Journal of Development Economics 171 (2024) 103339

Contents lists available at ScienceDirect

Journal of Development Economics

journal homepage: www.elsevier.com/locate/devec

Regular article

Roads, competition, and the informal sector☆

Elena Perra [a], Marco Sanfilippo [a,b,*], Asha Sundaram [c]

[a] Department of Economics and Statistics "Cognetti de Martiis", University of Turin, Italy
[b] Collegio Carlo Alberto, Italy
[c] Department of Economics, University of Auckland, New Zealand

ARTICLE INFO

JEL classification:
O18
O17
O55

Keywords:
Infrastructure
Informal sector
Productivity
Ethiopia

ABSTRACT

We examine the impact of competition from better connectivity to domestic markets on formal and informal firms. Combining geolocalized information on road improvements under a large infrastructure investment programme with data on manufacturing firms in Ethiopia between 2001 and 2013, we show that an increase in competition is associated with higher labour productivity, capital-intensity, investment in physical capital and wages in the formal sector. On the contrary, there is no associated increase in labour productivity or wages in the informal sector. In fact, increased competition results in lower capital-intensity and investment, a shift in composition towards workers without primary education and a lower likelihood of operating in the informal sector. We thus highlight that the benefits of infrastructure improvement programmes may not accrue uniformly in the economy.

1. Introduction

It is well-established that public investments in transportation infrastructure are crucial for economic growth (Duranton and Turner, 2012). What is less studied is the impact of infrastructure investments on firms, particularly in the context of heterogeneous firms that may benefit differently from such investments. This paper is an attempt to fill this gap. We investigate the effect of increased competition from an improvement in road connectivity to other domestic markets brought about by an extensive infrastructure development programme in Ethiopia. Our analysis focuses on formal and informal manufacturing firms and highlights that the benefits of road infrastructure development programmes may not accrue uniformly to firms in both sectors.

Our focus on informal firms is important for various reasons. Recent empirical work has emphasized the important role of the informal sector in job creation and structural transformation in developing countries (La Porta and Shleifer, 2014; Ulyssea, 2018; Dix-Carneiro et al., 2021). In the manufacturing sector, though formal firms contribute more to productivity growth, a majority of firms are informal and employ a larger share of the workforce (Diao et al., 2021; Kruse et al., 2022). In spite of the prominent role played by informal firms in generating jobs, few analyses have asked how policy reforms or other external shocks shape the composition of the manufacturing sector in terms of formal and informal firms.

Informal sector firms face unique constraints that influence their adaptation strategies. While formal sector firms can respond to competition by investing in better technologies, R&D and increasing efficiency (De Loecker and Goldberg, 2013; Topalova and Khandelwal, 2011), informal firms lack such capacity. Factors such as inadequate access to credit and information, and low education levels among informal entrepreneurs (La Porta and Shleifer, 2014) can hamper their ability to respond to increased competition from firms in connected areas. Besides, the increase in competition is likely to impact firm selection as in Melitz (2003), as the least productive firms exit. The dynamics of firm adjustment may differ between the formal and informal sectors, given the informal sector's role in providing a means of survival in developing countries.

To examine the impacts of competition from an increase in connectivity, we combine granular, geolocalized, information on road improvements under the Ethiopian Road Sector Development Programme

☆ We would like to thank the Editor, Jeremy Magruder, and two anonymous reviewers for helpful suggestions and for guiding us through the revisions of the paper. Mohammad Amin, Antonio Spilimbergo, Alan Finkelstein Shapiro, and participants to the Nordic Conference on Development Economics in Helsinki, the 2022 IMF Economic Review Conference in Rabat, the Annual Bank Conference on Development Economics at the World Bank (online), the Annual Conference of the Italian Economic Association in Turin, the Jobs and Development Conference in Cape Town, the MIASA Policy Conference in Accra and seminars at DEGIC (World Bank) and the Policy Studies Institute, Addis Ababa provided useful comments on early draft versions of the paper. We are grateful to Nicola Limodio for sharing the data on the geographic distribution of bank branches in Ethiopia.

* Corresponding author at: Department of Economics and Statistics "Cognetti de Martiis", University of Turin, Italy.
E-mail addresses: elena.perra@unito.it (E. Perra), marco.sanfilippo@unito.it (M. Sanfilippo), a.sundaram@auckland.ac.nz (A. Sundaram).

https://doi.org/10.1016/j.jdeveco.2024.103339
Received 30 March 2023; Received in revised form 25 June 2024; Accepted 8 July 2024
Available online 15 July 2024

(RSDP) with firm level data from the formal and informal sectors of Ethiopia between 2001 and 2013. We utilize the Small Scale Industries Survey (SSIS), covering small and informal firms, and the Large and Medium Manufacturing Industries Survey (LMMS), which provides data on firms in the formal sector. We explore a range of firm outcomes, such as the likelihood of operating informally, choice of technique, investment in physical capital, labour productivity and the composition of workers given their level of education.

Ethiopia is an excellent case for various reasons. First, the RSDP was a massive road improvement programme commencing in July 1997 to improve connectivity in the country. New roads were built and existing roads upgraded in quality, generating variation across time and space in improvements in the road network and reductions in travel time. Reductions in travel time occurred not just because the road network expanded, but also because roads were upgraded and better quality roads (such as paved, relative to gravel) allow greater speeds. Second, roads dominate transport in Ethiopia, which is characterized by an almost complete lack of substitutes.[1] During the period covered by our study, the Ethiopian road network accounted for 90%–95% of total inter-urban freight (Worku, 2011). Therefore, road improvements and the resulting expansion of the transport network and reduction in travel time produce sizeable changes in trade costs for firms. The availability of granular data on road improvements and firm location, complemented with detailed data on formal and informal firms allows us to isolate the effects of road improvements on firm outcomes in a quasi-experimental setting.

Our empirical analysis follows the existing literature (Donaldson and Hornbeck, 2016; Huang and Xiong, 2018; Jedwab and Storeygard, 2021; Fiorini et al., 2021) and adopts the spirit of the market-access approach to construct a measure of competition from greater road connectivity. We begin by treating each Ethiopian district (*woreda* in the Ethiopian context) as a local market. For each district *and* industry of a firm in a given year, we construct a weighted average of the inverse of travel times to all other districts given the road network and travel speed (which depends on the quality of the road), where the weights are total production in the district and industry. Variation in this measure captures both variation in production in the firm's industry in connected markets, and variation in travel times as roads are expanded and improved. It is thus a time-varying measure of changes to competition faced by the firm as the road network evolves.

Borrowing from the literature (Donaldson, 2018; Huang and Xiong, 2018), we refer to this measure as consumer market-access (CMA) to convey the idea that it measures the availability of alternatives for consumers in a particular industry. Take a given sector, furniture, as an example. Consumers in a remote district A have few options other than to buy furniture from local firms. But, with improvements in connectivity with other districts where furniture firms are active (i.e. an increase in the CMA for furniture firms in district A), consumers in A have better access to furniture products produced in connected markets. Thus, the increase in CMA translates into more competition for furniture firms in District A. Note that in line with our focus on firms, by construction, this measure can only be applied to districts and industries in which at least one active firm is present.

The CMA measure is different to the more widely-used measure of market-access that captures the idea that as travel times decrease with the improving road network, firms will have better access to consumers in connected markets. We call this an improvement in firm market access (FMA). We account for improved FMA with fixed effects that vary by district and year because, arguably, this effect is uniform across industries. In the presence of these district-year fixed effects, the CMA measure primarily captures competition from producers in connected markets. We then relate CMA to firm outcomes such as productivity, capital-labour ratio, investment, wages and skill composition of the workforce.

Identification of the causal effects of competition from road connectivity improvements on firm outcomes is susceptible to the canonical concern of endogeneity bias. Particularly, the changes in travel time given the road network are likely to be endogenous. As reported by Gebresilasse (2023), Moneke (2020), Fiorini et al. (2021) for the Ethiopian case, road construction under the RSDP was potentially non-random, both in terms of timing and placement. It is plausible that policy makers made endogenous investment allocation decisions, motivated by an array of considerations. These range from economic, such as higher economic and social potential of particular districts to political, such as favouritism in the allocation of public investments (Burgess et al., 2015).[2]

For this reason, our identification strategy relies on an instrumental variable approach that exploits road improvements occurring outside "exclusion areas" (as in Jedwab and Storeygard, 2021). Exclusion areas are defined as the surroundings of the pre-RSDP (1969) Ethiopian road network. The rationale is that incremental investments in roads in such exclusion areas are expected to be endogenous, given their proximity to pre-existing road arteries and connections between towns. By excluding road improvements that are most likely to be endogenously determined, our instrument isolates exogenous variation in road improvements. We check that our results are robust to utilizing different definitions of the pre-RSPD network (for instance, the road network in 1996 instead of in 1969) and to an alternative instrumental variable, similar to the one employed by Moneke (2020).

During Ethiopia's occupation, road arteries were built to connect capitals of former ancient kingdoms to each other and to major ports. Road construction was almost exclusively motivated by military considerations to facilitate colonial conquest, and ignored features of the terrain. We use the road map from Ethiopia's occupation to construct a synthetic or hypothetical road network that varies over time as an instrument for actual road improvements under the RSDP. The instrument isolates variation in road improvements stemming from straight line, orthogonal distances to roads constructed during occupation, and a budgeting algorithm whereby districts are sequentially connected to the road network with the one closest to the artery connected first. The instrument is therefore less susceptible to endogenous economic and political considerations driving public investments in road infrastructure.[3]

We first document a positive relationship between CMA and competition in local markets. More specifically, we show that an increase in CMA is associated with reduced firm markups and a lower likelihood of operating in the informal sector, consistent with an increase in competition and pointing to potential selection effects, where less productive informal firms exit.[4]

[1] Ethiopia has no direct access to the sea, nor does its territory contain any transportation substitutes to roads, such as navigable rivers, canals or railroads, with the exception of a single railroad line to Djibouti, which was not functioning during our study period.

