Gains from market integration:
Welfare effects of new rural roads in Ethiopia*

Hundanol A. Kebede†

June 27, 2022

Abstract

This paper estimates the welfare gains from the construction of rural roads that connect agricultural villages to market centers. I take theoretical predictions from Ricardian trade models to a rich high spatial resolution micro data on agricultural production from Ethiopia, which coincides with a period of extensive rural road construction. I estimate that this road construction resulted in an approximately 13% increase in real agricultural income, on average, and show that this increase is attributed to the mechanisms suggested in the Ricardian trade model: the prices of villages' comparative advantage crops increased, and villages reallocated land for these crops following decreases in trade costs.

Keywords: Market Integration, Ricardian Trade Models, Rural Development, Rural Roads, Trade Costs. JEL Codes: F14, H54, O12, O13, O18, Q12, R12

*I am indebted to Kerem Cosar, Sheetal Sekhri, John McLaren and James Harrigan for their guidance in writing this paper. I would like to thank Jonathan Colmer and Sandip Sukhtankar for their helpful comments. I thank Caglar Ozden for his helpful suggestions and encouragement. This paper has also benefited from comments received from participants in UVA's Trade and Development workshops and the WADES DC . conference.

†Southern Illinois University Carbondale. Email: hundanol.kebede@siu.edu or hundanolkebede@gmail.com

1 Introduction

Roads play a key role in economic development by facilitating the movement of goods, people and ideas across locations (Ali et al., 2015). Developing countries not only have poor road infrastructure and high trade costs (Atkin and Donaldson, 2015), but also often face a trade-off in terms of investing in highways and railroads that interconnect cities compared to rural roads that connect rural villages to nearby towns. The recent trade literature has extensively documented the effects of highway and railroad expansions on the spatial distribution of economic activities and welfare in different contexts. Examples include Faber (2014) in China, Ghani et al. (2016) and Donaldson (2018) in India, Cosar and Demir (2016) in Turkey, and Allen and Arkolakis (2014, 2020) in the US.

However, the welfare effects of rural roads are less understood, and it is likely that these effects can be different from those of highways both quantitatively and qualitatively due to differences in the economic roles they play.[1] Few studies in the development literature document reduced-form evidence on the effects of rural roads on different outcomes of interest, such as migration and agricultural productivity (e.g., Asher and Novosad (2020), Shamdasani (2021), and Gebresilasse (2018)). This reduced-form evidence lacks a theoretical framework to pin down the deep mechanisms through which roads affect the outcomes of interest.

In this paper, I empirically analyze the welfare effects of the construction of new rural roads. The novel feature of this paper is that it empirically documents the margins of responses to improvements in road infrastructure. The agricultural setup and access to high-spatial-resolution micro data allow me to conduct a tight validation of the deep mechanisms in the Ricardian trade model suggested in the theoretical section. In most papers exploring the gain from a decrease in trade costs in agriculture, the model has usually been calibrated or estimated in the cross-section, and the impacts of trade cost reductions have been simulated (or the adjustments have not been directly considered). This paper shows these mechanisms at play in a relatively short period of time.

[1]Highways and railroads primarily serve the urban populations and the manufacturing sector, while rural roads favor the rural populations (agrarians) and the agricultural sector by facilitating better access to agricultural input and output markets.

To understand the mechanisms through which roads affect welfare in rural economies, I develop a multicrop multivillage Ricardian model of trade with heterogeneous land quality within and across villages. On the demand side, a village maximizes utility by choosing optimal quantities of varieties of different crops to consume, given local crop prices. On the production side, the village makes a decision on how to optimally allocate its limited farmland across the potential crops, given local crop prices and the productivity of the village in different crops. Villages also engage in costly trade with one another, very similar to how countries trade in workhorse trade models such as that of Eaton and Kortum (2002).

The model provides a number of sharp predictions on the mechanisms through which improvements in road infrastructure affect village welfare. First, decreases in trade costs lead to increases in the relative prices of villages' comparative advantage crops (CA-crops). Second, decreases in trade costs lead to the reallocation of farmland from a village's comparative disadvantage crops (CD-crops) toward its CA-crops. Third, the size of welfare gains from roads depends on the crop composition of the village's consumption basket vis-a-vis the fraction of village land allocated to these crops. For example, the model predicts that a village that has CA in cashcrops should benefit more from decreases in trade costs than should a village that has a CA in cereal because the latter faces a significant increase in the relative costs of its consumption basket following decreases in trade costs. I use the model structure to derive a measure of how much of the total welfare gain from roads estimated with reduced-form regression is explained by those mechanisms suggested in the Ricardian trade model.

The model predictions can be easily tested given panel data on village-level land utilization, crop prices, and a shock to transport infrastructure, which I obtain from Ethiopia. Ethiopia provides a unique setting to study the welfare effects of rural roads. First, almost all of the rural villages were not accessible by modern transport as recently as 2010. Second, to alter this situation, the Ethiopian government launched a large-scale rural road expansion project called the Universal Rural Road Access Program (URRAP), which aimed to connect all rural villages by all-weather roads in just five years, between

2011 and 2015.[2] The program led to the doubling of the total road length in the country by 2015. Third, unlike most sub-Saharan African countries, Ethiopia collects a very rich micro data on agricultural production and crop prices. My agricultural data cover over 2,000 nationally representative rural villages (locally named *Kebele*s, which are the lowest-level administrative units) and all crops. Approximately of half of these villages received road connections under the URRAP, and in most cases, their first roads passable by vehicles.[3]

To address the potential endogeneity of road placements (selection of villages for the URRAP program), I construct a counterfactual road network predicted solely based on cost considerations (land gradient and the location of rivers and lakes). I use this counterfactual road network to construct an instrumental variable for the actual road network. However, it is important to mention that some of the main results in this paper, specifically those concerning the effects of road connections on land reallocation across crops and the relative prices of comparative advantage crops within villages, are less sensitive to these potential endogeneity issues because they are based on triple-difference variation (i.e., crop-village-year variation).[4]

At the core of my empirical exercises is to define a village's comparative advantage crop(s). This requires information on yield estimates for each crop in each village. My agricultural survey data include village-level crop yield estimated by trained enumerators.[5][6] Given the yield estimates, I define a village's comparative advantage crops using the following procedure. First, I calculate the village's yield relative to the national average for each crop. Next, I rank crops within each village based on their yield relative to the national average. I define crops in the top 20% of the ranking based on relative yield as

[2]While the URRAP program was launched in 2011, the first year was spent on capacity building and most of the road construction was not commenced until 2012.

[3]Some of the villages had poor-quality dirt roads that were passable by vehicles only during the dry season. Such roads are often constructed manually by local communities.

[4]This can be observed from the comparison of OLS and IV results in section 4.3.

