University of Oxford

ROADS TO DEVELOPMENT?

THE DISTRIBUTIONAL EFFECTS OF MARKET ACCESS ON WAGES: EVIDENCE FROM KENYA 1992 - 2010

May 3, 2018

Luke Milsom, University College

Submitted in partial fulfilment of the requirements for the degree of the Master of Philosophy in Economics

WORD COUNT: 24,053

Abstract

This thesis analyses the heterogeneous and distributional effects of changes in transport infrastructure on city level wages in Kenya. To do this I develop a consistent theory, dichotomising industries into *tradeable* and *non-tradeable*, and adopting a sufficient statistic approach, using *market access*. To empirically evaluate these questions, I construct two novel data sets digitising historical Statistical Abstracts and Michelin maps between 1992 and 2010. This new data on wages by industry by year by city allows unique insights into the spatial distribution of economic activity in a Sub-Sahara African country: Kenya. I find that transport investment positively affects wages in Kenya, but that most of the impact is attributed to increases in tradeable sector wages, and thus to already relatively more developed cities. I also consider the direct effect of market access on development indicators and find consistent evidence of negative impacts. These results indicate that transport infrastructure projects may not be the silver bullet they are often purported to be. Finally, I consider a counterfactual scenario: building the proposed Mombassa-Malaba standard gauge railway. I find that although the project will raise wages in Kenya, it will also increase the spatial wage inequality already present, with the lion's share of benefits going to Nairobi and Mombassa.[1].

[1]The author would like to thank Dr. Ferdinand Rauch for excellent supervision, Sam Altmann for careful proof reading, participants of the Oxford trade seminar series for helpful comments, and Kai Xuan Ong for endless patience with my spelling and grammar. All errors remain my own.

Contents

1 Introduction

Little is known about the effect of transport investments on labour market outcomes in Sub-Saharan Africa (SSA). SSA is, however, the least urbanised and least connected region in the world[2]. Therefore, we expect, in both dimensions, large gains in the future. This motivates the importance of understanding the impact changes in the transport network will have on labour market outcomes. However, in order to answer these questions basic knowledge of what people do, and how much they are paid to do it, in the hinterland cities of SSA is required. It is a testament to the data limitations in this area that previous research at the intra-country level has focused on city size (Jedwab and Storeygard, 2016) or night lights as proxies for economic activity (Krugman and Venables, 1995; Henderson et al., 2012, 2017; Storeygard, 2016). This thesis goes further by considering the effect of changes in a SSA country's transport network on city level wages, and how this varies by industry type. This allows us to consider the impact of transport investments on inter-industry wage inequality, and in the case of Kenya's recent history, spatial wage inequality.

Cities, as hotbeds of creativity and productivity, are known to be drivers of economic growth. Therefore, successful urbanisation is crucial for national economic development (Glaeser, 2012). However, there is evidence to suggest that this conventional wisdom may not hold in a SSA setting. Recent research on urbanisation in SSA suggests it may be *different* (Henderson et al., 2013) and that SSA may be urbanising without industrialising (Gollin et al., 2016). This thesis sheds light on this discussion by analysing the effects changes in the transport network have on wages in cities, and how they may be driving industry and spatial inequality, contributing to the dysfunctionality of SSA cities.

A major contribution of this thesis is in the construction of two new datasets, allowing empirical analysis of these questions. Firstly, by digitising historical statistical abstracts from the Kenyan government I construct a balanced panel

[2]Its Urbanisation rate is one third compared to the worlds one half (Nations, 2014), and the number of paved roads per 1,000 people is a fifth of the global average (Bofinger, 2011).

of employment and wages of *major*[3] Kenyan cities every year from 1992 to 2010 by nine industries[4]. This allows us, for the first time, to study what people are doing in the hinterland cities of a Sub-Sahara African country and to look at changes over time. We find the labour force is surprisingly diverse: almost all cities have some significant employment in manufacturing, and indeed in all industries[5]. Conventional wisdom suggested that all higher-level industry, such as manufacturing, will be centred in Nairobi and Mombassa, following a core-periphery mindset. However, I find using this new data, that manufacturing can be found in almost all of Kenya's major hinterland cities, challenging the conventional wisdom. There is also some evidence of specialisation in specific industries in some cities, especially in manufacturing, and in general there exists considerable variation in industry shares and wages across cities.

I complement this dataset with a network dataset of the entire transport network of Kenya, Uganda and Tanzania between 1992 and 2010 constructed by digitising Michelin road maps and consulting historical records. Using these novel data sources, I can consider how changes in the transport network affected wages at the city-industry level within Kenya between 1992 and 2010. To inform my empirical work, I develop a general theory based on New Economic Geography (Fujita et al., 1999), combining the reduced-form market access (MA) approach of Donaldson and Hornbeck (2016) with a tradeable, non-tradeable industry dichotomy due to Venables (2017). From this novel theoretical approach, I derive a set of general predictions which can be tested in my specific empirical setting. However, to take these predictions to the data, I require a calculable market access variable. Market access is a measure of the size of the market a city has access to and is related to the cost-weighted sum of other city populations. Such costs are calculated by finding the least-impedance path between each city along the constructed transport

[3]A major city is defined as one with at least 1,000 people in wage employment or self-employment in 1973. 22 such cities are identified, see appendix B for a full list.

[4]See section 2 and appendix B for a full description of the data

[5]Almost all cities have employment in all industries except mining which only employs persons in a subset of urban centres close to mines in Kenya.

network using Dijkstra's algorithm. The market access variable captures all the complex linkages and spillovers between cities and can be derived from standard trade theory. The model delivers a reduced form empirical relationship between wages and market access and predicts that wages will be positively associated with market access, but that the effect on tradeable industries will be larger than that on non-tradeable ones.

Four key predictions can be generated from the model; motivated by these, I construct a simple empirical specification relating market access to wages. As with any measure of a transport network we maybe worried that changes in said network are related to local outcomes, such concerns surrounding the endogeneity of changes in the transport network are addressed by considering only 'far away' variation. That is, although one might think that the construction of a road to settlement X is endogenously determined with city X's economic outcomes, it is unlikely that the construction of a road 100km away from X would be[6]. I find, using this strategy, that market access has a positive impact on tradeable industry wages with an elasticity of approximately 0.28, and that the impact on non-tradeables is significantly lower with an elasticity of approximately 0.06. It is well known that transport projects will have spatially heterogeneous effects on economic outcomes, indeed such projects are often championed specifically because they are believed to lead to economic development in the immediate vicinity of said project. However, the interaction of spatial and industry heterogeneity will lead to unintended distributional effects. Even cities close to changes in the transport network will see little impact on wages if they have a low tradeable industry employment share. This finding is particularly important in the SSA setting as some cities may be stuck in a low tradeable equilibrium with further advances in the transport network only serving as to increase the gap between them and the already more developed cities. Through this mechanism changes in the transport network of SSA countries will have the effect as to increase spatial spatial wage inequality above that which is implied by it's geographic location.

[6]Examples of this identification strategy or variations thereof can be found in Donaldson and Hornbeck (2016) or Jedwab and Storeygard (2016).

The finding that changes in the transport network, through variation in market access, have the effect of increasing wage inequality may have serious ramifications on development outcomes and by extension the success of cities. In order to directly study this issue I develop a third novel dataset combining the results from six surveys[7] to construct consistent development indicators at the city level. I regress these against market access to find their conditional correlational association with wages. I find that market access has a negative impact on all indicators, highlighting the validity of our concern.

Finally, I utilise the structural nature of the estimated model to study a counterfactual scenario of particular importance. I consider the potential impact of the proposed (and in construction) building of the Mombassa-Malaba standard gauge railway, a £7million project purported to transform the economy of the region (as well as that of Uganda). I find that this project may have unintended distributional effects as the majority of the benefits will be felt in the already developed and industrialised Nairobi and Mombassa, further increasing spatial inequality in Kenya.

1.1 Contribution to the literature

To derive the main contributions of this thesis to the literature, I first provide a short synopsis of the existing relevant works. The main motivation for this thesis lies within the field of urban economics. Lucas (1988) said that "cities are the fundamental driving force behind economic development" and this sentiment has been echoed many times, notably by Glaeser (2012) in *Triumph of the city*. The idea that cities are places of high productivity where innovation happens and countries develop has become commonly accepted as fact. Therefore, the study of cities, and what determines their success, is crucial for promoting growth and development at the national level. This thesis attempts to delve deeper into the effects of the city-to-city transport network on city level outcomes and to shed light on the effectiveness

[7] Three censuses and three DHS surveys, that is all available nationally representative surveys in Kenya geolocated at sufficient detail over the study period 1992-2010.

of infrastructure projects on promoting development within cities.

To study the relationship between transport infrastructure and city outcomes in a clear and general manner, I utilise formal models grounded in the seminal works of New Economic Geography (Krugman, 1991; Fujita et al., 1999). This literature attempts to explain the spatial distribution of economic activity and the phenomenon of cities through agglomeration effects. I take this structure as my starting point and follow Donaldson (2010), Donaldson and Hornbeck (2016) and Hering and Poncet (2010) in deriving a parsimonious sufficient statistic approach to summarise the effects of spatial factors on wages. I develop this further by modelling tradeable and non-tradeable industries separately following Venables (2017). Venables (2017) points out that if there are two sectors, tradeables and non-tradeables, available to each city, one which responds to world prices and demand and the other only to local, then increasing returns to tradeable firms (agglomeration effects) will lead to multiple equilibrium. It is possible one city could be in a bad, high non-tradable equilibrium and another ex-ante identical city in a good, high tradeable equilibrium.

The empirical literature on using New Economic Geography models to explain geographical wage differences is sparse, firstly because of the relative infancy both of the theory and the empirical methods required to robustly test it, and secondly due to the large data requirements. Hanson (2005) documents a positive correlation between market access and wages, Head and Mayer (2006) find a similar effect across regions in the EU, Paillacar (2006) corroborates these findings in Brazil but stresses the importance of education as well as MA in wage determination. Perhaps closest to this thesis is the work by Hering and Poncet (2010) on the effects of market access on wages in China. Hering and Poncet (2010) find, that their estimated positive relationship (with elasticity between 0.1 and 0.2) of market access on wages only holds for certain firm types: foreign and private firms, which are more likely to be in the tradeable sector. Similarly, Fallah et al. (2010) find a positive effect of market access on wages in US metropolitan areas, but that this effect is much larger for those at the top of the wage distribution, and subsequently that market access drives wage inequality. As higher wage industries are more likely to be in the tradeable sector results from this thesis can explain Fallah's findings as well as provide a channel through which the effects found in Hering and Poncet could move, generalising both results within a consistent framework. These comparisons are also notable for their geographic dispersion, evidence from China, America and now Africa support the model and results developed in this thesis.

This thesis, however, focuses on Sub-Saharan Africa, where much less is known. Previous research in this area has focused on using colonial rail roads and crude development proxies such as population or night lights. Jedwab et al. (2015) find positive and persistent effects of railroads on economic growth in Sub Saharan Africa. Corroborating this, Storeygard (2016) finds positive effects of decreases in trade costs on night lights for cities closer to ports. This literature remains small and limited to asking basic questions, due to the severe lack of data availability in SSA. However, it is here where answering such questions is most important and has the potential to have the largest impact on welfare. It is also of particular importance to rigorously address questions relating large transport infrastructure spending to welfare outcomes due to the burgeoning literature suggesting that urbanisation in Africa may be 'different', and therefore, that the results derived from more developed countries may not be generalisable to this region. Henderson et al. (2013) suggest exactly that urbanisation in SSA is different, and that SSA responds differently to both agriculture and manufacturing shocks. Gollin et al. (2016) coined the phrase "Urbanisation without industrialisation", which highlights how, unlike European or American cities, those in SSA are not experiencing large increases in industrialisation as they urbanise.

The main contributions to the literature of this thesis lie in increasing our understanding of how changes in the transport network affect economic outcomes (through MA) in a SSA country and the heterogeneity of these effects both across industry type as well as geography. One can derive four main contributions of this thesis:

- Creating large, novel and detailed data sets on the spatial distribution, across time, of economic activity in a SSA country (Kenya). This unique data set allows us, for the first time, to answer basic facts about what people do in hinterland cities in Kenya and rigorously analyse the heterogeneous impacts of transport investment on wages.

- A theoretical contribution of combining the market access approach of Donaldson and Hornbeck within a tradeable, non-tradeable industry dichotomy of Venables. This novel and general NEG model allows us to derive testable predictions to inform and validate our empirical work.

- Find evidence of the heterogeneous effects of market access (and so transport infrastructure investments) on wages across different industry types in Kenya. Explore the distributional implications of these findings.

- Look directly (using another new dataset) at the development implications of increasing connectivity. Give evidence and theory-based policy recommendations regarding the (non-obvious) effects of future transport projects on wages, especially concerning the unintended impacts on spatial wage inequality.

Appendix C considers the broader historical context of Kenya and the East African Community from independence[8] to the present day, but with particular reference to the period under consideration: 1992-2010. Any serious empirical work in economics requires a holistic understanding of the institutions and history under consideration. Therefore, although relegated to an appendix, I consider this work of crucial importance in assuring the validity of my empirical estimates.

The remainder of this thesis is organised as follows: section 2 gives a detailed description of the three new datasets I have constructed, section 3 considers the main facts we learn from this novel data, section 4 develops a novel theory of the potentially heterogeneous effects of market access by industry type, section 5

[8]That is independence from colonial British rule.

constructs the empirical methodology and identification strategies, section 6 reports the empirical results and section 7 concludes.

2 Data

In order to study the central questions of this thesis I require city level, geo-located, temporal and industry specific data, in an area of the world (Sub Saharan Africa, SSA) in which less than half of the countries have ever had a census. Such granular data is unheard of in SSA. To construct the novel dataset used in this thesis, I combine three main types of data from 33 sources, summarised below.

1. Data on labour market outcomes, wages and employment, by industry, by city, by year. This data was constructed by digitising 21 historical Kenyan Statistical Abstracts.

2. A full description of the transport network of Kenya, including variation in road type as well as transport type, over time. I construct this data by digitising six historical Michelin maps, maps of the rail network and accounts of the Lake Victoria ferries.

3. Measures of development, by city, by year. To formulate such specifically geo-located development data I merge three Kenyan censuses and three DHS surveys.

Below I provide a more detailed description of each dataset and its construction.

2.1 Labour market data

Kenyan Statistical Abstracts are produced annually by the Central Bureau of Statistics (CBS) in Kenya and consists of a collection of surveys reporting on the state of the Kenyan economy in the previous year. In this thesis I am particularly interested in the *Labour* section of the abstract, which is derived from the Annual Enumeration of Employees and Self-employed Persons. One feature of this survey is a set of tables containing the number of individuals in wage employment and their total earnings for all *major* towns in Kenya by nine industry classifications. This data is available yearly over the period under consideration, 1992-2010.

The survey defines major towns[9] as those with 1,000 or more persons engaged[10] in 1972, a list of the 22 cities adhering to this definition is printed in appendix B. Note that the smallest city, Webuye, is still very much urbanised, it is densely populated with a population of over 20,000 in the 1999 census. The survey covers all establishments operating from a permanent address[11] in urban areas, this list of establishments is considered complete as it is a legal requirement to register and it would be difficult for a firm to operate without registering. The enumeration occurs in June with surveys sent some time prior, non-responders are sent the surveys a second time and if they still do not respond firms with over 20 employees are visited personally by a CBS worker. Due to this the overall rate of respondents is high at over 60%. In the case of non-responses, the previous year's data is taken, if a firm hasn't responded for two consecutive years, it is assumed that employment estimates follow an average pattern of firms in the same sector and location which have responded for three consecutive years or more.

The survey defines those in wage employment as including casual employees, part-time workers, directors and partners but does not include self-employed persons or family workers who do not receive a regular wage. Notably employment in the informal sector is excluded, as is small scale agricultural work. These two categories are likely to constitute a large proportion of the working population, even in urban centres in Kenya and their exclusion is significant. Earnings are defined as covering all cash payments, this includes the basic salary as well as cost of living allowances, bonuses and the value of free board, however they exclude pension contributions. Earnings are calculated as those received in June multiplied by 12.

[9]This thesis will refer to major towns satisfying this above definition as *cities* and all other urban agglomerations as *settlements*.

[10]A person is engaged if they are in wage employment or are self-employed.

[11]Businesses in Kenya which operate from a permanent address are required to register a licence from the Kenyan government.

Employment and earnings are disaggregated by city and industry. Nine industries are considered broadly following the standard ISIC codes. These industries are:

1. Agriculture, forestry and fishing (Agriculture). Although small scale agricultural activity is not included in the survey, large scale farms with permanent employees are and consist of a surprising fraction of employment in some cities.

2. Mining and quarrying (Mining). Kenya does not have any significant mineral deposits, and therefore mining constitutes a very small proportion of employment. No persons are employed in mining in many city-years.

3. Manufacturing. Manufacturing in Kenya, although being concentrated in Nairobi and Mombassa, consists of a surprisingly large proportion of economic activity across a wide variety of cities. However, the majority of manufacturing performed in Kenya is relatively small scale and unindustrial with over fifty percent of those employed in manufacturing working in food processing or textiles.

4. Electricity and water (Electricity). These industries are concerned with the supply of amenities and include waste collection.

5. Construction. Includes all persons involved in the building and design of structures both private and public.

6. Whole sale, retail trade, restaurants and hotels (Retail). In Kenya employment in this section is roughly distributed a quarter in wholesale, a quarter in restaurants and hotels and half in retail.

7. Transport and communication (Transport). Includes all transportation, storage, publishing, broadcasting or other communication activities. In Kenya roughly half of those employed in this category are employed in transportation and the other half in communication.

8. Finance, insurance, real estate and business services (Finance). Also includes data processing, technical services like engineering and book keeping/ auditing.

9. Community social and personal services (Services). Includes government services, educational services, medical services and domestic services.

Industry definitions are consistent over time and over cities. This data set is unique for SSA in its disaggregation both by geography and industry, and temporal consistency, for any country in Sub-Saharan Africa and thus allows researchers a unique view into Kenya. However, this survey is not without limitations. Although its coverage appears to be universal there always remain concerns over any data collected by survey, it is possible that firms may have had incentives to misreport employment numbers or earnings, and regardless it is likely both are measured with considerable noise. The glaring omission of the informal sector is also of importance as the majority of those in work in Kenya are likely to be doing so informally, empirical results should be read with this in mind.

2.2 Transport network

I require a complete representation of Kenya's transport network over time from 1992 to 2010 in order to study the effects of variation therein on labour market outcomes. Additionally, it is necessary to also consider variation in the transport networks of Uganda and Tanzania. This is for two main reasons, firstly the border between Kenya its two main trading partners Uganda and Tanzania, is permeable and set to become increasingly a formality as the East African Community (EAC[12]) grows closer. Therefore, it is likely that variation in the transport network in Uganda and Tanzania will impact Kenyan cities close to the border and thus Kenya as a whole, in order then to capture this important source of variation I must model the network in these countries in the same was as it is modelled in Kenya. Secondly as will be discussed in section 6, there may be concerns over endogeneity issues surrounding

[12] The EAC is an organisation comprising of six partner states including Kenya, Tanzania and Uganda. See appendix C.

the placement of roads, one way these can be tackled is to consider only variation in the transport network from outside Kenya, to do this one must necessarily model the whole network of Uganda and Tanzania.

Due to its lack of navigable rivers, Kenya's transport network consists of three elements. Firstly, it's rail network. Built circa 1900 by the British, it has since fallen into decline due to neglect and underinvestment. Today much of the colonial line is unusable, and that which remains[13] is painfully slow by modern standards, travel along the Mombassa-Nairobi line proceeding on average at the snail pace of 24.8km/h[14]. The usable length of rail road has not changed over the period under consideration, as is evidenced from the Kenyan Statistical Abstracts which detail rail track length, thus we take the rail roads, and stations, from Open Street Maps[15] and project this backwards until 1992. It is impossible, with the available data, to determine heterogeneity in travel times across lines or across time, however on both dimensions we expect little variation over the study period, much of the decline in Kenya's rail roads happened before 1992. Thus, it should be noted that no identifying variation comes from rail roads, it is important to bear this in mind when considering the validity of counterfactuals concerning rail road expansion.