[2] For instance, using the same data for Ethiopia, Perra (2022) provides evidence that areas connected ethnically and politically to the ruling party received more roads, and roads of better quality.

[3] Note that changes in the sectoral production weights, the other component of the CMA, may also be endogeneous, as they might evolve with the road network. In a robustness test, we employ an instrument where the weights are fixed at pre-sample levels of production for each market. These weights are based on data on the industrial distribution of workers from the 1994 population census of Ethiopia.

[4] The exit channel à la Melitz and Ottaviano (2008) would suggest that an increase in competition in the formal sector is associated with the least productive formal firms exiting into the informal sector (leading to entry into the informal sector) and the least productive informal firms exiting production. Given that our data are cross-sectional in nature and do not allow us to study exit of firms, we do not explicitly focus on this margin of adjustment in the paper. However, the lower likelihood of operating in the informal sector suggests that exit from the informal sector dominates entry from the formal to the informal sector.

Next, we find that the impact of CMA on firm outcomes varies substantially across formal and informal sector firms. Among formal firms, a one standard deviation increase in CMA corresponds to a 6.5% increase in labour productivity, an effect almost twice as large as the effect for the sample taken as a whole. To put this differently, for a formal firm in the furniture industry (ISIC code 3610) that moves its production from the district of Degua Temben in Tigray (at the 25th percentile of the CMA distribution in that sector) to the district of Akaki-Kalit in Addis (75th) our analysis reports a productivity gain of about 11%. For formal firms, increases in CMA are linked to improvements in capital-intensity, investment in physical assets and wages. Among informal firms, we find no relationship between CMA and labour productivity. If anything, the relationship is weakly negative. Results show that an increase in CMA reduces the capital-labour ratio and investment among informal firms. In addition, exploiting information on the level of education of each individual worker within a firm from the SSIS, we find that an increase in CMA is associated with a larger share of workers without primary education and a smaller share with higher education among informal firms.

These findings are consistent with a framework where firms in the formal sector compete nationally, while the informal sector operates locally.[5] An increase in competition from better intranational connectivity results in an increase in the elasticity of demand for each formal sector product variety. This leads to a decrease in product varieties produced, a decrease in firm markups and an increase in firm size in the formal sector (Desmet and Parente, 2010). The increase in firm size means that the returns to investments in physical capital typically associated with innovative activity and technological upgrading are larger, making such investments more attractive. In a capital-constrained environment, where formal firms can access formal credit, while informal firms resort to informal sources, the increase in demand for capital in the formal sector drives up the cost of capital for informal firms. This results in lower (higher) capital input, capital intensity, investments in physical capital and labour productivity in the informal (formal) sector. Note that exit of less productive formal firms into the informal sector and of informal firms from production due to the increase in competition would raise aggregate productivity in the formal and informal sectors. In the informal sector, it would work against the capital investment channel discussed above, thereby yielding ambiguous empirical results on labour productivity.

We test for evidence of the mechanism we propose in the paper. We show that the differential effects of CMA on labour productivity of formal versus informal firms hold particularly in (a) industries with greater reliance on credit, as proxied by the share of interest paid over sales; and (b) in districts without any bank branches.

Finally, we conduct a battery of tests to ensure that our results are qualitatively robust to alternate specifications and definitions of the variables of interest and to control for potential confounders at the district and industry levels. Overall, our results highlight that competition from better road connectivity due to road infrastructure improvements may disadvantage the informal sector, as it disciplines the formal sector.

This paper speaks to the literature looking at the economic benefits of public investments in infrastructure (Duranton and Turner, 2012) and at the literature studying the role of geography in influencing firm choices (Redding, 2020). We augment this literature by explicitly focusing on the informal sector and underlining the differential effects of such investments on informal firms, whose adaptation strategies and constraints differ from those of formal firms. Our paper is thus related to key studies exploring the impacts of infrastructure on informality in developing countries. Focusing on an infrastructure maintenance programme in rural India that bundles transport and electricity, Chaurey and Le (2022) find a positive effect on local employment, which is driven by an increase in the number of (mostly informal) micro-enterprises. Chatterjee et al. (2021), find that access to roads in India has positive productivity effects on formal, but not on informal firms. While their approach is based on proximity to a specific corridor, ours exploits connectivity improvements as the entire Ethiopian road network evolves over time.

Though we look at reductions in *intranational* trade costs and an increase in domestic competition, our study is related to the literature on the impact of trade liberalization on informality, which has not yet reached a consensus (Goldberg and Pavcnik, 2003; Nataraj, 2011; McCaig and Pavcnik, 2018; Becker, 2018; Dix-Carneiro et al., 2021). In the African setting, McMillan and McCaig (2020) find that trade liberalization in Botswana was associated with an increase in the prevalence of working in an informal firm and in self-employment. Erten et al. (2019) show that workers in districts facing larger tariff reductions in South Africa experience a significant decline in both formal and informal employment in the tradable sector.

Finally, we contribute to a small but growing strand of evidence on the implications of the RSDP for Ethiopia. While previous work has investigated the impacts of roads on firm productivity (Fiorini et al., 2021; Shiferaw et al., 2015), agricultural productivity (Adamopoulos, 2019; Gebresilasse, 2023), local economic development (Alder et al., 2022) and the structural transformation of local labour markets (Moneke, 2020; Fiorini and Sanfilippo, 2022), no evidence has so far been available on the informal sector. In this context, our findings reasonate with those by Diao et al. (2021), who find that in Ethiopia, the productivity benefits of global integration accrue disproportionately to formal firms at the top of the distribution, with gains concentrated mainly in productivity than in employment. We thus emphasize the tension between inclusive employment growth and enhancements in productivity.

The remainder of the paper is organized as follows: Section 2 provides an overview on the context of the Ethiopian RSDP; Section 3 describes the data employed in the analysis and outlines the empirical approach adopted for this study; Section 4 reports evidence on the implications of CMA for competition in local markets. The main results, including mechanisms, are discussed in Section 5, while Section 6 provides a list of robustness checks. Section 7 concludes.

2. The road sector development programme

The Road Sector Development Programme (hereafter, RSDP) is an ambitious investment programme implemented in Ethiopia since 1997 and still ongoing, with the objectives of rehabilitating existing Ethiopian roads and constructing new networks. The Ethiopian road network has increased from 26 550 km in 1997 to 113 066 km in 2016, while the proportion of the country's rehabilitated roads has increased from 22% to 72%. Therefore, road density per 100 sq. km has risen significantly from 21.1 km in 1997 to 102.8 km in 2016 (ERA, 2016; Bank, 2021). The main authorities in charge of its implementation were the Ethiopian Roads Authority (ERA) and the Regional Roads Authorities (RRAs). This large-scale development project has attracted particular interest among researchers analysing the impact of infrastructural investments. Evidence has shown so far that the RSDP was a key driver of agricultural productivity (Adamopoulos, 2019), spurred business activity (Shiferaw et al., 2015; Fiorini et al., 2021) and stimulated structural transformation from agriculture to the services sector (Fiorini and Sanfilippo, 2022; Moneke, 2020).

The RSDP is considered to be the largest infrastructural investment projects ever implemented by the Ethiopian government and one of the most ambitious infrastructure programmes in the entire region (Shiferaw et al., 2015). Its implementation has required a significant investment in foreign currency, with an estimated cost over the fourteen years of around US$7.08 billion (Worku, 2011; Shiferaw et al., 2015).

[5] In Appendix B we provide a more detailed discussion, including a simple theoretical model, regarding the conceptual framework.

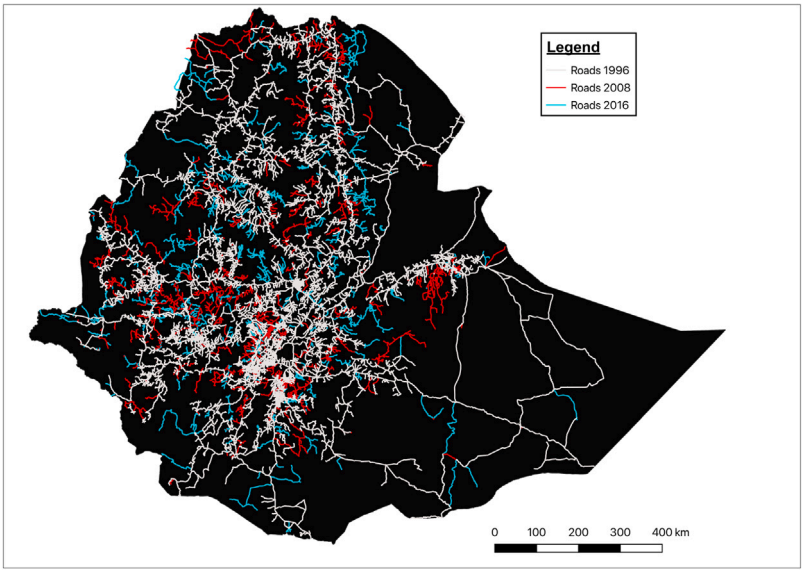

Fig. 1. Ethiopian road network under the RSDP. *Notes*: Roads in white represent the state of the road infrastructure in Ethiopia before the start of the RSDP. Roads in red and blue show upgrades completed during different phases of the implementation of the program. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
Source: Authors' calculation on RSDP data.

The five-year plan of the RSDP has been implemented thorough annual action plans, closely supervised and influenced by the government (Shiferaw et al., 2015). Although ERA has assigned different criteria for road upgrading projects, it is not clear which specific variables it employs to operationalize them (Worku, 2011; Shiferaw et al., 2015).

Road improvements under the RSDP are documented in Fig. 1, which shows the remarkable advancement of the road network between 1996 (the baseline) and two successive periods, 2008 and 2016. Improvements in multidimensional aspects of road accessibility indicators are summarized in Table A.1 in the Appendix, which shows a general increase in the proportion of roads in good and serviceable condition. Improvements in road quality are captured by the registered drop in travel time needed to cross each segment of the network. The data on road quality improvements are aggregated in the speed matrix proposed by ERA (2009) and reported in Table A.2, which describes the average travel speed as a function of the particular road surface. Figure A.1 in the Appendix reports visual evidence on how the average bilateral travel time for each Ethiopian district with other districts has improved over time and across the country.