[5]The enumerators use a method called *crop cut* in which they take a sample of plots (each with an area of 4 square meters) and conduct crop cuts to obtain a yield estimate.

[6]As a robustness check, I use FAO-GAEZ data on agroclimatically attainable yield of crops in each village. These data use a number of agroecological, soil and climatic factors and sophisticated agronomic models to provide yield estimates at 5 arc-minute resolution (see, for instance, Nunn and Qian, 2011; Costinot and Donaldson, 2012; Costinot and Donaldson, 2016; and Donaldson and Hornbeck, 2016 for the application of these datasets in economics studies).

my baseline comparative advantage crops. I relax this baseline threshold to the top 30%, top 40%, etc.., to see the sensitivity of my results; that is, as the thresholds relax, the effects should become weaker.

There are three main results in this paper. First, using reduced-form estimation, I show that the road expansion resulted in an approximately 13% increase in welfare (real agricultural income per hectare), on average. Second, using the model structure, I show that the estimated welfare gain from the reduced-form is attributed to the expansion of trade among the villages. Third, I empirically validate the deep mechanisms suggested in the trade model; i.e., road expansion led to increases in the relative prices of CA crops by approximately 2% and increases in the fraction of land allocated to CA-crops by approximately 4%, for a village with an average increase in market access. These results are robust to an alternative method of addressing the endogeneity of road placement, a matching-based DID strategy.

The estimated gains from rural roads documented in this paper are quantitatively larger than those documented in previous studies quantifying the welfare gain from highways or railroads. This finding is particularly interesting given the short time horizon after road construction and the significantly low-cost nature of the roads considered in this paper and is perhaps attributed to either the fact that the vast majority of roads studied connect rural villages that were previously inaccessible by modern means of transport systems to the nearest urban centers or because these roads play different economic roles than highways and railroads. In contrast, most of the roads studied in previous studies include upgrading or providing alternatives to existing transport infrastructure, such as trunk roads, highways, and railroads.

An emerging body of literature has studied the gains from intranational market integration, particularly in the agricultural sector, using many-location many-goods Ricardian trade models with heterogeneous factors of production (Costinot and Donaldson 2016, Donaldson 2018, Sotelo 2020, Allen and Atkin 2016, and Adamopoulos 2018). The papers most closely related the current paper are Adamopoulos (2018), Donaldson (2018), and Sotelo (2020). Adamopoulos (2018) found a 13.6% increase in aggregate agricultural yield following road expansion and upgrading that reduced trade costs between Ethiopian districts and the location of national grain market centers. In contrast, current paper focuses on the effect of *rural* roads that connect village centers to roads or nearby towns instead of a decrease in trade cost between district capitals and Addis Ababa or other major urban centers. Donaldson (2018) developed a multisector multiregion Ricardian model in which land is treated as homogeneous within a region to study the gains from railway expansion in colonial India. Sotelo (2020) introduced heterogeneous land quality to study how falling trade costs due to the (counterfactual) paving of roads increase agricultural productivity and welfare in Peru. The current paper builds on these two papers for the theoretical part and contributes to this literature by estimating the effect of low-cost rural road construction on crop prices, land allocation, and welfare.

This paper also relates to the broad literature in development economics on the effects of rural roads on various aspects of the rural economy in developing countries (Aggarwal et al. 2018; Gebresilasse 2018, Shamdasani 2021, Aggarwal 2021, Shrestha 2020, and Asher and Novosad 2020).[7] Asher and Novosad (2020) exploited the strict implementation rule of India's massive rural road expansion project, called Pradhan Mantri Gram Sadak Yojana (Prime Minister's Village Road Program, or PMGSY) to identify the program's causal effect using a fuzzy regression discontinuity design. They found that the roads' main effect is to facilitate the movement of people out of agriculture, with little or no effect on agricultural income and consumption. However, the paper relied on proxies, instead of direct measures, for agricultural outcomes due to a lack of data at a fine geographic level. The current paper uses a large household-level agricultural survey and price surveys at a detailed geographic level to construct real agricultural income and consumption. Shamdasani (2021) studied the effect of a large road-building program in India and found that remote farmers who had access to roads diversified their crop portfolio by starting to produce noncereal hybrids, adopted complementary inputs and improved technologies, and hired more labor. Gebresilasse (2018) studies how rural roads complement the agricultural extension program, a program that trains farmers on how to use best agricultural practices

[7] See also Nakamura et al. (2020), Gafarso and Tufa (2019), and Wondemu et al. (2012) for the impact of rural roads on wider indicators of socioeconomic outcomes.

and technology adoption, to increase farm productivity in Ethiopia, while Aggarwal et al. (2018) studied the effect of a counterfactual decrease in transport costs on the adoption of improved agricultural inputs in Tanzania. Shrestha (2020) found that a 1% decrease in distance to roads due to the expansion of highways resulted in a 0.1 - 0.25% increase in the value of agricultural land in Nepal. The current paper uses a theoretical model structure to identify the mechanisms through which rural roads affect welfare in village economies.

The rest of the paper is organized as follows. In Section 2, I present the data, identification issues, and empirical evidence showing how URRAP roads have improved market integration to motivate the model presented in Section 3. Section 4 presents the estimation of key model parameters, welfare gains, and evidence supporting the detailed mechanisms in the theoretical model. Section 5 explores heterogeneity in welfare gains and Section 6 concludes the paper.

2 Data

2.1 Sources

Agricultural production data: I primarily use the Ethiopian Agricultural Sample Survey (AgSS), which is the largest annual agricultural survey in the country covering over 40,000 nationally representative farm households in over 2,000 villages. While this dataset goes back as far as 1995, villages were resampled every year until 2010, which makes tracking a village over time difficult. Staring in 2010, the Central Statistical Agency (CSA) has kept the sample of villages fixed but takes a random sample of approximately 20 farmers per village every year. This dataset includes detailed production information: areas of land covered by each crop, application of fertilizer and other inputs, and quantities of harvest. Moreover, every three years starting from 2009/2010, CSA has also gathered crop utilization information, i.e., the fraction of crop production used for personal consumption, the fraction sold, the fraction used to pay wages, the fraction used for seeds, etc,., for all crops.

Consumption data: To estimate some preference parameters of the model (the elasticity of substitution between crops), I need consumption information. I use Ethiopian Socioeconomic Survey (ESS) data, which are exceptionally detailed panel data of approximately 4,000 nationally representative farm households for 2011, 2013 and 2015. The main advantage of the ESS dataset is that it includes consumption information disaggregated by crops.[8] A large caveat of this dataset is that it covers households in only approximately 330 villages.

