The role of inter-city transport costs in the formation of small companies across Kenya

Master Thesis in Economics and Business
Specialization: International Economics

ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Name student: Léon Kamenický
ID Number: 531716

Supervisor: Laura Hering
Second assessor: Aksel Erbahar

16/04/2021

Abstract

The effect of inter-city road transport costs on within-country economic development is still a topic of a current debate. The aim of this thesis is to scrutinize the relationship between the inter-city road transport costs and the formation of small companies across 95 cities in Kenya with the intention to contribute to the proportionally few studies on economic development done for most of the countries located in the SSA region so far. The main results from the panel data analysis show, that if the oil price increases significantly (as it did between 2009 and 2011), if a city A is farther away from the closest economic centre than an initially identical city B, less companies registered will be observed in city A than in city B after the increase in the oil price.

Key words: Kenya, Transport costs, Small companies, Roads, Infrastructure

List of Contents

Abstract 2

List of Figures 4

List of Tables 4

1 Introduction 5

2 Theoretical background 9

2.1 Impact of the location and inter-city transport costs on economic development 9

2.2 Small companies, their importance in economic development and the locations where they tend to emerge 11

2.2.1 Is an increase in the number of small companies only an indicator of positive economic development? 11

2.3 The challenges small companies face in Kenya 13

2.4 Summary 14

3 Data 14

3.1 Dependent variable: Number of small companies registered 15

3.1.1 Should a shorter time period be considered? 15

3.1.2 The location of the small companies 16

3.1.3 Small companies registered by city and sector 18

3.2 Explanatory variable: The transport costs measure to the closest economic centre 19

3.2.1 Fuel prices 20

3.2.2 Distance to the closest economic centre 21

3.3 Control variables 22

3.4 Potential omitted variables 22

3.5 Summary 23

4 Empirical Specification 24

5 Results & Discussion 26

5.1 Baseline specification: Full period (1992 – 2016) 26

5.2 Baseline specification: Reduced sample (2009 – 2016) 28

5.3. Baseline specification: Inclusion of city-sector fixed effects and estimation with sector specific measures on small companies registered 30

5.2 Further robustness checks 31

5.2.1 Omitted variables: location-specific factors 31

5.2.2 Robustness checks: Measuring inter-city transport costs to Nairobi, Mombasa or by using the median distance to any city 33

5.3 Limitations of current study and further research 35

6 Conclusion 37

References 38

List of Figures

Figure 1: The Physical Map of Kenya showing major geographical features like elevations, mountain ranges, deserts, Ocean, lakes, plateaus, peninsulas, rivers, plains, landforms and other topographic features. 8
Figure 2: Total number of small companies registered by year without and with Nairobi in the Sample 16
Figure 3: Location of the selected sample cities (100 most populous (2009 census data) in Kenya) and their relative size based on the total number of companies registered from 1992 to 2016 17
Figure 4: graphical representation of the divisions of the sample small companies registered between 2009 to 2016 into the different sectors (excl. those located in the 5 largest cities in terms of population) 19
Figure 5: Graphical representation of the evolution of diesel fuel prices in Kenya over time and the world oil price 21
Figure 6: Categorisation of the cities into isolation groups from 1 to 8 based on the minimum distance to any other city. 22

List of Tables

Table 1: Summary and descriptive statistics of variables used 23
Table 2: Main results and functional form of the explanatory variable 27
Table 3: Main results and functional form of the explanatory variable revised 29
Table 4: Dependent variable revised 31
Table 5: Omitted variables: geography and city specifics 33
Table 6: Distance to important economic centres revised 35

1 Introduction

Economic activity is seldom homogenous across countries. In many low-income nations, some of the hinterland regions are disadvantaged on their path to economic prosperity (Henderson et al., 2012). Kenya, with a population of 51.393 million (World Bank, 2018) and one of the largest economies in the Sub-Saharan African (SSA) region, is no exception. Despite the GDP of Kenya is growing from year to year, a large proportion of the population lives in extreme poverty (35.6% of people live below the poverty line; World Bank, 2016) with Nairobi, the capital, being home to one of the largest slums in SSA. In combination with high levels of corruption, i.e. ranking 137th among 190 countries (Transparency International, 2019), the formation of small businesses was certainly challenged a lot throughout the years.

Thus, the diverse systematic failures symptomatic for the region (corruption, poor infrastructure, ethnic clashes, etc.) mitigated by a number of recent structural improvements, coupled with favourable climate and tourism conditions make Kenya an interesting country for a challenging empirical research. The following geo-political factors, but not only, should be considered before proceeding to next Sections of this thesis:

1. Ethnic diversity and conflicts

   Post-election violence back in 2008, resulting from tensions between ethnicities, is assumed to have a major impact on the course of economic development (Mohajan, 2013). But despite this and a large number of ethnic groups living on the territory of Kenya, the number of conflicts is still moderate in contrast with the rest of SSA countries. Therefore, the country's political system is rather stable, but it is unclear what mechanisms maintain the status quo. Nevertheless, it is worth to consider that Kenya scores low in many categories related to the Index of Economic Freedom such as fiscal health and government integrity, partly due to high corruption (Heritage Foundation, 2020).

2. Geography and climate

   The geography of Kenya is slightly different from the rest of the SSA countries as it is very mountainous. This makes connecting of some cities with the rest of the country quite challenging and might explain why many of them are located in the South and South-West of the country (see Figure 1 below for better clarity). The climate is

favourable to produce coffee, one of the main exports of the country. In addition, Kenya's climate is optimal for agriculture, making growing crops and raising livestock a popular occupation of the country: 61.1% of labour force is employed in this sector (CIA World Factbook, 2020).

3. Tourism

The combination of good climate and the mountainous terrain attracts many tourists from Europe and North America every year. However, research has been inconclusive in determining whether tourism significantly promotes economic growth (e.g. Kibara et al.; 2012).

One of the reasons to such doubts is the darker side of tourism, related to Kenya's unsustainable way of promoting it. For instance, mass tourism has negative environmental consequences leading to significant damages in natural parks (Akama and Kieti, 2007). In recent years, those damages became so excessive, that international tourists started to rather visit other safari-based tourist destinations resulting to a fall in the influx of visitors (Akama, 2000). The net benefit of tourism is further questioned by several other papers, adding that the revenues from the business are going primarily to foreign entities (Dieke, 1991; Sindiga, 1996) with local people not getting a chance to get well-paid managerial job opportunities (Sindiga, 1999; Manyara and Jones, 2005).

4. Oil production

Kenya started exporting oil after its discovery within the country's territory in 2019 (Kippra, 2020). This is important to consider, as the world oil price will be used as the time varying component of the explanatory variable used in the analysis. I will thus assume that Kenya has not a significant impact on oil price and that it did not benefit from increases in the world oil price before 2019 (e.g. in the form of higher tax revenues).

5. Development of infrastructure

The Kenyan National Highway Authority (2020) reports that many projects with the aim to renovate and build quality roads around the country are currently being implemented or planned in the future. However, based on information available from the Logistics Capacity Assessment (2020), the road quality is far from perfect. Additionally, armed banditry is a prominent concern along the most important

humanitarian, trade and tourist routes. When it comes to the current road classification, even the ones labelled as the best ones (i.e. A and B road classification), are not fully paved with several problematic sections. Most of the roads are only partially in a good or fair condition and some larger cities as Lodwar, Wajir and Garissa are hard to access from both Mombasa and Nairobi (Nairobi Kenyan Roads Board, 2018). For reader's reference, a map is presented in Figure 1, illustrating the location of major hubs in the country. It might be observed that the road network is denser in the south and south-west of the country, where many of the cities are located.

Given the above mentioned, the purpose of this thesis will be to focus on the topic of road transport costs and their role on the formation of small businesses across Kenyan cities. Recent research, which significantly inspired this thesis, found evidence that in times when the fuel prices increase significantly, hinterland regions inside many SSA countries experience a slower economic development due to the increasing transport costs they face when accessing larger markets (Storeygard, 2016). Since many economists believe that small businesses are an important stimulus of low-income countries' economies (e.g. Fox and Murray, 2004; Edmiston, 2007), I decided to test, whether the number of small companies registered in city A would be negatively affected by an increase in fuel prices as opposed to an identical city B which is located closer to a larger market. Profiting from a relatively high number of city-level observations on small companies registered in Kenya in the Orbis database[1], I choose an empirical strategy based on an econometric analysis of panel data.

This thesis is structured as follows. In the next section, an overview of relevant literature is presented, discussing the impact of small companies on economic development, the role of transport costs and peculiarities around small companies' incorporation in Kenya. Subsequently, section 3 explores the dataset used in the empirical part, while section 4 defines the empirical strategy. Finally, Section 5 discusses the results including diverse robustness checks followed by Section 6 which concludes the thesis.

[1] Entries on small companies in the Orbis database for Kenya are significantly higher, if compared to other large countries in the SSA region.

Source: 2015 Ezilon.com Regional Maps

Figure 1: The Physical Map of Kenya showing major geographical features like elevations, mountain ranges, deserts, Ocean, lakes, plateaus, peninsulas, rivers, plains, landforms and other topographic features.

2 Theoretical background

The theoretical background covers the following topics. Firstly, the existing literature on the impact of the location and inter-city transport costs on economic development is discussed. Next, the link between small companies and economic development is scrutinized based on existing literature. Even though business activity seems to have long-term positive effects on development of low-income countries, some researchers argue that an increasing number of a certain type of companies must not necessarily sign only positive development. Therefore, the pros and cons of a high number of small companies are also discussed, since it might be very relevant in the context of Kenya. Finally, the last sub-section presents results from surveys that aimed to identify the major obstacles small entrepreneurs face in Kenya based on their answers.

2.1 Impact of the location and inter-city transport costs on economic development

In recent years, geographical factors gained some popularity in economic research and they started to be a regular part of the investigation in the relationship between trade and economic growth (e.g. Gallup et al., 1999; Redding and Venables, 2004). Consequently, geographical features of a particular region and access to the market became a crucial part in the analyses of the impact of transport costs on economic development, which have different findings in different contexts. Primarily, this effect has been mostly studied by using cross-country data (e.g. Limao and Venables, 2001) or by analysing the impact of a construction of a new road in a specific country (e.g Atack et al., 2010; Baum-Snow et al., 2020).