Secondly, and most importantly for our analysis, Kenya has an extensive road system, which since the continued decline of the railways has been increasingly important both for transporting persons and goods. I construct the road network over time for Kenya, Uganda and Tanzania using an approach common in the literature of digitising historic Michelin road maps[16]. I first download Open Street Map data for the region. OSM maps are extremely detailed covering every known road (and settlement) at the time of downloading (2nd Nov 2017). However, OSM has two important drawbacks, firstly information on road quality is inconsistent and

[13]The main line from Mombassa to Uganda and some subsidiaries remain, see figure 1

[14]Information on rail speeds has been taken from the World Food Programme Logistics.

[15]Open street map is a comprehensive map of the world built by the community, it is effectively an often more detailed, significantly more up to date and freely downloadable version of Google Maps.

[16]See Jedwab et al. (2015), Jedwab and Storeygard (2016) and Burgess et al. (2015) for examples of this approach.

incomplete, secondly it provides a snap-shot of the network only and does not have information on changes over time. To address both drawbacks I couple OSM data with digitised historic Michelein maps. Michelin maps are created by French cartographers and so provide an independent view of the road network. These maps are also only as useful as they are accurate and so Michelin has a strong incentive to provide accurate maps. The main sources from which Michelin base their maps are their previous maps, ad-hoc revisions to existing maps from drivers who use Michelin tyres and report inaccuracies and knowledge of larger infrastructure projects (Jedwab and Storeygard, 2016). Michelin maps are likely to provide an accurate representation of the road network under consideration, Michelin also use a consistent categorisation of road types thus allowing variation by road quality. This is of particular importance as over the study period road improvements, rather than road building, is the main source of variation in the road network. Michelin maps, although an exciting new data source allowing a previously impossible view into the history of transport networks, have two main limitations. Firstly, Michelin maps were not produced every year, only maps in the years 1992, 1998, 2000, 2003, 2007 and 2012 are relevant for this study. Secondly, variation within category is not observable[17], this is particularly important over the time period considered as it is likely that there was considerable, unobserved, variation along this dimension.

Lastly, to complete the transport network, I consider possible routes across Lake Victoria. Africa's largest lake, Lake Victoria, is bordered by Uganda, Tanzania and Kenya. It therefore constitutes an opportunity for integration and trade between these three, already close, countries. However, it has been underutilised, the only real large scale connection across its waters being a small fleet of rail ferries which connected the rail networks of all three countries. Operating since the mid-20th century, these provided a small but important connection, however they fell out of use in 2006. I therefore add these connections to the transport network completing it. Figure 1 shows the complete transport network in 2010. Thick black to thin grey

[17]That is resurfacing and general maintenance of roads rather than improvements.

lines represent the road network, largest paved roads being the darkest and thickest. Green lines denote railways and red lines national borders.

Figure 1: COMPLETED TRANSPORT NETWORK

Notes: this figure shows the complete transport network of Uganda, Tanzania and Kenya in 2010. Red lines denote country borders, green lines railways and black lines roads. Thick, dark black lines denote primary improved roads, dark grey secondary improved roads, light grey partially improved roads and thin light grey dirt tracks.

2.3 Development indicators

Previous research in this area has always used some non-perfect proxy for development, either night lights or population. A major contribution to the literature of this paper is its novel addition of more concrete development indicators. Population, or even light intensity increases, need not imply advances in development or welfare of the area under consideration. Specifically, with such measures it is impossible to distinguish between mere density increases and actual development advances. To combat this gap in previous research I construct a novel dataset of actual development indicators using data from every available geo-located survey between 1989 and 2010 in Kenya. In total I use data from 6 individual surveys either from the development health survey (DHS) or national censuses.

From these 6 surveys consistent measures can be derived for the following variables.

- Roof/ wall/ floor construction material. House construction material, and particularly roof covering, can be highly indicative of development levels in SSA countries.
- Literacy rate and level of education.
- Access to sanitation (i.e. flushing toilet) and availability of drinking water.

This data suffers from the usual problems of survey data, which can be especially relevant in SSA. Censuses in SSA countries are notoriously unreliable, there may be political or economic reasons to alter survey results. These concerns are likely to be lessened in the DHS surveys, although responder errors are possible. In addition to this, development indicators are only available for six out of a possible 21 years. Considering all these limitations, results from this section are suggestive only.

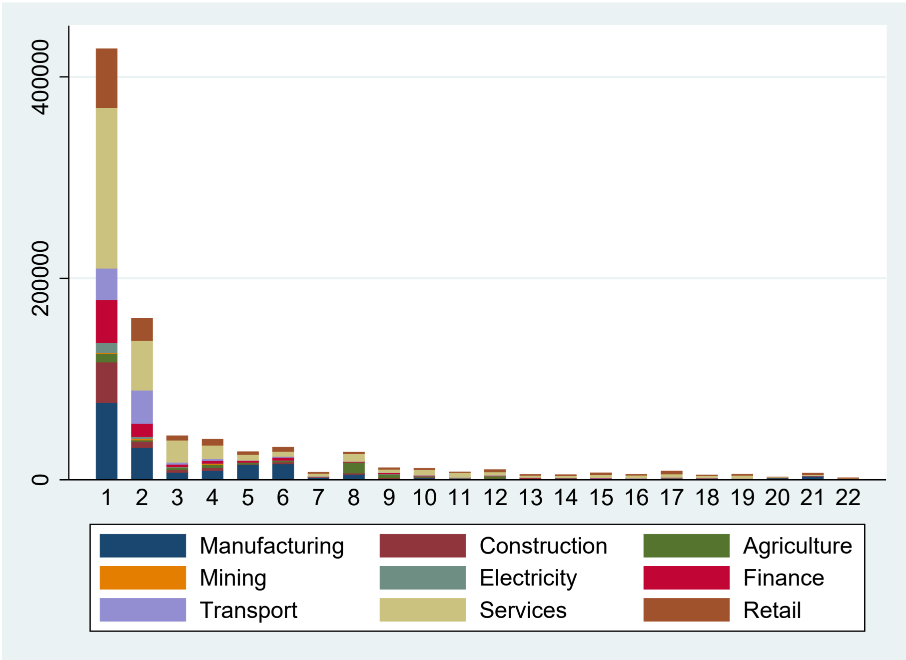
To summarise, combing all possible sources of data at the city level, over time, in Kenya, I have formed the most comprehensive dataset on city level outcomes for any Sub-Sahara African country. Combining known data sources with new data derived from the digitisation of historical maps and records. This novel data set allows us, for the first time, to ask basic questions concerning the spatial distribution of economic activity in Kenya, as well as study the more complex questions asked in this thesis concerning the distributional effects of transport investments on wages.

3 The spatial distribution of economic activity in Kenya over time: facts

With this new and comprehensive dataset in hand, we can turn to answering the central questions of this thesis. However, in order to study how changes in the transport network may affect wages in different Kenyan cities and industries, we need first to understand what labour markets look like in the hinterland cities of Kenya, and how they have evolved over time. Using this data, we can answer questions regarding how industry employment shares vary with city size and how large and small cities differ from one another in a Sub Sahara African country. These facts are interesting in and of themselves and a valuable contribution to the literature. Additionally building up a picture of what urban labour markets look like in Kenya, and how employment is spatially distributed, is key to understanding the implications of the empirical analysis conducted in this thesis. The facts presented here motivate the need for a formal model developed in the next section in order to test, in a general sense, how changes in Kenya's transport network affects wages across space and industries. I report only the most relevant facts, but further descriptive figures are presented in appendix D.

First, a well-known stylised fact, is the dominance of Nairobi and Mombassa in terms both of total employment but also in terms of employment in higher level industries such a manufacturing. I document and confirm this fact with figure 2 which averages employment within each industry over the sample period for each city. Cities are ordered by their 1990 employment rank, thus city one is the largest city (Nairobi) city two the second largest (Mombassa) and so on. Nairobi dominates, with a wage-employed population significantly over double that of Mombassa, indicating possible evidence of excess primacy in Kenya.

Figure 2: AVERAGE WAGE EMPLOYMENT ACROSS CITIES

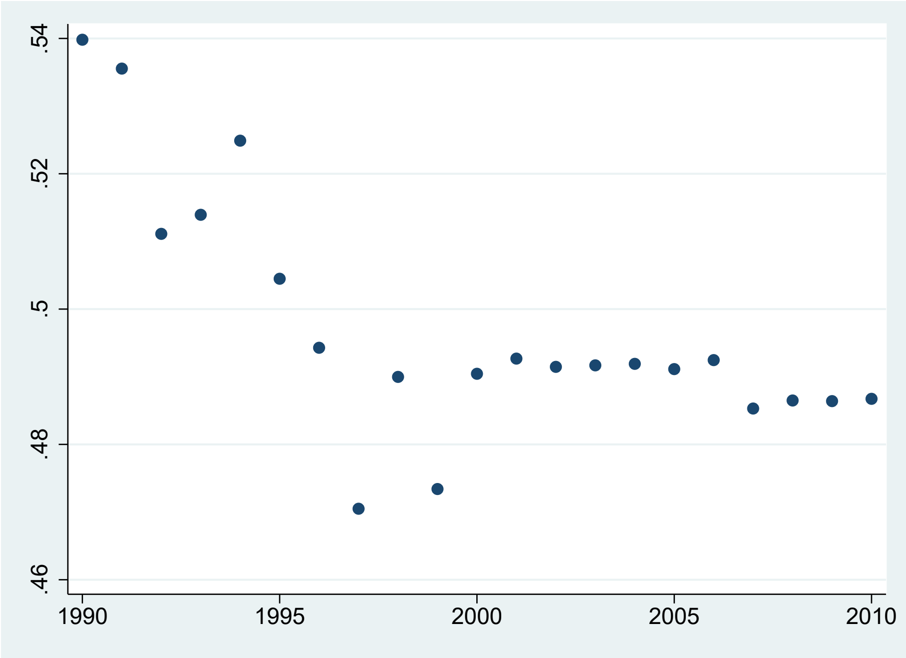


Notes: This figure compares the average (over the period 1992-2010) number of persons in wage employment in each of the 22 cities considered disaggregating by employment type. Note cities are ranked by their 1992 number of wage employed, thus city 1, Nairobi is the largest, city 2 Mombassa the second largest and so on.

Nairobi dominates, but its position is changing. In figure 3 we look at the changing dominance of Nairobi over time. As documented above, the Urban population in Nairobi has expanded massively over the study period, Nairobi has grown with this increase, but as figure 3 shows, at a slower rate. In 1990 54% of the wage employed worked in Nairobi, however by 2010 this had decreased to 48.5%, Nairobi's working population has relatively decreased.

Figure 2 and 3 are uninformative regarding the employment shares of smaller cities. Knowing what people do in these hinterland cities would be interesting for this study and generally little is currently known about employment patterns in the hinterland cities of SSA. Using this new dataset, we can observe the spatial distribution of economic activity across cities by industries. We can look at what exactly it is people do in the hinterland cities of Sub Saharan Africa. Figure 4 shows the average proportion (over the sample period) employed in each industry for each city. Certainly, it is true that the majority of manufacturing occurs in Nairobi

Figure 3: NAIROBI'S RELATIVE WAGE EMPLOYMENT

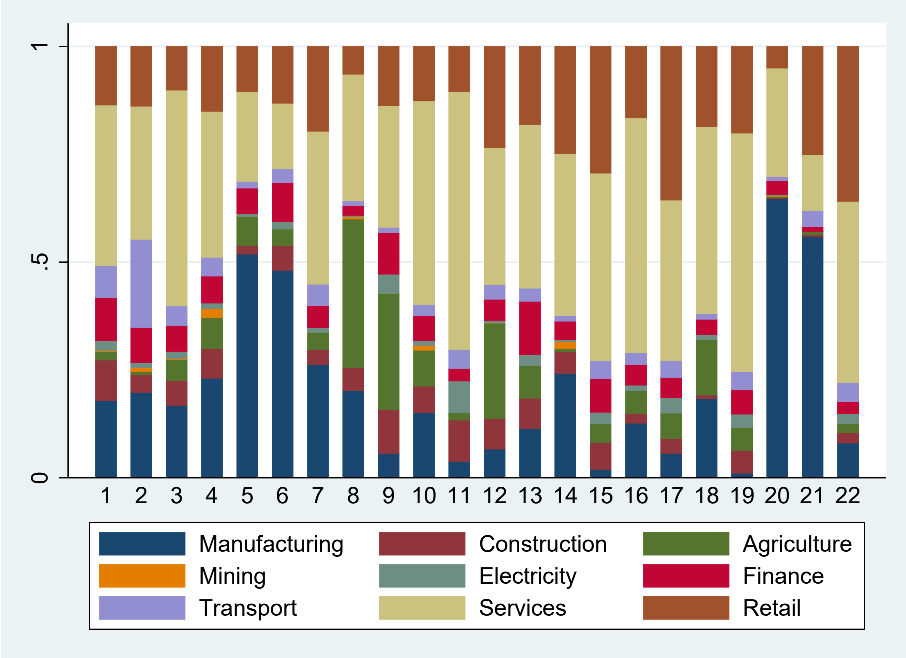


Notes: This figure shows the number of wage employed in Nairobi as a percentage of the total number of wage employed in all 22 cities of this sample.

and Mombassa, by virtue of their dominating populations, but interestingly figure 4 shows that every city in our sample engages in some degree of manufacturing. In general, figure 4 shows the perhaps surprising result that people seem to engage in a wide variety of economic activities in every city sampled.

This figure maybe surprising to some, it seems to provide evidence somewhat against the consumption cities hypothesis, or the bad-equilibrium outcome of Venables (2017) where cities are stuck with no tradeable sector. The variance in manufacturing is high, but we see evidence of activity in almost every city, however on closer inspection the picture becomes clearer. Manufacturing in Kenya is not what is traditionally thought of as manufacturing. Most activity is in *soft* manufacturing: textiles or food processing. Additionally, one can pick out some specific cities of interest. It appears that cities ranked 5,6,20 and 21 have an unusually high proportion of individuals working in manufacturing. The fifth ranked city, Thika, is a food processing hub, the sixth Eldoret is the main centre of textile production in Kenya. The 20th and 21st, Athi River and Webuye respectively specialise in cement production and paper and sugar processing. There is, therefore, some evidence of

Figure 4: SPATIAL DISTRIBUTION OF EMPLOYMENT BY INDUSTRY

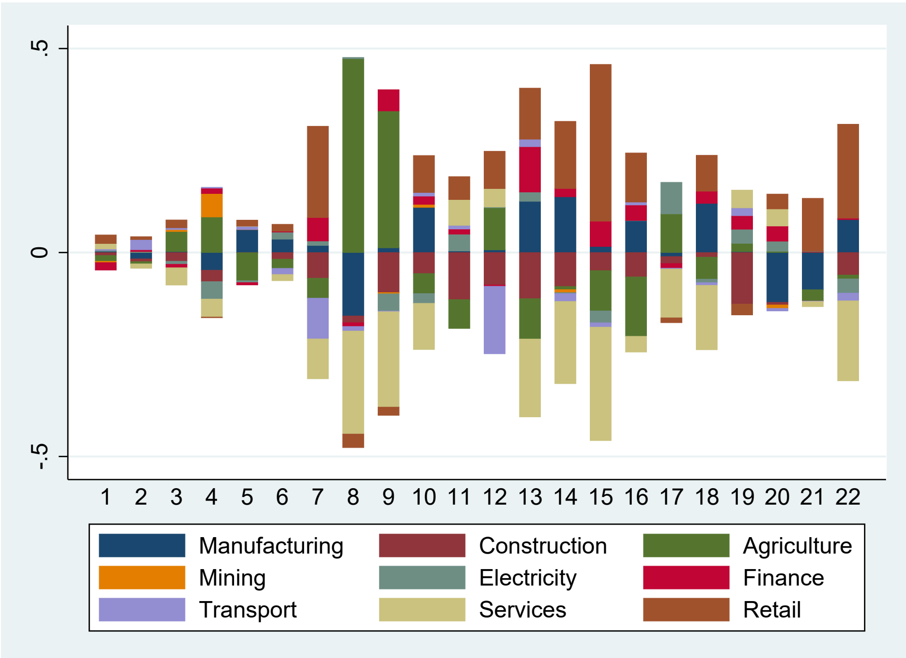


Notes: This figure shows the proportion employed in each industry in each city, averaging over the study period 1992-2010. Note cities are ranked by their 1992 number of wage employed, thus city 1, Nairobi is the largest, city 2 Mombassa the second largest and so on.

returns to specialisation and agglomeration economies at work. Also, of interest are cities ranked 8 and 9, which have an unusually high proportion working in wage employment within the agricultural sector. These cities, Nyeri and Kericho are both located in highly fertile countryside and are agricultural centres as reflected in the data. Finally, the city ranked 17, Malindi, has the highest proportion of retail workers of any city in the sample. This is because Malindi, located on the coast close to Mombassa is a popular tourist destination, and can therefore support more retail shops than other cities.

Figure 4 averages over the whole sample period, also of interest is how employment shares have changed between 1992 and 2010 in each city, figure 5 shows this. Two trends are notable. Firstly, corroborating the findings of figure 14, there are no systematic increases in the proportion employed in manufacturing, suggesting again that Kenya is not industrialising at any faster pace than it is urbanising. Secondly there is a general trend for an increase in the proportion employed in the retail industry at the expense of the service industry, this is an interesting trend and may give evidence in the support of the rise of consumption cities. However, the mechanisms behind this trend are not well understood, this is an area where further research could be fruitful.

Figure 5: CHANGE IN THE SPATIAL DISTRIBUTION OF EMPLOYMENT BY INDUSTRY



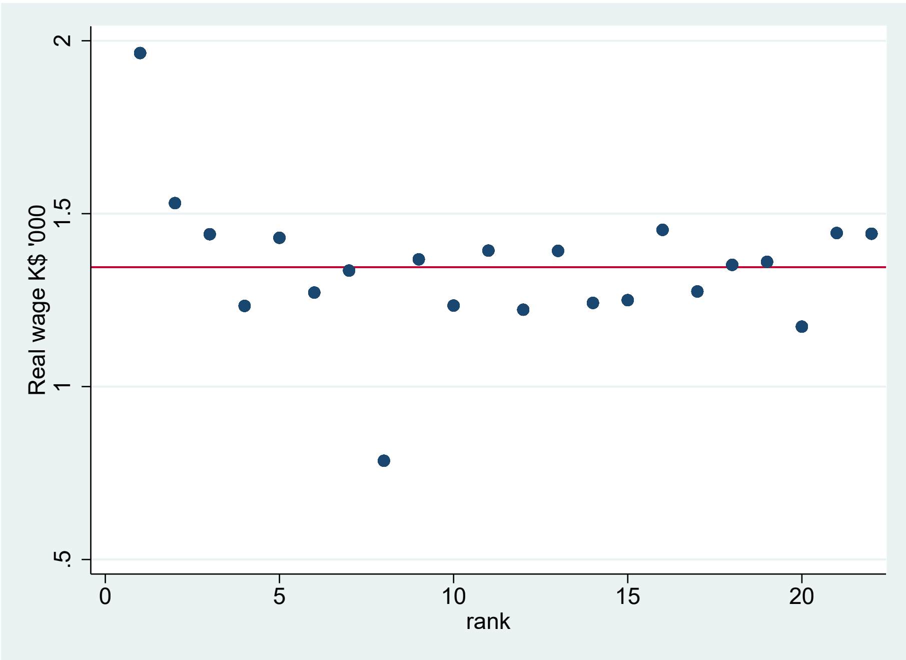
Notes: This figure shows how the proportion employed in each industry in each city has changed between 1992 and 2010. Positive bars indicate that the proportion employed in that industry in that city increased between 1992 and 2010, negative bars represent the corresponding decrease in a different industry. Note cities are ranked by their 1992 number of wage employed, thus city 1, Nairobi is the largest, city 2 Mombassa the second largest and so on.

