Roads and Jobs in Ethiopia*

Matteo Fiorini[†] Marco Sanfilippo[‡]

July 24, 2022

Abstract

We look at how improving roads can affect jobs and structural transformation. We use a novel geocoded dataset covering the universe of Ethiopian roads and match this information with individual data to identify the relation between improvements in road infrastructure and labor market outcomes over the 1994–2013 period. We find that at the district level, greater market access due to better roads correlates with the process of structural transformation in Ethiopia. Improvements in market access are related to reductions in the share of agricultural workers and increases in that of workers in the services sector, but not in manufacturing. We show some heterogeneity across industries, gender, education level and age cohorts. Finally, investigating the underlying mechanisms, we show that patterns of internal migration and changes in economic opportunities can help rationalize our findings.

Keywords: structural transformation; jobs; roads; Ethiopia
JEL Classification: L16; O18; O55; R4

* We would like to thank the editor, Eric Edmonds, three anonymous reviewers, Margaret McMillan, Ali Sen, Kunal Sen, Victor Stolzenburg, participants of the annual UNU-WIDER Conference at ESCAP in Bangkok, "The Future of Industrial Work" Workshop at UNIDO in Vienna, the Jobs & Development Conference (online), the Nordic Conference on Development Economics (online) and seminars at the Universities of Catania and Napoli Parthenope for valuable comments on preliminary draft versions of the paper. We also thank Francesco Iacoella for valuable research assistance. Support from Fitsum Mulugeta and the Ethiopian Economic Association is gratefully acknowledged for facilitating the collection of data from the Central Statistical Agency. Matteo Fiorini acknowledges financial support from the European Union's Horizon 2020 research and innovation program under grant agreement No. 770680 (RESPECT).

† Robert Schuman Center for Advanced Studies, European University Institute. Mail: matteo.fiorini@eui.eu

‡ Department of Economics and Statistics, University of Turin and Collegio Carlo Alberto. Mail: marco.sanfilippo@unito.it

Introduction

Developing countries can hardly embark on economic development and structural transformation without first reducing transport costs (Gollin and Rogerson, 2014). Greater connectivity can improve the lives of individuals, widening their work and educational opportunities, while fostering transition to more productive activities. In addition, improving domestic transport infrastructure can reduce some of the constraints that affect the private sector in many low-income countries, allowing firms to better connect to local and international markets. This, in turn, can improve the efficiency of firms and enable them to offer better jobs. Understanding whether policies supporting the construction of transport infrastructure can affect jobs in low-income countries is therefore a question of high policy relevance.

In this paper, we look at whether and how developments in road infrastructure interact with labor market outcomes in Ethiopia. We take advantage of the collection of very granular information on a recent large-scale program, the Road Sector Development Programme (RSDP). The RSDP started in 1997, with the aim of improving connectivity across the country through the rehabilitation of existing roads and the construction of new ones.[1] In the space of just a decade, the improvements due to the RSDP have been remarkable. Road density rose from 24.1 per 1000 km^2 when the program started to 44.4 in 2010 (when an evaluation of the first three rounds of the program was completed; Ethiopian Road Authority, 2011). Over the same period, the proportion of the road network in good condition increased from 22% to 56%.

Our analysis uses geo-localized information on the Ethiopian road network, for which we track specific road-segment improvements undertaken through the RSDP. We match information on the road network with information at the level of individuals. Individual information was taken from the 1994 Population Census and the Ethiopian National Labour (NLF) Survey, a nationally representative survey of Ethiopian workers available for the years 1999, 2005 and 2013. We use the district (or *woreda*, the third administrative unit level in Ethiopia) as the unit of analysis. To better explore how transport infrastructures affect labor demand, we further combine road data with additional information on the activity of firms.

The case of Ethiopia is particularly relevant for our purposes. Beginning with the agricultural-development-led industrialization (ADLI) strategy in 1995, and later with growth and transformation plans, a large emphasis has been attributed to structural transformation. This policy agenda promotes entrepreneurship and diversification into highly productive activities. Improving connectivity both within the country and with external markets occurs in parallel with the pursuit of structural transformations and economic upgrading (Ali, 2019).[2] Existing evidence from Ethiopia shows that high transport costs have so far represented barriers to market integration (Atkin and Donaldson, 2015; Gunning et al., 2018) and labor supply (Franklin, 2018).

We study the impact of road infrastructure as an indicator of market access. This allows us to

[1] Note that roads represent the main transport infrastructure in Ethiopia over the period covered by our analysis. The railway, connecting Addis Ababa to Djibouti, was in fact re-established in 2017.

[2] The recent efforts to develop industrial parks in the country is consistent with the idea that employment creation in non-agricultural sectors and the resulting pattern of structural transformation are indeed dependent on reliable infrastructure. Note, however, that our sample covers a period during which none of the industrial parks were operating.

account for the direct and indirect effects of roads investments that took place all over the country under the RSDP. Improved access to markets makes locations more attractive to production and consumption, raising population density, the relative price of non-tradable goods (Fajgelbaum and Redding, 2018) and, more generally, fostering economic activity (for instance, Storeygard, 2016; Alder, 2019; Chiovelli et al., 2019; Eberhard-Ruiz and Moradi, 2019).

Changes in market access alter the economic environment for both firms and workers, affecting the labor market. Improvements in road infrastructure reduce firms' transport costs, increasing market opportunities while lowering the cost of sourcing inputs. This can trigger private sector development through increased entry, and higher performance, and ultimately generates an increase in labor demand.[3] However, better roads also increase competitive pressures faced by firms, with potentially opposite implications on labor demand. On the supply side, roads can contribute to pushing workers out of agriculture, which is still the prevalent source of employment in the country. This happens primarily through improvements in farm productivity (due, for instance, to greater access to new and imported inputs).[4] In addition, lower transport costs reduce constraints to migration choices (Morten and Oliveira, 2017; Lagakos, 2020).[5]

Understanding how labor demand and supply interact in the Ethiopian context in response to improved road infrastructure and the consequences of this on employment is therefore an empirical question that we try to address in the paper.

In our empirical analysis we exploit the time-series dimension of improvements in roads within each district. This feature of the data allows us to run a regression with district fixed effects and region-specific time trends, to control for time-contingent shocks and to partial out confounding heterogeneity across districts[6]. Many factors can simultaneously concur with improvements in market access, both within and outside a district's borders. In all of our specifications, we try to minimize endogeneity by controlling for improvements to roads within each district that are orthogonal to changes in the district's connectivity (similarly to Donaldson and Hornbeck, 2016). Still, this does not ensure that our results can be interpreted in a causal way, given that factors unrelated to the characteristics of individual districts can also play a role (e.g., programs targeting remote locations in a process of regional convergence). Hence, in the rest of the paper we are careful to avoid interpreting these as causal effects. Nevertheless, we feel that the relationships are sufficiently interesting and, importantly, policy-relevant to justify our analysis.

Our results show that some of the changes occurring in the labor market of Ethiopian districts are associated with improvements in road infrastructure. There is no evidence that districts im-

[3]For the case of Ethiopia, Fiorini et al. (2021) show that improvements in market access due to the RSDP are a necessary condition for firms to experience productivity gains from trade liberalization.

[4]Recent evidence provides support linking improvements in connectivity under the RSDP with increases in agricultural productivity (Adamopoulus, 2020; Gebresilasse, 2020).

[5]According to a report by the recently established Ethiopian Jobs Commission (Jobs Creation Commission Ethiopia, 2019), increases in migration (mostly rural–urban) do exert a pressure on urban labor markets, with likely consequences on wages, unemployment and the size of the informal sector.

[6]Most of the existing studies on the impact of infrastructures employ a difference-in-differences approach (see, for instance, the review by Redding, 2020). Dercon et al. (2009), Mu and van de Walle (2011), Faber (2014), and Storeygard and Jedwab (2020) use a specification in first differences that wipes out any fixed effects in the level of the economic outcome of interest. Closer to the empirical approach adopted here, the papers by Alder (2019), Gebresilasse (2020), Aggarwal (2018) or Khandker et al. (2009) employ instead specifications in levels controlling for location and time fixed effects.

proving market access experience increases in employment, but we observe changes in the sectoral composition of the workforce. This happens through a reduction of agricultural workers and an increase of workers in the services sector, but not in manufacturing. Improvements in roads seem to go hand in hand with a pattern of structural transformation without manufacturing, which is consistent with the findings of other studies looking at the dynamics of structural transformation in the region (Rodrik, 2016; Baccini et al., 2021). At a more disaggregated level, improvements in roads are associated with increases in jobs due to the provision of market services. We also do not find evidence of changes in the composition of jobs due to increases in the construction sector or in government-related activities.

In the second part of the analysis, we account for the heterogeneity in individual characteristics. First, we show the existence of gender-specific patterns in our empirical framework. Within the services sector, women seem to respond more quickly to opportunities from improved market access compared to men, a result that confirms previous evidence on the gender-specific benefits of infrastructure (e.g. Dinkelman, 2011; Lei et al., 2019). Second, we find evidence that larger proportions of the working-age population with a higher education are in districts where investments in road infrastructure provide increased market access, as well as increased participation for the school-age population. Third, disaggregating by age cohorts, we show that the main dynamics particularly involve the youngest workers.[7]

Finally, we investigate some of the potential economic mechanisms that can help to understand our results. By looking at migration patterns, we find evidence of domestic migration to areas characterized by higher market access, along with increases in migrant employment in modern activities. These findings can help explain the structural transformation response to the road infrastructure reform in Ethiopia. Next, using data on both formal and informal manufacturing firms, we find that the response of manufacturing firms to greater market access includes improvements in productivity as well as a relative increase in the number (and wages) of non-production workers. On the other hand, firms operating in trade-related services (i.e., wholesalers and retailers) respond to higher market access by increasing their average size, but with no change in their productivity.

Our work is related to a growing body of research using micro data to investigate the drivers of structural transformation in developing countries (see Lagakos and Shu, 2021, for a recent review of the existing evidence). Among those drivers, infrastructure has been widely studied due to its persistent effects on urbanization and the distribution of economic activities across space, as well as due to second-order advantages related to lowering the costs of migration (e.g., Adam et al., 2018; Bryan and Morten, 2018; Hjort and Poulsen, 2019; Khandker et al., 2009; Adukia et al., 2020; Asher and Novosad, 2020; Storeygard and Jedwab, 2020).[8]

[7]The main patterns identified in our analysis are robust to several checks, including different specifications, various cuts of the data and alternative definitions of the variables of interest.

[8]Among these studies, Hjort and Poulsen (2019); Adukia et al. (2020); Asher and Novosad (2020) look specifically at how connection to infrastructures can affect structural transformation. Similar to our findings, Hjort and Poulsen (2019) show that this effect is likely driven by the rise of a more dynamic private sector in treated locations. Asher and Novosad (2020) and Adukia et al. (2020) exploit rich information on the construction of roads in rural villages in India. They offer a more nuanced set of results. Investments in roads do stimulate the reallocation of workers out of agriculture and higher investment in education but do not significantly increase local economic activities in treated areas.

Most of all, we contribute to a small strand of evidence looking at the consequences of infrastructural investment in Ethiopia, and in particular under the RSDP. Shiferaw et al. (2015) provide evidence on the positive effects of the RSDP on business dynamism, finding evidence of more entry in the formal sector. Fiorini et al. (2021) show that the reduction in transport costs enabled domestic manufacturing firms to take advantage from trade liberalization, increasing their productivity. Their findings are in line with our mechanisms relative to the manufacturing sector firms. The lack of employment growth in the formal manufacturing sector that we find in our analysis despite gains in productivity is consistent with findings by Diao et al. (2021), who attribute it to the diffusion of capital-intensive techniques related to global trends in technology.