Evidence from previous cross-country trade and economic development research suggests that the lack of access to the sea in landlocked developing countries has negative effect on their economic growth when compared to the non-landlocked ones (e.g. Faye et al., 2007; MacKellar et al., 2000). In countries like Kenya, where people from many ethnic groups are split into the different regions, the ones being located farther from the sea might face similar challenges as inhabitants of landlocked countries. Moreover, the poor relationships between the different ethnic groups (see Section 2.3. for further details) might even more cause the within-country development not to be very heterogenous.

The question of the role of the distance and resulting transport costs from a particular city or region to the country's main economic hubs was discussed in several papers, focusing on

countries like South Africa (Havenga, 2010), China (Faber, 2014) or Indonesia (Rothenberg, 2013). One of the arguments, why the location and quality of the transportation network matter is that the better the access to the market from a country's hinterland region to an important economic centre, the easier is the transportation of industrial goods, as well as trading of a large variety of manufacturing and agricultural goods (e.g. Khander et al., 2009). This in turn might be one of the reasons for huge income inequalities between the different regions within many SSA countries, as they are often very large in terms of surface and the low quality of the transportation infrastructure might therefore sustain the negative effect on economic development of a city located farther away from economic hubs. In general, most of economic activity is centralized in one large city in a vast majority of SSA countries leading to a huge lack of opportunities for people residing in hinterland regions leading to many young people relocating to the one economic centre. Subsequently, this results in excessive and unsustainable population growth of cities such as Lagos in Nigeria, Kinshasa in the DRC, Johannesburg in South Africa, and finally including Nairobi in Kenya as well.

Nevertheless, not much research on the effect of within-country intercity transport costs and its impact on regional development was done in the context of SSA. As Storeygard (2016) notes, there are only few authors that presented a high-quality empiric research on this topic yet up. He therefore presents a study, where he examines the role of intercity transport costs in determining the income of SSA cities in countries whose largest city is a port. The empirical strategy is based on a panel data fixed effects estimation, where the dependent variable represents yearly average city night light activity, a proxy used to measure city-level development as proposed by Henderson et al. (2012). The explanatory variable, a function of the world oil price and road distance to the largest city is used in the research as an exogenous measure for intercity transport costs. As a result, Storeygard (2016) finds significant and negative relationship between his variable of intercity transport costs and night light data, suggesting that cities located further from the main economic hub are disadvantaged in terms of economic growth in times when a sudden shock to the world oil price occurs. In conclusion, the empirical approach is quite convenient for the practical part of this thesis as it uses an explanatory variable that deals with the issue of lack of data and succeeds to present convincing results.

2.2 Small companies, their importance in economic development and the locations where they tend to emerge

Many researchers believe that the formation and survival of small companies in developing countries is a crucial part on the way to economic growth (e.g. Fox and Murray, 2004; Edmiston, 2007). In many African countries, activities by small companies are significantly contributing to the GDP and offering many job opportunities to local people (Advani, 1997; Kayanula and Quartey, 2002; Abor and Quartey, 2010). In fact, 80% of jobs created in Kenya in 2018 are attributed to the small and micro enterprises (Kenyan SME Finance Survey, 2018). In addition, small companies stimulate the economy also by acting as consumers in need of different local goods and services to maintain their operations.

However, the distribution of economic activity in many low-income countries is spatially concentrated, with some of the regions left behind (McDaniel, 2001). The factors leaving some regions underdeveloped are not only geographical, but also socio-cultural, political and infrastructural (UNIDO, 1999; Mohajan, 2013). Moreover, some of the factors might even influence each other. For example, in case the country's infrastructure is underdeveloped, the geographical location of a city is of very high importance since isolated cities might be less attractive for small manufacturers or tourism (Barkema and Drabenstott, 2000). Therefore, it might be assumed that remoteness from economic hubs which defines the magnitude of transport costs from a particular location, may serve as a significant barrier to entry for businesses in spatially large low-income countries. For instance, Melo et. al (2010) argue that international companies may choose not to open new subsidiaries in places where transportation of resources needed to successful business making is too expensive. They therefore highlight the importance of maintaining national roads, since they find out that if remote areas are connected by high quality roads with their economic centres, the likelihood of new firms being established proportionally across such country is significantly higher.

2.2.1 Is an increase in the number of small companies only an indicator of positive economic development?

Banerjee and Duflo (2012) in their book discuss the paradox of the poor and their small businesses. They point out several findings and observations leading to the following question:

Is an increasing number of newly established small businesses in low-income countries necessarily an indicator of positive economic development?

Firstly, many of the small businesses are unprofitable operated by people that most probably cannot afford to pay their few employees. These businesses often run by a single family whose unpaid members must invest a lot of energy for almost no return. This in turn has also negative effects on the education of kids from such families, since they are often expected to stop their educational process and help out with the business.

Secondly, the businesses of poor have very limited potential of growing into middle-sized companies. On contrary, most of them do not survive more than 5 years despite often having access to limited micro-funding. As was also found by Daniels and Mead (1998) and Atieno (2009) there is strong evidence that the financial constraints are only one of the many issues why small businesses operated by the poor are not able to survive and bring enough wealth to their families.

Third, as also mentioned by Tambunan (2008), there is a specific pattern observed in many low-income countries on the formation of small businesses, where the human development is low, and a significant part of the population lives below the poverty line. In fact, an abnormally high number of small businesses formed might potentially point to negative economic development rather than positive, since the small businesses represent the only source of income, thus escape from poverty.

Last, but not least, it is also important to mention that other sized businesses are also important in the economy. As will be discussed in Section 3.1, the observed number of other sized companies registered, other than micro and small is very low across Kenyan cities and can be mostly found only in large cities, such as Nairobi and Mombasa. Thus, there is evidence that there is most probably a huge disproportion of formal and informal jobs in the rural areas of Kenya. Since only small companies are observed in most cities across Kenya, it is highly possible that only informal jobs are available and the people have to relocate to larger cities in order to get a formal job, often available in a large company or governmental institution.

Thus, all of the above mentioned should be considered as well, since on one hand, small companies might be of great contribution to regional GDP, but on the other one, an extremely high number of small businesses operated by the poor is not a sign of a healthy economy.

2.3 The challenges small companies face in Kenya

In general, there are several survey-based studies discussing the challenges small entrepreneurs face in Kenya. The major problems could be divided into three groups: lack of finances, issues related to human capital and infrastructure around the country. Some of the papers' findings concerning financial and educational constraints are summed up in the next paragraphs.

The absence of financial resources together with the constraints related to money borrowing were mentioned as one of the most important issue that affects the creation and performance of small businesses in survey-based studies conducted by Daniels and Mead (1998) and Kamunge et al. (2014). This, among other factors, leads to a large mortality rate as most small companies do not survive more than two years on average. However, the findings by Daniels and Mead (1998) do not necessarily point out that the financial constraints are the most significant factor for small companies to die out. In their study, they compared small companies that received financial support and those that did not, concluding that the difference in their overall performance is negligible. The same is concluded also by Atieno (2009), identifying that joining an association was way more efficient, since it benefits small entrepreneurs with building a relevant network and gives them the opportunity to share best practices in fields such as financial management.

In all, a common thread across the literature is the negative impact of the absence of business education. Cant et al. (2003) and Njorge et al. (2013) found that most owners of small companies in Kenya, do not have any appropriate education about business fundamentals. Therefore, they face the issue of developing and executing the plans by themselves with no prior experience. Surprisingly however, entrepreneurs are still able to perform a basic market research and to successfully promote their agricultural products (Njorge et al.; 2013).

Thus, it is unclear to what extent the lack of finances and education play a role in the opening of new businesses across the country. It might be, that Kenyan entrepreneurs attribute higher importance to factors, such as education and financial resources, overlooking other potential

obstacles which might also play a significant role in the success of the businesses. Nonetheless, Kenyan government has recorded success in the field of education, with 5.2% of GDP invested in the sector resulting in one of the highest literacy rates in SSA - 81.5% (CIA World Factbook, 2020). As such, other factors may be overlooked as significant contributors to the hurdles of conducting business in Kenya, such as infrastructure and distance from economic hubs.

2.4 Summary

In summary, micro companies, as well as small and medium enterprises, face various challenges when doing their business in low-income countries. Kenya, as discussed in the previous sections has already recorded development in several segments of the economy (such as education and governance), but the country is still on the verge of the transition into a developed economy. Even today, many cities are not inter-connected by paved roads making it hard for any kind of entrepreneurs to access the market of goods from the hinterland parts of the country.

3 Data

The regression sample is restricted to the hundred cities of Kenya having a population of at least 30,000 based on census data from 2009. The dependent variable, number of small companies incorporated by city and year, is a measure based on observations from the Orbis database which collects private company data all over the world since 1970s. In this study, the time period for which the data are collected is restricted to 1992 – 2016. However, as will be discussed further, even this time period will finally need to be re-evaluated because of the issue with potentially missing observations between 1992 to 2008.

In addition, a matrix calculating the distances of all 100 cities from each other was created by accessing the Google Maps API, using the administrative city centre as the point of reference for the distance estimation. It is used as a base to calculate the explanatory variable as well as several control variables. Further details on all data used, their sources and descriptive statistics are presented in the following sub-sections and then summed up in Table 1.

3.1 Dependent variable: Number of small companies registered

The measure on the number of small companies registered by city and year will be referred to as SCR_{ct} in this thesis (where c is an index referring to the city and t to year). The measure is based on data collected from the Orbis database; the only source of information used to build it. To understand what a "small company" actually is, Orbis defines it as one, that has less than 15 employees, its revenue is below one million EUR (1.3 million USD), and total assets are below 2 million EUR (2.6 million USD). These companies are labelled with an "S" in the database signalizing their small size. The city-year measure was constructed as follows. Firstly, all the available information on small companies registered in Kenya between 1992 to 2016 was downloaded (i.e. their name, their location and year when incorporated, thus, officially registered). Second, only observations on companies residing in the hundred selected cities were kept. Finally, the observations were grouped by city and year, using the date of incorporation as the reference for the year when the small company was most likely established. Thus, SCR_{ct} equal to 57 for a city A and year 2008, would indicate that approximately 57 small companies were registered in this city in 2008.

