Department of Economics and Statistics "Cognetti de Martiis"
Campus Luigi Einaudi, Lungo Dora Siena 100/A, 10153 Torino (Italy)
www.est.unito.it

Dipartimento Economia e Statistica Cognetti de Martiis

Working Paper Series

21/22

ROADS, COMPETITION, AND THE INFORMAL SECTOR

ELENA PERRA, MARCO SANFILIPPO and ASHA SUNDARAM

UNIVERSITÀ DEGLI STUDI DI TORINO

The Department of Economics and Statistics "Cognetti de Martiis" publishes research papers authored by members and guests of the Department and of its research centers. ISSN: 2039-4004

Roads, Competition, and the Informal Sector

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

[a]*Department of Economics, University of Florence,*
[b]*Department of Economics and Statistics, University of Turin, and Collegio Carlo Alberto,*
[c]*Department of Economics, University of Auckland,*

December 4, 2022

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, 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.

Keywords: Infrastructure, Informal Sector, Productivity, Ethiopia
JEL: O18, O17, O55

*We would like to thank 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) and the Annual Conference of the Italian Economic Association in Turin for useful comments on early draft versions of the paper. Perra: `elena.perra@unifi.it`; Sanfilippo: `marco.sanfilippo@unito.it`; Sundaram: `a.sundaram@auckland.ac.nz`.

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 deferentially 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.

Furthermore, 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 (RSDP) with firm level data from the formal and informal sectors of Ethiopia. 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 infrastructure 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, both in its local and in connected markets 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. Of course, as travel times decrease with the improving road network, firms will also have better access to consumers in connected markets (an improvement in the firm's market-access (FMA), following Donaldson (2018). We account for this improved FMA with fixed effects that vary by district and year. Arguably, this effect is uniform across industries, so that 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. We discuss each of the two components of our CMA measure: production and travel time. We argue that production in connected markets is largely exogenous to local firm outcomes, given that each district is a small economy relative to all other Ethiopian districts. Travel time given the road network is more likely to be endogenous. As reported by Gebresilasse (2020), Moneke (2020),

[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.

and 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). 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.

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 an alternative instrumental variable employed by Moneke (2020) and Gebresilasse (2020). During Ethiopia's colonization, 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 Ethiopia's colonial road map 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 colonial roads and a budgeting algorithm whereby districts are sequentially connected to the road network with the one closest to the colonial artery connected first. The instrument is therefore less susceptible to endogenous economic and political considerations driving public investments in road infrastructure.

We find that the impacts of competition, measured by CMA, vary 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 big 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 25^{th} percentile of the CMA distribution in that sector) to the district of Akaki-Kalit in Addis (75^{th}) 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. Additionally, the likelihood of operating informally decreases with an increase in CMA.

These findings are consistent with a framework where firms in the formal sector compete nationally, while the informal sector operates locally. 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, the increase in demand for capital in the formal sector drives up the cost of capital for informal firms, lowering capital input, capital intensity, investments in physical capital and labor productivity in the informal sector. However, exit of less productive firms in the informal sector also means that surviving informal firms are of higher productivity, countering the negative effect to a certain extent. 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, 2020) 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 are in line with 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 organised as follows: Section 2 outlines a conceptual framework, Section 3 provides overview on the context of the Ethiopian RSDP; Section 4 describes the data employed in the analysis and outlines the empirical approach adopted for this study; Section 5 reports the results, with robustness checks discussed in Section 6. Section 7 concludes.

2. Conceptual Framework

We describe a conceptual framework to trace out the impact of competition from better connectivity to intranational markets on formal and informal firms. The economy is composed of two sectors, formal and informal. Consumer utility is two-tiered, with upper-tier utility being Cobb-Douglas with expenditure shares μ_I and μ_F for the informal and formal good respectively. For simplicity, we focus on the informal sector, where we assume that second tier utility is of the Dixit-Stigliz type with constant elasticity of substitution between differentiated product varieties given by σ^I. Firms in the informal sector compete under monopolistic competition[2]. We assume that the informal sector is local, while goods produced by the formal sector are traded nationally[3].

Demand for an informal firm producing variety x_I is given by

$$x_I = \frac{p_I^{-\sigma^I}}{P_I^{1-\sigma_I}} \mu_I M \tag{1}$$

where p_I is the price of the product variety, P_I is the aggregate price of the informal good and M is aggregate income. The firm uses capital and labor to produce output. The production function is Cobb-Douglas, with constant returns to scale as follows

[2]We posit that informal firms produce products differentiated by quality, product attributes or even along the spatial dimension in the presence of high commuting costs.

[3]This assumption is supported in our data. The two most recent waves of the SSIS suvery that we use to identify informal firms ask firms whether they sell domestically and/or abroad and, in the former case, if their market is mostly local or national. More than 99% of the firms sell within Ethiopia. Out of them, slightly more than 91% declare that their market is local, and the rest also sell elsewhere in the country. The LMMS data do not provide information on the geographic scope of the national market for firms with more than 10 persons engaged.

$$x_I = K^{1/2}L^{1/2} \tag{2}$$

Profit maximization is given by

$$\max_{x_I} \Pi = (\frac{\mu_I M}{P_I^{1-\sigma_I}})^{1/\sigma_I} x_I^{(\sigma-1)/\sigma} - 2(r_I w_I)^{1/2} x_I \tag{3}$$

where r_I is the cost of capital for informal firms and w_I is the informal wage. We assume that this is positively associated with the cost of capital for formal firms r_F. Specifically, formal and informal firms operate in different but related capital markets. This assumption is not unreasonable. Informal firms typically resort to informal sources of credit since they lack both the collateral and the networks necessary to access formal credit such as bank loans.

Solving the profit maximization problem gives optimal output as

$$x_I^* = (2(r_I w_I)^{1/2})^{-\sigma} (\frac{\sigma_I - 1}{\sigma_I})^{\sigma_I} (\frac{\mu_I M}{P_I}) \tag{4}$$

Optimal capital, labor and the capital-labor ratio are given by

$$L = (\frac{r_I^{1-\sigma}}{w_I^{1+\sigma}})^{1/2} (1/2^\sigma) (\frac{\sigma_I - 1}{\sigma_I})^{\sigma_I} (\frac{\mu_I M}{P_I}) \tag{5}$$

$$K = (\frac{w_I^{1-\sigma}}{r_I^{1+\sigma}})^{1/2} (1/2^\sigma) (\frac{\sigma_I - 1}{\sigma_I})^{\sigma_I} (\frac{\mu_I M}{P_I}) \tag{6}$$

$$\frac{K}{L} = \frac{w_I}{r_I} \tag{7}$$

Following Desmet and Parente (2010), we posit that in the formal sector firms choose to undertake investments in physical capital at the cost of r_F. Intuitively, we conceptualize this investment as a fixed cost of technological upgrading or a cost that is necessary to support innovative activity. We assume that the investment lowers the marginal cost of production.

We now consider the impact of an increase in competition resulting from lower trade costs due to better connectivity from improved road infrastructure. Given our simplifying assumption, competition impacts the formal sector whose products are traded intranationally. As shown by Desmet and Parente (2010), an increase in competition results in an increase in the elasticity of demand for each formal product variety. This leads to a decrease in product varieties produced, a decrease in firm markups and an increase in firm size. The increase in firm size means that the effect on profits when the marginal cost drops from capital investments is larger, making such investments in formal capital more attractive. In a capital constrained environment, the increase in demand for formal capital drives up the cost of capital in both the formal and informal sectors (r_I and r_F).

