta < 1$.

Firm's Problem when $\sigma = 1$

Firms seek to solve the following problem.

$$max_{\{L,T\}>0}\, P_tA^{'}L_{tk}^{\beta}T_{tk}^{\theta} - wL_{tk} - \gamma T_{tk} \tag{3.C.3}$$

Markets are competitive so the firm takes the price of transportation services or other raw materials γ and that of labour w.

FOCs

Firm's profit maximizing conditions in equilibrium are that the marginal revenue product of labour and transportation services or raw materials equal their prices.

$$\frac{\partial Y_{tk}}{\partial L_{tk}} = P_t\beta A^{'}L_{tk}^{\beta-1}T_{tk}^{\theta} = w \tag{3.C.4}$$

$$\frac{\partial Y_{tk}}{\partial T_{tk}} = P_t\theta A^{'}L_{tk}^{\beta}T_{tk}^{\theta-1} = \gamma \tag{3.C.5}$$

At the optimal, the quantity of transportation services or raw materials demanded is $T*$ and this is obtained using FOCs.

$$T* = \frac{w}{\gamma}\frac{\theta}{\beta}L_{tk} \tag{3.C.6}$$

To obtain the input demand function for Labour substitute equation (3.C.6) into equation (3.C.4)

$$L^*_{tk}\left(P_t, w, \gamma\right) = \left(P_t A^{'}\right)^{\frac{1}{1-\beta-\theta}} \left(\frac{\beta}{w}\right)^{\frac{1-\theta}{1-\beta-\theta}} \left(\frac{\theta}{\gamma}\right)^{\frac{\theta}{1-\beta-\theta}} \tag{3.C.7}$$

To derive the input demand function for transport services or raw materials substitute equation (3.C.7) into equation (3.C.6)

$$T^*_{tk}\left(P_t, w, \gamma\right) = \left(P_t A^{'}\right)^{\frac{1}{1-\beta-\theta}} \left(\frac{\beta}{w}\right)^{\frac{\beta}{1-\beta-\theta}} \left(\frac{\theta}{\gamma}\right)^{\frac{\theta}{1-\beta-\theta}} \tag{3.C.8}$$

To derive the firm output supply function substitute the two input demand function equations $(3.C.7)$ and $(3.C.8)$ into the production function equation $(3.C.2)$.

$$Y^*_{tk}\left(P_t, w, \gamma\right) = P_t^{\frac{\beta+\theta}{1-\beta-\theta}} A^{'\frac{1}{1-\beta-\theta}} \left(\frac{\beta}{w}\right)^{\frac{\beta}{1-\beta-\theta}} \left(\frac{\theta}{\gamma}\right)^{\frac{\theta}{1-\beta-\theta}} \tag{3.C.9}$$

Proposition 1

A fall in γ the price of transportation services has a positive effect on firm output, the demand for labour and other raw materials.

Constant Elasticity of Substitution Production Function

To provide an understanding of the effect of improved roads on firm substitutability of factor inputs, I employ a CES production function with hicks-neutral technological change. The CES production function shows how easy firms can substitute labour or other raw materials for transportation services when the price of transportation services falls.

$$Y_{tk} = A^{'}[\beta L^{\rho}_{tk} + \theta T^{\rho}_{tk}]^{\frac{\psi}{\rho}}$$

$$Y_{tk} = A^{'}[\beta L_{tk}^{\frac{\sigma-1}{\sigma}} + \theta T_{tk}^{\frac{\sigma-1}{\sigma}}]^{\frac{\psi\sigma}{\sigma-1}} \tag{3.C.10}$$

Where ψ is the returns to scale parameter. $\psi = 1$ is constant returns to scale, $\psi > 0$ is increasing returns to scale and $\psi < 0$ is decreasing returns to scale.

$\sigma =$ elasticity of substitution of inputs and lies between $0 < \sigma < \infty$. $\rho =$ extent of substitution between labour and transportation services lies between$-\infty < \rho \leq 1$. The relationship of ρ to σ is ; $\rho = \dfrac{\sigma - 1}{\sigma}$and $\sigma = \dfrac{1}{1-\rho}$. When $\rho \to 1, \sigma \to \infty$. This indicates perfect substitutability and the production technology becomes linear. If $\rho \to -\infty, \sigma \to 0$ indicating no substitutability and the production technology becomes L-shaped, Leontief production function. If $\rho \to 0, \sigma \to 1$ indicating the production technology becomes convex, the Cobb-Douglas production function.

Theoretically, the elasticity of substitution measures the percentage change in the factor ratio due to a percentage change of the ratio of marginal products of inputs along a given isoquant curve (Helm, 1987). The elasticity of substitution is constant but is not always equal to unity. It is given by the formula below.

$$\sigma = \frac{\partial \left(\frac{T}{L}\right) \setminus \frac{T}{L}}{\partial \left(\frac{MPL}{MPT}\right) \setminus \frac{MPL}{MPT}} \tag{3.C.11}$$

Firm's problem when $\sigma \neq 1$

Firms seek to generate maximum profit by combining factors of production at minimal cost. With a fixed level of output given, profit maximization implies minimizing costs to produce the chosen firm output level. The decision to select the least cost mixture of factor inputs involves calculating the marginal rate of technical substitution and the elasticity of substitution of factor inputs.

$$max_{\{L,T\}>0} P_t A^{'} [\beta L_{tk}^{\rho} + \theta T_{tk}^{\rho}]^{\frac{\psi}{\rho}} - wL_{tk} - \gamma T_{tk} \tag{3.C.12}$$

FOCs

$$\frac{\partial Y_{tk}}{\partial L_{tk}} = P_t A^{'} \psi \left(\beta L_{tk}^{\rho} + \theta T_{tk}^{\rho}\right)^{\frac{\psi-\rho}{\rho}} \beta L_{tk}^{\rho-1} = w \tag{3.C.13}$$

$$\frac{\partial Y_{tk}}{\partial T_{tk}} = P_t A^{'} \psi \left(\beta L_{tk}^{\rho} + \theta T_{tk}^{\rho}\right)^{\frac{\psi-\rho}{\rho}} \theta T_{tk}^{\rho-1} = \gamma \tag{3.C.14}$$

The marginal products of Labour and other raw materials are decreasing in their relative abundance.

$$\frac{MPL}{MPK} \Rightarrow \left(\frac{L}{T}\right)^{\rho-1} = \frac{w}{\gamma}\frac{\theta}{\beta}$$

Recall $\rho = \dfrac{\sigma - 1}{\sigma}$

$$\frac{MPL}{MPK} \Rightarrow \left(\frac{L}{T}\right)^{\frac{1}{\sigma}} = \frac{w}{\gamma}\frac{\theta}{\beta}$$

$$\frac{T}{L} = Y = f\left(A, L, K, T\right)\left(\frac{\gamma}{w}\frac{\beta}{\theta}\right)^{-\sigma} \tag{3.C.15}$$

As more Labour is substituted for transportation services or raw materials, the ratio of the quantity of labour to transportation services or raw materials falls. The Marginal rate of technical substitution (MRTS) of transportation services to labour also falls.

The optimal demand for transportation services T_{tk} obtained using the FOCs.

$$T^{*}_{tk} = \left(\frac{\gamma}{\theta}\right)^{\frac{1}{\rho-1}}\left(\frac{w}{\beta}\right)^{-\frac{1}{\rho-1}} L_{tk} \tag{3.C.16}$$

Raise both sides of the above equation by ρ, then substitute equation (3.C.16) into equation (2.C.10) to get the constant-output input demand for labour below. In terms of σ you get:

$$L^{*}_{tk}\left(Y_t, \gamma, w\right) = \frac{\left(\frac{Y_{tk}}{A'}\right)^{\frac{1}{\psi}}\left(\frac{w}{\beta}\right)^{-\sigma}}{\left[\beta\left(\frac{w}{\beta}\right)^{1-\sigma} + \theta\left(\frac{\gamma}{\theta}\right)^{1-\sigma}\right]^{\frac{\sigma}{\sigma-1}}} \tag{3.C.17}$$

The constant-output input demand for transportation services or other raw materials is as below.

$$T^{*}_{tk}\left(Y_t, \gamma, w\right) = \frac{\left(\frac{Y_{tk}}{A'}\right)^{\frac{1}{\psi}}\left(\frac{\gamma}{\theta}\right)^{-\sigma}}{\left[\beta\left(\frac{w}{\beta}\right)^{1-\sigma} + \theta\left(\frac{\gamma}{\theta}\right)^{1-\sigma}\right]^{\frac{\sigma}{\sigma-1}}} \tag{3.C.18}$$

To get the output supply function below substitute the constant-out input demand functions equations (3.C.17) and (3.C.18) into FOC equation (3.C.13)

$$Y^{*}_{tk}\left(P_t, w, \gamma\right) = \left(P_t A\psi\right)^{-\frac{\psi}{\psi-1}}\left[\beta\left(\frac{w}{\beta}\right)^{1-\sigma} + \theta\left(\frac{\gamma}{\theta}\right)^{1-\sigma}\right]^{-\frac{\psi}{(\sigma-1)(\psi-1)}} \tag{3.C.19}$$

To get the unconditional input demand function for labour, substitute equation (3.C.19) into equation (3.C.17)

$$L^{*}_{tk}\left(P_t, w, \gamma\right) = \left(P_t A\psi\right)^{-\frac{1}{\psi-1}}\left(\frac{w}{\beta}\right)^{-\sigma}\left[\beta\left(\frac{w}{\beta}\right)^{1-\sigma} + \theta\left(\frac{\gamma}{\theta}\right)^{1-\sigma}\right]^{\frac{\sigma-1-\psi\sigma}{(\sigma-1)(\psi-1)}} \tag{3.C.20}$$

To get the unconditional demand function for other raw materials and transportation services substitute equation (3.C.19) in to equation (3.C.18)

$$Y = f(A, L, K, T)$$

$$T^{*}_{tk}\left(P_t, w, \gamma\right) = \left(P_t A\psi\right)^{-\frac{1}{\psi-1}} \left(\frac{\gamma}{\theta}\right)^{-\sigma} \left[\beta\left(\frac{w}{\beta}\right)^{1-\sigma} + \theta\left(\frac{\gamma}{\theta}\right)^{1-\sigma}\right]^{\frac{\sigma-1-\psi\sigma}{(\sigma-1)(\psi-1)}} \tag{3.C.21}$$

To study the substitution effect of a fall in prices of transportation services γ, on Labour, other raw materials and output we take the derivative of the unconditional input demand functions $L^{*}_{tk}\left(P_t, \gamma, w\right)$, $T^{*}_{tk}\left(P_t, \gamma, w\right)$ and $Y_{tk}\left(P_t, \gamma, w\right)$with respect to γ.

$$\begin{aligned}\frac{\partial L^{*}_{tk}\left(P_t, \gamma, w\right)}{\partial\left(\gamma\right)} &= \frac{\sigma-1-\psi\sigma}{\psi\sigma-\sigma-\psi+1}\left(1-\sigma\right)\left(P_t A'\psi\right)^{-\frac{1}{\psi-1}} \\ &\left(\frac{w}{\beta}\right)^{-\sigma}\left(\frac{\sigma}{\theta}\right)^{-\sigma}\left[\beta\left(\frac{w}{\beta}\right)^{1-\sigma} + \theta\left(\frac{\sigma}{\theta}\right)^{1-\sigma}\right]^{-\frac{\psi}{(\sigma-1)(\psi-1)}}\end{aligned} \tag{3.C.22}$$

$$\begin{aligned}\frac{\partial T^{*}_{tk}\left(P_t, w, \gamma\right)}{\partial\gamma} =& \left(P_t A'\psi\right)^{-\frac{1}{\psi-1}}\left(P_t A'\psi\right)^{-\frac{1}{\psi-1}} \\ &\left\{\begin{array}{l} -\sigma\left(\frac{\gamma}{\theta}\right)^{\sigma-1}\frac{1}{\theta}\left[\beta\left(\frac{w}{\beta}\right)^{1-\sigma} + \theta\left(\frac{\gamma}{\theta}\right)^{1-\sigma}\right]^{\frac{\sigma-1-\psi\sigma}{(\sigma-1)(\psi-1)}} + (1-\sigma) \\ \frac{\sigma-1-\psi\sigma}{(\sigma-1)(\psi-1)}^{-2\sigma}\left[\beta\left(\frac{w}{\beta}\right)^{1-\sigma} + \theta\left(\frac{\gamma}{\theta}\right)^{1-\sigma}\right]^{\frac{2(\sigma-1)-\psi(2\sigma-1)}{(\sigma-1)(\psi-1)}} \end{array}\right\}\end{aligned} \tag{3.C.23}$$

$$\begin{aligned}\frac{\partial Y^{*}_{tk}\left(P_t, \gamma, w\right)}{\partial\left(\gamma\right)} &= -\frac{\psi\left(1-\sigma\right)}{\left(\sigma-1\right)\left(\psi-1\right)}\left(P_t A'\psi\right)^{-\frac{\psi}{\psi-1}}\left(\frac{\gamma}{\theta}\right)^{-\sigma} \\ &\left[\beta\left(\frac{w}{\beta}\right)^{1-\sigma} + \theta\left(\frac{\sigma}{\theta}\right)^{1-\sigma}\right]^{\frac{\psi\sigma+\sigma-1}{\psi\sigma-\sigma-\psi+1}}\end{aligned} \tag{3.C.24}$$

Proposition 2

An increase in the price of transportation services γ increases the utilisation of labour in production. However, the extent to which firms substitute labour for transport services or other raw materials depends on σ the constant elasticity of substitution. An increase in the price of transportation services decreases the utilisation of transportation services and other raw materials in production and thus reduces value added output.

4 Improved Highways, Transportation Costs and Trade: Firm Panel Data Evidence from Zambia

4.1 Introduction

Transport infrastructure is widely recognised as crucial for economic development (Gibbons et al., 2019; Gurara et al., 2018 and Sequeira, 2011). Good quality transport infrastructure reduces transportation costs and promotes economic growth (Bhattacharya et al., 2016). The literature comparing economic infrastructure in developing economies to that of advanced economies reveals that developing economies lag behind in terms of quality, quantity and accessibility of infrastructure (Gurara et al., 2018). Comparisons made among developing economies show that African countries lag behind developing countries elsewhere on every infrastructure measure, and this contributes to its low rate of economic growth (Yepes, Pierce and Foster 2008).[1] Thus, in the last two decades, countries in Africa have been investing considerable resources in economic infrastructure, especially transport infrastructure, as a way to stimulate trade and increase economic growth (Jedwab and Storeygard, 2017; the Economist, 2015).[2] However, despite these large capital investments in transport infrastructure, there are few studies in Africa analysing the actual economic impacts of such investments.

This essay provides empirical evidence of the impact of upgrading a highway leading to a seaport on the tradable sector in a landlocked country in Southern Africa. To the best of my knowledge, no study currently explores the impacts of upgraded highways on trade at a microeconomic level in Africa. Other factors motivating this research include the following. There are few studies on this topic in developing countries, especially Africa, because of data availability challenges. Further, Africa trades 'too little' both with itself and with the rest of the World (Limao and Venables, 2001). The reasons behind Africa's poor trade performance include

[1] Africa has a high level of infrastructure inadequacy mainly in energy, transport, telecommunications, water and sanitation, and irrigation which have contributed to the continent's low economic performance, low intra-African trade (11.9% of total trade in 2017) and weak integration into global value chains (Juma, 2015).

[2] According to Ashiagbor et al (2018) from the European Investment Bank, infrastructure investment in Africa stood at approximately USD 63 billion or 3.5 per cent of the continent's GDP in 2016. They report that infrastructure financing from national governments accounts for 42% of the total investment. African governments, donors and other multilateral development organisations together provide over 85% of financing and this increases to 96% when financing from China is included (Ashiagbor et al.,2018).

protectionist trade policies (Collier, 1995; Collier and Gunning, 1999) and high transport costs due to poor infrastructure (Amjadi and Yeats, 1995, Buys et 2006).[3] Poor infrastructure holds back intra-regional trade activity and undermines firm competitiveness (Ashiagbor et al., 2018). According to the World Bank, isolation from regional and international markets contributes significantly to poverty in many countries in Sub-Saharan Africa (Buys et al., 2006). Further, with the progressive dismantling of tariffs in the 2000s, some researchers and policymakers believe transport costs have replaced tariffs as the main barrier to trade in some parts of the developing world (Sequeira, 2011). Thus, to increase the participation of African countries in trade, it is important to know the extent to which improved highways impact a country's participation in international trade. Based on this, the research question addressed in this study is, "what are the effects of reduced transport costs on firms engaged in international trade in developing economies?". To answer this question, I apply a triple difference estimation method on firm panel data from Zambia for the period 2013 to 2017. The year 2013 is the pre-treatment period, and 2017 is the post-treatment period.[4]

Disentangling the empirical relationship between infrastructure development and economic outcomes is essential to the design of transport policy. Furthermore, given the importance of roads in the movement of people and goods, the correct identification of the impact of road investments is important for economic policy as a whole (Sanchis-Guarner, 2012). In this essay, the identification of impacts is based on two facts. First, the improved highway affects firms in the tradable sector located close to the road differently than it affects those in more distant locations. This difference allows for the use of a difference-in-differences (DID) estimation strategy with time and location fixed effects. The validity of the DID methodology rests on the assumption of common trends in outcome variables in the treatment and control groups in the absence of treatment. Data on the outcome variables is required for more than one pre-treatment period to check for common trends, but this is not available. Thus, taking advantage of the fact that the impacts of road development projects take a few years to appear, time series data covering the period of highway improvement works is used to check for common trends. The time series data shows parallel trends in almost all but two outcome variables – labour costs and freight transport costs in the treatment and control groups. To deal with this problem, I use a triple difference estimation strategy that uses an additional within-state control group (the non-tradable sector) to remove location-specific time-varying confounding factors.[5] Second, as in the previous chapter, I treat the highway improvement project as a natural experiment in districts laying between the two endogenously chosen capital cities, Lusaka and Lilongwe.[6] This

[3]Remoteness and poor communication tied with poor transport infrastructures impede trade (Limao and Venables, 2001).

[4]The highway improvement project began in December 2013 and ended in August 2017.

[5]Olden and Moen (2020), report that the difference between two difference-in-differences estimators does not require two parallel trend assumptions to have a causal interpretation, but only one. Their reasoning is that the difference between two biased difference-in-differences estimators will be unbiased (as long as the bias is the same in both estimators) because the bias is differenced out when the triple difference is computed.

[6]Chandra and Thompson (2000), Michaels (2008) and Datta (2012), treat the construction of new roads as a natural experiment to intermediate areas laying between two endogenously chosen locations. The road project in this paper involves the improvement of an existing highway which too is prone to the

is because the Nacala road corridor was built and improved to promote regional integration and economic growth in the Southern Africa Development Community region by connecting the two capital cities to the Port of Nacala at the shortest distance. The placement and improvement of the highway is exogenous to areas laying in-between the two capital cities because these areas were included in the treatment because they are located along the shortest driving route between the two cities and not because of any factors related to the outcome variables at baseline. Due to reverse causality concerns, estimates of the impact of the highway on firm data that includes Lusaka city are likely to be upward biased. Thus, I run estimations on two samples, the sample excluding Lusaka city (the cleanest sample under identification) to estimate the impact of the highway on firms outside the capital city (the peripheral areas) and the sample including Lusaka city to study influences of the centre.

Results show that the improved international highway brings gains in trade of increased exportable output and reduced inventory costs. Firms in the tradable sector compared to those in the non-tradable sector experience a large increase in gross profit, sales, fuel costs and labour costs after the highway is improved. The semi-elasticity of gross profit and that of sales with respect to the improved highway are the highest at 369.93 and 188.81 respectively.[7] The semi-elasticity of labour costs with respect to the improved highway is 136.69 higher in the tradable sector compared to the non-tradable sector. However, the improved highway has inconclusive effects on freight transport costs. All effects discussed above are observed in the peripheral areas.

This paper makes three major contributions to the literature. It is the first to investigate effects of reduced transport costs arising from an improved international highway on firms engaged in international trade in a country in Southern Africa. Existing studies in Africa are mostly cross-country level studies that show that high transport costs arising from poor transport infrastructure impede trade (Amjadi and Yeats, 1995; Buys et al., 2006; Limao and Venables, 2001).[8] This paper takes a different approach by conducting a within country microeconomic study on trade using firm-level panel data. The paper analyses the impact of reduced transport costs arising from the upgrade of an international highway on trade unlike the opposite of analysing the impact of high transport costs arising from poor road infrastructure done in previous studies. The second contribution of this paper is that, unlike other studies, it estimates the impact of the improved highway on actual firm trade costs of transportation using freight transport and fuel costs data. Studies by Donaldson (2018) infer transportation trade

endogeneity concern of reverse causality of road projects. According to Van de Walle (2009), roads are built or rehabilitated in certain locations and not in others, for reasons that have to do with the attributes of those locations. This highway could have been improved because of high economic activities in the two capital cities that led to a greater demand to improve it.

[7] To give the semi-elasticities a per centage interpretation, multiply them by 100. Thus, semi-elasticities of 369.93 and 188.81 mean that after the highway is improved, firms in the tradable sector near the highway experience a growth in gross profit and sales that is 36,993 per cent higher and 18,881 per cent higher respectively, compared to firms in the non-tradable sector.

[8] Limao and Venables (2001) further run simulations to estimate effects on trade that would result if countries with poor transport infrastructure in Africa improved their level of infrastructure by moving from the 75th to the top 25th per centile of countries in Africa with better infrastructure.

costs from differences in location prices of homogeneous products that can only be produced in one location while, Limao and Venables (2001) estimate trade transport costs using shipping company quotations (that do not show actual cost expenditures) and cross section data of CIF/FOB ratios that are subject to several measurement errors. Importing countries report the CIF value of imports from partner countries, inclusive of carriage, insurance and freight (CIF) while, exporting countries report their value free on-board (FOB), which measures the cost of the imports and all charges incurred in placing the merchandise aboard a carrier in the exporting port (Limao and Venables, 2001). Quotations are based on shipping an item from Baltimore in the United States (US) to various destinations in Africa, which makes generalisation of these results difficult (Limao and Venables, 2001). The International Monetary Fund CIF/FOB ratios are used, but measurement errors are present as exports are not always accurately reported. For example, some countries report CIF import values lower than the corresponding FOB export value, which would imply (CIF/FOB) negative costs (Limao and Venables 2001). The essay's final contribution is that it uses new firm panel data. Unlike previous studies in the region, the economic data used in this essay is defined at the 2-digit division and the 4-digit class, the finest disaggregation level of economic activity of the International Standard of Industrial Classification (ISIC) Revision 4. Information on industry-level economic activity aids in the assignment of firms to the tradable and non-tradable sectors.

The essay structure is as follows. Section 4.2 reviews the related literature. Section 4.3 considers the data and defines the sectors. Section 4.4 lays out the theoretical framework. Section 4.5 presents the methodology. Section 4.6 discusses the results, and Section 4.7 sets out the concluding remarks.

4.2 Related Literature

The existing literature identifies an adequate supply of infrastructure services as a key ingredient for economic development (Ashiagbor et al., 2018;Gurara et al., 2018). Infrastructure development not only promotes growth and equity but also helps reduce poverty (Bhattacharya et al., 2016). Thus, transport policy for large-scale transport infrastructure investments aims at combining national efficiency with regional equity objectives; the assumption is that falling trade costs promote national growth as well as the diffusion of economic activity to peripheral regions (Faber, 2014). However, despite the important benefits to an economy of falling trade costs, few studies in the literature investigate the impact on trade of transport infrastructure at a microeconomic level.

The few studies analysing the effects on trade of transport infrastructure include the following. Duranton (2015) analyses the effects of major roads on the weight and value of bilateral trade between Colombian cities. Duranton et al. (2014) investigate the effects of new highways on the weight and value of bilateral trade between and within large US cities. Faber (2014) studies the impact of trade-cost reductions arising from China's trunk highway system on the diffusion of industrial and economic activity between the peripheral regions and the large metropolitan centres. Martincus et al. (2017) estimate the impact of new highway infrastructure on exports of firms in Peru. Studies specific to Africa include Amjadi and Yeats (1995) on the impact of transport costs on trade flows in Sub-Saharan Africa, Limao and Venables (2001) examine the impact of poor highways on trade and Buys et al. (2006) estimate the trade growth that would result from the removal of bad highways, difficult border crossings and roadblocks in Sub-Saharan Africa.

The existing literature on this topic in relation to Africa are mainly macroeconomic studies. These point out that road conditions in Africa are poor despite roads being the continent's main mode of transportation, transporting approximately 80 per cent of goods and 90 per cent of passengers (Yepes, Pierce, and Foster 2008; Ashiagbor et al., 2018).[9] Further, inadequate infrastructure limits trade within the continent and internationally. It contributes to the continent's low economic performance, low intra-African trade and weak integration into global value chains (Juma, 2015).[10] Poor transport infrastructure accounts for 40 per cent of predicted transport costs for coastal countries and up to 60 per cent for landlocked countries (Limao and Venables, 2001).[11] Transport costs and other costs of doing business are important determinants of a country's ability to participate fully in the world economy (Limao and Venables,2001). The lower transport costs are as a proportion of the total cost of a good, the more likely, that the goods will be traded internationally (Sachs and Larrain, 1993). Therefore, new or improved highways between cities should directly foster trade with anticipated indirect consequences of higher employment, wages and other desirable outcomes (Duranton, 2015).

General findings from studies on transport infrastructure and trade activity include the following:

1. Poor quality highways increase transport costs and limit trade and economic growth (Limao and Venables, 2001 and Buys et al, 2006).
2. Africa has a high level of infrastructure inadequacy in comparison to other continents. As reported by Jedwab and Storeygard (2017), Africa's road density is less than a third of South Asia's and only a quarter of the network is paved (Vivien and Briceño-Garmendia, 2010), against two thirds in China (World Bank, 2015) and 60 per cent in India (Government of India, 2016).
3. Africa trades too little with itself and internationally (Limao and Venables, 2001).
4. The endogeneity problem regarding the placement of roads complicates the estimation of the impact of transport infrastructure on various economic outcomes (Datta, 2012). Old studies barely controlled for this, and their results are thus highly questionable.

[9] Africa has the least developed transport network compared to other developing countries elsewhere (Jedwab and Storeygard, 2017).

[10] Inadequate infrastructure in energy, transport, telecommunications, water and sanitation, and irrigation.

[11] A 10-per centage point increase in transport costs typically reduces trade volumes by approximately 20 per cent (Limao and Venables, 2001).

5. The empirical studies analysing the nexus between transport infrastructure and various economic outcomes capture road effects use a market-access index, the distance to the road and instrumental variables. The calculation of the market-access index varies across studies though it is mainly based on travel time, the distance between two locations and other economic and geographic factors existing between the two location points (Faber, 2014; Storeygard, 2016; Jedwab and Moradi, 2011). Studies using the distance to the road include Datta (2012) and Ghani et al. (2016).

The work in this essay contributes to the literature on the effects of transport infrastructure on trade in developing countries. The literature on Africa examines the effects of poor-quality roads on trade volumes; however, the effect of improved highways on the performance of firms engaged in international trade is unknown.

4.3 Data and Definition of Variables

4.3.1 Firm and highway Data

Similar to the study in the previous chapter, the firm panel data used in this chapter comes from the Zambia Revenue Authority (ZRA) online tax system. The firm data provides information on the seven dependent variables in the study, which are gross profits, sales, total employment expense, transport costs, fuel and lubricants, cost of sales and closing raw materials inventory. Sales are the gross receipts of the firm from products sold. Gross profit is the difference between sales and the cost of goods sold. The cost of goods sold includes all costs to manufacture the final product, such as labour, utilities and raw materials. The cost of sales includes costs to sell the products, such as advertising, promotions and distribution costs. Total employment expenses are labour costs, which constitute wages of labour directly engaged in production, salaries of labour not directly engaged in production, pension contributions and all other employment-related costs. Closing raw materials inventory is the cost of closing inventory comprising raw materials and work in progress. Transport costs include expenses of freight by land, air or sea and costs pertaining to handling, packaging and warehousing goods, insurance, and seaport and border clearance. Fuel and lubricant costs are expenses on fuel and other lubricants used in the company vehicles.

Other information in the firm database includes import purchase values, firm ID and ISIC name at the 4-digit-class level. The firm data is district referenced; the exact location of firms is not provided for confidentiality reasons. To capture the variation in effects arising from the improved highway in the firm dataset, I calculate the driving distance to the improved highway from each district's central business area where the local authority, post office or market is located (whichever is present in the listed order) using Google Maps.[12] The dependent

[12] Distance matters because it affects transportation costs (Hummels, 1999). Distance and geography are measures of transport costs often used in economic literature and these measures assume that transport costs

variables in this essay are transformed into inverse hyperbolic sine to retain zeros and to make the interpretation of coefficients easier as semi-elasticities.

Information on the highway project comes from the Road Development Agency (RDA) in Zambia. Works on the 360 km Nacala Road Corridor physically commenced in January 2014 and were completed in August 2017. Completed road sections were immediately opened to traffic as works continued, with the first road section opened in 2015.[13]

Further, the district population statistics used as control variables are projected figures from the Central Statistical Office of Zambia.

Information from all the above sources is combined to create an unbalanced panel data set for the years 2013 and 2017. In this dataset, the sample excluding Lusaka has 1,185 observations in forty-one districts, while the sample including Lusaka has 3,206 observations in forty-two districts.[14]

4.3.2 Sample Attrition

Table 4.1 provides descriptive statistics of firms by industry using ISIC Aggregation –A10 classification. The table indicates that there is sample attrition over the 5-year period. In the full sample, the number of firms drops by 410 from 1,808 in 2013 to 1,398 in 2017.[15] In the small sample, the number of firms drops by 213 from 699 in 2013 to 486 in 2017.[16] There is a net change in the number of firms in the full sample of negative 22 per cent while the small sample experiences a net change of negative 30.47 per cent. In both samples, attrition is highest in the wholesale and retail trade and the transport-services sector, accounting for 76.15 per cent and 69.23 per cent of total firm attrition in the full and small sample, respectively.[17] Sample attrition is caused by the failure of some firms to consistently submit their annual financial returns to the ZRA, firms shutting down operations and missing values. Firms once registered with the ZRA do not downgrade to a lower annual-sales band to be exempted from making submissions to ZRA. They only completely stop submitting financial returns if they shut down in operations.

increase with distance, and decrease with adjacency (Nordas and Piermartini, 2004). Improved highway quality also reduces travel time and vehicle operating costs. Both distance and quality of highway infrastructure directly influence travel time which ultimately affects firm transportation costs (Albarran et al., 2013).