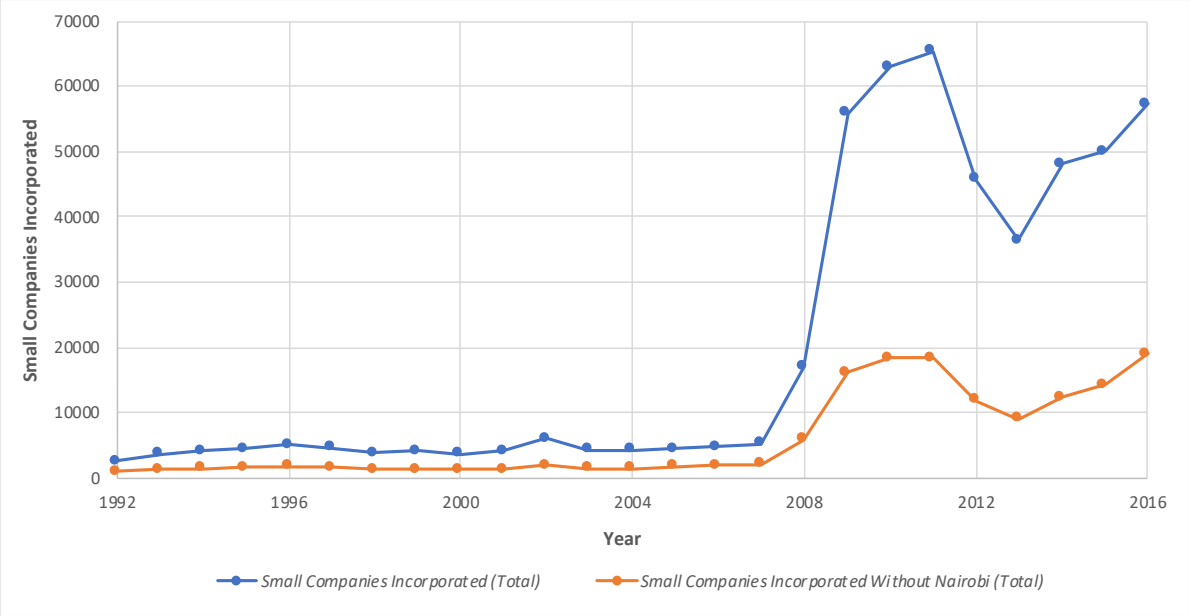
Some interesting observations can be drawn from the data itself. In particular, from the 507,000 small companies registered from 1992 to 2016, around 357,000, thus 70.5% were registered in Nairobi. The high number of companies legally founded in Nairobi itself points to the importance of the city as an economic driver of the country. Notably, Nairobi was also the only city in which other than small companies were regularly observed throughout the years. These might have an impact on the very high number of small companies registered in the city, as the large companies might attract them as observed by some researchers (e.g. Tambunan, 2008). The remaining observations are divided into three sub-sections as they deserve further attention.

3.1.1 Should a shorter time period be considered?

The evolution of the number of small companies registered over the sample time period is depicted in Figure 2. The blue line is based on data with Nairobi in the sample and the orange one without it, respectively. It is visible that both lines follow the same path over the years and that the number of small companies registered significantly increased after 2008 and reached its peak somewhere around 2010 and 2011 followed by a drop in 2013. After that an increase in the number of small companies registered is observed again.

However, the jump in the number of small companies registered after 2008 might not necessarily indicate a sudden 'boom' in opening of small businesses. A possible explanation could also be that data until 2008 could not be correctly recorded (e.g. lack of reliable information) and that some corrective measures were performed enabling a better tracking of company activities from 2009. Therefore, the estimation should also be done for a restricted period, where the accuracy of collected data might be assumed to be higher (i.e. 2009 – 2016).

Regarding the upper limit of the time period selected, it is worth to mention, that despite observations on small companies in Kenya after 2016 being available in the Orbis database, the total number of newly registered companies did not correspond to the official statistics as presented by the Ministry of Industrialization, Trade and Enterprise Development of Kenya (2020), where a significantly higher number of small companies registered was presented in years 2017 and onwards. On the other side, aggregate data on small companies in the period of 1992 – 2016 seem to be inter-coherent between the two sources.



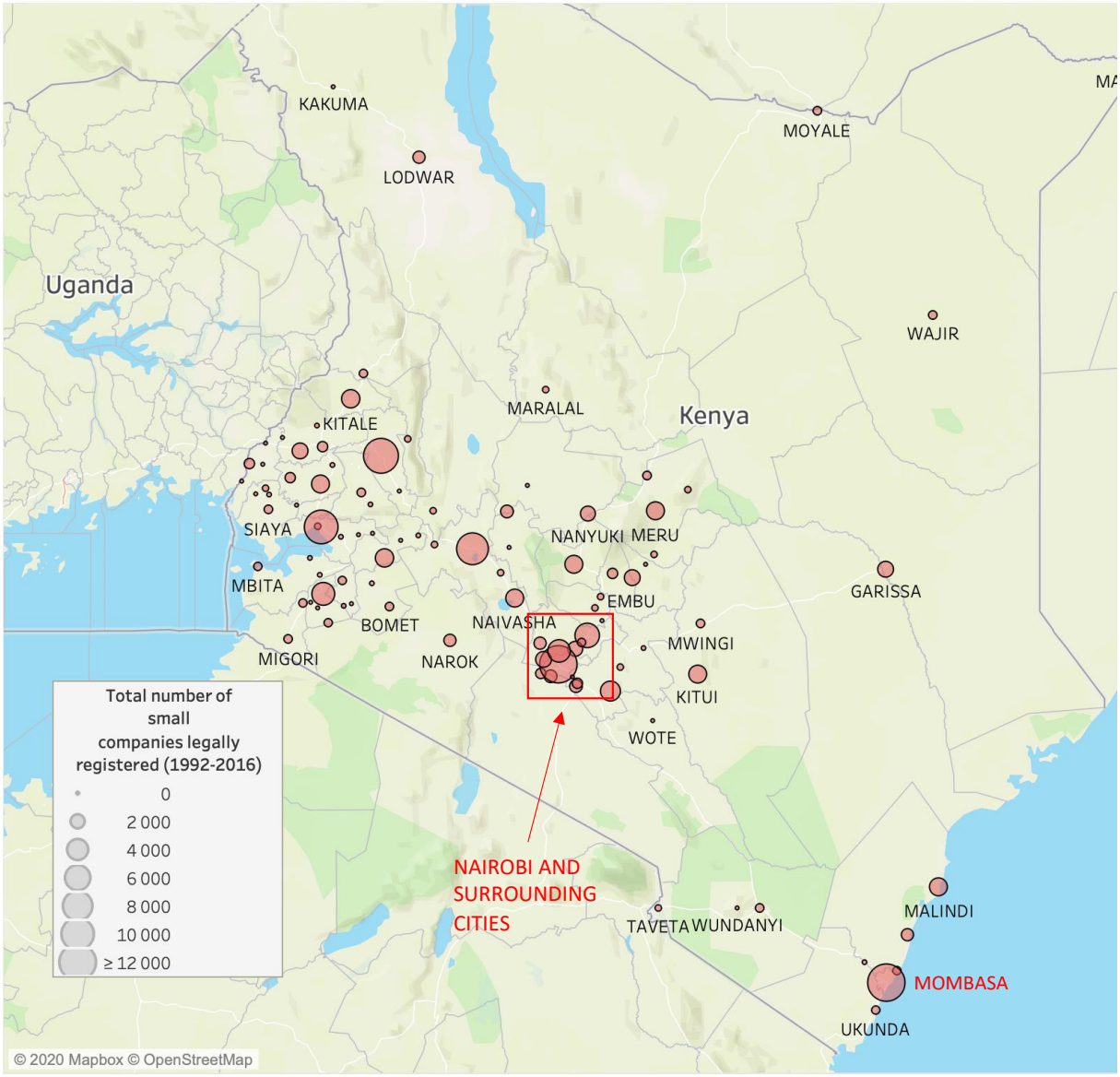
Source: author's elaboration

Figure 2: Total number of small companies registered by year without and with Nairobi in the Sample

3.1.2 The location of the small companies

The map in Figure 3 is a helpful tool to see where the sample cities in Kenya are located. The red dots represent the relative size of the cities based on the number of small companies

registered between 1992 to 2016. Interestingly, the cities are quite proportionally spread in the south of Kenya. In many cities located in the South-Western part of the country, more than 8,000 small companies were registered in total throughout the years, despite being located in areas that are not well accessible by national roads. This is mainly true for those located to the west from Nairobi close to the borders with Uganda. The high number of small companies registered from 1992 to 2016 in these cities, namely Nakuru, Eldoret and Kisumu, are the first indication that Nairobi might not necessarily be the only economic centre of the country. This concept will be further discussed throughout the next sections having important implications for the empirical research, such as defining of the explanatory variable.