Price data: The main price data are from the Agricultural Producer Price Survey (AgPPS), which is a monthly survey of farm-gate prices at a detailed geography (villages) for almost all crops and many other agricultural products.[9] These data cover over 500 representative villages that can be tracked over the period since 2010. I also use the Retail Price Survey (RPS), which is a monthly survey of prices of almost all crops and nonagricultural commodities in major urban centers throughout the country. This dataset covers over 100 urban centers across all administrative zones of the country. Importantly, the agricultural products covered in both datasets overlap almost fully.

Rainfall and agroclimatic data: I use the Food and Agricultural Organization's (FAO) Global Agro-Ecological Zones (GAEZ) data for agroclimatically attainable yield for low/intermediate input use to construct villages' crop suitability, as a robustness check to yield measures in AgSS data. These data cover 19 crops. However, they miss some of the endemic crops that are widely grown in Ethiopia, such as enset and teff. As a result, I use these data only for the robustness exercise. The rainfall data come from Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS), which provides a rainfall dataset starting in 1981. CHIRPS incorporates 0.05° resolution satellite imagery with station data to create a gridded rainfall time series for trend analysis and seasonal drought monitoring. It is widely used to monitor drought in East Africa (Funk et al.,

[8]The consumption information is based on a seven-day recall of basic consumption items, which are predominantly crops. However, household crop utilization information also provides information on how much of each crop produced is consumed within the household.

[9]CSA claims that the prices in this survey can be considered *farm-gate* prices because they are collected at the lowest market channel where the sellers are the producers themselves, i.e., no intermediaries involved.

2015).

Road data: I use administrative data on the entire road network in the country. These data include the attributes of the roads (such as surface type), the role of the road (trunk road, link road, etc.), and ownership (federal government, regional government, etc.). In this paper, I use a large-scale rural road expansion under the URRAP as a source of variation to villages' access to roads/markets. Over the period 2011–2015, the Ethiopian government exclusively focused on the URRAP and constructed over 62,413 km of new all-weather roads connecting village centers to the nearest road or the nearest town, whichever is shorter. Figure 1 shows a map of the road network before and after the URRAP.

The main objective of the URRAP was to improve villages' access to product and input markets. The program increased the overall road density per 1,000 square-km from 44.4 in 2010 to 100.4 in 2015 (Ethiopian Road Authority, 2016). Although the URRAP was launched in 2011, very few roads were commenced in 2011, which is officially considered a capacity building year. Almost all the rural roads constructed under this program were started and completed between 2013 and 2015.

2.2 Identification of the effect of roads

There are two main issues that must be addressed to identify the causal effects of roads on village outcomes. The first is heterogeneity in treatment intensity and potential spillover effects. That is, villages that are connected to a dense network may gain more from the road than may those that are connected to a sparse network. Moreover, road connections in a village may have spillover effects on other villages that are not directly connected. When a village is connected to the preexisting road network or to the nearest urban center, all its neighbors also have improved access to the market via the connected village. As a result, nonconnected villages may not serve as a control group in the identification of the effects of road connections. The second concern is selection bias - villages are selected for the road program based on some demographic, geographic, social, and economic factors.

In particular, due to limited budgets, officials may prioritize villages that would gain more from the road when deciding which villages to connect.[10] However, it is important to mention that some of the estimations in this paper are less sensitive to these potential endogeneity issues because they use triple-difference variation (i.e., crop-village-year variation). This can be observed by comparing the OLS and IV estimation results for the effects of road connections on land reallocation across crops and the relative prices of comparative advantage crops within villages, which are the main results in this paper.

To address the heterogeneity in treatment intensity and spillover effects, I use market access measures derived from general equilibrium trade models (see Donaldson and Hornbeck 2016) that are calculated using the entire road network and the distribution of the population across Ethiopian villages. A change in market access captures the treatment benefit from both direct and indirect road connections and accounts for the density of the network to which a village is connected. See Appendix A for the details on the construction of the market access measure.

I address the potential endogeneity of road placement using an instrumental variable (IV) estimation strategy. To construct the IV, I first obtain the road network predicted based solely on construction costs. The road construction cost is a function of the land gradient and location of rivers and lakes. Next, I construct a market access measure for each village based on the predicted road network - the *predicted* market access measure. I use this *predicted* market access measure as an instrument for the *actual* market access measure. Because the *predicted* market access is based on a road network that is predicted based solely on exogenous cost factors (without taking into account potential benefits of the roads or population settlements), it is a valid IV for the *actual* market access measure.

I use the spatial distribution of the population before the road program, and the entire

[10] Unfortunately, there is no official guideline as to which villages should be selected for the URRAP in a given year. Even though the project is fully funded by the federal government, the implementation of the URRAP has been completely decentralized to regional governments. Within each regional government, districts propose a list of villages that should get a road during a particular year, and the regional governments approve villages based on the available regional budget.

road network of the country pre- and postprogram to construct the market access measure.

$$MarketAccess_{ot} = \sum_{d} \tau_{odt}^{-\theta} Population_d \tag{1}$$

where $Population_d$ is the destination village population from the 2007 census (before the onset of the URRAP). Using the pre-URRAP population distribution is necessary because the population distribution is likely to respond to improvements in road infrastructure. θ is the inverse land heterogeneity parameter (which governs trade elasticity), which I estimate in section 4. τ_{odt} denotes the freight costs of transporting one ton of cargo from origin village o to destination village d along the least cost path, before ($t = 0$) and after ($t = 1$) the construction of URRAP roads. See Appendix B for the details on the construction of the IV.

Note that the market access measure is constructed using road networks and population settlements over the entire country, not just based on villages for which agricultural surveys are available. Thus, although we have market access measures for each of the over 15,000 villages in the country, our analysis is limited to those for which agricultural data are available.

Table 1 reports the descriptive statistics on the market access measure. The first column shows that the market access measure increased by 45.6%, on average, between 2012 and 2016. Column 2 shows that the market access measure increased more for villages that are directly connected compared with villages that are not directly connected by 42%. Finally, the last column reports the first-stage regression results. The market access measure constructed from the *predicted* road network is strongly correlated with the market access measure constructed based the *actual* road network. The estimate in Column 3 shows that the correlation between within-village changes in predicted and actual market access measures is 0.56, with a standard error of 0.024. As a result, the first-stage F-statistics in all specifications are above 1,000.

To check the sensitivity of my results, I use a matching-based difference-in-differences (MB-DID) estimation as an alternative strategy to address endogeneity. That is, I first obtain a matched sample of treated and nontreated villages based on their pre-URRAP observable characteristics that may be relevant for the selection of villages for the URRAP. I match treated and nontreated villages based on their similarity in distance to the nearest town, distance to the preexisting road network, population size, average slope of land in the village, average elevation in the village, and average rainfall over the 1990 - 2010 period. I then conduct a DID estimation based on these matched samples of treated and nontreated villages. Combining matching with the DID strategy is a powerful approach to address the selection problem. The matching step enables me to compare treated villages with nontreated villages that have similar observed characteristics and, hence, similar treatment probability. The DID strategy on these matched samples helps me wash out unobserved time-invariant village characteristics that may confound the treatment effect. See Appendix E for details.