From equation (4), an increase in r_I is associated with a decrease in production x_I. From, equations (5), (6) and (7), labor, capital and the capital-labor ratio decrease, with the decrease in labor more attenuated than the decrease in capital (since we expect some substitution away from capital to labor). To summarize, our hypotheses are that an increase in competition from better road connectivity is associated with greater output, capital investments (captured by capital input, capital intensity and investments in physical capital) in the formal sector and lower output, capital input, capital intensity and investments in the informal sector. If labor productivity is positively related to capital intensity and investments in physical capital, we expect competition to be positively (negatively) associated with labor productivity in the formal (informal) sector.

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. 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 and the least productive informal firms exiting production.[4] This exit channel would raise aggregate productivity in the formal and informal sectors. In the informal sector, it would work against the capital investment channel we discuss in this section, thereby yielding ambiguous empirical results on labor productivity. In the next sections, we explore a range of outcome variables to examine these channels.

3. The Road Sector Development Programme

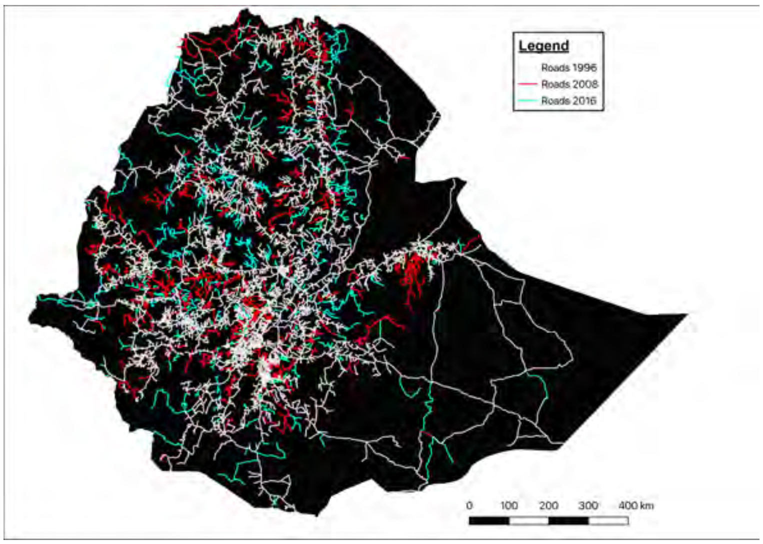
The Road Sector Development Programme (herafter, RSDP) is an ambitious investment programme implemented in Ethiopia since 1997 and still ongoing, with the objectives of rehabilitating existing Ethiopian roads and constructing new networks. The Ethiopian road network has increased from 26550 km in 1997 to 113066 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 (Ethiopian Road Authority, 2016; World 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

[4] Table 7 shows evidence for this channel by establishing that the likelihood of operating in the informal sector declines with better road connectivity. This result is consistent with exit from the informal sector dominating entry from the formal to the informal sector.

in foreign currency, with an estimated cost over the fourteen years of around US$7.08 billion (Worku, 2011; Shiferaw *et al.*, 2015). 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).

Figure 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. (Source: Authors' calculation on proprietary road data).

Road improvements under the RSDP are documented in Figure 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 road. The data on road quality improvements are aggregated in the speed matrix proposed by ERA (2009) and reported in Table 1, which describes the average travel speed as a function of the particular road surface.

Table 1: ERA Travel Speed Matrix

	Average Travel Speed	
Pavement Type	Before Upgrading	After Upgrading
Asphalt Roads	50 km/h	70 km/h
Major Gravel Roads	35 km/h	50 km/h
Minor Gravel Roads	25 km/h	45 km/h
Earth Surfaced Roads	20 km/h	30 km/h

Source: Authors' calculation based on ERA (2009).

4. Data

4.1. 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) 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.2 reports precise 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.[5]

[5] See Table A.3 in the Appendix for an overview of the role played by the small business sector in Ethiopia.

We combine the two datasets (LMMS and SSIS), obtaining information at the firm, industry and *woreda* level. Note that 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[6]. 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).

From Table 2, and in line with expectations, an informal firm appears to be on average smaller, younger, less capital intensive and less export/import oriented compared to a formal firm. This is reflected in Figure 2, which shows the productivity[7] distributions of formal and informal firms. The wider right tail of the formal firm distribution 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.1 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.

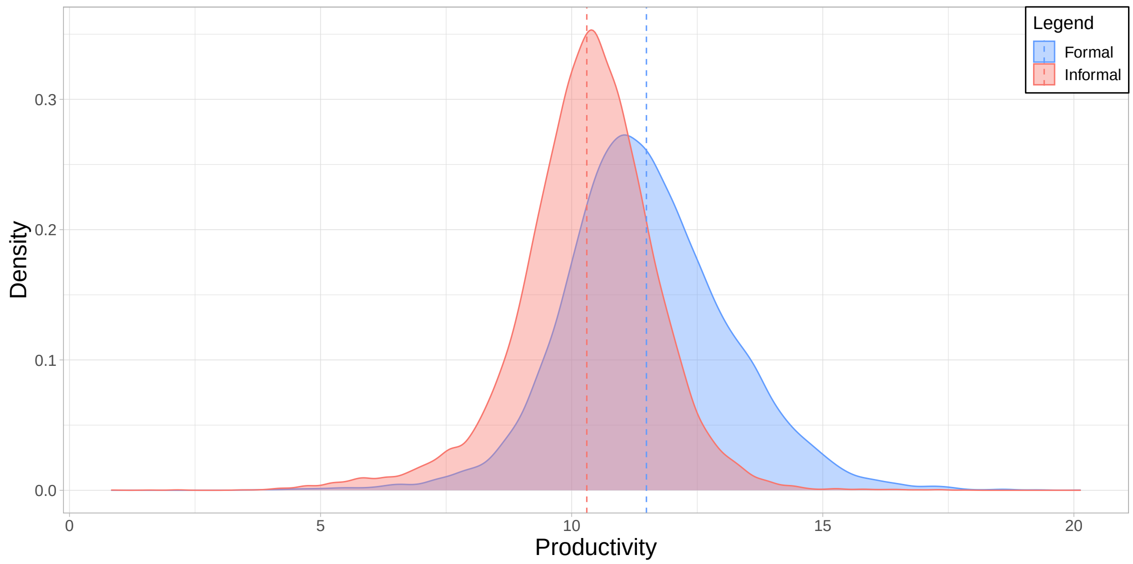
Table 2: Summary Statistics

	Informal	Formal
Variable	Mean	Mean
Labour Productivity	87.2	599.8
Capital Intensity	39.4	191.7
Capital	103.6	12030.5
Investment	23.3	1908.8
Employees	3.6	50.9
Wage_{pc}	7.3	28.0
Age	5.9	9.4

Notes: Values are expressed in thousands ETB, except for Employees and Age. Labour Productivity is calculated as the ratio between the value of production and the number of employees; Capital is measured as the book value of fixed assets at the beginning of the period; Capital intensity is the ratio of capital on assets; Wage per capita is calculated dividing the total wage bill of the firm by the number of employees; Investment includes the value of expenditures to buy fixed assets; Age measures the number of years since the firm was established.

[6] In Section 6.1 we also investigate alternative definitions of informality as a robustness check.
[7] Productivity is defined as the ratio of production to the number of employees.

Figure 2: Productivity Distribution for formal and informal firms



Notes: Productivity is calculated as the simple average across formal and informal firms over the whole sample period. (Source: Authors' calculation on LMMS and SSIS data.)