Source: author's elaboration

Figure 3: Location of the selected sample cities (100 most populous (2009 census data) in Kenya) and their relative size based on the total number of companies registered from 1992 to 2016

3.1.3 Small companies registered by city and sector

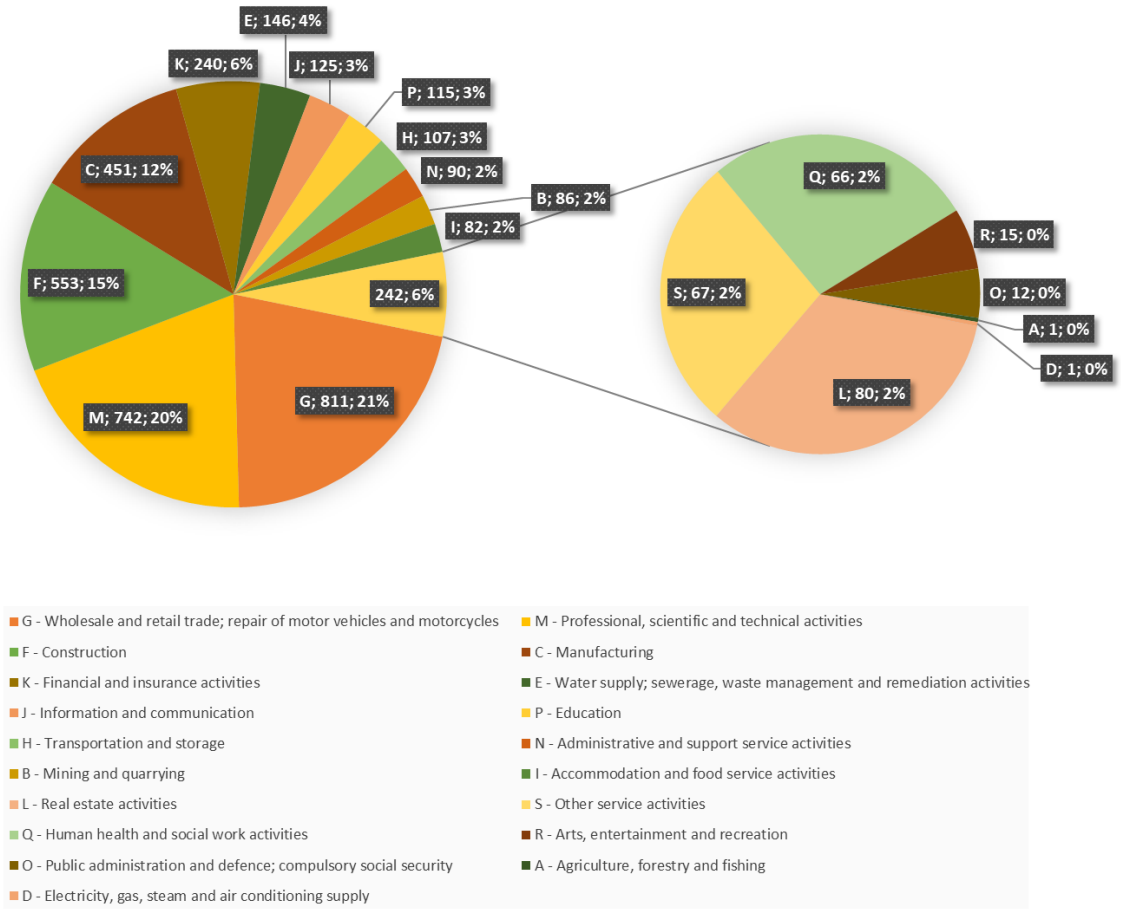
As the information on the sector of small companies in Kenya included in the Orbis database is very poor. In this case, I directly restrict myself to a shorter time period, where a significantly larger amount of information on small companies is provided in the database (i.e. from 2009 to 2016) in order to build a measure of small companies registered by city, year and sector. However, even after taking a shorter and more recent time period, the information on the sector in which the companies are operating is available only for around 10-15% of all observations available.

Figure 4 which is a graphical representation of the divisions of the sample small companies into the different sectors (excl. those located in the 5 largest cities in terms of population). It might be observed that some sectors are represented by more companies than others. As discussed in Section 2.2.1, a healthy economic development should most probably consist of a diversified portfolio of small businesses what seems not to be the case in Kenya. One might argue that some sectors are more specific than the other ones, what is basically right; however, on the other side, some could be more represented. As suggested by Figure 4[2], most businesses for which the sector information was observed appear to be engaged in Wholesale and retail trade (incl. repair of motor vehicles and motorcycles), following by Professional, scientific and technical activities, Construction, Manufacturing and Financial and insurance activities.

On the other hand, even though agricultural goods are a crucial part of Kenya's exports and GDP, there are almost no small companies in the sector Agriculture, forestry and fishing observed. Thus, several questions arise. First, the information on a small business might be generally more indicated for a specific type of businesses. Next, the agricultural activities could potentially be operated mainly by big international corporations (e.g. coffee production) and small farmers can directly sell their cultivated crops to them, without any need to operate officially as a small company. In this case, the question of who benefits from the rich soils of Kenya's mountainous regions arises. It would be no surprise, if those would be mainly the international corporations. Another possible explanation could be that small companies active in the agricultural sector might have been registered in a greater amount before 2009. Finally, the data used to construct the graph presented in Figure 4 do not include observations on small

[2] The legend lists the 19 sectors from most to least represented for easier understanding.

companies from Mombasa, one of the few sample cities located near the sea, thus a potential place where the fishing business could be represented. In summary, even though it is not clear why there are almost no agricultural small businesses registered between 2009 to 2016, the finding itself is quite interesting in order to hypothesize about the potential causes.



Source: author's elaboration

Figure 4: graphical representation of the divisions of the sample small companies registered between 2009 to 2016 into the different sectors (excl. those located in the 5 largest cities in terms of population)

3.2 Explanatory variable: The transport costs measure to the closest economic centre

In the context of Kenya, more complicated measures of transport costs that would rely on diverse qualitative data (e.g. road criminality and mortality rate) is almost impossible considering that they are not collected often. I therefore proceed with the same approach as Storeygard (2016) did when determining the magnitude of road transport costs from a specific city located in one of the selected SSA country to the respecting economic hub. Thus, the explanatory variable is decomposed into two components: (1) fuel prices, representing the time-

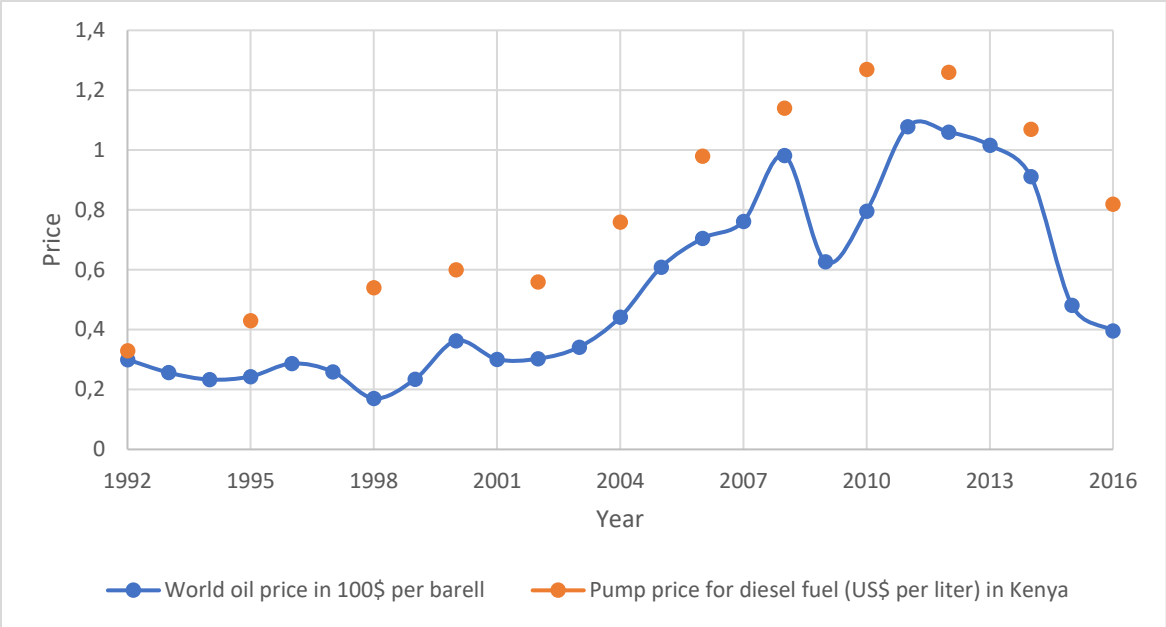
varying component, and (2) the road between a city and an economic centre, representing the time-unvarying component, respectively.

3.2.1 Fuel prices

To measure the fuel prices I use the world oil price based on the Europe Brent Spot Price FOB measured in dollars per barrel and deflated by using the US consumer price index (World Bank, 2010 = 100).[3] The main reason why I use this variable instead of the yearly average pump price of diesel in Kenya is that the source of information is more reliable, the observation is available for every sample year, and that it is a convenient proxy which is exogenous, assuming that Kenya is not a big exported of oil and neither has some influence on the world oil price as discussed in Section 2.2.4. Figure 5 is a graphical representation of the evolution of diesel fuel prices in Kenya over time (represented by orange dots) and the world oil price (represented by connected blue dots). From the information available, it appears that the two variables follow the same path over the time, supporting the assumption that the fuel prices in Kenya are dependent from the world oil price and reflect sudden oil price shocks.

In general, the world oil price fluctuates increases significantly from 2001 to 2009, followed by a negative shock in the price in 2009. From 2009 to 2016 the world oil price increases again; however, declines regularly afterwards. Thus, the fluctuations of the world oil price between 2009 and 2016 are quite significant, leading to fluctuation transport costs that small entrepreneurs face when they want to access the closest market.

[3] The oil price is represented by Europe Brent Spot Price FOB measured in dollars per barrel and deflated by using the US consumer price index (World Bank, 2010 = 100)



Source: author's elaboration

Figure 5: Graphical representation of the evolution of diesel fuel prices in Kenya over time and the world oil price

3.2.2 Distance to the closest economic centre

Even though, the economic activity in Nairobi seems to be multiple times higher, than in any other city of Kenya, some other cities might still be more interesting in terms of accessing larger markets. As shown in Figure 1 in the Introduction, Kenya is a spatially large country and trading certain type of goods (e.g. manufacturing) solely in Nairobi must therefore not necessarily be of high priority for entrepreneurs located in cities closer to another relatively large market. I therefore selected 5 largest cities by population as indicated by the 2009 census, namely Nairobi, Mombasa, Nakuru, Kisumu and Eldoret. As all of them, have a population over 500,000 inhabitants, they can be seen as relatively large and interesting markets for small entrepreneurs. In fact, if an entrepreneur located in a small city whose closest economic centre is Eldoret would want to trade their goods in Nairobi, they would not only cross Eldoret, but Nakuru as well on their way. Thus, considering the spread of the sample cities around the country, measuring transport costs of many cities to Nairobi would not make much sense for many entrepreneurs (results from a regression, where the distance to Nairobi is considered only were found to be statistically not significant, as will be further discussed in Section 5.1).

As a result, the measure of road transport costs from a specific city c to the local economic centre in year t is determined as follows: $Distance(CEC)_c * Price(Oil)_t$.