2.3 Evidence on the improvement in market integration

To motivate the theoretical model in the next section, I first present some evidence on how road construction under the URRAP has improved market integration among rural villages.

Farmers face considerable barrier to trade: ESS data include direct questions about transport costs. I use this survey to estimate transport costs. The ad valorem trade cost (transport cost per value of transaction) on vehicles is very high (the median is 6.5%, and the mean is 11.4%). The size of this cost is comparable to international trade costs estimated by Hummels (2007) for US and New Zealand imports, although in my data the median distance traveled is just 12 kilometers.

The URRAP has decreased trade costs: The main objective of URRAP roads is to integrate rural villages and market centers (Ethiopian Road Authority, 2016). If URRAP roads truly integrated rural villages with local market centers, then we would see the price gap between the rural villages and the market centers decrease for villages that receive road connections relative to those that do not receive road connections. I test this by looking at the absolute value of the difference in log crop prices between zone capitals and the villages within the zones using two rich price surveys, AgPPS and RPS. I run the following regression:

$$|lnP^k_{zmt} - lnP^k_{zvmt}| = \alpha_1 Post_t + \alpha_2(Post_t * URRAP_v) + \gamma_v + \gamma^k_m + \gamma_t + \varepsilon^k_{zvmt}$$

where k denotes a crop, v is a village, z is a zone capital, m is the month, t is the year, $Post$ equals zero for all month-years before the URRAP and one for all month-years after the URRAP, $URRAP_v$ is a dummy variable representing whether a village received a URRAP road between 2012 and 2015, and γ^k_m is crop-month fixed effect that captures the possible seasonality of crop prices.

The results are reported in Table 2. This shows that road connections significantly decreased the urban – rural price gap. The first column pools all 56 crops for which data are available on both urban and rural prices and shows that trade costs, as proxied by absolute-value difference in log crop prices between zone capitals and the villages within the zones, decreased by approximately 0.022 (about 6% of the average absolute log price difference) for villages that received road connections, relative to those that did not. In Column 2, the estimation is restricted to perishable products, vegetables and fruits. The point estimate is more than twice the estimate for all crops: the trade cost for vegetables and fruits decreased by approximately 0.047 (about 7% of the average absolute log price difference for vegetables and fruits).

Informed by this evidence on the effect of roads on market integration, in the next section, I develop a multisector multivillage Ricardian trade model to analyze the mechanisms through which road construction affects village welfare.

3 Theoretical framework

The model builds on Donaldson (2018), Sotelo (2020) and Galle et al. (2021).[11] Consider an economy composed of V villages indexed by $v = 1, ..., V$, and each village is represented by a representative household. A village derives utility from the consumption of K crops

[11]The model is also similar in spirit to Costinot et al. (2012).

indexed by $k = 1, ..., K$, which can be potentially locally produced or imported from other villages. Each crop k comes with a continuum of varieties indexed by $j \in \Delta_k$, where Δ_k is the mass of varieties in crop k.

Preferences: The village spends all its income on crops, and its preference is given by

$$U_v = \prod_k \left(q_v^k\right)^{\mu_n^k}, \quad \text{where} \tag{2}$$

$$q_v^k = \left(\int_{\Delta_k} c_v^k(j)^\sigma dj\right)^{\frac{1}{\sigma}} \tag{3}$$

where U_v is utility in village v, $c_v^k(j)$ is the quantity of variety j of crop k consumed by the village, μ_r^k is the share of expenditure on crop k in zone n such that $\sum_k \mu_n^k = 1$, and σ is the elasticity of substitution among varieties of a crop.

Technology: Similar to Sotelo (2020) and Allen and Atkin (2016), I assume that the farmer's technology is constant returns to scale. I also assume, for simplicity of exposition, that land is the only input.[12] Each village has L_v amount of land, which is divided into a continuum of plots of size one indexed by $\omega \in \Omega_v$, where Ω_v is the set of plots in village v such that $\int_{\Omega_v} d\omega = L_v$. Each plot is potentially different in terms of how well it is suited to growing different crops, which I denote as $z_v^k(\omega)$. The suitability of a plot for a particular crop depends on a number of agronomic variables, including soil type. I assume that land suitability for a given crop does not vary across varieties of the crop. Assuming that a given plot can only be used to grow one crop at a time (plots cannot be divided), the production function is given as follows:

$$y_v^{k(j)}(\omega) = t_v^{k(j)} z_v^k(\omega)$$

where $y_v^{k(j)}(\omega)$ is the quantity of variety j of crop k per unit of plot. $t_v^{j(k)}$ is the technology to produce variety j of crop k in village v. This represents villages' know-how about

[12]The model can easily be extended to include labor without altering any of the analysis in this section but at a cost of introducing new notations. Hence, I abstract from introducing labor in this section.

yield-improving or disease-/drought-resistant varieties. A representative farmer in a village draws $t_v^{j(k)}$ from a Fréchet distribution with a level parameter T_v^k and shape parameter γ, with the following cumulative distribution function:

$$\mathcal{D}_v^j = Pr(t_v^{j(k)} < a) = \exp(-(T_v^k)^{\gamma} t^{-\gamma})$$

A farmer also draws $z_v^k(\omega)$ independently for each plot and crop from a Fréchet distribution with a level parameter A_v^k and shape parameter θ. [13] A_v^k can be interpreted as the average productivity of village v in crop k (constant across varieties in crop k). For villages with agroclimatic conditions that make it impossible to produce crop k, A_v^k is set to zero. θ is the (inverse) measure of dispersion in the productivity of land in a village, and it is constant across villages and crops.

Trade: Villages operate in a perfectly competitive crop market. There is an iceberg trade cost of $\tau_{vv'}^k \geq 1$ between villages v and v' in crop k. Motivated by the result in table 2, which shows that spatial price variation differs across crops, trade costs are assumed to vary across crops to reflect that some crops, such as vegetables, are more costly to trade (e.g., perishable) than are others, such as cereals. I assume that $\tau_{vv}^k = 1, \forall k$, and impose the standard assumption of triangle inequality on trade costs, $\tau_{vv'}^k \times \tau_{v'v''}^k \geq \tau_{vv''}^k, \forall k$.

I assume no arbitrage condition, so that for any two villages v and v', equilibrium crop prices satisfy $p_{vv'}^{k(j)} = \tau_{vv'}^k p_{vv}^{k(j)}$ where $p_{vv'}^{k(j)}$ is the price in village v' of variety j of crop k originating from village v, and $p_{vv}^{k(j)}$ is the price in village v of variety j of crop k originating from the same village v.