4.2. 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 categorised 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 dimension of market access, which we label "Consumer

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

Market Access (CMA)". CMA accounts for how proximity to competitors 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 market access, in which changes in travel time τ are weighted using the level of total production in each woreda-industry pair jx at time t:

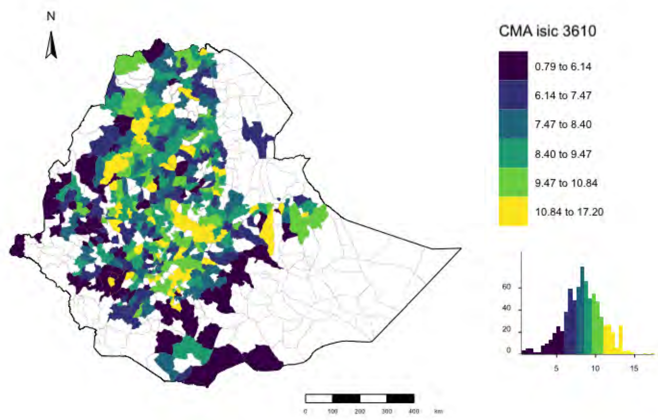
$$CMA_{jxt} = \sum_{d \neq x} Production_{jdt} * \tau_{jxdt}^{-\theta} \tag{8}$$

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. 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[9]. In robustness checks, reported in Section 6.2, we show that results are robust to different values of θ.

Figure 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). We show also that CMA is a valid proxy for competition. Figure A.2 in the Appendix plots the correlation of CMA with the Herfindahl-Hirschman Index (HHI) computed at the level of each market. The figure shows that higher levels of CMA do indeed correspond to more competitive, i.e. less concentrated, markets.

[9]Duranton *et al.* (2014) adopt a cost-distance elasticity for the US of 1.27, meaning that $1.27 * 3 \approx 3.8$

Figure 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 on SSIS and LMMS data.)

4.3. Identification Strategy

Our empirical analysis is based on estimating the following relation:

$$Y_{ijxt} = \beta_1 CMA_{jxt} + \beta_2 X_{it} + \theta_{jx} + \mu_{rt} + \epsilon_{ijxt} \tag{9}$$

which relates CMA to indicators of performance (Y_{ijxt}) 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 industry-woreda (jx) fixed effects. Woreda-year (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 acounting for whether the firm was originally surveyed in the SSIS or in the LMMS[10]. Standard errors are clustered at the market (district-industry) level.

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

[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.

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 (2).

By digitizing the CIA's map of Ethiopia (United States Central Intelligence Agency (CIA), 1969)[11], 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 digitised roads as of 1969. We then define exclusion areas, where the road network will remain frozen, exactly as it was before the RSDP (in 1996). Outside the exclusion zones, the road network changes over time according to RSDP improvements. As shown in Figure 4, the 50 km buffer includes all major city centers 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 buffer at 1996 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 over time[13]. Our approach thus enables us to account for not only endogenous

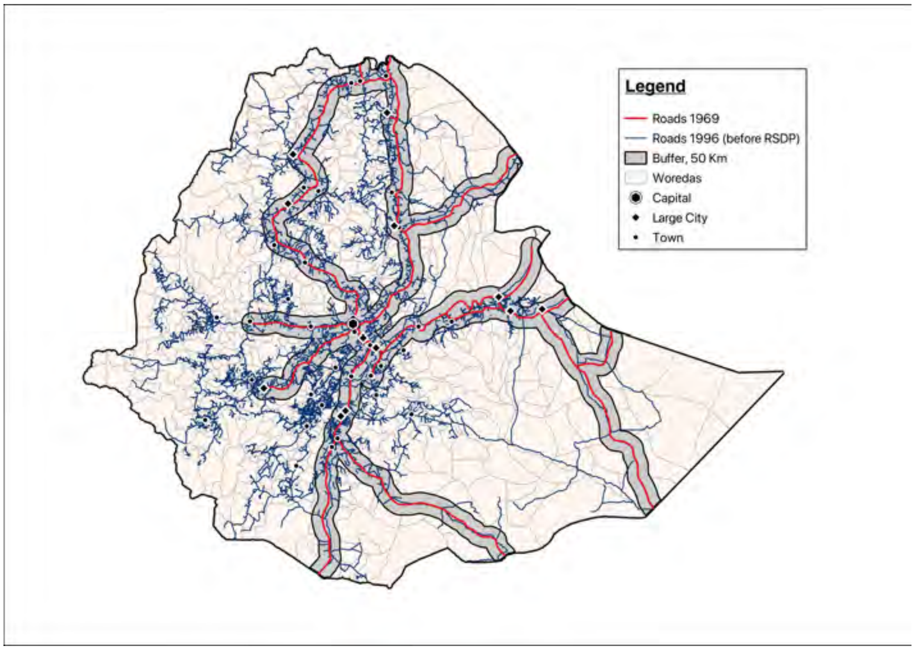
[11] Reported in Figure A.3 in the Appendix

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

[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 7,000 km of roads, 3,450 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

local road improvements, but also for those improvements targeted at connecting principal Ethiopian economic hubs and cities.

Figure 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 (Source: Authors' calculation on United States Central Intelligence Agency (CIA) (1969) and proprietary data on the road network).

5. Results

5.1. Formal and informal firms productivity

In Table A.5 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. 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

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 7,304 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).

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.[14]

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 subsection 4.3. 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.6 and shows that our instrument is a strong predictor of CMA. Similarly, the first stage F-statistic appears to be strong, confirming the relevance of our instrumental variable. It is also reassuring to see that results of the reduced-form (reported in Table A.9 in the Appendix) remain consistent. 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, 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).

5.2. Heterogeneous Effects

The results presented in Section 5.1 point to no significant impact of competition on firms in the informal sector. However, we expect heterogeneity in response to competition among informal firms as compared to the group of more structured, formal firms. In this section, we explore potential heterogeneity along two different dimensions. First, by distinguishing firms operating in markets with high vs low degree of informality. Second, by showing results across size cohorts of informal firms.

Size of the informal sector. In Appendix Table A.7 and A.8, we explore heterogeneous effects across districts with above and below median share of informal firms on the total num-

[14] 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 equation 8 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.12. There is no significant relationship between this altered CMA measure and productivity, suggesting that changes in industrial activity are a crucial determinant of the relationship of interest.

Table 3: Results IV: Labour Productivity

	WHOLE SAMPLE		INFORMAL		FORMAL	
VARIABLES	(1) Labor Prod.	(2) Labor Prod.	(3) Labor Prod.	(4) Labor Prod.	(5) Labor Prod.	(6) Labor Prod.
CMA	0.0781***	0.0758***	0.00311	-0.00263	0.144***	0.145***
	(0.0290)	(0.0217)	(0.0240)	(0.0241)	(0.0311)	(0.0269)
Obs.	23,662	23,662	14,066	14,066	8,685	8,685
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-test	3298	3292	3188	3189	2533	2524
Mean DV	10.78	10.78	10.30	10.30	11.49	11.49