[13]Appendix figure 4.A.1 shows the location of the highway and distribution of firms in the districts. The improved part of the Nacala Road Corridor is between point B Luangwa bridge and point C Mwami border.

[14]I lose 5217 out of 6595 observations in the sample excluding Lusaka because of missing values. 53.9 per cent of missing values are on closing raw material inventory, 52 per cent on import purchases and 48 per cent on freight transport costs.

[15]The full sample includes Lusaka city

[16]The net change in sample attrition is the overall change in the number of firms over five years taking into account attriting and entering firms.

[17]In table 4.1, five sectors in the full sample experience a negative change in the number of firms of -545 and four sectors have a positive next change of 135 firms. In the small sample four sectors experience a negative change of -234 and two sectors experience a positive net change of 21 firms.

Table 4.1: Aggregation-A10 Industry Distribution of Firms

ISIC Rev 4, Aggregation A10	Division	Full Sample			Sample Excluding Lusaka		
		No of Firms		Change	No of Firms		Change
		2013	2017	(2017-13)	2013	2017	(2017-13)
Construction	41-43	88	114	26	22	23	1
Financial and insurance activities	64-66	44	56	12	5	5	0
Information and communication	58-63	18	22	4	4	2	-2
Manufacturing, mining and quarrying a	07-38	250	197	-53	112	82	-30
Other service activities	90-99	42	135	93	15	35	20
Professional, scientific, technical	69-82	233	186	-47	88	57	-31
Public administration and defence, ed	85-88	57	32	-25	19	10	-9
Real estate activities	68	21	16	-5	4	4	0
Wholesale and retail trade, transport	45-56	1,055	640	-415	430	268	-162
Total		1,808	1,398	-410	699	486	-213

Note: The net change in sample attrition is the overall change in the number of firms over five years considering attriting and entering firms.

Table 4.2 provides attrition statistics over the 5-year period for both samples. Using firm time series data for the period 2013 to 2017, I identify firms in the database that comply with financial-return submission requirements, those that are inconsistent in their submissions, those that enter the market and those that exit the sample. I define new firm entrants as those with no firm data in the first four periods of the 5 years, while exiting firms are those without data in the last four periods of the 5 years. Inconsistent firms are those that have data in at least two periods in the 5 years, and compliant firms are those with data for all five periods. Table 4.2 shows that the highest rate of attrition is caused by non-submission of financial returns by inconsistent firms; this accounts for 74.86 per cent and 62.39 per cent of the total sample attrition in the full and small sample, respectively. Firms with missing values are the second-largest contributors to attrition, and firms exiting are the lowest contributors in both samples.[18]

[18] I find that 1500 and 570 firms are consistent in financial return submissions over the 5 year period in the big and small sample respectively. However, out of the 1500 consistent firms in the big sample, 594 appear only in 2013, 513 appear only in 2017 with 393 firms appearing in both years. In the small sample, out of the 570 consistent firms, 237 appear only in 2013, 167 appear only in 2017 and 166 firms appear in both years. Consistent firms do not appear in both time periods in the two samples because of missing values. There are also some inconsistent firms that have missing values, 7 firms are in the big sample and 2 firms in the small sample.

Table 4.2: Attrition Statistics

	Full Sample		Sample Excluding Lusaka	
Panel Data (2017-2013)	Unbalanced Panel	Balanced Panel	Unbalanced Panel	Balanced Panel
New Entrants	118		41	
Exiting Firms	49		16	
Inconsistent Firms	1137	9	390	2
Consistent	1500	393	570	166
Total	2804	402	1017	168

Note	
New	No data in the first 4 periods of the 5 years
Exit	No data in the last 4 years of the 5 years
Compliant	Firm has data in all 5 periods
Inconsistent	A firm has data in at least two periods of the 5 years
Inconsistent with missing values	Firm pattern shows it appears in 2017 and 2013, but the firm is seen only in 2013 or 2017.
Consistent with missing values	Firm pattern shows it appears in all 5 years, but the firm is seen only in 2013 or 2017.

	Full Sample	Sample Excluding Lusaka
Sample Attrition (Total no. of firms)	545	234
Exiting Firms	49	16
Inconsistent Firms	408	146
Inconsistent Firms with missing values	7	2
Consistent Firms with missing values	81	70
Exiting Firms %	8.99	6.84
Inconsistent Firms %	74.86	62.39
Inconsistent firms with missing values	1.28	0.85
Consistent firms with missing values %	14.86	29.91

A major concern in this study regarding sample attrition is that if the firms that drop out of the sample because of non-submission of financial returns, missing values or exit are consistently different to those that remain in the sample, then the surviving sample is no longer representative of the initial population.[19] Regression estimates from this surviving sample would be attrition biased and thus, impede the ability to draw valid inferences on econometric analysis (Cheng et al., 2015). Conversely, if attrition is random, then regression results would not be attrition biased, and there would be no need to adjust the data for sample attrition (Baulch and Quisumbing, 2011). I test whether attrition is random in the two samples using the attrition probit test proposed by Fitzgerald, Gottschalk and Moffitt (1998) and the BGLW pooling test proposed by Becketti, Gould, Lillard and Welch (1988). I estimate probit equations for the likelihood of attrition using as covariates all baseline outcome variables and the treatment variable because these are likely to affect attrition. In the attrition probit tests, the treatment variable result is key. If the results indicate it is independently statistically significant, this would mean

[19] Attrition may cause major differences between the surviving sample and the initial sample called attrition bias or between the control and treatment groups referred to as selective attrition bias.

attrition bias may be present and is significantly induced by the improved highway. This would confound estimates of the effects of the improved highway on firm transportation costs and other outcomes.[20] For the BGLW tests, I regress baseline values of the dependent variable on the treatment variables, the attrition variable and the interaction terms of the attrition variable with all the covariates. The BGLW test determines whether the baseline values of the remaining firms in the non-attriting sample differs sufficiently in its model parameters from the firms in the initial sample (sample consisting of both attritors and non-attritors), and in so doing, tests of equality of coefficients are carried out both individually and jointly (Outes-Leon and Dercon, 2009).[21] Both the attrition probit test and the BGLW tests indicate that sample attrition between 2013 and 2017 is largely random, meaning there is no attrition or selective attrition bias affecting the econometric analysis.

Table 4.3 shows attrition probit test results conducted on the small sample where the treatment variable is $d400km$.[22] Panel A gives a Pseudo R-squared of 4.75 per cent. This indicates that baseline variables and the treatment dummy variable d400km explain about 4.75 per cent of sample attrition between 2013 and 2017. Therefore, 95.25 per cent of the attrition in the sample is the unexplained random part. According to Outes-Leon and Dercon (2009), a very low Psuedo R-Squared in the attrition probit models indicates that attrition overwhelmingly remains a random phenomenon.[23] The Z-statistics and p-values in the last two columns of Table 4.3 show that two of the eight variables, namely, employment expense and freight transport cost, are individually statistically different from zero at the 5 per cent and 1 per cent level, respectively. This shows that the higher the employment expenses and freight transport costs, the less likely it is for a firm to attrite. The t-test results give a Chi-squared statistic of 162.81, which shows that all eight variables are jointly statistically different from zero at the 1 per cent level (the P-value is 0.000). This indicates that all these variables jointly significantly predict attrition.

Panel B in Table 4.3 shows the BGLW test results. The F-statistics and p-values indicate that out of the seven dependent variables, the coefficients on freight transport are not equal in the two samples (the surviving firms compared to the sample consisting of both attritors and non-attritors). However, this result is weakly statistically different from zero with an F-statistic of 2.48 and a p-value of 0.0996. I interpret this result as inconclusive because of the large p-value. Therefore, I do not adjust the data for attrition bias given that the attrition probit tests also indicate that all baseline variables and the treatment variable explain non-random attrition as being under 5 per cent. Results in this table fail to reject the null hypothesis of equality of coefficients in the sample before and after the attrition on six dependent variables at the

[20] The treatment variable is the highway driving distance in kilometres.

[21] An F-test of the joint significance of the attrition dummy and the interaction variables is conducted to determine whether covariate coefficients differ between the two samples, the full sample at baseline compared to surviving firms at baseline. (Outes-Leon and Dercon, 2009).

[22] To see attrition test results for the full sample see appendix table 4.B.1 Results in the full sample are not so different especially in the attrition probit tests.

[23] Pseudo R-squared from attrition probits indicate the proportion of attrition that is non-random (Outes-Leon and Dercon, 2009).

baseline.

Table 4.3: Attrition Probit Results Statistics

Panel A: Attrition Probit Test

Probit regression	Number of obs	=	699
	Wald chi2(8)	=	162.81
	Prob > chi2	=	0.0000
Log pseudolikelihood = -450.44033	Pseudo R2	=	0.0475

(Std.Err. adjusted for 34 clusters in Location)

Variables (2013 Values)	Coefficient	Robust Std. Err.	z	P-value
ihs Gross profit	-0.011389	0.0145536	-0.78	0.434
ihs Sales	-0.0175495	0.0327749	-0.54	0.592
ihs Tot. Employment Expense	-0.029767	0.0148732	-2.00	0.045
ihs Freight Transport	-0.0369789	0.0131649	-2.81	0.005
ihs Fuel & Lubricants	0.0057992	0.0131022	0.44	0.658
ihs Cost of Sales	-0.0189317	0.0197076	-0.96	0.337
ihs CLS. Raw Materials	-0.0104832	0.0112957	-0.93	0.353
d400km_Rd	0.029398	0.130853	0.22	0.822
Constant	0.41654	0.0999072	4.17	0.000
t-test				
chi2(9)	162.81			
Prob > chi2	0.0000			

Panel B: BGLW Pooling Tests

Variables (2013 Values)	F test	P-value
ihs Gross profit	1.18	0.3203
ihs Sales	2.09	0.1401
ihs Tot. Employment Expense	0.75	0.481
ihs Freight Transport	2.48	0.0996
ihs Fuel & Lubricants	1.77	0.1866
ihs Cost of Sales	2.15	0.1322
ihs CLS. Raw Materials	0.86	0.2543

4.3.3 Definition of Sectors

4.3.3.1 The Tradable and Non-Tradable Sector

Traditionally, all goods-producing industries are considered as the tradable sector, with the service industry as the non-tradable sector (Mano and Castillo, 2015; and Uy et al., 2013).[24] However, this classification approach has the problem of industries overlapping across the broad, tradable and non-tradable sectors. For instance, services traditionally classified as non-tradable are now increasingly internationally traded (Fieleke, 1995). At the same time, a large number

[24]Industries frequently classified as tradable include agricultural food products, hunting, forestry, fishing, mining, quarrying, manufacturing, textiles and chemicals; while construction, health services and education are classified as the non-tradable sector (Mano and Castillo, 2015; and Lombardo and Ravenna, 2012).

of goods and services are traded internationally in relatively negligible amounts, such that the literature extends the definition of non-tradable goods to include goods in the traded category provided they are traded below a certain threshold (De Gregorio et al., 1994).[25]A good benchmark of tradability should take into account the extent to which goods and services are actually traded in an industry. One limitation in this study is that the database used does not include export values to allow for the calculation of industry export-intensity values. To identify firms in the tradable sector, I use information on the top-10 exports in Zambia in 2018.[26] The top-10 exporting industries constitute twenty industries at the ISIC 4-digit level.[27] This information allows me to map the industries in the firm panel dataset. In 2018, the top-ten exports accounted for 89.1 per cent of the overall value of Zambia's global shipments (Workman, 2019).[28] Further, a trade performance review carried out between 2002 and 2012 shows that Zambia is one of the world's most concentrated exporters, with copper and cobalt accounting for 80 per cent of its formal exports (World Bank, 2014).[29] This gives comfort that the composition of industries in the top-ten remains stable over the study period.

In this essay, the tradable sector is made up of firms in the top-10 exporting industries and the entirety of the mining and manufacturing industries.[30]The definition of tradability includes firms in the manufacturing and mining industries that do not export or import but whose final products compete with imports.[31] Definition of the tradable sector should include importable and exportable industries. An industry is importable if a significant proportion of its output replaces imports in the domestic market and exportable if a significant proportion of its output is actually exported (Australian Bureau of Statistics, 2010). I do not conduct single-industrial-sector analyses of the combined industries making up the tradable sector because of the problem of there being very few observations in some industries.

Table 4.4 shows the number of firms in the two broad sectors analysed.[32] Most of the firms in both samples are classified as in the non-tradable sector. The definition of the non-tradable

[25]Lombardo and Ravenna (2012) measure tradability of an industry as a sum of exports and imports relative to its gross output, and classifies an industry as tradablethat the literature extends the definition of non-tradable goods to include goods in the traded category provided they are traded below a certain threshold if its tradability measure is above 10 per cent. Zeugner (2013) uses the relative share of exports to total output per sector while, Mano and Castillo (2015) use the average export to value added ratio.

[26]See the world's top exports http://www.worldstopexports.com/zambias-top-10-exports/

[27]Linking the top 10 export data information to the firm panel data, I find that 8 out of the 20 industries are not present in the firm dataset .

[28]Appendix table 4.B.2 gives more information on the top 10 exporting industries.

[29]Export diversification is constrained in Zambia by high trade costs which arise from complying with numerous non-tariff regulatory measures and lengthy administrative procedures at border crossings. Costs of complying with these measures can be high for the exporter, increasing their costs and creating barriers to trade (World Bank, 2014 report).

[30]I exclude the agriculture industry in this analysis because agricultural activities are highly affected by weather patterns. In Zambia crop activity is mainly rain-fed, which makes output volatile to changes in rainfall patterns in different parts of the country. The exclusion is done to avoid rain shocks confounding results by affecting districts that lay within 400 km of the highway differently compared to control areas far away.

[31]The list of top 10 imports in Zambia in 2018 account for 74 per cent of total imports. See Appendix table 4.B.3 for more details. Further information is at http://www.worldsrichestcountries.com/top-zambia-imports.html.The data on imports shows that industrial sectors that significantly import are mining and manufacturing.

[32]Appendix table 4.B.4 shows the distribution of firms by the tradable sub-sectors and the non-tradable sector.

sector includes firms that do not import or export in the services industry. I identify and exclude all importing firms in the services industry. I classify firms as importing using two approaches. In the first approach, which I refer to as the firm-level approach, I consider importing firms all those with an average import-purchase value greater than zero over time. In the second approach, which I refer to as the sector intensity approach, importing firms are all those firms in industrial sectors with an import intensity greater than 10 per cent.[33] The use of two import-status classification techniques yields two distinct control groups for non-tradable service firms. The firm-level approach yields a control group of 853 in the sample that excludes Lusaka, and 2,370 observations in the sample including Lusaka. The sector-intensity approach yields a control group of 896 and 1,488 observations in the sample excluding Lusaka and the full sample, respectively.

Table 4.4: Tradable and Non-Tradable Sectors

	ISIC Rev.4 Division	Name	Sample	No. of Firms (2013)	No. of Firms (2017)	Change (2017-13)	No. of Districts within 400km	No. of Districts outside 400km
Trading Sector								
Exporting all	07-33, 58	Manufacturing mining and publishing activities	With Lusaka	244	190	-54	8	15
			Without Lusaka	109	80	-29	7	15
Non-Tradable Sector								
Firm level Approach	35-99	Services	With Lusaka	1332	1038	-294	11	27
			Without Lusaka	511	342	-169	10	27
Sector Intensity Approach	35-99	Services	With Lusaka	757	731	-26	10	22
			Without Lusaka	554	342	-212	7	15

Note: Exporting all tradable sector includes all firms in industrial sectors under the top 10 exporters in Zambia, including all firms in the mining and manufacturing sector.
Firm level approach classifies all firms with an average import purchase value greater than zero over time as importing with, firms with an import purchase value equal to zero classified as non-importing. Firms with an import purchase value of zero that do not export are classified as the non-tradable sector firms under the firm level approach.
The sector intensity approach classifies all firms in industrial sectors with an import intensity greater than 10 percent as importing. Those firms with an import intensity less than 10 percent and are non-exporting form the non-tradable sector under the sector intensity approach.

4.3.3.2 Why Study the two Broad Sectors

The effects on economic activity of reduced transport costs arising from improved highways are categorised for the tradable and non-tradable sectors. A more fine-grained sectoral analysis was not possible because there are very few firms in many of the industrial sectors. Further,

[33] Sector import intensity is the import intensity of firms averaged at 4-digit industry level. A firm in a sector with sector intensity greater than 10 per cent is classified as an importing firm. Firm import intensity is the ratio of import purchase to sales values.

analysing sectoral dynamics in terms of tradable versus non-tradable sectors allows the consideration of the implications for exports (Piton, 2017). The tradable sector is an important source of foreign exchange. It provides foreign exchange to import intermediate goods, increasing capital formation and stimulating output growth (Awokusa, 2008; Balassa, 1978; Esfahani, 1991). Therefore, many governments prioritise investments in export industries to improve the balance of payments, stimulate trade and economic growth.[34]

4.4 Theoretical Framework

4.4.1 Monopolistic Competition with Heterogenous Firms

Poor quality transport infrastructure increases transport costs (Nordas and Piermartini, 2004). It causes friction in the movement of goods and labour from one location to another (Naude and Matthee, 2007).[35]

Correspondingly, the improvement of an international highway should decrease transport costs and increase trade. I use Melitz (2003) trade model with heterogeneous firms to explain how reduced transport costs affect the performance of firms in the tradable and non-tradable sectors. The Melitz model is suited to explain the empirical relationships investigated in this essay for several reasons. The model explains how under firm heterogeneity, a decrease in trade-related variable costs leads to an increase in exports through existing exporters exporting more and new firms entering the export market. The model also explains how increased trade forces the least productive firms to exit. It further describes how increased trade increases foreign product variety in the domestic market. Lastly, the model also explains how incumbent firms incur a loss in domestic sales due to increased trade, with exporting firms making up for the loss through increased exports while non-exporting firms incur the full loss.

In the theoretical framework of this essay, I adopt all assumptions made in the Melitz model. This essay studies an economy existing in a world that comprises some number of identical countries.[36] The market structure is monopolistic competition with heterogeneous firms. Firm entry into the domestic market requires an investment in overhead costs of $f > 0$ by firms in both the tradable and non-tradable sector. Tradable sector firms can export to any country, although entry into each export market requires another fixed cost investment of $f_x > 0$ measured in labour units. Non-tradable sector firms do not export because their fixed costs to enter

[34] Prioritising investment in the tradable sector promotes economic growth and policy makers and researchers look at the success of the outward-oriented Asian economies in contrast to the economic failures of inward-oriented developing countries like India and Latin America as evidence (Awokuse, 2008).

[35] When the movement of goods, services and inputs in production are within the borders of a particular country, the transport costs are called domestic transport costs, whereas when movement crosses borders there is an additional element of international transport costs (Naude and Matthee, 2007). Roy Harrod states that virtually all commodities are tradable within some area, with the extent of the area determined by transportation costs (De Gregorio et al., 1994). Non-tradable goods are those goods that are not traded internationally, due to prohibitive transportation costs (Obstfeld and Rogoff, 1996)

[36] This assumption is made to ensure factor price equalization across countries to allow for the analysis of firm selection effects that are independent of wage differences (Melitz, 2003).

the export markets for these firms are infinite. Firms in the tradable and non-tradable sectors face the same domestic fixed costs but have different productivity levels ψ. All exporting firms in the tradable sector service the domestic market; however, there are some tradable sector firms that do not export. These firms together with the non-tradable sector firms exclusively service the domestic market. Firm mark ups are exogenously fixed by the constant elasticity of substitution between varieties.[37] The pricing rule for an individual firm in the tradable and non-tradable sector in the domestic market is $p_d(\psi) = \frac{w}{\rho\psi} \iff \frac{1}{\rho\psi}$ where $\dfrac{1}{\rho} = \dfrac{\sigma - 1}{\sigma}$ is the profit maximising mark up. w is the wage cost normalised to one for simplicity. Exporting firms in the tradable sector set prices higher in the foreign markets to reflect the marginal cost τ of servicing those markets, $p_x(\psi) = \dfrac{\tau}{\rho\psi} = \tau p_d(\psi)$. Firm profit from the export market is given by π_x while that from the domestic market is given by π_d. Below are the profit equations from the export and domestic markets:

$$\pi_x = \frac{r_x(\psi)}{\sigma} - f_x \tag{3.C.1}$$

$$\pi_d = \frac{r_d(\psi)}{\sigma} - f \tag{3.C.2}$$

Here σ is the constant price elasticity of demand which is the same for tradable and non-tradable varieties. The variables, $r_x(\psi)$ and $r_d(\psi)$ are the revenue from exports and domestic sales, respectively, with export revenue expressed as $r_x(\psi) = \tau^{1-\sigma} r_d(\psi)$.[38] The endogeneously determined productivity level is ψ is drawn from a common distribution $g(x)$, which lies between $0, \infty$ and has a continuous cumulative distribution $G(x)$. The profit for a non-tradable firm is given by $\pi_{nt} = \pi_{nt}(\psi_{nt})$ while the combined profit for a tradable sector firm is given by:

$$\pi_t(\psi_t) = \pi_{td}(\psi_{td}) + max\{0, n_x \pi_{tx}(\psi_x)\} \tag{3.C.3}$$

n_x represents the number of countries to which firms export. A firm from the tradable or non-tradable sector will produce in the domestic market if when entering the market, the initial productivity level it draws from the ex-ante exogenous distribution $g(x)$ is above the cutoff productivity level ψ^* for its sector. The cutoff productivity is the lowest productivity level at which a firm breaks even, $\pi(\psi^*) = 0$. The cutoff productivity level in a sector determines the domestic equilibrium average productivity level of that sector $\tilde{\psi}(\psi^*)$. Accordingly, a tradable sector firm will export if its productivity level lies above the cutoff productivity level for exporting firms ψ_x^*, such that $\pi_{tx}(\psi_x^*) = 0$. If $\psi_x^* = \psi_{td}^*$ all firms in the tradable sector export. If $\psi_x^* > \psi_{td}^*$ then firms in the tradable sector with productivity levels equal to or above ψ_{td}^* but less than ψ_x^* will produce for the domestic market. These firms do not export as their export profits would be negative. They do, however, earn non negative profits from domestic sales. Tradable

[37] The price elasticity of demand of any variety of good is unresponsive to changes in the number or price of any competing variety of good. There is constant elasticity of substitution among product varieties.

[38] See Melitz (2003) for the derivation of this equation and all other equations under this section.

sector firms with productivity levels above ψ_x^* earn positive profits both from domestic and export sales. The partitioning of firms in the tradable sector by export status occurs because of fixed and variable trade costs that are greater than the domestic sunk costs, $\tau^{1-\sigma} f_x > f$. A firm's decision to export is based on its ability to meet the fixed-investment and variable-trade costs. The decision to enter the export market is made after each firm knows its productivity level, which remains constant. The export cutoff productivity level ψ_x^* can be expressed in terms of the tradable sector domestic cutoff productivity level ψ_{td}^* using the profit condition as follows:

$$\frac{\frac{r_x(\psi_x^*)}{\sigma} - f_x}{\frac{r_d(\psi_{td}^*)}{\sigma} - f} = \frac{\tau^{1-\sigma} r_d\left(\psi_x^*\right) - f_x}{r_d\left(\psi_{td}^*\right) - f} \iff \psi_x^* = \psi_{td}^* \tau \left(\frac{f_x}{f}\right)^{\frac{1}{\sigma-1}} \tag{3.C.4}$$

When an international highway is improved, a decrease in transport costs affects variable trade costs for tradable sector firms. Firms in the non-tradable sector do not export and only service the district market where they are located. These firms source inputs in their local districts and thus, rely on district roads for business and remain unaffected by the improved international highway. In equation 4.4 the parameter τ, represents variable trade costs. A fall in τ implies a fall in the export cutoff productivity level ψ_x^*. This generates an opportunity for firms in the tradable sector who initially were unable to do so because of the high export cutoff productivity level to enter the export market. The expectation of future positive profits is the reason firms make investments to enter the domestic and export market. Thus, the average profit function $\bar{\pi}_t$ per firm in the tradable sector expressed in terms of the cutoff productivity level ψ^* is:

$$\bar{\pi}_t = \pi_{td}\left(\tilde{\psi_{td}}\right) + p_x n_x \pi_{tx}\left(\tilde{\psi_x}\right) \iff f k\left(\psi_{td}^*\right) + p_x n_x f_x\left(\psi_x^*\right) \tag{3.C.5}$$

Where p_x is the fraction of firms that export and $k\left(\psi^*\right) = \left[\frac{\tilde{\psi}\left(\psi^*\right)}{\psi^*}\right]^{\sigma-1} - 1$. A fall in transport costs entails an increase in p_x the fraction of exporting firms as the export productivity cutoff level ψ_x^* falls as shown in Equation 4.4. Equation 4.5 indicates that an increase in p_x implies an increase in $\bar{\pi}_t$ the equilibrium average profit per firm or the zero cutoff profit condition in the tradable sector. This induces a higher cutoff productivity level $\psi_{td}^{**} > \psi_{td}^*$ for tradable sector firms in the domestic market.[39] Thus, firms in the tradable sector with productivity levels below the new cutoff productivity level ψ_{td}^{**} exit the domestic market as they start to experience negative domestic profits in the new equilibrium. The relationship of the total number of tradable sector firms in relation to the number of exporting firms is given by:

$$M = \frac{R}{\bar{r}} = \frac{L}{\sigma\left(\bar{\pi} + f + p_x n_x f_x\right)} \tag{3.C.6}$$

In equilibrium, total revenue is $R = L = L_p + L_e$. The model assumes labour is the only input

[39]Refer to Melitz (2003) for more details on the determination of the unique equilibrium ψ^* and $\bar{\pi}$ determined by the zero cutoff profit and the free condition curves.

in production and revenue should cover the cost of productive labour L_p and investment labour L_e. The equilibrium number of tradable sector firms M can be determined by dividing the total revenue R by the average revenue per firm $\bar{r}$ as shown above. Equation 4.6 shows that an increase in the fraction of exporting firms p_x in the tradable sector decreases the equilibrium number of tradable sector firms M in the economy. Although the number of tradable sector firms decrease in the economy due to increased trade, the quantity of product variety available for consumption in the tradable sector increases. An increase in foreign-product variety compensates for the reduction in the number of domestic tradable firms. The equilibrium-level total variety consumed in the tradable sector of an economy is given by M_t:

$$M_t = M + n_x p_x M \iff M_c = (1 + n_x p_x)\, M \tag{3.C.7}$$

M is the total amount of domestic product variety produced by the tradable sector and $n_x p_x M$ is the total amount of foreign product variety available in an economy. Equation 4.7 implies that a decrease in the number of tradable sector firms M is dominated by an increase in $n_x p_x M$ the number of new foreign tradable sector firms or imported products, which increase the total product variety in the tradable sector of an economy.

Further, because increased trade induces an increase in the domestic cutoff productivity level in the tradable sector from ψ_{td}^{*} to ψ_{td}^{**}, the equilibrium average productivity level $\tilde{\psi}_t$ also increases.[40]

$$\tilde{\psi}_t = \left\{ \frac{1}{M_t} \left[M \tilde{\psi}_t^{\sigma-1} + n M_x \left(\tau^{-1} \tilde{\psi}_x \right)^{\sigma-1} \right] \right\}^{\frac{1}{\sigma-1}} \tag{3.C.8}$$

M_t represents the total number of firms in the tradable sector in an economy. M represents domestic tradable sector firms and M_x represents exporting firms in the tradable sector. A fall in τ in Equation 4.8 implies an increase in the equilibrium average productivity level $\tilde{\psi}_t$. This implies a fall in aggregate domestic prices in the tradable sector $p_{td}(\psi_t) = M_c^{\frac{1}{1-\sigma}} \frac{1}{\rho \tilde{\psi}_t}$ and a reduction in revenue. Therefore, all tradable sector firms in the domestic market, both exporting and non-exporting will experience a loss in revenue due to increased trade. Non-tradable sector prices remain unaffected as the average productivity level in this sector is not affected by increased trade.

The main conclusions from the theoretical model above are that:

1. Decreased transport costs increases output in the tradable export sector. The now lower equilibrium export productivity cutoff level attracts new firm entrants into the export market, and incumbent exporting firms experience increased profits.
2. Increased trade forces inefficient firms in the tradable sector to exit the domestic market. The cutoff productivity level increases in the domestic market, and firms with productivity below that level experience negative profits and exit the market.

[40] By symmetry, the equilibrium combined average productivity encompassing all firms in the tradable sector (domestic and foreign) increases. See Melitz (2003) aggregation section 6.2 for further explanations.

3. Incumbent tradable sector firms experience a fall in revenue, but exporting firms make up for the loss in domestic revenue with export revenue. A high cutoff productivity level in the domestic market reduces the aggregate domestic prices and revenue for firms in the tradable sector.
4. Increased trade exposure increases foreign product variety in the tradable sector of an economy.

The following are the empirical predictions:

1. Sales, gross profits and employment expenses– Increased growth in the tradable sector compared to the non-tradable sector after the road is improved because of decreased transport costs and increased access to foreign markets.
2. Freight transport and fuel costs – Increased growth in the tradable sector relative to the non-tradable sector because of increased utilisation of the highway by tradable sector firms.
3. Inventory Costs – Decreased growth in the tradable sector compared to the non-tradable sector because of increased access to wider markets.
4. Cost of Sales – Increased growth in the tradable sector compared to the non-tradable sector. Increased exports to new markets and increased product variety in the domestic tradable sector market also entail increased advertising and marketing costs.