Similarly figure 6 shows the wage distribution by city, the horizontal red line showing the average. As expected this figure shows that wages are higher in Nairobi and Mombassa thereafter they drop off. This evidence shows that our data supports the stylised fact that larger cities command higher wages Behrens and Robert-Nicoud (2015) suggest that wages and city size have an elasticity between 2% and 10%. We can test this more formally in our data by regressing log real wages on log city size, proxied by log wage employment. To this end I estimated equation 1 below which includes city and year fixed effects C_c and T_t respectively.

$$\ln\left(w_{ci,t}\right) = \beta \ln\left(N_{ci,t}\right) + C_c + T_t + \varepsilon_{ci,t} \tag{1}$$

I estimate a coefficient of 0.091 significant at the 1% level for β, indicating that wages increase by almost 0.1% for each percent increase in the number of wage employed. The estimated elasticity falls within the range anticipated in the literature. I do not give a causal interpretation to this result, issues surrounding the endogeneity of employment are rife and not addressed here, however even as a conditional correlation the result that Kenyan cities adhere to this stylised fact is interesting.

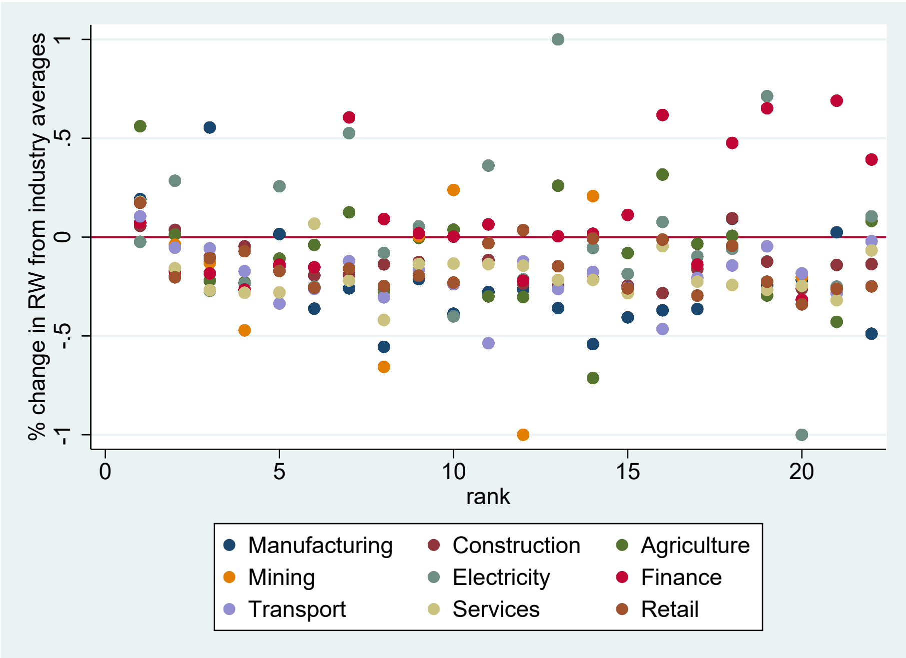
Figure 6: AVERAGE REAL WAGE BY CITY



Notes: This figure shows the average real wage (in 1992 prices) of each city, averaging over all nine industries and the time period considered. Note cities are ranked by their 1992 number of wage employed, thus city 1, Nairobi is the largest, city 2 Mombassa the second largest and so on.

There are two possible reasons for higher wages in larger cities. Firstly, these cities could have a higher proportion employed in higher paying industries. Secondly, the same industry could be paid more in bigger cities. Figure 7 sheds some light on this matter. Figure 7 plots, for each city the wage of each industry within that city, relative to the industry wide average, all averaged over the sample period. That is if an industry is above the red line it means that those working in that city, in that industry, are paid relatively more than the average for that industry over all cities. Here there are a couple of points to note. Firstly, almost all industries in Nairobi are paid above average, this supports the hypothesis that even within industry there is a city-size wage premium. This does not exclude the possibility that high paid industries are disproportionately represented in larger cities but gives evidence to suggest that this at least isn't the whole story. Secondly, due to the size of Nairobi and the fact that wages are relatively higher there, most of the relative wages in other cities must lie below the red line. Thirdly, the variation in relative wages, is notably large and present at all city size levels.

Figure 7: RELATIVE INDUSTRY REAL WAGE BY CITY



Notes: This figure shows the wage rate of each industry in each city relative to that industries average wage across all cities. Thus, dots above the red line indicate that in that city that industry is paid more than the average across all cities. All quantities are calculated by averaging over the time period considered. Note cities are ranked by their 1992 number of wage employed, thus city 1, Nairobi is the largest, city 2 Mombassa the second largest and so on.

These reported facts, as well as being a valuable contribution in and of themselves, provide us with a deeper understanding of the underlying variation in employment and wages, across space, time and industries. We find significant variation across both space and industries, implying significantly imperfect labour mobility in Kenya and thus scope for changes in effective distances between cities (that is changes in the transport network) to affect wages. This considerable variation motivates a deeper study into the underlying causes, here we focus on understanding how changes in the transport network of Kenya may be driving the observed differences. The large industry share and wage inequalities across space suggests that differing transport/movement costs may play an important determining role. To understand the mechanisms behind these differences, it is necessary to formally model the effect of a transport network on wages, to do this I develop a general and novel model, taking a market access approach.

4 Theoretical underpinnings

I develop a formal model to increase our understanding of how transport costs affect wages in different cities and industries, and inform my empirical work, as well as nest it within a general framework. Following Hering and Poncet (2010), I draw on Fujita et al. (1999) and Redding and Venables (2004) in setting up a New Economic Geography (NEG) model of wage determination in a spatial setting. This model expands on the existing literature in a novel manner, by combining the general sufficient statistic result of taking a market access approach to summarising spatial effects in an economy, with heterogeneous industries which vary in their responsiveness to external markets. From this novel and general model, I derive testable predictions which inform my subsequent empirical analysis.

Consider $\mathcal{C}$ city locations and two types of goods: tradeables $\mathcal{T}$ and non-tradeables or *local* goods $\mathcal{L}$[18]. The demand and supply of local goods, by definition, is only determined within-city, that is by the local market. Examples of such goods are abundant, for example; one would not consider hairdressers in London to be impacted by a new high-speed rail connection to Birmingham through any mechanism other than the effect such a connection may have on the size of the local market in London. We model consumer utility as a Cobb-Douglas aggregate over the two types of goods $U = X_{\mathcal{T}}^{\theta} X_{\mathcal{L}}^{1-\theta}$ where a fraction θ of income is spent on tradeables. Take $\theta \in (0, 1)$, that is consumers spend a non-zero proportion of their income on both tradeable and nontradeable goods. Local goods, conditional on the size of the

[18]Throughout this thesis I will use the terms non-tradeable and local interchangeably to refer to the group of industries which do not produce goods that can be traded on markets outside the city in which they are produced.

home market, are not affected by changes to the transport network of a country and therefore require a separate model to tradeables. We turn first to modelling tradeables.

Firms have increasing returns to scale and produce differentiated tradeable goods modelled in the Dixit-Stiglitz manner. Consumers have utility increasing in the number of varieties modelled in a CES form $U_A = \left(\int_0^n m(i)^\rho di\right)^{\frac{1}{\rho}}$ where n is the measure of varieties and $\sigma = \frac{1}{1-\rho}$ is the elasticity of substitution between such varieties. The aggregate CES price index for each location $c \in \mathcal{C}$ as $G_c = \left(\int_0^n p_{ci}^{1-\sigma} di\right)^{1/(1-\sigma)}$ where p_{ci} is the price of variety i in location c. Solving the consumers maximisation problem assuming CES preferences as outlined above we find that the demand from location c for goods produced in r is given by.

$$D_{cr} = p_{cr}^{-\sigma} \frac{E_r}{G_r^{1-\sigma}} \tag{2}$$

Where $E_r/G_r^{1-\sigma}$ is the price-weighted total expenditure on manufacturing goods in location r and $p_{cr}^{-\sigma}$ is the price of a variety from r in c. This price is modelled as a mark-up on the mill price of production in r given by the cost of transport modelled in the standard iceberg manner thus denote $p_{cr} = \tau_{cr} p_r$. p_r is the mill price in r and $\tau_{cr} > 1$ captures the increased cost of transporting goods between c and r, which we assume to be symmetric. Equation 2 can easily give us the effective demand faced by a firm in c from location r, taking into account transport costs we have

$$x_{cr} = \tau_{cr} p_{cr}^{-\sigma} G_r^{\sigma-1} E_r = p_c^{-\sigma} T_{cr}^{1-\sigma} G_r^{\sigma-1} E_r \tag{3}$$

We further assume for simplicity that the differential cost of varieties is so small that each firm will produce only one variety, and on the supply side that firms face a fixed cost F_c and marginal cost m_c of producing. Thus, increasing returns to scale come through the location specific fixed cost F_c. Denote q_{ci} as the amount supplied by firm c to market (that is location) i, then each firm maximises profits in each market given by

$$\pi_{ci} = (p_c - m_c)\tau_{ci}q_{ci} \tag{4}$$

Solving the maximisation problem, we find that the mill price is a simple mark up over marginal costs $p_c = \frac{\sigma}{1-\sigma}m_c$, therefore all varieties produced in any given location are valued at the same price. The gross profit in each individual market i for a firm in location c is thus given by $\pi_{ci} = p_c x_{ci}/\sigma$. Total *net* profit for a firm in location c is giving by summing over all markets i.

$$\Pi_c = \sum_i p_c \frac{x_{ci}}{\sigma} - F_c = \frac{1}{\sigma} m_c^{1-\sigma} \sum_i \tau_{ci}^{1-\sigma} G_i^{\sigma-1} E_i - F_c \tag{5}$$

We follow convention and define market access for a given location c as

$$MA_c = \sum_i \tau_{ci}^{1-\sigma} G_i^{\sigma-1} E_i \tag{6}$$

The variable costs which set profits equal to zero can be found by setting 5 to zero and rearranging, doing this we find.

$$m_c = \left(\frac{MA_c}{\sigma F_c}\right)^{\frac{1}{\sigma-1}} \tag{7}$$

Thus, one can see that marginal costs of production are a decreasing function of the market access associated with a location. I assume that labour is the only factor of production therefore, m_c is a function of wages w_c only. Fitting a Cobb-Douglas functional form we have.

$$m_c = w_c^{\beta} \tag{8}$$

Assume further that fixed costs are proportional to marginal costs $F_c = fm_c$, subbing this into equation 8 and rearranging we find.

$$m_c = \left(\frac{MA_c}{\sigma f}\right)^{\frac{1}{\sigma}} \tag{9}$$

Subbing in our form for marginal cost, taking logs and rearranging we discover a key relationship.

$$\ln(w_c) = \kappa_0 + \kappa_1 \ln(MA_c) \tag{10}$$

Equation 10 is referred to as the wage equation where $\kappa_0 = -\frac{1}{\beta\sigma}\ln(\sigma\rho)$, and $\kappa_1 = \frac{1}{\beta\sigma}$. This equation states that all of the complex forward and backward linkages due to the underlying trade and transportation structure in this spatial model can be summarised by the market access of that location. Note that the coefficient on log market access is positive as $\beta > 0$ and $\sigma > 0$, thus the model predicts that market access will have a positive effect on wages.

Next, we turn to modelling local goods. We drop the city subscript as the supply and demand of local goods depend only on the city in which they are produced. I follow Venables (2017) in this exposition, whilst also building upon Venables (2017) by allow for imperfect labour mobility and edogenising tradeables. Local goods are produced under constant returns to scale, we choose units such that one unit of labour produces one unit of output of local goods, therefore $p_{\mathcal{L}} = w_{\mathcal{L}}$, that is the price of local goods equals the wage. Therefore, the value of the supply of such goods is $w_{\mathcal{L}}N_{\mathcal{L}}$ where $N_{\mathcal{L}}$ is the number of people employed in producing local goods within any given city. Similarly denote $N_{\mathcal{T}}$ as the number of people employed in tradeable industries. We assume that this dichotomisation partitions the employed within a city such that the total number of employed is equal to the sum of those employed within each industry type $N = N_{\mathcal{L}} + N_{\mathcal{T}}$. The total demand for local goods in any given city is equal to the fraction spent on such goods within the city $(1-\theta)$ times the total city income, which is equal to the city wage bill, plus any hinterland demand for local goods.

$$(1-\theta)\left(N_{\mathcal{T}}w_{\mathcal{T}} + N_{\mathcal{L}}w_{\mathcal{L}}\right) + p_{\mathcal{L}}h\left(p_{\mathcal{L}}\right) \tag{11}$$

Where the value of hinterland demand for a city's local goods is given by total hinterland expenditure on such goods $p_{\mathcal{L}}h(p_{\mathcal{L}})$, $h(\cdot)$ is the hinterland demand function assumed to be decreasing in price. Recall that $p_{\mathcal{L}} = w_{\mathcal{L}}$, using this fact, and assuming a linear form for hinterland demand $h(w_{\mathcal{L}}) = \beta - \alpha w_{\mathcal{L}}$, we can equate supply and demand and solve for local wages assuming such wages are positive.

$$w_{\mathcal{L}} = \frac{\beta - \theta N_{\mathcal{L}} + \sqrt{(\beta - \theta N_{\mathcal{L}})^2 + 4\alpha(1-\theta)N_{\mathcal{T}}w_{\mathcal{T}}}}{2\alpha} \tag{12}$$

Note first that local wages are determined by local demand which depends on the terms $N_{\mathcal{L}}$ and $N_{\mathcal{T}}w_{\mathcal{T}}$ and local supply which only depends on $N_{\mathcal{L}}$, and secondly that although local goods are separate from the global economy, changes in MA will impact $w_{\mathcal{T}}$ and therefore the wages of local good producers. The intuition here is simple, as market access increases wages for those working in the tradeable sector increase, as these people consume a constant proportion of their income as local goods, the demand for local goods also increases. Therefore, the price and wages offered in local industries rises also. Through this channel, and this channel alone, changes in access to non-local markets does affect wages of those employed in local industries. Note that the salience of this effect is determined by the number employed in tradeable industries, if $N_{\mathcal{T}} \approx 0$ market access will have no impact on local industry wages. This is logical as in this case the city has no connection to the outside world. The intuition of the model is developed further in section 4.2, but first we consider adding the temporal dimension.

4.1 Dynamic model and labour mobility assumptions

To summarise the above discussion, from strong theoretical foundations, we have formulated two equations determining the wages in tradeable and non-tradeable, or local, industries in a static model. We transform the above into a dynamic model in the simplest possible manner, by assuming that the static model is repeated in each time period. Therefore, I simply include a time subscript and formulate the following two equations which determine wages in each industry type in each city c.

$$\ln(w_{ct}^{\mathcal{T}}) = \kappa_0 + \kappa_1 \ln(MA_{ct}) + \kappa_2 \ln(G_{ct}) \tag{13}$$

$$w_{ct}^{\mathcal{L}} = \frac{\beta - \theta N_{ct}^{\mathcal{L}} + \sqrt{(\beta - \theta N_{ct}^{\mathcal{L}})^2 + 4\alpha(1-\theta)N_{ct}^{\mathcal{T}} w_{ct}^{\mathcal{T}}}}{2\alpha} \tag{14}$$

When considering wage determination over time, labour mobility assumptions are extremely important. There are four dimensions on which labour can move:

1. Between industries, within a city.

2. Between cities, within industries.

3. Between the hinterland and tradeable industries in a city.

4. Between the hinterland and non-tradeable industries in a city.

In section 3 we observed variation in wages both across industries within a city and across cities within an industry, implying imperfect labour immobility on these dimensions. There are large migration frictions between cities within Kenya, owing to factors specific to Kenya (and other developing countries) i.e. large transport costs, large information gaps (i.e. unlikely to know of a job opening at a city far away) and particularly strong familial/ tribal ties. Add to these are the usual costs of moving, and an assumption of perfect labour mobility between Kenyan cities, even over a long-time frame, is unlikely to be reasonable. Similarly, due to standard reasons of skill/ education acquisition, career advancement and habit forming it is also unlikely that perfect labour mobility between industry types is a reasonable assumption. Therefore I allow imperfect labour mobility between cities and industries (dimensions 1 and 2) even across long time periods[19]. Finally, we assume that whereas movement from the hinterland to producing local goods is possible, moving directly from rural areas to producing tradeables in the city is not.

[19]Note also that I do not consider a one off shock, but rather a continuous change in the transport network, therefore, even if adjustment occurs relatively quickly the system will be continuously shocked.

Note finally that if tradeable and local good producers have differing inter-city labour mobility rates we will not be able to separately identify this from differing impacts of MA on each industry keeping labour shares constant. However, I see this as a strength rather than a weakness of my approach, as we are interested in the overall impact of an expanding (or improving) transport network on wage outcomes, differing rates of labour mobility is an import aspect of this that should be taken into account. However, the extent to which differences in labour mobility rather than market access are driving results will have implications for policy responses. These issues are discussed further in section 6.3.

4.2 Predictions from the model

Equations (13) and (14) lead directly to predictions that can be taken to the data. Note that we can rewrite equation (14) in logarithms,

$$\ln\left(w_{ct}^{\mathcal{L}}\right) = -\ln(2\alpha) + \ln\left(\beta - \theta N_{ct}^{\mathcal{L}} + \sqrt{\left(\beta - \theta N_{ct}^{\mathcal{L}}\right)^2 + 4\alpha(1-\theta)N_{ct}^{\mathcal{T}} w_{ct}^{\mathcal{T}}}\right)$$

From this and the preceding equations, we can generate four predictions from the model that can be tested directly in the data:

1. Directly from equation (14) we see that log wages in tradeable industries increase with log market access as $\kappa_1 > 0$. This is because, for tradeable industries, increases in market access constitute increases in demand, which drive up prices. These price increases are compensated by increases in wages to maintain the zero-profit condition.

2. As $\ln\left(w_{ct}^{\mathcal{L}}\right)$ increases with $\ln\left(w_{ct}^{\mathcal{T}}\right)$ it follows that log wages in local industries also increase in log MA. Intuitively one expects these wages to increase at a slower rate than those in tradeable industries as market access affects wages in local industries only *through* it's impact on tradeable wages. Showing this is however non-trivial.