3. Data

3.1. Firm-level data

We combine two sources of microdata covering the whole manufacturing sector in Ethiopia. The first is the Large and Medium Manufacturing industry Survey (LMMS), an annual census of firms published by the Central Statistical Agency (CSA). Data cover all firms with at least 10 persons engaged and that use electricity in their production process. Firms are required to respond to this census every year; therefore, this source includes the universe of 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 of the workforce and on the location of each firm. Firms also provide details on sales values and quantity produced for the domestic and international market for each product, as well as information on raw materials, both domestic and imported, employed at the firm level for the production processes, and their share in total firm expenditure. Manufacturing industries are defined at the 4-digit level according to the ISIC Rev. 3 classification.

The second dataset is the Survey of Small-scale manufacturing Industries (SSIS). We combine all existing waves of the SSIS, covering the years 2001, 2004, 2007, 2010, 2013. This is a survey that covers small (with less than 10 persons engaged) and mainly informal firms operating 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 sectors and the grain mills industry), sampled in similar proportions across regions. Due to 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.

Table A.3 reports figures for the years in which the SSIS and the census were run simultaneously. On average and consistently over time, small and informal firms represent the large majority of all manufacturing establishments, approximately half of total manufacturing employment, but a much smaller share in terms of total production, wage bill and capital expenditures.[6]

We combine the two datasets (LMMS and SSIS), obtaining information at the firm, industry and *woreda* level. Since both datasets are based on a similar questionnaire, we can reasonably compare indicators across them. Based on the pooled data, we define informal firms as those who do not keep books of accounts and have less than 10 employees.[7] This conceptualization of informality is the closest to the official definition that the CSA provides of an informal firm, that is firms: "*that do not keep complete books of accounts; mainly engaged in market oriented production; do not register the enterprise and its employees; and have a very limited number of persons engaged (less than ten persons) in the enterprise; have no license*" (Siba, 2015).

Descriptive statistics show that an informal firm appears to be on average smaller, younger, less capital intensive and offering lower wages compared to a formal firm (see Table A.4 in the Appendix). This is reflected in Fig. 2, which shows the productivity distributions of formal and informal firms. The wider right tail of the formal firm distribution

[6] See Table A.5 in the Appendix for an overview of the role played by the small business sector in Ethiopia.

[7] In Section 6.3 we also investigate alternative definitions of informality as a robustness check.

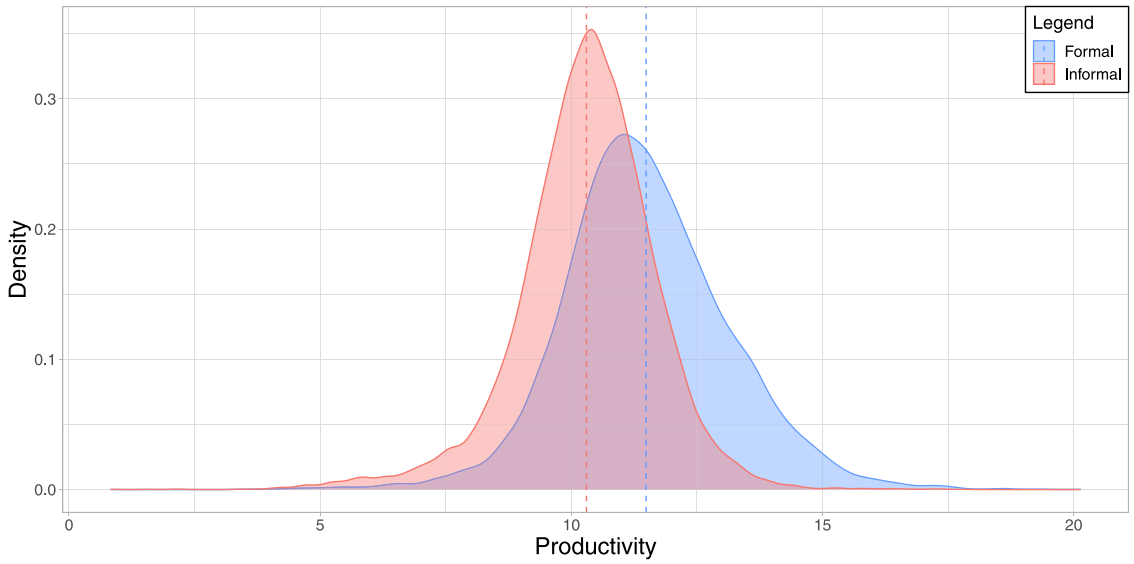

Fig. 2. Productivity distribution for formal and informal firms. *Notes*: Productivity is calculated as the simple average of firms' labour productivity (output per employees) over the whole sample period and it is reported in logarithms.
Source: Authors' calculation on LMMS and SSIS data.

confirms their productivity advantage over informal firms, which are concentrated on the left-hand side of the distribution. Focusing on their spatial distribution, Figure A.2 in the Appendix shows that they appear to be equally spread across Ethiopia, with a higher concentration of both types of firms in the areas surrounding Addis Ababa. Yet, over time, firms have spread out across the country, especially in Amhara, Oromia, Tigray and SNNP (see Figure A.3).

3.2. Roads

This paper employs rich geolocalized data on the Road Sector Development programme (RSDP), which spans 1996 to 2016. This database consists of a time series of shape files of the Ethiopian road network, describing the incremental improvements in terms of road surface (earth surface, minor gravel, major gravel and asphalt) and travel time needed to cross each road segment. Moreover, it provides details on whether a road-segment is categorized as not-rehabilitated, rehabilitated or completely new. This exhaustive data represents a significant source of information, especially in the light of limited availability of time series data on transport infrastructure in low-income developing countries. Information on improvements in the road surface allow us to calculate enhancements in the average travel time needed to cross each road segment in accordance with the speed matrix reported by the ERA.[8]

In the rest of the paper we employ a market access approach based on Donaldson and Hornbeck (2016) and Jedwab and Storeygard (2021). Market access is a useful indicator to account for various dimensions of the role of road improvements, including: (i) its capacity to reduce the cost of transporting goods, allowing firms to sell their products to larger markets in the country; and (ii) its capacity to increase competition between firms. To some extent, an increase in market access can be viewed as a domestic shock akin to trade liberalization. In the remainder of the paper, we focus on one specific dimension of market access, which we label "Consumer Market Access (CMA)". For a firm i in industry j, CMA accounts for how proximity to competitors (i.e. firms operating in the same industry j) based elsewhere in the country changes with improvements in the road network (Huang and Xiong, 2018).

The definition of CMA is based on a modified version of the market access approach, in which changes in travel time τ are weighted using the level of total production in each woreda-industry pair jx at time t:

$$CMA_{jxt} = \sum_{d \neq x} Production_{jdt} * \tau_{jxdt}^{-\theta} \quad (1)$$

The minimum distance in hours τ, is calculated employing Dijkstra's algorithm; θ is the elasticity measuring the decrease in trade volumes as travel time increases. Empirical papers using a market access approach have resorted to different values of θ, usually in the range of 1 (the market potential approach originally proposed by Harris (1954)) to about 10 (Donaldson and Hornbeck, 2016). For our analysis we rely on a value of θ equal to 3.8. This is the same value adopted by Jedwab and Storeygard (2021) in their paper looking at the effects of road improvements on urbanization in Africa.[9] In robustness checks, reported in Section 6.4, we show that results are robust to different values of θ.

Fig. 3 depicts the distribution of CMA for one manufacturing industry with a wide coverage of both the formal and informal firms, the furniture sector (ISIC code 3610). Appendix Table A.6 reports the variation of average CMA by year. It shows a steady increase in CMA over time (column 2), consistent with markets becoming more integrated with road development. It also reports the growth rate of CMA (column 3), which increases between 2004–07 and then decreases. Figure A.4 shows the geographical dimension of changes in CMA over time. While initial increases in CMA are close to the centre (and therefore, to Addis), by 2013, CMA is increasing in more remote regions. Finally, Table A.7 shows that, on average, CMA for formal firms is higher than for informal firms, and this holds across all years.

3.3. Identification strategy

Our empirical analysis is based on estimating the following relation:

$$Y_{ijxt} = \beta_1 CMA_{jxt} + \beta_2 X_{it} + \theta_{xj} + \phi_{xt} + \epsilon_{ijxt} \quad (2)$$

[8] The same matrix has been employed also by Shiferaw et al. (2015), Jedwab and Storeygard (2021), and Fiorini et al. (2021).

[9] Jedwab and Storeygard (2021) obtained this value using the estimated cost-distance elasticity for Nigeria and Ethiopia from Atkin and Donaldson (2015), which is 3 times larger the one found by Duranton et al. (2014) for the US. Duranton et al. (2014) adopt a cost-distance elasticity for the US of 1.27, meaning that $1.27 * 3 \approx 3.8$.