4.5 Methodology

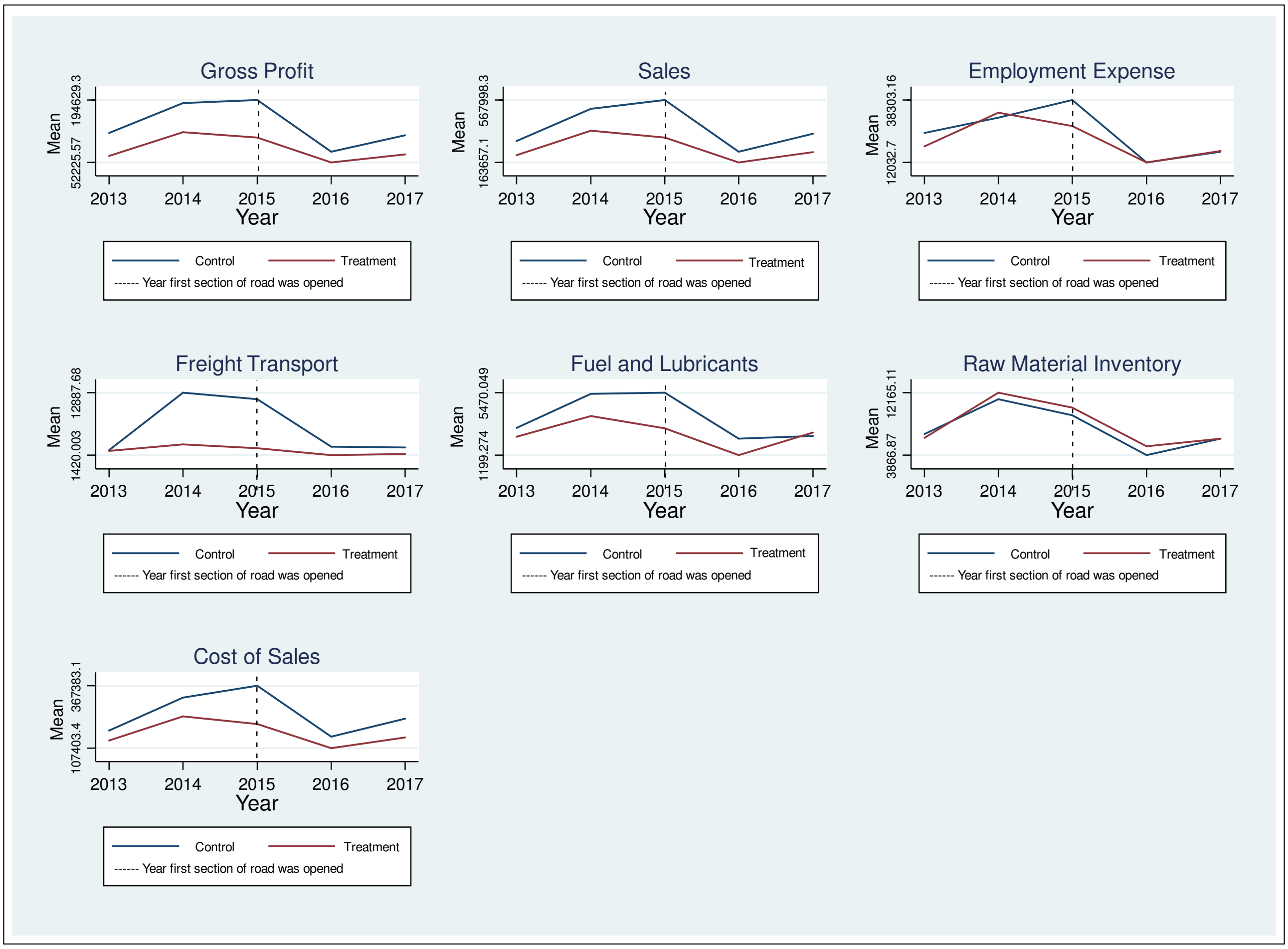
The aim is to estimate the effect of an improved highway on the performance of firms engaged in international trade. I estimate the effects on costs and output of firms in the tradable sector relative to firms in the non-tradable sector of reduced transport costs arising from an improved international highway. I use a triple difference estimation method that compares the double difference in the tradable sector to the double difference in the non-tradable sector. This methodology controls for location unobserved fixed differences, time-varying factors that affect all locations equally and time-varying effects unrelated to the treatment that affect locations distinctively, all of which can bias the estimation of average treatment effects.

4.5.1 Identification Strategy

4.5.1.1 Common Trends Assumption

The validity of the DID estimation strategy requires that, in the absence of treatment, unobserved differences and trends in the treatment and control groups remain fixed over time. Violation of this assumption results in biased causal estimates. The triple difference estimation strategy, however, controls for time-varying factors that change differently across groups. Data is required for more than one pre-treatment period to check whether outcome variables in the control and treatment groups follow parallel trends pre-treatment; this data is, however, not available. Taking advantage of the fact that effects of highway-construction improvement take a few years to appear, I use firm time series data covering the highway construction period from 2013 to 2016 to check for pre-trends in outcome variables. The highway improvement project was completed in 2017, while the first complete section (the Nyimba-Petauke 34 km road) was opened to traffic in December 2015. Effects of the improved highway are expected to appear after 2016. Results in Figure 4.1 show trends of the seven dependent variables in the firms located near the improved highway compared to firms located far away from it during the construction period. Trends on five variables are parallel in the control and treatment groups, except for employment expenses and transport costs. However, this problem is controlled by using the triple difference estimation methodology.[41]

Figure 4.1: Parallel Trends Trends in Outcome Variables by Distance



[41]When the parallel trends assumption is violated because of some time varying confounder that changes differentially across states, researchers resolve this problem by using a DDD design to look for a within-state not exposed to treatment but exposed to the time varying confounder (Wing et al, 2018). Results in Appendix figure 4.A.2 displays trends in outcome variables by sector treatment. It also shows that some outcome variables do not show parallel trends in the pre-treatment period in the control and treatment groups.

4.5.1.2 Nacala highway Project as A Natural Experiment

Road construction and improvement projects are prone to the endogeneity problem of reverse causality. Roads are built or improved in areas where economic activity is growing very fast or lagging in order to stimulate development. Based on source documents from the RDA, the improvement of the highway is related to boosting economic activity in Southern African Development Community by providing an efficient transport infrastructure. The aim of the corridor is to reduce transport costs by connecting Lusaka in Zambia and Lilongwe in Malawi to the Port of Nacala in Mozambique using an efficient transport link at the shortest driving distance. However, the improved highway is anticipated to have a different impact in districts along the highway compared to the endogenously chosen cities that face different development initial conditions and growth trajectories. It is expected that estimates showing the impact of the improved highway will be biased upwards in the sample including firms in Lusaka city. This is because the placement and improvement of the highway is related to the high economic activity in Lusaka at baseline, while this is not the case for areas located in-between Lusaka city and Lilongwe city. Further, locations between the two cities were treated only because they lay along the shortest driving route between the two cities. This identification strategy allows me to run estimations on two samples, one sample excludes Lusaka city and the other does not. This allows for the study of the impact of the improved highway in the peripheral region.[42]

4.5.2 Model Specification

The empirical model below is used to estimate the impact of the improved highway in the tradable sector.

$$\begin{aligned} Y_{ijst} = \alpha_0 + \beta_1 Post_t + \beta_2 d400km_j + \beta_3 d400km_j * Post_t + \beta_4 Tradable_s + \\ \beta_5 Tradable_s * Post_t + \beta_6 d400km_j * Tradable_s * Post_t + \\ \beta_7 Tradable_s * d400km_j + \gamma X_{jt} + \tau_j + \epsilon_{ijst} \end{aligned} \tag{3.C.1}$$

In the above model, Y_{ijst} is the outcome of the ith firm in district j sector s at time t. $Post_t$ is a binary variable taking the value one if the year is 2017 and zero otherwise.[43] $d400km_j$ is a binary variable taking the value one if j indexes a district within 400 km of the improved highway and zero otherwise.[44] $Tradable_s$ is a binary variable taking the value one if s indexes a firm in the tradable sector and zero otherwise. X_{jt} are other observable district factors affecting the outcome variable. τ_j are location fixed effects and ϵ_{ijt} is the idiosyncratic error term. The parameter of interest is β_6 which shows the differential change in outcomes of firms within the

[42] The term peripheral refers to districts outside the capital city of Lusaka while the centre refers to the city of Lusaka.
[43] $Post_t$ has no group subscript because time period does not vary across groups
[44] This dummy does not have a time subscript because group membership is time invariant

tradable sector compared to the differential change in outcomes of firms in the non-tradable after the highway is improved. The aim is to observe growth in the tradable sector compared to the non-tradable sector of firms located near the highway after it is improved. The control group is the non-tradable sector which constitutes service firms that do not import or export. In all regressions, I use location and time fixed effects to control for potential time-invariant group-specific characteristics and time-varying group-invariant confounding factors. With location fixed effects, highway effects are identified entirely from variation within locations over time, after accounting for common shocks that affect all districts.[45] I also interact the post variable with the industry variable and the city size variable (population) separately to ensure results are not driven by variation in certain industries or cities of certain size over time.

I study effects of the improved highway on the tradable sector located within 400 km compared to the tradable sector outside this radius over time vs the non-tradable sector within 400 km compared to the non-tradable sector outside this radius because, 400 km is the closest distance bandwidth to the improved highway with a reasonable number of firms to allow for the estimation of consistent highway effects.[46]However, I do run estimations on the average driving distance of firms to the improved highway as a robustness check to observe changes in the performance of tradable sector firms as the distance from the improved highway is increased.[47]

Table 4.5 shows the distribution by distance band of firms in the two broad sectors in the two time periods. The number of treated firms within 400 km is considerably higher than the number in other distance bands much closer to the improved highway. Nevertheless, asymptotic distribution properties are still not met in this sample as there are only a few treated firms in few treated clusters.[48] This raises two issues. The first is that t-tests based on cluster-robust standard errors will severely over-reject the null hypothesis when the number of clusters is small (Mackinnon and Webb, 2018). Secondly, when the number of observations in the treated clusters is small relative to the number of observations in the control group (heterogeneous clusters) the homoskedastic assumption is violated and the null hypothesis is over-rejected (Ferman et al, 2019). To address both concerns, I use Monte Carlo permutation tests of the t-statistic to estimate p-values because this method does not rely on asymptotic properties.[49]

[45]The location-level analysis captures effects working both through changes within firms, and through entry, exit and relocation of firms (Gibbons et al, 2019).

[46]I could not run estimations on incremental driving distance of 100 km bandwidths or less than 400 km from the improved highway because there are not enough tradable sector firms in these bandwidths. Further, some distance bandwidths do not have any tradable sector firms in 2017. Firms are sparsely distributed in the dataset and the average firm driving distance to the improved highway is 518.76.

[47]In back end calculations, I estimate the impact of a change in accessibility to Port Nacala on firms in the tradable sector compared to firms in the non-tradable sector and results remain insignificant as in chapter three. See appendix 4.B.5, 4.B.6 and 4.B.7 for rank results.

[48]Tradable sector firms are present in 7 districts within 400 km of the improved highway. Table 4.5 panel A indicates that there are three firms in the tradable sector within 300 km of the improved highway in 2013. The number increases to 18 firms when the distance from the improved highway is increased to 400 km.

[49]Monte Carlo permutation tests determine the significance of the observed value of a test statistic by rearranging the order (permuting) of the observed values of a variable. When the variability implied by the confidence interval makes conclusions difficult, you may increase the number of replications to determine more precisely the significance of the test statistic of interest. In other words, the p-value will converge

I run 1000 simulations in order to get more accurate small p-values to avoid obtaining p-values that equal zero when the observed test statistic is greater than those in the replications.

Table 4.5: Distribution of Firms by Distance

	Number of Firms (2013)		Number of Firms (2017)	
Distance	Tradable Sector	Non-Tradable	Tradable	Non-Tradable
Sample Excluding Lusaka				
100km	1	10	0	2
200km	1	12	0	3
300km	3	15	3	9
400km	18	49	12	35
500km	48	188	31	107
600km	103	457	75	298
600km>	6	54	5	44
Full Sample				
100km	1	10	0	2
200km	1	12	0	3
300km	138	836	113	705
400km	153	870	122	731
500km	183	1009	141	803
600km	238	1278	185	994
600km>	6	54	5	44

4.6 Empirical Results

4.6.1 Firms within 400 km

4.6.1.1 Transport and Fuel Costs in the Tradable Sector

Table 4.6 shows the results for freight transport, fuel and lubricant costs. As explained earlier, freight transport costs reflect the costs to the firm of transporting raw materials and finished goods by road, air or sea. Freight transport costs also include costs of vehicle maintenance, insurance, roadblocks, and border and seaport clearance costs. Fuel and lubricant costs capture costs associated with the use of fuel and other oils by the firm's fleet. I run regressions to

to the true permutation p-value as the number of replications gets arbitrarily large (see stata manual on https://www.stata.com/manuals13/rpermute.pdf).

I was not able to use the subcluster bootstrap (ordinary wild cluster bootstrap) which is recommended when faced with few treated clusters because it does not perform well under few treated heterogeneous clusters. (Mackinnon and Webb, 2018).

estimate how the improved highway affects the firm's freight transport and fuel costs. The term of interest is $d400km_j * Post_t * Tradable_s$, which is the triple difference term capturing growth effects after the highway is improved of firms in the tradable sector relative to those in the non-tradable sector. As indicated in the previous section, the inference from the results is based on p-values estimated using Monte Carlo permutation tests of the t-statistic; this is necessary because of the problem of there being few treated heterogeneous clusters.

Panel A in Table 4.6 shows regression results generated from the sample excluding Lusaka, which is the cleanest sample under identification. Results under the firm-level approach indicate that, compared to those in the non-tradable sector, firms in the tradable sector experience semi-elasticity in freight transport costs with respect to the improved highway that is 5.53 higher. [50] A semi-elasticity of 5.534 means that after the highway is improved, firms in the tradable sector experience an increase in freight transport costs that is 553.4 per cent higher than for firms in the non-tradable sector. This result is inconclusive because the effect size is large; however, the p-value is greater than the 0.10 significance level. This suggests that based on the existing data, it cannot be concluded that the improved road has no effect on freight transport costs. The result under the sector intensity approach is larger in size (semi-elasticity of 8.37 based on the coefficient value of 2.238) but equally inconclusive.[51] Large positive estimates of freight costs signal that a fall in transport costs when the highway is improved causes firms in the tradable sector to increase their use of the highway. Theoretically, a fall in transport costs allows tradable sector firms who previously only supplied products to the domestic market to enter the export market, with incumbent exporting firms increasing their export quantities. These two activities increase the utilisation of the improved highway, which manifests through an increase in freight transport costs. On the contrary, insignificant growth effects could suggest that transport costs associated with road quality are a small and insignificant part of freight costs, while border and port clearance costs and delays are probably the significant part of freight transport costs faced by exporting firms in landlocked countries in Africa.[52]

Further, I run regressions to observe the effects of the improved highway on trade costs of fuel and lubricants. Under the firm-level approach in the sample excluding Lusaka, I find that the semi-elasticity of fuel and lubricant costs with respect to the improved highway is 53.65 (coefficient 4.001) for firms in the tradable sector. This result means that firms in the

[50] In regression models with log transformed outcomes, the impact of a one-unit change in an independent variable (X) is calculated by exponentiating the coefficient and subtracting one from the number (Mitnik et al., 2018). In this case, $(exp\,(\beta_j) - 1) = exp\,(1.877) - 1 = 5.534$. To give it a per centage interpretation multiply this result by 100. When the estimated coefficient β_j is less than 0.10 the interpretation that a unit increase in X is associated with an average of $100 * \beta_j$ per cent increase in Y works well (Mitnik et al., 2018). I refer to the exponentiated coefficients throughout the paper. Semi-elasticity refers to the per centage change in the dependent variable Y with respect to a unit change in the independent variable X, $\%\Delta y = [100 * \{(exp\beta_j) - 1\}]\Delta x$.

[51] The sector intensity approach classifies all firms in industrial sectors with an import intensity greater than 10 per cent as importing while the firm level approach considers all firms with an average import purchase value greater than zero as importing.

[52] Naude and Matthee (2007) report that many firms in Africa face high domestic transport costs to get their products to the ports, which are more than 500 km away. Firms in landlocked countries in Africa transit goods through various borders to get to the seaport and thus transport cost reductions derived from improving the highway quality may be neutralised by border and port clearance costs and delays.

tradable sector near the highway experience a growth in fuel and lubricant costs of 5,365 per cent higher compared to firms in the non-tradable sector after the highway is improved. This estimate is large and statistically significant at the 10 per cent level, and the result under the sector intensity approach is not so different. The results for fuel costs suggest that firms in the tradable sector take advantage of the improved transport infrastructure by utilising it more. This is directly reflected in positive growth in fuel costs. Improved highways decrease transportation costs, which induces more firms in the tradable sector to utilise the improved highway as they access new markets.

Panel B in Table 4.6 shows regression results from the sample that includes Lusaka. The semi-elasticity of freight transport costs and of fuel costs with respect to the improved highway are both statistically insignificant and are much smaller than those under panel A. This suggests that, in this sample, effects are largely driven by firms located in Lusaka. The double difference term $d400km_j * Post_t$ captures differential changes for firms in the non-tradable sector located near the improved highway as compared to changes for non-tradable-sector firms located further away. This term is free from improved highway treatment effects but captures other economic or policy effects that differentially affect districts over time. With regards to fuel and lubricant costs, Columns 2 and 4 show that the non-tradable sector firms in Lusaka seem to be on a higher growth path, unlike firms in the same sector outside Lusaka. As indicated earlier, the sample which includes Lusaka city is expected to have estimates that are upward biased because of the endogeneity problem of improving the highway because of the high economic activity in Lusaka city. As shown in Table 4.6, results in the sample that includes Lusaka are not as expected. This could be because Lusaka, being the capital city of Zambia, is already connected to other efficient transport links. As a result, the improvement of the Nacala road corridor does not bring significant transport cost reductions to induce growth in transport and fuel costs for firms in the tradable sector.[53]

The summary of findings in Table 4.6 is that firms in the peripheral area in the tradable sector, relative to those in the non-tradable sector, experience positive growth in fuel costs after the highway is improved. A fall in transport costs associated with the improved highway reduces variable trade costs, which causes firms in the tradable sector to increase their use of the transport link. A fall in transport cost leads to an increase in the number of exporting firms in the tradable sector as tradable sector firms that previously only serviced domestic markets due to high transport costs are now able to enter the export market when transport costs decrease. On the other hand, the impact of the improved highway on firm freight transport costs are inconclusive given the available data. According to the World Trade Organisation, freight transport costs are a crucial determinant of a country's ability to participate in global trade (Nordas and Piermartini, 2004). Thus, inconclusive effects of the improved highway on firm freight transport costs in this essay suggest the need for additional studies on this relationship in Southern Africa for a better understanding.

[53]See map of Lusaka and its transport port links in appendix figure 4.A.3.

Table 4.6: Transport and Fuel Costs

	(1) Transport	(2) Fuel & Lubricants	(3) Transport	(4) Fuel & Lubricants
	Firm Level		Sector Intensity	
Panel A: Sample Excluding Lusaka				
d400km*Post	-1.498*	0.199	-1.681**	-0.067
	(0.796)	(1.055)	(0.819)	(0.977)
Post*Tradable	8.892***	12.495***	9.338***	12.457***
	(0.331)	(0.619)	(0.370)	(0.864)
d400km*Post*Tradable	1.877	4.001***	2.238	3.965***
	(2.420)	(0.733)	(2.454)	(0.860)
Observations	1042	1042	1085	1085
R^2	0.2967	0.2762	0.2564	0.2438
Few Treated Heterogeneous Clusters Robustness				
Monte Carlo permutation tests (p-value)	0.3190	0.0890	0.2550	0.1030
Panel B: Full Sample				
d400km*Post	-0.031	1.061**	0.531	1.576*
	(0.429)	(0.448)	(0.539)	(0.928)
Post*Tradable	0.953	4.257***	1.597**	3.972***
	(0.795)	(0.327)	(0.601)	(0.476)
d400km*Post*Tradable	0.063	-0.379	0.021	-0.182
	(0.625)	(0.828)	(0.636)	(0.574)
Observations	2804	2804	1922	1922
R^2	0.1567	0.1712	0.1860	0.1983
Few Treated Heterogeneous Clusters Robustness				
Monte Carlo permutation tests (p-value)	0.9440	0.7180	0.9720	0.8600

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All variables are in inverse hyperbolic sine. Standard errors in parentheses and clustered at district level. Columns 1 and 2, the sample excluding Lusaka has 41 clusters while the full sample has 42 clusters. Column 3 and 4, the sample excluding Lusaka has 40 clusters while the full sample has 36 clusters. The definition of import status affects the composition of the control group (service firms that do not import or export). The firm level approach considers all firms with an average import purchase value greater than zero as importing. The sector intensity approach classifies all firms in industrial sectors with an import intensity greater than 10 percent as importing. All regression estimations include location and time fixed effects, in addition to an interaction term of the industry dummy and the post variable to control for differential trends by industry overtime in the treatment and control group. To control for differential trends in districts of different sizes overtime in the treatment and control groups, I include an interaction term of district population and the post variable as a control in all regressions. Post is a dummy variable that equals 1 when the year is 2017 and 0 otherwise. d400km is a dummy variable that equals 1 when a firm is located within 400km of the improved highway and 0 otherwise. Road driving distance is calculated using google maps with the point of origin being the district city council, post office or market to the nearest point on the improved highway. Tradable is a dummy variable that equals 1 when a firm belongs to the tradable sector and 0 otherwise. The interaction term d400km*Post gives the differential change of firms in the non-tradable sector located within 400 km compared to firms in the same sector located outside this radius. The interaction term Post*Tradable gives the differential change in firms in the tradable sector compared to firms in the non-tradable sector overtime. The interaction term d400km*Post*Tradable gives the triple difference estimate. This coefficient measures the differential change in firms in the tradable sector relative to firms in the non-tradable sector over time and space. Permutation tests estimate p-values for the triple difference observed value of a test statistic by rearranging the order of the observed values of a variable.

4.6.1.2 Firm Output and Employment Costs in the Tradable Sector

Table 4.7 shows regression results for the impact of the improved highway on gross profits, sales and employment costs.[54] Panel A under the firm level approach shows that the semi-elasticity of gross profits with respect to the improved highway is 177.75 higher for firms in the tradable sector than for firms in the non-tradable sector, while the semi-elasticity of sales is 84.20 higher, although this is inconclusive. [55] This means that firms in the tradable sector, compared to firms in the non-tradable sector, experience growth in gross profits and sales of 17,775 per cent and 8,420 per cent, respectively, after the highway is upgraded. Results under the sector intensity approach give very similar results for gross profits while the estimate on sales increases slightly and becomes weakly significant.[56] These results suggest that firms in the tradable sector experience higher growth in sales and gross profits after the highway is improved than firms in the non-tradable sector. A fall in transport costs reduces trade-related variable costs. This causes tradable-sector firms to begin exporting that were previously unable to do so due to prohibitive transport costs. It also causes incumbent exporting firms to increase export quantities. Thus, increased demand from newly accessed markets results in an increase in the growth of sales and gross profits in the tradable sector.

Results on employment expenses in panel A indicate that the semi-elasticity of employment expense with respect to the improved highway is 107.20 (coefficient 4.684) and 136.69 (coefficient 4.925) higher in the tradable sector relative to the non-tradable sector under the firm level and the sector intensity approaches respectively. This means that firms in the tradable sector relative to firms in the non-tradable sector experience an increase in employment expenses ranging between 10,720 and 13,669 per cent after the highway is upgraded. The coefficients on the term $d400km_j * Post_t$, indicate the growth in employment expenses over time unrelated to the improved highway for firms in the non-tradable sector located within 400 km of the highway compared with those in the non-tradable sector outside this radius. The semi-elasticities of -0.88 (coefficient -2.14) and -0.89 (coefficient -2.21) under the firm-level and sector intensity approach, respectively, are significant at the 5 per cent level (Monte Carlo estimated p-values of 0.048 and 0.045). These results suggest a shift of labour in the peripheral area over time from the non-tradable to the high-performing tradable sector.

In Table 4.7, the results in Panel B show that the improved highway has a statistically insignificant effect on firm output (sales and gross profits) and firm labour costs (total employment expenses). The expectation is for coefficients in the sample that includes Lusaka to be upward biased because of the high level of economic activity in this capital city. The results indicate otherwise, and as previously mentioned, this may be because firms in Lusaka are already enjoy-

[54] As defined earlier, total employment expenses are labour costs which constitute wages of labour directly engaged in production, salaries of labour not directly engaged in production, pension contributions and all other employment related costs.

[55] Recall that when a model has a log transformed dependent variable, a unit change in the independent variable is calculated by the formula $exp\,(\beta_j) - 1$. See (Mitnik et al, 2018). Thus, coefficients 5.186 and 4.445 in column 1 and 2 of panel give semi-elasticities of 177.75 and 84.20 respectively.

[56] As stated earlier, monte carlo p-values are used for inference in this subsection.

ing the optimised benefits of reduced transport costs arising from existing good-quality highway links such that the impact of the improved highway is insignificant.

In summary, compared with firms in the non-tradable sector, firms in the tradable sector in the periphery experience positive growth in output and labour costs after the highway is improved. The growth in employment expenses ranges from 10,720 per cent to 13,669 per cent in the tradable sector. This implies a large increase in the real incomes of individuals in districts outside Lusaka who are employed in the tradable sector. Trade gains of increased sales by exporting firms when transport costs fall increases the demand for workers in the tradable sector. The results in this study regarding employment expenses are close to the findings of Donaldson (2018), who reports that in India, new railways bring social welfare benefits as they cause a large increase in real agriculture incomes.

Table 4.7: Firm Output and Employment Costs

	(1) Gross profit	(2) Sales	(3) Employment Expense	(4) Gross profit	(5) Sales	(6) Employment Expense
		Firm Level			Sector Intensity	
Panel A: Excluding Lusaka						
d400km*Post	0.655	-1.395	-2.144**	0.047	-1.947	-2.210**
	(1.832)	(1.403)	(0.883)	(1.996)	(1.709)	(0.975)
Post*Tradable	30.187***	8.372***	11.761***	30.600***	7.882***	11.232***
	(0.534)	(0.620)	(0.471)	(0.555)	(0.452)	(0.380)
d400km*Post*Tradable	5.186**	4.445***	4.684***	5.916***	5.246***	4.925***
	(1.968)	(1.301)	(0.909)	(2.031)	(0.956)	(0.940)
Observations	1042	1042	1042	1085	1085	1085
R^2	0.3009	0.2792	0.2881	0.2752	0.2496	0.2546
Few Treated Heterogeneous Clusters Robustness						
Monte Carlo (p-value)	0.0930	0.1190	0.0570	0.0680	0.0730	0.0450
Panel B: Full Sample						
d400km*Post	1.167	0.686	0.321	2.450***	2.500***	1.166**
	(0.704)	(0.423)	(0.461)	(0.761)	(0.555)	(0.566)
Post*Tradable	-0.738	-1.264	1.490**	0.492	0.376	2.445**
	(1.232)	(1.029)	(0.567)	(1.883)	(1.870)	(1.144)
d400km*Post*Tradable	1.548	0.627	-0.311	1.411	0.259	-0.376
	(1.040)	(0.738)	(0.519)	(1.040)	(0.467)	(0.368)
Observations	2804	2804	2804	1922	1922	1922
R^2	0.2086	0.2046	0.1939	0.2576	0.2666	0.2358
Few Treated Heterogeneous Clusters Robustness						
Monte Carlo (p-value)	0.2490	0.5980	0.7620	0.3190	0.8280	0.7330

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All variables are in inverse hyperbolic sine. Standard errors in parentheses and clustered at district level. Columns 1,2 and 3, the sample excluding Lusaka has 41 clusters while the full sample has 42 clusters. Column 4, 5 and 6, the sample excluding Lusaka has 40 clusters while the full sample has 36 clusters. The definition of import status affects the composition of the control group (service firms that do not import or export). The firm level approach considers all firms with an average import purchase value greater than zero as importing. The sector intensity approach classifies all firms in industrial sectors with an import intensity greater than 10 percent as importing. All regression estimations include location and time fixed effects, in addition to an interaction term of the industry dummy and the post variable to control for differential trends by industry overtime in the treatment and control group. To control for differential trends in districts of different sizes overtime in the treatment and control groups, I include an interaction term of district population and the post variable as a control in all regressions. Post is a dummy variable that equals 1 when the year is 2017 and 0 otherwise. d400km is a dummy variable that equals 1 when a firm is located within 400km of the improved highway and 0 otherwise. Road driving distance is calculated using google maps with the point of origin being the district city council, post office or market to the nearest point on the improved highway. Tradable is a dummy variable that equals 1 when a firm belongs to the tradable sector and 0 otherwise. The interaction term d400km*Post gives the differential change of firms in the non-tradable sector located within 400 km compared to firms in the same sector located outside this radius. The interaction term Post*Tradable gives the differential change in firms in the tradable sector compared to firms in the non-tradable sector overtime. The interaction term d400km*Post*Tradable gives the triple difference estimate. This coefficient measures the differential change in firms in the tradable sector relative to firms in the non-tradable sector over time and space. Monte Carlo permutation tests estimate p-values for the triple difference observed value of a test statistic by rearranging the order of the observed values of a variable.

4.6.1.3 Cost of Sales and Raw Material Closing Inventory Costs in the Tradable Sector

Table 4.8 shows the regression results concerning the impact of the improved highway on the costs of sales and raw materials inventory.[57] Panel A indicates that the semi-elasticity of the cost of sales with respect to the improved highway is inconclusive. However, the semi-elasticity of raw materials inventory costs with respect to the improved highway is significant and is approximately 0.98 less in the tradable sector than in the non-tradable sector under both the firm-level and sector-intensity approaches. This means that, relative to firms in the non-tradable sector, firms in the tradable sector experience a growth in inventory costs that is 98 per cent less after the highway is upgraded. The reduction in inventory costs for firms in the tradable sector suggests that the improved highway causes firms to increase their production capacity in response to increased demand from new export markets. In addition, a reduction in inventory costs for firms in the tradable sector may also signal that the improved highway increases access to cheaper input markets. Consequently, firms in the tradable sector start to purchase their raw materials from the cheaper markets. This result resonates with the finding of Datta (2012) that improved highways reduce inventory-holding days of Indian firms and that many firms change their supplier arrangements.