Theorem 1. *Log wages in industries producing local goods are less sensitive to changes in MA as wages in tradeable industries.*

Proof. We want to prove that $\frac{\partial \ln\left(w^{\mathcal{L}}\right)}{\partial \ln(MA)} < \frac{\partial \ln\left(w^{\mathcal{T}}\right)}{\partial \ln(MA)} = \kappa_1$. To show this we first note that as

$$\frac{\partial \ln\left(w^{\mathcal{L}}\right)}{\partial \ln(MA)} = \frac{\partial \ln\left(w^{\mathcal{L}}\right)}{\partial \ln\left(w^{\mathcal{T}}\right)} \cdot \frac{\partial \ln\left(w^{\mathcal{T}}\right)}{\partial \ln\left(MA\right)}$$

It suffices to show that $\frac{\partial \ln\left(w^{\mathcal{L}}\right)}{\partial \ln\left(w^{\mathcal{T}}\right)} < 1$. Next note by laws of differentiation and the differential of the natural logarithm that

$$\begin{aligned}\frac{\partial \ln\left(w^{\mathcal{L}}\right)}{\partial \ln\left(w^{\mathcal{T}}\right)} &= \frac{1}{\frac{\partial \ln(w^{\mathcal{T}})}{\partial \ln(w^{\mathcal{L}})}} \\ &= \frac{1}{\frac{\partial w^{\mathcal{T}}}{\partial \ln(w^{\mathcal{L}})}\frac{1}{w^{\mathcal{T}}}} \\ &= \frac{\partial \ln\left(w^{\mathcal{L}}\right)}{\partial w^{\mathcal{T}}} w^{\mathcal{T}}\end{aligned}$$

This significantly simplifies our exposition as we already have $\ln(w^{\mathcal{L}})$ as an explicit function of $w^{\mathcal{T}}$, taking this derivative we find

$$\frac{\partial \ln\left(w^{\mathcal{L}}\right)}{\partial w^{\mathcal{T}}} = \frac{\frac{1}{2}4\alpha(1-\theta)N^{\mathcal{T}}}{(\beta-\theta N^{\mathcal{L}})\left((\beta-\theta N^{\mathcal{L}})^2+4\alpha(1-\theta)N^{\mathcal{T}}w^{\mathcal{T}}\right)^{1/2}+(\beta-\theta N^{\mathcal{L}})^2+4\alpha(1-\theta)N^{\mathcal{T}}w\mathcal{T}}$$

Multiplying the above expression by $w^{\mathcal{T}}$ we can clearly define the inequality that we want to prove:

$$\frac{\partial \ln\left(w^{\mathcal{L}}\right)}{\partial w^{\mathcal{T}}} w^{\mathcal{T}} = \frac{w^{\mathcal{T}}2\alpha(1-\theta)N^{\mathcal{T}}}{(\beta-\theta N^{\mathcal{L}})\left((\beta-\theta N^{\mathcal{L}})^2+4\alpha(1-\theta)N^{\mathcal{T}}w^{\mathcal{T}}\right)^{1/2}+(\beta-\theta N^{\mathcal{L}})^2+4\alpha(1-\theta)N^{\mathcal{T}}w^{\mathcal{T}}} < 1$$

Define $A := 4w^{\mathcal{T}}\alpha(1-\theta)N^{\mathcal{T}}$ and $B := \beta - \theta N^{\mathcal{T}}$, we assume both quantities are positive. Then the inequality in above set of equations can be shown to hold if $0.5A < B(B^2 + A)^{0.5} + B^2 + A$ that is if $A > 0$ and $B > 0$ (note that only one of these inequalities need be strict).

Firstly: $A = 4\alpha(1-\theta)N^{\mathcal{T}}w^{\mathcal{T}} > 0$ which is true whenever a positive number of workers are employed in the tradeable industry, as we shall see in our data this holds for every city at every time period[20]. Secondly consider $B = \beta - \theta N^{\mathcal{L}} > 0$, multiplying both sides by $w^{\mathcal{L}} > 0$ we find that $\beta w^{\mathcal{L}} > \theta N^{\mathcal{L}} w^{\mathcal{L}}$. That is the share of non-tradeable producers income spent on tradeable goods $\theta N^{\mathcal{L}} w^{\mathcal{L}}$ must be less than the maximal value of demand for local goods by the hinterland. As local producers buy and sell local goods from each other the only possible source of funding for their purchasing of tradeable goods comes from the hinterland demand of their local goods. Therefore, certainly the value of local good producers demand for tradeable goods must be less than the maximal value of the hinterland demand for their goods. So as we have shown $A > 0$ and $B > 0$ the result follows. □

3. Log local industry wages are increasing in the log of the population employed in tradeable industries. The intuition here is that if more people are employed in tradeable industries but the number employed in local industries remains constant, the demand for local goods has increased without seeing any changes in value of supply, thus prices and so local wages must increase.

4. Log local industry wages are decreasing in log of the population employed in local industries. This can be shown mathematically employing the same techniques used above.

$$\frac{\partial \ln\left(w^{\mathcal{L}}\right)}{\partial \ln\left(N^{\mathcal{L}}\right)} = \frac{\partial \ln\left(w^{\mathcal{L}}\right)}{\partial N^{\mathcal{L}}} N^{\mathcal{L}} = \frac{-\theta - \theta\left(\beta - \theta N_{\mathcal{L}}\right)\left(\left(\beta - \theta N^{\mathcal{L}}\right)^2 + 4\alpha(1-\theta)N^{\mathcal{T}}w^{\mathcal{T}}\right)^{-1/2}}{\left((\beta - \theta N^{\mathcal{L}})^2 + 4\alpha(1-\theta)N^{\mathcal{T}}w^{\mathcal{T}}\right)^{1/2}}$$

Note that the denominator is positive and thus the sign of the above object

[20] We also assume that $\theta \neq 1$, that is workers do not spend all of their income on tradeable goods.

is equal to the sign of the numerator, as discussed above $\beta - \theta N^{\mathcal{L}} > 0$ and therefore this partial derivative is negative and log local industry wages are indeed decreasing in log of the population employed in local industries. Again, the intuition here is simple. Although increases in the population employed in local industries does increase the demand for local industry goods it cannot increase the demand by the magnitude of the increase in supply if non-tradeable industry workers spend any of their pay on tradeable goods.

Therefore, to conclude, we have used a NEG model to facilitate a sufficient statistic approach to determining the impact of changing the transport network on wages and expanded this to include two types of industry, tradeable and non-tradeable. From this model we reach four main conclusions which are directly testable in the data. Firstly, that log wages in tradeable industries are increasing in log market access. Secondly, that log wages in non-tradeable industries will also be increasing in log market access, but to a lesser extent. Thirdly, that log non-tradeable industry wages are increasing in log of the population employed in tradeable industries but, fourthly, it will be decreasing in log of the population employed in non-tradeable industries. This formal modelling framework and derived general predictions both inform and bring external validity to empirical work. A strength of this theoretical approach is in its simplicity, relying on only one key variable: market access. This simplicity allows us to construct a theory-consistent but assumption-light and plausibly causal empirical strategy as discussed in the next section.

5 Empirical methodology

Having constructed a novel dataset, gained an understanding of the underlying spatial and industry-level wage and employment share variation in Kenya over time and derived a formal model with testable implications relating transport infrastructure to wages, it only remains to develop an empirical strategy. In this section we are concerned with combining the formal theory above with the specific empirical setting of Kenya described in sections 2, 3 and appendix C, to elucidate the causal impact of changes in the transport infrastructure of wages in Kenya. One of the main challenges of this endeavour is the possible endogeneity of our key market access variable. This and related issues are discussed in detail in this section, however first we need to translate the theoretical equations into empirical specifications. Consider the theoretical equations (13) and (14) derived above relating log wages to log market access and log of the number of people working in each industry type, these equations can be combined into a simple empirical specification.

$$\begin{aligned}\ln\left(w_{cit}\right) &= \alpha + \beta_1 \ln\left(MA_{ct}\right) + \beta_2 \ln\left(N_{ct}^{\mathcal{T}}\right) + \beta_3 \ln\left(N_{ct}^{\mathcal{L}}\right) \\ &+ \gamma_1 \mathbb{1}_{[\mathcal{L}=1]} \ln\left(MA_{ct}\right) + \gamma_2 \mathbb{1}_{[\mathcal{L}=1]} \ln\left(N_{ct}^{\mathcal{T}}\right) + \gamma_3 \mathbb{1}_{[\mathcal{L}=1]} \ln\left(N_{ct}^{\mathcal{L}}\right) + \rho X_{cit} + \varepsilon_{cit} \quad (15)\end{aligned}$$

Where $\mathbb{1}_{[\mathcal{L}=1]}$ denotes an indicator variable equal to one when considering local industries and zero otherwise. $X_{cit} = (C_c, T_t, D_{cit})$ is a matrix of dummy variables taking into account city fixed effects C_c, year fixed effects T_t and removing other potentially spurious variation in D_{cit} such as, for example, the effect of the 2007/8 election violence on tourism in 2008[21]. The four predictions from the theory can be mapped directly onto predictions on the coefficients in this empirical model.

1. Log wages in tradeable industries increase with log market access if $\beta_1 > 0$
2. Log wages in local industries increase with log market access but to a lesser extent than log tradeable wages if $0 < \beta_1 + \gamma_1 < \beta_1$.
3. Log wages in local industries increase with log of the population employed in tradeable industries if $\beta_2 + \gamma_2 > 0$.
4. Log wages in local industries decrease with log of the population employed in local industries if $\beta_3 + \gamma_3 < 0$.

This equation is however not currently estimable, recall that $MA_c = \sum_i \tau_{ci}^{1-\sigma} G_i^{\sigma-1} E_i$ where τ_{ci} is the transportation cost from c to i, σ is the elasticity of substitution

[21]See Appendix C for details.

between varieties of tradeable goods, G_i is the CES price level over tradeable goods in city i and E_i is the total expenditure of city i. i indexes every city connected to c. This object, central in our theory and empirical work as a sufficient statistic, is not directly observable. We follow Donaldson and Hornbeck (2016) in creating a first order approximation which is observable:

$$MA_{ct} = \sum_i \tau_{ci,t}^{1-\sigma} N_{it} \tag{16}$$

Where N_{it} is the population of city i at time t. Therefore as N_{it} are observable from census records it only remains to calculate τ. This is a non-trivial task, as it involves calculating the transport costs from every city to every other city in every year, that is finding the least-cost path across a complex and evolving transport network. This is a large number of calculations, in order to perform them in a timely manner I appeal to Dijkstra's algorithm for finding the least-cost route along a network. Detailed discussion on the construction of market access is postponed until section 5.1.

Even with an observable first order approximation to the market access term, the above equation has endogeneity concerns. Firstly, the population of a city is endogenously determined with the wage rate, therefore we omit own-city population from our definition of market access in order to side step endogeneity bias on this parameter. In general changes in population maybe endogenous determined and do not reflect the exogenous variation we wish to capture: changing transportation costs. Therefore, in order to isolate exogenous variation in market access we keep the population of each city fixed at the 1989 census level, and omit own-city population from our calculation of market access. Finally, we require a value of σ, following Donaldson and Hornbeck (2016) we use $\sigma = 4.2$.

There remains concern over the endogeneity of the above described market access term. It is possible that, for example, a road built connecting two cities, was constructed precisely because either city was expanding, or for the reason of stimulating growth in otherwise lagging regions. For this reason, variation in τ may not be exogenous to changes in wages. In order to deal with this my identification strategy focuses on estimating *far away* variation in transport costs and therefore market access. I consider excluding variation from changes to the transport network 20km, 50km and 100km from each city as well as only considering cross-border variation, that is variation which occurs in Uganda and Tanzania. This strategy satisfies endogeneity concerns as it is decreasingly likely that city outcomes are endogenous to roads built far way. A further discussion of this identification strategy is postponed to section 6.3.

Using this identification strategy, we consider our measured market access variable to be strictly exogenous with respect to wage changes, and thus β and γ in the following equation to be identified and estimable.

$$\ln(w_{cit}) = \alpha + \beta \ln\left(\widetilde{MA}_{ct}\right) + \gamma \mathbb{1}_{[\mathcal{L}=1]} \ln\left(\widetilde{MA}_{ct}\right) + \rho X_{cit} + \varepsilon_{cit} \tag{17}$$

However, our theory and equation 15 stress the importance of $N^{\mathcal{L}}$ and $N^{\mathcal{T}}$ in determining local wages. We have already discussed the possible endogeneity of population, as employment is highly correlated with population it is likely that these variables are also endogenous. We maintain the assumption that $N^{\mathcal{L}}$ and $N\mathcal{T}$ are predetermined and uncorrelated with any city or year fixed effects. Thus, we can use lagged values of these variables as instruments in a GMM setting in order to estimate equation 15.

We define our observable, exogenous, market access term as

$$\widetilde{MA}_{ct} = \sum_{i \neq c} \left(\hat{\tau}^{r}_{ci,t}\right)^{-3.2} N_{i,1989} \tag{18}$$

Where $\hat{\tau}^{r}_{ci,t}$ denotes the cost of transportation between c and i at time t keeping the transport network within $r$km of c constant at 1990 levels. In estimating this equation, it is not expected that spatial correlation is likely to be a big issue as the unit of spatial measurement, the city, are geographically sparsely located. That is no two cities are adjacent, nor within 30km of each other, therefore usual problems resulting from spatial correlation of the errors which are common in a chess board approaches, are not of concern here. However, heteroskedasticity and serial correlation in the errors are both probable in this empirical approach. To address these concerns, I use bootstrapped standard errors. In all specifications I included a dummy for the retail industry in 2008, due to the potential effects the election violence could have had on this industry, see section C for historical details.

5.1 Construction of market access

It only remains to describe the procedure followed to calculate our measure of market access from the raw data. In order to achieve this, I construct a temporal network dataset in ArcGIS, consisting of nodes (settlements) and arcs (transport routes). To each node is assigned a constant population, and to each arc an impedance which varies over time. The impedance of each arc is determined by two variables, a time invariant distance d and a time variant type T. The type of an arc determines the speed of movement along said arc, arcs are categorised into one of eight types, outlined below.

1. Railway.
2. Ferry.
3. Road: motorway.
4. Road: primary improved.
5. Road: secondary improved.
6. Road: partially improved.
7. Track.
8. Does not exist.

As discussed in section 2, I digitise historical Michelin road maps to find variation in road type and existence over the study period, railway lines are constructed from OSM and ferries from historical records. For any settlement whose extent doesn't include a known road, I introduce a track from said settlement to the nearest known road. Roads that are not present in earlier maps but come into existence only after a certain date, are categorised as non-existent prior to that date.

The length of each arc is calculated in ArcGIS. However, to find the impedance of each arc, it is necessary to know the relative cost of moving along each type of arc. The Kenyan Statistical Abstracts report statistics on the rail network, including information on the average cost of transporting a kilometre-ton of goods. Therefore, we use kilometre-ton as the unit of cost measurement and calibrate this to the average real cost on railways which are given as K\$0.04[22]. Using this information, we calculate the kilometre-ton cost of every arc type by assuming the cost is proportional to travel speed along each type of arc. Transport speeds for each arc type are taken from a variety of sources. Speeds along road types are found using Google Map's API, taking the average speed across several routes consisting solely of that type of road. Railway speeds are taken from World Food Programme Logistics[23] who maintain information on the speed at which humanitarian assistance can be delivered in times of crises. Ferry travel speed information is taken from historical documents from the East African Railways Corporation[24], who maintain records on travel times between various ports on Lake Victoria. Finally, dirt tracks are assumed to have an average travel speed of 15km/h and non-existent roads have a travel speed of zero. This information in summarised in table 1 below.

Table 1: Impedances by travel type

	Speed (km/h)	Cost (K$)	Source of speed estimates
Motorway	72.0	0.0138	Google API
Primary improved	58.2	0.0170	Google API
Secondary improved	44.6	0.0222	Google API
Partially improved	34.4	0.0288	Google API
Dirt track	15.0	0.1984	
Railway	24.8	0.0400	World Food Programme Logistics
Ferry	26.5	0.0374	East Africa Railways Corporation

The second element in the network dataset used to construct our measure of market access is the nodes. In order to capture an accurate measure of the market which a given city c in my sample of 22 major Kenyan cities, $\mathcal{C}$, has one needs to consider not only the market access c derives from being connected to the other 22

[22]Calculated as the average of the deflated reported cost over the sample period.
[23]https://www.wfp.org/logistics
[24]http://www.mccrow.org.uk/EastAfrica/EastAfricanRailways/indexEAR.htm

cities in my sample, but also that which it derives from every other settlement in the region. Therefore, in order to create an accurate measure of market access, and to more fully exploit the variation in network quality, we require data on the location and population of all other urban centres in Kenya (as well as Uganda and Tanzania when looking at cross border variation). I define a settlement s as a point in space satisfying two conditions:

1. There exists some observable build cover within 5km of s in 2014.

2. There exists a known settlement at s in the OSM data.

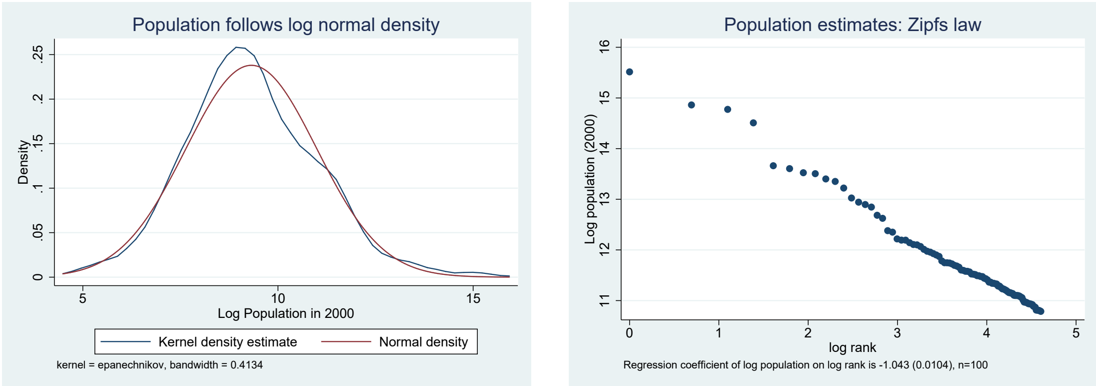
The set of all such settlements is denoted $\mathcal{S}$. We find 497 settlements satisfying this definition across Kenya, Uganda and Tanzania. Build cover data is taken from the European Commission Global Human Settlement Layer. This data source uses high-resolution satellite imagery, census data and volunteered geographic information with spatial data mining and machine learning techniques to construct a map of the world at 1km resolution, where each pixel is defined as built-up or not (Pesaresi et al. (2015)). I overlay this map with a map downloaded from OSM which shows every known settlement location. Note that the set of cities in my sample (22) is a subset of the set of settlements (497), these settlements become the nodes of the network. Finally, in order to calculate market access, we require an estimate of population for each settlement above defined. I find this by overlaying a raster of the gridded population of the world (GPW) from CIESIN Columbia University (2016). The GPW dataset uses spatially disaggregated censuses and other surveys to estimate the population of the world in 30arc-second grids. Over the build cover extent associated with every settlement I take the sum of the associated population in each 30 arc-second grid as given in the gridded population of the world dataset in 2000. This is taken as the population of each settlement[25]

Having constructed this dataset of settlements with associated populations, I perform some sense checks. A common stylised fact in the city population distri-

[25]Data is not available pre-sample as it is for the 22 cities in our sample. In order to prevent under weighting cities (as opposed to settlements) by taking earlier population estimates we deflate calculated settlements populations by the population urban growth rate between 1989 and 2000.

bution literature is that all settlements in a country or region should follow a log normal distribution and, relatedly, the largest cities follow Zipf's law (Ioannides and Overman, 2003). We test both of these predictions below in figure 8, and find that this new dataset conforms tightly to both[26], giving us some confidence in its method of construction. To the aforementioned features I add ports and stations, the loca-

Figure 8: SETTLEMENT DATA SENSE CHECKS



Notes: The left hand figure shows the log distribution of city populations in our sample in comparison with a log-normal distribution. Note the empirical density function is calculated with the automatic Epanechnikov kernel with bandwidth 0.4134. The right hand figure shows the log-rank, log-population plot for the largest 100 cities in our sample. The coefficient $\hat{\beta}$ from running $log(pop_i) = \alpha + \beta log(rank_i) + \varepsilon_i$ is estimated to be -1.043 with standard error 0.0104.

tions of which have been lifted from OSM. It is assumed that one can travel along the intersection of any road by any other, but that for railways one must keep on the line and cannot switch at intersections. Changing between modes of transport is only permitted at stations or ports and has an associated fixed cost of K$4[27]. The different elements of this network data set are brought together by enforcing upon it a topology. We ensure that each station is coincident with both a road and a railway, likewise each port is coincident with a road and a ferry route. Similarly, each settlement must lie a road, if a there does not exist a road within a settlements build cover a track is created from the settlement to the nearest road.

Our network is, however, not complete. To generate an accurate measure of market access for all cities in our sample we cannot consider Kenya in isolation.

[26]The coefficient on the log-log regression $\ln(pop_i) = \alpha + \beta \ln(rank_i) + \varepsilon_i$ is reported as -1.043 with bootstrapped standard error 0.0104. This regression is run on a truncated sample of the 100 largest cities.

[27]This fixed cost is taken as proportionally the same as the fixed cost imposed in Donaldson and Hornbeck (2016).