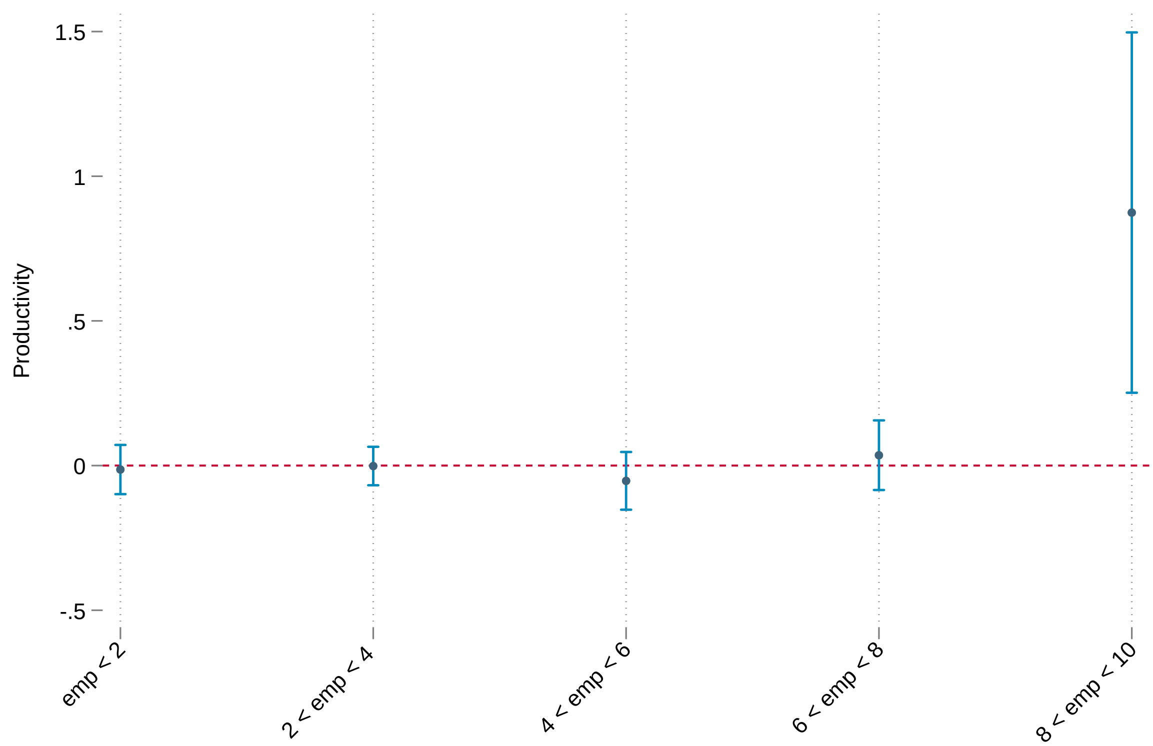
Notes: The dependent variable is labor productivity at the firm level, while the independent variable is CMA, which varies at the woreda, industry and year level. Results are reported for the whole sample, for the informal and the formal sector, respectively. Firm level control variables include firm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

ber 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.7 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 areas with below median share of informal firms. From Table A.8, 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.7 and A.8 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. Figure 5 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.

Figure 5: Productivity for different informal employment cohorts.



Notes: The figure reports the point estimate for the coefficient of interest (CMA) based on separate regressions for sub-samples of informal firms according to bins of size cohorts. The full set of results is presented in Table A.13 in the Appendix.

5.3. Mechanisms

In this section, we explore channels through which competition arising from better connectivity due to improvements in the road network impact labour productivity in the formal and informal sectors. We do this by looking at how firms respond to changes in competition by adjusting their capital stock, as well as production and employment. We finally look at the effects of CMA on the composition of manufacturing in terms of formal and informal firms.

5.3.1. Capital and employment

In light of our conceptual framework, 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, as shown in Table 5. 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. We can do this only for the sub-sample of informal firms covered by the SSIS[15], since this dataset includes a module in which firms report characteristics for each worker (including any working owner), including their level of education. Table 6 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[16]. 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 seen in column (2) of Table 6.

[15] 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.

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

Table 4: Capital Intensity, Capital and Investment

	WHOLE SAMPLE			INFORMAL			FORMAL		
VARIABLES	(1) Capital Intensity	(2) Capital	(3) Investment	(4) Capital Intensity	(5) Capital	(6) Investment	(7) Capital Intensity	(8) Capital	(9) Investment
CMA	0.0991***	0.166***	0.129**	-0.0568*	-0.0655**	-0.205***	0.182***	0.273***	0.310***
	(0.0281)	(0.0349)	(0.0613)	(0.0294)	(0.0300)	(0.0707)	(0.0405)	(0.0463)	(0.0934)
Obs.	20,606	20,606	23,662	11,672	11,672	14,066	8,059	8,059	8,685
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-test	3075	3075	3292	3083	3083	3189	2413	2413	2524

Notes: The dependent variables are the value of capital intensity (in log), 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. All regressions include woreda-industry and woreda-time fixed effects. Robust standard errors in parentheses are clustered at the woreda-industry level *** p<0.01, ** p<0.05, * p<0.1.

Table 5: Production, Employment, Wage per capita

	WHOLE SAMPLE			INFORMAL			FORMAL		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
VARIABLES	Production	Employment	Wage per capita	Production	Employment	Wage per capita	Production	Employment	Wage per capita
CMA	0.141***	0.0656***	0.0471***	-0.00679	-0.00417	-0.00468	0.236***	0.0909***	0.0622***
	(0.0280)	(0.0140)	(0.0153)	(0.0274)	(0.00921)	(0.0198)	(0.0349)	(0.0205)	(0.0187)
Observations	23,662	23,662	15,430	14,066	14,066	7,178	8,685	8,685	7,486
R-squared	0.314	0.456	0.126	0.029	0.033	0.021	0.296	0.397	0.089
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-test	3292	3292	2657	3189	3189	2233	2524	2524	2735

Notes: The dependent variables are the total value of Production (in log), employment (the log number of employees) and the value of per capita wages (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 firms in the SSIS. All regressions include woreda-industry and woreda-time fixed effects. Robust standard errors in parentheses are clustered at the woreda-industry level *** p<0.01, ** p<0.05, * p<0.1.

Table 6: CMA and the composition of employment in informal firms

VARIABLES	(1) Share of workers without EDU	(2) Share of workers with primary EDU	(3) Share of workers with secondary and above
CMA	0.00899**	-0.00753	-0.0117*
	(0.00359)	(0.00774)	(0.00708)
Observations	13,695	13,695	13,695
R-squared	0.003	0.000	0.001
District-Industry FE	Y	Y	Y
District-year FE	Y	Y	Y
Controls	Y	Y	Y
F-test	3282	3282	3282

Notes: The dependent variables are the share of workers without education, with primary education and with secondary education or beyond as a share of total. The regressions are run for a sub-sample of informal firms included in the SSIS. CMA varies at the woreda, industry and year level. Control variable include the age of the firms in each specification. All regressions include woreda-industry and woreda-time fixed effects. Robust standard errors in parentheses are clustered at the woreda-industry level *** p<0.01, ** p<0.05, * p<0.1.

5.3.2. Market composition

An increase in competition could lead to a restructuring of markets in which formal and informal firms compete. On the one hand, less productive formal firms may move into the informal sector whereas less productive informal firms may leave the market. However, given that informal sector firms tend to be less productive than formal firms, we expect greater exit in the informal sector. Overall, we expect an increase in competition to be associated with a formalization of firms. We explore this empirically following Nataraj (2011). We construct a binary variable that takes a value of one if the firm is in the informal sector and zero otherwise. We then estimate the relationship between CMA and this indicator of informality using the instrumental variable estimation strategy. Results are presented in Table 7. Column (1) includes no control variables, while columns (2) and (3) introduce some firm controls sequentially. Results show that the coefficient on CMA is negative and statistically significant. Thus, there is evidence that greater competition resulting from better connectivity to other intra-national 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 coefficients of interest are 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.

Table 7: Effects of CMA on Informality

VARIABLES	(1) Informal	(2) Informal	(3) Informal
CMA	-0.0145	-0.0136**	-0.0114*
	(0.0104)	(0.00644)	(0.00636)
Observations	23,662	23,662	23,662
R-squared	0.001	0.242	0.249
District-Industry FE	Y	Y	Y
District-year FE	Y	Y	Y
Controls	N	Y	Y
F-test	3298	3292	3292

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; in column 3 we also add labor productivity as a control. All regressions include woreda-industry and woreda-time fixed effects. Robust standard errors in parentheses are clustered at the woreda-industry level *** p<0.01, ** p<0.05, * p<0.1.

6. Robustness checks

In this section, we test robustness of our baseline findings to an alternative definition of informality, varying trade elasticity numbers and an alternate instrumental variable. As we show in the following sections, our results remain qualitatively robust.