Adamopoulus (2020) and Gebresilasse (2020) link improvements in connectivity under the RSDP to increases in agricultural productivity using a panel of Ethiopian districts over a similar period to the one we cover[9]. Their findings offer a complementary perspective to ours, providing evidence on a mechanism that we cannot test with our data, i.e., increased productivity in agriculture due to reduced transport costs and higher market access.[10] The findings of Adamopoulus (2020) are especially relevant for us. He shows that due to a fall in transport costs production shifts to more productive areas that specialize in the production of cash crops for the export markets, and concentrate in larger farms. This reduces the labor required for food production and generates a structural shift to non-agricultural activities. An important difference with our work is that these two papers look at rural areas and more specifically at the effects of the Universal Rural Road Access Program (URRAP), which was introduced in 2011 with the aim of connecting rural villages to all-weather roads.

Much closer to the spirit of the current paper is the work by Moneke (2020), which also looks at the role of infrastructural investments for structural transformation in Ethiopia. There are both differences and similarities among the two studies. First, the scope of his work is broader than ours. His paper investigates existing complementarities in road construction and electrification and uses a quantitative model to understand their welfare implications. Second, there are differences in how the papers measure road improvements. While Moneke (2020) uses a dummy variable indicating the presence of all-weather roads in a given district, we adopt a market-access approach that allows us to account for both the direct and indirect benefits of road construction and expansion. This allows us to test some specific mechanisms, i.e. migration and firms' expansion, that are more likely to depend on country-wide, rather than merely local, road improvements[11]. Still, despite these differences he also finds that road investments alone promote structural transformation out of agriculture and towards services but not into manufacturing.

The remainder of the paper is organized as follows. Section 1 presents the data and Section 2 describes the empirical strategy. Section 3 introduces the core results, while some extensions and the robustness checks are discussed in Section 4. Section 5 offers an empirical investigation of

[9] The paper by Dercon et al. (2009), though not explicity evaluating the RSDP, shows that access to all-weather roads has important effects in terms of both consumption growth and poverty reduction among Ethiopian rural households.

[10] Though they do not test this specifically, there is considerable evidence showing that increases in productivity cause a drop in this sector's employment and, ultimately, structural transformation (Bustos et al., 2016)

[11] Moneke (2020) finds no evidence of local roads' improvements on internal migration, but shows evidence on positive selection of migrants driven by roads. No comparable analysis is performed on firm level data.

the mechanisms driving our findings. Section 6 offers some concluding remarks.

1 Data

Individuals. Individual-level data are obtained by combining two sources that provide complementary information. The first is the Ethiopian National Labour Force (NLF) survey. This is a representative survey of both urban and rural areas administered by the Central Statistical Agency (CSA), with the objective of monitoring the economic and social conditions of the economically active population. The information provided in the survey includes, among others, the demographic characteristics of the individuals, their education and working conditions. The NLF includes information on whether respondents report a previous residence different from the current, thus allowing the identification of internal migrants, as well as on the formal or informal nature of an individual's current job. We use all existing waves of the NLF, covering the years 1999, 2005 and 2013.[12]

A limitation of the NLF surveys is that they do not cover the period before the RSDP. To address this issue, we combine the NLF with the 1994 population census, which also provides details on the distribution of workers across industrial sectors.[13]

Once the NLF surveys and the census datasets were harmonized,[14] we collapsed all of the information at the district-year level using sample weights to recover information on the underlying population.

Table 1 reports the distribution of labor shares over the 1994–2013 period, computed at the national level using the sample of working-age population. While employment is on the rise, there is also evidence of the process of structural transformation occurring in the country. Over time, workers are less engaged in agriculture and more active in the services sector. However, despite a visible increase in its employment share, the manufacturing sector remains small. The data also show that the relative position of women in the labor market has improved over time (they represent about 47% of total workers in 2013, up from 43.8% in 1994), especially in the services sector.

Roads. The main source of information on road infrastructure is a proprietary geo-spatial database consisting of coded reports by the Ethiopian Road Authority (ERA) covering all road construction and/or rehabilitation sites that were opened under the different phases of the RSDP. The data are organized as a time series of shapefiles of the Ethiopian road network, reporting

[12] The NLF surveys are representative at the national level and use regions, the first administrative units, as the main sample domains. They cover all urban and rural areas of the country except the non-sedentary areas in the Somali region. The sampling frame to select enumerator areas is provided by the population census (the 1994 census for the 1999 and 2005 NLF waves and the 2007 census for the 2013 wave). All of the relevant information on the sampling procedures, coverage and full descriptive statistics are available in the survey reports published by the CSA (2004, 2006, 2014).

[13] This information is not included in the 2007 population census, which we do not use given the specific purposes of our analysis.

[14] NLF survey data are not geocoded but include identification codes for each location, including region, zone and district. To combine the different waves of data, we used the definition of district (woreda) provided by IPUMS that matches districts using their names when the geographic definition of borders differed between the 1994 and 2007 censuses. Overall, the final estimation sample covers, on average, about 80% of the estimated total population in each wave.

Table 1: Sector composition of employment

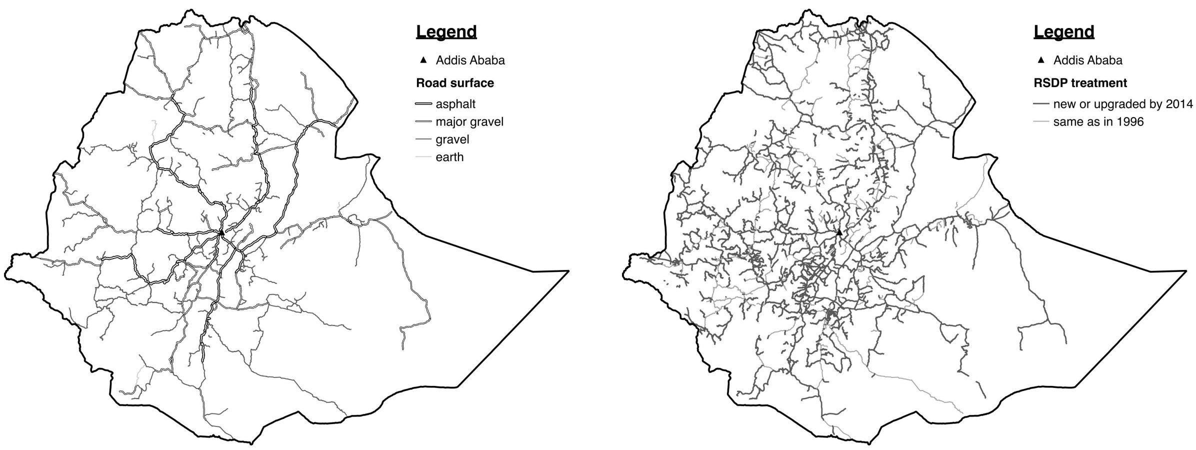
Year	Employment	Agriculture	Manufacturing	Construction	Services
1994	76.21%	87.95%	1.99%	0.35%	9.62%
1999	75.69%	77.67%	4.80%	1.05%	16.35%
2005	81.09%	77.89%	5.35%	1.65%	14.99%
2013	80.83%	70.84%	5.17%	2.35%	21.11%

Source: Authors' elaboration on CSA data. *Notes:* The first column reports the proportion of employed persons in the working-age (15–64) population. Following the NLF report, a worker is defined as a person who declared at least 1 hour of work during the week preceding the interview. The following columns report the proportion of sectoral workers out of the total number of employed persons in the specific year. All data have been weighted before collapsing information at the national level.

two main attributes for each geo-localized road segment: the type of road surface and the road's condition.[15]

Figure 1a presents the network of federal and regional roads in 1996 by surface type. Figure 1b shows the same types of roads in 2014, distinguishing between segments that existed in 1996 and were not rehabilitated by 2014 (light-grey segments on the map) and roads that were either newly constructed or rehabilitated during the first three phases of the RSDP. A visual inspection of the two maps shows a substantial expansion of the road network between 1996 and 2014. Moreover, road development does not appear to be geographically concentrated but, rather, spans over different administrative areas across the country. The information on surface type and condition can be aggregated to compute the average travel speed for each road segment at each point in time. This is done following a standard speed matrix proposed by the ERA and reported in Table A.1.[16]

Figure 1: Federal roads, regional roads and the RSDP



(a) RSDP roads in 1996 by surface type

(b) New and upgraded RSDP roads from 1996 to 2014

We employ an indicator of market access (Donaldson and Hornbeck, 2016) to measure the economic effects of infrastructural development in the context of a formal structural gravity trade model. In the context of the present paper, and similarly to Storeygard (2016), market access

[15]There are four types of road surface in the data: earth surface, minor gravel (which identifies regional rural roads with a gravel surface), major gravel (federal gravel roads) and asphalt. As for road conditions, the database distinguishes between two categories: not rehabilitated and new or rehabilitated.

[16]The same speed matrix has been used by Shiferaw et al. (2015) and Storeygard and Jedwab (2020).

captures the structure of road connections between a geographically defined area and all other markets in the country, weighted by the intensity of their economic activity.

For each district i, market access is defined as the weighted sum of income in each district z different from i, with weights equal to the iz bilateral transport cost scaled down by a trade-elasticity parameter. Formally:

$$\text{Market Access}_{it} = \log\Big(\sum_{z \neq i} D_{iz,t}^{-\theta} L_z\Big) \tag{1.1}$$

$D_{iz,t}$ is the minimum distance in hours of travel between district i and district z given the road network in place at t. Bilateral distances in travel hours are computed applying the Dijkstra algorithm on the network of Ethiopian districts (the nodes are set at each District's centroid) connected by federal and regional Ethiopian roads (links). L_z is an indicator of economic activity based on night light intensity in z. θ is the trade-elasticity coefficient.

There are different values of θ in the literature, ranging from 1 to 10 (Donaldson and Hornbeck, 2016; Chiovelli et al., 2019) depending on the context. In this paper, we use a trade elasticity of 3.12. We obtain this value following the procedure adopted by Storeygard and Jedwab (2020), i.e., combining the estimated trade elasticity with respect to roads for the US (1.27) with the difference in cost–travel time elasticity that has been estimated for Ethiopia by Atkin and Donaldson (2015) (2.46 times the US value).[17] In Section 4 we show, however, that our main results remain consistent with the adoption of different values of θ.

While some papers—including Donaldson and Hornbeck (2016)—use population data in the computation of market access, we employ nightlight intensity data as in Storeygard (2016), Chiovelli et al. (2019), Baum-Snow et al. (2018) and Alder (2019).[18] This is particularly appropriate given that nightlight is a better indicator of local economic development, in contrast to population, which provides improved information on the size of an area but lacks information on its purchasing power (Chiovelli et al., 2019). With specific reference to Ethiopia, satellite imagery has a better capacity to catch local economic development and population dynamics especially in lowland areas (about 60% of the country's territory) where part of the population lives in nomadic, semi-nomadic or pastoral way, so that official data are less likely to provide precise information.[19]

Information on nightlight intensity is sourced from NOAA National Geophysical Data Center (2018) and is available at the level of 0.86 km^2 grid cells over the whole country area. For each cell the nightlight intensity score can vary from 0 to 63.[20] For each district we compute the

[17] In their paper, which looks at the effects of market access on urbanization in Africa, Storeygard and Jedwab (2020) obtain a value of 3.8 because they use the estimated cost–distance elasticity for Nigeria, which is 3 times larger than that of the US (1.27*3=3.8). Since the paper by Atkin and Donaldson (2015) provides detailed estimates on the cost–distance ratio for Ethiopia (see their Table 4), we use that value.

[18] Note that other works (e.g., Alder, 2019) use market access computed with beginning of sample period weights as an instrumental variable (IV) for the market access using time-varying night light density weights.