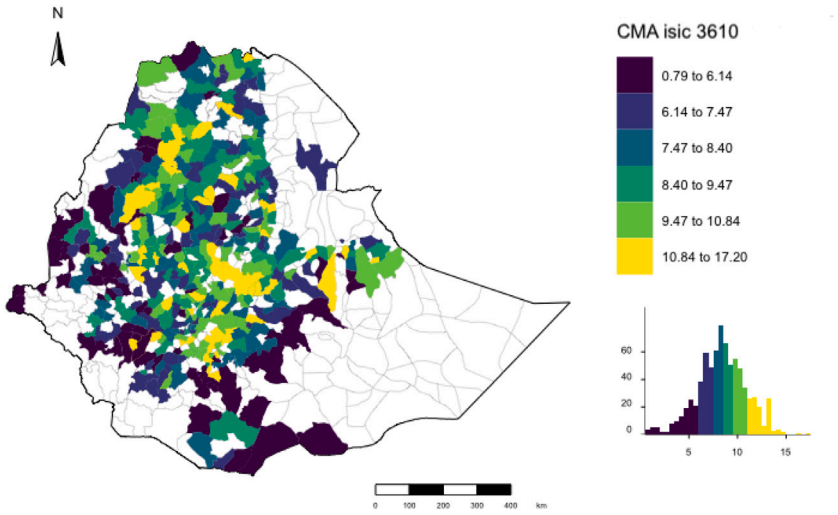

Fig. 3. Consumer Market Access (CMA) in the furniture industry. *Notes*: The map reports the simple average of CMA across different districts during the sample period in the furniture industry, which corresponds to the ISIC code 3610.
Source: Authors' calculations.

which relates CMA to indicators of performance (Y_{ixjt}) of (formal and informal) firms. Since we employ a repeated cross-section of firms, our identification strategy exploits changes within markets over time. For this reason we add woreda-industry (θ_{xj}) fixed effects. Woreda-year (ϕ_{xt}) fixed effects are added to account for time specific changes occurring within each district over time, and are important to account for shocks to economic activity that can affect the relationship of interest. X_{it} includes firm specific controls, such as the age of the firm and a dummy variable accounting for whether the firm was originally surveyed in the SSIS or in the LMMS.[10] To account for the presence of spatial correlation in the residuals, we estimate our model by introducing a spatial HAC correction of standard errors based on the Conley method, using the code proposed by Colella et al. (2019). For the main specification, we consider a radius of 200 km around each district's centroid. This radius roughly corresponds to the distance that separates each woreda and 95% of its neighbours, travelling at an average speed of 45 km/h, or the speed of a federal gravel road after rehabilitation. It is also wide enough to cover, especially in the central part of the country, a large number (109) of contiguous districts, thereby accounting for spatial autocorrelation and heteroskedasticity extending far beyond the immediate neighbourhood of each observation.[11]

We now deal with endogeneity issues regarding the placement of transport infrastructure. Indeed, the choice of where to build infrastructure is not exogenous, since random assignment of route placements is implausible. It is reasonable to assume that planners decide to allocate investments with specific goals — for instance, where high growth is expected or in specific peripheral areas connecting target nodes (Asher and Novosad, 2020; Duflo and Pande, 2007). Moreover, domestic income and trade shocks that determine infrastructure investments are likely to be spatially correlated (Chandra and Thompson, 2000). Omitted variable bias is a further challenge. Indeed, market access may be driven by population changes across locations, by changes in the road network connecting them, or by natural impediments and other unobserved factors. All these are likely to influence the allocation decisions of road investments. Moreover, roads may be built in anticipation of benefits from growth prospects of neighbouring cities or nearby economic hubs (Jedwab and Storeygard, 2021). Finally, policymakers may compete to attract larger shares of infrastructure investments to their region, which could be correlated across locations.

In the specific context of the RSDP in Ethiopia, any anticipation should be mitigated by the structure of the programme, which is linked to a five-year investment plan (Shiferaw et al., 2015). However, in order to deal with residual concerns, we propose an instrumental variable approach. This approach involves creating alternative synthetic (or hypothetical) road networks that are used as instruments to actual roads under the RSDP in the market access indicator in equation (9).

By digitizing the 1969 CIA's map of Ethiopia (Figure A.5), we apply a strategy similar to the one adopted by Jedwab and Storeygard (2021) and Fiorini et al. (2021). We identify a 50 km buffer along digitized roads as of 1969. We then define exclusion areas, where the road network will remain frozen, exactly as it was before the RSDP. Outside the exclusion zones, the road network changes over time according to RSDP improvements. As shown in Fig. 4, the 50 km buffer includes all major city centres with a population larger than 50,000 people in 1994.[12] We compute the bilateral distances between an origin centroid of woreda x, and a destination centroid of woreda d, without taking into account road changes in the exclusion zone. This new road network is a synthetic road network, which we use to instrument the actual road network.

"Freezing" roads in areas inside the 1969 buffer enables us to exclude from the analysis all enhancements under the RSDP that are more likely to be affected by endogenous drivers. The exclusion zones reflect principal Ethiopian thoroughfares connecting economic hubs and main cities, which we argue are most attractive for long-term investments. As shown by Bertazzini (2022), transport networks and concentration of economic activity in Ethiopia remain fairly persistent

[10] There is in fact some degree of overlap among the two surveys. Firms that do not keep books of accounts are also present in the LMMS. Moreover, the LMMS includes some firms reporting less than 10 persons employed. The opposite also happens in the case of SSIS. Overall, about 7% of firms in the census are classified as informal, while almost 20% of those in the SSIS can be accounted for as formal according to the definition used in the paper. In some robustness checks reported in Section 6.3 we consider alternative definition of informality based on either just the survey instrument the firm belongs to or the size threshold, and find that such differences in the sample composition do not affect our results.

[11] In additional analyses, we also check for the robustness of our results using two alternative radii (see Section 6.6): a more restrictive distance of 100 km, corresponding to the distance in km when considering the average travel time between woredas (142.87 min) and an average speed of 45 km/h, and a larger specification (300 km), which encompasses the distance between each woreda and 99.9% of its neighbours at the same speed.

[12] 1994 is the year of the population census, and it is also three years before the official start of the RSDP.

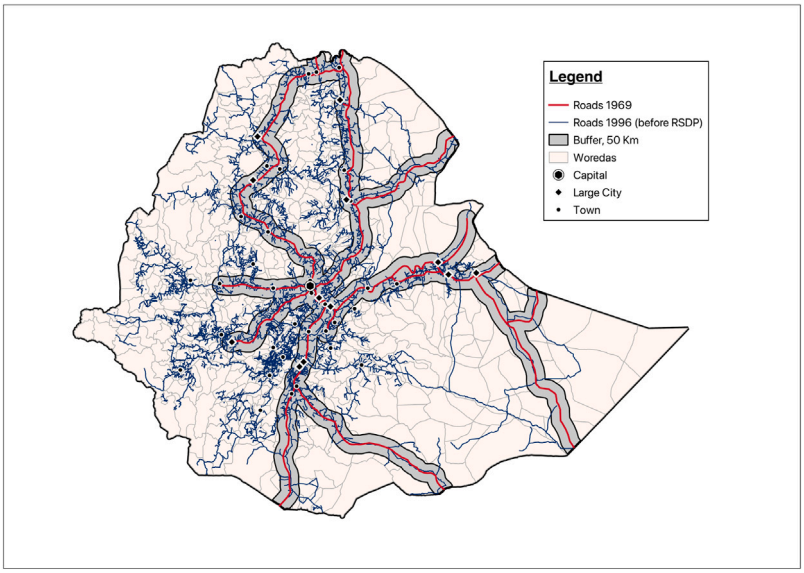

Fig. 4. Buffer of 50 km around the 1969 road network. *Notes*: This figure provides a graphical representation of the IV strategy employed in the analysis. The 50 km buffer is represented in grey in the map, and follows the trajectories of the digitalized road network of 1969, represented in red. The geographical areas of the buffer include the main Ethiopian city centres. Inside the buffer, roads are frozen as at 1996, while outside, they change in accordance with the upgrades of the RSDP. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
Source: Authors' calculation on United States Central Intelligence Agency (CIA) (1969) and RSDP data on the road network.

over time.[13] Our approach thus enables us to account not only for endogenous local road improvements, but also for those improvements targeted at connecting principal Ethiopian economic hubs and cities.

4. Roads and competition

In this Section we look at the implications of an increase in CMA on competition, which is the first order effect underlying the impacts on firm productivity. We bolster our argument that an increase in CMA triggers greater competition in three ways.

First, we provide a more descriptive piece of evidence by computing an Herfindahl–Hirschman Index (HHI) at the level of each market and plotting its relationship with CMA. Figure A.7 in the Appendix shows that higher levels of CMA do indeed correspond to more competitive, or less concentrated, markets. Next, we employ a regression framework to examine the effect of CMA on the distribution of formal and informal firms (Section 4.1), and on firm markups as a more direct measure of competition (Section 4.2).

4.1. Market composition

An increase in competition could lead to a restructuring of markets in which formal and informal firms compete. Less productive formal firms may move into the informal sector, while less productive informal firms may leave the market. Overall, given that informal sector firms tend to be less productive than formal firms, we expect greater competition to be associated with greater exit in the informal sector and greater formality. We explore this empirically, while acknowledging that a detailed analysis of entry and exit dynamics requires panel data on formal and informal firms, which we do not have. Hence, we follow Nataraj (2011), and construct a binary variable that takes a value of one if the firm is in the informal sector and zero otherwise.[14] We then estimate the relationship between CMA and this indicator of informality using the instrumental variable estimation strategy. Results are presented in Table 1. Column (1) includes no control variables, while column (2) introduces firm controls. Results show that the coefficient on CMA is negative and statistically significant only in the conditional regression in column (2). Thus, there is evidence that greater competition resulting from better connectivity to other internal markets brought about by road improvements is associated with a lower likelihood of a firm operating in the informal sector. Differently from Nataraj (2011), our coefficient of interest is (slightly) significant, reinforcing the argument that the increase in connectivity is at the expense of the informal sector, strengthening the exit mechanisms on the left-hand side of the firm productivity distribution.

4.2. Markups

We argue that our CMA measure captures increases in competition faced by firms as better roads improve connectivity to other markets, thereby increasing exposure to competitors. To test this channel, we construct firm-level markups using the approach by De Loecker and Warzynski (2012) and estimate our relationship of interest (Eq. (2)) with the markup as a dependent variable.[15] Results are reported in Table 2 and demonstrate a negative relationship between CMA and markups, consistent with an increase in CMA being associated with

[13] Bertazzini (2022) shows how proximity to colonial roads, and therefore lower transport costs, attracted economic activity until the 1960s. In turn, this generates a positive feedback mechanism, driving investments in more economically developed areas with greater potential to reap the benefits of increasing returns of scale. During the Italian occupation between 1935–1941, the Italian road programme built a total of 7000 km of roads, 3450 km of which were tarred (Baker, 1974), which where primarily designed to serve military purposes. Colonial roads were kept operational after Liberation until 1951. Only after 1960 were other major road arteries built. In particular, the Imperial Highway Authority launched a major series of infrastructure projects, which from 1951 to 1968 led to the construction and improvement of 7304 km of highway, with the goal of reaching all the main cities and towns linked to the capital and improving access to the Lakes Region and the coffee-producing areas (Baker, 1974).