The world market to which Kenya has access too must also be considered, otherwise cities close to the border will have a relatively underestimated market access, biasing results. Kenya is connected to the rest of the world in two main ways. Firstly, through its borders with Uganda and Tanzania (see section C for a discussion on the East African Community), and secondly, it's connection with the rest of the world through the port city of Mombassa. We first deal with cross-border trade by constructing a network, exactly as set out above, for Uganda and Tanzania, by far Kenya's largest African trading partners. Although major changes in Kenya-Uganda and Kenya-Tanzania border permeability may have occurred since 2010 with the introduction of the customs union and common market across the EAC, there is little reason for variation in border permeability within the study period. Therefore, we assume a constant cost of crossing the border equal to that of transferring transport type K$4. This may appear low but these countries are fairly well integrated and share a common language.

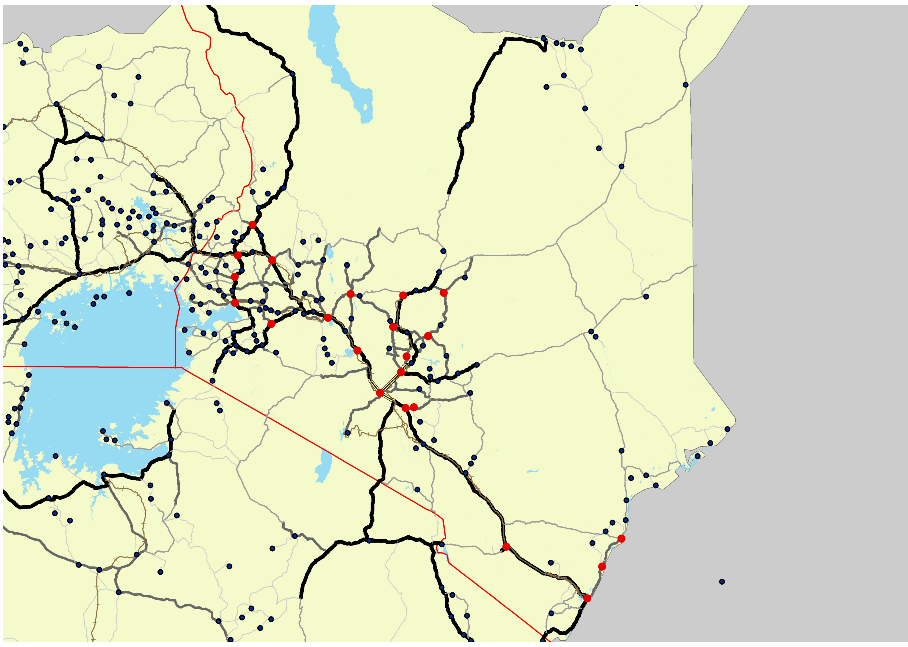
Secondly, we turn to considering Kenya's trade with the rest of the world. Almost all of Kenya's non-African trade comes through the port of Mombassa, highlighting its importance for the regional and national economy of Kenya. Omitting the rest of the world from our market access calculations would therefore relatively underestimate Mombassa and surrounding cities market access. Modelling the entire transport network and its interconnectedness for the whole world between 1990 and 2010 is beyond the scope of this study, instead we take a much simpler first order approximation. The rest of the world is visualised as a city lying just off the coast by Mombassa which is only and costlessly connected to Mombassa port[28]. As with other settlements we keep the population of this hypothetical *rest of the world* settlement fixed at 2000 levels. It therefore only remains to estimate its population size. From the Kenyan Statistical Abstracts, I extract the value of exports to the rest of the world in 2000 and times this by Kenya's GNP per capita in 2000 to find the effective population of the rest of the world in terms of Kenya's economy. Using

[28]Note that Mombassa port is not costlessly connected to Mombassa as a K$4 transfer still applies.

this method, we estimate the rest of the world is equivalent to a city of population $\approx 2,000,000$ lying just off the coast of Mombassa.

Putting all the above elements together we can create a complete network in the EAC area over the period 1990 to 2010, estimating the constant-rate population of each settlement and including gains in market access due to access to international markets. Figure 9 shows the completed map zoomed in on Kenya in 2010. Red lines denote country borders, brown lines railways, grey to black lines roads of increasing size, black dots are settlements and red dots are the cities in our sample. Having

Figure 9: COMPLETED TRANSPORT NETWORK 2010



Notes: This figure shows the completed transport network including settlements in 2010 focusing on Kenya. Red lines denote national borders, brown lines railways and black lines roads. Thick, dark black lines denote primary improved roads, dark grey secondary improved roads, light grey partially improved roads and thin light grey dirt tracks. Black dots denote settlements and large red dots cities which constitute our main sample.

constructed the full network dataset I calculate the market access for each city $c \in \mathcal{C}$ using equation 18. Populations are taken directly as described above, however it remains to construct $\widehat{\tau}$ for each city in each time period. To do this I use Dijkstra's algorithm to calculate the least-cost path from every city to every settlement, taking into account the relative costs of travelling on each arc type as described above.

6 Results

Having constructed a robust, theory consistent empirical strategy and created a novel dataset to test this strategy we can finally turn to the results. We first report results from running a simplified specification: equation 19 where $X_{ci,t}$ are the set of controls which vary across specifications, results are reported in table 2. In all the results presented below we omit the mining industry as in 73% of cases there is no one employed in mining in a given city-year, results are qualitatively similar when mining is included.

$$\ln(w_{ci,t}) = \beta \ln\left(\widetilde{MA}_{c,t}\right) + X_{ci,t}\Gamma + \varepsilon_{ci,t} \tag{19}$$

Table 2: AVERAGE EFFECT OF MARKET ACCESS ON WAGES

	[1]	[2]	[3]	[4]	[5]	[6]	[7]
Log MA	0.1622***	0.0850*	3.124***	0.1824***	0.0668	0.1052***	0.1257
	(0.0470)	(0.0468)	(0.2043)	(0.0375)	(0.5049)	(0.0367)	(0.5648)
Year FE		X			X	X	X
City FE			X		X		X
Industry FE				X		X	X
Observations	3206	3206	3206	3206	3206	3206	3206

Standard errors are bootstrapped.

Notes: This table reports results from estimating equation (19) using the full sample of cities and industries over the whole time period. Bootstrapped standard errors are reported in parenthesis; * p< 0.1, ** p< 0.05, *** p< 0.01. The dependent variable in all regressions is log real wages. The independent variable in column [2] includes year fixed effects, column [3] city fixed effects, column [4] industry fixed effects, column [5] year and city fixed effects, column [6] year and industry fixed effects and column [7] year, city and industry fixed effects. In each regression a dummy variable to account for possible disruption to the retail industry in 2008 is included.

In column [1][29] $X_{ci,t} = (1)$ a vector of ones, that is a constant, in column two $X_{ci,t} = T_t$ a matrix of year fixed effects dummies, in column [3] $X_{ci,t} = C_c$, city fixed effects dummies and in column [4] $X_{ci,t} = I_i$ industry fixed effects dummies. Preceding columns are combinations of the previous, for example in column [5],

[29] Note that the maximum number of city-industry-years is 3344, however due to some industries, mainly agriculture, having zero employment in some city-years only 3206 observations are available. Therefore, we do not observe 4% of the possible wages.

$X_{ci,t} = [T_t, C_c]$. Columns without city fixed effects report a constant and significant estimate of around .15 for the average effect of market access on wages, note that here we are estimating a log-log specification and so this coefficient is interpreted as an elasticity. However, once we include city fixed effects in columns [5] and [7] this coefficient becomes insignificantly different from zero, but remains positive. This is primarily due to a lack of temporal variation across the sample period I focus on. This simple specification sets a bench mark, reporting similar point estimates to those found in the literature.

Table 3 turns to estimating heterogeneous effects by tradeable and non-tradeable industries, testing the fundamental logic forwarded by this thesis that non-tradeable industries should not react as sensitively to changes in a cities access to external markets. Non-tradeable industries are defined as agriculture, construction and retail although results here are robust to any reasonable categorisation, see Appendix A for results using alternative categorisations. Results from table 3, use an augmented version of equation 19, allowing the coefficient on market access to vary by industry categorisation. Table 3 is based on equation 22 where $\mathbb{1}_{[\mathcal{L}=1]}$ is an indicator variable taking the value one when the industry is categorised as non-tradeable (or local). $X_{ci,t}$ takes the same form as in equation 19.

$$\ln(w_{ci,t}) = \beta_1 \ln\left(\widetilde{MA}_{c,t}\right) + \beta_2 \mathbb{1}_{[\mathcal{L}=1]} \ln\left(\widetilde{MA}_{c,t}\right) + X_{ci,t}\Gamma + \varepsilon_{ci,t} \tag{20}$$

Table 3 confirms the main hypothesis of this thesis, that the effects of market access on wages differ by industry type, and specifically that the effects of market access are significantly less for non-tradeable industries. Estimates in the first row of table 3 mirror, but is in all cases are greater than, the estimates in table 2. Giving strong evidence in favour of the hypothesis of differing means, and in particular that tradeable industries react, on average, more sensitively to changes in market access. The second row in table two confirms this, reporting consistent estimates of the interaction effect of market access on non-tradeable industries of around -0.14. Overall the effect of market access on non-tradeable industries is normally positive

Table 3: HETEROGENEOUS EFFECTS OF MARKET ACCESS ON WAGES

	[1]	[2]	[3]	[4]	[5]	[6]	[7]
Log MA	0.2335***	0.1589***	3.187***	0.2095***	0.1318	0.1348***	0.1607
	(0.0742)	(0.0470)	(0.2050)	(0.0416)	(0.4797)	(0.0417)	(0.6732)
Log MA *	-0.1472**	-0.1540**	-0.1787**	-0.0766	-0.1882**	-0.0839	-0.1150*
non-tradeable	(0.0509)	(0.0734)	(0.0757)	(0.0718)	(0.0940)	(0.0417)	(0.0672)
Year FE		X			X	X	X
City FE			X		X		X
Industry FE				X		X	X
Observations	3206	3206	3206	3206	3206	3206	3206

Standard errors are bootstrapped.

Notes: This table reports results from estimating equation 22 using the full sample of cities and industries over the whole time period. Bootstrapped standard errors are reported in parenthesis; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable in all regressions is log real wages an indicator for non-tradeable industries and their interaction. The independent variable in column [2] includes year fixed effects, column [3] city fixed effects, column [4] industry fixed effects, column [5] year and city fixed effects, column [6] year and industry fixed effects and column [7] year, city and industry fixed effects. In each regression a dummy variable to account for possible disruption to the retail industry in 2008 is included.

but small, and certainly considerably (and statistically significantly) less than the effect on tradeable industries. This result is apparent across all specifications considered and gives strong evidence in support of the core logic of this thesis, that non-tradeable industries shouldn't react as sensitively to changes in market access as tradeable industries. This is also a result borne out of the theory and confirmed here.

There is concern about endogeneity bias in the estimates reported in tables 2 and 3. This is due to the possibility that roads are built/upgraded endogenously with respect to city outcomes. Road construction may occur to/from cities that would otherwise have experienced wage growth. We exploit a feature of our definition of market access, that much of the variation in any given cities market access is due to changes far from that city, to isolate plausibly exogenous variation in the market access variable. To achieve this, we draw circles of diameter d around each city and freeze the transport network for that city within the circle, therefore any variation in market access will occur solely due to *far away* changes in the transport network which are unlikely to be endogenously determined with local city level outcomes. See section 6.3 for a discussion of this approach. Table 4 shows the results from implementing this empirical strategy. Column [1] repeats the results from column [5] in table 3 for comparison. Columns [2], [3] and [4] report results controlling for variation within 20km, 50km and 100km radii. Finally, column [5] uses a similar identification strategy, considering only cross-border variation in market access. That is, for estimation in column [5], the transport network in Kenya is assumed constant at 1990 levels, and the only identifying variation comes from changes in the transport network in Uganda and Tanzania and the knock-on effect that has on domestic market access in Kenya.

Table 4: IDENTIFICATION STRATEGIES

	[1]	[2]	[3]	[4]	[5]
Log MA	0.1318	0.340	0.3631	0.0762	0.4749
	(0.5450)	(0.4971)	(0.5388)	(0.5216)	(0.3013)
Log MA * non-tradeable	-0.1882***	-0.1884***	-0.1886**	-0.1879**	-0.3376***
	(0.0735)	(0.0660)	(0.0807)	(0.0839)	(0.0752)
Year and City FE	X	X	X	X	X
Observations	3206	3206	3206	3206	3206

Standard errors are bootstrapped.

Notes: This table reports results from estimating equation 22 using the full sample of cities and industries over the whole time period. Bootstrapped standard errors are reported in parenthesis; * p< 0.1, ** p< 0.05, *** p< 0.01. The dependent variable in all regressions is log real wages an indicator for non-tradeable industries and their interaction. Each regression includes year and city fixed effects. Column [1] reports estimates from column [5] of table 4 for comparison. Column [2] reports estimates controlling for variation in the transport network within 20km of each city, column [3] within 50km and column [4] within 100km. Column [5] reports estimates where only variation due to changes in the Ugandan and Tanzanian have been utilised. In each regression a dummy variable to account for possible disruption to the retail industry in 2008 is included.

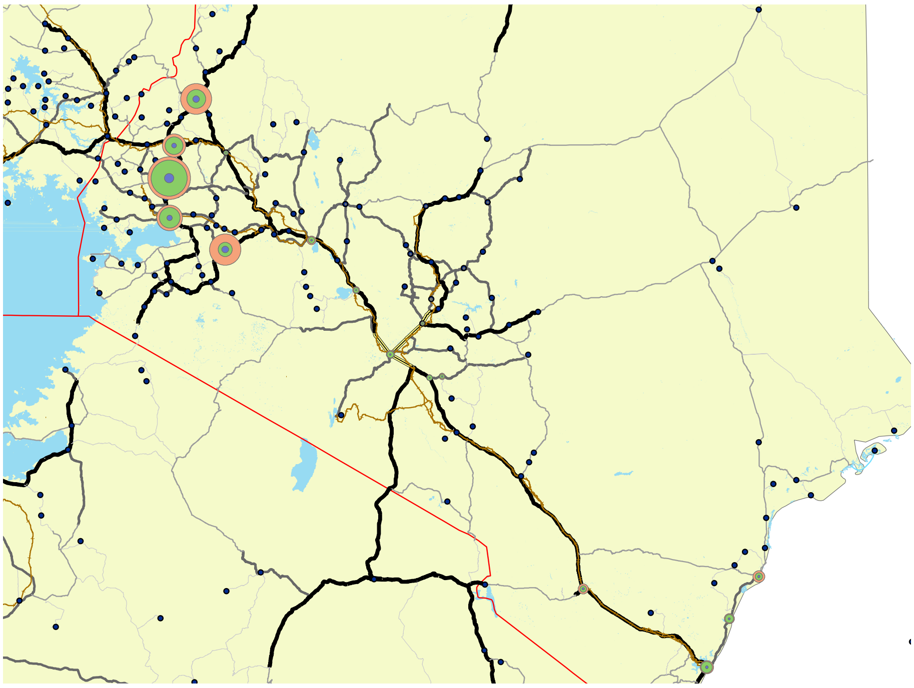
Results from table 4 suggest my previous estimates are somewhat robust to identification concerns. The qualitative interpretation of the regressions remains the same, market access positively affects tradeable industries (albeit imprecisely estimated), and affects non-tradeables in a significantly less positive manner. Note that in this table only the preferred specification of year and city fixed effects is considered. Results for other specifications follow a similar pattern as observed in

3, and are presented in appendix A.

We take the average of the coefficients over these strategies as our core results, that is the effect of log market access on log wages in tradeable industries is ≈ 0.3 and the effect on non-tradeables is ≈ 0.06. Using these estimates, we can consider the spatial heterogeneity of effects over the study period. For each city I calculate the causal impact changes in the transport network between 1992 and 2010 had on wages in tradeable and non-tradeable industries. Using this method, we find that on average tradeable real wages increased by 0.7% with standard deviation of 0.66 as a result of increases in market access and non-tradeable wages increased by 0.14% with standard deviation of 0.14. By weighting changes in each industry type by the average proportion of the population employed in that industry over the sample period, we can deduce the overall change in wages due to changes in market access (and therefore changes in the trasnport network), we find that on average market access caused wages to increase by 0.5% over the study period, with standard deviation 0.48. In order to better understand how effects vary spatially, the relative percentage increases in wages are mapped onto figure 10 as three concentric circles showing the percentage change in tradeable (orange) and non-tradeable (green) wages as well as the overall percentage change (blue) due to changes in market access.

From figure 10 the heterogeneous nature of the effect of market access on wages over the study period is evident, due to most of the changes to the transport network occurring in the west of the country this is where the gains from such changes translate into wage increases. These results paint a positive picture for spatial inequality, relatively increasing wages in the Rift Valley area as opposed to in Nairobi or Mombassa. Such distributional effects act as to reduce regional wage inequality in Kenya, however it should be noted this is a direct result of the geographical location of changes in the transport network. Changes in the transport network can increase or decrease spatial inequality depending on where new roads are built, these results therefore suggest that over the last twenty years Kenya's road

Figure 10: EFFECT OF MARKET ACCESS ON WAGES



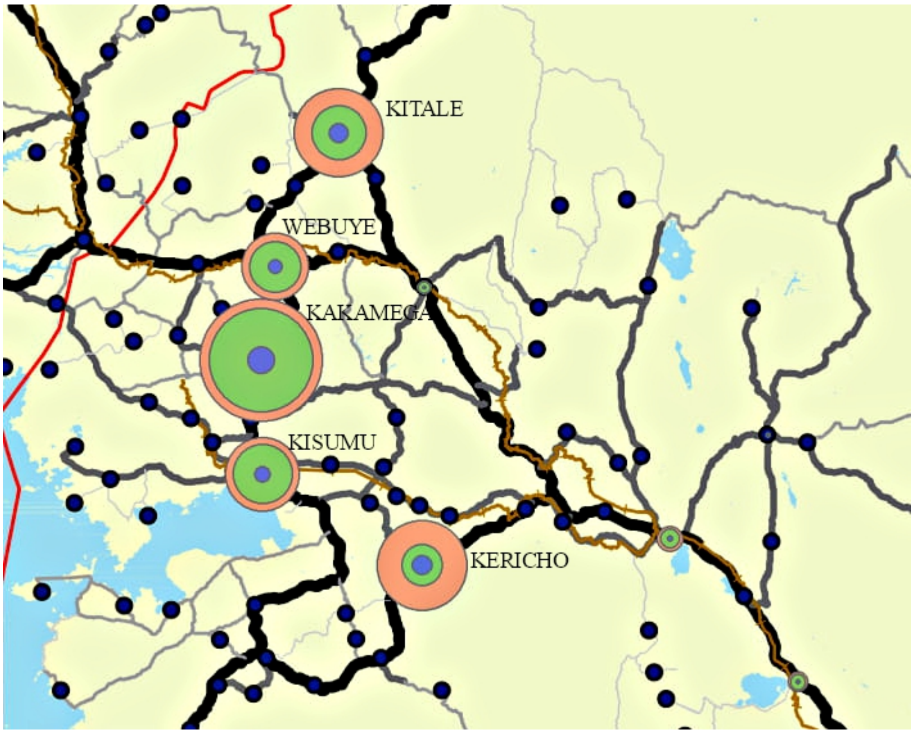
Notes: This figure shows how changes in the road network between 1992 and 2010 have affected wages in Kenya's major towns. The diameter of orange circle denotes relate change in tradeable wages for city on which the circle is centred, similarly the diameter of blue circles indicate the relative change in non-tradeable wages. The diameter of the green circle represents the relative overall change in wages and is a weighted average of the two other diameters. Red lines denote national borders, brown lines railways and black lines roads. Thick, dark black lines denote primary improved roads, dark grey secondary improved roads, light grey partially improved roads and thin light grey dirt tracks. Black dots denote settlements and large red dots cities which constitute our main sample.

building projects were in locations with relatively lower income levels[30], and thus acted as to decrease spatial inequality. This highlights the potentially important impact role road building location decisions can have on spatial inequality, and that this maybe a channel through which governments can tackle primate city dominance - by improving the transport network far away from the main city. However even among the cities which saw the largest increases in market access, there is significant heterogeneity in wage changes. This can be seen in figure 11 which zooms in on the most affected cities from figure 10.