6.1. Alternative Definition of Informality

In order to assess the reliability of our results, we conduct our baseline analysis to check if results are sensitive to different definitions of informality. In this exercise, we consider the following two alternative definitions: (1) firms that do not keep books of accounts, meaning those not registered, independently of their size; and (2) firms split according to the specific survey, the SSIS and the LMMS. The results are presented in Table A.10 in the Appendix and appear to be consistent with the baseline results in Table 3.

6.2. 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.11 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.

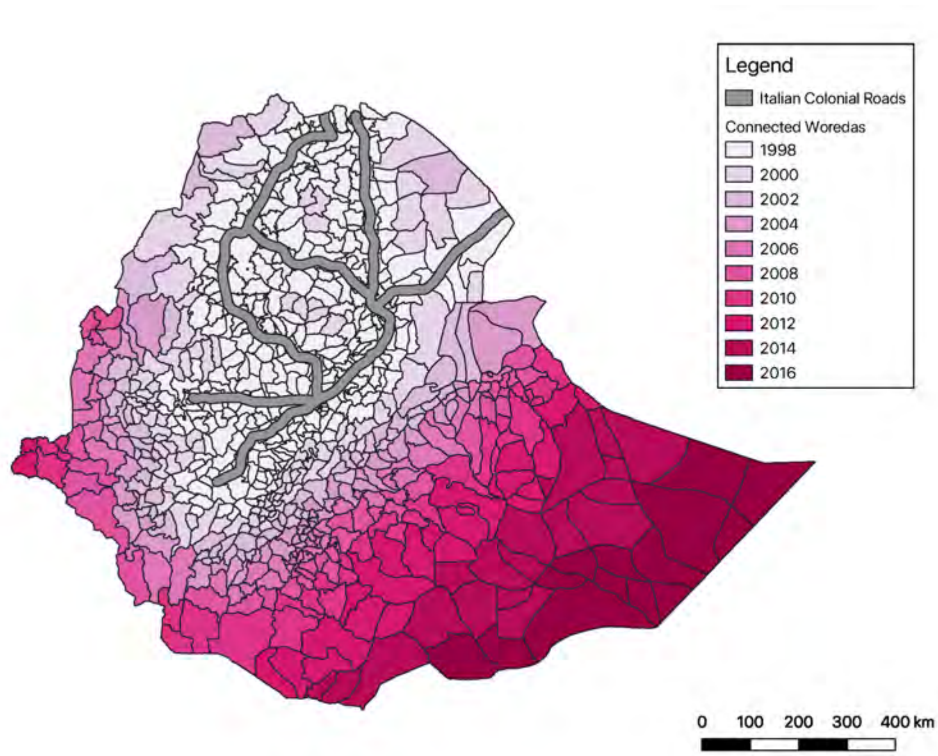
6.3. Alternative Instrument

Finally, in Table 8, we present results using an alternate instrument. Following the insights of Moneke (2020) and Gebresilasse (2020), we construct an additional instrumental variable. First, 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 6.

Note that 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 35km/h) to each artificially added road segment[17] . 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 8, remain qualitatively similar.

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

Figure 6: Connected Districts under the Synthetic Road Network: 1998-2016



Note: Graphical representation of the alternative IV employed in the analysis. Darker shading indicates districts that were connected to the synthetic road network in more recent years, following the artificial regional budgeting algorithm.

Table 8: Alternative IV Strategy

	WHOLE SAMPLE		INFORMAL		FORMAL	
VARIABLES	(1) Labor Prod.	(2) Labor Prod.	(3) Labor Prod.	(4) Labor Prod.	(5) Labor Prod.	(6) Labor Prod.
CMA	0.0363	0.0439	-0.0613	-0.0689	0.115**	0.133***
	(0.0395)	(0.0315)	(0.0504)	(0.0507)	(0.0484)	(0.0424)
Obs.	23,662	23,662	14,066	14,066	8,685	8,685
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-test	140	140	53.68	53.42	97.88	97.60

Notes: The dependent variable is labor 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. All regressions include woreda-industry and woreda-time fixed effects. Robust standard errors in parentheses are clustered at the woreda-industry level *** p<0.01, ** p<0.05, * p<0.1.

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.

References

Adamopoulos, Tasso (2019). *Spatial Integration, Agricultural Productivity, and Development: A Quantitative Analysis of Ethiopia's Road Expansion Program.* 2019 Meeting Papers. Society for Economic Dynamics.

Asher, Sam and Novosad, Paul (2020). "Rural Roads and Local Economic Development". *American Economic Review* 110.3, 797–823.

Atkin, D. and Donaldson, D. (2015). "Who's Getting Globalized? The Size and Implications of Intra-National Trade Costs". *Kilts Center at Chicago Booth Marketing Data Center Paper Series.*

Baker, J. (1974). "Developments in Ethiopia's Road System". *Geography* 59.2, 150–154.

Becker, Dennis (2018). "Heterogeneous firms and informality: the effects of trade liberalization on labour markets". *Oxford Economic Papers* 70.1, 47–72.

Bertazzini, Mattia (2022). "The long-term impact of Italian colonial roads in the Horn of Africa, 1935-2000". *Journal of Economic Geography* 22.1, 181–214.

Burgess, Robin, Jedwab, Remi, Miguel, Edward, Morjaria, Ameet, and Miquel, Gerard Padró i (2015). "The Value of Democracy: Evidence from Road Building in Kenya". *American Economic Review* 105.6, 1817–51.

Chandra, Amitabh and Thompson, Eric (2000). "Does public infrastructure affect economic activity? Evidence from the rural interstate highway system". *Regional Science and Urban Economics* 30.4, 457–490.

Chatterjee, Santanu, Lebesmuehlbacher, Thomas, and Narayanan, Abhinav (2021). "How productive is public investment? Evidence from formal and informal production in India". *Journal of Development Economics* 151, 102625.

Chaurey, Ritam and Le, Duong Trung (2022). "Infrastructure maintenance and rural economic activity: Evidence from India". *Journal of Public Economics* 214, 104725. ISSN: 0047-2727.

De Loecker, Jan and Goldberg, Pinelopi Koujianou (2013). *Firm Performance in a Global Market.* Working Paper 19308. National Bureau of Economic Research.

Desmet, Klaus and Parente, Stephen L. (2010). "Bigger is Better: Market Size, Demand Elasticity, and Innovation". *International Economic Review* 51.2.

Diao, Xinshen, Ellis, Mia, McMillan, Margaret S., and 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, and Ulyssea, Gabriel (2021). *Trade and Informality in the Presence of Labor Market Frictions and Regulations.* NBER Working Papers 28391. National Bureau of Economic Research, Inc.

Donaldson, Dave (2018). "Railroads of the Raj: Estimating the Impact of Transportation infrastructure". *American Economic Review* 108.4-5, 899–934.

Donaldson, Dave and Hornbeck, Richard (2016). "Railroads and American Economic Growth: A Market Access Approach". *The Quarterly Journal of Economics* 131.2, 799–858.

Duflo, Esther and Pande, Rohini (2007). "Dams". *The Quarterly Journal of Economics* 122.2, 601–646.

Duranton, Gilles, Morrow, Peter M., and Turner, Matthew A. (2014). "Roads and Trade: Evidence from the US". *The Review of Economic Studies* 81.2 (287), 681–724.

Duranton, Gilles and Turner, Matthew A. (2012). "Urban Growth and Transportation". *The Review of Economic Studies* 79.4, 1407–1440.

ERA (2009). *SDP Performance: 12 Years Later*. Tech. rep. Ethiopian Road Authority (ERA).

Erten, Bilge, Leight, Jessica, and Tregenna, Fiona (2019). "Trade liberalization and local labor market adjustment in South Africa". *Journal of International Economics* 118, 448–467. ISSN: 0022-1996.