[14] Since SSIS is a survey, this exercise is limited to the sample used in each wave.

[15] For each firm i in our sample we compute markups as follows:

$$\mu_i = \alpha_i^v(\beta_i^v)^{-1}, \quad (3)$$

Table 1
Effects of CMA on informality.

	(1) Informal	(2) Informal
CMA	−0.0145	−0.0136*
	(0.0117)	(0.0056)
Observations	23 662	23 662
R^2	0.001	0.242
Dist-Ind. FE	Y	Y
Dist-Year FE	Y	Y
Controls	N	Y
F-stat	1143.536	1139.506

Notes: The dependent variable is a dummy that takes a value of one if the firm operates in the informal sector and zero otherwise. The variable of interest is CMA, which varies at the woreda, industry and year level. Estimates in column 2 include firm age and a binary variable taking value of 1 for firms in the SSIS. Conley standard errors are reported in parenthesis. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 2
Results IV: Markups.

	Whole sample (1) Markup	Informal (2) Markup	Formal (3) Markup
CMA	−0.0288***	−0.0066	−0.0185***
	(0.0090)	(0.0105)	(0.0038)
Observations	21 257	12 776	7649
R^2	0.160	0.029	0.176
Dist-Ind. FE	Y	Y	Y
Dist-Year FE	Y	Y	Y
Controls	Y	Y	Y
F-stat	1112.538	1494.270	983.309

Notes: The dependent variable is the markup at the firm level, while the independent variable is CMA, which varies at the woreda, industry and year level. Results are reported for the whole sample, for the informal and the formal sector, respectively. Firm level control variables include firm age and a binary variable taking value of 1 for all firms observed in the Survey of Small-scale manufacturing Industries (SSIS). Conley standard errors in parenthesis *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

increased competition. The relationship is statistically significant for formal firms, supporting the view of CMA as a driver of competition in domestic markets.

These results also allow us to rule out other channels through which an increase in exposure to other firms operating in the same industry may impact firm outcomes. For instance, more exposure to firms in the same industry may increase learning and technology transfer, some of which may stem from people movement. This channel is much harder to pin down in the absence of travel or migration data specific to the industry-woreda. However, we would expect such learning effects to lower costs and increase markups in the absence of competitive pressures. Our finding that an increase in CMA is associated with lower markups suggests that the competition channel is a dominant one.

5. Results

5.1. Formal and informal firms productivity

In Table A.8 in the Appendix, we use ordinary least squares (OLS) to estimate the relationship between labour productivity at the firm level and consumer market access (CMA), which captures competition and varies at the woreda, industry and year level. All regressions include woreda-time and woreda-industry fixed effects and report Conley standard errors. Columns (1) and (2) report results for the whole sample of informal and formal firms. Columns (3)–(4) and (5)–(6) report results for the informal and formal sector firms respectively. While columns (1), (3) and (5) do not include control variables, columns (2), (4) and (6) include controls. We find that the coefficients of interest are positive and significant for firms in the formal sector. This is consistent with the idea that increased competition is associated with exit of less productive formal firms, resulting in greater aggregate productivity. In addition, competition can spur firms to invest in better technology, become more efficient and increase productivity. The relationship between competition and labour productivity is negative for firms in the informal sector, but the coefficient is not statistically significant.

Next, we provide results based on an instrumental variable approach using two stage least squares (2SLS). Second stage results are reported in Table 3, by exploiting instrumental variables constructed as discussed in Section 3.3. In order to isolate the exogenous improvements to the road network, we remove all upgrades in road construction in the exclusion zones around 1969 roads in the construction of our instrument. The first stage regression is reported in Table A.7 and shows that our instrument is a strong predictor of CMA. Similarly, the first stage F-statistic appears to be high, confirming the relevance of our instrumental variable. This very high first stage could be due to the fact that, like in the case of our CMA measure, the instrument uses time-varying industry production weights. In fact, when we freeze weights at the pre-sample year in both the CMA measure and the instrument, both the first stage coefficient and the F-test drop (see Section 6.5). Table 3 shows that our results are consistent with the OLS results, which nonetheless report a small downward bias. This (slight) downward bias is consistent with results in Jedwab and Storeygard (2021) with a similar instrumental variable strategy, and may be because the effects of CMA are larger for areas away from main directories. Our coefficient of interest implies that one standard deviation increase in CMA corresponds to a 6.5% increase in productivity of formal firms and a 3.9% increase for the whole sample. Alternatively, a given firm in the furniture industry (ISIC code 3610) moving from a district at the 25th percentile of the CMA distribution (such as Degua Temben in the Mehakelegnaw zone of the Tigray region) to one at the 75th percentile (Akaki - Kalit in Addis Abeba), will see an associated gain of 5.9% in productivity (the gain is almost 11% for formal firms).[16]

5.2. Heterogeneous effects

Size of the informal sector. In Appendix Table A.11 and A.12, we explore heterogeneous effects across districts with above and below median share of informal firms on the total number of firms (columns (1) and (2)) and as a share of total employment (columns (3) and (4)) for informal and formal firms respectively. The idea is to explore whether competition from better connectivity due to road improvements impacts labour productivity in the informal and formal sectors differently when the relative size of the informal sector in the area is high.

We find in Table A.11 that in areas with above median share of informal firms, CMA is associated with lower labour productivity and the coefficient is now statistically significant. This is not the case in

where $\alpha_i^v = \frac{\partial Q_{i(\cdot)}}{\partial V_i}\frac{V_i}{Q_i}$ is the output elasticity of variable inputs (labour and materials) and $\beta_{it}^v = \frac{p_{it}^v V_{it}}{P_{it}Q_{it}}$ is the share of expenditure on those inputs in total revenue. Each firm is assigned with an output elasticity based on their 2-digit industry (ISIC) code that we obtain from Fiorini et al. (2021), who estimate physical productivity (TFPQ) using a production function approach utilizing a fraction of the firm data that we use in this paper and for which panel data are available, covering the period 1996–2009.

[16] In order to reinforce the argument that our measure of market access (CMA) captures a competition effect that is industry specific, we construct an alternative measure of market access based on Eq. (1) in which we replace the value of total production with a value of 1. In this case, variation in CMA is solely driven by changes in roads and is independent of changes in industrial activity. Results of this exercise are reported in Table A.10. There is no significant relationship between this altered CMA measure and productivity, suggesting that changes in connectivity to other production hubs in the same industry are a crucial determinant of the relationship of interest.

Table 3
Results IV: Labour productivity.

	Whole sample		Informal		Formal	
	(1) Labour Prod.	(2) Labour Prod.	(3) Labour Prod.	(4) Labour Prod.	(5) Labour Prod.	(6) Labour Prod.
CMA	0.0781*	0.0758**	0.0031	−0.0026	0.1438***	0.1446***
	(0.0430)	(0.0327)	(0.0218)	(0.0211)	(0.0361)	(0.0286)
Observations	23 662	23 662	14 066	14 066	8685	8685
R^2	0.002	0.078	0.000	0.014	0.007	0.058
Dist-Ind. FE	Y	Y	Y	Y	Y	Y
Dist-Year FE	Y	Y	Y	Y	Y	Y
Controls	N	Y	N	Y	N	Y
F-stat	1143.536	1139.506	1527.226	1529.068	1040.476	1036.031
Mean DV	10.762	10.762	10.299	10.299	11.480	11.480

Notes: The dependent variable is labour productivity at the firm level, while the independent variable is CMA, which varies at the woreda, industry and year level. Results are reported for the whole sample, for the informal and the formal sector, respectively. Firm level control variables include firm age and a binary variable taking value of 1 for all firms observed in the Survey of Small-scale manufacturing Industries (SSIS). Conley standard errors in parenthesis * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

areas with below median share of informal firms. From Table A.12, we find that this heterogeneous impact is unique to the informal sector. In the formal sector, the increase in labour productivity with competition is largely uniform across areas with above and below median labour productivity. Thus, results in Table A.11 and A.12 lend further support to the channel whereby competition resulting from better road connectivity exerts differential effects across the formal and informal sectors, with negative productivity effects concentrated in the informal sector.

Size cohorts. Firms in the informal sector may operate differently based on their size. Firms with very few persons employed are more likely to be family firms focused on subsistence or survival and may react less to competition. On the other hand, firms close to the upper bound of the size distribution are more likely to behave like formal firms. Appendix figure A.6 reports the distribution of the coefficient of interest from a set of separate regressions based on sub-samples of informal firms of different size cohorts. Results show that the effects of competition are indeed heterogeneous across size groups. More specifically, for firms closer to the size threshold of 10 employees (which are a minority in the informal sector), we see a positive relationship between competition and CMA. Results for the majority of informal firms that are small echo those in Table 3.

5.3. Mechanisms

In Appendix B we introduce a simple theoretical framework where firms produce differentiated products and formal firms sell intranationally, while informal firms sell locally. An increase in competition from improved roads induces formal firms to invest in technology that requires capital investments. This is because greater competition is associated with lower markups and greater output as firms face a more elastic demand curve for their product variety. Greater output increases the returns to investment that lowers marginal cost, leading formal firms to demand more capital and switch to more capital-intensive techniques of production. This raises the cost of capital for informal firms that face constraints in accessing formal sources of credit, leading to lower investments and capital intensity among informal firms. In what follows, we empirically test some of the predictions of the framework, showing evidence on: (1) capital, investment and employment responses to changes in CMA (Section 5.3.1); and (2) differential effects based on access to credit (Section 5.3.2).