The results in Panel B show that the effects of the improved highway on the costs of sales and closing inventory are smaller and insignificant compared to those in Panel A. This result is similar to those observed in Tables 4.7 and 4.6 after Lusaka is added to the sample. In this sample, firms within 400 km in the non-tradable sector seem to experience higher growth than non-tradable-sector firms outside this radius. Thus, the triple difference estimate, in which the double difference in the non-tradable sector is deducted from the double difference in the tradable sector, yields a smaller coefficient when compared to the triple difference estimate in Panel A.

[57] Cost of sales are costs associated with the selling of products such as advertising and promotions.

Table 4.8: Cost of Sales and Inventory Costs

	(1) Cost of Sales	(2) Raw Materials	(3) Cost of Sales	(4) Raw Materials
	Firm Level		Sector Intensity	
Panel A: Sample Excluding Lusaka				
d400km*Post	-1.332	0.307	-1.579	0.211
	(1.508)	(0.864)	(1.706)	(0.854)
Post*Tradable	5.881***	14.608***	5.310***	21.009***
	(0.813)	(0.345)	(0.577)	(0.469)
d400km*Post*Tradable	4.262***	-3.808	4.607***	-3.784
	(1.575)	(2.856)	(1.185)	(2.890)
Observations	1042	1042	1085	1085
R^2	0.3040	0.4054	0.2825	0.3744
Few Treated Heterogeneous Clusters Robustness				
Monte Carlo permutation tests (p-value)	0.1410	0.0530	0.1360	0.0460
Panel B: Full Sample				
d400km*Post	0.530	0.316	2.537***	0.226
	(0.477)	(0.225)	(0.470)	(0.353)
Post*Tradable	-1.050*	-0.544	0.512	-0.674
	(0.619)	(0.683)	(1.238)	(0.629)
d400km*Post*Tradable	0.416	0.506	-0.161	0.569
	(0.947)	(0.688)	(0.671)	(0.630)
Observations	2804	2804	1922	1922
R^2	0.2252	0.2883	0.2938	0.3462
Few Treated Heterogeneous Clusters Robustness				
Monte Carlo permutation tests (p-value)	0.4770	0.7490	0.9350	0.4440

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All variables are in inverse hyperbolic sine. Standard errors in parentheses and clustered at district level. Columns 1 and 2, the sample excluding Lusaka has 41 clusters while the full sample has 42 clusters. Column 3 and 4, the sample excluding Lusaka has 40 clusters while the full sample has 36 clusters. The definition of import status affects the composition of the control group (service firms that do not import or export). The firm level approach considers all firms with an average import purchase value greater than zero as importing. The sector intensity approach classifies all firms in industrial sectors with an import intensity greater than 10 percent as importing. All regression estimations include location and time fixed effects, in addition to an interaction term of the industry dummy and the post variable to control for differential trends by industry overtime in the treatment and control group. To control for differential trends in districts of different sizes overtime in the treatment and control groups, I include an interaction term of district population and the post variable as a control in all regressions. Post is a dummy variable that equals 1 when the year is 2017 and 0 otherwise. d400km is a dummy variable that equals 1 when a firm is located within 400km of the improved highway and 0 otherwise. Road driving distance is calculated using google maps with the point of origin being the district city council, post office or market to the nearest point on the improved highway. Tradable is a dummy variable that equals 1 when a firm belongs to the tradable sector and 0 otherwise. The interaction term d400km*Post gives the differential change of firms in the non-tradable sector located within 400 km compared to firms in the same sector located outside this radius. The interaction term Post*Tradable gives the differential change in firms in the tradable sector compared to firms in the non-tradable sector overtime. The interaction term d400km*Post*Tradable gives the triple difference estimate. This coefficient measures the differential change in firms in the tradable sector relative to firms in the non-tradable sector over time and space. Permutation tests estimate p-values for the triple difference observed value of a test statistic by rearranging the order of the observed values of a variable.

4.6.2 Firms Located within 519 km

Theoretically, improving the quality of the highway reduces transportation costs for firms in the tradable sector using the highway. The benefits of decreased transport costs associated with an improved highway will affect firms in the tradable sector differently depending on their driving distance to the highway. Further, firms located in districts far from the improved highway or in areas through which the improved highway does not pass will also, for some part of their journey, use the improved highway to transport goods and services and thus will still benefit to some extent. Benefits are expected to be highest for tradable-sector firms located in districts through which the improved highway passes, with the size of effects decreasing the farther away from the highway a firm is located.

The earlier estimations in this study focus on the impact of the improved highway on tradable sector firms located within 400 km in comparison to those located outside this radius. Coefficients from these regressions are expected to underestimate the impact of the improved highway because of the large driving distance from the improved highway considered. As stated earlier, the distance band of 400 km was chosen because of the problem that there are few firms in the tradable and non-tradable sectors in distance bins less than 300 km. Table 4.5 provides further details.

I run a regression where the distance treatment variable is the average firm driving distance to the improved highway to test the hypothesis that the effects of the improvements decrease the farther away from the highway a firm is located. The average distance of 519 km is used in the analysis because at any greater distance firms are too dispersed in the dataset. At a distance less than 519 km firms are clustered closer to the improved highway in average terms.[58] I expect coefficients in the regression using a distance bandwidth of 519 km to be smaller than those obtained with a distance bandwidth of 400 km, which is closer to the improved highway. The results, shown in Table 4.9, are generated from the sample that excludes Lusaka, which is the cleanest sample considered. Table 4.9 shows what happens when the driving distance is extended from 400 to 519 km under the firm-level approach and the sector-intensity approach. The direction of effects of the improved highway on tradable sector firms compared to non-tradable sector firms are the same for both treatments and both approaches. However, the size of estimates is smaller in the 519 km distance bandwidth. This shows that improved highway impacts reduce in size the farther away firms are located from it. Coefficients that are insignificant in Table 4.9 are interpreted as inconclusive because they are large. This shows that there is insufficient evidence to confidently conclude that the improved highway has no impact on firms in the tradable sector in comparison to those in the non-tradable sector for certain observed outcomes.

[58] The average distance of 519 km is the farthest distance that can be used without it being considered as too dispersed in the data. The average distance treatment variable will give the lowest estimate of the impact of the improved highway on firms considered not too dispersed in terms of the average distance in the dataset.

Table 4.9: Results Extending the Distance Band

	(1) Gross profit	(2) Sales	(3) Cost of Sales	(4) Employment Expense	(5) Freight Transport	(6) Fuel	(7) Raw Materials
Panel A: Firm Level Approach							
d400km*Post*Tradable	5.186*	4.445	4.262	4.684*	1.877	4.001*	-3.808*
	(1.968)	(1.301)	(1.575)	(0.909)	(2.420)	(0.733)	(2.856)
d519km*Post*Tradable	3.954*	3.603**	3.635*	4.146***	0.513	3.072**	-0.016
	(1.644)	(0.973)	(0.989)	(0.803)	(1.976)	(0.517)	(1.590)
Observations	1042	1042	1042	1042	1042	1042	1042
Panel B: Sector Intensity Approach							
d400km*Post*Importing	5.916*	5.246*	4.607	4.925**	2.238	3.965	-3.784*
	(2.031)	(0.956)	(1.185)	(0.940)	(2.454)	(0.860)	(2.890)
d519km*Post*Importing	4.172**	3.716*	3.411*	4.164***	0.827	2.593*	-0.142
	(1.787)	(1.129)	(1.025)	(0.925)	(1.977)	(0.814)	(1.525)
Observations	1085	1085	1045	1085	1085	1085	1085

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses and clustered at district level. Monte Carlo permutation test estimated pvalues to determine significance of an observed value of a test statistic by rearranging the order of the observed values of a variable. Cluster robust standard errors in parentheses. All variables except the treatment variable and the sector dummy variables are in inverse hyperbolic sine.

4.7 Conclusion

This essay investigated the impact of reduced transport costs that arise from upgrading an international highway on firms engaged in international trade in a developing country in Southern Africa. Specifically, the study analysed the impact of the improved highway on firms in the tradable sector in Zambia, a country reported as among the top-five countries in Southern Africa experiencing high transport costs because of its landlocked status and the extent of its poor-quality roads. A triple difference estimation strategy is used – the double difference in the tradable sector is compared to that in the non-tradable sector – to investigate firm outcomes of output, employment costs, transport and fuel costs, cost of sales and raw materials inventory costs. The difference-in-difference-in-differences estimation strategy controls for location-permanent unobserved differences, time-varying effects that impact the entire dataset equally and time-varying effects that affect groups distinctively. the Melitz (2003) trade model with heterogeneous firms is the theoretical framework used to explain the mechanisms through which reduced transport costs arising from improved highways affect firms in the tradable and the non-tradable sector.

Empirical results in this essay show that compared to those in the non-tradable sector, firms in the tradable sector experience positive growth in fuel and lubricant costs, gross profits, sales and employment expenses when the international highway is improved. The following explains these results. First, positive growth in gross profits and sales suggests the improved highway increases accessibility of new international output markets and cheap input markets for tradable sector firms. Second, the positive growth in fuel costs implies that firms in the tradable sector respond to reduced transport costs arising from the improved highway by increasing their utilisation of the transport infrastructure. This finding is similar to that in the first essay in Chapter 3 that compared to those further away, firms within 400 km of the improved highway experience a growth in fuel and lubricants that is 41,449 per cent (semi-elasticity of 414.49) higher after the road is upgraded. In this essay, compared to non-tradable sector firms, tradable-sector firms experience growth between 5,356 per cent and 5,172 per cent higher (semi-elasticities ranging between 53.65 and 51.72, respectively) after the highway is upgraded. Furthermore, the positive-growth effects reported in this chapter are only seen in the peripheral area; firms in those areas have access to only one international highway. As in the first essay in Chapter 3, the results in this chapter on firms' freight transport costs are inconclusive. The inconclusive results on freight transport costs in both chapters resonate with existing literature on Africa that transport costs on the continent associated with poor infrastructure are an insignificant portion of freight transport costs; rather, corruption, port inefficiencies and roadblocks inflate freight transport costs (Naude and Matthee, 2007; Sequeira, 2011). Thus, reductions in transport costs associated with the improved highway are neutralised by high transport costs associated with corruption and port and border inefficiencies.

This essay contributes to the literature on trade and transport costs in Africa. Unlike previous studies, it uses a within-country microeconomic approach to estimate improved highway effects on actual firm transport cost data, which is difficult to obtain. Evidence in this essay points to investment in highway infrastructure as a policy for promoting international trade in developing countries in Africa. Upgrading highways stimulates the performance of exporting firms in peripheral areas.

4.A Appendix: Figures

Figure 4.A.1: Map of the Nacala Corridor and the Distribution of Firms

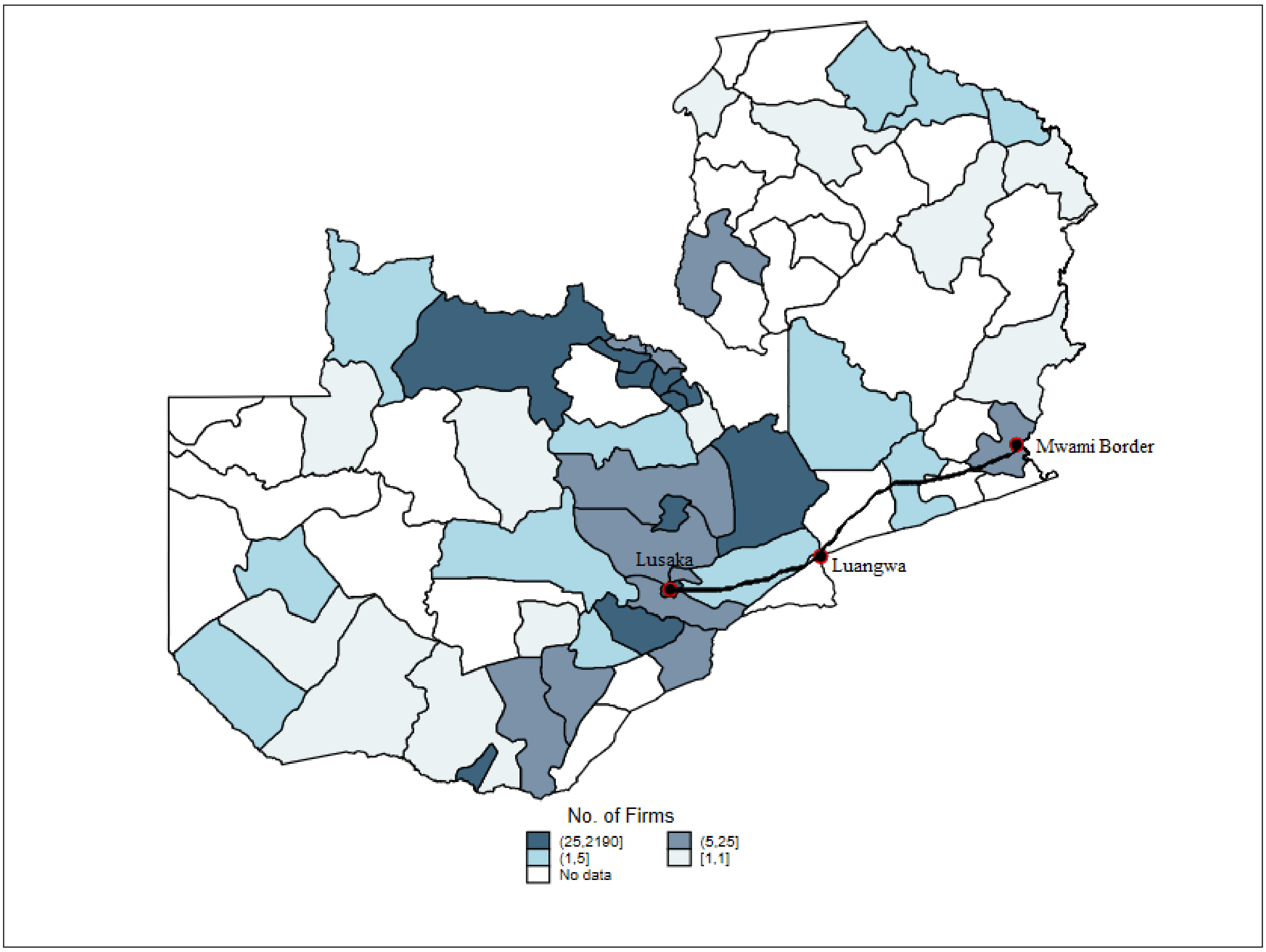


Figure 4.A.2: Parallel Trends by Sector

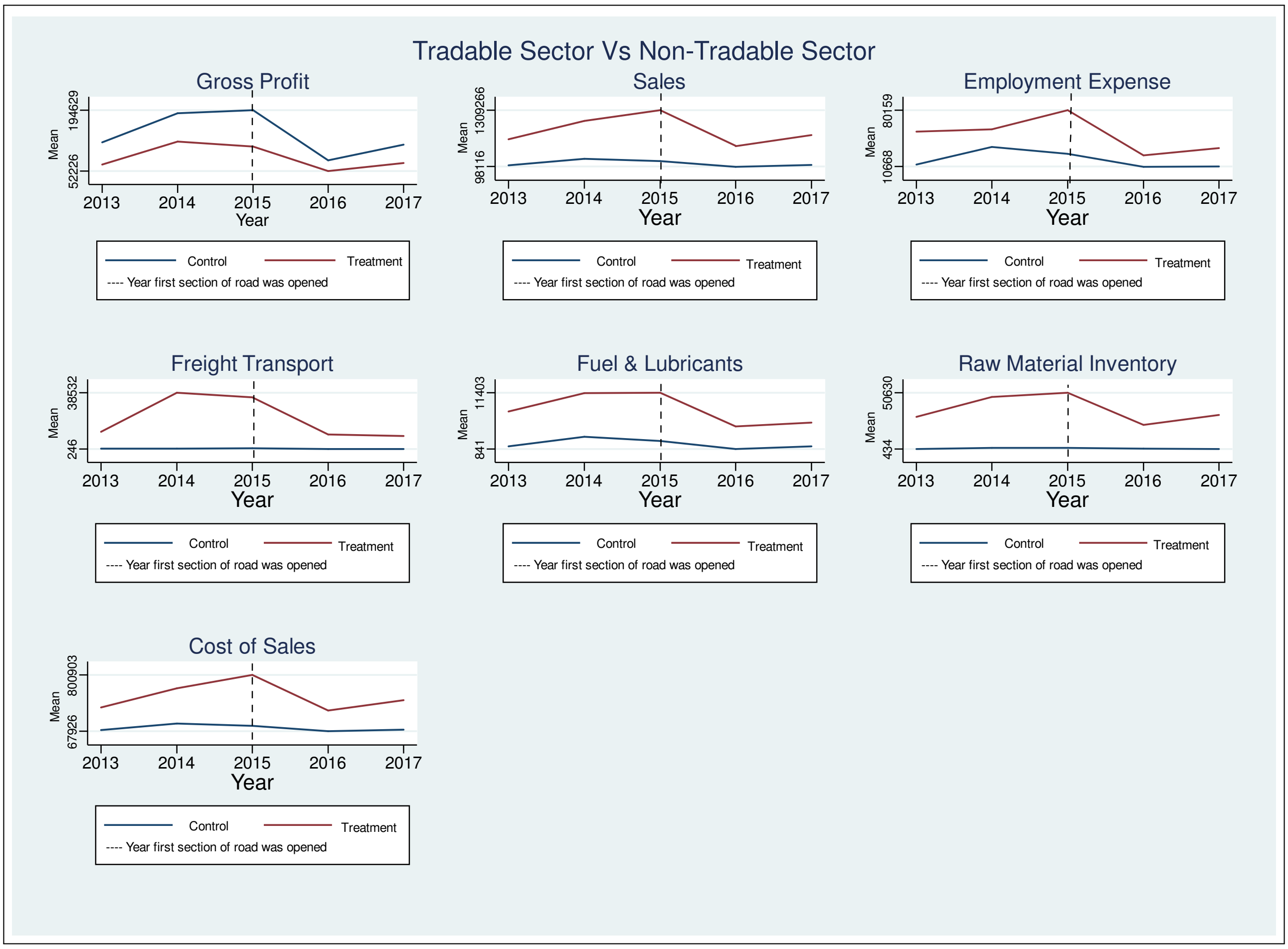
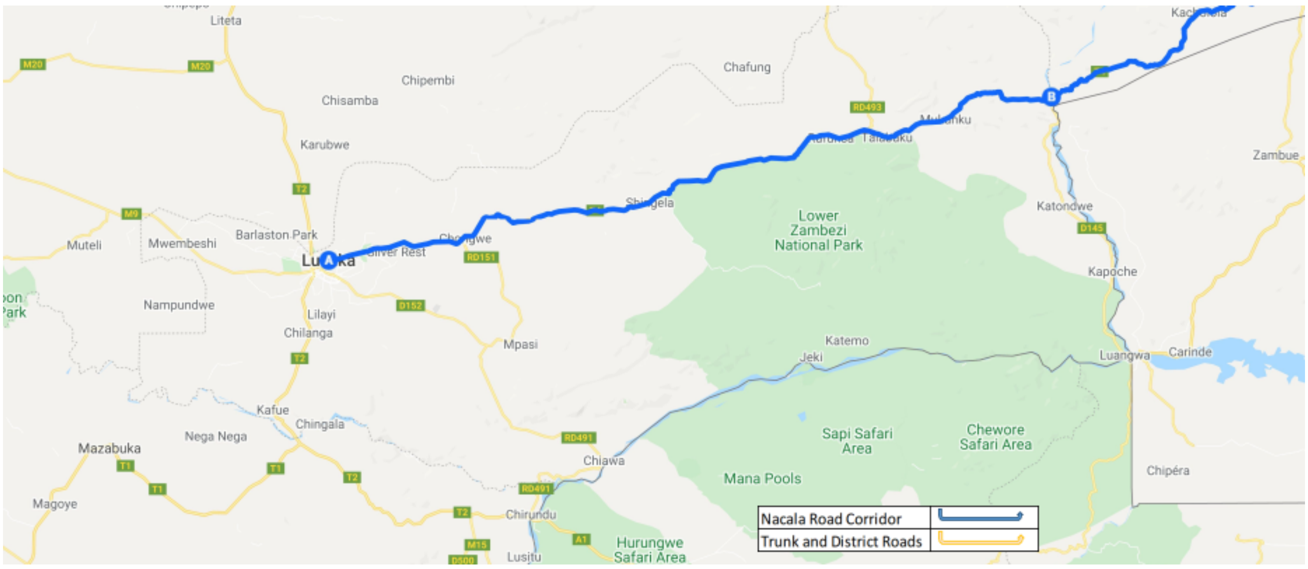


Figure 4.A.3: Lusaka Transport Link



4.B Appendix: Tables

Table 4.B.1: Attrition Test - Full Sample

Panel A: Attrition Probit Test

Probit regression		Number of obs	= 1808
		Wald chi2(9)	= 991.80
		Prob > chi2	= 0
Log pseudolikelihood	= -1186.5594	Pseudo R2	= 0.0407

(Std. Err. adjusted for 35 clusters in Location)

Variables (2013 Values)	Coef.	Robust Std. Err.	z	P>z
ihs Gross profit	-0.0004729	0.008148	-0.06	0.954
ihs Sales	-0.0407189	0.0138002	-2.95	0.003
ihs Tot. Employment Expense	-0.0255616	0.0050327	-5.08	0.000
ihs Freight Transport	-0.0297832	0.006763	-4.40	0.000
ihs Fuel & Lubricants	0.0074303	0.0057071	1.30	0.193
ihs Cost of Sales	0.0014923	0.007631	0.20	0.845
ihs CLS. Raw Materials	-0.0254581	0.0085954	-2.96	0.003
d400km_Rd	0.0681268	0.0415863	1.64	0.101
Constant	0.3442627	0.0528824	6.51	0.000
t-test				
chi2(9)	991.80			
Prob > chi2	0.0000			

Panel B: BGLW Pooling Tests

	Full Sample	
Variables (2013 Values)	F test	P-value
ihs Gross profit	1.35	0.2736
ihs Sales	2.71	0.0806
ihs Tot. Employment Expense	1.43	0.2538
ihs Freight Transport	1.91	0.1639
ihs Fuel & Lubricants	1.22	0.3085
ihs Cost of Sales	2	0.1507
ihs CLS. Raw Materials	0.94	0.4008

Table 4.B.2: Top 10 Exports

SN	Top 10 Exports	Industrial Class	4-digit ISIC	Contribution to Total Exports %	Missing in Firm Panel Data
1	Copper	Mining of metal ores	720	75.1	N
2	Gems, precious metals	Mining of other non-ferrous metal ores	729	1.2	N
3	Salt, sulphur, stone, cement	Quarrying of stone, sand and clay	810	2.1	N
		Mining of chemical and fertilizer minerals	891		Y
		Extraction of salt	893		Y
		Other mining and quarrying n.e.c.	899		N
4	Sugar, sugar confectionery	Manufacture of sugar	1072	1.4	N
		Manufacture of cocoa, chocolate and sugar confectionery	1073		N
5	Tobacco, manufactured substitutes	Manufacture of tobacco products	1200	1.2	Y
6	Inorganic chemicals	Manufacture of basic chemicals	2011	2.2	N
7	Stone, plaster, cement, asbestos	Manufacture of cement, lime and plaster	2394	1.2	Y
		Manufacture of articles of concrete, cement and plaster	2395		Y
		Cutting, shaping and finishing of stone	2396		Y
8	Other base metals	Manufacture of basic iron and steel	2410	1.3	N
		Manufacture of basic precious and other non-ferrous metals	2420		N
		Casting of iron and steel	2431		N
		Casting of non-ferrous metals	2432		Y
9	Machinery including computers	Manufacture of computers and peripheral equipment	2620	1.4	Y
10	Books, newspapers, pictures	Publishing of newspapers, journals and periodicals	5813	2.1	N
		Other publishing activities	5819		N
				89.2 %	

Table 4.B.3: Top 10 Imports

SN	Top 10 Imports	Division Name	2-digit ISIC	Contribution to Total Imports %
1	Machinery	Manufacture of machinery and equipment n.e.c.	28	14.7
2	Mineral fuels including oil	Mining of coal and lignite	5	14.2
3	Ores, slag, ash	Mining & Manufacturing	7,23	13.4
4	Vehicles	Manufacture of motor vehicles, trailers and semi-trailers	29	8.5
5	Electronic equipment	Manufacture of electrical equipment	27	6.2
6	Fertilizers	Manufacture of chemicals and chemical products	20	4.6
7	Iron or steel products	Manufacture of basic metals	24	3.8
8	Plastics	Manufacture of rubber and plastics products	22	3.3
9	Pharmaceuticals	Manufacture of pharmaceuticals, medicinal chemical and botanical products	21	2.9
10	Other Chemical goods	Manufacture of chemicals and chemical products	20	2.4
				74.0 %

Table 4.B.4: Sector Distribution of Firms

	No of Districts	No of Firms
Panel A: Sample Excluding Lusaka	**Treatment = Within 400km**	
Tradable - Manufacturing	7	29
Tradable- Mining	1	1
Tradable - Top 10	6	18
Tradable - Mining, Manufacturing & Top 10	7	30
Non- Tradable (Services)	10	84
Panel B: Full Sample		
Tradable - Manufacturing	8	201
Tradable- Mining	2	71
Tradable - Top 10	7	132
Tradable - Mining, Manufacturing & Top 10	8	275
Non- Tradable (Services)	11	1601
Panel C: Both Samples	**Panel B: Treatment = Outside 400km**	
Tradable - Manufacturing	13	105
Tradable- Mining	11	54
Tradable - Top 10	13	72
Tradable - Mining, Manufacturing & Top 10	15	159
Non- Tradable (Services)	27	769

Table 4.B.5: Rank Results - Transportation Costs

	(1)	(2)	(3)	(4)
	Transport	Fuel & Lubricants	Transport	Fuel & Lubricants
	Firm Level		Sector Intensity	
Panel A: Treatment is Rank 2 to 3				
Rank2_3*Post	0.875	1.509	0.488	2.146**
	(0.715)	(1.037)	(0.922)	(0.902)
Post*Tradable	4.725***	5.652***	4.376***	6.199***
	(1.136)	(1.754)	(1.149)	(1.468)
Rank2_3*Post*Tradable	-1.879	-1.291	-1.531	-1.477
	(1.136)	(1.754)	(1.149)	(1.468)
Observations	955	955	1002	1002
R^2	0.2823	0.2710	0.2401	0.2398
Few Treated Heterogeneous Clusters Robustness				
Monte Carlo (p-value)	0.3800	0.6120	0.4160	0.5920
	Firm Level		Sector Intensity	
Panel A: Treatment is Rank 4				
Rank4*Post	-0.137	2.224	-0.786	1.273
	(1.095)	(2.276)	(1.303)	(2.101)
Post* Tradable	1.597	-3.188	5.312	3.612
	(3.293)	(2.740)	(3.356)	(3.106)
Rank4*Post*Tradable	0.790	3.872	2.326	3.612*
	(3.314)	(2.239)	(3.040)	(1.828)
Observations	192	192	185	185
R2	0.7495	0.6393	0.7248	0.6196
Few Treated Heterogeneous Clusters Robustness				
Monte Carlo (p-value)	0.9560	0.5080	0.6540	0.5560

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses. All variables are in inverse hyperbolic sine. Results calculated on sample excluding Lusaka only. The definition of import status affects the composition of the control group (service firms that do not import or export). The firm level approach considers all firms with an average import purchase value greater than zero as importing. The sector intensity approach classifies all firms in industrial sectors with an import intensity greater than 10 percent as importing. All regression estimations include location and time fixed effects. I include an interaction term of the industry dummy and the post variable to control for differential trends by industry overtime in the treatment and control group in all regressions. To control for differential trends in districts of different sizes overtime in the treatment and control groups, I include an interaction term of district population and the post variable as a control in all regressions. Post is a dummy variable that equals 1 when the year is 2017 and 0 otherwise. Rank2_3 and Rank4 are dummy variables that equal 1 when a firm is located in a district where Port Nacala is ranked as the 2nd and 3rd or 4th nearest seaport and 0 if ranked 5th to 8th. Road driving distance from each district to various seaports is calculated using google maps with the point of origin being the district city council, post office or market. Tradable is a dummy variable that equals 1 when a firm belongs to the tradable sector and 0 otherwise. The interaction term Rank2_3*Post or Rank4*Post give the differential change of firms in the non-tradable sector with Rank 2_3 or Rank 4 compared to firms in the non-tradable sector with Rank from 5 to 8. The interaction term Post*Tradable gives the differential change in firms in the tradable sector compared to firms in the non-tradable sector overtime. The interaction term Rank*Post*Tradable gives the triple difference estimate. This coefficient measures the differential change in firms in the tradable sector relative to firms in the non-tradable sector over time and space. Monte Carlo permutation tests estimate p-values for the triple difference observed value of a test statistic by rearranging the order of the observed values of a variable.