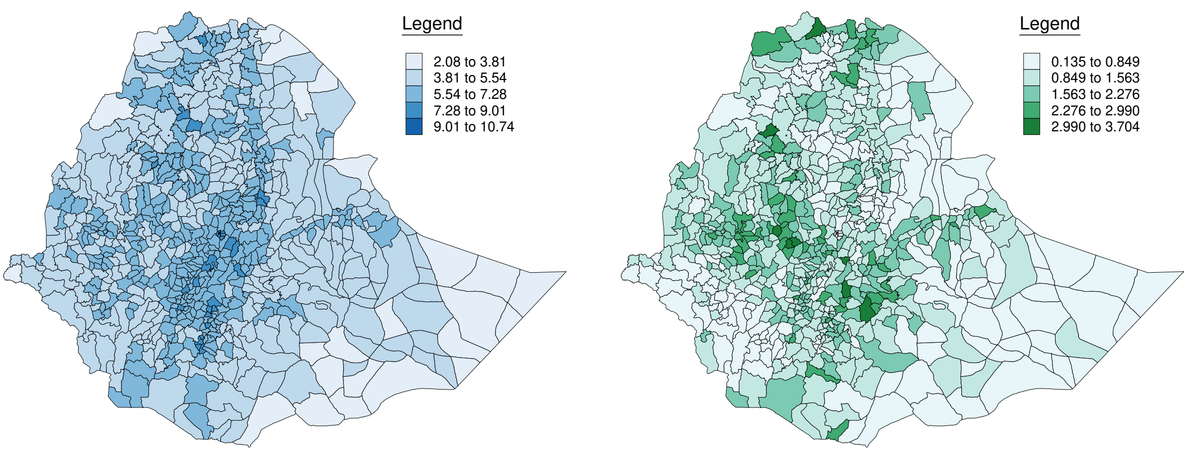
[19] See, for instance, a recent analysis by the World Bank, available at the following link: `http://devseed.com/ethiopia-docs/`

[20] Following Eberhard-Ruiz and Moradi (2019), we use scores from raw satellite images, instead of processed images with stable nightlights, as more reliable proxies of economic activity in small and medium African urban

sum of the nightlight intensity scores across all the cells within the district's border. We fix the weight L_z at the beginning of the sample period (1994) to exclude potential correlation between changes in destinations' economic activity and our outcome variables. The mean value of L_z is 6176, the median is 3435 while the minimum and maximum values are 0 and 83408 respectively.[21] A potential drawback of using nightlight instead of population is that the former includes many zeros. Since our unit of analysis—the district—usually includes both urban and rural areas, this is less of a concern compared to more granular settings.

Figure 2a plots the value of the market access indicator at the beginning of our baseline estimation sample (1996) for all Ethiopian districts covered in our estimation sample. Figure 2b shows the change in market access between 1996 and 2014 for each woreda. Focusing on Figure 2a, dark blue woredas near the center of the country close to Addis Ababa reveal higher market access in this area. Figure 2b shows a larger increase in market access for less-connected districts away from the center, suggesting that improvements in road infrastructure occurred over the time period of our analysis.

Figure 2: Market access. Starting point and change by woreda



(a) Market Access$_{r,1996}$

(b) Market Access$_{r,2014}$−Market Access$_{r,1996}$

By combining information on travel time expressed in hours with data on nightlight intensity, single values of the market access variable cannot be interpreted in isolation. To get a better sense of the quality of road connections behind a certain value of market access, we can look at relevant statistics computed on the distribution of travel times from the origin district and in the year corresponding to that value. For instance, the median year–district observation in terms of market access is 1996 Ganta Afeshum in the northern Tigray region, with a value of 6.01968. The average travel time from Ganta Afeshum to all other districts in our sample in 1996 was 39.5 hours, with a minimum of 3.5 hours. The observation with the highest value of market access in our sample is 2014 Arada (with a value of 10.946), followed by 2014 Lideta (10.904). Arada

areas.

[21] Measuring economic activity via nightlight has been subject to criticism in the literature. For instance, two recent papers (Asher et al., 2021; Gibson et al., 2021) point out that nightlight lacks temporal consistency and is thus less suited to time-series analyses. We note that these concerns are mitigated in our case since we use cross-section variation in nightlight to weight our market access measure.

and Lideta are two districts in the region of Addis Ababa. The average travel time to connect them with all other districts in our sample in 2014 was approximately the same for both and is equal to 13.7 hours, with minima of less than 20 minutes. Finally, the observation with the lowest value for market access is 1996 Moyale (3.163). Moyale is a district in the Somali Region. This woreda includes the southernmost point of the whole country, on the border with Kenya. The average travel time from Moyale to all other districts in our sample in 1996 was 55 hours, with a minimum of 19 hours.

2 Empirical Specification

The objective of our empirical analysis is to study the link between improvements in connectivity, captured by variation in market access, and district-level labor market outcomes in Ethiopia. For each outcome variable, we propose the following baseline specification:

$$y_{it} = \beta \text{ Market Access}_{it} + \gamma_i + \rho_{rt} + \epsilon_{it}, \tag{2.1}$$

where y captures a generic outcome variable among those included in our panel of Ethiopian districts i across years t. These include the proportion of the population employed and the percentage of employment in agriculture, manufacturing and services. The term Market Access$_{it}$ is the measure of connectivity between district i and relevant economic activity in the rest of the country at time t. Each specification includes district fixed effects and region-specific time trends. District fixed effects are important to control for all the time-invariant characteristics of the district (e.g., geophysical features, such as soil quality and elevation) that can simultaneously affect the decision to invest in roads and labor market outcomes. Region-specific time trends account for common changes (e.g., regional policies, or changes in regional budget on roads) that can confound the relationships among the outcomes and the treatment.

In estimating equation 2.1, standard errors are clustered at the regional level.[22] Due to the small number of regions (n=11), all estimation tables in the paper report wild cluster bootstrap standard errors (Cameron et al., 2008),[23] although we show in Section 4 that the results are consistent with different clustering strategies. Finally, all regressions are weighted by each district's population.

Our estimation sample consists of an unbalanced panel of 1,573 observations covering 506 districts. Taken together, these observations account for over 80% of the total population and total jobs in the country. Table A.2 in the Appendix reports descriptive statistics of our outcomes of interest and of the main regressors.

A potential threat to identification in our empirical setting is the endogeneity of the main regressor of interest, Market Access$_{it}$. While the fixed effects capture the main sources of omitted-variable bias, potential confounding heterogeneity at the district–time level remains an active

[22] In this we follow Abadie et al. (2017), and we cluster according to the sampling strategy of the NLF surveys, which are representative at the regional level.

[23] We implement this in STATA using the boottest routine written by Roodman et al. (2019)

source of endogeneity in our specification. Reverse causality can also play a role, with time-contingent shocks to local employment and/or economic activity shaping incentives for investment in local roads. For instance, geographic areas with relatively larger (smaller) agricultural or service sectors might be systematically more (less) interested and successful in attracting infrastructure investment for the improvement of local roads in the district.

We follow an identification strategy similar to the one proposed by Donaldson and Hornbeck (2016). More precisely, we exploit the fact that variation in each district's market access is determined by improvements to the whole national road network (as we keep market size fixed, market access does not respond to changes in economic activity over time). Moreover, when captured in the market access measure, improvements in local road segments reflect not only the higher road coverage and/or travelling speed but also their contribution to the district connectivity relevant to economic activity in the rest of the country. We can therefore partial out the changes in the quality of the *local* road network, which are a major source of endogeneity concerns. Indeed, one might hypothesize that political economy forces at the local level lead to both changes in aggregate and/or sectoral employment within the district and to investments in local roads. We capture the district-level infrastructure developments through a weighted sum of the distance covered by each road segment within the district area, with weights equal to the speed allowed by the type of surface and the road condition. We denote this variable as *Local Roads.*[24]

While it is fair to assume a positive relationship between market access and local roads,[25] the linkages between local roads and indicators of structural transformation are not trivial. When fixing a district's connectivity with respect to relevant economic activity in the rest of the country, it is not clear if economic forces within the district, activated by improvements of the local road network, are sufficient to trigger labor reallocation.[26] On the other hand, after controlling for local roads, the coefficient of the Market Access variable identifies the relationship between within-district measures of structural transformation and any change in the Ethiopian road network that (i) affects district-level connectivity with relevant economic activity in the country, and (ii) is in principle orthogonal to the mere expansion of the road network within the district.

[24]Our strategy based on partialling out follows Donaldson and Hornbeck (2016). However, its rationale is also fully consistent with other works (see, for instance, Storeygard and Jedwab, 2020) that build a measure of market access without factoring in improvements in local roads and use this measure as an instrument for overall market access. The literature on the effects of transport infrastructure has advanced other solutions to address endogeneity concerns, including identification strategies relying on time-invariant instruments, such as historical or planned infrastructural networks (see Redding and Turner, 2015; Redding, 2020, for a review). However, those strategies are more likely to capture variation in the location of infrastructure, rather than the evolution of investment over time (Storeygard and Jedwab, 2020).

[25]The empirical correlation between market access and local roads is positive and statistically significant. In a simple univariate linear regression of market access on local roads, the estimated coefficient is equal to 0.076 with a robust standard error of 0.004. When the same relationship is estimated including district fixed effects and region-specific time trends as in (2.1), we get an estimate of 0.022 with a standard error of 0.004.

[26]Existing work looking at the role of local roads has mostly done so at the level of individual towns or villages (e.g., Asher and Novosad, 2020).

3 Results

3.1 Jobs and Structural Transformation

We start by introducing a set of results linking market access to the number and sectoral composition of jobs in Ethiopian districts. The dependent variable y_{it} is, in turn, the share of total jobs in the working-age population in district i at time t and the share of jobs in each of the main sectors of the economy over the number total jobs. For each dependent variable, Table 2 reports estimates from two specifications. The first is the baseline regression featuring market access on the right-hand side, in addition to district and region–year fixed effects (equation 2.1). The second is the same specification augmented with a measure of road segment improvements within the district (Local Roads), as discussed in Section 2. Standard errors clustered at the region level are reported in parentheses below each estimated coefficient while p-values based on wild cluster bootstrap standard errors are reported at the end of the table.

There is a small difference in both the size and the precision of the estimated coefficients for market access when comparing the two models. This suggests that what matter in the relation between market access and the outcomes of interest is not district-specific changes in the length and speed allowed on local roads but changes at the country-level road network that increase the district's connectivity to economic activity in the rest of the country. As for local roads, the Wild p-values reported under columns 2, 4, 6 and 8 in Table 2 suggest that improvements in the local road network that are orthogonal to changes in the district's connectivity have little implications for structural transformation[27].

Table 2: Roads and jobs

Outcome var.	Jobs		Agriculture		Manufacturing		Services	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Market Access	0.008	0.012	-0.046	-0.049	0.007	0.008	0.033	0.035
	(0.012)	(0.013)	(0.006)	(0.004)	(0.003)	(0.003)	(0.004)	(0.005)
Local Roads		-0.002		0.001		-0.000		-0.001
		(0.001)		(0.002)		(0.000)		(0.001)
Observations	1,573	1,573	1,573	1,573	1,573	1,573	1,573	1,573
Wild p-value Market Access	0.578	0.438	0.0117	0.0127	0.348	0.238	0.0107	0.00293
Wild p-value Local Roads		0.203		0.496		0.297		0.604
Mean DV	0.808	0.808	0.790	0.790	0.0409	0.0409	0.154	0.154
Quantification	0.00828	0.0132	-0.0496	-0.0532	0.00722	0.00858	0.0356	0.0382

Notes: The dependent variables measure, respectively, the ratio between the number of jobs and the working-age population in each district (Jobs); the share of agricultural workers on total workers (Agriculture); the share of manufacturing workers on total workers (Manufacturing); the share of service workers on total workers (Services). All regressions include district fixed effects and region-specific time trends. All regressions are also weighted by the size of the district population. Standard errors clustered at the region level are reported in parentheses below each estimated coefficient. Wild p-values indicate the p-value for wild cluster bootstrap standard errors at the region level. Mean DV is the sample mean of the dependent variable. Quantification reports the change in the dependent variable associated with an increase in Market Access of 1 sample standard deviation.