Ethiopian Road Authority (2016). *Road Sector Development Program 19 years performance assessment*. Tech. rep. Ministry of Transport, Ethiopian Road Authority (ERA).

Fiorini, Matteo and Sanfilippo, Marco (2022). "Roads and Jobs in Ethiopia". *The World Bank Economic Review* 36.4, 999–1020.

Fiorini, Matteo, Sanfilippo, Marco, and Sundaram, Asha (2021). "Trade liberalization, roads and firm productivity". *Journal of Development Economics* 153, 102712. ISSN: 0304-3878.

Gebresilasse, Mesay M. (2020). "Rural Roads, Agricultural Extension, and Productivity". *Boston University* Job Market Paper.

Gli Annali dell'Africa Italiana (1937-1943). *Piccioli, Angelo*. Tech. rep. Mondadori.

Goldberg, Pinelopi and Pavcnik, Nina (2003). "The response of the informal sector to trade liberalization". *Journal of Development Economics* 72.2, 463–496.

Harris, Chauncy D. H (1954). "The Market as a Factor in the Localization of Industry in the United States". *Annals of the Association of American Geographers* 44.4, 315–348.

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

Jedwab, Rémi and Storeygard, Adam (2021). "The Average and Heterogeneous Effects of Transportation Investments: Evidence from Sub-Saharan Africa 1960–2010". *Journal of the European Economic Association* 20.1, 1–38.

Kruse, Hagen, Mensah, Emmanuel, Sen, Kunal, and De Vries, Gaaitzen (2022). "A manufacturing renaissance? Industrialization trends in the developing world". *IMF Economic Review*.

La Porta, Rafael and Shleifer, Andrei (2014). "Informality and Development". *Journal of Economic Perspectives* 28.3, 109–26.

McCaig, Brian and Pavcnik, Nina (2018). "Export Markets and Labor Allocation in a Low-Income Country". *American Economic Review* 108.7, 1899–1941.

McMillan, Margaret S. and McCaig, Brian (2020). "Trade Liberalization and Labor Market Adjustment in Botswana". *Journal of African Economies* 29 (3), 236–270.

Melitz, Marc J. and Ottaviano, Gianmarco I. P. (2008). "Market Size, Trade, and Productivity". *The Review of Economic Studies* 75.1, 295–316.

Melitz, Marc (2003). "The Impact of Trade on Intra-Industry Re-Allocation and Aggregate Industrial Productivity". *Econometrica* 71, 1695–1725.

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". *Journal of International Economics* 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.

Shiferaw, Admasu, Söderbom, Måns, Siba, Eyerusalem, and Alemu, Getnet (2015). "Road Infrastructure and Enterprise Dynamics in Ethiopia". *The Journal of Development Studies* 51.11, 1541–1558.

Siba, Eyerusalem (2015). "Returns to Physical Capital in Ethiopia: Comparative Analysis of Formal and Informal Firms". *World Development* 68.

Topalova, Petia and Khandelwal, Amit (2011). "Trade Liberalization and Firm Productivity: The Case of India". *The Review of Economics and Statistics* 93.3, 995–1009.

Ulyssea, Gabriel (2018). "Firms, Informality, and Development: Theory and Evidence from Brazil". *American Economic Review* 108.8, 2015–47.

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

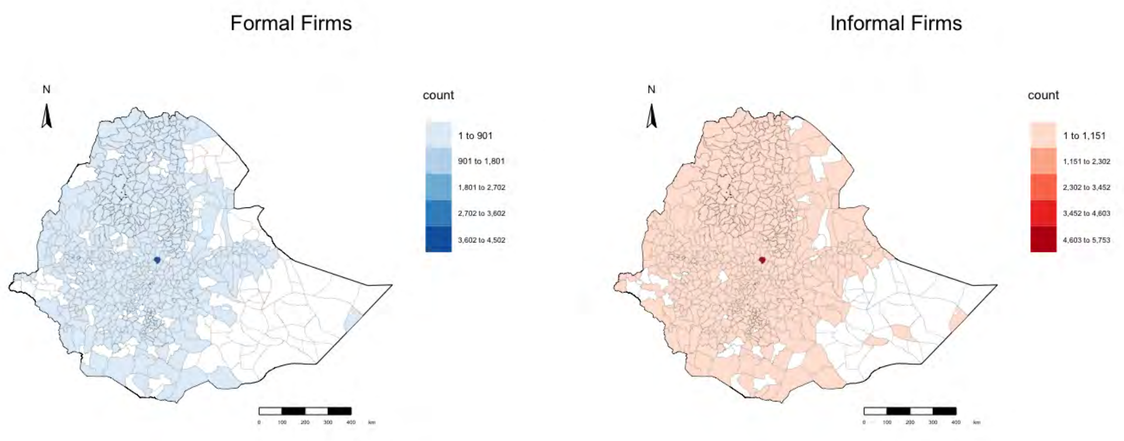
Worku, Ibrahim (2011). "Road Sector Development and Economic Growth in Ethiopia". Addis Ababa, Ethiopia: Ethiopian Development Research Insitute.

World Bank (2021). *Ethiopia-Road Sector Development Program APL4*.

A. Appendix

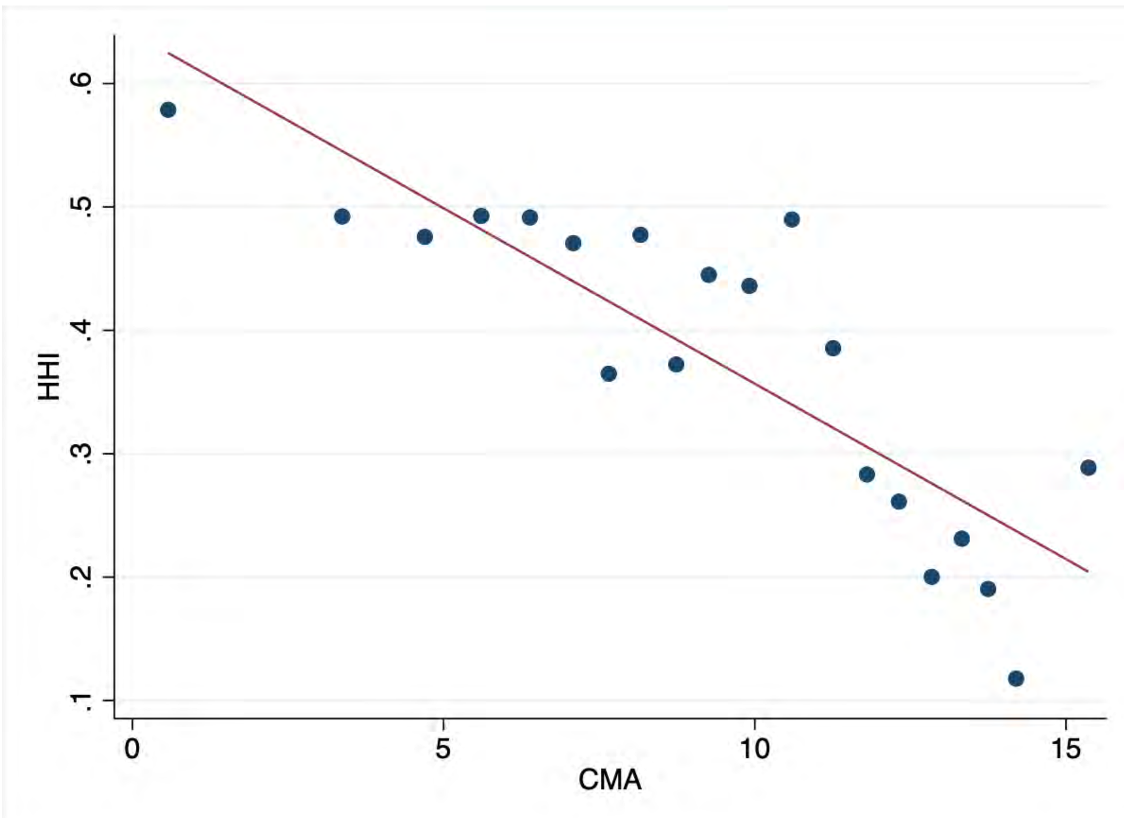
Appendix Figures

Figure A.1: Geographic Distribution of Formal and Informal firms on the Ethiopian territory



Notes: The maps report the count of formal and informal firms per district during the period covered by our sample.(Source: Authors' calculations on SSIS and LMMS data.).