5.3.1. Capital and employment

We check whether capital endowments of firms change in response to increases in CMA. Table 4 reports estimations based on the IV regressions using capital stock, capital intensity and investments in fixed assets as dependent variables. We find that, for the formal sector, CMA has a positive effect on all of them. This suggests that formal sector firms are increasingly capital-intensive and competition is associated with increased investment, in line with improvements in technology. These findings are corroborated by the positive effects on total production, employment and, especially, wages (see Table A.14 in the Appendix). Conversely, we find no evidence for the informal sector. Rather, we find that greater competition from better connectivity is associated with a lower capital-labour ratio and less investment. These results echo the idea that as labour is released from exiting formal firms due to competition, it is potentially absorbed in informal sector firms that become less capital-intensive.

We probe this channel further using detailed data on worker education levels in informal firms. We explore how changes in CMA affect the composition of the workforce by education. This can be done only for the sub-sample of informal firms covered by the SSIS,[17] since this dataset includes a module in which firms report characteristics for each worker (including any working owner), including their level of education. Appendix Table A.15 reports estimates linking CMA to the total number of persons engaged (this includes both employees and working owners) and the share of persons engaged with (a) no education; (b) primary education; and (c) secondary education and above.[18] Results show that an increase in competition from higher CMA is associated with an increase in the share of less educated workers (workers without primary education) in informal firms. This finding complements our earlier results showing an increase in investment and capital-intensity among formal firms in response to competition. Adoption of technology is likely to be intensive in high-skilled labour, potentially shifting worker composition towards more educated workers in the formal sector. We would expect to see a corresponding shift towards less educated workers among informal firms, as shown in column (1) of Appendix Table A.15.

5.3.2. Access to finance

To test the channel highlighted in our conceptual framework, we split our sample into two industry categories — industries that have a high and low reliance on credit. To do so, we draw on pre-sample information on firms from earlier available years of the LMMS data, specifically, between 1996 and 2000. We then utilize interest payments made by firms divided by total sales for each 4-digit industry as a measure of the industry's dependence on credit or external sources of finance.[19] Figure A.8 in the appendix shows that while it is mostly capital-intensive industries, such as manufactures of basic chemicals (ISIC code 2411) or basic metals (3140) that have a larger share of interest payments to sales, industries such as manufacture of dairy (1520) and rope (1723) in the more labour-intensive food and textile sectors also rank high.

We replicate our main regression by splitting the sample along the median value of the measure (the red line in Figure A.8). If competition from better road connectivity induces formal firms to make capital investments, making capital more expensive and therefore less accessible for informal firms, we expect this mechanism to be stronger in sectors that typically face a high cost of external financing due to their inherent characteristics, such as nature of technology, demand or organization. Indeed, results in Appendix Table A.16 align with our expectations. From columns (1) - (3), an increase in consumer market

[17] An equivalent module including all the individual worker characteristics is not included in the LMMS. The analysis that follows is therefore based on the sub-sample of informal firms included in the SSIS only.

[18] Primary education corresponds to grades 1 to 7 in the Ethiopian system, secondary to grades 8–12, and tertiary above 12.

[19] Results of this exercise do not change if we use 1996 data alone. Using 1996 helps us avoid concerns related to the endogeneity of access to finance to road placement, since the RSDP commenced after 1996. However, the number of industries covered in 1996 is smaller.

Table 4
Capital intensity, Capital and investment.

	Whole sample			Informal			Formal		
	(1) Capital intensity	(2) Capital	(3) Investment	(4) Capital intensity	(5) Capital	(6) Investment	(7) Capital intensity	(8) Capital	(9) Investment
CMA	0.0991**	0.1662***	0.1289	−0.0568***	−0.0655***	−0.2045**	0.1818***	0.2729***	0.3096***
	(0.0450)	(0.0613)	(0.0843)	(0.0079)	(0.0164)	(0.0899)	(0.0603)	(0.0743)	(0.0730)
Observations	20 606	20 606	23 662	11 672	11 672	14 066	8059	8059	8685
R^2	0.117	0.295	0.010	0.033	0.033	0.010	0.079	0.231	0.011
Dist-Ind. FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Dist-Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y
F-stat	1147.544	1147.544	1139.506	1649.012	1649.012	1529.068	992.651	992.651	1036.031

Notes: The dependent variables are the value of capital intensity (in log), measured as capital over employees; the value of capital (in log), measured as the book value of fixed assets at the beginning of the period; and the value of investment in fixed assets (in log) at the firm level. The variable of interest is CMA, which varies at the woreda, industry and year level. Results are reported for the whole sample, for the informal and the formal sector, respectively. Firm level control variables include firm age and a binary variable taking value of 1 for all firms observed in the SSIS. Conley standard errors in parenthesis * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

access is associated with an increase in labour productivity in both formal and informal sector firms in industries that face a lower cost of external financing, while columns (4) - (6) show that this positive relationship is statistically and economically insignificant for informal firms (but positive and significant for formal firms) in sectors that face a higher cost of external financing.[20] These results underscore capital investments as a key channel through which competition from better road connectivity impacts the productivity of formal and informal firms.

A potential shortcoming of our measure of the cost of external financing is that it does not account for firm-level heterogeneity in the cost of borrowing. For instance, large firms that can provide more collateral may be able to borrow more at relatively lower interest rates. This could lead to an interest payment-to-sales ratio comparable to that of a small firm which pays a higher interest rate on a small loan, introducing noise in the measure. To tackle this shortcoming, we employ an alternative measure of access to finance. Table 5 estimates our baseline regressions separately for districts that in 1996 had no bank branches and districts with at least one bank branch present.[21] The idea is that access to finance is most likely to be severely constrained in districts without any banks relative to other districts. It is in such districts that we would expect to see informal firms rationed, and therefore hampered from realizing the benefits of road improvements. Indeed, this is what we find. In districts with at least one bank, an increase in CMA is associated with an increase in labour productivity for both formal and informal firms. In districts with no banks, the increase in CMA is associated with an increase in labour productivity only for formal firms and not for informal ones, echoing the baseline result.[22]

6. Robustness checks

6.1. Firm Market Access (FMA)

We estimate a version of our baseline regression in which we include firm market access (FMA) as an additional control variable. The variable FMA is intended to capture improvements in market access for firms, as better roads allow them to access consumers in connected markets. The variable is constructed as a weighted average of bilateral travel times among each pair of districts in the country, the weights being the size of population in the district of destination.[23] Given its construction, FMA varies at the woreda-time level and therefore, we exclude woreda-year fixed effects from these regressions. Results are presented in Table A.17 and show that accounting for the expansion in market access for firms from better roads does not alter the findings on the impacts of competition operating through CMA. We find that the coefficient on FMA estimated on the whole sample is positive, suggesting that greater market access is associated with higher labour productivity. This is consistent with the idea, explored in related literature on the role of infrastructures on firms (e.g. Alder et al., 2022; Nataraj, 2011; Shiferaw et al., 2015) that greater scale can incentives investments in productivity enhancements. However, the coefficient is statistically insignificant in regressions for formal and informal firms. On the other hand, the coefficient on CMA is positive and statistically significant for formal firms, though its magnitude is about one-third lower in comparison to the baseline estimate. Taken together, this suggests that the impact of CMA (a measure of competition) still dominates, but that FMA explains a relevant part of the overall effect of improvements in market access.

6.2. Import competition

It is possible that not accounting for changes in exposure to international trade stemming from improvements in connectivity results in inconsistent estimates of the impact of domestic competition captured by CMA on firm outcomes. Differential changes in CMA may in fact be correlated with changes in access to international markets and hence to import/export opportunities. Exporting is rare among Ethiopian manufacturing firms. On average across the years we consider, only about 4% of firms from the census report a positive value for exporting. To account for the confounding effects of changes in exposure to imports induced by increased connectivity, we introduce an additional control variable in our baseline estimation that captures import competition.

The control variable we generate combines information on changes in travel time from each district i to the district of Afambo in the Afar region, and interacts this variable with data on import flows at the 4-digit ISIC level.[24] This strategy is motivated by the specific configuration on how trade happens in Ethiopia. Most trade comes via the port of Djibouti,[25] and the district of Afambo borders the village of Galafi, the official border crossing from Djibouti to Ethiopia (henceforth we label this variable "Galafi"). It therefore represents a necessary

[20] However, a test comparing the coefficients of columns 1–4, 2–5, 3–6 fails to show that the coefficients of interest are statistically different from each other.

[21] Information on the distribution of bank branches in Ethiopia comes from Limodio and Strobbe (2023). For each branch, they collect information on the year of establishment and its location, which we map onto districts in which firms are located.

[22] Tests comparing the coefficients of columns 1–4, 2–5, 3–6 show that all coefficients of interest are statistically different from one another.

[23] Data on population come from Rose et al. (2020). As in Eq. (1) we use an elasticity of 3.8.

[24] Data on import flows is obtained by the World Bank World Integrated Trade Solution (WITS).

[25] The World Bank estimates that about 93% of total Ethiopian trade transits through the port of Djibouti (World Bank, 2013).

Table 5
Presence of banks in 1997 (Woredas with banks vs. No banks).