The lack of statistical significance for the estimated coefficients reported in columns (1) and (2) of Table 2 suggest that job creation is not correlated with within-district improvement in market access over time. On the other hand, we find some evidence of correlations between market access

[27] The results presented in Table 2 remain robust when removing the population weights, and working on a balanced sample of districts (i.e. excluding those changing borders or denomination over the sample period). The estimates derived in these robustness tests are not reported for reasons of space, but are available upon request.

and structural transformation. Indeed, there is evidence of lower shares of agricultural workers in districts reporting increases in market access. The decrease in the share of agricultural jobs seems occur in relation to an increase in the services sector, rather than manufacturing.[28] This pattern is not uncommon in low-income countries and echoes existing evidence on the direction of structural change, which shows a reallocation of workers out of agriculture towards services.

The estimated coefficients are sizeable. According to the estimates in columns (4) and (8), an improvement of market access by 1 sample standard deviation correlates with a reduction of about 5 percentage points (p.p.) in the share of agricultural workers and an increase of about 4 p.p. in the share of workers employed in the services sector. These numbers are economically significant as they represent 7% and 25% of the sample average shares of agricultural and service workers, respectively. Note that a 1 standard deviation change in market access corresponds to moving from the sample median of market access to its 80$^{\text{th}}$ percentile, i.e., from the starting sample value of Ganta Afeshum in the Tigray region and almost at the border with Eritrea, to the end sample value of Kalo, a district of the Amhara region in the middle of the motorway connecting Addis Ababa to Mekele (the capital of Tigray).

Industry heterogeneity. To unpack the baseline results presented above, we further ask the question of which specific service sectors are most affected by roads. Table 3 provides a summary of regression estimates after grouping relevant subsets of 2-digit industries. We find that improvements in market access mainly correlate with higher shares of workers in private services, a group of services including trade-related (wholesale and retail) activities, financial and business services. Conversely, we do not find evidence of road-driven improvements in industries that can be directly connected to road construction. This is also relevant for identification purposes. It shows that our results are not mechanically driven by an increase in jobs in the construction sector itself or in related services. An important issue in our context is the fact that investment in infrastructure is often accompanied by additional public services (e.g., maintenance, security, provision of utilities) that can create new jobs directly linked to the infrastructure investment. Provided that these services belong to the public sector, the statistically insignificant coefficient in column 2 of Table 3 shows that public services are not driving the baseline results in our sample. This confirms that the pattern of structural change captured in our results reflects improvements in access to markets rather than the increase of non-market services provided by the government.

3.2 Heterogeneity

Gender. We test for potential heterogeneity in our main results when distinguishing individuals by gender. The estimates reported in Table A.3 of the Appendix show that the patterns described

[28] Working with the shares of sectoral employment does not allow us to draw an unequivocal conclusion about the nature of the reallocation we observe in the data. Two additional tests that we have run can help to understand the process. First, we have replicated the analysis using absolute values (i.e., using the total number of jobs by district). The results of this exercise (not included but available upon request) are not conclusive either. The direction of all sectoral coefficients is consistent with the main findings, but only the manufacturing sector displays a significant (positive) value. Second, when disaggregating by age cohorts, we show that most of the changes we observe can be explained by young people entering the job market. The latter finding is discussed in more detail in the next subsection.

Table 3: Unpacking the services sector

Dep var.	Private	Public	Utilities	Construction	Others
	(1)	(2)	(3)	(4)	(5)
Market Access	0.013	0.000	0.001	0.005	0.010
	(0.001)	(0.002)	(0.001)	(0.004)	(0.006)
Local Roads	-0.000	0.001	0.000	-0.000	-0.001
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
Observations	1,573	1,573	1,573	1,573	1,573
Wild p-value Market Access	0.0312	0.848	0.299	0.213	0.467
Wild p-value Local Roads	0.359	0.0801	0.368	0.576	0.463
Mean DV	0.0585	0.0265	0.00231	0.0131	0.0233

Notes: The dependent variables measure, respectively, the share of workers employed in private services (Private), in public services (Public), in the utilities sector (Utilities), in construction (Construction), and in other services (Others). Private services include the following industries: trade, financial, real estate, and transport. Public services include the following industries: public administration, education, and health. Others is a residual category including personal services. All regressions include district fixed effects and region-specific time trends. All regressions are also weighted by the district population size. Standard errors clustered at the region level are reported in parentheses below each estimated coefficient. Wild p-values indicate the p-value for wild cluster bootstrap standard errors at the region level. Mean DV is the sample mean of the dependent variable.

in the previous section are largely confirmed across genders. While men are more likely to leave agricultural jobs and join both services and manufacturing, the coefficient of services denotes a larger propensity of women to be driven into modern activities following changes in market access. The coefficient of market access is statistically different across the two specifications for the gender-specifc share of services jobs.[29] Evidence from developing countries shows that women face greater difficulties in the labor market compared to men and are disproportionately affected by infrastructural bottlenecks. Improving connectivity can reduce some of these constraints, saving time spent in unpaid activities and enabling opportunities beyond the local community (Lei et al., 2019).

Education. Increases in market access can shape educational investment decisions. They may increase the number of people returning to education on the one hand, while increasing the opportunity cost of schooling on the other (Adukia et al., 2020). In our empirical framework, we explore whether and how changes in market access affect educational choices. Educational indicators cover information on the highest grade completed and the current grade attended. The number of individuals with some level of education has been growing over time. However, only a very small fraction of individuals report an education level higher than primary school (grades 1 to 8). We run two different exercises with education data. The first measures changes in the share of employed individuals with different levels of education. The estimates are presented in Table A.4 and show that participation in the labor market by better-educated workers is significantly increasing in areas with improving market access over the sample period.[30] The

[29]To test whether the estimated coefficients are statistically different across gender-specific specifications, we have appended the data for female- and male-specific versions of each dependent variable. Then, we have estimated our baseline model augmented with (i) a gender indicator for the dependent variable, and (ii) an interaction with that indicator and any other term on the right hand side of the model. The coefficient for the interaction between the gender indicator and market access is statistically different from 0 (p-value equal to 0.032) in the specification for the share of services sectors in total jobs.

[30]Though our data do not allow further inferences, these results suggest a higher return to education in areas with better connections and are consistent with the recent work by Adukia et al. (2020) linking investment in roads to educational outcomes in rural India.

second exercise analyzes specific cohorts of individuals, namely those that are of school age, i.e., from 7 to 18.[31] By carrying out this analysis, we find some evidence of a positive correlation between higher levels of education for children in districts experiencing greater market access (see Appendix Table A.5).

Age cohorts. Related to the previous exercise, another important source of heterogeneity is the demographic composition of the working-age population. Since a large share of the working population is young, it is important to understand whether economic opportunities are most likely to involve young workers or not. Hence, we spilt our sample and replicate our baseline analysis for the following three age cohorts: 15–19, 20–39 and 40–65.[32] Results, reported in Table A.6, show that some of the change is indeed occurring in the younger cohorts of workers. The youngest cohort, in particular, is also likely to experience increases in their employment rate (though starting from lower levels) in relation to an increase in market access.

4 Robustness

Growth hubs. As noted by Faber (2014) and Storeygard and Jedwab (2020), growth in economic hubs might drive the location of road placement and implementation. Hence, we run some robustness checks in which we remove potential growth hubs from the sample. We first exclude the woredas in the Addis Ababa special administrative division (six in total). Next, we also exclude the districts where regional capitals are located.[33] Finally, we replicate our estimates excluding all districts belonging to the Tigray region, which hosts the majority of the Tigrayan ethnic group that was in political power until 2018. This test is motivated by the political economy argument, according to which co-ethnicity can drive public investment choices (Burgess et al., 2015). Results, summarized in Table A.7 in the Appendix, show that our main findings hold across all of these altered datasets.

Alternative measure of market access. We explore whether our results are robust to an alternative measure of market access. Specifically, following the discussion in Section 1, we experiment with different values of trade elasticity (θ). Following the existing empirical evidence, we use the three alternative values of θ: (1) a value of 1, as originally proposed by Harris (Harris, 1954); (2) a value of 1.5, which was adopted by Gebresilasse (2020) in his work measuring the effects of market access under the RSDP and the URRAP programs in rural Ethiopia; (3) a value of 8.22, which was used by Donaldson and Hornbeck (2016) in their work on railroads in the US. Finally, we also report results based on a definition of market access that replaces nightlights with population as an indicator of economic activity in destination markets[34]. The

[31] Since 1994, Ethiopia has had a 8-2-2 formal education structure. Primary school has an official entry age of seven and a duration of eight grades. Secondary school is divided into two cycles: lower and upper, which consist of grades 9–12. Students sit the Primary School Certificate Examination at the end of grade 8, the General Secondary Education Certificate Examination at the end of grade 10, and the Higher Education Entrance Certificate Examination at the end of grade 12 (the source of this information is UNESCO).

[32] The three age groups represent, respectively, 26.3%, 50.1% and 23.6% of the population in our sample.

[33] The capital of the Oromia region was moved to Addis Ababa in 2005. Still, for the purpose of this exercise we use the old capital, Adema. As in the case of Addis Ababa, we also include all districts (2 in total) belonging to the special administrative zone of Dire Dawa.

[34] Data on population is taken from 1994 census.

results are reported in Table A.8 of the Appendix. While the magnitude and precision of the estimated coefficients change, the results do not provide significant evidence that contradicts the qualitative pattern suggested by our baseline analysis.

Alternative definition of local roads. In the spirit of Donaldson and Hornbeck (2016), controlling for changes to local roads allows us to isolate the variation in market access that is orthogonal to investments in each district. However, our main definition of local roads—given by the total (speed-weighted) length of all roads within the borders of the district—may not capture investments developed in nearby areas outside of the district's administrative borders that still reflect district-specific incentives or are undertaken in expectation of a district-specific payoff. A recent paper by Storeygard and Jedwab (2020) argues that choosing closer rather than farther buffers will introduce a trade-off in terms of excludability versus strength in the identification strategy. To address this issue, we augment our baseline specification by controlling for improvements in all roads, as captured by the same variable computed considering the district area extended by buffers of 10 to 50 km. We replicate our baseline estimates by adding the resulting controls in the same regression including our main measures of market access and local roads. Table A.9 reports one of these estimates, including a buffer of 50 km, which largely confirms the baseline patterns discussed above.

Different clustering of the standard errors. In this section, we check whether the results survive different clustering strategies of the standard errors. These include clustering standard errors at the district level (i.e., the level of treatment) or using heteroskedastic robust methods. In addition, we check whether the possible presence of spatial correlation in the residuals can affect the results. To do this, we estimate our model by introducing a spatial HAC correction of standard errors based on the Conley method, using the code proposed by Hsiang et al. (2011). We impose no constraints on the temporal decay of the weights and test the robustness of our specification to different lengths of the radius (respectively, from 100 to 500 km) for the spatial kernel. Table A.10 in the Appendix reports the results. For the Conley method, we only report results based on a 150 km cutoff, but standard errors are generally smaller when considering greater distances (especially in the specifications for services) and our results remain qualitatively stable.[35]

Omitted variables and reverse causation. Although the nature of our results is mostly descriptive, a potential issue of concern is the bias in the estimated coefficient that can occur. This may be due to either the omission of time-invariant variables at the district level affecting the relationship between roads and labor market outcomes and/or the reverse causation among the two. To address the former issue, we run a specification that includes several controls. One is nightlight intensity, which is a commonly adopted proxy for the level of economic activity at the subnational level. We also account for the number of conflicts occurring in each district on a yearly basis. Finally, we include information on weather conditions, namely the level of yearly precipitation (in millimeters) and the average monthly air temperature (in degrees Celsius)[36].