Equilibrium: In equilibrium each plot is allocated to the crop that maximizes the return from the land, that is, the rental rate of the land. Let r_v^k be the average rental rate of the plots allocated to crop k in village v, which is determined in equilibrium. The unit cost of production of variety j of crop k in village v is $c_v^{k(j)} = \frac{r_v^k}{t_v^{k(j)}}$, which is stochastic because it is a function of stochastic productivity $t_v^{k(j)}$. The price at which village v

[13]The cumulative distribution function for $z_v^k(\omega)$ is $F_v^k(z) = Pr(z_v^k(\omega) < z) = \exp(-(A_v^k)^{\theta} z^{-\theta})$.

supplies variety j of crop k to village v', $p_{vv'}^{k(j)} = \frac{r_v^k}{t_v^{k(j)}} \tau_{vv'}^k$.

Following steps similar to those in Eaton and Kortum (2002), the probability that village v is the cheapest supplier of any variety j of crop k to village v' (the probability that village v's productivity draw for any variety j of crop k, adjusted for trade costs and rental rates, is the highest compared with all other potential villages trading with v') is as follows:

$$\begin{aligned}\pi_{vv'}^k &= \Pr\Big[p_{vv'}^{k(j)} \le \min_n \{p_{nv'}^{k(j)}\}, \quad \text{for some} \quad j \in \Delta_k\Big] \\ &= \frac{(T_v^k)^\gamma (r_v^k \tau_{vv'}^k)^{-\gamma}}{\sum_n (T_n^k)^\gamma (r_n^k \tau_{nv'}^k)^{-\gamma}}\end{aligned}$$

which is increasing in the average productivity of village v in crop k, T_v^k, and decreasing in the trade cost $\tau_{vv'}^k$ and the rental rate in village v, r_v, relative to other villages.

The probability that a village is the cheapest supplier of any variety j of crop k for itself is

$$\begin{aligned}\pi_{vv}^k &= \Pr\Big[p_{vv}^{k(j)} \le \min_{n \neq v} \{p_{nv'}^{k(j)}\}, \quad \text{for some} \quad j \in \Delta_k\Big] \\ &= \frac{(T_v^k)^\gamma r_v^{-\gamma}}{\sum_n (T_n^k)^\gamma (r_n \tau_{nv}^k)^{-\gamma}}\end{aligned}$$

A village is more likely to self-produce a variety j of crop k if the village is more productive in the crop relative to other villages and/or if there are high trade costs between the village and other villages.

The price index of crop k is given by

$$p_v^k = \Gamma\Big(\sum_{v'=1}^{V} (T_{v'}^k)^\gamma (r_{v'}^k \tau_{v'v}^k)^{-\gamma}\Big)^{\frac{-1}{\gamma}} \tag{4}$$

Given the preferences in 2 and the expression for p_v^k given in equation 4, the price index faced by village v is given as follows:

$$P_v = \prod_k \left(p_v^k\right)^{\mu_n^k} \tag{5}$$

Equilibrium land allocation: I assume that there is a competitive land rental market. Each village decides how to allocate its farmland across different crops given prices p_v^k and the suitability of the village land for various crops.[14] Revenue maximization implies that each plot of land is allocated to a crop that yields the highest return:

$$R_v(\omega) \equiv \max_k \{p_v^k z_v^k(\omega)\} \tag{6}$$

where $R_v(\omega)$ is revenue from plot ω. Together with the Fréchet distribution, this implies the following land allocation rule:

$$\eta_v^k = \frac{(p_v^k A_v^k)^\theta}{(\Phi_v)^\theta}, \quad \text{where} \quad \Phi_v = \left(\sum_l \left(p_v^l A_v^l\right)^\theta\right)^{\frac{1}{\theta}} \tag{7}$$

where η_v^k is the fraction of land in village v allocated to variety j of crop k. It increases with the price of the crop and the average productivity of the village in the crop, relative to all other crops.

Revenue per plot and equilibrium rental rate: In Appendix C, I derive the conditional distribution of land productivity $\mathcal{Q}_v^k(z) \equiv \mathcal{P}\Big(z_v^k(\omega) < z|\omega \in \Omega_v^k\Big)$, i.e., the distribution of productivity of a plot conditional on the plot being used for crop k, which gives the following distribution function:

$$\mathcal{Q}_v^k(z) = \exp\Big(- \Big(\frac{\Phi_v}{p_v^k}\Big)^\theta z^{-\theta}\Big)$$

which is Fréchet with the expected value of $\frac{\Phi_v}{p_v^k}$.

Suppose that crop k is the crop that maximizes revenue from plot ω so that optimal revenue from plot ω is given by $R_v(\omega) \equiv p_v^k y_v^k(\omega) = p_v^k z_v^k(\omega)$. The conditional distribution of revenue from a plot conditional on the plot being used for crop k, $\mathcal{P}\Big(R_v(\omega) < R|\omega \in \Omega_v^k\Big)$, is also Fréchet with the expected value of Φ_v because revenue is just the productivity

[14] Note that because land suitability does not vary across varieties of a given crop, the land allocation problem is solved at the crop level using crop price instead of at the variety level. Once a decision is made regarding which crop is planted on a given plot, the specific variety of the crop to be planted is chosen based on the technology of the farmer $t^k(j)_v$, i.e., the unit cost of the crop variety.

term multiplied by a nonstochastic price p_v^k. Moreover, given the assumption of a competitive land rental market, the rental rate per plot is equal to revenue per plot. Thus, the conditional distribution of the rental rate per plot is the same as the conditional distribution of revenue per plot (note from equation 6 that $r_v(\omega)|\omega \in \Omega_v^k = p_v^k z_v^k(\omega)$ which has a Fréchet distribution with parameter Φ_v).

Note that the equilibrium rental rate is equalized across crops. That is, if you take the set of plots that are covered by crop k and another set of crops covered by crop k', the average rental rate is equal across both sets of plots, $r_v^k = r, \forall k$. This result follows from the Fréchet functional form.