	No banks			At least one bank		
	(1) Whole sample	(2) Informal	(3) Formal	(4) Whole sample	(5) Informal	(6) Formal
CMA	0.0160	−0.0365	0.0791***	0.1299***	0.0336***	0.1919***
	(0.0136)	(0.0236)	(0.0176)	(0.0154)	(0.0059)	(0.0172)
Observations	11 973	7808	3394	11 689	6258	5291
R^2	0.042	0.008	0.043	0.100	0.019	0.066
Dist-Ind. FE	Y	Y	Y	Y	Y	Y
Dist-Year FE	Y	Y	Y	Y	Y	Y
Controls	Y	Y	Y	Y	Y	Y
F-stat	7235.366	5483.334	18 633.749	8345.071	5881.158	18 908.719

Notes: The dependent variable is labour productivity at the firm level. The variable of interest is CMA, which varies at the woreda, industry and year level. Results are reported for the whole sample, for the informal and the formal sector, respectively. Firm level control variables include firm age and a binary variable taking value of 1 for all firms observed in the SSIS. Conley standard errors in parenthesis (in column (1) Conley distance is 190 km), *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

passage for most imported goods before reaching their destination in Ethiopia. Measuring changes to the distance to Afambo allows us to account for the increase of import competition that is related to the improvements of the road network. Results of this specification are reported in Table A.18. All the estimations also include a firm level control for exports. Results demonstrate that the qualitative flavour of our baseline results is unchanged, indicating that while improved roads may change the nature of competition firms face from abroad, the impact of domestic competition remains stable. In the same table, we report a similar exercise in which we interact the distance in travel time to Addis with import intensity at the industry level. This is to account for the fact that some imports may transit first to the dry port of Modjo, near the capital, before being shipped elsewhere (Fiorini et al., 2021). Also in this case, results do not change qualitatively.

6.3. Alternative definition of informality

We estimate our baseline regression focusing on labour productivity as a dependent variable using alternative definitions of informality. Our goal here is not only to assess the reliability of our results, but also to investigate whether it is informality per se, or simply firm size that determines differential responses to competition from better connectivity. We consider three alternative scenarios: (1) define as informal firms that do not keep books of accounts, independently of their size; (2) define as informal firms that do not keep books of accounts, but restrict the sample to firms employing less than 10 workers; and (3) define as informal firms that are in the SSIS survey, and as formal firms that are in the LMMS survey. The results are presented in Table A.19 in the Appendix.

From columns (1) and (2), classifying firms as formal based on whether they maintain books of accounts reiterates our previous result in Table 3. Thus, our baseline result to robust to this alternative definition of informality. More importantly, from columns (3) and (4), restricting firms to less than 10 workers does not change the flavour of our main result. If our main result was not about informality and formality, but simply about firm size, we would expect to see no impact of competition from better connectivity on formal firms, given that these formal firms employ less than 10 workers and are classified informal as per the size definition. What these results indicate is that formality is more than just size and includes business practices followed by formal firms, such as maintaining accounts. Finally, results in columns (5) and (6) show that our result is robust to defining informality based on the survey source.

6.4. Alternative trade elasticities

We test the sensitivity of our results to different values of θ, the distance decay parameter used to define CMA. θ captures the non-linear impact of distance on trade, and through its value is likely to be context specific, it is normally estimated in a range going from 1 to 10. Hence, we replicate our results using values of θ equal to 1, 3.124 and 8.22. A value of 1 corresponds to the canonical definition of market potential, as provided originally by Harris (1954). We got to a value of 3.124 if we replicate our baseline scenario using the trade-travel time differential between Ethiopia and the US, that is estimated to be approximately 2.46 times by Atkin and Donaldson (2015). Last, 8.22 is the elasticity estimated in Donaldson and Hornbeck (2016). From Table A.20 in the Appendix, we find that independently of the value of θ, results appear consistent and in line with the baseline. This indicates that infrastructural investments are associated with higher productivity in the formal sector, and not statistically significantly related to productivity in the informal sector.

6.5. Alternative instruments

Alternative base period. We probe whether our results are sensitive to the choice of buffer that we employ in our main instrument. Specifically, we use the 1996 road network to construct the 50 km buffer instead of the 1969 road network. Since many major roads were built between 1969 and 1996, choosing 1996 as the buffer allows us to define an exclusion zone that takes these major roads into account. Table A.21 presents results and shows that they remain qualitatively robust. Further, we employ this same instrument (i.e. with the 50 km buffer around 1996 roads) after excluding those firms that existed prior to 1996, since roads built before 1996 could have targeted these firms. Table A.22 shows that our key results endure.

Endogeneity of production weights. Next, we tackle the potential endogeneity of the production weights employed in our CMA measure. It is in fact possible that changes in the production over time are connected to changes in the road network, affecting our estimation. We proceed as follows. We construct an alternative instrument based on our main approach in which we replace current production weights with pre-sample weights that are specific to each woreda-industry pair. This is not possible using the firm data, given that the earliest wave of the SSIS survey is in 1999, two years after the start of the RSDP. We thus use information on the distribution of sectoral employment at the woreda level from the 1994 population census. The census tracks employment of individuals in their working age (15–64) and provides information on the industry of their occupation. This information is available at the 3 digit level of the ISIC classification. This covers both formal and informal employment, providing a detailed picture of the distribution of economic activity in the country. Results of this new instrumental variable estimation exercise, reported in Table A.23 once again confirm our key findings.[26]

[26] The first stage regression, reported in Table A.24 in the Appendix, shows that the correlation of the instrument with the CMA is smaller compared to the one reported in Table A.9. The CMA weights in these two versions are no longer the same.

Table 6
Alternative IV strategy.

	Whole sample		Informal		Formal	
	(1) Labour Prod.	(2) Labour Prod.	(3) Labour Prod.	(4) Labour Prod.	(5) Labour Prod.	(6) Labour Prod.
CMA	0.0363	0.0439	−0.0613	−0.0689	0.1145***	0.1328***
	(0.0429)	(0.0447)	(0.0519)	(0.0580)	(0.0409)	(0.0397)
Observations	23 662	23 662	14 066	14 066	8685	8685
R^2	0.001	0.078	−0.001	0.012	0.006	0.058
Dist-Ind. FE	Y	Y	Y	Y	Y	Y
Dist-Year FE	Y	Y	Y	Y	Y	Y
Controls	N	Y	N	Y	N	Y
F-stat	70.988	71.096	60.800	59.920	58.226	58.532

Notes: The dependent variable is labour productivity at the firm level. The variable of interest is CMA, which varies at the woreda, industry and year level. Results are reported for the whole sample, for the informal and the formal sector, respectively. Firm level control variables include firm age and a binary variable taking value of 1 for all firms observed in the SSIS. Conley standard errors in parenthesis * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Colonial roads. We employ an alternative instrumental variable strategy based on the work by Moneke (2020) to address further aspects of the endogeneity of CMA. We digitize the historical Italian colonial road network from Gli Annali dell'Africa Italiana (1937-1943) and calculate the Euclidean distance to Italian colonial roads for the centroid of each district. Next, we distribute the entire length of roads under the RSDP to districts, subject to the constraint of connecting them to the road network by the end of the sample period using an artificial regional budgeting algorithm. In particular, we sort district centroids on the basis of their proximity to the digitized colonial roads and gradually connect them until the annual mileage per region of road construction has been achieved. The algorithm operates with a least-cost logic: districts closer to the road network at time t get connected at time $t + 1$ first, until the annual regional budget is exhausted. For every subsequent year, we connect each district's centroid to its closest Italian artery, or with the artificial roads constructed at the previous step whichever is closer. In this way, districts far away from the colonial arteries will be connected at later time periods in the synthetic network, as shown in Figure A.9 in the Appendix.

The way RSDP road improvements enter our CMA measure is through changes in travel time. Hence, differently from Moneke (2020), the spatial and temporal variation in our artificial road network is obtained by allocating a speed increment (equal to 35 km/h) to each artificially added road segment.[27] Therefore, for every subsequent year, each newly connected district is assigned a travel speed that it will maintain in future time periods. The temporal variation then comes from the change in travel time that is experienced by each road segment artificially added to the Italian colonial artery. We thus generate a time-varying instrument that derives its exogenous variation from the straight line distance to Italy's digitized colonial road map. Results from all regression employing this additional instrument, reported in Table 6, remain qualitatively similar.

6.6. Alternative specifications

Industry-year fixed effects. To control for unobserved, time-varying shocks such as technological growth that are specific to industries, we include industry-year fixed effects in our baseline estimation. The advantage of this more rigorous specification is that it allows us to rule out an alternative explanation for our results. Suppose that greater exposure to firms operating in the same industry through improved connectivity results in industries coordinating lobbying efforts. An outcome of this process is that industries may receive greater protection over time, which will be captured by these fixed effects. As shown in Table A.25, results retain their qualitative flavour.

[27] We assign 35 km/h as the speed increment since it represents the average travel speed of major gravel roads before the upgrading of the RSDP.

Additional controls. We account for factors that may underlie firm performance by including control variables for the total number of firms and total employment at the district-industry level. These are important control, as they can approximate any agglomeration forces driven by changes in roads. Results are presented in Table A.26 and demonstrate that our results are qualitatively robust to the addition of these additional control variables.

Analysis at market level. We also conduct our baseline analysis at the market level, instead of at the firm level, to look at the impacts of CMA on aggregate market productivity. We achieve this by collapsing the firm level data to the market (woreda-industry) level and re-estimating our baseline regressions. Results are presented in Table A.27. The fundamental insights carry through in the aggregate. Specifically, an increase in CMA is associated with an increase in labour productivity, but only in the formal sector.

Alternative clustering. In Table A.28 we provide estimates based on Conley standard errors based on buffers of 100 or 300 km, respectively. We find no qualitative changes to the main results. Table A.29 adds that our results are also qualitatively robust to clustering at the woreda and at the woreda-industry level, respectively.

7. Conclusions

This study connects two areas of high priority for public policy: the informal sector, which is a pervasive feature of the developing world and plays a central role in structural transformation; and infrastructure investments, which are critical drivers of economic growth. Understanding the responses of the formal and informal sectors to large infrastructure projects is key for policy design and implementation. Although governmental provision of public goods should be seen as productivity-enhancing, it is plausible that these benefits vary significantly across formal and informal firms. Ignoring such heterogeneity in the response of firms to road infrastructure improvements can create distortions in resource allocation.