[35] Results of the additional specifications based on different cutoffs are not included due to space considerations but are available upon request to the authors.

[36] Data on conflicts and weather is provided by Aiddata GeoQuery. Conflict data is originally sourced from ACLED Conflict Events; data on precipitation is from the UDel Precipitation dataset (v.5.01); data on tempera-

Results, reported in Table A.11, show that the overall findings remain largely unaffected by the inclusion of these controls.

We then run some exercises to understand to what extent the presence of pre-trends, or the influence of initial conditions, may affect our findings. First, we estimate the relationship between the overall sample change in our treatment and baseline values of the outcome variables. More precisely, we run the following regression:

$$\Delta \text{ Market Access}_i = \boldsymbol{\beta}'\mathbf{X}_i + \phi_i + \epsilon_i, \tag{4.1}$$

where the dependent variable is the change in market access from 1996 (i.e., 1 year before the beginning of the RSDP) to 2013 and $\mathbf{X}_i$ is a vector of initial characteristics of district i. These include our main outcome variables, i.e., initial employment and the shares of agriculture, manufacturing, and services on total jobs. Initial characteristics are computed using information included in the 1994 census. We also run an alternative version of the previous equation in which the delta of market access is computed comparing each period t with the previous, and regress this value on employment and sectoral shares measured in $t-1$. Estimates of both exercises are reported in panels A and B of Table A.12 in the Appendix and show no evidence that the main outcomes of interest are driving subsequent investment in road infrastructure.

Second, to better deal with pre-trends, we run an exercise in which interaction terms between time trends and initial values of the outcome variables are included as additional regressors. This should help alleviate the concern that districts with, for instance, high initial agricultural employment prior to the RSDP may experience differential structural transformation trajectories over the period of the program.[37] Results, reported in Table A.13, show that the inclusion of initial values interacted with time dummies does not alter the size or the direction of the findings.

5 Mechanisms

Our main results provide a detailed characterization of the role of infrastructure reforms in the process of structural transformation in Ethiopia. In particular, we find a correlation between improvements in market access and an employment transition from agriculture towards services (especially for women). We also find some evidence of higher investment in education in more intensively treated locations. In this section, we try to harmonize these pieces of evidence by testing some of the potential underpinning mechanisms. There are several channels through which improvements in market access can affect changes in the composition of the labor force: directly, by lowering transport costs and therefore reducing some of the typical frictions affecting labor mobility and internal migration in developing countries, and indirectly, by levelling the playing field through increased economic opportunities and competition in the treated locations.

ture is from the UDel Air Temperature dataset (v.5.01).

[37] This is a commonly used strategy in papers examining the implications of trade liberalization (e.g., Dai et al., 2021).

5.1 Internal migration

Migration, especially from rural to urban areas, is related to changes in economic development and normally fuels the processes of structural transformation and urbanization in developing countries (Gollin and Rogerson, 2014; Storeygard and Jedwab, 2020). However, high transportation costs and a lack of economic opportunities might hamper migration (Lagakos, 2020). Hence, we ask whether improvements in market access, by driving down transportation costs and increasing economic opportunities at destinations, can facilitate the movement of workers towards areas with better economic prospects. While this is intuitive, there is very little empirical evidence supporting this relationship. An exception is the recent work by Morten and Oliveira (2017) showing how improved market access had large role on abating migration costs in Brazil.

We investigate this potential channel in two ways. First, we check whether improvements in roads are conducive to more migration in treated districts. Information available in both the population census and the NLF data allows us to track past changes in the respondents' places of residence. An individual is classified as a (internal) migrant in the year of the survey if their birthplace is different from the place where they currently reside. For this exercise, we can only track migrants at the destination and not at the origin, as information on the former is only available at a more aggregated geographic area (the zone). While migration could be motivated by several reasons, the search for work opportunities is—according to the qualitative information provided by the NLF surveys—the main one, and it was also on the rise over the period examined (see Table A.14 in the Appendix). Results in Table A.15 in the Appendix seem to confirm that locations with higher levels of market access are likely to attract a larger number of migrants, irrespective of whether they work or not. In columns (2) and (3), we split migration according to whether their location within a given district is urban or rural. The results are statistically significant only for urban migration, confirming that improvements in roads are more likely to make urban locations more attractive.

Second, we replicate our main specification replacing total workers with migrant workers. This shows whether changes in the labor market outcomes that are correlated with improvements in market access reflect the increase in the relative share of migrant workers. Results are reported in Table 4 and show that migrant workers are (a) more likely to be employed following improvements in market access and (b) more likely to be engaged in non-agricultural activities, especially in manufacturing and services.

5.2 Economic Opportunities

We now turn to the channel that explains the role of market access for structural transformation through its ability to shape economic opportunities and labor demand. To frame this analysis, we match the road data with firm-level datasets covering manufacturing sectors and trade-related services.[38]

Manufacturing firms. Data on manufacturing firms come from two sources. The first is the annual census of large and medium manufacturing establishments, published by the CSA. Man-

[38]Table A.16 reports the summary statistics on all the variables used in the firm level analysis.

Table 4: Migrant Workers

Variables	Employment	Agriculture	Manufacturing	Construction	Services
	(1)	(2)	(3)	(4)	(5)
Market Access	0.015	-0.004	0.003	0.001	0.015
	(0.009)	(0.004)	(0.002)	(0.002)	(0.005)
Local Road	0.001	0.002	-0.000	-0.000	-0.000
	(0.002)	(0.001)	(0.000)	(0.000)	(0.001)
Observations	1,573	1,573	1,573	1,573	1,573
Wild p-value Market Access	0.00781	0.455	0.00391	0.676	0.0371
Wild p-value Local Roads	0.398	0.0107	0.453	0.424	0.961
Mean DV	0.128	0.0751	0.00875	0.00340	0.0445

Notes: The dependent variables measure, respectively, the share of migrants in the total work force (Employment); the share of migrants working in the agricultural sector (Agriculture); the share of migrants working in the manufacturing sector (Manufacturing); the share of migrants working in the construction sector (Construction); the share of migrants working in the services sector (Services). All regressions are weighted by the size of the district population. Standard errors not corrected for the wild cluster procedure are reported in parentheses below each coefficient. The p-values for wild cluster bootstrap standard errors at the region level are reported at the end of the table.

ufacturing industries are defined at the 4-digit level according to the ISIC Rev. 3 classification. Data cover all formal firms with at least 10 persons employed and that use electricity in their production process.[39] These firms are required to respond to this census every year; therefore, it reports on all large and medium firms in the manufacturing sector. The census records provide information on the characteristics of each establishment, as well as detailed information on the size and composition (including by skills and gender) of the workforce and on the location of each firm.[40] Our data cover yearly information over the 1998–2009 period, ending a few years earlier than the analysis conducted so far.

The second dataset is the survey of small-scale manufacturing industries (SSIS). We combine all existing waves of the SSIS, covering the years 2002, 2006, 2008, 2010 and 2014. This is a survey that covers small (i.e., those employing less than 10 persons) and informal firms in the manufacturing sector. The sample is single-stage stratified, considering six main industries (textiles and garments, metal work, wood work, leather and leather products, other manufacturing sector and grain mills industry), sampled in similar proportions across regions. Due to the lack of a proper sample frame, it is not necessarily representative of the sector but provides considerable information on the activities of smaller firms, which comprise the majority of firms in the country. Over 95% of the firms surveyed in the different waves of the SSIS do not keep a book of accounts (or declare it incomplete) and are hence informal. On average and consistently over time, small and informal firms represent the large majority of all manufacturing establishments, approximately half of total manufacturing employment, and about one third of the value added produced in the sector. Table 5 reports precise figures for the two years in which the SSIS and the census were run simultaneously. Information included in the SSIS is based on a similar questionnaire to the census of larger firms, allowing for the comparison of some outcomes.

We start with the census data by showing whether improvements in market access at firm locations are related to the dynamism of the private sector and several dimensions of firm performance. In all regressions, we control for firm and region–year fixed effects as well as for a measure

[39]The number of persons refers to employees as well as working owners.

[40]Note that information on the location of manufacturing firms is slightly more precise than that used in this paper, so we have computed market access at the level of the (urban) town, rather than at the district level.

Table 5: Share of informal manufacturing sector

Year	% of firms	% of employment	% of value added
2006	97.14%	51.42%	38.77%
2008	96.74%	59.64%	31.06%

Source: Authors' elaboration on CSA data. *Notes:* All values represent the share of informal manufacturing firms on the total values. The latter is given by the sum of informal and formal firms' annual totals. Information on informal firms is calculated using sample weights provided by CSA. We report the information only for the two years in which the SSIS and the firm census were run simultaneously. Value added is computed as the total value of production minus production costs.

of local roads. Results are presented in Table A.17 of the Appendix. In the first two columns, we collapse the data at the location–year level and show that while improving road infrastructure is not correlated with firm entry, it correlates with the entry of foreign firms. It has been shown that foreign investors normally generate higher quality jobs, pay higher salaries and create more links with local firms (including in the services sector).[41] Moving to the firm-specific changes, those in locations that have improved market access are found to have experienced gains in a few dimensions, including an increase in (labor) productivity. Moreover, we find some (though statistically weak) evidence of a compositional shift towards non-production workers, whose real per capita wages are also positively correlated with increases in market access.

Next, we replicate the previous set of exercises using data on informal manufacturing firms. This is particularly relevant since informal firms account for a large majority of all manufacturing firms, as well as for a (slight) majority of the sector's employment (see Table 5). Due to the lack of a panel dimension in the SSIS data, we run our analysis using district, industry (at the four-digit level of the ISIC classification) and region–year fixed effects to understand how aggregate and average firm indicators have changed over time within the same district under differential changes in market access. The results, reported in Table A.18 in the Appendix, are to some extent similar to those obtained by looking at formal firms. These findings suggest that improvements in market access are related to improvements in informal firm productivity. On the other hand, there is no clear evidence regarding the composition and wages of workers.

Services firms. No equivalent information is available for the service sectors. Our analysis of services firms is therefore based on the Ethiopian Distributive Trade Survey (DST), which covers firms in trade-related services (i.e., retailers and wholesalers) and is available in a cross-sectional setting for the years 2003, 2009 and 2011. Moreover, this survey only covers urban areas and, therefore, is not representative at the national level.[42] However, it includes information on the districts in which firms are located, as well as other basic information about their activity such as size, sales, capital and wages of employees.

Based on these data, we conduct a similar analysis to the one discussed for manufacturing firms. Estimates presented in Table A.19 show that service firms in districts experiencing improvements in market access do experience increases in their average size. There is no evidence of corresponding changes in wages or labor productivity. This suggests an increase in employment

[41] Looking at the local impact of FDI in Ethiopia, recent work by Abebe et al. (2019) provides sound evidence that the entry of FDI generates high spillovers on domestic firms and workers.

[42] Urban areas covered in the survey correspond to fifteen major urban centers (regional capitals and other major towns) and 106 towns.

in private service activities and is therefore consistent with the findings reported in Table 3.

6 Conclusions

In this paper, we have studied the relation between road infrastructure development on the size and composition of jobs in Ethiopian districts. We have taken advantage of novel geocoded information covering the Ethiopian road network, which we combined with individual information from population censuses and nationally representative labor force surveys. Our analysis has generated the following results. Higher market access at the district level due to road development is related to a process of structural transformation characterized by a reduction in the share of agricultural workers in favor of services. We do not find evidence of improvements in market access being correlated to more jobs in the manufacturing sector. We also show that such changes are most likely to benefit women and younger individuals, and that better road infrastructure brings about a potential upgrading of the labor force through higher participation in education. We highlight our results by showing that higher economic activity induced by road investments stimulates both the demand from firms, through increases in their size and productivity, and supply from workers, who are more likely to migrate towards areas in the country with greater market access.