3.3 Control variables

Two control variables are included in the baseline specification which is discussed in Section 4. The first control variable, represented by the interaction $ln(DistanceCEC)_c * ln(pOil)_t * 1Nairobi_c$ controls for the fact that small entrepreneurs whose closest economic centre is Nairobi might be hit differently by a sudden change in the magnitude of transport costs due to on oil price shock. Thus, the dummy variable $1Nairobi_c$ takes a value of 1 in case the closest economic centre is Nairobi itself. The second control variable, represented by the interaction $\ln(DistanceCEC)_c * \ln(pOil)_t * IsolationLevel_c$ aims to control for the fact that some cities are more isolated than the other ones. Therefore, small entrepreneurs in cities that are more isolated could be more affected by an increase in the magnitude of transport costs. The categorical variable $IsolationLevel_c$ groups the sample cities into 8 different categories based on the minimum distance to any other city. A city A would take a value of 1, in case the distance to any other city is lower than 10km and 8 in case the closest city is farther than 200km away. For further reference on the categorisation of the cities based on the minimum distance to any other city, refer to Figure 6.

Isolation level	Number of cities	Distance interval
1	9	Below 10km
2	24	10 - 20km
3	20	20 - 30km
4	16	30 - 40km
5	9	40 - 50km
6	9	50 - 100km
7	5	100 - 200km
8	3	Above 200km
Grand Total	**95**	

Figure 6: Categorisation of the cities into isolation groups from 1 to 8 based on the minimum distance to any other city.

3.4 Potential omitted variables

Finding data on potential omitted variables on city-year level is quite challenging as neither county, nor country level data are collected on a regular basis. Most of the ones available, are freely accessible from databases such as the Kenya Open Data, with most of them being collected only after 2010. Moreover, most of them were collected just once and therefore no

evolution over time might be evaluated. I therefore decided to at least include city-level dummy variables that could be potentially correlated with the explanatory variable for the purpose of robustness checks. They are all listed in Table 1 including their description, source and corresponding descriptive statistics.

3.5 Summary

Table 1 is a summary of all variables used in the analysis. Descriptive statistics are also reported for each variable on the right (computed by excluding the 5 largest cities in the sample). The standard deviation of the dependent variable is large suggesting that there is a significant variation in the number of small companies registered across Kenyan cities throughout the years.

Table 1*: Summary and descriptive statistics of variables used*

Variable Name	Description	Source	Mean	Standard Deviation	Min	Max	Observations
Dependent Variable(s)							
SCR_{ct}	Represents the number of small companies registered in city c and year t.	Orbis database	36.34	82.468	0	1,063	2,375
SCR_{cst}	Represents the number of small companies registered in city c, sector s and year t.	Orbis database	0.262	1.208	0	45	14,480
Explanatory Variable ($PriceOil_t$ * $DistanceNairobi_c$)							
$pOil_t$	Represents the yearly average Europe Brent Spot Price FOB measured in dollars per barrel and deflated by using the US consumer price index (World Bank, 2010 = 100).	U.S. Energy Information Administration and the World Bank	0.526	0.295	0.171	1.079	2,375
$DistanceCEC_c$	Is the quickest road distance (measured in thousand kilometres) of a city *c* to one of the five closest economic centres (i.e. Nairobi, Mombasa, Nakuru, Eldoret and Kisumu) taking into consideration main transportation corridors.	Google Maps API	0.124	0.145	0.017	0.962	2,375
Control variables							
$1Nairobi_c$	Is a dummy variable depicting whether a city's closest economic centre is Nairobi.	Google Maps API	0.280	0.449	0	1	2,375

Isolation Level$_c$	Is a variable that categorizes cities into five groups based on the minimum distance to any other sample city.	Google Maps API	3.568	1.828	1	8	2,375
Robustness checks: omitted variables and other measures of transport costs							
Elevation$_c$	Is the elevation of a city as documented in the SRTM Data from NASA measured in thousands of meters.	SRTM Data (NASA)	1.449	0.572	0.001	2.489	2,375
1Good Road$_c$	Is a dummy that equals 1 if a city is located on a relatively good quality road (mostly along the main transportation corridors).	The Logistics Capacity Assessment (LCA)	0.230	0.421	0	1	2,375
1Railway$_c$	Is a dummy that equals 1 if a city has a train station that connects it with by rail to Nairobi which can be used for freight transportation.	The Logistics Capacity Assessment (LCA)	0.310	0.463	0	1	2,375
1Coffee Region$_c$	Is a dummy that equals of 1 if a city is located in one of the coffee regions that remained after 1990 in Kenya.	Royal News (June 2005)	0.230	0.421	0	1	2,375
1Natural Park$_c$	Is a dummy indicating that equals 1 in case a city is less than 30 km away from a natural park.	Kenya Wildlife Service and Google Maps	0.190	0.392	0	1	2,375
Distance Nairobi$_c$	Is the quickest road distance (measured in thousand kilometres) of a city c to Nairobi taking into consideration main transportation corridors.	Google Maps API	0.289	0.183	0	1.034	2,375
Distance Mombasa$_c$	Is the quickest road distance (measured in thousand kilometres) of a city c to Mombasa taking into consideration main transportation corridors.	Google Maps API	0.674	0.249	0	1.212	2,375
MediumDistance(Any city)	Is the medium distance of transport costs from a city c to any other sample city.	Google Maps API	0.358	0.169	0.189	1.099	2,375

Source: author's elaboration
Note: Logarithmic and square root transformations are not included in this table as the main purpose is to present the main

4 Empirical Specification

Combining the information gathered from reviewing different papers, as well as from initial analysis of the variables itself, I decided to use a fixed effects econometric analysis of panel

data.[4] In a scenario, where the lack of reliable data does not allow to include a sufficient number of control variables into a regression, the inclusion of fixed effects is an approach that might mitigate the impact of a potential omitted variable bias. Thus, considering the panel I created for my research, I can include year[5] and city[6] fixed effects, which can control for unobserved effects.

My baseline specification is:

$$lnSCR_{ct} = \beta_0 + \beta_1 \ln(DistanceCEC)_c * \ln(pOil)_t + \beta_2 \ln(DistanceCEC)_c * \ln(pOil)_t * 1Nairobi_c + \beta_3 \ln(DistanceCEC)_c * \ln(pOil)_t * + \delta_1 \ln(pOil)_t * 1Nairobi_c + \delta_2 (pOil)_t * IsolationLevel_c + \gamma_c + \gamma_t + \epsilon_{ct} \quad (1)$$

where $lnSCR_{ct}$ is the dependent variable representing the number of small companies registered in city *c* and year *t* in logarithm, the interaction term $ln(DistanceCEC)_c * ln(OilPrice)_t$ represents the explanatory variable of interest measuring the approximate magnitude of transport costs from a city *c* to the closest economic centre in year *t*, γ_c is a city-fixed effect and γ_t is a year-fixed effect. The interaction terms $\ln(DistanceCEC)_c * \ln(pOil)_t * 1Nairobi_c$ and $\ln(DistanceCEC)_c * \ln(pOil)_t * IsolationLevel_c$ represent control variables. Double interactions terms of $\ln(pOil)_t * 1Nairobi_c$ and $(pOil)_t * IsolationLevel_c$ have to be included as well, considering that the control variables consist of a triple interaction. Double interactions between two time invariant variables (e.g. $ln(DistanceCEC)_c * 1Nairobi_c$) are not included in the specification since they would be collinear with the city fixed effects. Standard errors are clustered at the city level.

The motivation behind the using of log-transformed variables in the baseline specification, is the relatively high variation in the dependent as well as explanatory variables (as presented in the descriptive statistics in standard deviation in Table 1). The measure $ln(DistanceCEC * pOil)_{ct}$ in this case cannot be used, since it is equal to $ln(Distance)_c + ln(pOil)_t$ and therefore would be collinear with the city and year fixed effects, respectively.

[4] For more technical details on the estimation methods suitable for Panel Data, refer to textbooks, such as Wooldridge, J. M. (2006) Introductory econometrics: a modern approach. 3rd ed.

[5] Potential year fixed effects: pre-election violence and civil wars, nationwide natural disasters and pandemics, and locust infestations

[6] Potential city fixed effects: climate, infrastructure, presence of natural resources, and constant population growth

Even though the empirical approach by Storeygard (2016) was used as the basis for the empirical analysis, I adapted and expanded it in several ways. The first adaptation is the dependent variable itself (see Section 3.1), as my main purpose is to analyse the relationship between transport costs and the formation of small businesses across Kenya. Next, instead of measuring the access to larger markets in terms of road transport costs to the largest economic centre (i.e. Nairobi), I use closest economic centres (see Section 3.1 for further details). When it comes to expanding the baseline specification, I include two control variables to account for factors as discussed in Section 3.3. In case the two control variables would not be included in the specification, they could both represent omitted variables since they might be correlated with one of the components of the explanatory variable, i.e. $\ln(DistanceCEC)_c * \ln(pOil)_t$.

Robustness checks are mainly in the form of the inclusion of an interaction of the explanatory variable with city specific dummies into the baseline specification. The purpose is to verify whether there are some omitted variables that might be correlated with the transport costs measure and therefore bias the result, leading to an over- or underestimation of the regression coefficients of interest. Such city specific dummy variables are for example proximity of a major national road or presence of railroads. In addition, specific regional 'advantages' that might attract companies to settle farther from economic centres (such as natural parks and coffee production), are also tested by interacting city-level dummies with the explanatory variable (for more details on such dummy variables refer to Table 1 in Section 3.5).

5 Results & Discussion

5.1 Baseline specification: Full period (1992 – 2016)

Table 2, Column 7 represents regression results for the baseline specification as presented in equation (1). The regression coefficient of interest, -0.224 on $ln(distanceCEC)_c * ln(pOil)_t$, is negative and statistically significant at ten percent level. However, omitting the two control variables, as specified in equation (1), does not yield a statistically significant coefficient as shown in Column 5. Neither do modifications of the functional form of the explanatory variable, e.g. by not using log-transformed forms of the measures on distance and oil price, lead to any statistically significant coefficients as shown in Columns 1 to 4 and 6. This suggest, that the coefficient of interest on $ln(DistanceCEC)_c * ln(pOil)_t$ might suffer from an omitted variable bias when the control variables are not included. On the other hand, if the number of small

companies registered by city and year before 2009 is low mainly because of an unproper recording of data, a potential bias caused by a measurement error might be present as well. A shown in Figure 2, Section 3.1.1, the number of observations on small companies registered in the period after 2009 is significantly higher and it is worth to repeat the estimation strategy by reducing the sample to a shorter time period, i.e. from 2009 to 2016.