We find that increased competition resulting from a decrease in intranational trade cost as connectivity improves from road improvements is associated with higher productivity and wages in the formal sector. This is contrary to what we find for the informal sector, where labour productivity, if anything, decreases with greater competition. We also find opposite effects of an increase in competition on capital-labour ratios and investments, with positive (negative) effects for the formal (informal) sector. Greater competition is also associated with a lower likelihood of a firm operating informally and of informal firms employing a greater share of workers with less than primary education. We thus highlight that the benefits of road infrastructure improvements accrue unevenly across formal and informal sectors in developing countries and that mitigating strategies might be called for to address these differential gains.

Our paper can be seen as a first attempt to shed light on the differential impacts of improved infrastructure on the performance of the formal and informal sectors in developing countries. We call for further data on firms in the informal sector, especially of a panel nature, to further probe the dynamics of firm adjustments to large infrastructure programmes.

CRediT authorship contribution statement

Elena Perra: Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Marco Sanfilippo:** Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Asha Sundaram:** Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Data availability

Some of the data were obtained after signing a confidential agreement with the Ethiopian Statistica Agency. We can share the data with authors that will sign the NDA themselves.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jdeveco.2024.103339.

References

Adamopoulos, Tasso, 2019. Spatial integration, agricultural productivity, and development: A quantitative analysis of Ethiopia's road expansion program. In: 2019 Meeting Papers. Society for Economic Dynamics.

Alder, Simon, Croke, Kevin, Duhaut, Alice, Marty, Robert Andrew, Vaisey, Ariana Brynn, 2022. The Impact of Ethiopia's Road Investment Program on Economic Development and Land Use : Evidence from Satellite Data. Policy Research Working Paper 10000, The World Bank.

Asher, Sam, Novosad, Paul, 2020. Rural roads and local economic development. Amer. Econ. Rev. 110 (3), 797–823.

Atkin, D., Donaldson, D., 2015. Who's Getting Globalized? The Size and Implications of Intra-National Trade Costs. NBER Working Paper N. 21439.

Baker, J., 1974. Developments in Ethiopia's road system. Geography 59 (2), 150–154.

Bank, World, 2021. Ethiopia-road sector development program APL4.

Becker, Dennis, 2018. Heterogeneous firms and informality: the effects of trade liberalization on labour markets. Oxf. Econ. Pap. 70 (1), 47–72.

Bertazzini, Mattia, 2022. The long-term impact of Italian colonial roads in the horn of africa, 1935–2000. J. Econ. Geogr. 22 (1), 181–214.

Burgess, Robin, Jedwab, Remi, Miguel, Edward, Morjaria, Ameet, Padró i Miquel, Gerard, 2015. The value of democracy: Evidence from road building in Kenya. Amer. Econ. Rev. 105 (6), 1817–1851.

Chandra, Amitabh, Thompson, Eric, 2000. Does public infrastructure affect economic activity? Evidence from the rural interstate highway system. Reg. Sci. Urban Econ. 30 (4), 457–490.

Chatterjee, Santanu, Lebesmuehlbacher, Thomas, Narayanan, Abhinav, 2021. How productive is public investment? Evidence from formal and informal production in India. J. Dev. Econ. 151, 102625.

Chaurey, Ritam, Le, Duong Trung, 2022. Infrastructure maintenance and rural economic activity: Evidence from India. J. Public Econ. 214, 104725.

Colella, Fabrizio, Lalive, Rafael, Sakalli, Seyhun Orcan, Thoenig, Mathias, 2019. Inference with Aribitrary Clustering. IZA DP No. 12584, IZA Institute of Labour Economics.

De Loecker, Jan, Goldberg, Pinelopi Koujianou, 2013. Firm Performance in a Global Market. NBER Working Paper 19308.

De Loecker, Jan, Warzynski, Frederic, 2012. Markups and firm-level export status. Amer. Econ. Rev. 102 (6), 2437–2471.

Desmet, Klaus, Parente, Stephen L., 2010. Bigger is better: Market size, demand elasticity, and innovation. Internat. Econom. Rev. 51 (2).

Diao, Xinshen, Ellis, Mia, McMillan, Margaret S., Rodrik, Dani, 2021. Africa's Manufacturing Puzzle: Evidence from Tanzanian and Ethiopian Firms. Working Paper 28344, National Bureau of Economic Research.

Dix-Carneiro, Rafael, Goldberg, Pinelopi K., Meghir, Costas, Ulyssea, Gabriel, 2021. Trade and Informality in the Presence of Labor Market Frictions and Regulations. NBER Working Papers 28391.

Donaldson, Dave, 2018. Railroads of the Raj: Estimating the impact of transportation infrastructure. Amer. Econ. Rev. 108 (4–5), 899–934.

Donaldson, Dave, Hornbeck, Richard, 2016. Railroads and American economic growth: A market access approach. Q. J. Econ. 131 (2), 799–858.

Duflo, Esther, Pande, Rohini, 2007. Dams. Q. J. Econ. 122 (2), 601–646.

Duranton, Gilles, Morrow, Peter M., Turner, Matthew A., 2014. Roads and trade: Evidence from the US. Rev. Econ. Stud. 81 (2 (287)), 681–724.

Duranton, Gilles, Turner, Matthew A., 2012. Urban growth and transportation. Rev. Econ. Stud. 79 (4), 1407–1440.

ERA, 2009. SDP Performance: 12 Years Later. Tech. Rep., Ethiopian Road Authority (ERA).

ERA, 2016. Road Sector Development Program 19 Years Performance Assessment. Tech. Rep., Ethiopian Road Authority (ERA).

Erten, Bilge, Leight, Jessica, Tregenna, Fiona, 2019. Trade liberalization and local labor market adjustment in south africa. J. Int. Econ. 118, 448–467.

Fiorini, Matteo, Sanfilippo, Marco, 2022. Roads and jobs in ethiopia. World Bank Econ. Rev. 36 (4), 999–1020.

Fiorini, Matteo, Sanfilippo, Marco, Sundaram, Asha, 2021. Trade liberalization, roads and firm productivity. J. Dev. Econ. 153, 102712.

Gebresilasse, Mesay M., 2023. Rural roads, agricultural extension, and productivity. J. Dev. Econ. 162.

Gli Annali dell'Africa Italiana, 1937-1943. Piccioli, Angelo. Tech. Rep., Mondadori.

Goldberg, Pinelopi, Pavcnik, Nina, 2003. The response of the informal sector to trade liberalization. J. Dev. Econ. 72 (2), 463–496.

Harris, Chauncy D.H., 1954. The market as a factor in the localization of industry in the united states. Ann. Assoc. Am. Geogr. 44 (4), 315–348.

Huang, Yuxiao, Xiong, Wentao, 2018. Geographic Distribution of Firm Productivity and Production: A Market Access Approach (Harvard PhD dissertation).

Jedwab, Rémi, Storeygard, Adam, 2021. The average and heterogeneous effects of transportation investments: Evidence from sub-saharan africa 1960–2010. J. Eur. Econom. Assoc. 20 (1), 1–38.

Kruse, Hagen, Mensah, Emmanuel, Sen, Kunal, De Vries, Gaaitzen, 2022. A manufacturing renaissance? Industrialization trends in the developing world. IMF Econ. Rev..

La Porta, Rafael, Shleifer, Andrei, 2014. Informality and development. J. Econ. Perspect. 28 (3), 109–126.

Limodio, Nicola, Strobbe, Francesco, 2023. Liquidity requirements, bank deposits and financial development. Economica 90 (357), 240–270.

McCaig, Brian, Pavcnik, Nina, 2018. Export markets and labor allocation in a low-income country. Amer. Econ. Rev. 108 (7), 1899–1941.

McMillan, Margaret S., McCaig, Brian, 2020. Trade liberalization and labor market adjustment in Botswana. J. Afr. Econ. 29, 236–270.

Melitz, Marc, 2003. The impact of trade on intra-industry re-allocation and aggregate industrial productivity. Econometrica 71, 1695–1725.

Melitz, Marc J., Ottaviano, Gianmarco I.P., 2008. Market size, trade, and productivity. Rev. Econ. Stud. 75 (1), 295–316.

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

Nataraj, Shanthi, 2011. The impact of trade liberalization on productivity: Evidence from India's formal and informal manufacturing sectors. J. Int. Econ. 85 (2), 292–301.

Perra, Elena, 2022. Road to Division: Ethnic Favouritism in the Provision of Road Infrastructure in Ethiopia. Working Papers - Economics, Universita' degli Studi di Firenze, Dipartimento di Scienze per l'Economia e l'Impresa.

Redding, Stephen J., 2020. Trade and Geography. Working Paper 27821, National Bureau of Economic Research.

Rose, Amy, McKee, Jacob, Sims, Kelly, Bright, Edward, Reith, Andrew, Urban, Marie, 2020. LandScan Global 2020. Oak Ridge National Laboratory, Oak Ridge, TN.

Shiferaw, Admasu, Söderbom, Måns, Siba, Eyerusalem, Alemu, Getnet, 2015. Road infrastructure and enterprise dynamics in Ethiopia. J. Dev. Stud. 51 (11), 1541–1558.

Siba, Eyerusalem, 2015. Returns to physical capital in ethiopia: Comparative analysis of formal and informal firms. World Dev. 68.

Topalova, Petia, Khandelwal, Amit, 2011. Trade liberalization and firm productivity: The case of India. Rev. Econ. Stat. 93 (3), 995–1009.

Ulyssea, Gabriel, 2018. Firms, informality, and development: Theory and evidence from Brazil. Amer. Econ. Rev. 108 (8), 2015–2047.

United States Central Intelligence Agency (CIA), 1969. Map of Ethiopia. Library of Congress Geography and Map Division Washington, D.C. 20540-4650 USA dcu, Retrieved from the Library of Congress.

Worku, Ibrahim, 2011. Road Sector Development and Economic Growth in Ethiopia. In: EDRI Working Paper 004: Ethiopia Support Strategy Program II – International Food Policy Research Institute, Ethiopian Development Research Institute, Addis Ababa, Ethiopia.

World Bank, 2013. Transport and Logistics in Djibouti: Contribution to Job Creation and Economic Diversification. Tech. Rep., World Bank, Washington, DC..