Table 4.B.6: Rank Results - Firm Output and Labour Cost

	(1) Gross profit	(2) Sales	(3) Employment Expense	(4) Gross profit	(5) Sales	(6) Employment Expense
		Firm Level			Sector Intensity	
Panel A: Treatment is Rank 2 to 3						
Rank2_3*Post	0.307	0.340	-0.537	0.738	0.694	-0.482
	(1.716)	(1.624)	(1.051)	(1.857)	(1.879)	(1.042)
Post*Tradable	24.888***	2.676	3.260	25.384***	2.682	2.369
	(4.180)	(2.797)	(2.896)	(4.231)	(3.016)	(2.885)
Rank2_3*Post*Tradable	-2.977	-0.100	0.094	-3.776	-0.600	-0.099
	(4.180)	(2.797)	(2.896)	(4.231)	(3.016)	(2.885)
Observations	955	955	955	1002	1002	1002
R^2	0.3078	0.2890	0.2931	0.2783	0.2509	0.2544
Few Treated Heterogeneous Clusters Robustness						
Monte Carlo (p-value)	0.3900	0.9200	0.9840	0.2900	0.8100	0.9940
Panel B: Treatment is Rank4						
		Firm Level			Sector Intensity	
Rank4*Post	0.661	-2.425	-2.764	0.738	0.694	-0.482
	(5.255)	(2.619)	(1.770)	(1.857)	(1.879)	(1.042)
Post*Tradeable	5.620	-1.383	1.880	25.384***	2.682	2.369
	(5.601)	(2.238)	(1.929)	(4.231)	(3.016)	(2.885)
Rank4*Post*Tradable	4.504	5.513**	3.385	-3.776	-0.600	-0.099
	(6.076)	(2.592)	(2.204)	(4.231)	(3.016)	(2.885)
Observations	192	192	192	1002	1002	1002
R^2	0.5669	0.6943	0.6250	0.2783	0.2509	0.2544
Few Treated Heterogeneous Clusters Robustness						
Monte Carlo (p-value)	0.4580	0.3720	0.5300	0.4680	0.3220	0.4980

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses. All variables are in inverse hyperbolic sine. Results calculated on sample excluding Lusaka only. The definition of import status affects the composition of the control group (service firms that do not import or export). The firm level approach considers all firms with an average import purchase value greater than zero as importing. The sector intensity approach classifies all firms in industrial sectors with an import intensity greater than 10 percent as importing. All regression estimations include location and time fixed effects, in addition to an interaction term of the industry dummy and the post variable to control for differential trends by industry overtime in the treatment and control group. To control for differential trends in districts of different sizes overtime in the treatment and control groups, I include an interaction term of district population and the post variable as a control in all regressions. Post is a dummy variable that equals 1 when the year is 2017 and 0 otherwise. Rank2_3 and Rank4 are dummy variables that equal 1 when a firm is located in a district where Port Nacala is ranked as the 2nd and 3rd or 4th nearest seaport and 0 if ranked 5th to 8th. Road driving distance from each district to various seaports is calculated using google maps with the point of origin being the district city council, post office or market to the seaport administrative office. Tradable is a dummy variable that equals 1 when a firm belongs to the tradable sector and 0 otherwise. The interaction term Rank2_3*Post or Rank4*Post gives the time trend in the non-tradable sector over time. It gives the differential change of firms in the non-tradable sector with Rank 2_3 or Rank 4 compared to firms in the non-tradable sector with Rank 5 to 8 respectively. The interaction term Post*Tradable gives the differential change in firms in the tradable sector compared to firms in the non-tradable sector overtime. The interaction term Rank*Post*Tradable gives the triple difference estimate. This coefficient measures the differential change in firms in the tradable sector relative to firms in the non-tradable sector over time and space. Monte Carlo permutation tests estimate p-values for the triple difference observed value of a test statistic by rearranging the order of the observed values of a variable.

Table 4.B.7: Rank Results - Cost of Sales and Raw Materials

	(1) Cost of Sales	(2) Raw Materials	(3) Cost of Sales	(4) Raw Materials
	Firm Level		Sector Intensity	
Panel A: Treatment is Rank 2 to 3				
Rank2_3*Post	0.097	0.771	0.403	1.007
	(2.001)	(0.922)	(2.314)	(0.966)
Post*Tradable	-1.253	-2.205	-1.461	4.699
	(3.048)	(2.858)	(3.287)	(2.817)
Rank2_3=1*Post*Tradable	0.038	2.205	-0.242	2.195
	(3.048)	(2.858)	(3.287)	(2.817)
Observations	955	955	1002	1002
R^2	0.3102	0.3780	0.2800	0.3469
Few Treated Heterogeneous Clusters Robustness				
Monte Carlo (p-value)	0.9660	0.2500	0.9460	0.3160
	Firm Level		Sector Intensity	
Panel A: Treatment is Rank 4				
*Rank4*Post*	-3.748	-0.002	-3.317	-0.088
	(2.407)	(2.497)	(2.451)	(2.693)
*Post*Tradable*	-10.316^{***}	13.679^{***}	0.700	13.603^{***}
	(2.832)	(2.746)	(2.990)	(2.448)
*Rank4*Post*Tradable*	5.196	-1.768	5.664	-1.870
	(4.644)	(5.620)	(4.402)	(5.648)
Observations	192	192	185	185
R2	0.7101	0.7349	0.7284	0.7290
Few Treated Heterogeneous Clusters Robustness				
Monte Carlo (p-value)	0.4380	0.6320	0.3780	0.5760

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses. All variables are in inverse hyperbolic sine. Results calculated on sample excluding Lusaka only. The definition of import status affects the composition of the control group (service firms that do not import or export). The firm level approach considers all firms with an average import purchase value greater than zero as importing. The sector intensity approach classifies all firms in industrial sectors with an import intensity greater than 10 percent as importing. All regression estimations include location and time fixed effects, in addition to an interaction term of the industry dummy and the post variable to control for differential trends by industry overtime in the treatment and control group. To control for differential trends in districts of different sizes overtime in the treatment and control groups, I include an interaction term of district population and the post variable as a control in all regressions. Post is a dummy variable that equals 1 when the year is 2017 and 0 otherwise. Rank2_3 and Rank4 are dummy variables that equal 1 when a firm is located in a district where Port Nacala is ranked as the 2nd and 3rd or 4th nearest seaport and 0 if ranked 5th to 8th. Road driving distance from each district to various seaports is calculated using google maps with the point of origin being the district city council, post office or market to the seaport administrative office. Tradable is a dummy variable that equals 1 when a firm belongs to the tradable sector and 0 otherwise. The interaction term Rank2_3*Post or Rank4*Post gives the time trend in the non-tradable sector over time. It gives the differential change of firms in the non-tradable sector with Rank 2_3 or Rank 4 compared to firms in the non-tradable sector with Rank 5 to 8 respectively. The interaction term Post*Tradable gives the differential change in firms in the tradable sector compared to firms in the non-tradable sector overtime. The interaction term Rank*Post*Tradable gives the triple difference estimate. This coefficient measures the differential change in firms in the tradable sector relative to firms in the non-tradable sector over time and space. Monte Carlo permutation tests estimate p-values for the triple difference observed value of a test statistic by rearranging the order of the observed values of a variable.

5 Estimating the Impact of Upgraded Highways Using Global Nighttime Light Data in Zambia.

5.1 Introduction

Africa is home to six out of ten of the fastest growing economies in the world (IMF, 2019). Unfortunately, inadequate infrastructure limits the continent's economic growth potential (Lmao and Venables, 2001; Amjadi and Yeats, 1995). The region's poor transport infrastructure increases transportation costs and potentially depresses economic activity.[1] There are many studies that report that investing in transport infrastructure stimulates economic growth (Ghani et al., 2016; and Chandra and Thompson, 2000).[2] However, none empirically explore the effects of reduced transport costs induced by improved highways on economic growth measured at the local level in Africa. This is mainly because statistical data on economic indicators at a fine geographic level is scarce.[3] This hampers the robust empirical research important for policy making. To overcome such challenges, researchers have resorted to using satellite remote sensing data to study human economic activity as captured from outer space. One type of remote sensing data is nighttime light data. This data, captured by weather satellites and made available by the Earth Observation Group on a global scale, provides an efficient way to study human economic activity at any geographic level.[4] The data capture the artificial lights on earth at night and reflects the use of public and commercial lighting, which is strongly associated with the state of the economy and is an indicator of population and industry development (He et al., 2005; Deren and Xi, 2015).[5] Studies show that nighttime light data is correlated with economic activity.[6]

[1] Poor transport infrastructure accounts for 40 percent of the transport costs in coastal countries and 60 percent in landlocked countries (Lmao and Venables, 2001).

[2] Transport systems boost economic activity by; connecting previously unconnected areas to facilitate forward and backward linkages crucial for sustained economic growth, create employment which assists to alleviate poverty, and stimulate production and other agglomeration activities in the opened up areas (Gibbons et al., 2019).

[3] Economic data for subnational administrative regions such as provinces, districts or municipalities are unavailable for most developing countries or are of poor quality if they exist (Bruederle and Hodler, 2018).

[4] Nighttime light data is a useful substitute for economic indicators when data is missing or is of low quality in a study area (Chen and Nordhaus, 2011).

[5] As income rises, so does light usage per person, in both consumption activities and many investment activities (Henderson, Storeygard, & Weil, 2012). When household income increases, light usage also increases, as lighting is a normal good (Cuong and Noy, 2018).

[6] Mellander et al, 2015; Hodler and Raschky; 2014 and Henderson, Storeygard, & Weil, 2012.

Further, it is reported that nighttime light brightness reveals the level of economic development in a region more accurately than other estimation techniques that depend on limited data (Donaldson and Storeygard, 2016 and Lee, 2018).[7]

Based on the above, this essay estimates the impact of an upgraded highway on economic growth measured using nighttime light data in a landlocked country in Southern Africa. Conceptually, satellite sensors detecting nighttime lights will capture the economic growth driven by road infrastructure development by recording an increase in luminosity from new streetlights and lights from vehicles, homes and factories; these all reflect increased human economic activity in an area. In addition, countries at the early stages of development tend to build infrastructure that generates lights at night, such as buildings and roads. By contrast, countries at more advanced stages tend to focus more on technological innovation, which is less associated with lights at night (Hu and Yao, 2019). Given the above, nighttime light data is very useful in assessing and augmenting measures of real gross domestic product (GDP) in low and middle-income countries compared to high-income countries (Hu and Yao, 2019).

The aim of this essay is to investigate the research question, "do upgraded highways promote economic growth at a local level?" The spatial units of analysis used in this study are grid cells of 0.1 decimal degrees (approximately $11.1\,km^2$ at the equator). To capture the effect of the upgraded highway on nighttime lights at the grid cell level, I use two treatment variables. One is a dummy variable that takes the value one if a grid cell experiences reduced transport costs because it is located close to the intervened highway and zero if the grid cell is not exposed to reduced transport costs associated with the intervened highway because it is located near a highway not upgraded. Grids are grouped into distance bins or buffers around highways using straight line distance information. Only grid cells accessible using the existing road network are included in the analysis.[8] Second, I use a continuous treatment variable to measure the impact of the intervened highway on grided nighttime lights. This continuous variable uses the minimum travel cost data.[9]

Effects are identified based on the variation of nighttime light brightness within grid cells over time. I use the difference-in-differences (DID) estimation methodology with location and time fixed effects to address the location-specific endogeneous concerns of areas receiving upgraded highways being systematically different to those that do not and to capture location invariant time trends.[10] Further, the evidence provided in this essay relies on the identification assumption that the upgrade of the highway is not correlated to any pattern in nighttime lights at the grid cell level. I use a graph to check for parallel trends in the outcome variable in the control

[7] Obtaining accurate and up-to-date information on the spatial distribution of GDP is important to better understand a country's social and economic condition (Yue et al, 2014).

[8] I calculate the straight line distance from the centroid of each grid cell to its nearest highway link. This gives the straight-line least cost path from each grid cell to its nearest highway.

[9] I calculate the minimum travel cost from the centroid of each grid cell to the intervened highway using the shortest network path tool in QGIS.

[10] Highways are placed or upgraded in certain locations because of the existing economic conditions. Highways maybe improved in areas with high economic activity so as to service the increasing transport demands in those areas. At the same time, highways can be upgraded in areas lagging in economic activity so as to open up the area for business.

and treatment groups in the five years before the highway upgrade. As a robustness check, I conduct a panel event study that allows for the inspection of trends in the outcome variable in the treatment groups prior to treatment and to estimate the temporal effects of the improved highway during and after construction.

The results presented in this essay using the harmonised nighttime light data are robust to the two empirical specifications used in the analysis. These results suggest that after the highway is upgraded, grid cells close to the upgraded highway experience higher growth in light intensity than those close to highways not upgraded. After the highway is upgraded, the average growth in nighttime lights in grid cells within 10 km of the improved highway is 28.02 per cent in the DID regression and 27.38 per cent in the event study regression. Using the elasticity of gross cell product (GCP) with respect to nighttime light of 0.756, I approximate the effect of the improved highway to translate into grid-GDP growth of 21 per cent in grid cells located within 10 km of the improved highway after completion of the highway construction works.[11] Further, results show that larger impacts are concentrated in distance bandwidths within 10 km of the intervened highway and that a one-unit increase in the minimum travel cost from the intervened highway leads to a reduction in nighttime light at the grid cell level of 1.92 per cent. Findings in this essay have implications for policy as they suggest that upgraded highways promote economic growth at a local level.

This study contributes to the existing literature on the economic impacts of improved transport infrastructures in developing economies. It is the first in Africa to investigate the effects of changes in transport costs resulting from an improved highway on economic growth measured at a fine geographic level using nighttime light data. A similar study in the region by Storeygard (2016) analyses the effects of changes in transport costs induced by increases in world oil prices on economic activity measured using nighttime lights in coastal countries over a period of 17 years.[12] My essay differs from Storeygard (2016) as follows. First, I measure the effects of changes in transport costs associated with upgraded highways, whereas they analyse changes in transport costs induced by a variation in world oil prices. Second, I investigate local effects of upgraded highways by conducting my analysis at the grid cell level; this level goes beyond subnational administrative boundaries of cities, districts and wards. Third, I provide evidence of the effects on economic growth in a non-coastal developing country of reduced transport costs arising from the upgrade of a highway. Most studies focus on coastal countries in continents facing an economic geography very different to that in Africa[13] For example Mitnik et al. (2018) investigate the effects of rehabilitated primary roads on economic activity at the communal aand pixel level in Haiti using lights data, and Asher and Novosad (2018) use nighttime lights to

[11] Gross cell product is gross value added and it aggregates across all cells in a country to gross domestic product (Nordhaus et al, 2006). The 21 percent translated grid-GDP growth is the same using either the DID or the event study coefficients.

[12] Storeygard (2016) identifies transport costs by interacting distance to the port and oil prices. He uses Defense Meteorological Satellite Program - Operational Linescan System (DMSP-OLS) lights to proxy city-level-income in 15 countries in Southern Africa.

[13] Sub-Saharan Africa unlike other continents has a large number of landlocked countries (15) and poor quality road systems linking interior areas to the coast (Buys et al., 2010).

estimate the economic impact of new rural roads at a village level in India.

This chapter proceeds as follows. Section 5.2 discusses the related literature. Section 5.3 provides the conceptual framework of highways and economic growth. Section 5.4 describes the data and variables used in the study. Section 5.5 presents the methodology. Section 5.6 reports the main findings, and Section 5.7 summarises the results and concludes.

5.2 Related Literature

The use of nighttime light data in research increased after the development of new processing methods and the distribution of a digital archive by the National Oceanic and Atmospheric Administration in the 1990s and 2000s (Elvidge et al., 1997). In economics, the paper spiraling the use of nighttime light data is by Henderson, Storeygard, and Weil (2012). They use nighttime light data in the form of a cross-country lights panel dataset and find that lights are a very good proxy for GDP.[14] Studies subsequently mushroomed. Dai et al. (2017) study the suitability of the Defense Meteorological Satellite Program – Operational Linescan System (DMSP-OLS) and Visible Infrared Imaging Radiometer Suite (VIIRS) lights to estimate GDP at the provincial and city level in China. Mellander et al. (2015) use DMSP-OLS lights data to predict economic activity in Sweden, and Shi et al. (2014) use VIIRS data to estimate GDP and electric power consumption in provincial and prefecture units of mainland China. Nighttime light data has primarily been valuable in allowing research to be conducted in areas for which there is otherwise no data.

The use of nighttime light data in research is not restricted to the study of GDP for different geographic units.[15] Zhang and Seto (2011) and Elvidge et al (2012) use nighttime lights to study the pace and type of urbanisation. Townsend and Bruce (2010) use nighttime lights to measure electricity use and energy consumption, while Doll et al. (2000) use it to measure greenhouse gas emissions. There are, however, few studies in the literature that use nighttime light data to measure the economic impacts of developed transport infrastructure. These include Deng et al. (2020) who study differential impacts of high-speed railways at the station level in mainland China. Studies focusing on road development include Asher and Novosad (2018) on the impact of new rural roads on village economic outcomes in India, Mitnik et al. (2018) on the impact of rehabilitated primary roads in Haiti and Storeygard (2016) on the impact on economic growth in Sub-Saharan Africa of changes in transport costs. There is no existing study investigating the effects of reduced transport costs arising from improved highways on economic activity analysed at the grid level in Africa.

Notable facts in the literature using nighttime light data include the following. Nighttime

[14] Henderson, Storeygard, & Weil (2012) use a long difference methodology and finds that the lights-GDP relationship has a correlation coefficient of 0.3 in a sample of low and middle-income countries. In Sub-Saharan Africa lights explain about 20 percent of the variation in log GDP net of country and year fixed effects (Storeygard, 2016).

[15] Studies focusing on GDP include; Sutton et al. (2007), Florida et al. (2012) and Chen and Nordhaus (2011).

light data reflects real economic activities and is correlated with true GDP (Hu and Yao, 2019). It provides a robust indicator of economic activity, although the relationship seems to be statistically stronger for developing economies than developed ones (Mellander et al, 2015).[16] This is because countries with high real GDP per capita tend to have existing nighttime lights that is bright enough to reach the saturation level of satellite sensors, and thus, the data may not adequately reflect variations in economic activities (Hua and Yao, 2019). Nighttime light data is subject to measurement error; there may be a difference between the true light emanating into space and what the satellite records (Henderson, Storeygard, & Weil, 2012). Background noises must be removed, and thus there are studies that discuss the process of cleaning nighttime light data (Shi et al., 2014 and Elvidge et al., 1997). Lastly, nighttime light data is available globally and at a greater degree of geographic fineness than any standard data of income products (Henderson, Storeygard, & Weil, 2012).

5.3 Conceptual Framework

Highways have both economic and spatial effects. Economic effects are mainly triggered by transport cost reductions that arise after the highway is improved. Upgraded highways are hypothesised to affect economic growth through transport costs reductions. Since transport costs are an input in the production of goods and services, reduced transport costs increase firm profits by lowering the per-unit cost of output. Reduced transport costs also increase firm accessibility to cheaper input markets. In addition, when a highway is improved, competitive firms expand output and increase profit margins by accessing a wider, competitively priced output market. Furthermore, firm theory postulates that high industry profits in an area provide an incentive for new firms to enter the market provided there are no barriers to entry. Based on this, areas surrounding an upgraded highway are expected to experience a boom in economic activity and populace due to reduced transport costs and high firm profits. Further, the incomes of individuals living in the surrounding areas are also positively affected during and after construction. The positive income effects seen during the construction phase are linked to investments made locally by construction and engineering firms who employ local labour and purchase local building supplies, which have multiplicative effects on the economy (Rephann et al., 1994). Meanwhile, the positive income effects seen after the construction period arise mainly from growth in the demand for labour in the tertiary sectors of manufacturing, finance, real estate, transportation and public utilities (Rephann et al., 1994).

As stated earlier, the upgrade of a highway also generates spatial effects. Spatial effects may range from the project site level to the national level. At the project site, effects of the upgraded highway are experienced in the whole or a measurable portion of the project site. At the local level, the impacts of the upgraded highway extend to areas adjacent to the site, such as the neighbourhood, small town or environments up to 15 km away from the project site (EIS report,

[16] The elasticity steadily decreases as real GDP per capita increases, reflecting different developing modes at different stages of economic development (Hu and Yao,2019).

2010). At a regional level, impacts of the upgraded highway extend to outlying areas of the city and adjoining towns. National effects entail the impact of the upgraded highway extending to as far as national boundaries. In general, differences in highway impacts depend on the degree of openness of a particular location and its economic or industrial structure (Rephann et al.,1994).

5.4 Data and Variables

5.4.1 Nighttime Light Data

Traditionally, there are two kinds of nighttime light data, those obtained from DMSP-OLS sensors and the ones from the VIIRS sensors. The DMSP-OLS data is the early generation lights data and provides annual data covering the period 1992 to 2013. Each pixel of the DMSP-OLS light image is a 30 arc-second grid (about 1,000 m), and it is associated with a numerical value of radiance or "digital number" ranging from 0 to 63 (Hu and Yao, 2019).[17] The digital number 0 represents dark areas, while 63 indicates the brightest areas. Each digital number denotes an average for lights on all nights after sunlight, moonlight, aurora, bush fires, gas flares and clouds are removed algorithmically, leaving mostly man-made light from human settlements (Storeygard, 2016).[18] A follow up to the DMSP-OLS data is the VIIRS monthly data, which covers the period from April 2012 to date. The VIIRS data feature a higher spatial resolution of 15 arc-second grids which translates to about 500 m (Shi et al., 2014). VIIRS data does not have the issue of over-saturation that exists in the DMSP-OLS data, since it uses a wider radiometric detection range. VIIRS data employs on-board calibration (not available for the DMSP-OLS data) which increases the data quality, comparability and continuity (Shi et al., 2014). The monthly VIIRS Cloud Mask Product (VCM) average radiance data has been filtered to remove stray light but is not filtered to remove features associated with aurora, gas flares, fires, boats and background noise associated with the reflectance of light from bright surfaces (Dai et al. 2017). Light noises can limit the accuracy and reliability of the nighttime light data in predicting GDP, and therefore, confounding noises that are irrelevant to real economic activities must be removed (Dai et al., 2017). Shi et al. (2014) proposes a VIIRS data pre-processing procedure that is widely used in research.

The full-year VIIRS data available from 2013 to date is appropriate for use in this study as it is in sync with the road development project considered in the study, which was launched in 2013 and completed in 2017. However, since the VIIRS data is available only from 2013, there is no data to analyse pre-treatment trends to show that the control and treatment groups were not trending on different paths prior to the upgrade of the highway, which is comforting

[17] Radiance is defined as the electrical power (Watt) per steradian per area unit (nWcm−2 sr−1) (Bergs and Issa, 2018).

[18] More details on filtering process of extraneous features on DMSP-OLS lights is available at https://www.ngdc.noaa.gov/eog/gcv4_readme.txt

and reassuring in the interpretation of the DID estimates as causal. To overcome this, I use the harmonised global nighttime lights data, a new dataset made available by Li et al. (2020) which covers a longer time period from 1992 to 2018.[19] Li et al. (2020) harmonise the inter-calibrated DMSP-OLS data from 1992 to 2013 with the VIIRS data from 2014 to 2018.[20] The framework used by Li et al. (2020) to create the harmonised global nighttime light data consists of three major steps. In the first step, the composite cloud-free monthly VIIRS radiance data are aggregated to annual data and then noises from aurora, fires and boats are removed using a threshold approach. In the second step, VIIRS radiance data are spatially aggregated to the same radiometric resolution as DMSP data of 30 arc-second geographic grid using a kernel density approach. Next, the VIIRS data are converted to DMSP-like data using a sigmoid function. In the final step, the above steps are applied at the global scale to obtain the consistent global harmonised nighttime light data of calibrated DMSP data (1992–2013) and DMSP-like simulated data from VIIRS data (2014–2018). The advantage of using the harmonised global nighttime lights is that DMSP data have been calibrated, the DMSP-like data from VIIRS data have been filtered to remove stray light, temporal lights and other background noises, and the radiometric resolution of the VIIRS data has been adjusted to be the same as DMSP data. This makes the two datasets comparable and offers a consistent nighttime light dataset of a period covering 26 years. Table 5.1 shows the key specifications of the three types of nighttime light data.

Table 5.1: Types of Nighttime Light Data

	DMSP - OLS	VIIRS	Harmonized Global NTL
Time Span	1992-2013	2012-Present	1992-2018
Spatial Resolution	1km	500m	1km
Geographical Scope	Globe (75N/65S/180E/180W)	Globe (75N/65S/180E/180W)	Globe (75N/65S/180E/180W)
Radiation Resolution	6 bits	14 bits	6 bits
Calibration	No	on-board calibration	on-board calibration
Filtered to remove stray light, lunar illumination, lightning and cloud cover	Yes	Yes	Yes
Filtered to remove aurora and temporal lights from fires and boat lights etc	Yes	No	Yes
Satellite overpass time	9:00 PM	1:30 AM	9:00 PM

Note: DMSP- OLS is Defense Meteorological Satellite Program /Operational Linescan System. VIIRS is Visible Infrared Imaging Radiometer Suite. Earths Observation Group makes both datasets available at https://eogdata.mines.edu/download_dnb_composites.html. Li et al (2020) make harmonized Global nighttime lights data available at https://figshare.com/articles/Harmonization_of_DMSP_and_VIIRS_nighttime_light_data_from_1992-2018_at_the_global_scale/9828827/2

The nighttime light variable used in this study is transformed into inverse hyperbolic sine.

[19] Harmonized data is available at https://figshare.com/articles/Harmonization_of_DMSP_and_VIIRS_nighttime_light_data_from_1992-2018_at_the_global_scale/9828827/2.

[20] Li et al (2020) simulates the VIIRS data to be like the DMSP data to make the two datasets comparable and to provide a consistent nighttime light dataset covering a longer period.

This is done in order to make the data less skewed and to preserve numbers with zero values which logarithmic transformation does not do.[21] The IHS transformation is expressed as $ihs(y) = log(\sqrt{y^2+1}+y)$ Where $ihs(y)$ represents the transformed variable y. IHS transformed variables have similar interpretation to log transformed variables (Friedline, 2015; Mitnik et al, 2018).

5.4.2 Road Data

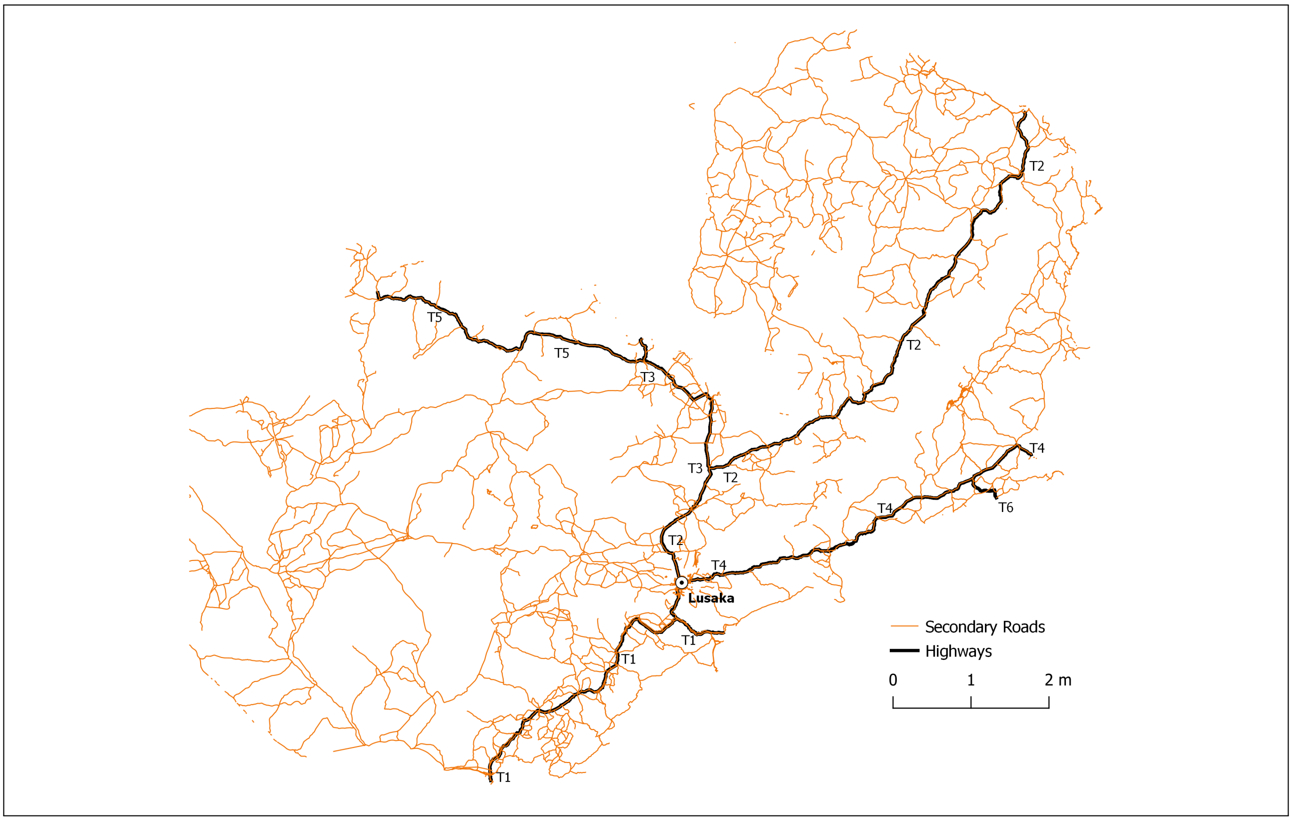
The information on the Nacala Road Corridor comes from the African Development Bank (AfDB) and the Road Development Agency (RDA) in Zambia. The geo-referenced information on the highway network in Zambia comes from the World Bank.[22] The information on secondary roads in Zambia is downloaded from Diva-GIS.[23] Using the two shapefiles of highways and secondary roads in Zambia, I am able to merge the road network information with the geo-referenced grided nighttime light data using the QGIS platform. Figure 5.1 shows the network of highways and secondary roads in Zambia.

[21] Friedline et al (2015) applied inverse hyperbolic sine (IHS) to wealth data to deal with skewness and to retain zero and negative values, which allows researchers to explore sensitive changes in the distribution and avoids stacking and disproportionate misrepresentation.

[22] The World Bank highway data on Zambia is available as at 2006. The vector line shapefile of highways in Zambia is downloaded from the World Bank at https://datacatalog.worldbank.org/dataset/zambia-roads [accessed on 12th July 2020].

[23] The secondary roads shapefile was downloaded at https://www.diva-gis.org/datadown [accessed on 27th November 2020]. Diva-GIS 7.5 version was developed by Robert Hijmans. DIVA-GIS offers information on administrative boundaries, roads, railways, altitude, land cover and population density for all countries in the world.