Table 2: Main results and functional form of the explanatory variable

	Dependent variable							
	(1) $\ln(SCR)_{ct}$	(2) $\ln(SCR)_{ct}$	(3) $\ln(SCR)_{ct}$	(4) $\ln(SCR)_{ct}$	(5) $\ln(SCR)_{ct}$	(6) $\ln(SCR)_{ct}$	(7) $\ln(SCR)_{ct}$	(8) $\ln(SCR)_{ct}$
Explanatory variable								
$(DistanceCEC)_c * (pOil)_t$	0.552							
	(0.590)							
$((DistanceCEC)_c * (pOil)_t)^2$		0.605						
		(0.982)						
$(DistanceCEC)_c * \ln(pOil)_t$			0.413					
			(0.314)					
$\ln(DistanceCEC)_c * (pOil)_t$				0.028				-0.411*
				(0.124)				(0.227)
$\ln(DistanceCEC)_c * \ln(pOil)_t$					0.027		-0.224*	
					(0.068)		(0.124)	
$(\ln(DistanceCEC)_c * \ln(pOil)_t)^2$						0.309		
						(0.270)		
Control variables								
$ExplanatoryVariable_{ct} * 1NotNairobi_c$							0.092	0.213
							(0.132)	(0.245)
$ExplanatoryVariable_{ct} * IsolationLevel_c$							0.060	0.101
							(0.022)	(0.041)
Observations	1896	1896	1896	1896	1896	1896	1896	1896
Years	1992-2016	1992-2016	1992-2016	1992-2016	1992-2016	1992-2016	1992-2016	1992-2016
Model	FE	FE	FE	FE	FE	FE	FE	FE
Fixed effects	Year, City	Year, City	Year, City	Year, City	Year, City	Year, City	Year, City	Year, City
Cities	95	95	95	95	95	95	95	95

Note 1: Each column is a separate fixed effects regression. Regression coefficients for double interactions included in regressions for which results are presented in Columns 7 and 8 are not reported since they are not very informative. Robust standard errors, clustered by city, are in brackets. *,** and *** represent significance at the ten, five, and one percent level, respectively. More details on the variables used in the regressions are available in Table 1 (Summary of data used and corresponding descriptive statistics).
Note 2: Other functional forms of the explanatory variable with control variables were tested as well. However, none of the coefficients on the different functional forms of the explanatory variable were statistically significant at any level of interest. The corresponding regression results are therefore not shown in the table.

Source: author's elaboration

5.2 Baseline specification: Reduced sample (2009 – 2016)

Based on the finings presented in previous Section 5.1., Table 3 reports results for estimations where a reduced sample restricted to observations from 2009 to 2016 is used. The coefficient of interest on $ln(DistanceCEC)_c * ln(pOil)_t$ is negative and statistically significant as supported by Column 7 of Table 3 which reports estimates of the baseline specification in equation (1). The coefficient of interest remains statistically significant and negative also in case in which the two control variables are omitted from equation (1). Moreover, it does also remain negative and statistically significant when the functional form of the explanatory variable is modified, e.g. by not using the log transformed forms of the measures on distance and oil price, as shown in Columns 1 to 5 of Table 3. Thus, reducing the sample to a shorter time period could have helped with the potential measurement error, despite a loss of observations.

As a result, if considering the statistically significant coefficient of interest, -0.234 on $ln(DistanceCEC)_c * ln(pOil)_t$, there is evidence, that if the oil price increases significantly (as it did between 2009 and 2011), and a city A is farther away from the closest economic center than an initially identical city B, less companies registered will be observed in city A than in city B after the increase in the oil price. This negative relationship between an increase in transport costs and number of small companies registered would be in line with existing literature pointing that rural areas suffer from large distance to bigger markets from the economic perspective (see Section 2.2. for theoretical support).

Notably, given the functional form of the explanatory variable, it is quite complicated to exactly quantify the negative effect of an increase in fuel prices on the formation of small companies located in cities farther from economic center. Nevertheless, the findings presented in Table 3 are already giving some interesting insights on the research.

Table 3: Main results and functional form of the explanatory variable revised

	Dependent variable							
	(1) $\ln(SCR)_{ct}$	(2) $\ln(SCR)_{ct}$	(3) $\ln(SCR)_{ct}$	(4) $\ln(SCR)_{ct}$	(5) $\ln(SCR)_{ct}$	(6) $\ln(SCR)_{ct}$	(7) $\ln(SCR)_{ct}$	(8) $\ln(SCR)_{ct}$
Explanatory variable								
$DistanceCEC_c * pOil_t$	-1.123*							
	(0.582)							
$DistanceCEC_c * (pOil_t)^2$		-2.015**						
		(0.918)						
$DistanceCEC_c * \ln(pOil)_t$			-0.804*					
			(0.449)					
$\ln(DistanceCEC)_c * pOil_t$				-0.204**				-0.325***
				(0.092)				(0.116)
$\ln(DistanceCEC)_c * \ln(pOil)_t$					-0.140**		-0.234***	
					(0.070)		(0.087)	
$(\ln(DistanceCEC)_c * \ln(pOil)_t)^2$						0.221		
						(0.204)		
Control variables								
$ExplanatoryVariable_{ct} * 1NotNairobi_c$							0.151	0.192
							(0.147)	(0.194)
$ExplanatoryVariable_{ct} * IsolationLevel_c$							0.026	0.033
							(0.021)	(0.027)
Observations	724	724	724	724	724	450	724	724
Years	2009-2016	2009-2016	2009-2016	2009-2016	2009-2016	2009-2016	2009-2016	2009-2016
Method	FE	FE	FE	FE	FE	FE	FE	FE
Fixed Effects	Year, City	Year, City	Year, City	Year, City	Year, City	Year, City	Year, City	Year, City
Cities	95	95	95	95	95	95	95	95

Note 1: Each column is a separate fixed effects regression. Regression coefficients for double interactions included in regressions for which results are presented in Columns 7 and 8 are not reported since they are not very informative. Robust standard errors, clustered by city, are in brackets. *,** and *** represent significance at the ten, five, and one percent level, respectively. More details on the variables used in the regressions are available in Table 1 (Summary of data used and corresponding descriptive statistics).

Note 2: Other functional forms of the explanatory variable with control variables were tested as well. However, none of the coefficients on the different functional forms of the explanatory variable were statistically significant at any level of interest. The corresponding regression results are therefore not shown in the table.

Source: author's elaboration

5.3. Baseline specification: Inclusion of city-sector fixed effects and estimation with sector specific measures on small companies registered

Table 4 reports regression results where the dependent variable, thus number of small companies registered, is sector specific as described in Section 3.1.3. Column 1 reports regression results for an equation where the dependent variable used is SCR_{cst}, i.e. small companies registered in city *c*, sector *s* and year *t*. Since many observations on SCR_{cst} are equal to *1* and *ln(1) = 0*, it is not log-transformed in this case. The regression coefficient of interest on $ln(DistanceCEC)_c * ln(pOil)_t$, is negative and statistically significant at ten percent level. As a result, the inclusion of city-sector fixed effects leads to identical results as presented in previous Section 5.2 above.

Columns 2 to 6 report regression results where the dependent variable is restricted to companies in one of the five sectors, for which the highest number of them registered from 2009 to 2016 were observed.[7] The statistically significant and negative coefficient on $ln(DistanceCEC)_c * ln(pOil)_t$ presented in Column 3, suggests that if the oil price increases significantly, if a city A is farther away from the closest economic centre than an initially identical city B, less companies registered in the 'Professional, scientific and technical activities (M)' sector will be observed in city A than in city B after the increase in the oil price. It appears, that their businesses are quite dependent from accessing larger markets and are their decision to open a small company in this sector is therefore negatively hit by a sudden increase in transport costs in remote areas.

The coefficients of interest on the explanatory variable shown in Columns 2 and 4 to 6 are not statistically significant at any level of interest. Therefore, not much can be said about the effect of an increase of transport costs on small companies registered in the Wholesale and retail trade (G), Construction (F), Manufacturing (C) and Financial and insurance activities (K) sectors.

[7] 1. G - Wholesale and retail trade; repair of motor vehicles and motorcycles; 2. M - Professional, scientific and technical activities; 3. F – Construction; 4. C manufacturing; and 5. K - Financial and insurance activities (see Figure 4 in Section 3.1.3 for further details)

Table 4: Inclusion of city-sector fixed effects and estimation with sector specific measures on small companies registered

	Dependent variable					
	(1) SCR_{cst}	(2) $SCRG_{ct}$	(3) $SCRM_{ct}$	(4) $SCRF_{ct}$	(5) $SCRC_{ct}$	(6) $SCRK_{ct}$
Explanatory variable						
$\ln(DistanceCEC)_c * \ln(pOil)_t$	-0.065*	-0.312	-0.705**	-0.001	-0.092	0.105
	(0.034)	(0.220)	(0.318)	(0.139)	(0.229)	(0.108)
Control variables						
$ExplanatoryVariable_{ct} * 1Nairobi_c$	-0.038	-0.078	-0.077	-0.093	-0.121	-0.295**
	(0.043)	(0.277)	(0.471)	(0.132)	(0.245)	(0.131)
$ExplanatoryVariable_{ct} * IsolationLevel_c$	0.025	0.134*	0.191	0.038	-0.008	0.043
	(0.009)	(0.069)	(0.092)	(0.027)	(0.064)	(0.031)
Observations	14,440	760	760	760	760	760
Years	2009-2016	2009-2016	2009-2016	2009-2016	2009-2016	2009-2016
Method	FE	FE	FE	FE	FE	FE
Fixed Effects	City-Sector, Year	City, Year	City, Year	City, Year	City, Year	City, Year
Cities	95	74	75	75	64	53
Sector	All (19)	G	M	F	C	K

Note 1: Each column is a separate fixed effects regression. Robust standard errors, clustered by city, city-sector (Column 1) are in brackets. *,** and *** represent significance at the ten, five, and one percent level, respectively. More details on the variables used in the regressions are available in Table 1 (Summary of data used and corresponding descriptive statistics).
Note 2: Other functional forms of the explanatory variable as well as estimation without control variables was tested. However, no statistically significant coefficients on the explanatory variable were found and the results are therefore shown only for the specifications closest to equation (1).
Note 3: The letters indicated the following sectors: G - Wholesale and retail trade; repair of motor vehicles and motorcycles; M - Professional, scientific and technical activities; F – Construction; C manufacturing; and K - Financial and insurance activities

Source: author's elaboration

5.2 Further robustness checks

5.2.1 Omitted variables: location-specific factors

Table 5 considers omitted variables, in the form of location-specific factors that might be potentially correlated with $ln(DistanceCEC)_c * ln(pOil)_t$: the elevation as well as the presence of a national road, railroads, coffee plantations and natural parks.