Comparative statics: Following Dekle et al. (2008) and Galle et al. (2021), we obtain the comparative statics of the effect of proportionate change in trade costs between two villages v and v' where $v \neq v'$. We denote $\hat{x} \equiv \frac{x'}{x}$ as proportionate change (the ratio of the new to the old value of a variable). Using this notation, for a given proportionate change in trade cost $\hat{\tau}_{vv'}^k = \frac{\tau'^k_{vv'}}{\tau^k_{vv'}}$ the resulting proportionate change in the endogenous variables is given as follows:

$$
\begin{aligned}
\hat{\Phi}_v &= \left[\sum_k \eta_v^k (\hat{r}_v^k \hat{\tau}_{vv'}^k)^\theta \right]^{1/\theta} \\
\hat{\pi}_{vv'}^k &= \frac{(\hat{r}_v^k \hat{\tau}_{vv'}^k)^{-\gamma}}{\sum_v \pi_{vv'}^k (\hat{r}_v^k \hat{\tau}_{vv'}^k)^{-\gamma}} \\
\hat{\eta}_v^k &= \frac{(\hat{r}_v^k \hat{\tau}_{vv'}^k)^\theta}{\sum_l \eta_v^l (\hat{r}_v^k \hat{\tau}_{vv'}^k)^\theta} \\
\hat{p}_v^k &= \left[\sum_v \pi_{vv'}^k (\hat{r}_{v'}^k \hat{\tau}_{vv'}^k)^{-\gamma} \right]^{-1/\gamma}
\end{aligned}
\tag{8}
$$

In deriving the above results, we assume that the technology parameter T_v^k and the land suitability parameter A_v^k are constant.

Welfare effects: Because land is the only factor of production in the model, the average real rental rate per plot, which is also equal to the average real revenue per plot, can be used as a measure of village welfare: $W_v \equiv \frac{r_v}{P_v} = \frac{R_v}{P_v}$ where r_v is average rental rate per plot, R_v is the average revenue per plot, and P_v is the village price index.

Given data on the quantities of each crop produced in a village, village crop prices, and model parameters needed to construct the village price index, one can construct village real revenue $\frac{R_v}{P_v}$ and compare its changes over time in villages that obtain new road connections with villages whose road status did not change in a difference-in-differences (DID) estimation. However, this DID estimate does not tell us how much of the change in welfare due to roads is attributed to trade expansion among villages. Moreover, I do not observe trade flows between villages to directly measure the magnitude of change in trade flow among the villages.

To address this challenge, I rely on the theoretical results from the model. Note that the proportionate change in real income for a given proportionate change in trade costs is given by $\hat{W}_v = \frac{\hat{R}_v}{\hat{P}_v}$. Using equations 5 and 8 and taking logs, we have the following:

$$\ln\hat{W}_v = \sum_k \frac{\mu_n^k}{\theta}\ln\hat{\eta}_v^k - \sum_k \frac{\mu_n^k}{\gamma}\ln\hat{\pi}_{vv}^k \tag{9}$$

Our main interest is with the term $-\sum_k \frac{\mu_n^k}{\gamma}\ln\hat{\pi}_{vv}^k$, which is a measure of welfare gain from trade. This term is very similar to the sufficient statistics measure of the welfare gain from workhorse trade models (see Costinot et al. (2012), and Arkolakis et al. (2012)). The size of the welfare gain from trade is governed by the size of the parameter γ, which is the traditional trade elasticity parameter. A larger γ implies that villages are more homogeneous in terms of their productivity in different varieties of crops, and hence, there are lower gains from a decrease in trade costs among villages. I construct this measure of welfare gain as follows. Given data on the quantities and prices of each crop produced in a village before and after the road program, I construct panel data of real agricultural income (the left-hand side expression in 9). Second, given village-level data on the allocation of land across crops (η_v^k) before and after the road program and estimated values of parameters μ_n^k and θ, I construct the first term on the right-hand side of equation 9. Then, I use the difference between these two terms as my measure of the welfare gain attributed to the expansion of trade among villages, $-\sum_k \frac{\mu_n^k}{\gamma}\ln\hat{\pi}_{vv}^k$.

The following propositions summarize the key empirically testable mechanism through which roads affect village welfare in this model.

Proposition 1. *Decreases in trade costs lead to reallocation of farmland to a village's comparative advantage crops, resulting in greater specialization of villages.*

Proof. The elasticity of the land share of a crop in a village to the village's productivity is given as $\frac{d\ln\eta_v^k}{d\ln A_v^k} = \theta(1-\eta_v^k)$. Differentiating with respect to trade costs, we obtain $\frac{d^2\ln\eta_v^k}{d\tau_{vv'}^k d\ln A_v^k} = -\theta\frac{d\eta_v^k}{dp_v^k}\frac{dp_v^k}{d\tau_{vv'}^k}$. The term $\frac{d\eta_v^k}{dp_v^k}$ is always positive (see equation 7). Consider two villages, v and v', and suppose that the price of crop k in village v' is normalized so that the price in village v is defined relative to the price in village v'. This implies that $p_v^k = \tau_{vv'}^k$ if k is a comparative disadvantage (CD) crop in village v (i.e., $p_v^k > p_{v'}^k = 1$) or $p_v^k = \frac{1}{\tau_{vv'}^k}$ if k is a CA crop in village v (i.e., $p_v^k < p_{v'}^k = 1$). Thus the term $\frac{dp_v^k}{d\tau_{vv'}^k}$ has a positive sign if crop k is a CD crop in the village and a negative sign otherwise.[15] This implies that as trade costs decrease, villages reallocate more land to their CA crops. □

The intuition is simple. As trade costs decrease, a village's CA crops ('export' crops) become relatively more expensive, and CD crops ('import' crops) become relatively less expensive at local markets. This makes growing CA crops relatively more attractive and growing CD crops relatively less attractive, which induces reallocation of land to these CA crops.

Proposition 2. *The size of the welfare gain from roads depends on the fraction of land allocated to different crops in a village vis-a-vis the consumption composition of the village.*

Recall that welfare is given by $\frac{r_v}{P_v} = \frac{\left(\sum_{k=1}^{K}(p_v^k A_v^k)^\theta\right)^{\frac{1}{\theta}}}{\prod_k \left(p_v^k\right)^{\mu_n^k}}$. Taking logs and differentiating this with respect to p_v^k gives the following:

$$d\ln W_v = \sum_k \left(\eta_v^k - \mu_n^k\right) d\ln p_v^k \tag{10}$$

This equation implies that a decrease in trade costs of CA crops increases village welfare by a larger magnitude if $\eta_v^k >> \mu_n^k$. A testable implication of this proposition

[15]Recall that, from the no-arbitrage and $\tau_{vv'}^k \geq 1$ conditions, $p_v^k = p_{v'}^k \tau_{vv'}^k$, if $p_v^k > p_{v'}^k$ or $p_v^k = p_{v'}^k/\tau_{vv'}^k$, if $p_v^k < p_{v'}^k$

is that a village that specializes in cereals gains less from road connections compared with a village that specializes in cashcrops. This is because, while both groups of villages gain from the increase in the prices of their CA crops, the villages specializing in cereals experience an increase in their consumption expenses, whereas those that specialize in noncereals experience the exact opposite.

4 Estimation of parameters and welfare gains

4.1 Estimation of model parameters

To construct the empirical measure of welfare gains from roads that is attributed to trade expansion among villages, we need to obtain estimates for the parameters of the model: the preference parameters μ_n^k, and the measure of homogeneity of plots in a village θ. We do not need to estimate the trade elasticity parameter γ because our procedure above would give us an estimate of the term $-\sum_k \frac{\mu_n^k}{\gamma} \ln \hat{\pi}_{vv}^k$.