Overall, our results show that investments in road infrastructure under the RSDP can support the process of job creation and structural transformation since they can contribute to reducing some of the typical frictions affecting the labor market in Ethiopia. Yet, the weak role of roads in increasing the manufacturing sector's capacity to generate jobs is concerning, especially in view of the country's high political focus on industrialization. However, it is important to highlight that our analysis does not cover the most recent years, when most of the industrial parks (e.g., the one established in Hawasa) have started large-scale activities and for which it is well known that infrastructure investments are key. Further research is needed to investigate the effects of infrastructure in more recent periods and to gain a better understanding of the relationship between market access and structural transformation in low–income countries.

References

Abadie, Alberto, Susan Athey, Guido W. Imbens, and Jeffrey Wooldridge, "When should you adjust standard errors for clustering?," NBER WP N. 24003 2017.

Abebe, Girum, Margaret McMillan, and Michel Serafinelli, "Foreign Direct Investment and Knowledge Diffusion in Poor Locations: Evidence from Ethiopia," mimeo 2019.

Adam, Christopher, David Bevan, and Douglas Gollin, "Rural–Urban Linkages, Public Investment and Transport Costs: The Case of Tanzania," *World Development*, 2018, *109*, 497–510.

Adamopoulus, Tasso, "Spatial Integration and Agricultural Productivity: Quantifying the Impact of New Roads," mimeo 2020.

Adukia, Anjali, Samuel Asher, and Paul Novosad, "Educational Investment Responses to Economic Opportunity: Evidence from Indian Road Construction," *American Economic Journal: Applied*, 2020, *12* (1), 348–376.

Aggarwal, Shilpa, "Do Rural Roads Create Pathways Out of Poverty? Evidence from India," *Journal of Development Economics*, 2018, *133*, 375–395.

Alder, Simon, "Chinese Roads in India: The Effect of Transport Infrastructure on Economic Development," mimeo 2019.

Ali, Seid Nuru, "Infrastructure and economic transformation in Ethiopia," in "The Oxford Handbook of the Ethiopian Economy," Oxford University Press, 2019.

Asher, Samuel and Paul Novosad, "Rural Roads and Local Economic Development," *American Economic Review*, 2020, *110* (3), 797–823.

_ **, Tobias Lunt, Ryu Matsuura, and Paul Novosad**, "Development Research at High Geographic Resolution: An Analysis of Night-Lights, Firms, and Poverty in India Using the SHRUG Open Data Platform," *The World Bank Economic Review*, 2021, *https://doi.org/10.1093/wber/lhab003.*

Atkin, David and Dave Donaldson, "Who's Getting Globalized? The Size and Implications of Intra-national Trade Costs," MIT, mimeo July 2015.

Baccini, Leonardo, Matteo Fiorini, Bernard Hoekman, and Marco Sanfilippo, "Services and Economic Development in Africa," IGC Working Paper N. F-20086-NOC-1 2021.

Baum-Snow, Nathaniel, Vernon Henderson, Matthew A. Turner, Quinghua Zhang, and Loren Brandt, "Does investment in national highways help or hurt hinterland city growth?," *Journal of Urban Economics*, 2018, *https://doi.org/10.1016/j.jue.2018.05.001.*

Bryan, Gharad and Melania Morten, "The aggregate productivity effects of internal migration: Evidence from Indonesia," *Journal of Political Economy*, 2018.

Burgess, Robin, Remi Jedwab, Edward Miguel, Ameet Morjaria, and Gerard Padro i Miquel, "The Value of Democracy: Evidence from Road Building in Kenya," *American Economic Review*, 2015, *105* (5), 1817–1851.

Bustos, Paula, Bruno Caprettini, and Ponticelli Jacopo, "Agricultural Productivity and Structural Transformation. Evidence from Brazil," *American Economic Review*, 2016, *106* (6), 1320–1365.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller, "Bootstrap-Based Improvements for Inference with Clustered Errors," *The Review of Economics and Statistics*, 2008, *90* (3), 414–427.

Chiovelli, Giorgio, Stelios Michalopoulos, and Elias Papaioannou, "Landmines and Spatial Development," mimeo 2019.

CSA, "Analytical Report of the National Labour Force Survey," Statistical Bullettin of the Central Statistical Authority 2004.

_ , "Analytical Report of the National Labour Force Survey," Statistical Bullettin of the Central Statistical Authority 2006.

_ , "Analytical Report of the National Labour Force Survey," Statistical Bullettin of the Central Statistical Authority 2014.

Dai, Mi, Wei Huang, and Yifan Zhang, "How do households adjust to tariff liberalization? Evidence from China's WTO accession," *Journal of Development Economics*, 2021, *150*, 102628.

Dercon, Stefan, Daniel O. Gilligan, John Hoddinott, and Tassew Wodehanna, "The Impact of Agricultural Extension and Roads on Poverty and Consumption Growth in Fifteen Ethiopian Villages," *American Journal of Agricultural Economics*, 2009, *91* (4), 1007–1021.

Diao, Xinshen, Mia Ellis, Margaret McMillan, and Dani Rodrik, "Africa's Manufacturing Puzzle: Evidence from Tanzanian and Ethiopian Firms," HKS Working Paper No. RWP21-018 2021.

Dinkelman, Taryn, "The Effects of Rural Electrification on Employment: New Evidence from South Africa," *American Economic Review*, December 2011, *101* (7), 3078–3108.

Donaldson, Dave and Richard Hornbeck, "Railroads and American Economic Growth: A "Market Access" Approach," *The Quarterly Journal of Economics*, 02 2016, *131* (2), 799–858.

Eberhard-Ruiz, Andreas and Alexander Moradi, "Regional Market Integration and City Growth in East Africa: Local but no Regional Effects?," *Journal of Development Economics*, 2019, *140*, 255–268.

Ethiopian Road Authority, "Road Sector Development Program 13 Years Performance and Phase IV," ERA 2011.

Faber, Benjamin, "Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System," *The Review of Economic Studies*, 2014, *81* (3), 1046–1070.

Fajgelbaum, Pablo and Stephen J. Redding, "Trade, Structural Transformation and Development: Evidence from Argentina 1869-1914," NBER Working Paper, 20217 October 2018.

Fiorini, Matteo, Marco Sanfilippo, and Asha Sudaram, "Trade Liberalization, Roads, and Firm Productivity," *Journal of Development Economics*, 2021, *153.*

Franklin, Simon, "Location, Search Costs and Youth Unemployment: Experimental Evidence from Transport Subsidies," *Economic Journal*, 2018, *128* (614), 2353–2379.

Gebresilasse, Mesay, "Rural Roads, Agricultural Extension, and Productivity," mimeo 2020.

Gibson, John, Susan Olivia, Geua Boe-Gibson, and Chao Li, "Which night lights data should we use in economics, and where?," *Journal of Development Economics*, 2021, *149.*

Gollin, Douglas and Richard Rogerson, "Productivity, transport costs and subsistence agriculture," *Journal of Development Economics*, 2014, *107* (2014), 38–48.

Gunning, Jan Willem, Pramila Krishnan, and Andualem T. Mengistu, "Fading Choice: Transport Costs and Variety in Consumer Goods," CSAE WPS/2018-05-2 2018.

Harris, Chauncy D., "The, Market as a Factor in the Localization of Industry in the United States," *Annals of the Association of American Geographers*, 1954, *44* (4), 315–348.

Hjort, Jonas and Jonas Poulsen, "The Arrival of Fast Internet and Employment in Africa," *American Economic Review*, 2019, *109* (3), 1032–1079.

Hsiang, Solomon, Kyle Meng, and Mark Cane., "Civil Conflicts Are Associated with the Global Climate," *Nature*, 2011, *476*, 438–41.

Jobs Creation Commission Ethiopia, "State of Jobs 2019," Federal Democratic Republic of Ethiopia, Addis Ababa 2019.

Khandker, Shahidur R., Zaid Bakht, and Gayatri B. Koolwal, "The Poverty Impact of Rural Roads: Evidence from Bangladesh," *Economic Development and Cultural Change*, 2009, *57* (4), 685–722.

Lagakos, David, "Urban-Rural Gaps in the Developing World: Does Internal Migration Offer Opportunities?," *Journal of Economic Perspectives*, 2020, *34* (3), 174–192.

_ **and Martin Shu**, "The Role of Micro Data in Understanding Structural Transformation," STEG Pathfining Paper 2021.

Lei, Lei, Sonalde Desai, and Reeve Vanneman, "Women's Non-agricultural Employment in India: The Conditioning Role of Community Gender Context," *Feminist Economics*, 2019, *DOI: 10.1080/13545701.2019.1655162.*

Moneke, Niclas, "Can Big Push Infrastructure Unlock Development? Evidence from Ethiopia," mimeo 2020.

Morten, Melanie and Jaqueline Oliveira, "The Effects of Roads on Trade and Migration: Evidence from a Planned Capital City," mimeo 2017.

Mu, Ren and Dominique van de Walle, "Rural Roads and Local Market Development in Vietnam," *The Journal of Development Studies*, 2011, *47* (5), 709–734.

NOAA National Geophysical Data Center, "Nighttime lights time series," Available at `https://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html` 2018.

Redding, Stephen J., "Trade and Geography," in "Handbook of International Economics," Routledge, 2020.

_ **and Matthew A. Turner**, "Transportation Costs and the Spatial Organization of Economic Activity," in "Handbook of Regional and Urban Economics," Elsevier, 2015.

Rodrik, Dani, "Premature Deindustrialization," *Journal of Economic Growth*, 2016, *21* (1), 1–33.

Roodman, David, Morten Ørregaard Nielsen, James G. MacKinnon, and Matthew D. Webb, "Fast and wild: Bootstrap inference in Stata using boottest," *The Stata Journal*, 2019, *19* (1), 4–60.

Shiferaw, Admasu, Måns Söderbom, Eyerusalem Siba, and Getnet Alemu, "Road Infrastructure and Enterprise Dynamics in Ethiopia," *The Journal of Development Studies*, 2015, *51* (11), 1541–1558.

Storeygard, Adam, "Farther on down the Road: Transport Costs, Trade and Urban Growth in Sub-Saharan Africa," *The Review of Economic Studies*, 04 2016, *83* (3), 1263–1295.

_ **and Remi Jedwab**, "The Average and Heterogeneous Effects of Transportation Investments: Evidence from sub-Saharan Africa 1960-2010," NBER WP N. 27670 2020.

Appendix

Table A.1: The ERA travel-speed matrix (km per hour)

Surface	Condition	
	Not rehabilitated	Rehabilitated or new
Asphalt	50	70
Major gravel	35	50
Minor gravel	25	45
Earth	20	30

Notes: The table reports average travel speed estimated by ERA as a function of the surface and condition of the road segment. Speed is measured in kilometers per hour.

Table A.2: Summary statistics

Variable	mean	sd	min	max	obs
Outcomes					
Empl. on pop.	.8076	.1273	.2698	1	1,573
Agr. share	.7901	.2156	0	1	1,573
Manuf. share	.0409	.0619	0	.6471	1,573
Services share	.1536	.1688	0	.9851	1,573
Regressors					
Roads	6.176	1.0791	3.163	10.940	1,573
Local roads	12.03	5.225	0	16.57	1,573

Table A.3: Roads and jobs by gender

Outcome var:	Male				Female			
	Jobs	Agri	Manuf	Ser	Jobs	Agri	Manuf	Ser
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Market Access	-0.010	-0.030	0.004	0.013	0.010	-0.020	0.003	0.022
	(0.013)	(0.015)	(0.001)	(0.002)	(0.013)	(0.013)	(0.003)	(0.004)
Local Roads	0.000	0.000	-0.000	0.000	-0.000	0.001	-0.000	-0.001
	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)
Observations	1,573	1,573	1,573	1,573	1,573	1,573	1,573	1,573
Wild p-value Market Access	0.455	0.0186	0.0137	0.0244	0.455	0.559	0.482	0.00195
Wild p-value Local Roads	0.551	0.672	0.393	0.783	0.551	0.604	0.344	0.363
Mean DV	0.541	0.455	0.0136	0.0607	0.459	0.335	0.0273	0.0929

Notes: The dependent variables measure, respectively, the ratio between the number of jobs and the working-age population in each district (Jobs); the share of agricultural workers on total workers (Agriculture); the share of manufacturing workers on total workers (Manufacturing); the share of service workers on total workers (Services). All regressions include district fixed effects and region-specific time trends. All regressions are also weighted by the size of the district population. Standard errors clustered at the region level are reported in parentheses below each estimated coefficient. Wild p-values indicate the p-value for wild cluster bootstrap standard errors at the region level. Mean DV is the sample mean of the dependent variable.