Figure 11 clearly shows how even if two cities have the same increase in their market access due to improvements in the transport network, how these in-

[30]That is relatively lower compared with Nairobi and Mombassa not to the North of the country which is much more sparsely populated and poorer.

Figure 11: DISTRIBUTIONAL EFFECTS OF INDUSTRY EMPLOYMENT SHARES



Notes: This figure shows how changes in the road network between 1992 and 2010 have affected wages in Kenya's major towns, focusing on changes in the Rift Valley area. The diameter of orange circle denotes relate change in tradeable wages for city on which the circle is centred, similarly the diameter of blue circles indicate the relative change in non-tradeable wages. The diameter of the green circle represents the relative overall change in wages and is a weighted average of the two other diameters. Red lines denote national borders, brown lines railways and black lines roads. Thick, dark black lines denote primary improved roads, dark grey secondary improved roads, light grey partially improved roads and thin light grey dirt tracks. Black dots denote settlements and large red dots cities which constitute our main sample.

creases translate into average wage gains is determined by the employment share of those working in tradeable industries. For example, Kericho saw a larger increase in its market access that Kisumu (as evidenced from the larger orange and blue circles in figure 11) but saw a *smaller* overall change in wages (smaller green circle). This is because Kisumu has a far larger proportion of its work force employed in tradeable industries as compared to Kericho. This interaction of spatial and industry heterogeneity implies that only cities with already relatively developed, tradeable, industries see the benefits, in terms of wage increases, of improvements to the transport network. Through this mechanism infrastructure projects can act as to drive wage inequality in developing countries to a greater extent than that which would be implied by their geographical location.

Above we confirm the core theoretical, and logical, result highlighted by this thesis: that the effect of market access on wages in heterogeneous in industry type, with tradeable industries having a larger a positive effect compared with non-tradeable industries. Now we turn to testing the four predictions from the full model, recall equation 15 re-printed below with our observed measure of market access included:

$$\begin{aligned}\ln\left(w_{cit}\right) = \alpha &+ \beta_1 \ln\left(\widetilde{MA}_{ct}\right) + \beta_2 \ln\left(N_{ct}^{\mathcal{T}}\right) + \beta_3 \ln\left(N_{ct}^{\mathcal{L}}\right) \\ &+ \gamma_1 \mathbb{1}_{[\mathcal{L}=1]} \ln\left(\widetilde{MA}_{ct}\right) + \gamma_2 \mathbb{1}_{[\mathcal{L}=1]} \ln\left(N_{ct}^{\mathcal{T}}\right) + \gamma_3 \mathbb{1}_{[\mathcal{L}=1]} \ln\left(N_{ct}^{\mathcal{L}}\right) \\ &+ \rho X_{cit} + \varepsilon_{cit}\end{aligned} \tag{21}$$

Where $\mathbb{1}_{[\mathcal{L}=1]}$ denotes an indicator variable equal to one when considering local industries and zero otherwise. $X_{cit} = (C_c, T_t, D_{cit})$ is a matrix of dummy variables taking into account city fixed effects C_c, year fixed effects T_t and removing other potentially spurious variation in D_{cit} such as, for example, the effect of the 2007/8 election violence on tourism in 2008. Recall from our discussion in section 5 that endogeneity concerns were overcome by using only exogenous (far-away) variation in our modified market access variable and taking a GMM approach, instrumenting $N^{\mathcal{L}}$ and $N^{\mathcal{T}}$ with lagged values[31]. This model is the empirical version of the theory set out in section 4, which made four clear predictions on the sign and relative magnitudes of the coefficients:

1. Log wages in tradeable industries increase with log market access if $\beta_1 > 0$

2. Log wages in local industries increase with log market access but to a lesser extent than log tradeable wages if $0 < \beta_1 + \gamma_1 < \beta_1$.

3. Log wages in local industries increase with log of the population employed in tradeable industries if $\beta_2 + \gamma_2 > 0$.

[31]Note that wage employment in each industry type in any given year is highly correlated with that the year before, and thus we consider these instruments to be *strong*. Note also that the Hansen test of overidentifying restrictions holds.

4. Log wages in local industries decrease with log of the population employed in local industries if $\beta_3 + \gamma_3 < 0$.

Table 5 summarises the results from estimating 21. Panel A of table 5 reports the results from estimating equation 21 and panel B summarises the implications for the predictions generated by the theory 1-4 described above. Although most results are imprecisely calculated, the point estimates conform to the theoretical predictions in three out of four cases. In the fourth case, prediction 2, the data correctly estimates the effect of market access on non-tradeable wages to be less than that on tradeable wages, but incorrectly estimates this effect to be negative, which is not supported by the model. Overall, we conclude that the data does support the theoretical results derived from the model, however due to a lack of identifying variation over the period we consider, most results cannot be precisely estimated.

Table 5: TESTING THE FULL MODEL

PANEL A	MA	MA* non-tradeables	N(tradeables)	N(tradeables)* non-tradeables	N(non-tradeables)	N(non_tradeables)* non-tradeables
Dependent variable:	0.1507	-0.1689***	0.0347	0.0010	-0.0274	0.0142
log wages	(0.5488)	(0.0636)	(0.0539)	(0.0285)	(0.0390)	(0.0313)

PANEL B

PREDICTION	ESTIMATE	INTERPRETATION
1	$\hat{\beta}_1 = 0.1507 > 0$	Effect of market access on tradeable industry wages is positive
2	$\hat{\beta}_1 + \hat{\gamma}_1 = -0.0182 < \hat{\beta}_1$	Effect of market access of non-tradeable industry wages is less than that in tradeable industries, and may even be negative.
3	$\hat{\beta}_2 + \hat{\gamma}_2 = 0.0357 > 0$	Wages in non-tradeable industries increase with the population employed in tradeable industries.
4	$\hat{\beta}_3 + \hat{\gamma}_3 = -0.0132 < 0$	Wages in non-tradeable industries decrease with population employed in non-tradeable industries.

Notes: This table reports results from estimating equation 21 using the full sample of cities and industries over the whole time period and taking a GMM approach, using lagged values of $N^{\mathcal{L}}$ and $N^{\mathcal{T}}$ as instruments for their current values. Bootstrapped standard errors are reported in parenthesis; * p< 0.1, ** p< 0.05, *** p< 0.01. Panel A reports results from running one regression with dependant variable log wages and independent variables: log market access, the number of people in wage employment in non-tradeable industries, the number of people employed in tradeable industries, an indicator variable taking the value one if an industry is categorised as non-tradeable and the interaction of the three other independent variables with this indicator. I also include city and year fixed effects and a dummy variable to account for possible disruption to the retail industry in 2008 is included. Panel B summarises the implications of the results reported in panel A on the four predictions made by the theoretical model.

In the above regressions we have considered market access as a whole, however in our construction of market access three channel were highlighted. Firstly, local market access which considers the Kenyan market cities have access to, secondly EAC market access, that is the market Kenyan cities can access from Uganda and Tanzania, and finally rest of the world market access. In this section we attempt to disentangle these three channels by constructing three market access variables. Firstly we consider $\widetilde{MA}^{D}_{ci,t}$, that is domestic market access calculated as described in section 5.1, but assuming Kenya is a closed economy. Secondly, I consider access to the EAC market by calculating the total cost in each period for each city to travel to the nearest border with Uganda or Tanzania, capturing each city's (cost-weighted) distance to Kenya's neighbours and how this varies over time as the Kenyan transport network varies. Thirdly, I measure access to the rest of the world in a similar manner, the vast majority of non-EAC trade with Kenya comes through the port of Mombassa, therefore I measure each cities access to the rest of the world by calculating it's least-cost route to Mombassa in each period. Table 6 summarises the results from decomposing market access in this manner. Note that for each regression in this table we implement our identification strategy of removing variation in the transport network within 100km of a given city.

Table 6 gives some potentially puzzling results. The first row follows a familiar pattern across this thesis of a positive effect of local market access on wages in tradeable industries and a significantly smaller effect on non-tradeable industries. However, column two appears to imply that the higher RoW market access, that is the closer a city is to Mombassa, the lower wages are. Results for EAC access follow a similar pattern. Note also that in columns [4], [5] and [6] when local market access and our measures of international market access are included in the same regression, the effect of local market access increases, and becomes significant in the last column. Together these results appear to suggest that although increasing access to local markets has a positive effect on wages, increasing access to international markets depresses wages. One possible explanation for this is that, by

Table 6: DECOMPOSING MARKET ACCESS

	[1]	[2]	[3]	[4]	[5]	[6]
MA (local)	0.0762			0.7559	0.5597	1.050**
	(0.4385)			(0.4938)	(0.5706)	(0.5394)
MA (local) *	-0.1879**			-0.2192***	-0.3107***	-0.4319***
non-tradeable	(0.0778)			(0.0838)	(0.1033)	(0.1166)
MA (RoW)		-13.32**		-15.61***		-13.73*
		(6.366)		(5.735)		(7.414)
MA (RoW) *		0.0350		0.0168		0.0204
non-tradeable		(0.0248)		(0.0240)		(0.0212)
MA (EAC)			-0.7890**		-0.9083**	-0.6273
			(0.3721)		(0.3572)	(0.4192)
MA (EAC) *			0.0346		0.0989**	0.1372***
non-tradeable			(0.0294)		(0.0403)	(0.0432)
Year and City FE	X	X	X	X	X	X
Observations	3206	3206	3206	3206	3206	3206

Standard errors are bootstrapped

Notes: This table reports results from estimating equation 22 using the full sample of cities and industries over the whole time period with three different measures of market access. Bootstrapped standard errors are reported in parenthesis; * p< 0.1, ** p< 0.05, *** p< 0.01. The dependent variable in all regressions is log real wages an indicator for non-tradeable industries and their interaction. Each regression includes year and city fixed effects. Column [1] reports estimates from column [5] of table 4 for comparison. Column [2] reports estimates using the rest-of-the-world measure of market access only and column [3] using the East African Community measure of market access only. Column [4] reports estimates from regressing real wages on local and RoW measures of market access together, column [5] local and EAC measures of market access together and column [6] all three measures. In each regression a dummy variable to account for possible disruption to the retail industry in 2008 is included.

being closer to international markets, a city has access to cheaper imports increasing competition on locally produced goods. This forces prices and so wages to become depressed in order for domestic firms to stay competitive. This explanation matches the pattern of larger negative effects in relation to RoW connectivity as compared to relatively smaller effects relating to being closer to the Uganda and Tanzania. Also, corroborative is the consistently positive and often significant, interaction effects of market access and non-tradeable industry dummy. One would expect any impacts of market access (positive or negative) to be diminished in non-tradeable industries. This explanation fits with the literature of increasing globalisation, and diminishing trade costs, lowering local wages in developing countries[32], but this is an area where more research is needed in the Sub-Saharan Africa context before strong conclusions can be reached.

Finally, we turn to looking for direct evidence of the effect of market access on development indicators. As described in section 2 we develop a consistent data set covering six main development indicators: percentage with no flushing toilet, percentage with straw roof percentage with straw or mud walls, percentage with piped water percentage literate and percentage with no schooling. We regress each of these against our measure of market access at the city level (note now we have no variation by industry) over the six years for which geolocated development data is available. Not all indicators are available for every city in every year, note that the maximum number of observations is 126 and therefore these results, especially as we include city and year fixed effects, are rather underpowered which goes some way to explaining the lack of significance reported. Table 7 summarises the results. Note

Table 7: CORRELATES WITH DEVELOPMENT INDICATORS

	% with no toilet [1]	% roofs straw [2]	% with straw or mud walls [3]	% with piped water [4]	% literate [5]	% with no schooling [6]
Log MA	0.7755 (0.5389)	0.4628 (0.4622)	0.8362 (0.7464)	-2.050 (1.602)	-0.1459 (0.5051)	0.1861 (0.4428)
Year and City FE	X	X	X	X	X	X
Observations	95	112	91	95	78	61
Standard errors are bootstrapped						

Notes: This table reports results from regressing development indicators against log market access using the full sample of cities and industries over the whole time period. Bootstrapped standard errors are reported in parenthesis; * p< 0.1, ** p< 0.05, *** p< 0.01. The dependant variable in column [1] is the percentage of residents in the city with no toilet, in column [2] it is the percentage of residents in the city with straw roofs, in column [3] it is the percentage of residents with straw or mud walls, in column [4] it is the percentage of residents with piped water, in column [5] it is the percentage of residents who are literate and in column [6] it is the percentage of residents with no formal schooling. Both year and city fixed effects are included in each regression.

firstly that we have not address possible issues of endogeneity in these regressions, and the models used here are entirely a-theoretical. Therefore, we consider these results to be reduced form in nature, relying on a fixed effects, difference in differ-

[32]See for example Acemoglu (2003), Akerman et al. (2013) and Helpman et al. (2017)

ences, methodology to remove sources of endogeneity. Thus, these regressions are not claimed to be causal. Note secondly that no coefficient is significant at the 5% level (although the F-test including all in one specification shows joint significance at the 5% level)f, this is due to our small sample size and subsequent low power. Despite this lack of significance, all coefficients reported tell the same story, that increases in the transport network, through market access, have a negative impact on urban development. One mechanism through which this result maybe explained is that highlighted in this paper of changes in the transport network driving spatial and industry inequality. Adding to this the unintended interaction effect, that less developed cities have a smaller share of workers employed in the tradeable sector and therefore are relatively worse off from changes in the network, and we can develop a story of lock-in and relative decline in the poorest areas which may contribute to the negative associations found. This is, however, an area where further research is needed before strong conclusions can be drawn.

In sum, market access affects wages positively on average in Kenya, but this hides considerable heterogeneity both geographically and by industry type. Non-tradeable industries are significantly less affected by changes in the transport network, this finding implies that transport projects can increase wage inequality in developing countries by only benefiting the already relatively developed cities. The model from section 4 is tested and appears to hold in the data. I also find suggestive evidence of heterogeneity by market access type; that access to more competitive markets can actually depress wages, and that market access is negatively associated with development indicators. These last two findings require further research in order to fully understand the mechanisms driving them.

6.1 Counterfactual analysis: Mombassa-Malaba railway

Using our estimated impact of market access on tradeable and non-tradable wages from table 4 we can estimate the impact of hypothetical changes in the transport network of Kenya (as well as Uganda and Tanzania) on wages in each city in each industry. In order to perform this analysis I take the average of the coefficients reported in table 4: the effect of log market access on log tradeable wages is thus ≈ 0.3 and the effect on non-tradables is ≈ 0.06. The credibility of any counter factual estimates lies in the assumption of linearity in the effect of market access on wages, and a similarity between the magnitude of the identifying variation and that which we consider in our counter factual analysis. First, we consider the linearity assumption, here we rely on the general theoretical framework we developed in section 4, which predicts a linear relationship. Secondly note that between 1992 and 2010 the percentage change in market access present in the data is on average 10% with standard deviation of this change of 5%. The bottom 5% experience a 5% increase in market access whereas the top 5% experience a 16% increase. I compare the magnitude changes with those found in the data described above.

I consider the possible impact of the proposed upgrade to the Mombassa-Malaba line. In 2014 Kenya received a large loan from China to begin construction on a new standard gage railway from Mombassa (on the coast) to Malaba (on the Ugandan border), as of 2017 the Mombassa-Nairobi stretch has already been completed. This is a serious infrastructure project expected to cost over K$1 trillion (£7 million), much of the financing being footed by loans from China. Kenya's national debt is only K$4 trillion, and therefore a possible debt to China of another K$1 trillion represents a large proportion of national debt and is testimony the importance and scale of this project[33]. In total this railway will be almost 1000km long, constituting the biggest infrastructure project in Kenya since independence, and the entire network is expected to be operational in December 2018. Large projects like this are often touted as bringing transformative economic outcomes to the region, however this research highlights the heterogeneous effects of such a project and it's potential to increase wage inequality. This thesis gives us a unique opportunity to assess the impact of this large project on average wages and their spatial variation. Travel on the current railway proceeds at an average

[33]See http://www.constructionkenya.com/2971/mombasa-malaba-sgr-project

speed of 24.8km/h on average, the new standard gauge railway will travel at an average speed of 120km/h, a 383% increase[34] in speed as compared to the current railway. This thesis does not and cannot provide a full appraisal of this project. Any large endeavour such as this railway will bring immediate employment for many Kenyans as well as training to many in the construction industry, in addition to this it will boost demand for many domestic products related to railway construction and provide continued employment for thousands of railway workers. Additionally, and specific to this case, much of Ugandan (and Sudanese) trade with the rest of the world passes through the port of Mombassa, therefore the construction of this railway is likely to increase these countries access to trade with the rest of the world. We do not attempt to take these effects, nor the knock-on effect of the increased importance of the port of Mombassa, into account. Finally, in all the analysis that follows we assume no change in the movement of people.

Maintaining our assumption of the linearity of transport costs with respect to transport speed, an increase in rail speed from 24.8km/h to 120km/h corresponds to a decrease in transport costs from K$0.04 km/ton to K$0.008 km/ton. I implement these changes into my network dataset by decreasing the impedance due to rail in a fictional 2011. The only additional modelling change employed is to, in a very reduced form sense, model capacity limits at Mombassa port. This is achieved simply by limiting the increase in market access to Mombassa due to this rail upgrade to a maximum of 10 times it's 1990 level. Implementing these changes causes market access to increase in all cities, however the effects are highly heterogeneous with most of the increases due to faster travel between Nairobi and Mombassa on the new railway. Increases in market access are measured relative to the actual 2010 network. The bottom 5% of cities market access increases by only 2%, the top 5% by 36% and Mombassa market access increases by 873%. Note that except for Mombassa the increases in market access are of the same magnitude as that found in the data, adding credibility to this counterfactual exercise. Note however

[34]See https://www.railway-technology.com/projects/mombasa-nairobi-standard-gauge-railway-project for details of this project.

that in my empirical work all variation in the data is in roads, the railway network remained constant, therefore no identifying variation came from changes in railways and yet this counterfactual exercise solely considers such variation. The extent to which we consider changes in transport costs through rail variation to affect wages differently from those due to road variation is the extent to which the validity of this exercise is questioned. We find it unlikely that wages react to these different types of variation differently, especially when we analyse changes in the transport network using a general, and theory backed, sufficient statistic of market access.