Estimation of μ_n^k: I use ESS data (round 2011) to estimate μ_n^k. Given data on households' consumption of crops, I estimate μ_n^k as the average share of expenditure on crop k in zone z. I allow this parameter to vary across 70 administrative zones in Ethiopia to account for spatial variation in food habits in the country. I use the pre-URRAP data to avoid the concern that these parameters may change in response to changes in trade costs.

Estimation of θ: I follow Sotelo (2020) in the estimation of θ. I rely on the AgSS village-level crop yield estimate that is constructed based on a random sample of crop cuts. To purge out the noise in yield estimates and fluctuations due to weather conditions, I take the average across all years (2010-2016) to obtain a time invariant measure of yield for a crop in a village.[16] I assume that the true crop productivity in village A_v^k is related

[16] Using GAEZ yield instead of time invariant yield constructed from AgSS gives a comparable estimate of θ, even though GAEZ data do not include some of the widely grown crops. This is because the time invariant yield constructed from AgSS is strongly correlated with the GAEZ yield measure for the crops that overlap in both datasets.

to the AgSS's village yield measure Y_v^k in the following equation:

$$A_v^k = \delta^k Y_v^k \exp(-u_v^k)$$

where $\exp(-u_v^k)$ is random noise and δ^k is a crop-specific constant. Plugging this for A_v^k into equation 7 and taking logs gives the following:

$$\ln\Big(P_v^k Y_v^k\Big) = \frac{1}{\theta}\eta_v^k + \ln\Phi_v - \ln\delta^k + u_v^k$$

The empirical counterpart of this is as follows:

$$\ln\Big(P_{vt}^k Y_v^k\Big) = \frac{1}{\theta}\eta_{vt}^k + \gamma_v + \gamma^k + \gamma_t + u_{vt}^k$$

where γ_v and γ^k are village and crop fixed effects, respectively.

I obtain a value of $\hat{\theta} = 2.7$ for productivity dispersion. This is larger than the estimate of Sotelo (2020) which is approximately 1.7.

4.2 Estimation of welfare gains

Recall that the measure of welfare W_{vt} in the model is real revenue per hectare $\frac{R_{vt}}{P_{vt}}$, which is also equal to real rental rate $\frac{r_{vt}}{P_{vt}}$. Given data on crop production and prices before and after the program, I can directly construct village revenue as $R_{vt} = \sum_k p_{vt}^k y_{vt}^k$. For the production of crops at the village level, I use AgSS data for 25 major nontree crops in the main analysis.[17] However, my price data (AgPPS data) do not cover all the villages for which agricultural production data are available (AgSS villages). I impute missing prices as follows. For each village in the AgSS data that received road connections under the URRAP, I find the nearest village in the AgPPS data that also received road under the URRAP. Similarly, for each village in AgSS data that did not receive a road under URRAP, I find the nearest village in the AgPPS data that did not receive a road under the URRAP. I use these imputed prices to calculate village revenue. The price index is

[17]Later, I show that the result is robust to including tree crops.

also calculated using the same data and the expenditure share of crops estimated in the previous subsection: $P_{vt} = \prod_k \left(p^k_{vt}\right)^{\mu^k_n}$. Note that I use these imputed prices for analysis only in this subsection, and I show that the estimated welfare gain using only those villages that have complete data on prices and agricultural survey is very similar in magnitude in a robustness exercise.

To estimate the welfare gain from road connectivity, I run the following regression:

$$\text{ln}W_{vt} = \alpha_1(Post_t * URRAP_v) + \alpha_2 Post_t + \delta\mathbf{X} + \gamma_v + \varepsilon_{vt} \tag{11}$$

where X includes a vector of village characteristics such as rainfall. I also run a similar regression using the market access measure instead of binary treatment using both OLS and IV estimation:

$$\text{ln}W_{vt} = \gamma_1 \text{lnMA}_{vt} + \delta\mathbf{X} + \gamma_v + \gamma_t + \varepsilon_{vt} \tag{12}$$

where the IV estimation uses the market access measure constructed using the counterfactual road network predicted from the land gradient and location of rivers and lakes. See section 2.2 and appendices A and B for details on the construction of the market access measure and IV.

Quantifying the welfare gains from trade How much of the welfare gain estimated in equations 11 or 11 is explained by the expansion of trade among villages? To answer this question, I run the following regression:

$$-\sum_k \widehat{\frac{\mu^k_n}{\gamma}\text{ln}\pi^k_{vvt}} = \alpha\text{lnMA}_{vt} + \delta\mathbf{X} + \gamma_v + \gamma_t + \varepsilon_{vt} \tag{13}$$

Note that in equation 9 the coefficient of $-\sum_k \widehat{\frac{\mu^k_n}{\gamma}\text{ln}\pi^k_{vvt}}$ is one, implying that $\hat{\alpha}$ in equation 13 should be very close to the estimated effect of lnMA on real agricultural income in equation 12.

Table 3 reports the estimation results for equations 11-13. Panel A reports the results for the estimation of the welfare effect; i.e., the dependent variable is log real agricultural income per hectare of land. Column 1 reports the results for the binary treatment approach, while Column 2 and 3 report the OLS and IV estimation results using the market access approach, respectively. Column 1 shows that real agricultural income per hectare of land (a measure of welfare) increased by 20% for treated villages compared with nontreated villages over the period of 2012 to 2016. Columns 2 and 3 report the elasticity of real revenue with respect to market access of 0.114 and 0.295 using OLS and IV estimation, respectively. For a village with an average increase in market access, these elasticities imply 5% and 13% increases in real agricultural income, respectively, for the OLS and IV estimations.[18] Panel B reports the results for the welfare effect that is attributed to the expansion of trade among villages; i.e., the dependent variable is $-\sum_k \widehat{\frac{\mu_n^k}{\gamma} \ln \pi_{vvt}^k}$. Column 1 uses the binary treatment approach, while Columns 2 and 3 report the result using the market access approach using OLS and IV estimation, respectively. The estimates in all three columns of Panel B are very similar to their counterparts reported in Panel A, which suggests that most of the estimated welfare gain from roads is actually attributed to the expansion of trade among the villages.

Table A.1 reports analogous results as table 3, but restricts the estimation sample to villages that have complete data both on agricultural production and crop prices (villages that overlap in both AgSS and AgPPS datasets). The estimated welfare effects in Column 1 are statistically significant and very similar in magnitude to their counterparts in table 3. The estimates in Column 2 are very similar to their counterparts in table 3, although they are less precisely estimated and statistically borderline insignificant, perhaps due to a significant drop in the sample size compared with table 3. Finally, the IV estimates in Column 3 are statistically significant and larger than their counterparts in table 3. Overall, comparing tables 3 and A.1 suggests that the welfare gain estimate in table 3 and the welfare gain attributed to the trade mechanism are not driven by the imputation of prices for AgSS villages that are not covered by the AgPPS price survey.