Kenya, as one of the few very mountainous countries in SSA has a very specific surface, since only it's capital city, Nairobi, has an elevation of 1,724 meters above sea level. This makes the transportation in the country more challenging, and elevation of a specific city might play a role in determining the magnitude of transport costs small entrepreneurs face there. The coefficient of the interaction of the explanatory variable with the city elevation in Column 1 of Table 5, is

positive, but not statistically significant at any level of interest, suggesting that on oil price increase has no differential effect on small companies registered in cities located significantly above the sea level farther from economic centers.

Neither a differential effect of an increase in the transport costs measure on the formation of small companies in cities located in a coffee producing region was identified as shown in Column 2. The reason might be that small companies whose business would consist of growing coffee beans could have been already registered many years before 2009. In addition, if the small farmers sell the fruits of their crops directly to someone that collects them on a regular basis, they might not really think about transport costs as a constraint to start with coffee plantation.

Cities located close to national parks might be interesting for entrepreneurs that want to sell some type of products or services to the incoming tourists. Analogous to the case with the coffee production, neither do the results presented in Column 3 of Table 5 suggest that an oil price increase would have a differential impact on the registration of small companies in remote cities located close to a natural park.

It is very likely that rail transport is less affected by oil price fluctuations for several reasons (e.g. fuel efficiency or fixed fuel prices). Table 5 Column 4 considers, along with the main effect, the differential impact of the transport costs measure on cities with a rail connection to one of the five economic centers. However, the additional rail effect is small and insignificant.

The quality of road network was relatively unchanged from 2009 till 2016 as major projects are still being implemented or have been completed in recent years (Kenya National Highway Authority, 2020). However, some major transportation corridors are maintained and therefore some cities relate to a better-quality road than the others making the transport costs more expensive for locations which have not an access to a paved road. Nevertheless, the results in Column 5 do not suggest that the additional better-quality road would be significant.

Results presented in Columns 6 and 7 are just for illustration of how the results would change in case more of the variables on location-specific factors together with control variables from equation (1) would be in one specification.

In summary, the main coefficient of interest on the explanatory variable, i.e. $ln(DistanceCEC)_c * ln(pOil)_t$ stays relatively robust to the inclusion of the hypothesized omitted variables as presented in Columns 1 to 3 of Table 5.

Table 5: Omitted variables: geography and city specifics

	Dependent variable						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	$\ln(SCR)_{ct}$	$\ln(SCR)_{ct}$	$\ln(SCR)_{ct}$	$\ln(SCR)_{ct}$	$\ln(SCR)_{ct}$	$\ln(SCR)_{ct}$	$\ln(SCR)_{ct}$
Explanatory variable							
$\ln(DistanceCEC)_c*\ln(pOil)_t$	-0.227*	-0.170**	-0.131*	-0.097	-0.091	-0.195	-0.220
	(0.131)	(0.065)	(0.067)	(0.083)	(0.085)	(0.137)	(0.140)
Control variables							
$\ln(DistanceCEC)_c*\ln(pOil)_t$ *$1NotNairobi_c$							0.095
							(0.196)
$\ln(DistanceCEC)_c*\ln(pOil)_t$ *$IsolationLevel_c$							0.023
							(0.021)
$\ln(DistanceCEC)_c*\ln(pOil)_t$ *$Elevation_c$	0.093					0.108	0.069
	(0.103)					(0.125)	(0.125)
$\ln(DistanceCEC)_c*\ln(pOil)_t$ *$1CoffeeRegion_c$		0.110				0.184	0.153
		(0.155)				(0.239)	(0.279)
$\ln(DistanceCEC)_c*\ln(pOil)_t$ *$1NaturalPark_c$			-0.090			-0.003	0.062
			(0.152)			(0.167)	(0.149)
$\ln(DistanceCEC)_c*\ln(pOil)_t$ *$1Railway_c$				-0.140		-0.103	-0.183
				(0.178)		(0.201)	(0.200)
$\ln(DistanceCEC)_c*\ln(pOil)_t$ *$1GoodRoad_c$					-0.188	-0.300	-0.248
					(0.135)	(0.196)	(0.210)
Observations	724	724	724	724	724	724	724
Years	2009- 2016	2009 -2016	2009-2016	2009-2016	2009-2016	2009-2016	2009-2016
Method	FE	FE	FE	FE	FE	FE	FE
Fixed effects	Year, City	Year, City	Year, City	Year, City	Year, City	Year, City	Year, City
Cities	95	95	95	95	95	95	95

Note: Each column is a separate fixed effects regression. Robust standard errors, clustered by city, are in brackets. *,** and *** represent significance at the ten, five, and one percent level, respectively. More details on the variables used in the regressions are available in Table 1 (Summary of data used and corresponding descriptive statistics).

Source: author's elaboration

5.2.2 Robustness checks: Measuring inter-city transport costs to Nairobi, Mombasa or by using the median distance to any city

Table 6 shows regression results where the distance component of the transport costs measure is modified. In Columns 1 and 2, instead of using the distance of a sample city to its closest economic center, the distances to Nairobi (Column 1) and Mombasa (Column 2) are used. These checks are done in order to contrast the approach applied by Storeygard (2016), since considering only the largest and most important city as the only place of trading opportunity

might not be the best approach in case of very large countries, where several big cities are located.

The coefficient of interest on $ln(DistanceNairobi)_c * ln(pOil)_t$ in Column 1 is positive, but not statistically significant at any level of interest. On a very similar note, when Mombasa, the largest port of Eastern Africa, is considered as the reference point for the distance component of the explanatory variable, no significant results were found as presented in Column 2 of Table 6. The same can be concluded when the median distance to any other sample city is considered for the distance component of the transport costs measure. It should also be noted that similarly to previous robustness checks presented, other functional forms of the explanatory variable were tested as well. However, no statistically significant coefficients at any level of interest were found.

In conclusion, in the context of the formation of small businesses a correlation with the transport costs measure can be found in case the distance to the closest economic center is considered as demonstrated by the results presented throughout this section. Even though Nairobi and Mombasa are certainly important economic centers of the country, increasing transport costs to those cities must not necessarily be a strong barrier to entry for small entrepreneurs. For many of them, the costs of accessing the closest ‘interesting’ market seem to be more important.

Table 6: Measuring inter-city transport costs to Nairobi, Mombasa or by using the median distance to any city

	Dependent variable		
	(1)	(2)	(3)
	$\ln(SCR)_{ct}$	$\ln(SCR)_{ct}$	$\ln(SCR)_{ct}$
Explanatory variable			
$\ln(DistanceNairobi)_c$ * $\ln(pOil)_t$	0.588		
	(0.544)		
$\ln(DistanceMombasa)_c$ * $\ln(pOil)_t$		0.457	
		(0.280)	
$\ln(MedianDistanceAnyCity)_c$ * $\ln(pOil)_t$			0.339
			(0.659)
Control variables			
Explanatory $variable_{ct}$ *$1NotNairobi_c$			-0.342
			(0.325)
Explanatory $variable_{ct}$ *$IsolationLevel_c$	-0.147	-0.091**	-0.128
	(0.078)	(0.041)	(0.077)
Observations	724	724	724
Years	2009-2016	2009-2016	2009-2016
Method	FE	FE	FE
Fixed Effects	Year, City	Year, City	Year, City
Cities	95	95	95

Note: Each column is a separate fixed effects regression. Robust standard errors, clustered by city, are in brackets. *,** and *** represent significance at the ten, five, and one percent level, respectively. More details on the variables used in the regressions are available in Table 1 (Summary of data used and corresponding descriptive statistics).

Source: author's elaboration

5.3 Limitations of current study and further research

Due to practical constraints, this paper cannot provide a comprehensive review of all contributing factors to the development of small businesses. Since throughout the thesis, I rely on new companies' formation as main dependent variable, this section summarizes suggestions as to how can the estimation be improved for this variable.

Firstly, a potential confounding factor could be the possible correlation between the failure rate of new companies and the change in road transport cost. Since Orbis database does not contain information on status changes of the enterprises for almost all companies residing in Kenya, this could not have been included into the regression.

On a similar note, every empirical research is prone to measurement errors due to the nature of the collected data. Despite collecting the data from the Orbis database, a reliable source of information, I had to finally work with a way shorter time period than initially planned in order to get some meaningful results as shown in Section 5.1. Even though, aggregate numbers have been cross validated with other public data sources, such as Kenyan National Chamber of Commerce and Industry (2020), it is not clear how the data on small companies in Kenya were collected for those incorporated before 2009. Therefore, the analysis could certainly be further developed in the future when more data will be included. In fact, based on the information available, recent years might be a turning point for the inter-city transportation in Kenya, since significant investments into the road as well as railroad systems were made with the help of the Chinese government (Bloomberg, 2019).

Furthermore, the paper focused only on Kenya, which could introduce country-specific bias to the estimation. Although conducting the study on the entire SSA region was out of the scope of the thesis, further research may concentrate on that to control against, country-specific heterogeneity, resulting in more robust insights. Nonetheless, this might be complicated due to the lack of historic data for a sufficiently big panel and imprecise recording of new incorporations. For instance, Nigeria's Orbis entry, spanning from 1990 to 2003 contains only zeros for subsequent years. This is highly unlikely due to the nation's economic activity rising significantly every year (CIA World Factbook, 2020).