Figure A.2: Correlation between CMA and the Herfindal Index



Notes: The graph represents a binscatter of the correlation between CMA and a measure of the Hirschman-Herfindal index calculated at the level of each market (a district-industry pair)(Source: Authors' calculations on SSIS and LMMS data.)

Figure A.3: Ethiopian Road Network 1969

<u>Notes</u>: Map of the Ethiopian Road Network 1969, which has been digitalized for the construction of the instrumental variable employed in the analysis. (Source: United States Central Intelligence Agency (CIA) (1969)).

Appendix Tables

Table A.1: Growth of the Ethiopian Road Network 1997-2016

Year	Paved		Major gravel		Minor gravel	
	Km	Growth (%)	Km	Growth (%)	Km	Growth (%)
1997	3,708		12,162		10,680	
1998	3,760	1.40	12,240	0.64	11,737	9.90
1999	3,812	1.38	12,250	0.08	12,600	7.35
2000	3,824	0.31	12,250	0.00	15,480	22.86
2001	3,924	2.62	12,467	1.77	16,480	6.46
2002	4,053	3.29	12,564	0.78	16,680	1.21
2003	4,362	7.62	12,340	-1.78	17,154	2.84
2004	4,635	6.26	13,905	12.68	17,956	4.68
2005	4,972	7.27	13,640	-1.91	18,406	2.51
2006	5,002	0.60	14,311	4.92	20,164	9.55
2007	5,452	9.00	14,628	2.22	22,349	10.84
2008	6,066	11.26	14,363	-1.81	23,930	7.07
2009	6,938	14.38	14,234	-0.90	25,640	7.15
2010	7,476	7.75	14,373	0.98	26,944	5.09
2011	8,295	10.96	14,136	-1.65	30,712	13.98
2012	9,875	19.05	14,675	3.81	31,550	2.73
2013	11,301	14.44	14,455	-1.50	32,582	3.27
2014	12,640	11.85	14,217	-1.65	33,609	3.15
2015	13,551	7.21	14,055	-1.14	30,641	-8.83
2016	14,632	7.98	13,400	-4.66	31,620	3.20

Notes: Table Table A.1 describes the growth of the Ethiopian road network broken down into three types of roads: paved, major gravel, and minor gravel. The kilometers of each type of road are from Table 19 of Ethiopian Road Authority, 2016. The growth rate was calculated as the percentage change in kilometers for a particular road type relative to the previous year.

Table A.2: Number of firms in SSIS and LMMS

Year	N firms surveyed (SSIS)	N firms surveyed (LMMS)
2001	8,054	759
2004	4,299	990
2007	11,314	1327
2010	3,882	1880
2013	11,307	2391
Total	38,856	7346

Source: Authors' summary of total number of firms surveyed by the Small Scale Industries Survey (SSIS) and of the Large and Medium Manufacturing industry Survey (LMMS)

Table A.3: Share of SSIS firms in the sample

Year	% Total Firms	% Total Emp.	% Total Prod.	% Total Wages	% Total Cap.
2001	97.67%	50.96%	2.54%	10.49%	11.18%
2004	97.53%	55.41%	9.63%	14.99%	11.76%
2007	97.02%	52.09%	9.29%	16.47%	12.05%
2010	96.47%	52.18%	9.55%	17.04%	11.41%
2013	97.99%	62.95%	16.56%	19.00%	10.68%

Notes: Data report the share of SSIS firms in the total number of firms, total employment, total production, total wage bill and total capital, respectively. Shares are measured combining information on the full sample of firms over the whole period covered in the analysis. Sample weigths available from the SSIS have been used to scale the small firm sector to its universe.

Table A.4: CMA of Formal and Informal Firms

	Informal					Formal				
year	mean	sd	min	max	n	Mean	Sd	Min	Max	N
2001	5.40	4.32	-6.90	16.71	5564	7.14	4.86	-3.45	16.71	1666
2004	6.16	4.80	-6.74	16.44	3337	8.04	4.85	-5.83	16.45	1436
2007	7.85	3.73	-7.17	17.08	8034	9.05	3.82	-8.01	16.52	2997
2010	8.64	3.35	-1.84	18.14	2901	10.19	3.80	-2.99	17.94	1920
2013	9.73	3.11	-5.33	18.78	7659	10.40	3.55	-4.24	18.78	4325

Notes: Authors' summary statistics of CMA for formal and informal firms over the period of the analysis.

Table A.5: Results, OLS

	WHOLE SAMPLE		INFORMAL		FORMAL	
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Labor Prod.	Labor Prod.	Labor Prod.	Labor Prod.	Labor Prod.	Labor Prod.
CMA	0.0729**	0.0718***	0.00280	-0.00328	0.137***	0.139***
	(0.0293)	(0.0215)	(0.0226)	(0.0227)	(0.0307)	(0.0265)
Constant	10.07***	10.88***	10.27***	10.94***	10.10***	10.41***
	(0.279)	(0.264)	(0.206)	(0.223)	(0.321)	(0.289)
Obs.	23,662	23,662	14,066	14,066	8,685	8,685
R^2	0.436	0.479	0.470	0.477	0.379	0.411
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

Notes: The dependent variable is labor 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. All regressions include woreda-industry and woreda-time fixed effects. Robust standard errors in parentheses are clustered at the woreda-industry level *** p<0.01, ** p<0.05, * p<0.1.

Table A.6: First Stage Regressions

	WHOLE SAMPLE	INFORMAL	FORMAL
	(1)	(2)	(3)
	CMA	CMA	CMA
CMA_{IV}	0.9651***	0.9806***	0.9576***
	(0.0168)	(0.01737)	(.01903)
Observations	23662	14066	8685

Notes: The table reports the results of the first stage regression. All regressions include firm specific controls, woreda-industry and woreda-time fixed effects. Robust standard errors in parentheses are clustered at the woreda-industry level *** p<0.01, ** p<0.05, * p<0.1.

Table A.7: Informal firms by Size of the Informal Sector

	Above Median	Below Median	Above Median	Below Median
	(1)	(2)	(3)	(4)
VARIABLES	Labor Prod.	Labor Prod.	Labor Prod.	Labor Prod.
CMA	-0.0830**	0.0376	-0.0683*	0.0340
	(0.0360)	(0.0408)	(0.0393)	(0.0404)
Observations	5,755	8,017	5,406	8,364
R-squared	0.009	0.016	0.008	0.017
District-Industry FE	Y	Y	Y	Y
District-year FE	Y	Y	Y	Y
Controls	Y	Y	Y	Y
F-test	1366	1663	1745	1905

Notes: The table shows the heterogeneous response by **informal** firms to CMA across districts with above and below median shares of informal firms on the total number of firms in a given market (columns 1 and 2) and the total number of employees (columns 3 and 4). Firm level control variables include firm age and a binary variable taking value of 1 for all firms observed in the SSIS. All regressions include woreda-industry and woreda-time fixed effects. Robust standard errors in parentheses are clustered at the woreda-industry level *** p<0.01, ** p<0.05, * p<0.1.