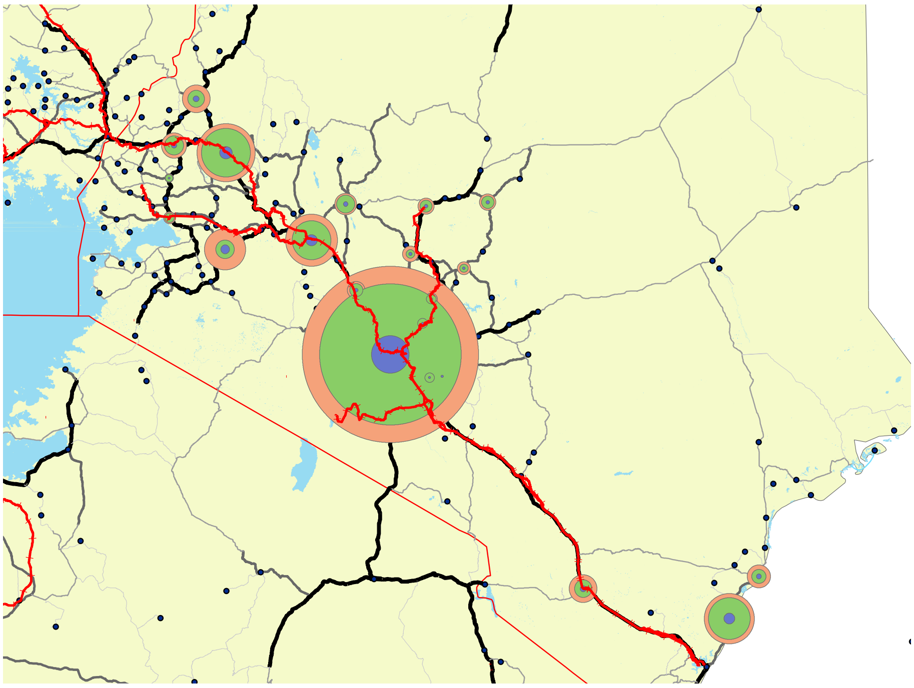
We find that replacing the existing rail road with modern standard gauge has the effect as to increase wages on average by 12.7% for tradeable industries and 2.7% for non-tradeable industries as compared to observed wages in 2010. These averages however mask large geographical variation, Mombassa is by far the most affected with its wages increasing by 242% and 52% respectively for tradeable and nontradeable industries. On the other hand wages in Thika only increase by 0.6% and 0.1% for tradeable and non-tradeable industries. Such variation becomes amplified once we consider changes to the city-wide wage bill, that is once differences in tradeable and non-tradeable employment shares across cities are considered. Nairobi and Mombassa, which benefit most from the new railway, also have among the highest proportion of individuals employed in tradeable industries exacerbating their relative gain. Overall changes in city wages by industry type are calculated by assuming constant worker composition[35], this geographic variation is summarised in figure 12. In figure 12 each city is represented by three concentric circles with radii representing percentage wage changes relative to observed 2010 levels in tradeable (pink) and non-tradeable (green) industries and overall (blue)[36]. Note that Mombassa has been omitted from this figure as it would dominate the image, wage

[35]In practice this is an unlikely assumption as we expect workers in tradeable industries to move to areas with higher market access, however we have not developed the required modelling tools to deal with this and so take a first approximation by keeping composition constant. Note also that although the proportion of workers employed in tradeable industries is likely to increase, this will put downward pressure on wages in this sector, implying that although our employment estimates are understated the wage effect may be overstated and so the product is more likely to be correct.

[36]Note that the same scale for circle diameters is used in this diagram as that which was employed in figure 10, facilitating comparison.

changes in Mombassa are a factor of magnitude larger than those in Nairobi.

Figure 12: EFFECT OF PROPOSED RAILWAY WORK



Notes: This figure shows how the construction of the proposed Mombassa-Malaba standard gauge railway may affect wages in Kenya's major towns. The diameter of the orange circles denote relative change in tradeable wages for city on which the circle is centred, similarly the diameter of blue circles indicate the relative change in non-tradeable wages. The diameter of the green circle represents the relative overall change in wages and is a weighted average of the two other diameters. Note that the circles surrounding Mombassa have been omitted as they dominate the picture. Circles are to the same scale as in figure 10 Red lines denote national borders, brown lines railways and black lines roads. Thick, dark black lines denote primary improved roads, dark grey secondary improved roads, light grey partially improved roads and thin light grey dirt tracks. Black dots denote settlements and large red dots cities which constitute our main sample.

Unsurprisingly the positive impacts of this railway are concentrated along the line itself. An unanticipated effect of this is that Nairobi and Mombassa, already the most developed cities in Kenya, see most of the wage increases. This has a negative impact on spatial inequality in Kenya, increasing existing differences and is likely only to contribute to Kenya's continued problem with excess primacy and deepen regional divides. This effect is exacerbated by Nairobi and Mombassa's relatively large tradeable industry shares, implying that these large increases in market access translate into large average wage increases. On the other hand a city such as Kericho which is also close to the railway, although experiencing large increases in market access sees only a small change in average wages. These relative distributional effects are important and should be considered when embarking on such a project, as they benefit cities with high tradeable employment shares more than those with low tradeable employment shares. Perhaps facilitating a ‘lock-in’ effect where ‘bad’ cities, with low tradeable employment shares, are left-behind.

6.2 Robustness checks

Here I summarise the robustness checks performed and their results. Full details of each test and results can be found in appendix A. The robustness checks I perform fall into three main categories, I analyse deviations from my main empirical specification, different definitions of tradeable industry and consider evidence in favour of the exogeneity of my measure of market access. In each test variations on my base specification are considered, using the identification strategy of maintaining a constant network within 100km of each city.

I estimate my base specification including province[37] fixed effects. We find this has no effect on the estimated coefficients and that none of the province fixed effects are significant. We also consider controlling for province fixed effects instead of individual city level fixed effects, we find that the estimates are qualitatively the same. Next, I consider including the interaction of a flexible fourth order polynomial in longitude and latitude coordinates with year fixed effects to flexibly control for temporal location-based effects. We find the coefficients stay stable and maintain a consistent interpretation. Lastly, we consider the possibility of outliers driving the results by running our main specification removing the top and bottom 10% of market access values, again we find that the qualitative interpretation of the coefficients remains the same.

Secondly, we consider alternative definitions of non-tradeable industries. Recall, wage data is disaggregated by nine industries, agriculture, mining, manufacturing, electricity, construction, retail, transport, finance and social. In the main

[37] Provinces are the largest administrative area in Kenya. There are eight such provinces which are further subdivided into 69 districts. However, in 2010 a new constitution in Kenya was ratified scrapping the province system in favour of a counties system.

specification agriculture, construction and retail are considered non-tradeable industries. This categorisation may seem somewhat arbitrary, to reassure the reader that this grouping has not been strategically chosen here we consider all reasonable possible categorisations of non-tradeable industries. Five of the nine industries maybe construed as non-tradeable: agriculture, electricity, construction, retail and transport. Retaining a base of agriculture and construction, I consider all possible combinations of the remaining three contenders for categorisation as non-tradeable. Again, across all specifications I find similar estimates and retain the same qualitative interpretation.

Finally, we consider a test for evidence in favour of the exogeneity of my measure of market access and identification strategy. To do this I regress market access on lagged values of wages (conditioning on year and city fixed effects). A null result supports my hypothesis of exogeneity of market access (whilst not proving it). I perform this test on my base specification and using my identification strategy, null results are reported in each case, with a higher p-value when using my identification strategy of only using far-away variation in market access. This gives evidence in support of the exogeneity of my measure of market access and my identification strategy.

6.3 Discussion

Despite vast data efforts in digitising historical Statistical Abstracts and road maps, the analysis presented in this thesis remains curtailed through data limitations. Firstly, over the period discussed there is less identifying variation in the Kenyan transport network available in the data than that which occurred, this prevents a detailed and concrete analysis of heterogeneous (by geography) effects and is the cause of many insignificant results reported. There are two ways which this can be improved on in the future. Firstly, by cross referencing information on changes in kilometres by type of road in the Kenyan Statistical Abstracts with differences from one Michelin map to the next it should be possible to determine the exact year each road change occurred. This is a significant improvement on the data in this thesis and other papers in the literature which use similar methods to construct network datasets over time. Secondly, data is available for both the transport network and wage outcomes in earlier years: all the way back to 1973. Between 1973 and 1992 there was significant variation in the transport network, additionally these extra years would increase our sample size, increasing the power of empirical tests employed. Time constraints have prevented the use of this extra data in this thesis, but this is an area for natural extension.

The wage data from the Kenyan Statistical Abstracts does however suffer from further limiting factors. Firstly, and most apparently, the number of cities available is small, 22, this impacts the power of any empirical strategies employed but may have more serious ramifications. The cities considered are all large cities in the south of Kenya. This may affect the generalisability of any results to smaller cities in Kenya (and thus our results applicability to the whole of Kenya), or indeed to other countries in SSA. Mitigating these concerns is the fully general model developed and tested in this thesis, and the consistent definition of sample cities employed. However, concerns remain, and the results should be read with them in mind. A related issue is that the wage data used in this thesis only considers the remuneration of those in wage employment and ignores the informal sector. This is particularly important given the relative size of the informal sector in Kenya and other SSA countries. Data on formal wages was extremely difficult to find, thus finding information on informal wages is practically impossible.

Although much has been done in my empirical methodology to account for the potential endogeneity of market access and in general variation in the transport network, endogeneity concerns will always remain. In order to identify causal variation in the transport network this thesis only considers changes in the network that occur far away from each city. The validity of this methodological approach rests on the assumption that the building/ improving of roads over 100km away from city A, is not endogenous to city A's outcomes. That is, although we may consider a road built connecting London and Birmingham to be endogenous to these cities outcomes (attempting to boost lagging midlands relative to booming London), we would not consider a road connecting York to Hull to be endogenous to Birmingham's outcomes. There are a few cases where one may expect this argument to break down.

Firstly, the case of regional initiatives. In England for example successive governments have championed a "northern powerhouse" suite of policies, which involve transport infrastructure development, for example HS2. In this scenario southern cities success (including high wages) relative to northern cities leads to building roads/railways in northern areas. In order to control for this potential source of endogeneity, I include province fixed effects as part of my robustness tests and find no significant change in coefficient estimates. Secondly long road/railway projects may cause concern. Using HS2 as an example again, although this method controls for the first 100km of route from London the remaining distance to Birmingham is not controlled for although it is clearly built for the same reasons as the first 100km. If one suspected the first 100km of being endogenous to London's outcomes certainly the remaining line is susceptible to the same argument. This case is, however, not an issue over the time period considered in Kenya as no single project is over 100km long. However, the concern remains when one considered projects quite-far but not 100km away from a city. In this scenario if a 50km road project moving away from a city begins 75k from said city the first 25km would be controlled for but the second 25km would not. This is a contrived case, and road projects that begin 75km from a city are themselves unlikely to be endogenous to that city, however this case maybe worrying and is a potential flaw in the current identification strategy. To address these concerns, I also consider only variation which occurs outside Kenya, and find that estimates have qualitatively the same interpretation.

Labour mobility and migration are not directly modelled, theoretically or empirically, in this thesis. Owing to data limitations the movement of individuals is impossible to elucidate, precluding any such analysis. However, how labour moves over each dimension[38] has important ramifications for the interpretation of the empirical results. Geographic variation in wages is only possible in a model with imperfect labour mobility, and therefore the extent to which results are found can in part be attributed to the extent to which labour is geographically (and inter-industry) un-mobile. This becomes particularly important if one considers the possibility of variable labour mobility across industries, perhaps those employed in tradeable industries are less mobile than those in non-tradeable industries and this fact alone drives the results found. However, as discussed in the main text, this interpretation does not affect the results per se, but influences the possible channel through which they are being driven, and therefore the appropriate policy response. Incorporating a dynamic, heterogeneous labour mobility into the NEG model developed in this thesis is a necessary step towards a rigorous analysis of these questions. Time constraints have prevented this thesis from exploring these areas, but this is a potentially fruitful area for future work.

However, some policy recommendations can be reached, these centre around the insight that transport infrastructure projects have a positive effect on wages, but also that there may be unintended distributional consequences of this. Some industries, specifically those in the tradeable sector, receive the lion's share of these benefits increasing inter-industry inequality. This can be mitigated by policies which enhance inter-industry labour mobility, i.e. policies which improve schooling/training or are pro labour force flexibility. Transport infrastructure projects may also have unforeseen ramifications for spatial inequality. This is because such projects are firstly most beneficial to nearby cities[39], but secondly improve wages in cities with a higher proportion of workers in tradeable industries[40]. That is, transport projects mainly benefit cities which are already advanced, in that they employ more people in tradeable industries, this acts as to increases wage inequality between these cities

[38]Between cities, between industries and between the hinterland and cities.

[39]Although this is likely to be an expected effect.

[40]This occurs through two mechanisms. Firstly, there are more people in tradeable industries which see higher wage rises due to changes in market access and secondly non-tradeable wages rise further also due to the larger increases in local demand caused by higher tradeable wages.

and those which maybe stuck only producing non-tradeable goods. This negative distributive effect is seen in the counterfactual considered. The wages increases implied by the proposed Nairobi-Malaba railway in the majority occur in Nairobi and Mombassa, further boosting their already high wages. To mitigate these effects governments in developing countries need to do more to encourage growth in the tradeable sector in hinterland cities. Policies that improve electricity provision and support firms locating outside the capital may go some way in addressing these issues. In sum, transport infrastructure projects are no silver bullet as they are often purported to be, they are likely to increase industry and spatial wage inequality, governments in developing countries should be aware of this and take policy actions to mitigate these effects.

7 Conclusion

This thesis brings evidence to answer the question: what are the effects of changes in a SSA country's transport network on city level wages by industry type, and thus, how such changes drive spatial wage inequality. The construction of three novel datasets, through the digitisation of historical statistical records and maps, allows us to answer these questions in detail as well as increase our understanding of the spatial distribution of employment in Kenya more generally. I develop a consistent theory of industry-type heterogeneity in the effect of changes in the transport network on wages, and from this derive an empirical specification based on a sufficient statistic, market access, approach. I identify causal effects in this framework by utilising variation in the transport network that occurs *far away* from each city, allaying endogeneity concerns. I find that market access is positively associated with wages. However, this relationship masks considerable heterogeneity, with tradeable industries seeing most of the gains as compared to non-tradeable industries. Thus, increasing transport infrastructure appears to increase inter-industry wage inequality, which in turn increases spatial wage inequalities due to differences in employment shares of cities.

I also report results looking at different types of market access and find evidence that although increases in the domestic market access have a positive effect on wages, increases in a cities access to international markets has a negative impact. This is explained by the increase in competition generated by easier access for importers to Kenya driving down local wages, this is consistent with the literature of globalisation depressing wages in developing countries. I also report the reduced form effect of market access on development indicators, here we find insignificant but negative effects of market access on all indicators. This suggests that the unintended impact of increased inequality maybe counter-productive for development, however this is also an area were more research is needed in the future.

Finally, I utilise the model to analyse the potential impact of the proposed Mombassa to Malaba standard gauge railway. I find that this expensive project will have the effect as to increase wages in Kenya for tradeable industries by 12.7% and for non-tradeables by 2.7%. However, the majority of the effects are found along the line itself, and particularly in Nairobi and Mombassa, areas which are already relatively well off. Thus, we find that this project will act as to significantly increase regional inequality in Kenya, benefiting the already better off areas and industries the most.

The Kenyan Statistical Abstracts are an incredibly rich data source, one which I have barely tapped into here. Future research using this data could uncover many new and interesting facts about a SSA economy which previously were unknown and unknowable due to data limitations. At a time when Africa is poised to massively urbanise, understanding the impact of improved inter-city connectivity and developing policy recommendations is key to ensuring cities work in Africa, and can produce the productivity miracles necessary for countries and the region as a whole, to develop. This thesis focuses how transport investments can drive spatial wage inequality to a far greater degree than that implied by their geographical location through their interaction with industry employment shares. Concluding that transport infrastructure projects are not the silver bullet policy makers often paint them to be.

References

Acemoglu, D. (2003). Patterns of skill premia. *The Review of Economic Studies*, 70(2):199–230.

Akerman, A., Helpman, E., Itskhoki, O., Muendler, M.-A., and Redding, S. (2013). Sources of wage inequality. *American Economic Review*, 103(3):214–19.

Behrens, K. and Robert-Nicoud, F. (2015). Agglomeration theory with heterogeneous agents. In *Handbook of regional and urban economics*, volume 5, pages 171–245. Elsevier.

Bofinger, H. (2011). *Africa's transport infrastructure: Mainstreaming maintenance and management*. World Bank Publications.

Burgess, R., Jedwab, R., Miguel, E., Morjaria, A., and Padró i Miquel, G. (2015). The value of democracy: evidence from road building in kenya. *American Economic Review*, 105(6):1817–51.

CIESIN Columbia University, f. (2016). Gridded population of the world, version 4 (gpwv4): Population count. *Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC)*.

Donaldson, D. (2010). Railroads of the raj: Estimating the impact of transportation infrastructure. Technical report, National Bureau of Economic Research.

Donaldson, D. and Hornbeck, R. (2016). Railroads and american economic growth: A "market access" approach. *The Quarterly Journal of Economics*, 131(2):799–858.

Edgerton, R. B. (1991). *Mau Mau: An African Crucible*. Ballantine Books.

Fallah, B. N., Partridge, M. D., and Olfert, M. R. (2010). New economic geography and us metropolitan wage inequality. *Journal of Economic Geography*, 11(5):865–895.

Fujita, M., Krugman, P. R., Venables, A. J., and Fujita, M. (1999). *The spatial economy: cities, regions and international trade*, volume 213. Wiley Online Library.

Glaeser, E. L. (2012). *Triumph of the city: How our greatest invention makes us richer, smarter, greener, healthier, and happier*. Penguin.

Gollin, D., Jedwab, R., and Vollrath, D. (2016). Urbanization with and without industrialization. *Journal of Economic Growth*, 21(1):35–70.

Hanson, G. H. (2005). Market potential, increasing returns and geographic concentration. *Journal of international economics*, 67(1):1–24.

Head, K. and Mayer, T. (2006). Regional wage and employment responses to market potential in the eu. *Regional Science and Urban Economics*, 36(5):573–594.

Helpman, E., Itskhoki, O., Muendler, M.-A., and Redding, S. J. (2017). Trade and inequality: From theory to estimation. *The Review of Economic Studies*, 84(1):357–405.

Henderson, J. V., Squires, T., Storeygard, A., and Weil, D. (2017). The global distribution of economic activity: nature, history, and the role of trade. *The Quarterly Journal of Economics*, 133(1):357–406.

Henderson, J. V., Storeygard, A., and Roberts, M. (2013). Is urbanization in sub-saharan africa different? *Policy Research Working Paper Series 6481, World Bank*.

Henderson, J. V., Storeygard, A., and Weil, D. N. (2012). Measuring economic growth from outer space. *American economic review*, 102(2):994–1028.

Hering, L. and Poncet, S. (2010). Market access and individual wages: Evidence from china. *The Review of Economics and Statistics*, 92(1):145–159.

Ioannides, Y. M. and Overman, H. G. (2003). Zipf's law for cities: an empirical examination. *Regional science and urban economics*, 33(2):127–137.

Jedwab, R. (2013). Urbanization without structural transformation: Evidence from consumption cities in africa. *George Washington University, Washington, DC. Processed.*

Jedwab, R., Kerby, E., and Moradi, A. (2015). History, path dependence and development: Evidence from colonial railroads, settlers and cities in kenya. *The Economic Journal.*

Jedwab, R. and Storeygard, A. (2016). The heterogeneous effects of transportation investments: Evidence from sub-saharan africa 1960-2010. *Working Paper.*

Krugman, P. (1991). Increasing returns and economic geography. *Journal of political economy*, 99(3):483–499.

Krugman, P. and Venables, A. J. (1995). Globalization and the inequality of nations. *The quarterly journal of economics*, 110(4):857–880.

Lucas, R. E. (1988). On the mechanics of economic development. *Journal of monetary economics*, 22(1):3–42.

Nations, U. (2014). World urbanization prospects: The 2014 revision, highlights. department of economic and social affairs. *Population Division, United Nations.*

Paillacar, R. (2006). Market potential and worker heterogeneity as determinants of brazilian wages. *University of Paris*, 1.

Pesaresi, M., Ehrilch, D., Florczyk, A. J., Freire, S., Julea, A., Kemper, T., Soille, P., and Syrris, V. (2015). Ghs built-up grid, derived from landsat, multitemporal (1975, 1990, 2000, 2014). *European Commission, Joint Research Centre (JRC).*

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

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.

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

Appendices

A Details of robustness checks

Here I report the results and exact specifications of the robustness tests performed. As discussed in the main body I consider three main classes of robustness checks, variations on the empirical specification, changes to the definition of non-tradeable industries and evidence in favour of the exogeneity of my measure of market access.