Figure 5.1: Road Network in Zambia



5.5 Methodology

5.5.1 Identification

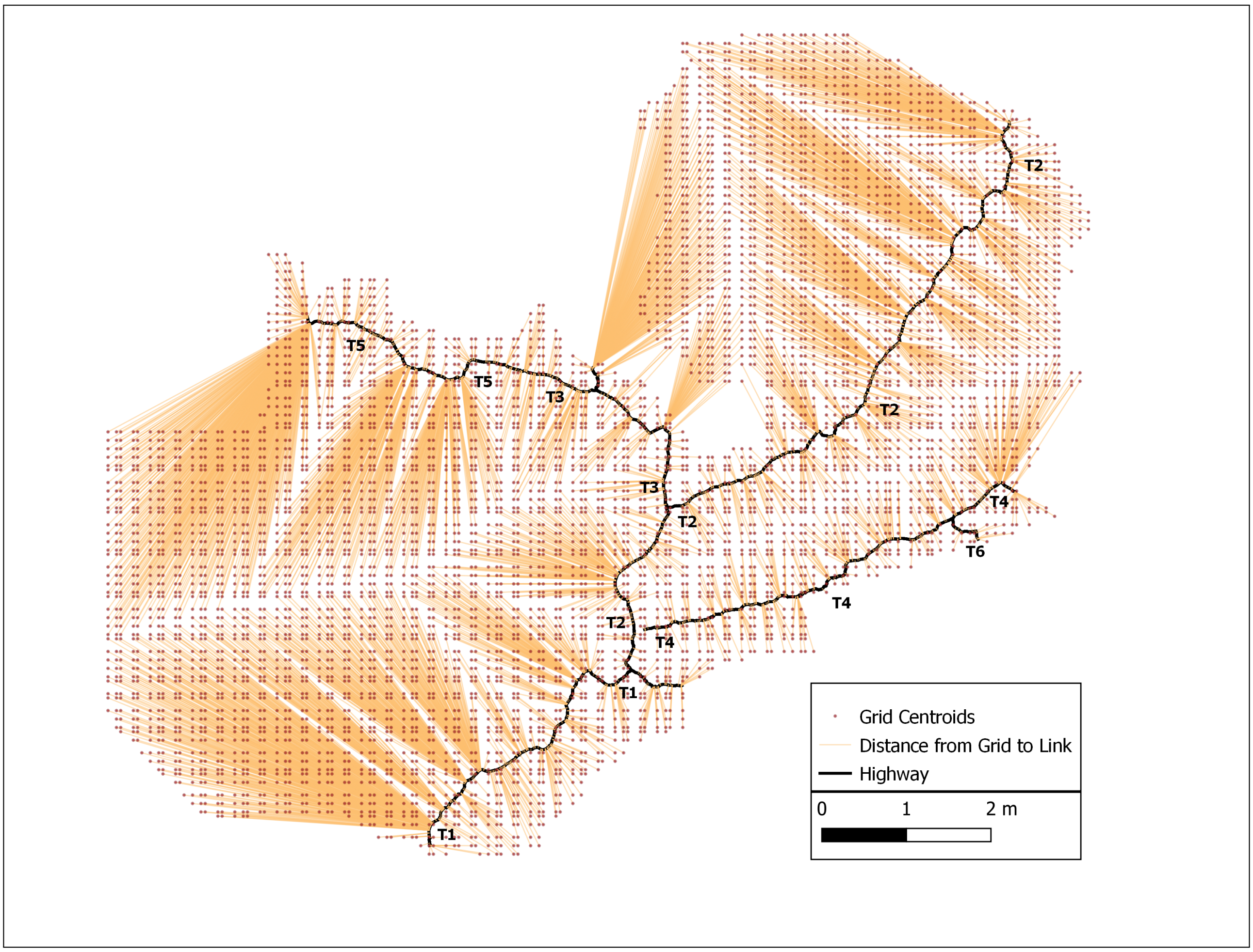
The aim is to identify the impact of the upgraded highway on nighttime lights measured at the grid-cell level. I investigate the impact using two treatment variables and two quasi-experimental approaches. The first treatment variable is a dummy variable that equals one when a grid cell is located within a specified distance band from the upgraded highway or zero if located equidistant to a highway that is not upgraded. I link the geo-referenced nighttime light data at the grid-cell level with the road network data in QGIS.[24] I calculate the shortest straight-line distance from the centroid of each grid cell to the nearest highway link.[25] Grids are grouped into distance buffers around their nearest highway using the straight line distance

[24] QGIS is used to derive the mean radiance light value in each grid cell. I clip nighttime light country raster information from the global nighttime light raster using a vector polygon of the country administrative boundary and create a mask layer. I then create grid cells of 0.1 decimal degrees on the nighttime light country mask layer. Using zonal statistics, I calculate the average radiance value (the total value of pixel lights divided by the pixel count in each grid cell) to get the light intensity per grid cell. Grid cells positioned outside the country administrative boundary are removed using the research processing tool and the geoprocessing tool in QGIS. The shape file for the map of Zambia comes from the database of Global Administrative Areas (GADM), which provides administrative shape files for all countries in the world. I download the country map of administrative boundary in "geopackage" format from https://gadm.org/data.html. This provides the vector polygon data used in QGIS. I download the nighttime light data in TIFF raster image format and this provides the raster data used in QGIS constituting the pixel values.

[25] This gives the straight-line least cost path from each grid cell to its nearest highway.

and only grid cells accessible using the road network are included in the sample.[26] Figure 5.2 shows the map of Zambia and the connection of grid centroids to the nearest highway link using the straight line distance.

Figure 5.2: Distance to Nearest Highway

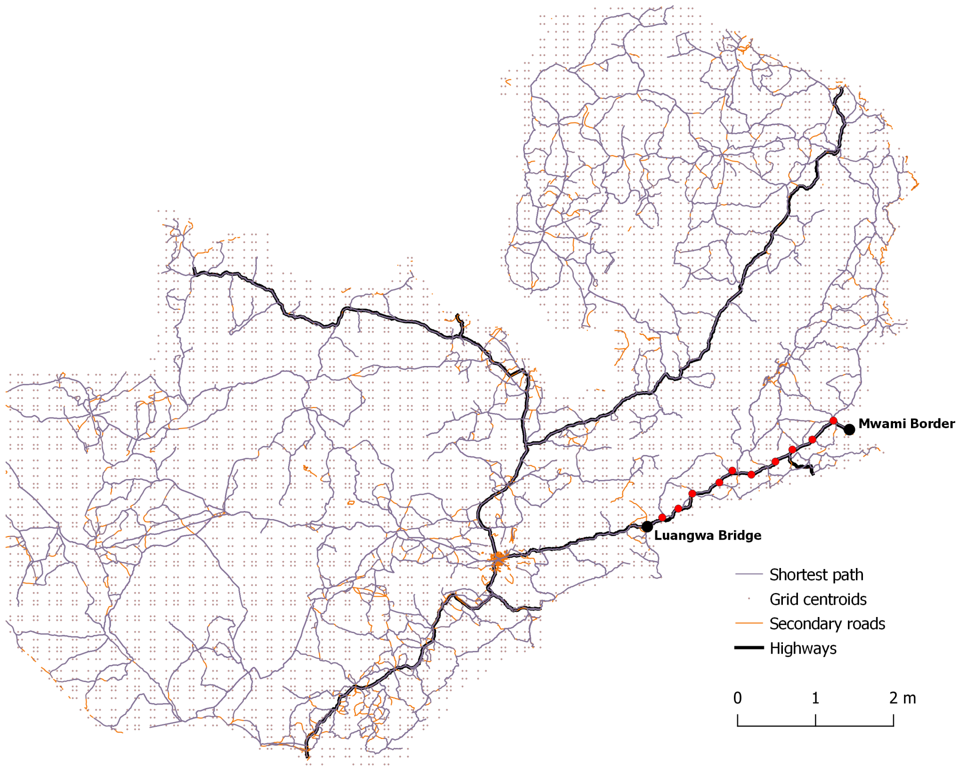


The second treatment variable is a continuous variable that measures the impact of the upgraded highway on a grid's nighttime lights using the minimum travel cost data. This continuous variable is derived by calculating the minimum travel cost from each grid centroid to the intervened highway.[27] Figure 5.3 is a map indicating the shortest path route analysis used to calculate the minimum travel cost to twelve random destination points on the improved highway from each grid centroid.

[26] To identify accessible grid cells using the road network, I overlay the road (secondary and highways) shapefiles in QGIS with the grid-lights centroids shapefile. Using the network analysis tool, I calculate the shortest distance path on the road network from each grid centroid to a point on the improved highway. Inaccessible grids using the road network unlike the accessible grids will show the minimum travel cost as blank from the grid centroid to any point on the improved highway. To group grid cells in 10 km distance bandwidths around a highway, I overlay the highway line shapefile with the vector shapefile of grided nighttime light data containing grid centroids. Using vector analysis command of distance to the nearest hub (line to hub) in QGIS, I calculate the nearest straight line distance from each grid centroid to the nearest highway link.

[27] The minimum travel cost is derived by calculating the shortest path (distance measured in layer units) from each grid centroid to 12 random destination points on the improved highway. Data for the lowest value of the minimum travel cost to the improved road is picked for each grid cell.

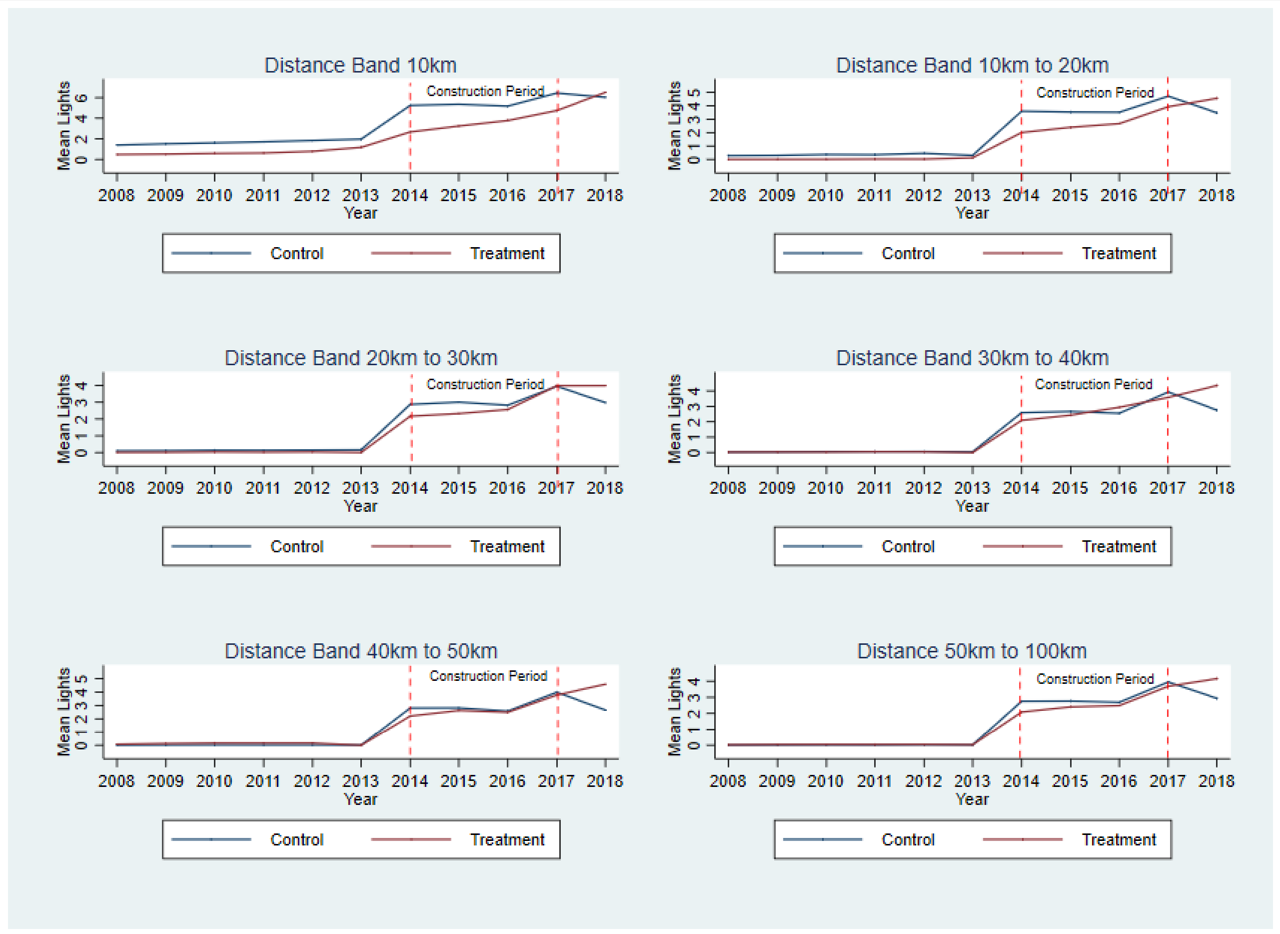
Figure 5.3: Shortest Path Route Analysis



The two quasi-experimental approaches used in this analysis are the DID and the panel event-study methodologies. Using the DID methodology, I compare differential changes in mean nighttime lights over time in grid cells located near the upgraded highway and those grid cells equidistant to highways not upgraded. Further, I estimate the effect of an increase in the minimum travel cost on the growth of nighttime lights at the grid level. The advantage of using the DID methodology is that the first difference controls for unobserved permanent differences in geographic areas, while the second difference controls for time trends (factors not related to the highway) that affect all geographic locations equally over time. In the DID regressions, I use panel nighttime light data for two periods, the years 2013 and 2018. The year 2013 is the pre-treatment period, while 2018 is the post-treatment period. The DID methodology provides unbiased casual estimates when the pre-treatment trends in the outcome variables do not differ in the control and treatment groups.[28] Given that the harmonised global nighttime lights data is available for a wider range of years, from 1992 to 2018, I am able to plot the average nighttime lights in grids by treatment group to check for pre-treatment trends. Figure 5.4 describes the trends in mean nighttime lights in the treatment and control groups in different distance bands pre-treatment.

[28] In the absence of treatment the difference between the 'treatment' and 'control' group must remain constant over time.

Figure 5.4: Parallel Trends in Outcome Variable



Inspection of Figure 5.4 shows that there is a drastic increase in mean nighttime lights from 2014. In the years before 2014, the DMSP-calibrated nighttime light data were used, while starting from 2014 to 2018, the DMSP-like data simulated from VIIRS data have been used. There is a mismatch in the lights trend in the two datasets for the following reasons. In the DMSP-calibrated data, unlike the DMSP-like data, low-light-intensity pixels are filtered away and set to zero (Henderson et al, 2012), with the maximum radiance light digital number capped at 63. Further, the simulated DMSP-like data have a larger range than the DMSP calibrated data. This is because of the aggregation procedure of converting VIIRS data from 15 to 30 arc-seconds using the kernel density method and because of the point-spread sigmoid function used to convert VIIRS data to DMSP-like data (Li et al., 2020)[29]. This leads to the large mismatch in light pixels between the two light-data types, as seen in Figure 5.4. To correct this, Li et al. (2020) exclude regions with low luminance using a threshold approach of 7, which brings the derived simulated DMSP-like data closer to the DMSP calibrated data in trend. When the threshold is increased to 20 and 30, most of the blooming effects are eliminated.[30] In this essay, I adopt a threshold approach of 6.5, which is the lowest digital number in the sigmoid function

[29] In the sigmoid function used to convert VIIRS data to DMSP-like data, the minimum digital number used was 6.5. See Li et al. (2020)

[30] See Li et al. (2020) for more information on testing the simulated DMSP-like VIIRS data against the DMSP data in the two overlapping years 2012 and 2013.

used by Li et al. (2020).[31] This gives a smoother continuous time series of the two light datasets. Figure 5.5 below shows the trends in nighttime lights using harmonised threshold corrected data. From 2008 to 2013, which is the last year before treatment, nighttime lights show parallel trends in all distance bandwidths.

Figure 5.5: Parallel Trends Threshold Corrected Data

Under the panel event study approach, I use the harmonised nighttime light data from 2008 to 2018. The event study methodology allows for the inspection of parallel trends in nighttime lights pre-treatment in the treated and control groups and, for the study of the temporal effects of the treatment during construction and post-construction. The existence of parallel trends in the outcome variable is a necessary condition that must hold for results to be interpreted as causal under this methodology.

A major issue in studies of road improvement projects is the endogeneity problem of improving

[31] The threshold approach of the digital number 7 was not used because of the following reason. The percentage of zeros in the 2013 calibrated DMSP data is 94 percent. However, when a threshold of 7 is applied to the 2018 simulated data the percentage of zeros increases to 96 percent in 2018. This entails the number of dark areas increase in 2018. Ideally, the number of dark areas should decrease or remain the same in future years provided there is no disaster like earthquakes or conflict that affects infrastructure and reduces the lights. In the country of study there was no negative shock to destroy the infrastructure and reduce lights during the study period. Thus, a threshold of 6.5 digital number is applied to the 2018 simulated data. When this threshold is used the percentage of zeros in the DMSP-like data drops to 91 percent. This suggests lit areas are increasing over time. Appendix table 5.B.1 shows zeros in the data per year.

roads only in locations with certain characteristics. In addition to controlling for this using location fixed effects in all estimations, the control group selected, like the treatment group, is equidistant in proximity to international highways; the only difference is that in the control group, highways did not receive an upgrade. This makes the comparison group very similar to the treatment group as compared to picking a control group located in a remote area with different economic, political and geographic conditions.[32]Another endogeneity concern is that of reverse causality. Source documents from RDA indicate that the rehabilitation of the Nacala Road Corridor is based on the deplorable state of the road and the need to have a reliable transport link connecting Zambia to a less decongested seaport at the shortest distance. In addition, the Zambian section of the Nacala Road Corridor is part of the roads built after independence in 1964. During this time, decisions on which roads to build were driven primarily by President Kaunda's "One Zambia One Nation" policy, which prioritised integrating the whole country (Raballand and Whitworth, 2011).[33] Therefore, the placement and improvement of the Nacala Road Corridor is not related to the variation in economic conditions at the grid cell level at baseline. Further, gridded data comprise cells fixed in time and space, insensitive to political boundaries and developments and completely exogenous to the likely features of interest such as spatial distribution of wealth and politics (Tollefsen et al., 2012).[34] In addition to the aforementioned factors driving road development in Zambia, grid cell data suggests some form of exogeneity in the road treatment variable. Another fact to consider is that the highway improvement project aimed to connect Lusaka in Zambia and Lilongwe in Malawi to a less congested seaport (Port Nacala) at the shortest distance for increased regional integration and economic growth in the Southern Africa Development Community region. Given this fact, areas located between these capital cities can be argued to be randomly included in the treatment since their inclusion is based solely on their location along the shortest path to the seaport. Therefore, I generate results on two samples, one in which grid cells located in Lusaka, the endogenously connected city, are excluded.[35] To further address the reverse-causality endogeneity problem, I conduct a Granger causality test to show that there is no systematic relationship between the highway improvement project and pre-treatment trends in nighttime lights in the grids studied. Granger causality testing checks that, conditional on location and year effects, past treatment predicts current outcomes while the future treatment does not (Angrist and Pischke, 2008).Event-study results also indicate that the growth in nighttime lights in grid cells near the improved highway occurs after the completion of the highway improvements and that prior to the intervention year, the control and treated groups

[32]Remote areas experience small transport cost reductions because of outlying proximity to highways, and thus are characterised by lower economic activity and higher poverty levels (Asher and Novosad, 2018).

[33]Thus, the resulting road network comprised both economically viable roads (connecting the main towns and to neighbouring countries) and low traffic volume roads designed to ensure all parts of the country were connected and to promote national unity (Raballand and Whitworth,2011).

[34]The stationary nature of the grid structure is a significant advantage, allowing for units of observation that are identical in shape and completely exogenous to the feature of interest (Tollefsen et al., 2012).

[35]To identify grid cells located in each district, I overlay the district shapefile with the nighttime lights grids with centroids shapefile in QGIS. Then using vector analysis I calculate the straightline distance from the centre of a grid cell to the centre of its nearest district. Appendix figure 5.A.1 shows the grids in each district of Zambia.

have parallel trends.

5.5.2 Empirical Specification

The empirical model below is used for the difference-in-differences estimation.

$$z_{jt} = \alpha_0 + \beta_1 Post_t + \beta_2 Distance_j + \beta_3 Distance_j * Post_t + \tau_j + \epsilon_{jt} \tag{3.C.1}$$

Here, z_{jt} is the mean annual lights in grid cell j at time t. α is a constant that indicates the mean nighttime lights at baseline in the control group. $Post_t$ is a dummy variable that equals one if the year is 2018 and zero otherwise, and τ_j are the grid fixed effects. $Distance_j$ is a dummy variable that equals one if grid cells are located within a specified distance band from the improved highway and zero if located equidistant to a highway that is not improved. In another specification, $Distance_j$ is a continuous variable representing the minimum travel time from each grid centroid to the upgraded highway, and ϵ_{jt} is the idiosyncratic error term capturing time-varying unobservable factors that affect nighttime lights. β_1 is the time trend in the control group and β_2 captures the differences between the control and treatment group at baseline.[36] I run multiple regressions using Equation (5.1) on different distance bandwidths.[37] Outcomes are compared in grid cells located near the upgraded highway in a specific distance band (grids within 10 km of it and those located in distance bandwidths greater than 10 km to 20 km, greater than 20 km to 30 km, greater than 40 km to 50 km and greater than 50 km to 100 km) with outcomes in grids in similar distance bands but located near a highway not upgraded. In all cases, the parameter of interest is β_3 which captures the differential change in mean lights in grid cells located in a specified distance bandwidth to the upgraded highway in comparison to grid cells located equidistant to highways not upgraded over time. In the specification where a continuous variable is used, β_3 captures the effect of a one unit increase in the minimum travel cost on nighttime lights at the grid cell level. All regressions include grid and time fixed effects. Time fixed effects control for factors that affect both the control and treatment groups equally, such as variation in worldwide satellite technology over time, changes in country economic conditions and country electricity-generation capacity. Grid fixed effects control for location-specific permanent unobserved differences. With grid fixed effects, relative variation in nighttime lights is identified entirely from within grid cells over time, after accounting for common shocks that affect all grid cells. Standard errors are clustered at the grid-cell level to account for potential within grid-cell serial correlation.

As stated earlier, I carry out reverse causality tests between the treatment variable and the outcome variable. Researchers can augment the standard DID regression model to include leading values of the treatment variable to examine the possibility that future treatment exposures are

[36]In all the regressions β_2 is omitted because it is collinear with a number of grid dummies.

[37]I run regressions on 6 distance bandwidths to understand the effect of the road in different geographic spaces. As seen in figure 2, Link T6 is a highway connecting to T4 an international highway. Grid cells within 17 km distance of T6 are taken as treated in the distance buffer of 50 km to 100 km around T4.

anticipated by current outcomes (Wing et al, 2018). If leads cause current outcomes, then the outcome happens before the treatment causing it. I use Equation (5.1) in the above empirical specification. But, I replace the dummy variable $Post_t$ with $TreatmentF$. $TreatmentF$ which is a dummy variable that equals one if grid cells in a specified distance band from the improved highway will receive the treatment in future and zero otherwise.[38] I expect the coefficient to be insignificant and close to zero. This indicates future treatment is not anticipated by current outcomes and that the growth in lights happens after treatment.

For the event study, I use the empirical model below.

$$z_{jt} = \alpha_0 + \sum_{n=2}^{N} \beta_n \left(lagn\right)_{jt} + \sum_{k=1}^{K} \gamma_k \left(Leadk\right)_{jt} + \lambda_t + \tau_j + \epsilon_{jt} \tag{3.C.2}$$

Lags and leads are dummy variables indicating the number of periods to treatment or past treatment in a grid cell located within a specified distance bandwidth to the improved highway at a specified time. In Equation (5.2), the first lag variable $n = 1$ is omitted to capture the baseline difference in outcomes between the treatment and control areas. Under this methodology, I use harmonised nighttime light data for the years 2008 to 2018. The baseline year is 2013 and 2014 is the event year. Grid cells connected to highways that are not upgraded comprise the counterfactual.[39] Lags and leads capture the difference in outcomes between treated and control grids in a year compared to the prevailing difference in the omitted baseline period (Clark and Schythe, 2020). In the event-study approach, the full set of event leads and lags allows for the inspection of parallel trends in the pre-treatment period and for the inspection of dynamic effects of the treatment. These indicate over time whether effects are growing, shrinking, transitory or permanent (Clark and Schythe, 2020).

To capture variation in transport costs induced by the improved highway in grid cells, I conduct the following. Firstly, I calculate the shortest route via the road network from each grid centroid to the nearest point on the improved highway. This allows for the identification of accessible grids using the road network. Thereafter, I calculate the shortest straight-line distance from each grid centroid to the nearest highway link. This allows me to group accessible grid cells into different distance buffers around their nearest highway link. Using these distance buffers, I estimate the average growth in nighttime lights in grid cells located in a specified distance bandwidth from the improved highway over time.

[38] Time series data is used covering the period 2008 to 2018. Grids to receive treatment after 2013 are coded 1 in years before treatment and 0 in years after treatment

[39] Grid cells in which the event never occurs are the pure controls and they have the eventtime variable left as blank in the data. Grid cells in which the event occurs (treated grids), have the eventtime variable set as 0 in the event occurrence year, set as negative numbers to indicate years to eventdate and positive numbers to indicate years post eventdate.

5.6 Results

5.6.1 Descriptive Statistics

Table 5.2 presents summary statistics for the lights data sample. The harmonised data has 936,144 pixels, and the spatial units of analysis are a total of 6,551 grid cells of 0.1 decimal degrees ($11\,km^2$).[40] Using the shortest path road network analysis tool in QGIS, 176 grid cells are inaccessible.[41] When the inaccessible grid cells are dropped from the full sample, 6,375 grid cells remain, which gives a total of 12,750 grid cell observations, as shown in Panel A of Table 5.2.[42] In Panel B, grid cells located in Lusaka have been removed from the sample to address the potential endogeneity problems associated with highway-improvement projects. Appendix Figure 5.A.1 shows a map of grid cells in each district in Zambia. Using the nearest neighbour analysis in QGIS, I connect each grid centroid to the centroid of its nearest district. I find that 38 grid cells are located in Lusaka. This translates to 76 observations that I drop in the full sample to retain 12,674 observations in the harmonised data sample, as shown in Panel B of Table 5.2.

Table 5.2 shows that the mean Euclidean distance is 134.68 km in the full sample and 135.39 km in the sample that excludes Lusaka. In both samples, the minimum travel cost measured in layer units from the improved section of the highway to the closest grid cell is 0.005, while the cost to the farthest grid cell is 10.78. The sample mean of the nighttime light digital number ranges between 0.469 and 0.406, with the minimum digital number being zero in both samples. The full sample has a maximum digital number of 60.29, while in the sample excluding Lusaka, the maximum digital number is 45.54. The maximum digital number is smaller in the sample that excludes Lusaka suggesting that none of the other cities is as industrialised as the capital.

[40] 6551 grid cells multiplied by 144 pixels in each grid cell equals 936,144 pixels.
[41] See figure 4.3 for a map indicating the shortest path route analysis.
[42] Full sample refers to the sample including Lusaka city.

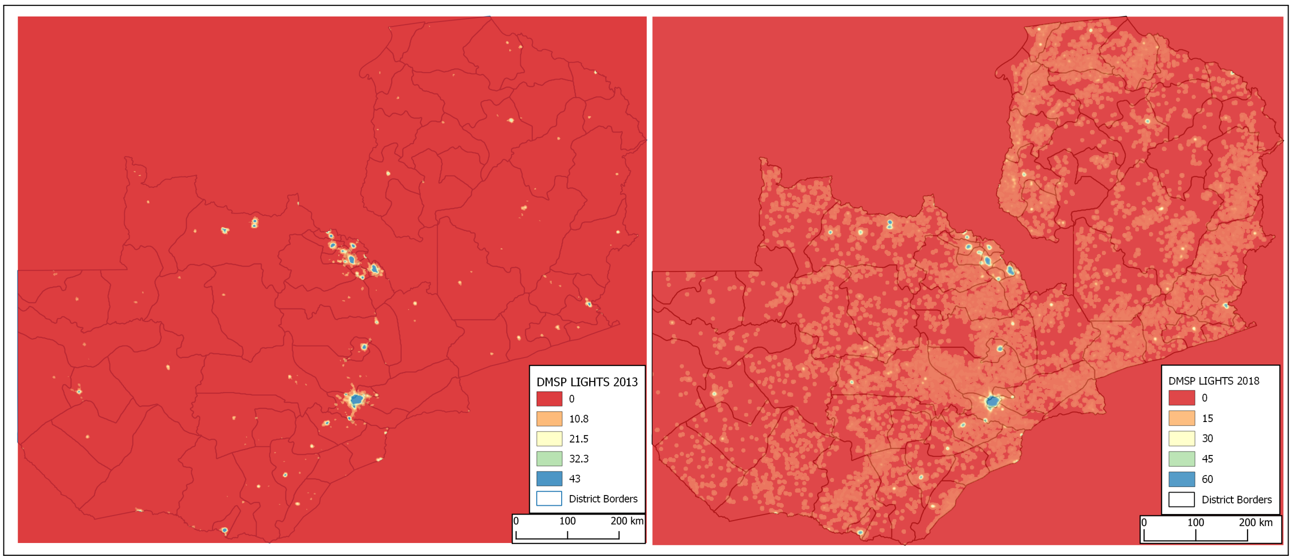
Table 5.2: Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Panel A: Full Sample					
Harmonized Data					
Minimum Travel Cost to Road	12,750	6.143	2.550	0.005	10.797
Euclidean Distance to Road	12,750	134.681	114.571	0.095	481.819
Mean Lights	12,750	0.469	2.497	0	60.292
ihs (Mean Lights)	12,750	0.152	0.619	0	4.792
Panel B: Sample Excluding Lusaka					
Harmonized Data					
Minimum Travel Cost to Road	12,674	6.166	2.539	0.005	10.797
Euclidean Distance to Road	12,674	135.393	114.541	0.095	481.819
Mean Lights	12,674	0.406	2.065	0	45.542
ihs (Mean Lights)	12,674	0.140	0.590	0	4.512

Note: There are 6551 grid cells of 0.1 decimal degrees (equivalent to 11.1 x11.1 km at the equator) in Zambia. Using network analysis in QGIS I remove 176 grid cells that are inaccessible. Minimum travel cost is the minimum shortest network path from a grid centroid to the improved highway. Euclidean distance to the road is the shortest straight-line distance from the centroid of a grid to its nearest highway. Mean lights indicate the sum of light pixels divided by the count of pixels in a grid cell. Each grid cell has 144 pixels. The harmonized data is cleaned - threshold corrected to remove outliers.