Table A.8: Formal firms by Size of the Informal Sector

	Above Median	Below Median	Above Median	Below Median
VARIABLES	(1) Labor Prod.	(2) Labor Prod.	(3) Labor Prod.	(4) Labor Prod.
CMA	0.227***	0.206***	0.184***	0.212***
	(0.0447)	(0.0692)	(0.0461)	(0.0587)
Observations	4,996	3,511	4,425	4,155
R-squared	0.059	0.059	0.043	0.063
District-Industry FE	Y	Y	Y	Y
District-year FE	Y	Y	Y	Y
Controls	Y	Y	Y	Y
F-test	9258	1909	11241	1957

Notes: The table shows the heterogeneous response by **formal** firms to CMA across districts with above and below median shares of informal firms on the total number of firms in a given market (columns 1 and 2) and the total number of employees (columns 3 and 4). Firm level control variables include firm age and a binary variable taking value of 1 for all firms observed in the SSIS. All regressions include woreda-industry and woreda-time fixed effects. Robust standard errors in parentheses are clustered at the woreda-industry level *** p<0.01, ** p<0.05, * p<0.1.

Table A.9: Reduced Form Specification

	WHOLE SAMPLE		INFORMAL		FORMAL	
VARIABLES	(1) Labor Prod.	(2) Labor Prod.	(3) Labor Prod.	(4) Labor Prod.	(5) Labor Prod.	(6) Labor Prod.
CMA_{IV}	0.0754***	0.0732***	0.00305	-0.00258	0.138***	0.138***
	(0.0283)	(0.0211)	(0.0235)	(0.0236)	(0.0305)	(0.0264)
Constant	10.08***	10.90***	10.27***	10.93***	10.15***	10.48***
	(0.256)	(0.252)	(0.203)	(0.223)	(0.304)	(0.277)
Obs.	23,662	23,662	14,066	14,066	8,685	8,685
R^2	0.436	0.479	0.470	0.477	0.379	0.411
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

Notes: The dependent variable is labor productivity at the firm level, while the independent variable is the instrumental variable, CMA_{IV}, 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. All regressions include woreda-industry and woreda-time fixed effects. Robust standard errors in parentheses are clustered at the woreda-industry level *** p<0.01, ** p<0.05, * p<0.1.

Table A.10: Alternative definition of informality

	Book of Account		Survey	
	Formal	Informal	Formal	Informal
	(1)	(2)	(3)	(4)
VARIABLES	Labor Prod.	Labor Prod.	Labor Prod.	Labor Prod.
CMA	0.192***	0.00779	0.143***	0.00794
	(0.0366)	(0.0249)	(0.0359)	(0.0229)
Observations	5,268	14,822	7,896	18,133
R-squared	0.014	0.009	0.016	0.007
District-Industry FE	Y	Y	Y	Y
District-year FE	Y	Y	Y	Y
Controls	Y	Y	Y	Y
F-test	2251	3004	2576	3264

Notes: The dependent variable is labor productivity at the firm level, while the independent variable is CMA, which varies at the woreda, industry and year level. Firm level control variables include firm age. Column (1) report results considering as informal firms those firms that do not keep books of account, while formal firms are those that keep them (column 2), independently of their size; columns (3) and (4) report results that define firms according to whether they were surveyed in the SSIS (Column 3) or in the LMMS (4). Firm level control variables include firm age and a binary variable taking value of 1 for all firms observed in the SSIS. All regressions include woreda-industry and woreda-time fixed effects. Robust standard errors in parentheses are clustered at the woreda-industry level *** p<0.01, ** p<0.05, * p<0.1.

Table A.11: Robustness to alternative Trade Elasticities

	WHOLE SAMPLE		INFORMAL		FORMAL	
VARIABLES	(1) Labor Prod.	(2) Labor Prod.	(3) Labor Prod.	(4) Labor Prod.	(5) Labor Prod.	(6) Labor Prod.
			θ =1			
CMA	0.149***	0.133***	0.00753	0.000553	0.228***	0.221***
	(0.0464)	(0.0346)	(0.0329)	(0.0329)	(0.0444)	(0.0397)
R-squared	0.004	0.080	0.000	0.014	0.011	0.062
F-test	178697	178748	251602	253388	100364	100170
			θ =3.12			
CMA	0.101***	0.0955***	0.00621	2.56e-05	0.172***	0.171***
	(0.0343)	(0.0254)	(0.0277)	(0.0278)	(0.0346)	(0.0302)
R-squared	0.002	0.079	0.000	0.014	0.008	0.060
F-test	4753	4747	5032	5034	3808	3798
			θ =8.22			
CMA	0.00919	0.0138	-0.00264	-0.00515	0.0328**	0.0389***
	(0.0113)	(0.00947)	(0.0114)	(0.0115)	(0.0142)	(0.0128)
R^2	0.000	0.077	-0.000	0.014	0.001	0.053
F-test	1261	1256	862	864.4	880	873.9
Obs.	23,662	23,662	14,066	14,066	8,685	8,685
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

Notes: The dependent variable is labor productivity at the firm level, while the independent variable is CMA, which varies at the woreda, industry and year level. We test the sensitivity of the results, by employing CMA with different values of θ (1, 3.12, 8.22), the distance decay parameter. 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. All regressions include woreda-industry and woreda-time fixed effects. Robust standard errors in parentheses are clustered at the woreda-industry level *** p<0.01, ** p<0.05, * p<0.1.

Table A.12: Results using Travel Time instead of CMA

	WHOLE SAMPLE		INFORMAL		FORMAL	
VARIABLES	(1) Labor Prod.	(2) Labor Prod.	(3) Labor Prod.	(4) Labor Prod.	(5) Labor Prod.	(6) Labor Prod.
Travel Time	0.0359	0.00353	0.0208	0.00339	0.0148	-0.0165
	(0.0301)	(0.0390)	(0.0393)	(0.0453)	(0.0316)	(0.0283)
Constant	10.87***	11.74***	10.37***	11.02***	11.52***	11.85***
	(0.0939)	(0.144)	(0.141)	(0.255)	(0.0746)	(0.0790)
Obs.	23,662	23,662	14,388	14,388	9,221	9,221
R^2	0.257	0.327	0.290	0.299	0.253	0.296
District FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y	Y	Y
Controls	N	Y	N	Y	N	Y

Notes: The dependent variable is labor productivity at the firm level, while the independent variable is a measure of travel time, obtained by giving value of 1 to sectoral production in the definition of CMA in equation 8. Firm level control variables include firm age and a binary variable taking value of 1 for all firms observed in the SSIS. All regressions include woreda and time fixed effects. Robust standard errors in parentheses are clustered at the woreda level *** p<0.01, ** p<0.05, * p<0.1.

Table A.13: Informal firms for different employment cohorts

	Employment cohorts				
	empl. ≤ 2	2 < empl. ≤ 4	4 < empl. ≤ 6	6 < empl. ≤ 8	empl. > 8
VARIABLES	(1) Prod.	(2) Prod.	(3) Prod.	(4) Prod.	(5) Prod.
CMA	-0.0137	-0.00165	-0.0528	0.0358	0.874***
	(0.0434)	(0.0339)	(0.0509)	(0.0614)	(0.318)
Obs.	3,450	5,662	2,023	629	239
R^2	0.015	0.008	0.016	0.023	0.130
Dist-Ind FE	Y	Y	Y	Y	Y
Dist-year FE	Y	Y	Y	Y	Y
Controls	Y	Y	Y	Y	Y
F-test	1851	2807	981.9	1034	1042

Notes: The dependent variable is labor productivity at the firm level, while the independent variable is the travel time, which varies at the woreda and year level. Firm level control variables include firm age. The table shows the heterogeneous response to CMA among informal firms across different employment cohorts. All regressions include woreda-industry and woreda-time fixed effects. Robust standard errors in parentheses are clustered at the woreda-industry level *** p<0.01, ** p<0.05, * p<0.1.