First we consider variations on my main empirical specification. Recall that my base specification is given by:

$$\ln(w_{ci,t}) = \beta_1 \ln\left(\widetilde{MA}_{c,t}\right) + \beta_2 \mathbb{1}_{[\mathcal{L}=1]} \ln\left(\widetilde{MA}_{c,t}\right) + T_t + C_c + \varepsilon_{ci,t} \tag{22}$$

Where I identify the parameters β_1 and β_2 using only *far-away* variation, that is the transport network within 100km of each city is kept constant. I consider four variations on this specification, including region fixed effects, replacing city fixed effects with region fixed effects, including a fourth order polynomial in longitude and latitude coordinates interacted with year fixed effects and finally removing outliers. Results are summarised in table 8 below. As we can see, although the coefficient on

Table 8: ROBUSTNESS: CHANGES TO THE SPECIFICATION

		Log MA	Log MA * non-tradeable
BASE		0.0762 (0.5216)	-0.1879 (0.0839)
Including province fixed effects.	[1]	0.0761 (0.4683)	-0.1877 (0.0755)
Replacing city fixed effects with province fixed effects.	[2]	0.2493 (0.0590)	-0.1662 (0.0825)
Including a fourth order polynomial in longitude and latitude interacted with year fixed effects.	[3]	0.0203 (1.085)	-0.1854 (0.0757)
Removing outliers	[4]	0.2099 (0.6228)	-0.0402 (0.1163)

Standard errors are bootstrapped and in brackets.

log market access varies somewhat, the interaction term is stable and the qualitative interpretation remains the same in each alternative specification.

Secondly we consider changes in the definition of non-tradeable industries.

We consider five candidate non-tradeable industries: agriculture, electricity, construction, retail and transport. We consider agriculture and construction to be core non-tradeables, and analyse how results vary when considering all remaining possible combinations. Results are summarised in table 9. From table 9 it is clear

Table 9: ROBUSTNESS: CHANGES TO THE DEFINITION OF NON-TRADEABLES

		Log MA	Log MA * non-tradeable
BASE (agriculture, construction and retail)		0.0762 (0.5216)	-0.1879 (0.0839)
Agriculture and construction	[1]	0.0302 (0.4717)	-0.1181 (0.0914)
Agriculture, construction and electricity	[2]	-0.0011 (0.5285)	0.0010 (0.0898)
Agriculture, construction and transport	[3]	0.0891 (0.5090)	-0.2231 (0.0897)
Agriculture, construction, electricity and retail	[4]	0.0193 (0.4693)	-0.0368 (0.0693)
Agriculture, construction, electricity and transport	[5]	0.0406 (0.0512)	-0.0636 (0.0868)
Agriculture, construction, retail and transport	[6]	0.1335 (0.5120)	-0.2472 (0.0810)
Agriculture, construction, electricity, retail and transport	[7]	0.0667 (0.4020)	-0.0933 (0.0877)

Standard errors are bootstrapped and in brackets.

that my results are not simply an artefact of my specific definition of non-tradeable industries. In all but one of the cases considered above the coefficients remain stable and the qualitative interpretation remains constant.

Finally I consider examining the preposition that market access is not endogenously determined with wage outcomes core to the validity of my empirical approach. Although this assumption is not directly testable we can consider examining the temporal direction of the effect, that is if indeed market access and wages are not endogeneously determined we wouldn't expect lagged values of wage outcomes to be correlated with market access. A null result doesn't prove exogeneity but gives evidence in support of it. Thus I regress log market access against log real wages conditional on year and city fixed effects and report robust standard errors. I consider this regression in two cases, firstly using all variation in market access and secondly employing my identification strategy by only using far-away variation in market access (that is changes in the transport network more than 100km from each city). Results are reported in table 10. As we can see lagged log real wages are

Table 10: ROBUSTNESS: ENDOGENEITY OF MARKET ACCESS

		Lagged log real wages
Base		-0.0007 (0.0009)
Base employing my identification strategy: only using far-away variation	[1]	-0.0005 (0.0009)
Standard errors are bootstrapped and in brackets.		

highly non-significant in both specification, and the associated p-values of 0.436 and 0.534 indicate that when using my identification strategy the coefficient becomes even less significant.

B List of major cities

List of "major" cities constituting the main sample used in analysis. A major city is defined as a city with at least 1,000 *engaged* persons in 1973 in Kenya. A person is considered engaged if they are in wage employment or self employed. There are twenty two such cities.

1. Athi River
2. Eldoret
3. Embu
4. Kakamega
5. Kericho
6. Kilifi
7. Kisumu
8. Kitale
9. Machakos
10. Malindi
11. Meru
12. Mombassa
13. Murang'a
14. Nairobi
15. Naivasha
16. Nakuru
17. Nanyuki
18. Nyahururu
19. Nyeri
20. Thika
21. Voi
22. Webuye

C Empirical setting

This thesis's empirical setting is primarily concerned with the geographic region of Kenya and its neighbours, between the years of 1992 and 2010. To fully contextualise this study and be confident in the validity of my empirical estimation, any important idiosyncrasies of this region over this time period must be properly dealt with. Thus, an understanding of the historical context of Kenya as a whole and in particular over the study period is necessary for any serious work in this area. However, it is beyond both the scope of this thesis and the capacity of the author to give a full historical account, and so the below is limited to the most important key facts.

Kenya gained independence from the UK in 1963 and has, since then, been relatively successful in its economic development. Kenya is not a *least developed country*[41], and despite being ethnically diverse has managed to avoid any major conflicts over its history. The only exceptions to this are the Mau Mau uprising 1952-1959,[42] and the 2007-08 election disruption, the latter of which which caused wide spread violence and looting across the country. The Mau Mau uprisings, although a significant event in Kenyan history, lie too far from the period under consideration to be relevant. On the other hand, the civil unrest caused by disputed elections in 2007, is of potential relevance. The nature of this unrest was in disputed elections between the incumbent Mwai Kibaki and Raila Odinga[43], and was tribal in nature. Over 200 people lost their lives and hundreds of thousands were displaced. Violence erupted almost immediately after Kibaki was announced as president, and was heavily concentrated in urban areas, particularly Nairobi and Mombassa, and continued throughout January 2008. Economic impacts during this time were significant with many businesses closing, however the long run impacts are likely to be small in all sectors, with the possible exception of tourism which suffered a large

[41] The UN states that "Least developed countries (LDCs) are low-income countries confronting severe structural impediments to sustainable development. They are highly vulnerable to economic and environmental shocks and have low levels of human assets." Kenya on the other hand is classified, by the UN, as a developing low-income country, one step up from being an LDC.

[42] See (Edgerton, 1991), for an overview of this conflict

[43] Kenya has a multi-party system but two main coalitions formed headed by Kibaki and Odinga thus this election effectively became a two horse race.

negative impact in the medium term due to extensive coverage (and exaggeration) of the violence in Western media. Violence tapered off in February and March, by April the country was back to normal. As the labour market data used in this paper is taken from an annual survey performed in June, this violence is unlikely to have a large impact on employment or wages in all industries except those relating to tourism. Thus, in empirical work we consider outcomes relating to industries most reliant on tourism in 2008 as anomalies, and so control for potentially spurious measurements of wages in the retail industry in 2008.

The Kenyan economy is still dominated by agriculture which employed 75% of the labour force, mainly in subsistence farming. However, services contribute to an ever increasing proportion of GDP, now around 62%. Kenya's main exports are cash crops: tea, coffee and picked flowers. Manufacturing development has been slow, accounting for only 14% of GDP[44] and almost solely concentrated around Nairobi, Mombassa and Kisumu. Perhaps fortunately for Kenya, it possesses no significant mineral deposits, and imports all its oil. This revelation is particularly interesting when considered with respect to recent work on *consumption* cities and their relation to natural resource extraction (Jedwab, 2013). This thesis will be able to shed some light on the conjecture of urbanisation without industrialisation in SSA countries, without relying on the presence of natural resources. Kenya also has a fairly well developed education system, with free primary schooling implemented by Kibaki in 2003 and heavily subsidised secondary education however, 38.5% of Kenya's remain illiterate, although this hides wide geographic heterogeneity between Nairobi and the north[45].

Of particular interest in this thesis is the development of Kenya's transport network. Lacking traversable rivers, the transport structure in Kenya can be categorised into three groups. Firstly, the rail network: Kenya has approximately 2,000km of rail roads, built by the British over 100 years ago. As a result of their age

[44]Statistics discussed in this section are taken from the Kenyan statistical abstracts unless otherwise stated

[45]87.1% in Nairobi compared to 8% in the North Eastern Province. Education statistics have been taken from the Kenya National Adult Literacy Survey, 2007

and lack of maintenance, much of the network is no longer commercially viable with total quantities of cargo and passengers steady declining over the last 30 years. In 2014 Kenya received $429m from China to repair and rebuild the Mombassa-Malaba line to modern standards. This is likely to cause somewhat of a transportation revolution in the area, and the results of this thesis are of particular importance when considering the effects of this future investment. In section 6.1, we use the whole model to consider the counterfactual of the construction of this rail road and analyse the possible impacts it may have of the wage structure across Kenya. I find that the proposed railway, although boosting national wages, is most beneficial in areas with already high wages, and for high wage industries, and therefore may have unintended negative distributional effects both geographically and by industry.

Secondly, and more importantly in recent years, we turn to Kenya's road network. Present day roads were mainly built prior to 1990, but after British rule. They shoulder most of the transportation needs of Kenya, in terms of the movement of both goods and people. Few new roads are built over the sample period, most variation is in roads being upgraded, from tracks to paved roads or from provincial to main roads, we observe these changes. Although we can measure road upgrading we cannot measure road maintenance, the fixing of pot holes etc. which increase transport times without changing the status of the road. Lastly, I consider the Lake Victoria ferries, these train ferries connected the lake ports of Kisumu (Kenya), Port Bell (Uganda) and Mwanza (Tanzania) and provided an alternative means of transporting goods (and passengers) between the three countries. These ferries have been in operation since before 1990, however after a period of decline in 2006 operations where completely terminated. In 2010 a US company Earthwise Ventures announced that it was going to bring a fleet of fast moving rail ferries to Lake Victoria bringing significant investment, however this is yet to materialise, but would be an interesting large infrastructure investment in the region, the potential impacts of which, once again, this paper can shed light on.

Also important for Kenya's development is the success of its neighbours.

Uganda and Tanzania are Kenya's main trading partners and have become increasingly integrated in the Kenyan economy in recent years with integration pushed by the East African Community. The East African Community (EAC) is a regional intergovernmental organisation which currently consists of six member states, Burundi, Rwanda, South Sudan, Kenya, Tanzania and Uganda, although only the final three play a significant role.

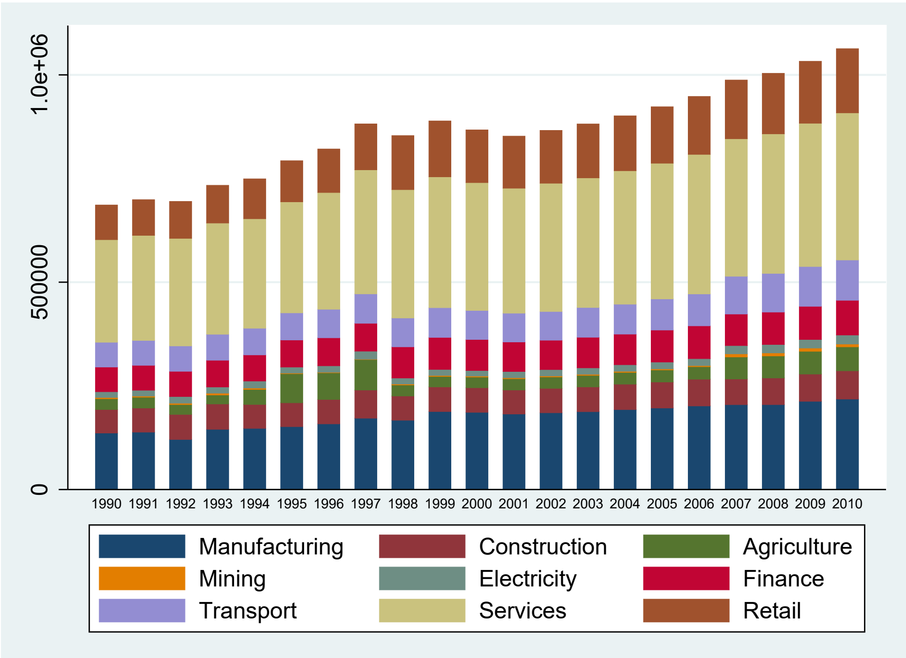
Kenya, Tanzania and Uganda exhibit significant variation along the dimensions of conflict (often relating to ethnic tensions), resource endowment and competence of political leadership (leading to institutional variation). However, they share many similarities also, all three are very poor, still mainly reliant on subsistence agriculture, speak the same languages, share a common colonial past, ageing rail infrastructure as a relic of this, and have made little progress towards industrialisation. Two further factors contribute to their homogeneity, firstly all three have access to the largest lake in Africa, Lake Victoria, and secondly, all three are founding and active members in the East African Community (EAC). The geography of Lake Victoria, and processes of the EAC need to be understood when studying the transport network in these countries. The stated aim of the EAC is the eventual creation of an East African Federation, which will be completed in four steps (1) customs union, (2) common market, (3) monetary union and (4) full federation. Thus far the EAC has implemented the first two stages of its four-step plan. In 2005 a customs union was ratified, agreeing on free internal trade and common external trade tariffs, this could potentially have large effects on Kenya, Tanzania and Uganda's economies, although effects are expected to be muted given the small size of their domestic markets. Despite passing into law in 2005, convergence to the new laws was slow with the union only becoming fully fledged in 2010. In 2010 the second stage was implemented, that of a common market. However similarly practical implementation has been slow and is outside the considered sample period.

D Further facts from the Kenyan Statistical Abstracts

In this appendix I document some additional descriptive figures derived directly from the digitisation of historical Kenyan statistical abstracts. These facts, although simplistic, bring evidence to support previously unknown, but supposed, facts about the distribution of the Kenyan work force over space and time.

Firstly, we document simply that the number of wage employed persons employed in the 22 cities of our sample, is increasing (figure 13), from 686,123 in 1990 to 1,063,298 in 2010 that is an increase of over 50% in 20 years. From 1989

Figure 13: WAGE EMPLOYMENT BY INDUSTRY OVER TIME



Notes: This figure shows how the number of wage employed has changed in the 22 major cities of Kenya studied over time.

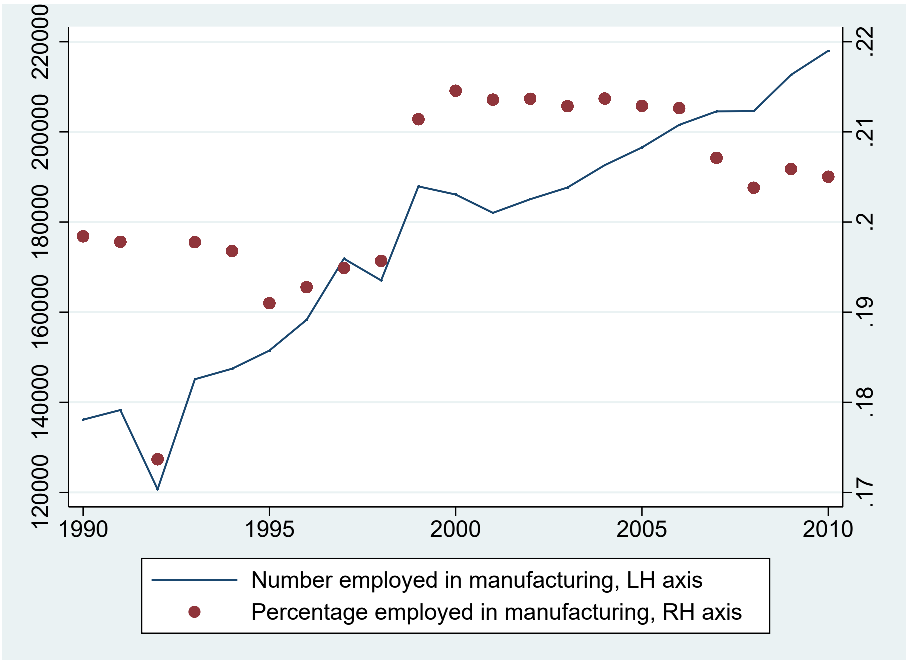
to 2009[46] the Kenyan population has increased from 21.4m to 38.6m, an increase of over 80%. Over this time period the labour force (those aged between 15 and 64) almost doubles from 10.5m to 20.7m, however those in wage employment only increased from 1.4m to 2m. One can see from these numbers that Kenya has a rapidly growing, and young, population. It is also useful to note that although the

[46] 1989 and 1999 are the relevant census years from which accurate population data is available.

data used on wage employment in this thesis only picks up $\sim 5\%$ of the population, this is $\sim 10\%$ of the labour force and $\sim 50\%$ of all individuals in wage employment.

Secondly, we can look for evidence of urbanisation without industrialisation a-la Jedwab (2013). Figure 14 plots the total number employed in manufacturing and the proportion of all wage employment in manufacturing over time. Figure 14

Figure 14: WAGE EMPLOYMENT IN MANUFACTURING OVER TIME

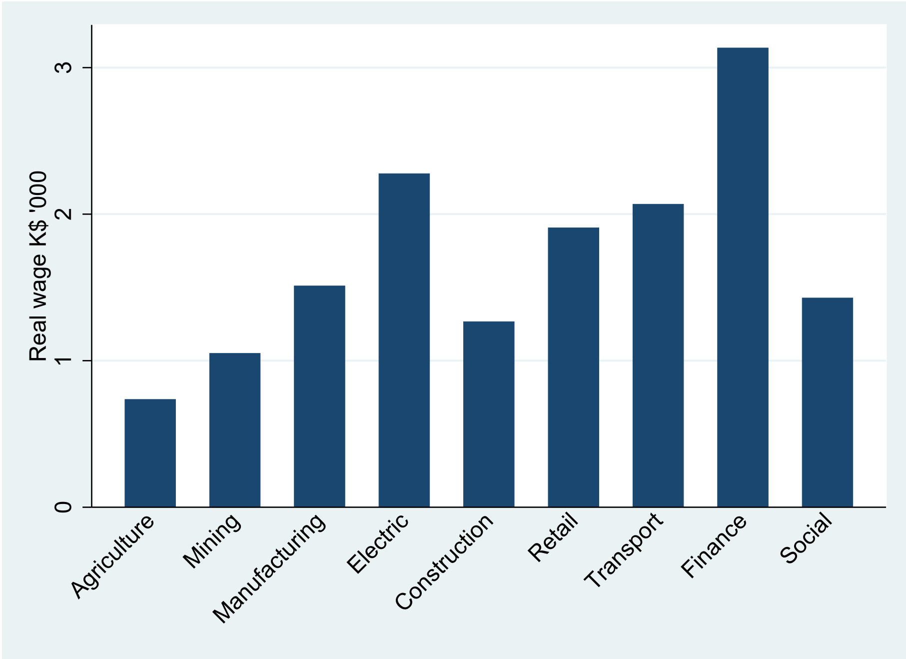


Notes: This figure compares the number employed in manufacturing (left hand axis) with the proportion employed in manufacturing (right hand axis), in the 22 major cities of Kenya considered.

shows that, although the number of people in wage employment in manufacturing has increased dramatically from 1990 to 2010, the proportion employed has stayed constant. This appears to support the thesis of urbanisation without industrialisation in Kenya, a figure of 20% being employed in manufacturing may seem high for a SSA country, but one must remember that this data only picks up those in wage employment. Given that most employed in the manufacturing sector will be in wage employment compared with the much larger agriculture sector, this figure is considerably inflated. As a percentage of the total labour force, the number of wage employed in manufacturing is $\sim 3\%$, much lower.

Turing our attention to wages, figure 15 shows the average wage over the time period and every city in the sample, by industry. As anticipated, wages in the agriculture and mining sectors are lowest, and wages in finance are the highest.

Figure 15: AVERAGE REAL WAGE BY INDUSTRY



Notes: This figure the average real wage in each city in 1992 prices, averaging over all 22 cities considered and the time period 1992-2010 considered.