What is more, one should be cautious about drawing causality implications from the obtained results. The regression coefficient could determine a causal relationship in case (1) all omitted variables were discovered and included in the regression, (2) there is no measurement error and (3) there is no evidence of reversed causality (Wooldridge, 2008). Nonetheless, the multiple robustness checks, yielding the same result provides enough support for the validity of the results in this country-specific case.

Ultimately, reversed causality is prominent in trade and economic development research as there is not a clear answer whether trade fosters economic prosperity or vice versa. In this case, reversed causality could be an issue if, for example, Kenya would decide where to build high-quality roads based on the number of companies residing in a specific region. To the best of my knowledge, this was not the case.

6 Conclusion

The main aim of this thesis was to test whether the increase in fuel prices has a differential effect on the formation of small companies across Kenyan cities based on their distance from their closest economic centre. For doing so, I regressed a self-developed variable on the number of small companies registered by city and year on a measure of transport costs to the closest economic centre. In order to determine the transport costs to the closest economic centre, I followed the methodology as presented in a recent paper by Storeygard (2016), where he decomposed the variable into two components: (1) fuel prices and (2) distance from city A to a particular economic centre within one of the sample SSA countries.

The regression results suggest, that in times when fuel prices increase significantly, less companies registered might be observed in cities located farther from their closest economic centre, if compared to identical cities located closer. Even though, the economic activity is highly concentrated in the largest African cities, it appears, that small entrepreneurs in Kenya are rather deciding on whether to open a business based on the transport costs to a local economic centre, an important factor that should be consider when evaluating the costs of accessing a larger market, where goods and service might be traded.

The results are consistent with previous research in which the impact of transport costs on regional economic development was studied (e.g. Barkema and Drabenstott, 2000; Storeygard, 2016). However, one might be cautious when evaluating the effects of transport costs on company activity in rural areas, since a higher number of small companies should not be necessarily associated with healthy economic development (e.g. Tambunan, 2008; Banerjee and Duflo, 2012).

In conclusion, this thesis provided evidence that the location of a city in Kenya matters for the development of small companies. Thus, recent investments in infrastructure might mitigate the costs of accessing larger markets from remote areas and potentially boost the number of small companies operating in cities located farther away. However, since inter-city trade is only one of the many factors having an impact on the survival and performance of small companies in Kenya, further research is needed to uncover them and quantify their impact. This is left for future students to scrutinize.

References

Abor, J., & Quartey, P. (2010). Issues in SME development in Ghana and South Africa. International research journal of finance and economics, 39(6), 215-228.

Advani, A. (1997). Industrial clusters: A support system for small and medium sized enterprises. Private Sector Development Department, World Bank.

Abhijit Banerjee, Esther Duflo (2012). Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty. Hachette UK. ISBN 1610391608, 9781610391603.

Akama, J. S. (2000). The efficacy of tourism as a tool for economic development in Kenya.

Akama, J. S. (1997). Tourism development in Kenya: Problems and policy alternatives. Progress in Tourism and Hospitality Research, 3(2), 95-105.

Akama, J. S., & Kieti, D. (2007). Tourism and socio-economic development in developing countries: A case study of Mombasa Resort in Kenya. Journal of sustainable tourism, 15(6), 735-748.

Atack, J., Bateman, F., Haines, M., & Margo, R. A. (2009). *Did railroads induce or follow economic growth? Urbanization and population growth in the American Midwest, 1850-60* (No. w14640). National Bureau of Economic Research.

Atieno, R. (2009). Linkages, access to finance and the performance of small-scale enterprises in Kenya (No. 2009/06). WIDER Research Paper.

Barkema, A. and Drabenstott, M. (2000). How Rural America Sees Its Future. The Main Street Economist, December, 1-4.

Baum-Snow, N., Henderson, J. V., Turner, M. A., Zhang, Q., & Brandt, L. (2020). Does investment in national highways help or hurt hinterland city growth?. *Journal of Urban Economics*, *115*, 103124.

Bloomberg (2019). China's Built a Railroad to Nowhere in Kenya. Retrieved March 31, 2021 from https://www.bloomberg.com/news/features/2019-07-19/china-s-belt-and-road-leaves-kenya-with-a-railroad-to-nowhere

Bosworth, G. (2009). Education, mobility and rural business development. Journal of Small Business and Enterprise Development.

Bureau van Dijk. (n.d.). [List of small companies incorporated from 1992 to 2016 in Kenya]. Retrieved May 15, 2020 from https://orbis.bvdinfo.com/

Central Bureau of Statistics (2004). Kenya Demographic and Health Survey, 2003, Central Bureau of Statistics, Nairobi, July 2004

Daniels, L., & Mead, D. C. (1998). The contribution of small enterprises to household and national income in Kenya. Economic Development and Cultural Change, 47(1), 45-71.

Devitt, C., & Tol, R. S. (2012). Civil war, climate change, and development: A scenario study for sub-Saharan Africa. Journal of Peace Research, 49(1), 129-145.

Dieke, P. U. (1991). Policies for tourism development in Kenya. Annals of Tourism Research, 18(2), 269-294.

Faye, M. L., McArthur, J. W., Sachs, J. D., & Snow, T. (2004). The challenges facing landlocked developing countries. *Journal of Human Development*, *5*(1), 31-68.

Fox, W. F., & Murray, M. N. (2004). Do Economic Effects Justify the Use of Fiscal Incentives?. Southern Economic Journal, 71(1).

Gallup, J. L., Sachs, J. D., & Mellinger, A. D. (1999). Geography and economic development. International regional science review, 22(2), 179-232.

Google. (n.d.). [Google Maps data on inter-city road distance on hundred most populous cities in Kenya]. Retrieved October 7, 2020, from https://www.google.com/maps

Havenga, J. (2010). Logistics costs in South Africa–The case for macroeconomic measurement. South African Journal of Economics, 78(4), 460-476.

Henderson, J. Vernon, Adam Storeygard, and David N. Weil. 2012. "Measuring Economic Growth from Outer Space." American Economic Review, 102 (2): 994-1028.DOI: 10.1257/aer.102.2.994

Johnson, A. T. (2013). University agency in peacebuilding: Perspectives on conflict and development in Kenya. Prospects, 43(3), 329-345.

Kamunge, M. S., Njeru, A., & Tirimba, O. I. (2014). Factors affecting the performance of small and micro enterprises in Limuru Town Market of Kiambu County, Kenya. International Journal of Scientific and Research Publications, 4(12), 1-20.

Kayanula, D., & Quartey, P. (2000). The policy environment for promoting small and medium-sized enterprises in Ghana and Malawi. University of Manchester.

Khadiagala, G. M. (2009). Regionalism and conflict resolution: Lessons from the Kenyan crisis. Journal of Contemporary African Studies, 27(3), 431-444.

Khandker, S. R., Bakht, Z., & Koolwal, G. B. (2009). The poverty impact of rural roads: Evidence from Bangladesh. Economic development and cultural change, 57(4), 685-722.

Kibara, O. N., Odhiambo, N. M., & Njuguna, J. M. (2012). Tourism and economic growth in Kenya: An empirical investigation. International Business & Economics Research Journal (IBER), 11(5), 517-528.

Kithae, P. P., Gakure, R., & Munyao, L. (2012). The Place of Micro and Small Enterprises in Achievement of Kenya's Vision 2030. Journal of US-China Public Administration, 9(12), 1432-1440.

Limao, N., & Venables, A. J. (2001). Infrastructure, geographical disadvantage, transport costs, and trade. The world bank economic review, 15(3), 451-479.

MacKellar, L., Wörgötter, A., & Wörz, J. (2000). Economic development problems of landlocked countries.

Manyara, G., & Jones, E. (2005). Development of an Indigenous Tourism SME Sector in Kenya. Tourism SMEs, service quality, and destination competitiveness, 59.

Mohajan, H. (2013). Poverty and economic development of Kenya.

McDaniel, K (2001). Small Business in Rural America. The Main Street Economist, May, 1-4.

Okpara, J. O., & Wynn, P. (2007). Determinants of small business growth constraints in a sub-Saharan African economy. SAM advanced management journal, 72(2), 24.

Okpara, J. O. (2011). Factors constraining the growth and survival of SMEs in Nigeria. Management Research Review.

Redding, S., & Venables, A. J. (2004). Economic geography and international inequality. Journal of international Economics, 62(1), 53-82.

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

Scheffran, J., Ide, T., & Schilling, J. (2014). Violent climate or climate of violence? Concepts and relations with focus on Kenya and Sudan. The International Journal of Human Rights, 18(3), 369-390.

Schumpeter, J.A. 1942. Capitalism, Socialism, and Democracy. New York: Harper & Row.

Sindiga, I. (1996). International tourism in Kenya and the marginalization of the Waswahili. Tourism Management, 17(6), 425-432.

Sindiga, I. (1999). Alternative tourism and sustainable development in Kenya. Journal of sustainable tourism, 7(2), 108-127.

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.

Tambunan, T. (2008). SME development, economic growth, and government intervention in a developing country: The Indonesian story. Journal of international entrepreneurship, 6(4), 147-167.

Thyne, C. (2006). ABC's, 123's, and the golden rule: The pacifying effect of education on civil war, 1980–1999. International Studies Quarterly, 50, 733–754.

UNIDO (2000). Annual Report UNIDO, 1999. United Nations Industrial Development Organization.

U.S. Energy Information Administration, (n.d.) [Europe Brent Spot Price FOB (Dollars per Barrel)]. Retrieved May 15, 2019 from https://www.eia.gov/dnav/pet/hist/rbrteD.htm

Wooldridge, J. M. (2006) Introductory econometrics: a modern approach. 3rd ed. Mason: Thomson/South-Western, c2006. ISBN 0-324-28978-2.

Yamakawa, R. (2007). Poverty reduction through tourism: The experiences in Asia (online). United Nations Economic and Social Commission for Asia and the Pacific. In Workshop on expanding the role of tourism in Poverty reduction. http://www. unescap. org/EPOC/meetings/WkshpPovertyReduction/English/Poverty-Reduction (RY). ppt (Vol. 375, No. 4).