Figure 5.6 shows the nighttime lights in 2013 and 2018 in Zambia generated using the harmonised raw data. The images indicate that the nighttime glow increased in 2018. Similarly, Table 5.3 shows positive changes in nighttime lights between the period 2013 and 2018 generated using the harmonised threshold corrected data which I use to run different regressions in this essay.[43]

Figure 5.6: Map of changes in Nighttime Lights



[43]The harmonized data has been cleaned to remove background noises using the threshold approach of digital number of 6.5.

Table 5.3 shows that the change in mean lights between 2013 and 2018 in the different treatment and control groups is positive. The change in nighttime lights is significant in all distance bandwidths which suggests that all areas in the dataset experience an increase in mean light intensity over the period under review. Further, the positive changes in mean light intensity are larger in grid cells close to the upgraded highway compared to grid cells close to highways not upgraded in almost all the distance bandwidths.[44] Grid cells within 10 km of a highway experience the greatest change in mean nighttime light. These results are not causal but only show mere associations. The change in mean nighttime light reduces as the distance bandwidth is increased from the highway up to the 20– 30 km distance band; thereafter, changes fluctuate. In the sample that excludes Lusaka, the correlation coefficient on ihs_distance and ihs_mean lights is −0.157, while the correlation coefficient on levels of these variables is −0.086. Both these statistics are significant at the 10 per cent level.[45] This suggests that a grid cell closer to a highway link will have a greater nighttime glow.

Table 5.3: Changes in Nighttime Lights

	Grid Cells			
Distance	Treated	Control	Treated	Control
Panel A: Full Sample			Harmonized Threshold Corrected Data	
Dist 10km	89	369	2.016**	1.529***
Dist 10km_20km	86	349	1.221***	0.718***
Dist 20km_30km	72	310	0.279*	0.508***
Dist 30km_40km	68	294	0.496**	0.372***
Dist 40km_50km	53	265	1.044***	0.463***
Dist 50km_100km	178	1059	0.781***	0.528***
Panel B: Sample Excluding Lusaka			Harmonized Threshold Corrected Data	
Dist 10km	84	360	1.743***	1.389***
Dist 10km_20km	80	342	1.027***	0.618***
Dist 20km_30km	71	306	0.283*	0.442***
Dist 30km_40km	65	292	0.416**	0.336***
Dist 40km_50km	53	264	1.044***	0.464***
Dist 50km_100km	178	1059	0.781***	0.528***

Note: denotes significance at 90%, ** at 95% and *** at 99%. Results in the above table are calculated using the t test. The t test examines the significance of the within-group differences over time or conducts a matched pair test of equality of means before and after the intervention in each group. Grid cells are 0.1 decimal degrees which approximately is equivalent to 11.1 square metre at the equator. Harmonized threshold corrected data has all digital numbers 6.5 and less set to zero.

5.6.2 Empirical Results

5.6.2.1 Difference-in-Differences Results

Table 5.4 presents results from the two period DID specification shown in Equation (5.1). This specification examines the differential changes overtime in nighttime light intensity in grid cells near the upgraded highway compared to those near highways not upgraded. I run regressions

[44] The exception is in the distance bandwidth of 20 km to 30 km. Grid cells near the upgraded highway represent the treated group whilst those close to highways not upgraded are the control group.

[45] The correlation coefficient on ihs_distance and ihs_mean lights is -0.179 while the coefficient on the levels of these variables is -0.098 and both are significant at the 10 percent level in the sample including Lusaka.

on different distance bandwidths (within 10 km, greater than 10 km to 20 km, greater than 20 km to 30 km, greater than 30 km to 40 km, greater than 40 km to 50 km and greater than 50 km to 100 km) to observe spatial impacts of the treatment. All regressions include grid and time fixed effects with standard errors clustered at the grid cell level. Conceptually, grid cells close to the highway that receives an upgrade are expected to experience a large growth in nighttime lights due to increased economic activity in the area because of a fall in transport costs that arises when the highway is upgraded. Further, highway effects on nighttime lights are expected to fade as the distance bandwidth is increased from the upgraded highway since the benefits of reduced transport costs become smaller. Column 1 in Table 5.4 shows results generated using the harmonised threshold-corrected data in the sample excluding Lusaka. Grid cells nearest to the improved highway have the largest positive mean light coefficient of 0.247. This means that, on average, grid cells located within 10 km of the improved highway experience growth in nighttime lights after completion of the highway improvements that is 28.02 per cent higher than those located within 10 km of highways not upgraded.[46] When the distance bandwidth exceeds 10 km on both sides of the improved highway, the effect of the improved highway decreases. The effect eventually becomes statistically insignificant from the distance bandwidth of 20 km to 30 km outwards. Coefficients in the distance bandwidth of 40 km to 50 km are large and indicate that grid cells at this distance experience growth in lights that is, on average, 25 per cent higher than in grid cells equidistant to highways not upgraded, though weakly significant. Growth in this distance bin is weakly significant; however, it may signal the establishment of new residential housing and businesses in the area of 40 km to 50 km away from the upgraded highway. Areas outside the prime space of economic activity tend to offer cheaper establishments once economic activity increases in areas along the highway (Shiferaw et al., 2012). However, grid cells in the distance bandwidth of 50 km to 100 km experience insignificant growth effects associated with the upgraded highway as these areas experience high transport costs to prime areas of economic activity. Column 2 shows results generated from the sample that includes Lusaka. The results are similar to those generated from the sample that excludes Lusaka in Column 1 in terms of the direction, size and significance of the effect in each distance band. For example, results within the 10 km distance bandwidth in Column 2 show a positive coefficient of 0.232, which is also significant at the five per cent level, just like the coefficient of 0.247 in Column 1. Column 2 indicates that after the highway is upgraded, grid cells located within 10 km of the upgraded highway experience growth in nighttime lights that is 26.11 per cent higher compared to grid cells located equidistant to highways not upgraded. [47]

[46] When regression models have log transformed outcomes the impact of a one-unit change in a covariate (X) is calculated by exponentiating the coefficient, and subtracting one from the number (Mitnik et al., 2018). In this case $(exp\,(\beta_j) - 1) = exp\,(0.247) - 1 = 0.2801$. To give it a percentage interpretation multiply this result by 100. When the estimated coefficient is less than 0.10 the interpretation that a unit increase in X is associated with an average of $100 * \beta_j$ percent increase in Y works well (Mitnik et al., 2018). I refer to the exponetiated coefficients throughout the paper unless indicated otherwise.

[47] However, I do not emphasize results from the full sample because of concerns of including Lusaka the endogenously connected city.

In summary, the harmonised data suggest that grid cells near the improved highway experience higher growth in nighttime light intensity over time compared to grid cells near highways not improved after the highway project is completed. Further, these effects are largest within the 10 km distance bandwidth.

Table 5.4: DID Results

	(1) Mean Lights Harmonized Data	(2) Mean Lights Harmonized Data
Post*Distance 10km	0.247**	0.232**
	(0.115)	(0.110)
Observations	888	916
R^2	0.1299	0.1366
Post*Dist 10km_20km	0.190*	0.204*
	(0.108)	(0.108)
Observations	844	870
R^2	0.0823	0.0962
Post*Dist 20km_30km	-0.033	-0.053
	(0.070)	(0.069)
Observations	754	764
R^2	0.0507	0.0579
Post*Dist 30km_40km	0.051	0.073
	(0.084)	(0.088)
Observations	714	724
R^2	0.0467	0.0543
Post*Dist 40km_50km	0.226*	0.227*
	(0.136)	(0.136)
Observations	634	636
R^2	0.0937	0.0935
Post*Dist 50km_100km	0.093	0.093
	(0.067)	(0.067)
Observations	2474	2474
R^2	0.0773	0.0773
Grid Fixed Effects	*Yes*	*Yes*
Sample with Lusaka	*No*	*Yes*

Note: * p < 0.10, ** p < 0.05, *** p < 0.01. All variables are in inverse hyperbolic sine. Standard errors in parentheses and clustered at grid cell level. Grid cells are grouped into distance bandwidths using the straight-line distance calculated from each grid centroid to the nearest highway link. Only accessible grid cells using the network path are included. The control group in each regression constitutes grid cells equidistant to highways like grid cells in the treatment group. Post is a dummy variable that equals one when the year is 2018 and zero otherwise. The interaction term Post*Distance indicates the growth in mean nighttime lights for grid cells laying within a specified distance bandwidth from the upgraded highway compared to grid cells laying within the same distance bandwidth to highways not upgraded. The harmonized data used is threshold corrected using the 6.5 digital number. The sample used in columns 2 excludes Lusaka the endogenously connected city.

Table 5.5 shows the DID results generated using the continuous treatment variable of the minimum travel cost from each grid cell to the improved highway. Results are similar in the full sample and in the sample excluding Lusaka.[48] The results in Table 5.5 indicate that, in

[48]The expectation is that results in the sample which includes the endogenously chosen city of Lusaka, will have larger coefficients compared to the sample excluding Lusaka. Table 5.5 shows that the results larger in the sample including Lusaka.

both samples, an increase in the minimum travel cost from the upgraded highway results in a reduction in nighttime lights at the grid cell level. As indicated in Panel A, a one-unit increase in the minimum travel cost leads to a reduction in nighttime lights at the grid cell level of 1.9 per cent. The conclusion here is that there is an inverse relationship between distance and nighttime lights at the grid-cell level. The luminosity of nighttime lights reduces as the distance from the improved highway increases.

As stated earlier, I test if future treatment affects current outcomes. Results are small and insignificant, suggesting that nighttime lights do not Granger cause the highway improvements.[49]

Table 5.5: DID Results - Minimum Travel Cost Analysis

	(1) Mean Lights Harmonized Data
Panel A: Sample without Lusaka	
Post	0.292***
	(0.027)
Post* Min travel cost	-0.019***
	(0.004)
Grid Fixed Effects	Yes
Observations	12674
R^2	0.0687
Panel B: Sample with Lusaka	
Post	0.311***
	(0.027)
Post* Min travel cost	-0.022***
	(0.004)
Grid Fixed Effects	Yes
Observations	12750
R^2	0.0722

Note: * p < 0.10, ** p < 0.05, *** p < 0.01. All variables are in inverse hyperbolic sine. Standard errors in parentheses and clustered at grid level. Post is a dummy variable that equals one when the year is 2018 and zero otherwise. The interaction term Post*Min travel cost captures the effect of a unit increase in the minimum travel cost on mean lights in a grid cell over time. The harmonized data used is threshold corrected at the 6.5 digital number.

5.6.2.2 Event Study Results

I analyse an event study with six periods before and four periods after the upgrade of the highway using the harmonised threshold-corrected data from 2008 to 2018. The year 2014 is the event year, and 2013 is the baseline reference year. The lag dummy variable representing the baseline year is omitted in all regressions.[50] As indicated earlier, I run regressions on different distance bandwidths, with location and time fixed effects, to study the spatial effects of the

[49] See appendix 5.B.2 for these results.

[50] A single lag or lead variable is omitted to capture the baseline difference between areas where the event does and does not occur (Clark and Schythe, 2020).

improved highway. The event study methodology also allows me to observe the effects of the treatment during the highway construction period and after construction is completed. Further, under this estimation strategy, if mean nighttime lights were increasing in the treated grid cells prior to the upgrade of the highway, then the estimated coefficients on the dummy variables indicating the period prior to the upgrade of the highway would be positive and individually or jointly significant. If mean nighttime lights started increasing after the highway was improved, then dummy variables indicating the post-treatment period would be positive and significant individually or jointly.

Figure 5.7 displays the event study results generated using the empirical specification shown in Equation (5.2) on the sample excluding Lusaka[51] The horizontal axis shows the event time, the number of periods to (or post) the upgrade of the highway, whereas the vertical axis indicates the coefficient of the dependent variable, mean nighttime lights. Each dot in the graph represents the estimated coefficient at a 95 per cent confidence interval. Figure 5.7 shows that coefficients on dummies indicating the periods prior to the upgrade of the highway are close to zero and insignificant in almost all the distance bands except in the distance bin of 30 km to 40 km.[52] This indicates common trends in the outcome variable before treatment in all the distance bands except that of 30 km to 40 km.

[51] Appendix Figure 5.A.2 shows that the graphical results hardly change even when grid cells located in Lusaka the endogenously chosen city are added to the sample.

[52] Graphs are normalized so that pre-change effects are centred on zero.

Figure 5.7: Event Study Graph Results - Small Sample

Table 5.6 shows coefficients on the six periods prior to the treatment or event date and the four periods post the treatment. Results shown are on the sample that excludes Lusaka. lead0 is the dummy variable representing the event year which is 2014. One dummy variable, lag1 representing the baseline reference year 2013, is dropped in estimations to avoid a dummy trap. In the distance bins of 20 km and less, the coefficients on dummies indicating the time to treatment (lag6 to lag2) are all small and individually statistically insignificant, as shown in Columns 1 to 2. This speaks to a causal interpretation of coefficients in the distance bandwidths of 20 km and less from the upgraded highway. [53] Grid cells in the distance bandwidths greater than 30 km from the improved highway have inconsistent results on the lag dummy variables. For instance, in Column 4, coefficients on lag5 to lag3 are positive and strongly significant, while in Column 6, the coefficients on the same lag variables are negative and weakly significant. This suggests the absence of common trends in the outcome variable for grid cells in the treated and control groups, for five to three periods pre-treatment in distance bandwidths of 30 km to 40km and 50 km to 100 km. Results in Column 5 are similar to results in Columns 1 and 2; they indicate common trends in the outcome variable in all periods prior to the treatment year 2014 *(lead0)*. Results in Columns 1, 2 and 5 suggest that the difference in mean lights at the grid

[53] Coefficients on dummies indicating the time to treatment, lag4 to lag2 in the distance bin of 20 km to 30 km are also small and insignificant. However, Lag5 variable shows a weakly significant coefficient, but since this happens 5 periods before treatment, it is quite far to influence treatment.

cell level between the control and the treatment groups in each period prior to the event date, compared to the prevailing difference in the baseline year, has no statistical effect to cause the upgrade of the highway. This implies that the effect on mean nighttime light post-dated the highway upgrade and that prior to this intervention, the treated and control groups shared a common underlying trend.

The lead dummy variables in Column 5 indicate that during the highway construction period, grid cells in the distance bandwidth of 40 km to 50 km start to experience an increase in nighttime light that begins immediately in the year in which construction of the highway starts. This effect is shown by the dummy variable *lead0*, which has a coefficient of 0.182 and is significant at the five per cent level. Highway effects, shown by the variable *lead2*, last until the year before construction ends. Effects disappear from the year in which construction is completed, as indicated by coefficients on the dummies *lead3* (−0.65) and *lead4* (0.218), which are insignificant. The dummy variable *lead3* represents the year 2017, the completion year of improvement works. Temporal changes in nighttime light experienced during the construction phase only in grid cells located between 40 km to 50 km can be attributed to the establishment of bituminous hot mix plants and extraction sites of quarries and borrow pits during construction; these tend to be located away from settlement and road reserve areas.[54] This is done to avoid the excessive noise and dust from blasting activities impacting the quality of life of nearby communities.[55]

In the year of completion of construction works, grid cells within 10 km of the upgraded highway experience significant negative effects in nighttime light, as shown by the coefficient of −0.313 (negative growth of 26.88 per cent) in Column 1.[56] However, after the highway improvement works are completed, grid cells located within 10 km of the upgraded highway start to experience positive effects in nighttime light, as shown by coefficient 0.242 which indicates a growth of 27.38 per cent in lights in this distance bandwidth.[57] I interpret the negative effects as transitory effects of the upgraded highway in the 10 km distance bandwidth. A year after the completion of the highway improvement project, the effects of the upgraded highway become positive and

[54] Results also show that grid cells located less than 40 km from the upgraded highway experience largely insignificant changes in nighttime lights during the construction period.

[55] See the EIS report (2010) for more details on the environmental impact assessment conducted on the rehabilitation of the Nacala Road Corridor in Zambia on https://www.eib.org/attachments/pipeline/20090648_eis_en.pdf . In Zambia, quarrying activities are done in conformity to the Mine and Minerals Act, No. 31 of 1995 which states that preferably locations should not be less than 500 m away from the closest households, unless households are reallocated and fairly compensated.

[56] This effect is seen in the 50 km to 100 km distance bin but because the area shows that there is no common trends in the outcome variable, these estimates show associations. See column 6.

[57] According to the EIS report (2010), highway construction works in Chipata district required the separation of cyclist and pedestrian traffic from the mainline traffic. Thus, construction works involved the relocation of existing property or utility services such as electricity poles, water pipes, bill boards, a bus stop and hawkers stands which interrupted business in the area. Further, as works on the project started from Luangwa bridge and ended in Mwami border, works in Chipata district located just before Mwami border occurred in the final year of the highway upgrade. This may explain the reduction in nighttime lights attributed to the upgrade of the highway in grid cells located within 10 km (where chipata district is located) of the upgraded highway in the year of highway construction ends.

statistically significant at the five per cent level.[58] Further, Column 2 shows that grid cells adjacent to the 10 km distance bandwidth (grids located at a distance from the highway of between 10 km to 20 km) also experience positive spill-over effects of 0.178 one year after construction ends; this indicates a growth of 19.48 per cent after the highway is upgraded. The positive effects in nighttime light seen after the upgrade in the area immediately surrounding the upgraded highway and its adjacent communities are attributed to an influx of economic activity and people to the area because of transport-cost reductions.[59] This result is similar to the finding of Chandra and Thompson (2000), that total incomes in counties adjacent to highways increase after the opening of a new interstate highway.

In summary, after the highway construction project is completed, grid cells nearest to the upgraded highway (within 10km) and those adjacent to it (10 km to 20km) experience positive effects in nighttime light. The effects appear a year after completion of the highway improvement project, with the largest effects occurring in grid cells located within 10 km of the improved highway. During the highway construction period, grid cells in the distance band of 40 km to 50 km experience positive effects in nighttime light from the year the construction begins. However, these effects disappear in the year construction is completed. This suggests this distance bandwidth is used for setting up raw-materials extraction sites during construction – these must be established far away from settlement areas to protect the communities from noise pollution and dust from blasting activities. Further, in the year in which construction ends, grid cells within 10 km of the improved highway experience negative effects in nighttime light associated with the highway improvement project. However, these are only transitory and not permanent effects associated with the upgraded highway; they disappear in the next year. The results in this essay provide empirical evidence of spatial development patterns that can occur in a developing landlocked country in Southern Africa when an international highway is upgraded.

[58] The post construction effects cannot be classified as transitory because I am not able to observe effects exceeding 2018 (lead4). This is because the harmonized data is currently available only for the years 1992 to 2018.

[59] Transportation cost reductions associated with improved highway service tend to re-draw trade and service boundaries in favor of highway counties (Briggs 1980; Blum 1982)

Table 5.6: Event Study Results

	(1) Mean Lights 10km	(2) Mean Lights 10km to 20km	(3) Mean Lights 20km to 30km	(4) Mean Lights 30km to 40km	(5) Mean Lights 40km to 50km	(6) Mean Lights 50km to 100km
lag6	-0.055	0.077	0.046	0.028	0.038	-0.073
	(0.143)	(0.093)	(0.035)	(0.035)	(0.050)	(0.053)
lag5	-0.068	0.117	0.061*	0.103***	0.042	-0.100*
	(0.136)	(0.085)	(0.034)	(0.029)	(0.058)	(0.057)
lag4	-0.087	0.094	0.056	0.088***	0.043	-0.097*
	(0.134)	(0.085)	(0.036)	(0.028)	(0.055)	(0.057)
lag3	-0.113	0.111	0.050	0.104***	0.046	-0.098*
	(0.132)	(0.081)	(0.034)	(0.031)	(0.054)	(0.057)
lag2	-0.100	0.053	0.040	0.032	0.054	-0.077
	(0.143)	(0.089)	(0.034)	(0.034)	(0.047)	(0.052)
lead0	-0.161	0.026	-0.075	-0.005	0.182**	-0.089
	(0.150)	(0.110)	(0.047)	(0.045)	(0.085)	(0.056)
lead1	-0.092	0.067	-0.002	-0.019	0.122*	-0.039
	(0.154)	(0.105)	(0.048)	(0.041)	(0.069)	(0.057)
lead2	-0.039	-0.076	0.131*	0.058	0.147**	-0.060
	(0.161)	(0.099)	(0.077)	(0.056)	(0.068)	(0.067)
lead3	-0.313**	-0.126	-0.038	0.029	-0.065	-0.134**
	(0.147)	(0.108)	(0.088)	(0.077)	(0.085)	(0.066)
lead4	0.242**	0.178*	-0.088	0.074	0.218	0.037
	(0.112)	(0.106)	(0.058)	(0.097)	(0.134)	(0.067)
Observations	4831	4662	4125	4023	3532	13644
R^2	0.0866	0.0523	0.0476	0.0422	0.0704	0.0520

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses and clustered at grid level. All variables are in inverse hyperbolic sine. Regression estimation run on a sample that excludes Lusaka the endogenously connected city. Lag are dummy variables indicating the number of periods to treatment and lead are dummy variables indicating the number of periods post treatment. Lead0 is the dummy variable indicating the treatment event year which is 2014. Lag1 (the year 2013) variable is the baseline reference period and it is omitted in the regression to avoid a dummy trap. Lag1 captures the baseline difference between areas where the event does and does not occur.

5.6.3 Economic Interpretation of Empirical Results

In this section, I empirically test this essay's main assumption that nighttime lights are a good proxy for economic growth. I also provide an economic interpretation of the growth in lights in areas near the upgraded highway discussed in the previous section. The Demographic Health Survey (DHS) geospatial data for Zambia for 2013 and 2018 provides gross cell product (GCP) and nighttime light data at a grid cell level of $110\,km^2$.[60] GCP is gross value added, and it aggregates across all cells in a country to gross domestic product (Nordhaus et al.,2006).[61] Using the DHS dataset, I compute the elasticity between nighttime lights and GCP. I regress GCP

[60] DHS geospatial data is available at http://spatialdata.dhsprogram.com/home/ [accessed 18 January 2021].

[61] Gross value added is equal to total production of market goods and services in a region less purchases from businesses (Nordhaus et, 2006). See Nordhaus et (2006) for detailed information on the GCP.

on nighttime lights controlling for time and grid fixed effects, to quantify the GDP predictive effect of nighttime lights in Zambia at a grid-cell level.[62] I find that annual grid changes in GCP are correlated with changes in nighttime lights with an elasticity of approximately 0.756. I assume this elasticity also holds at the grid cell level of 11 km^2. Furthermore, after completion of the improvements, grid cells located in close proximity to the intervened highway experience growth in nighttime lights of 28.02 higher in the sample that excludes Lusaka. Therefore, I infer that the upgraded highway could have generated a growth in GCP that is 21 per cent higher in grid cells located within 10 km of the upgraded highway than those of the same distance to highways not improved.[63] Event study results also show that after the construction of the highway is completed, grid cells located within 10 km of the improved highway experience growth in nighttime lights of 27.38 per cent. This implies that the improved highway could also have generated growth in GDP at the grid cell level of 21 per cent.

5.7 Conclusion

This essay analysed the impact of upgrading a highway on economic growth measured at a local level of grid cells of $11km^2$ in Zambia a landlocked country in Southern Africa. As reported in the existing literature, nighttime light is correlated with human economic activity and are a good proxy of GDP (Henderson et al, 2012). Based on this fact, this essay used harmonised global nighttime light data to proxy GDP. Due to its wide time frame, this dataset's coverage allowed for the inspection of trends in nighttime lights in the pre-treatment period, an important requirement to treat as causal the coefficients in the DID and the event study quasi-experimental approaches. Graphs showing the trend in nighttime lights in the pre-treatment period in the control and treatment groups indicate common trends in all groups.

The results of the DID estimations presented in this chapter show that grid cells located within 10 km of the improved highway experience greater growth in light intensity than locations in similar in proximity to highways not upgraded. The growth in grid light intensity is 28.38 per cent in grid cells located within 10 km of the intervened highway. These causal effects are estimated to translate into grid-GCP growth of 21 per cent in grid cells located within 10 km of the improved highway after completion of construction works. The translated estimates are based on the annual correlation elasticity of GCP to nighttime lights of 0.756.

The essay also investigated the spatial effects of the upgraded highway in various distance bandwidths. Both the DID results and the event study results show that the impact of the upgraded highway are largest in grid cells located within 10 km of the intervened highway.The event study results also show that grid cells within 40 km to 50 km experience positive temporal effects that start appearing in the year of treatment and last only up to the year before completion of construction. The positive temporal effects suggest that, as required by Zambian

[62]The grid fixed effects use the cluster ID.

[63]DID results of 28.02 percent translate into 21 percent growth in grid-GDP in the sample excluding Lusaka.

law, material-extraction sites are set up far away from settlement areas to protect the communities from noise pollution and dust from blasting activities. Results showing the effect of the upgraded highway using minimum travel cost data reveal that a one-unit increase in the minimum travel cost from the improved highway leads to a 1.92 per cent reduction in nighttime light.

Results in this essay have important policy implications and contribute to informing policymakers that upgraded highways increase not only regional integration but also support positive economic growth at a local level.

5.A Appendix: Figures

Figure 5.A.1: Grids in a District

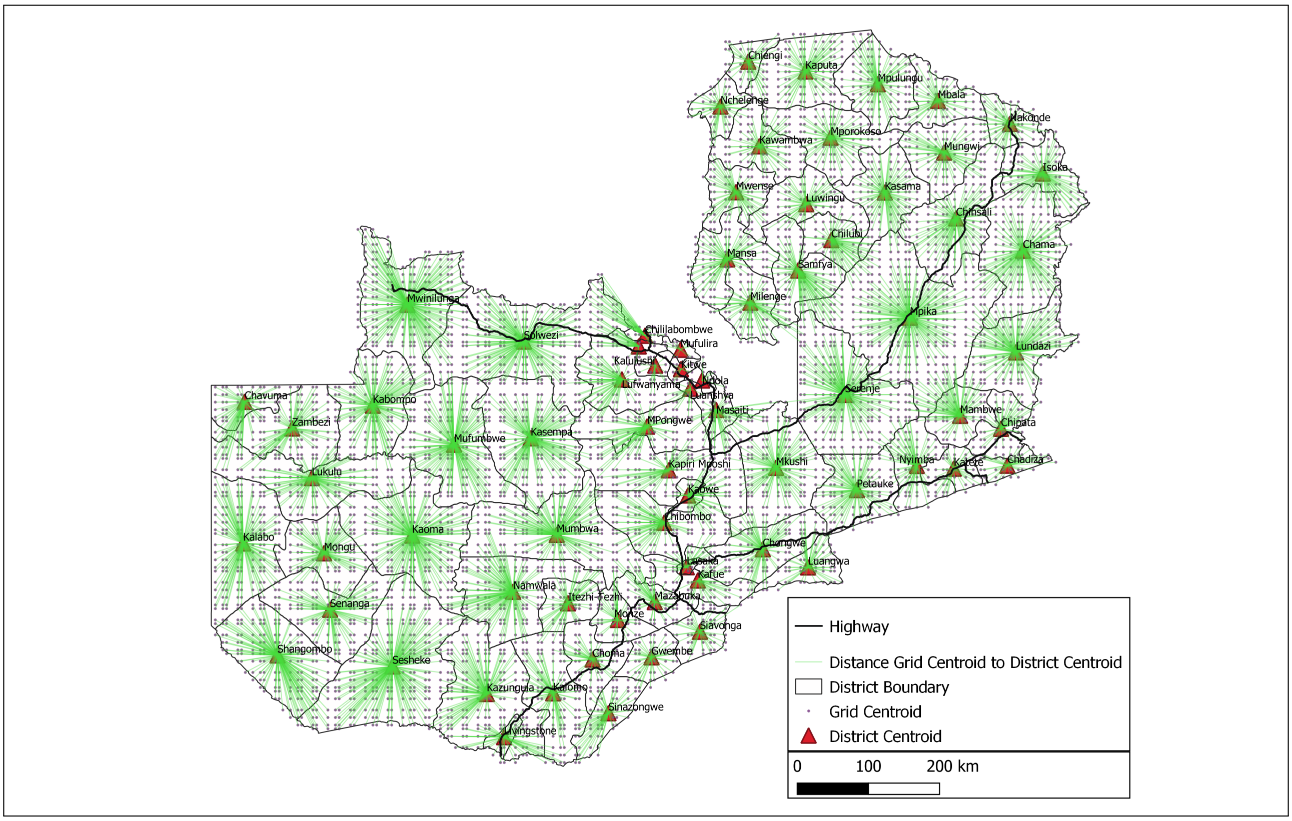
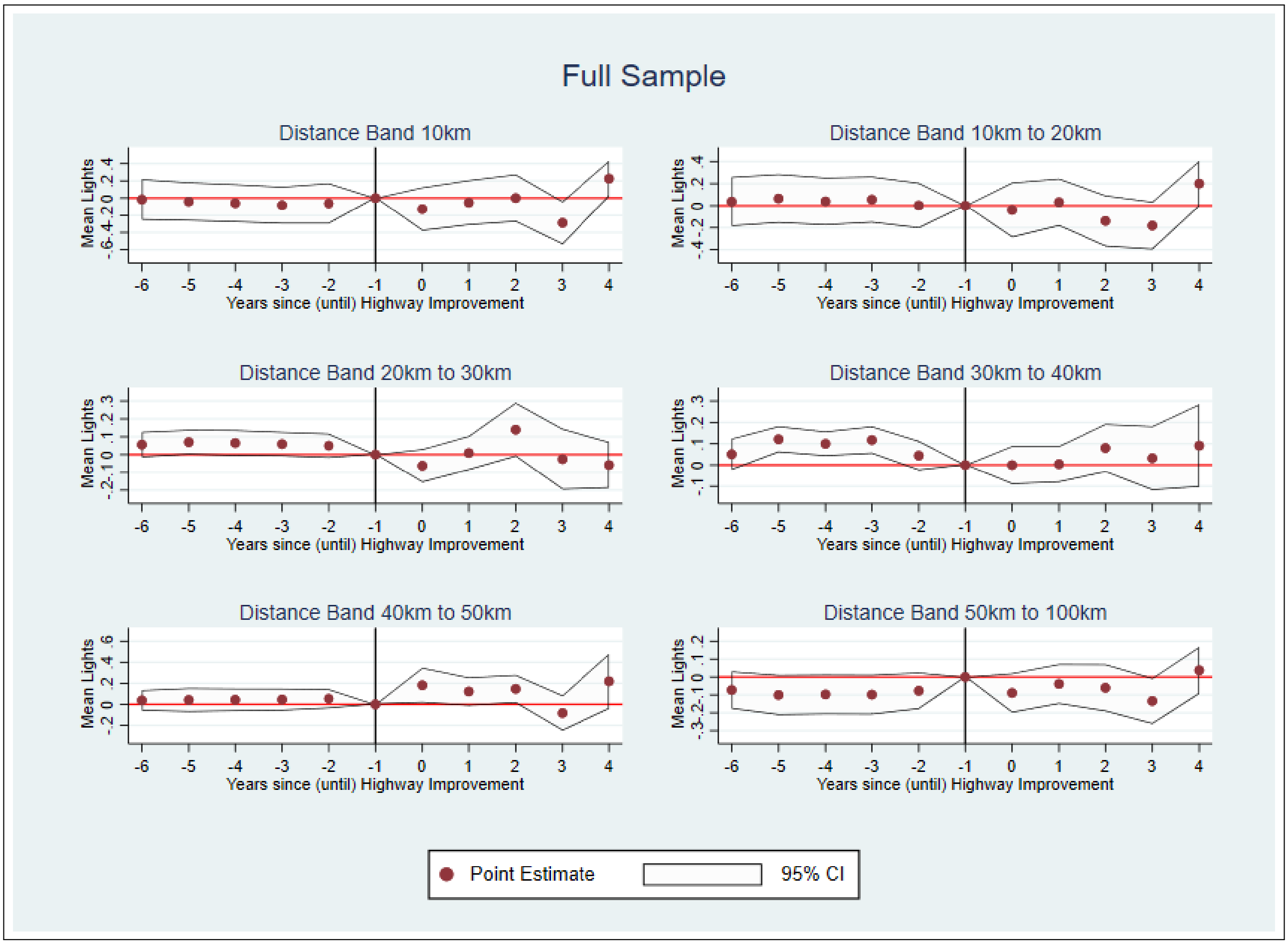