Table A.4: Roads and education, employed persons

Dep var:	Illiterate	Grade 1-8	Grade 9-12	Diploma	Degree
	(1)	(2)	(3)	(4)	(5)
Market Access	-0.025	0.005	0.004	0.003	0.003
	(0.008)	(0.001)	(0.001)	(0.001)	(0.001)
Local Roads	-0.001	0.000	0.000	0.000	-0.000
	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	1,573	1,573	1,573	1,573	1,573
Wild p-value Market Access	0.0879	0.0918	0.208	0.0625	0.0273
Wild p-value Local Roads	0.410	0.521	0.127	0.912	0.0488
Mean DV	0.648	0.0216	0.0136	0.00881	0.00368

Notes: The dependent variables measure, respectively, the share of illiterate individuals (Illiterate), the share of individuals who completed grades 1–8 (Grade 1–8), 9–12 (Grade 9–12), a diploma (Diploma) and a degree (Degree) in the working-age population. All regressions include district fixed effects and region-specific time trends. All regressions are also weighted by the size of the district population. Standard errors clustered at the region level are reported in parentheses below each estimated coefficient. Wild p-values indicate the p-value for wild cluster bootstrap standard errors at the region level. Mean DV is the sample mean of the dependent variable.

Table A.5: Roads and education, younger kids

Employment:	Illiterate	Grade 1-8	Grade 9-12
	(1)	(2)	(3)
Market Access	-0.007	0.003	0.007
	(0.008)	(0.005)	(0.001)
Local Roads	-0.003	0.003	0.000
	(0.000)	(0.001)	(0.000)
Observations	1,580	1,580	1,580
Wild p-value Market Access	0.366	0.639	0.0470
Wild p-value Local Roads	0.0107	0.00800	0.541
Mean DV	0.611	0.372	0.0192

Notes: The dependent variables measure, respectively, the share of illiterate individuals (Illiterate), the share of individuals who completed grades 1–8 (Grade 1–8) and 9–12 (Grade 9–12) on the population between 7 and 18 years of age. All regressions include district fixed effects and region-specific time trends. All regressions are also weighted by the size of the district population. Standard errors clustered at the region level are reported in parentheses below each estimated coefficient. Wild p-values indicate the p-value for wild cluster bootstrap standard errors at the region level. Mean DV is the sample mean of the dependent variable.

Table A.6: Roads and jobs by age cohort

Employment:	Jobs	Agriculture	Manufacturing	Services
	(1)	(2)	(3)	(4)
Panel A: 15-19 years old				
Market Access	0.029	-0.054	0.012	0.042
	(0.011)	(0.008)	(0.005)	(0.010)
Local Roads	-0.004	0.001	-0.001	-0.002
	(0.001)	(0.003)	(0.001)	(0.002)
Wild p-value Market Access	0.0391	0.00781	0.412	0.0146
Wild p-value Local Roads	0.0931	0.608	0.288	0.441
Mean DV	0.717	0.797	0.0367	0.154
Panel B: 20-39 years old				
Market Access	0.005	-0.052	0.007	0.038
	(0.014)	(0.005)	(0.004)	(0.005)
Local Roads	-0.001	0.001	-0.001	-0.001
	(0.001)	(0.002)	(0.000)	(0.001)
Wild p-value Market Access	0.732	0.0117	0.235	0.0176
Wild p-value Local Roads	0.0931	0.545	0.230	0.663
Mean DV	0.845	0.769	0.0435	0.170
Panel C: 40-65 years old				
Market Access	0.005	-0.034	0.005	0.023
	(0.016)	(0.007)	(0.003)	(0.008)
Local Roads	-0.001	0.002	-0.000	-0.001
	(0.001)	(0.001)	(0.000)	(0.001)
Wild p-value Market Access	0.780	0.0781	0.496	0.136
Wild p-value Local Roads	0.502	0.291	0.587	0.439
Mean DV	0.830	0.825	0.0400	0.120
Observations	1,573	1,573	1,573	1,573

Notes: The dependent variables measure, respectively, the ratio between the number of jobs and the working-age population in each district (Jobs); the share of agricultural workers on total workers (Agriculture); the share of manufacturing workers on total workers (Manufacturing); the share of service workers on total workers (Services). All regressions include district fixed effects and region-specific time trends. All regressions are also weighted by the size of the district population. Standard errors clustered at the region level are reported in parentheses below each estimated coefficient. Wild p-values indicate the p-value for wild cluster bootstrap standard errors at the region level. Mean DV is the sample mean of the dependent variable.

Table A.7: Robustness: Cuts to the data

Employment:	Jobs	Agriculture	Manufacturing	Services
	(1)	(2)	(3)	(4)
Panel A: Excluding Addis				
Market Access	0.012	-0.049	0.008	0.036
	(0.013)	(0.004)	(0.003)	(0.005)
Local Roads	-0.002	0.001	-0.000	-0.001
	(0.001)	(0.002)	(0.000)	(0.001)
Observations	1,547	1,547	1,547	1,547
Wild p-value Market Access	0.438	0.0117	0.242	0.00391
Wild p-value Local Roads	0.203	0.492	0.297	0.594
Panel B: Excluding Regional Capitals				
Market Access	0.011	-0.048	0.008	0.034
	(0.012)	(0.005)	(0.003)	(0.005)
Local Roads	-0.002	0.001	-0.000	-0.001
	(0.001)	(0.002)	(0.000)	(0.001)
Observations	1,546	1,546	1,546	1,546
Wild p-value Market Access	0.457	0.00977	0.234	0.00781
Wild p-value Local Roads	0.221	0.520	0.316	0.594
Panel C: Excluding Tigray				
Market Access	0.008	-0.050	0.009	0.038
	(0.013)	(0.005)	(0.003)	(0.005)
Local Roads	-0.002	0.002	-0.000	-0.001
	(0.001)	(0.002)	(0.000)	(0.001)
Observations	1,445	1,445	1,445	1,445
Wild p-value Market Access	0.578	0.0234	0.227	0.0195
Wild p-value Local Roads	0.305	0.367	0.312	0.461

Notes: The dependent variables measure, respectively, the ratio between the number of jobs and the working-age population in each district (Jobs); the share of agricultural workers on total workers (Agriculture); the share of manufacturing workers on total workers (Manufacturing); the share of service workers on total workers (Services). All regressions include district fixed effects and region-specific time trends. All regressions are also weighted by the size of the district population. Standard errors clustered at the region level are reported in parentheses below each estimated coefficient. Wild p-values indicate the p-value for wild cluster bootstrap standard errors at the region level. Mean DV is the sample mean of the dependent variable.

Table A.8: Robustness: Alternative definitions of Market Access

Employment:	Jobs	Agriculture	Manufacturing	Services
	(1)	(2)	(3)	(4)
Panel A: θ=1				
Market Access	0.047	-0.287	0.033	0.218
	(0.065)	(0.027)	(0.014)	(0.028)
Local Roads	-0.002	0.002	-0.000	-0.001
	(0.001)	(0.002)	(0.000)	(0.002)
Wild p-value Market Access	0.532	0.0170	0.309	0.0170
Wild p-value Local Roads	0.276	0.443	0.367	0.531
Panel B: θ=1.5				
Market Access	0.031	-0.165	0.021	0.123
	(0.040)	(0.015)	(0.007)	(0.016)
Local Roads	-0.002	0.001	-0.000	-0.001
	(0.001)	(0.002)	(0.000)	(0.001)
Wild p-value Market Access	0.470	0.0130	0.254	0.0170
Wild p-value Local Roads	0.246	0.450	0.356	0.551
Panel C: θ=8.22				
Market Access	0.002	-0.011	0.002	0.007
	(0.003)	(0.002)	(0.001)	(0.002)
Local Roads	-0.002	0.001	-0.000	-0.000
	(0.001)	(0.001)	(0.000)	(0.001)
Wild p-value Market Access	0.557	0.0150	0.509	0.000
Wild p-value Local Roads	0.264	0.648	0.346	0.720
Panel D: population weights				
Market Access	-0.006	-0.010	0.002	0.007
	(0.002)	(0.002)	(0.001)	(0.003)
Local Roads	-0.002	0.000	-0.000	0.000
	(0.001)	(0.002)	(0.000)	(0.001)
Wild p-value Market Access	0.323	0.0751	0.258	0.202
Wild p-value Local Roads	0.295	0.928	0.512	0.991
Observations	1,573	1,573	1,573	1,573

Notes: The dependent variables measure, respectively, the ratio between the number of jobs and the working-age population in each district (Jobs); the share of agricultural workers on total workers (Agriculture); the share of manufacturing workers on total workers (Manufacturing); the share of service workers on total workers (Services). All regressions include district fixed effects and region-specific time trends. All regressions are also weighted by the size of the district population. Standard errors clustered at the region level are reported in parentheses below each estimated coefficient. Wild p-values indicate the p-value for wild cluster bootstrap standard errors at the region level.

Table A.9: Robustness: Additional local roads

Employment:	Jobs	Agriculture	Manufacturing	Services
	(1)	(2)	(3)	(4)
Market Access	0.011	-0.048	0.006	0.036
	(0.014)	(0.008)	(0.004)	(0.007)
Local Roads	-0.002	0.001	-0.001	-0.001
	(0.001)	(0.002)	(0.000)	(0.001)
Observations	1,573	1,573	1,573	1,573
Year FE	YES	YES	YES	YES
Wild p-value Market Access	0.448	0.0180	0.498	0.0150
Wild p-value Local Roads	0.227	0.482	0.307	0.608

Notes: The dependent variables measure, respectively, the ratio between the number of jobs and the working-age population in each district (Jobs); the share of agricultural workers on total workers (Agriculture); the share of manufacturing workers on total workers (Manufacturing); the share of service workers on total workers (Services). All regressions include proxy measures of local road improvements occurring outside the district border (up to 50 km) as well as district fixed effects and region-specific time trends. All regressions are also weighted by the size of the district population. Standard errors clustered at the region level are reported in parentheses below each estimated coefficient. Wild p-values indicate the p-value for wild cluster bootstrap standard errors at the region level. Mean DV is the sample mean of the dependent variable.

Table A.10: Robustness: Alternative clustering of the SE

Employment:	Jobs	Agriculture	Manufacturing	Services
	(1)	(2)	(3)	(4)
Panel A: Robust SE				
Market Access	0.0122	-0.0493	0.00795	0.0354
	(0.0111)	(0.0173)	(0.00582)	(0.0138)
Local Roads	-0.00171	0.00127	-0.000474	-0.000882
	(0.00121)	(0.00130)	(0.000513)	(0.00102)
Panel B: District SE				
Market Access	0.0122	-0.0493	0.00795	0.0354
	(0.0113)	(0.0173)	(0.00522)	(0.0139)
Local Roads	-0.00171	0.00127	-0.000474	-0.000882
	(0.00126)	(0.00124)	(0.000464)	(0.000992)
Panel C: Conley SE				
Market Access	0.0122	-0.0493	0.00795	0.0354
	(0.0108)	(0.0150)	(0.00389)	(0.0120)
Local Roads	-0.00171	0.00127	-0.000474	-0.000882
	(0.000780)	(0.00114)	(0.000371)	(0.000900)
Observations	1,573	1,573	1,573	1,573

Notes: The dependent variables measure, respectively, the ratio between the number of jobs and the working-age population in each district (Jobs); the share of agricultural workers on the total (Agriculture); the share of manufacturing workers on the total (Manufacturing); the share of service workers on the total (Services). The regressor of interest (Roads) measures the log of market access. All regressions are weighted by the size of the district population. Panel A reports regressions employing heteroskedasticity-robust s.e.; Panel B reports regressions employing s.e. clustered at the district level; Panel C reports regressions using s.e. corrected for spatial autocorrelation, with a distance cutoff of 150 km. The latter regressions have been estimated using the user-written STATA command acreg (Colella et al., 2019).

Table A.11: Robustness: Adding controls

Employment:	Jobs	Agriculture	Manufacturing	Services
	(1)	(2)	(3)	(4)
Market Access	0.012	-0.049	0.008	0.036
	(0.011)	(0.005)	(0.003)	(0.005)
Local Roads	-0.002	0.001	-0.000	-0.001
	(0.001)	(0.002)	(0.000)	(0.001)
Nightlights intensity	0.001	-0.001	0.005	-0.006
	(0.001)	(0.003)	(0.001)	(0.002)
Number of conflicts	-0.017	-0.005	-0.003	0.007
	(0.007)	(0.012)	(0.001)	(0.013)
Precipitation	0.000	-0.000	-0.001	0.001
	(0.001)	(0.001)	(0.000)	(0.001)
Temperature	0.074	-0.019	0.022	-0.008
	(0.066)	(0.128)	(0.045)	(0.084)
Observations	1,538	1,538	1,538	1,538
Wild p-value Market Access	0.214	0.0137	0.252	0.00586
Wild p-value Local Roads	0.227	0.486	0.347	0.591

Notes: The dependent variables measure, respectively, the ratio between the number of jobs and the working-age population in each district (Jobs); the share of agricultural workers on total workers (Agriculture); the share of manufacturing workers on total workers (Manufacturing); the share of service workers on total workers (Services). All regressions include proxy measures of local road improvements occurring outside the district border (up to 50 km) as well as district fixed effects and region-specific time trends. All regressions are also weighted by the size of the district population. Standard errors clustered at the region level are reported in parentheses below each estimated coefficient. Wild p-values indicate the p-value for wild cluster bootstrap standard errors at the region level. Mean DV is the sample mean of the dependent variable.

Table A.12: Robustness: Initial conditions and changes in Market Access

Employment:	Jobs	Agriculture	Manufacturing	Services
	(1)	(2)	(3)	(4)
Panel A: Initial values				
X	-0.026	-0.149	0.034	0.261
	(0.938)	(0.523)	(0.961)	(0.527)
Observations	363	363	363	363
Wild p-value Market Access	0.938	0.523	0.961	0.527
Panel B: Lagged values				
X	0.366	0.139	-0.613	-0.042
	(0.168)	(0.234)	(0.105)	(0.770)
Observations	647	647	647	647
Wild p-value Market Access	0.168	0.234	0.105	0.770

Notes: The dependent variable in panel (a) measures changes in market access between 2013 and 1996; while in panel (b) it measures the changes in market access between any two consecutive periods. The independent variables are baseline levels (panel a) or lagged values (panel b) of the ratio between the number of jobs and the working-age population in each district (Jobs); the share of agriculture on total jobs (Agriculture); the share of manufacturing on total jobs (Manufacturing); the share of services on total jobs (Services). All regressions include region f.e. and are weighted by the district population. Standard errors clustered at the region level are reported in parentheses below each estimated coefficient. Wild p-values indicate the p-value for wild cluster bootstrap standard errors at the region level.

Table A.13: Robustness: Controlling for initial conditions

Employment:	Jobs	Agriculture	Manufacturing	Services
	(1)	(2)	(3)	(4)
Market Access	0.009	-0.055	0.007	0.041
	(0.011)	(0.012)	(0.003)	(0.011)
Local Roads	0.000	0.004	-0.001	-0.003
	(0.001)	(0.002)	(0.000)	(0.002)
Observations	1,269	1,269	1,269	1,269
Wild p-value Market Access	0.547	0.0234	0.283	0.00195
Wild p-value Local Roads	0.820	0.227	0.221	0.242

Notes: The dependent variables measure, respectively, the ratio between the number of jobs and the working-age population in each district (Jobs); the share of agricultural workers on total workers (Agriculture); the share of manufacturing workers on total workers (Manufacturing); the share of service workers on total workers (Services). All regressions include initial values of the dependent variables interacted for a time trend as well as district fixed effects and region-specific time trends. All regressions are also weighted by the size of the district population. Standard errors clustered at the region level are reported in parentheses below each estimated coefficient. Wild p-values indicate the p-value for wild cluster bootstrap standard errors at the region level. Mean DV is the sample mean of the dependent variable.

Table A.14: Reasons to migrate

1999		2005		2013	
Motivation	resp (%)	Motivation	resp (%)	Motivation	resp (%)
Search for work	25.22	Search for work	25.56	Search for work	35.55
Education	13.07	Education	24.33	To live with family	17.24
Job transfer	11.73	To live with family	10.24	Marriage arrangement	12.91

Source: Authors' elaboration on NLF data *Notes:* : The information reported in the table is drawn from responses to the NLF surveys. Persons declaring to be migrants also provide a motivation, which is coded by the CSA. Percentages reported in the table refer to individuals of working-age only.

Table A.15: Internal migration

VARIABLES	Migration	Urban migration	Rural migration
	(1)	(2)	(3)
Market Access	0.016	0.009	0.004
	(0.009)	(0.005)	(0.005)
Local Road	0.002	0.000	0.001
	(0.002)	(0.000)	(0.001)
Observations	1,573	1,573	1,573
Wild p-value Market Access	0.0215	0.0664	0.496
Wild p-value Local Roads	0.227	0.432	0.230
Mean DV	0.136	0.0324	0.0947

Notes: The dependent variables measure the share of migrants in the total population (Migrant), the share in urban areas (Urban migrant) and rural (Rural migrant) areas. The main control (Roads) measures the log of market access. All regressions are weighted by the size of the district population. All regressions are weighted by the district population. Standard errors clustered at the region level are reported in parentheses below each estimated coefficient. Wild p-values indicate the p-value for wild cluster bootstrap standard errors at the region level.

Table A.16: Summary statistics – Firm data

Variable	mean	sd	min	max	obs
Large manufacturing firms					
Entry	.1384	.2891	0	1	900
Foreign_entry	.0077	.0619	0	1	900
Productivity	10.99	.6679	3.5663	18.315	9,323
Empl	3.2946	.3718	0	8.4015	12,590
Non production emp	-.7439	1.024	-4.7274	3.7841	10,781
Wage	8.2978	.8638	.3142	12.5616	12,577
Wage non-prod	8.441	.9089	3.3928	12.7996	11,298
Wage prod.	8.2881	.7922	3.3928	12.8323	11,918
Small manufacturing firms					
Empl	.8404	.6942	-1.3863	2.9444	30,643
Productivity	10.775	1.272	2.524	19.591	17,988
Non production empl	.0374	.1127	0	1	30,642
Wage	7.328	2.992	0	12.5816	22,030
Wage prod	8.178	1.039	0	12.619	17,989
Wage non-prod	7.075	2.193	0	13.291	3,869
Services firms					
Employment	1.132	.7004	0	8.189	14,674
Wages	6.713	1.560	-2.485	12.577	4,772
Productivity	10.079	1.932	2.493	20.851	14,544

Table A.17: Manufacturing firms

VARIABLES	Entry	Foreign_entry	Productivity	Empl	Non production emp	Wage	Wage non-prod	Wage prod.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Market Access	0.00787	0.00266	0.163	0.0160	0.0583	0.00971	0.0792	0.00895
	(0.0155)	(0.00155)	(0.0510)	(0.0264)	(0.0334)	(0.0268)	(0.0393)	(0.0249)
Local Road 0.0726	-0.0206	-0.0822	-0.107	0.132	0.121	0.0274	0.174	
	(0.129)	(0.0137)	(0.115)	(0.0484)	(0.0677)	(0.0518)	(0.0875)	(0.0481)
Constant	-0.322	0.0914	9.232	3.296	-2.143	6.940	7.096	6.629
	(0.559)	(0.0595)	(0.980)	(0.463)	(0.543)	(0.467)	(0.708)	(0.452)
Observations	604	604	8,448	10,092	8,721	10,082	9,160	9,530
R-squared	0.537	0.406	0.698	0.907	0.630	0.793	0.738	0.689
Firm FE	NO	NO	YES	YES	YES	YES	YES	YES
Town FE	YES	YES	NO	NO	NO	NO	NO	NO
Region Year FE	YES	YES	YES	YES	YES	YES	YES	YES

Notes: The dependent variables measure, respectively, the entry rate, measured as the share of new firms at t on the total number of firms at $t-1$ in each town (Entry); the entry rate of foreign owned firms (Foreign_entry); firms' labour productivity, measured as value added on employment (Productivity); firms' (log of) total employment (Empl); Firms' share of non-production workers (Non-prod workers); the (log of) per capita wages for all employees (Wage), and for production (Wage prod.) and non-production workers (Wage non-prod.). All the regressions at firm level (columns 3-8) control for the (log) age of the firm. Standard errors are clustered at the town level and are reported in parenthesis below each coefficient.

Table A.18: Informal manufacturing firms

VARIABLES	Empl	Productivity	Non production empl	Wage	Wage prod	Wage non-prod
	(1)	(2)	(3)	(4)	(5)	(6)
Market Access	-0.0415	0.274	0.00211	-0.0847	0.0374	-0.395
	(0.0349)	(0.0929)	(0.00865)	(0.188)	(0.0513)	(0.373)
Local Roads	-0.000778	-0.0171	0.000555	-0.0460	-0.0103	0.0718
	(0.00579)	(0.0145)	(0.00103)	(0.0286)	(0.00710)	(0.0463)
Constant	1.018	8.840	0.0132	8.399	7.990	8.632
	(0.238)	(0.651)	(0.0572)	(1.333)	(0.341)	(2.552)
Observations	19,966	10,359	19,966	14,735	12,151	2,364
R-squared	0.238	0.286	0.173	0.219	0.412	0.372
District FE	YES	YES	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES	YES	YES
Region-Year FE	YES	YES	YES	YES	YES	YES

Notes: The dependent variables measure, respectively, firms' (log of) total employment (Empl); firms' labour productivity, measured as value added on employment (Productivity); Firms' share of non-production workers (Non-prod workers); the (log of) per capita wages for all employees (Wage), and for production (Wage prod.) and non-production workers (Wage non-prod.). All the regressions controls for firms' age. Standard errors are clustered at the district level and are reported in parenthesis below each coefficient.

Table A.19: Trade Services

VARIABLES	Employment	Wages	Productivity
	(1)	(2)	(3)
Market Access	0.611	1.762	-0.631
	(0.181)	(1.437)	(1.196)
Local Roads	-0.0126	-5.226	0.0199
	(0.0102)	(0.819)	(0.0360)
Constant	-2.944	60.31	14.24
	(1.265)	(13.15)	(8.275)
Observations	10,582	3,609	10,490
R-squared	0.134	0.311	0.360
District FE	YES	YES	YES
Region Year FE	YES	YES	YES

Notes: The dependent variables measure, respectively, the (log) number of employees (Employment); the (log of) wage per capita (Wage); the (log of) sales of employees (Productivity), all computed at the firm level. All variables have been deflated using the GDP deflator from the IMF. Standard errors are clustered at the district level and are reported in parenthesis below each coefficient.