Figure 5.A.2: Event Study Graph Results - Full Sample



5.B Appendix: Tables

Table 5.B.1: Percent of Zeros in Data

Year	Obs per Year	Zero Obs	Percent of zeros	Data
		Harmonized Data		
2008	6551	6287	96%	Calibrated Data
2009	6551	6256	95%	Calibrated Data
2010	6551	6167	94%	Calibrated Data
2011	6551	6195	95%	Calibrated Data
2012	6551	6211	95%	Calibrated Data
2013	6551	6166	94%	Calibrated Data
2014	6551	1558	24%	Simulated DMSP like Data
2015	6551	1239	19%	Simulated DMSP like Data
2016	6551	1405	21%	Simulated DMSP like Data
2017	6551	579	9%	Simulated DMSP like Data
2018	6551	1451	22%	Simulated DMSP like Data
		Harmonized Data Corrected		
2014	6551	6418	98%	Threshold 7 DN
2015	6551	6403	98%	Threshold 7 DN
2016	6551	6371	97%	Threshold 7 DN
2017	6551	6348	97%	Threshold 7 DN
2018	6551	6302	96%	Threshold 7 DN
2014	6551	6278	96%	Threshold 6.5 DN
2015	6551	6282	96%	Threshold 6.5 DN
2016	6551	6220	95%	Threshold 6.5 DN
2017	6551	5931	91%	Threshold 6.5 DN
2018	6551	5991	91%	Threshold 6.5 DN
		VIIRS Data		
2013	6551	10	0.15%	VIIRS Data
2018	6551	0	0%	VIIRS Data

Table 5.B.2: Reverse Causality Test

	(1) Mean Lights	(2) Mean Lights	(3) Mean Lights	(4) Mean Lights	(5) Mean Lights	(6) Mean Lights
TreatmentF*Dist10km	-0.001					
	(0.043)					
TreatmentF*Dist10km_20km		0.046				
		(0.040)				
TreatmentF*Dist20km_30km			0.044			
			(0.034)			
TreatmentF*Dist30km_40km				0.016		
				(0.035)		
TreatmentF*Dist40km_50km					-0.087	
					(0.054)	
TreatmentF*Dist50km_100km						-0.026
						(0.032)
Observations	4967	4805	4208	4060	3552	13653
R^2	0.0538	0.0418	0.0255	0.0175	0.0308	0.0340

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All variables are in inverse hyperbolic sine. Standard errors in parentheses and clustered at the grid cell level. Grid cells are grouped into distance bandwidths using the straight-line distance calculated from each grid centroid to the nearest highway link. Only accessible grid cells using the network path are included. The control group in each regression constitutes grid cells equidistant to highways like grid cells in the treatment group. TreatmentF is a dummy variable that equals one when the location will receive an improved highway in future and zero otherwise. The interaction term TreatmentF*Dist km indicates the anticipated growth in mean lights for grid cells laying within a specified distance bandwidth of a highway to be improved in future compared to those grids laying within the same distance band to other highways not expected to be improved in future.

6 Conclusion

6.1 Main Findings and Policy Implications

According to the existing literature, a major challenge constraining economic development in Africa is its inadequate and poor infrastructure. Evidence shows that globally, Africa has the worst road transport infrastructure, despite roads being the main mode of transportation on the continent. Further, investment targeted to address Africa's road-infrastructure needs has increased over the last two decades, but there remain few studies that analyse the impacts on the continent of these investments. Existing studies compare the impacts of poor road infrastructure across countries in Africa, but none include a within-country analysis of highway improvements. The three essays in this thesis provide a better understanding of how investments in existing international highways impact within-country economic activity in landlocked economies in Africa. Each essay focuses on a specific context and provides interesting insights on this topic.

The first essay in Chapter 3, "*Improved Roads and Firm Performance: Evidence from Zambia*," focuses on two changes in accessibility effects that arise when the international highway is upgraded. The first is the effect on firms in 64 economic sectors and located in close proximity to the highway, of a fall in transport costs induced by the improved international highway. The second is the effect on firms in Zambia of a fall in transport costs to a trade seaport. The essay shows that compared to those farther away, firms in close proximity to the improved highway experience positive growth in fuel and lubricant costs with negative growth in assets after the highway is upgraded. The study also highlights the differential impacts on firm output, labour costs, freight transport costs and raw materials inventory in various distance bandwidths, all of which are inconclusive. Lastly, the essay shows that a fall in transport costs to a trading seaport has inconclusive effects on firm performance. This essay contributes to the literature by using a new district-referenced, disaggregated firm panel dataset from a country in Southern Africa to study the spatial effects of improved roads on economic activity. It provides knowledge that the short-run impacts of improved roads may not be easy to observe on firm economic activity across multiple sectors.

The second essay in Chapter 4, *"Improved Highways, Transportation Costs and Trade: Firm Panel Data Evidence from Zambia,"* investigates the impact on firms engaged in international trade of reduced transport costs induced by upgrading a section of an international highway leading to a trading seaport. This essay considers the possibility that spatial proximity to an improved highway may, over time, have distinct differential impacts on firms in the tradable sector as compared to those in the non-tradable sector. The results show that, relative to firms in the non-tradable sector, firms in the tradable sector outside the centre and close to the upgraded highway experience gains in trade of increased export output and reduced inventory costs as a result of the highway improvements. The upgraded highway has insignificant effects on firms in the centre who before the intervention are already star connected to other efficient transport links. Results in this essay provide important information for consideration by African countries in designing effective policy targeted at increasing economic activity in the tradable sector outside metropolitan cities in the short run.

The third and final essay in Chapter 5,*"Estimating the Impacts of Upgraded Highways Using Global Nighttime Light Data in Zambia,"* ," measures the impact of upgrading the international highway on GDP measured at a local level of grid cells of11 km^2. Due to the absence of statistical economic data at this fine geographic level, the essay used the harmonised nighttime light remote sensing data to proxy GDP. The essay finds that grid cells surrounding the upgraded highway experience positive growth in nighttime lights after the highway upgrade. I also estimated the GDP-light elasticity using GCP values in the DHS geospatial data to provide an economic interpretation of the growth in nighttime lights at the grid level cell in terms of GDP at that level. This essay provides insight on the impact of reduced transport costs on economic growth measured at a local level using nighttime light data.

6.2 Future Research

Some questions still remain regarding the impact of reduced transport costs arising from road infrastructure development in Africa. The first two essays analysed the impact of upgraded highways on firm panel data. It would be interesting to know the direct impacts of such investments on household consumption, expenditures and other socio-economic welfare indicators. This would increase understanding of how upgraded highways affect different economic agents in a country. Country demographic and health-survey data can be used to extend the research on this topic to understand the impact of upgraded highways on individual socio-economic welfare.

Further, the essays in this thesis investigated the short-term impacts of upgraded highways on a country in Africa. The literature remains unclear regarding the long-term impacts of upgraded highways in Africa.

Finally, there are different types of road infrastructure – rural roads, secondary roads, bridges and so forth – and thus, more studies are needed to provide insights into how different roads impact various economic agents. Transport policy aimed at improving roads may have different economic outcomes than policy aiming to develop new roads. Research in this area would provide insight into what investments in various types of road infrastructure in Africa entail.

Bibliography

Aaron, H. J. (1990). Discussion of why is infrastructure important. Is there a Shortfall in Public Capital Investment [Online], 51-63.

Adenikinju, A. F. (2005). Analysis of the cost of infrastructure failures in a developing economy: The case of the electricity sector in Nigeria. [Online].

Adukia, A., Asher, S. and Novosad, P. (2017). Educational investment responses to economic opportunity: evidence from Indian road construction. [Online].

Albarran, P., Carrasco, R., & Holl, A. (2013). Domestic transport infrastructure and firmsâ€™ export market participation. Small Business Economics, 40(4), 879-898.

Amjadi, A., & Yeats, A. (1995). Have transport costs contributed to the relative decline of african exports? some preliminary evidence. Washington DC, World Bank Working Paper, (1559).

Angrist, J. D. and Pischke, J. (2008). Mostly Harmless Econometrics: An Empiricist's Companion. [Online]. Princeton university press.

Aschauer, D. A. (1989a). Does public capital crowd out private capital?. Journal of Monetary Economics [Online], 24(2), 171-188.

Aschauer, D. A. (1989b). Is public expenditure productive?. Journal of Monetary Economics [Online], 23(2), 177-200.

Asher, S. and Novosad, P. (2018). Rural Roads and Local Economic Development. [Online]. The World Bank.

Ashiagbor, D., Deiana, R., Kappeler, A., Minsat, A., & Nguyen-Quoc, T. (2018). 9. financing infrastructure in africa. Banking in Africa, , 201.

Australian Bureau of Statistics (2010). Australia's Tradable Sector. Australian Economic Indicators. 1996. [Online].

Awokuse, T. O. (2008). Trade openness and economic growth: Is growth export-led or import-led? Applied Economics, 40(2), 161-173.

Bailly, H. and Brinckerhoff, P. (1999). California life-cycle benefit/cost analysis model (Cal-B/C). [Online].

Balassa, B. (1978). Exports and economic growth: Further evidence. Journal of Development Economics, 5(2), 181-189.

Barro, R. J. (1990). Government spending in a simple model of endogeneous growth. Journal of Political Economy [Online], 98(5, Part 2), S103-S125.

Baulch, B., & Quisumbing, A. (2011). Testing and adjusting for attrition in household panel data. CPRC Toolkit Note,

Becketti, S., Gould, W., Lillard, L., & Welch, F. (1988). The panel study of income dynamics after fourteen years: An evaluation. Journal of Labor Economics, 6(4), 472-492.

Bergs, R., PRAC, B. S. and Issa, M. (2018). What do night satellite images and small-scale grid data tell us about functional changes in the rural-ur-ban environment and the economy?. [Online].

Bhattacharya, A., Meltzer, J. P., Oppenheim, J., Qureshi, Z., & Stern, N. (2016). Delivering on sustainable infrastructure for better development and better climate.

Blum, U. (2002). Effects of Transportation Investments on Regional Growth: A Theoretical and Empirical Investigation in Transport Infrastructure. [Online].

Bonaglia, F., et al, M. (2000). Public capital and economic performance: evidence from Italy. Giornale Degli Economisti e Annali Di Economia [Online], 221-244.

Briggs, R. (1980). The Impact of Interstate Highway System on Non Metropolitan Growth [Online].

Bruederle, A. and Hodler, R. (2018). Nighttime lights as a proxy for human development at the local level. PloS One [Online], 13(9), e0202231.

Buys, P., Deichmann, U., & Wheeler, D. (2006). Road network upgrading and overland trade expansion in sub-saharan africa The World Bank.

Cameron, A. C. and Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. Journal of Human Resources [Online], 50(2), 317-372.

Canning, D. and Pedroni, P. (2004). The effect of infrastructure on long run economic growth. Harvard University [Online], 1-30.

Central Statistical Office. (CSO). (2013). 2010 Census of Population and Housing Population and Demographic Projections 2011 – 2035. www.zamstats.gov.zm.

Chandra, A. and Thompson, E. (2000). Does public infrastructure affect economic activity?: Evidence from the rural interstate road system. Regional Science and Urban Economics [Online], 30(4), 457-490.

Chen, X. and Nordhaus, W. D. (2011). Using luminosity data as a proxy for economic statistics. Proceedings of the National Academy of Sciences of the United States of America [Online], 108(21), 8589-8594.

Cheng, T. C., & Trivedi, P. K. (2015). Attrition bias in panel data: A sheep in wolf's clothing? A case study based on the MABEL survey. Health Economics, 24(9), 1101-1117.

Chissokho, L. and Seck, A. (2013). Electric Power Outages and the Productivity of Small and Medium Enterprises in Senegal, Investment Climate and Business Environment Research Fund (ICBE-RF) [Online].

Collier, P. (1995). The marginalization of africa. Int'l Lab.Rev., 134, 541.

Collier, P., & Gunning, J. W. (1999). Explaining african economic performance. Journal of Economic Literature, 37(1), 64-111.

Cuong, N. V. and Noy, I. (2018). Measuring the impact of insurance on urban recovery with light: The 2010-2011 New Zealand earthquakes. [Online].

Dai, Z., Hu, Y. and Zhao, G. (2017). The suitability of different nighttime light data for GDP estimation at different spatial scales and regional levels. Sustainability [Online], 9(2), 305.

Datta, S. (2012). The impact of improved roads on Indian firms. Journal of Development Economics [Online], 99(1), 46-57.

De Gregorio, J., Giovannini, A., & Wolf, H. C. (1994). International evidence on tradables and nontradables inflation. European Economic Review, 38(6), 1225-1244.

Deng, T., et al. (2020). What caused differential impacts on high-speed railway station area development? Evidence from global nighttime light data. Cities [Online], 97, 102568.

Deren, L. and Xi, L. (2015). An overview on data mining of nighttime light remote sensing. Acta Geodaetica Et Cartographica Sinica [Online], 44(6), 591.

Doll, C. H., Muller, J. and Elvidge, C. D. (2000). Night-time imagery as a tool for global mapping of socioeconomic parameters and greenhouse gas emissions. AMBIO: A Journal of the Human Environment [Online], 29(3), 157-162.

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

Donaldson, D. and Storeygard, A. (2016). The view from above: Applications of satellite data in economics. Journal of Economic Perspectives [Online], 30(4), 171-198.

Dorosh, P., et al. (2010). Crop production and road connectivity in Sub-Saharan Africa: a spatial analysis. [Online].

Duranton, G. (2015). Roads and trade in colombia. Economics of Transportation, 4(1-2), 16-36.

Duranton, G., Morrow, P. M. and Turner, M. A. (2014). Roads and Trade: Evidence from the US. Review of Economic Studies [Online], 81(2), 681-724.

Duranton, G., & Turner, M. A. (2012). Urban growth and transportation. Review of Economic Studies, 79(4), 1407-1440.

Esfahani, H. S. (1991). Exports, imports, and economic growth in semi-industrialized countries. Journal of Development Economics, 35(1), 93-116.

Escribano, A., Guasch, J. L. and Pena, J. (2010). Assessing the Impact of Infrastructure Quality on Firm Productivity in Africa: Cross-Country Comparisons Based on Investment Climate Surveys from 1999 to 2005. [Online]. The World Bank.

Escribano, A., et al. (2008). Impact of Infrastructure Constraints on Firm Productivity in Africa. Africa Infrastructure Country Diagnostic Working Paper.Washington, DC: World Bank [Online].

Elvidge, C. D., et al. (2012). The Night Light Development Index (NLDI): a spatially explicit measure of human development from satellite data. Social Geography [Online], 7(1), 23-35.

Elvidge, C. D., et al. (1997). Mapping city lights with nighttime data from the DMSP Operational Linescan System. Photogrammetric Engineering and Remote Sensing [Online], 63(6), 727-734.

Faber, B. (2014). Trade integration, market size, and industrialization: evidence from China's National Trunk road System. Review of Economic Studies [Online], 81(3), 1046-1070.

Fedderke, J. W. and Bogetic, Z. (2006). Infrastructure and Growth in South Africa: Direct and Indirect Productivity Impacts of 19 Infrastructure Measures. [Online]. The World Bank.

Ferman, B., & Pinto, C. (2019). Inference in differences-in-differences with few treated groups and heteroskedasticity. Review of Economics and Statistics, 101(3), 452-467.

Fieleke, N. S. (1995). The soaring trade in'nontradables.'(includes related article). New England Economic Review, , 25-37.

Fitzgerald, J., Gottschalk, P., & Moffitt, R. A. (1998). An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics,

Finn, M. G. (1993). Is all government capital productive?. [Online].

Florida, R., Mellander, C. and Gulden, T. (2012). Global metropolis: assessing economic activity in urban centers based on nighttime satellite images. The Professional Geographer [Online], 64(2), 178-187.

Foster, V. and Briceño-Garmendia, C. M. (2009). Africa's Infrastructure: A Time for Transformation. [Online]. The World Bank.

Friedline, T., Masa, R. D. and Chowa, G. A. (2015). Transforming wealth: Using the inverse hyperbolic sine (IHS) and splines to predict youth's math achievement. Social Science Research [Online], 49, 264-287.

Ghani, E., et al. (2016). road to success: The impact of the Golden Quadrilateral project for the location and performance of Indian manufacturing. The Economic Journal [Online], 126(591), 317-357.

Gibbons, S., Lyytikäinen, T., Overman, H. G., & Sanchis-Guarner, R. (2019). New road infrastructure: The effects on firms. Journal of Urban Economics, 110, 35-50.

Graham, D. J. (2007). Agglomeration, productivity and transport investment. Journal of Transport Economics and Policy (JTEP) [Online], 41(3), 317-343.

Gurara, D., Klyuev, V., Mwase, N., & Presbitero, A. F. (2018). Trends and challenges in infrastructure investment in developing countries. International Development Policy| Revue Internationale De Politique De Développement, 10(10.1)

Government of India. (2016). Basic Road Statistics of India 2013-14 and 2014-15, Delhi, India: Ministry of Road Transport & Highways.

He, C., et al. (2005). The urbanization model and process in Bohai Sea surrounding area in the 1990s by using DMSP/OLS data. Acta Geographica Sinica [Online], 60(3), 409-417.

Henderson, J. V., Storeygard, A. and Weil, D. N. (2012). Measuring economic growth from outer space. American Economic Review [Online], 102(2), 994-1028.

Hodler, R. and Raschky, P. A. (2014). Regional favoritism. The Quarterly Journal of Economics [Online], 129(2), 995-1033.

Holl, A. (2016). Highways and productivity in manufacturing firms. Journal of Urban Economics, 93, 131-151.

Hu, Y. and Yao, J. (2019). Illuminating Economic Growth. [Online]. International Monetary Fund. Lee, Y. S. (2018). International isolation and regional inequality: Evidence from sanctions on North Korea. Journal of Urban Economics [Online], 103, 34-51.

Hummels, D. L. (1999). Toward a geography of trade costs. Available at SSRN 160533.

International Monetary Fund. (2019). IMF Regional Economic Outlook. Sus-Saharan Africa. Navigating Uncertainty [Online].

Ivanova, E. and Masarova, J. (2013). Importance of road infrastructure in the economic development and competitiveness. Economics and Management [Online], 18(2), 263-274.

Jedwab, R. and Moradi, A. (2011). Transportation infrastructure and development in Ghana. [Online].

Jedwab, R. and Storeygard, A. (2017). The Average and Heterogeneous Effects of Transportation Investments: Evidencefromsub-SaharanAfrica1960-2010. [Online].

Juma, C. (2015). Infrastructure for innovation", New African, 26 May 2015. [Online]. http://newafricanmagazine.com/infrastructure-for-innovation/.

Kessides, C. (1993). The Contributions of Infrastructure to Economic Development: A Review of Experience and Policy Implications. [Online]. Vol. 213. World Bank Publications.

Leduc, S., & Wilson, D. (2013). Roads to prosperity or Fedderke, J. W. and Bogeticbridges to nowhere? theory and evidence on the impact of public infrastructure investment. NBER Macroeconomics Annual, 27(1), 89-142.

Lee, Y. S. (2018). International isolation and regional inequality: Evidence from sanctions on North Korea. Journal of Urban Economics [Online], 103, 34-51.

Li, X., et al. (2020). A harmonized global nighttime light dataset 1992–2018. Scientific Data [Online], 7(1), 1-9.

Limao, N. and Venables, A. J. (2001). Infrastructure, geographical disadvantage, transport costs, and trade. The World Bank Economic Review [Online], 15(3), 451-479.

Lombardo, G., & Ravenna, F. (2012). The size of the tradable and non-tradable sectors: Evidence from inputâ€“output tables for 25 countries. Economics Letters, 116(3), 558-561.

Macchi, P. and Sequeira, S. (2009). Soft versus hard infrastructure in transport. Afrique Contemporaine [Online], 230, 69-82.

MacKinnon, J. G., & Webb, M. D. (2018). The wild bootstrap for few (treated) clusters. The Econometrics Journal, 21(2), 114-135.

Mano, R., & Castillo, M. (2015). The level of productivity in traded and non-traded sectors for a large panel of countries International Monetary Fund.

Martincus, C. V., Carballo, J. and Cusolito, A. (2017). Roads, exports and employment: Evidence from a developing country. Journal of Development Economics [Online], 125, 21-39.

Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. Econometrica, 71(6), 1695-1725.

Mellander, C., et al. (2015). Night-time light data: A good proxy measure for economic activity?. PloS One [Online], 10(10), e0139779.

Michaels, G. (2008). The effect of trade on the demand for skill: Evidence from the interstate road system. The Review of Economics and Statistics [Online], 90(4), 683-701.

Ministry of transport and communication, 2019. National Transport Policy. https://www.mtc.gov.zm/wp-content/uploads/2019/08/2019-National-Transport-Policy.pdf.

Mitnik, O. A., Yañez-Pagans, P. and Sanchez, R. (2018). Bright investments: Measuring the impact of transport infrastructure using luminosity data in Haiti. [Online].

Mody, A. and Srinivasan, K. (1996). Japanese and United States Firms as Foreign Investors: Do they March to the Same Tune?'Washington, DC: World Bank [Online].

Moro, A. (2007). Intermediate goods and total factor productivity. Economics Working Papers we076034, Universidad Carlos III, Depar $ Tamento De Economía.

Morrison, C. (1988). Quasi-fixed inputs in US and Japanese manufacturing: a generalized Leontief restricted cost function approach. The Review of Economics and Statistics [Online], 275-287.

Munnell, A. H. (1990). Why has productivity growth declined? Productivity and public investment. New England Economic Review [Online](Jan), 3-22.

Naudé, W., & Matthee, M. (2007). The significance of transport costs in africa.

Nordås, H. K., & Piermartini, R. (2004). Infrastructure and trade.

Nordhaus, W., Azam, Q., Corderi, D., Hood, K., Victor, N.M., Mohammed, M., Miltner, A. and Weiss, J., 2006. The G-Econ database on gridded output: methods and data. Yale University, New Haven, 6.

Obstfeld, M., Rogoff, K. S., & Wren-Lewis, S. (1996). Foundations of international macroeconomics MIT press Cambridge, MA.

Olden, A. and Møen, J., 2020. The triple difference estimator. NHH Dept. of Business and Management Science Discussion Paper, (2020/1).

Outes-Leon, I., & Dercon, S. (2009). Survey attrition and attrition bias in young lives.

Pereira, A. M. and Andraz, J. M. (2004). Public road spending and state spillovers in the USA. Applied Economics Letters [Online], 11(12), 785-788.

Pereira, A. M. and Andraz, J. M. (2013). On the economic effects of public infrastructure investment: A survey of the international evidence. Journal of Economic Development [Online], 38(4), 1.

Piton, S. (2017). Economic integration and the non-tradable sector: The european experience. Paper presented at the 25th CEPR European Summer Symposium in International Macroeconomics (ESSIM), London, UK,

Raballand, G. and Macchi, P. (2008). Transport prices and costs: the need to revisit donors' policies in transport in Africa. Bureau for Research & Economic Analysis of Development Working Paper [Online], 190.

Raballand, G. and Whitworth, A. (2011). The Crisis in the Zambian Road Sector [Online].

Rephann, T. and Isserman, A. (1994). New highways as economic development tools: An evaluation using quasi-experimental matching methods. Regional Science and Urban Economics [Online], 24(6), 723-751.

Robinson, J. A. and Torvik, R. (2005). White elephants. Journal of Public Economics [Online], 89(2-3), 197-210.

Rostow, W. W. (1959). The stages of economic growth. The Economic History Review [Online], 12(1), 1-16.

Road Development Agency. (2017). 2017 Annual Report and Financial Accounts. Roads Agency for the Ministry of Housing and Infrastructure Development.[Online].

Sachs, J. D., & Larrain, F. B. (1993). Macroeconomics in the global economy.

Sanchis-Guarner, R. (2012). Driving up wages: The effects of road construction in great britain. SERC Discussion Papers, (120)

Sequeira, S. (2011). Transport costs and firm behaviour: Evidence from Mozambique and South Africa. Where to Spend the Next Million? [Online], 123.

Schmidt, E., Dorosh, P., Wang, H., & You, L. (2010). Crop production and road connectivity in sub-saharan Africa: A spatial analysis The World Bank.

Shah, A. (1992). Dynamics of public infrastructure, industrial productivity and profitability. The Review of Economics and Statistics [Online], 28-36.

Shen, Z., et al. (2019). Measurement of blooming effect of DMSP-OLS nighttime light data based on NPP-VIIRS data. Annals of GIS [Online], 25(2), 153-165.

Shi, K., et al. (2014). Evaluating the ability of NPP-VIIRS nighttime light data to estimate the gross domestic product and the electric power consumption of China at multiple scales: A comparison with DMSP-OLS data. Remote Sensing [Online], 6(2), 1705-1724.

Shiferaw, A., et al. (2012). Road infrastructure and enterprise development in Ethiopia. International Growth Centre.

Shioji, E. (2001). Public capital and economic growth: a convergence approach. Journal of Economic Growth [Online], 6(3), 205-227.

Storeygard, A. (2016). Farther on down the road: Transport costs, trade and urban growth in sub-saharan africa. The Review of Economic Studies, 83(3), 1263-1295.

Sutton, P. C., Elvidge, C. D. and Ghosh, T. (2007). Estimation of gross domestic product at sub-national scales using nighttime satellite imagery. International Journal of Ecological Economics & Statistics [Online], 8(S07), 5-21.

The Economist. 2015. “African roads and rails: All aboard.” Print edition, 28 February.

Tollefsen, A. F., Strand, H. and Buhaug, H. (2012). PRIO-GRID: A unified spatial data structure. Journal of Peace Research [Online], 49(2), 363-374.

Townsend, A. C. and Bruce, D. A. (2010). The use of night-time lights satellite imagery as a measure of Australia’s regional electricity consumption and population distribution. International Journal of Remote Sensing [Online], 31(16), 4459-4480.

Uy, T., Yi, K., & Zhang, J. (2013). Structural change in an open economy. Journal of Monetary Economics, 60(6), 667-682.

Van de Walle, D. (2009). Impact evaluation of rural road projects. Journal of Development Effectiveness, 1(1), 15-36.

Vivien, F., & Briceño-Garmendia, C. (2010). Africa’s infrastructure: A time for transformation. Paper presented at the Foster, V. and C. Briceno-Garmendia (2010), Africa’s Infrastructure: A Time for Transformation, Africa Development Forum, World Bank, Washington DC,

Wang, X., et al. (2016). Roads to Innovation: Firm-Level Evidence from China. [Online]. Vol. 1542. Intl Food Policy Res Inst.

Wing, C., Simon, K., & Bello-Gomez, R. A. (2018). Designing difference in difference studies: Best practices for public health policy research. Annual Review of Public Health, 39

World Bank. (1997). Project to Support A Road Sector Investment Program. Staff Appraisal Report. Republic of Zambia.

World Bank. (2005). Implementation completion report (ida-29930 tf-29564) on a credit for us$ 70.0 million to the republic of Zambia for a road sector investment program support project.

World Bank. (2007). Evaluation of World Bank Support to Transport Infrastructure. World Bank Publications.

World Bank. (2014).The Republic Of Zambia Diagnostic Trade Integration Study (Dtis) Main Report. Draft prepared for the Enhanced Integrated Framework.

World Bank. (2018). Poverty and shared prosperity. Piecing Together the Poverty Puzzle. Ending Extreme Poverty: Progress, but Uneven and Slowing. [Online].

Workman, D. (2019). Zambia's Top 10 Exports. [Online]. http://www.worldstopexports.com/zambias-top-10-exports/

Yepes, T., Pierce, J., & Foster, V. (2008). Making sense of sub-saharan Africa's infrastructure endowment: A bench marking approach. AICD, World Bank, Washington, DC,

Yue, W., Gao, J. and Yang, X. (2014). Estimation of gross domestic product using multi-sensor remote sensing data: A case study in Zhejiang province, East China. Remote Sensing [Online], 6(8), 7260-7275.

Zeugner, S. (2013). Tradable vs. non-tradable: An empirical approach to the classification of sectors. European Commission, , 46-51.

Zhang, Q. and Seto, K. C. (2011). Mapping urbanization dynamics at regional and global scales using multi-temporal DMSP/OLS nighttime light data. Remote Sensing of Environment [Online], 115(9), 2320-2329.