In table A.6, I explore the robustness of the above results to alternative ways of

[18] These values are obtained by multiplying the estimated elasticity by the average increase in market access: 0.114*0.45 and 0.295*0.45.

addressing the endogeneity of roads – matching-based DID strategy. That is, I first obtain a matched sample of treated and nontreated villages based on a set of observed village characteristics that might affect the selection of villages for the program before conducting the DID estimation (see Appendix E for details of this procedure). Column 1 reports the results for binary treatment, while Column 2 reports the results for the market access approach. These estimates are very close to their counterparts in 3, implying that our results are robust to alternative ways of addressing the potential endogeneity of road placement.

Overall, the results in tables 3 and A.6 show that URRAP roads have led to significant increases in village welfare, and almost all of the welfare gain is attributed to the expansion of trade among the villages. In the next subsections, I explore this trade mechanism in more detail. In particular, for the welfare gain to be caused by the expansion of trade, two conditions are expected to be present. The first is that villages should experience an increase in the relative prices of their comparative advantage crops. Second, villages should reallocate more farmlands to these crops. We explore whether the data support these predictions in the next subsection.

4.3 Mechanisms

The theoretical results in section 3 suggest that decreases in trade costs lead to increases in the relative prices of a village's CA crops and reallocation of farmland toward these crops. I test this directly. To identify a village's CA crop(s), I primarily use village-level crop yield estimates provided in the AgSS. These yield estimates are conducted by trained enumerators using a method called *crop cut* where they take a sample plot of 4 square meters and conduct a crop cut to obtain yield estimates. However, there are two caveats to this yield estimate. First, yield estimates are provided only for the crops that are actually produced in the village. Second, such yield estimates are influenced by seasonal fluctuations in climatic factors such as rainfall and crop diseases. I overcome the second problem by averaging yield estimates across pre-URRAP years within a village. For the first problem, I assume a yield of zero for crops that are not produced in a village.

As a robustness check, I use FAO-GAEZ data on the agroclimatically attainable yield of crops in each village. These data use a number of agroecological, soil and climatic factors, and sophisticated agronomic models to provide yield estimates at 5 arc-minute resolution under three scenarios of intensity of input usage (low, medium, and high input intensity). I use yield estimates under the low input intensity, which is more likely to reflect the reality in Ethiopia. While the FAO-GAEZ data overcome the above two problems with the yield estimates provided in AgSS data, they cover only a partial list of crops produced in Ethiopia. In particular, these data miss some of the widely grown endemic crops in Ethiopia, such as enset and teff. Nevertheless, I use this alternative yield estimate for a robustness check. The correlation between the FAO-GAEZ yield estimates and the AgSS yield estimates, based on crops that exist in both datasets, is approximately 0.8, and all my results are robust across the alternative yield estimates used.

Given the yield estimates A_v^k, I define a village's CA crops using the following procedure. First, I calculate a village's yield relative to the national average for each crop, $\frac{A_v^k}{A^k}$. Next, I rank crops within each village based on their yields relative to the national average. I define crops in the top 20% of the ranking as my baseline village comparative advantage crops. I relax this baseline threshold to top the 30%, top 40%, etc., to see the sensitivity of my results. One issue is that some villages grow only a handful of the major crops considered for analysis and using the above procedure would end up classifying all or most of the crops grown in the village as CA crops in these villages.[19] For instance, if a village grows only 5 of, say, 25 crops in the data and we define CA crops as those in the top 20% in relative yield, then all 5 crops grown in the village are classified as CA crops. As a result, we do not discern any land reallocation because we are defining the set of all crops grown in the village as CA crops. To overcome this problem, I keep only crops that are grown in a village in at least one of the pre- or post-URRAP years and rank these crops in their relative productivity $\frac{A_v^k}{A^k}$ within the village. In this approach, for a village that grows only 5 of the 25 crops, the CA crop is the crop that is at the top in ranking of $\frac{A_v^k}{A^k}$ within the village.

[19] Approximately 4% of the villages grow five or fewer crops, and the maximum number of crops grown in a given village is 19 out of the 25 major crops.

4.3.1 New roads and crop prices

As trade costs between a village and its trading partners decrease due to new road construction, the relative prices of the village's CAc rops increase (or the relative prices of the village's CD crops decrease). This model prediction can be directly tested, given data on village-level prices of crops before and after the program and our CA crop definition from the previous subsection. I use the following regression to test this prediction:

$$\begin{aligned}\ln p^k_{vmt} &= \alpha_1 \ln\mathrm{MA}_{vt} + \alpha_2 CA^k_v + \alpha_3(\ln\mathrm{MA}_{vt} * CA^k_v) + \beta\mathbf{X} \\ &\quad + \gamma^k + \gamma_v + \gamma_{mt} + \varepsilon^k_{vt}\end{aligned}$$

where $\ln p^k_{vmt}$ denotes the log price of crop k in village v in month m of year t, $\ln\mathrm{MA}_{vt}$ is log market access, and CA^k_v is a dummy variable indicating whether crop k is among the village's CA crops. I include village, crop, and crop-month fixed effects. Crop-month fixed effects are included to account for seasonal fluctuations in crop prices. Including village-year fixed effects (instead of village fixed effects) has no significant effect on the estimated results and the R-square. I run this regression using both OLS and IV approaches.

I use the AgPPS data for the same 25 major nontree crops in the main analysis.These data cover approximately 450 villages (after cleaning for missing information).[20] The results are reported in table 4. Panel A reports the OLS results, and Panel B reports the IV results. In both panels, the estimated effects of market access on village CA crop prices are statistically and economically significant. Focusing on the IV estimation results in Panel B, the estimates in Column 1 imply that the prices of CA crops (relative to prices of CD crops) increased by 0.036 between 2012 and 2016 following a one log unit increase in market access, implying a 1.62% increase in prices of CA crops (relative to CD crops) for a village with an average (45%) increase in market access.[21]. However, this estimate significantly drops as we relax the cutoff for the definition of CA crops, eventually becoming statistically insignificant in Columns 3 and 4 of Panel B.

[20]Note that the analysis in this subsection uses only villages for which complete price data are available.

[21]Note that this interpretation ignores the coefficients of lnMA which are statistically insignificant and close to zero in both Panel A and Panel B.

I explore a number of robustness exercises. First, I include all 45 crops for which price data are available. The results are reported in table A.2. Clearly the results are very similar to those in table 4. Second, I use GAEZ data to define CA crops. The results are reported in table A.3. The estimated increases in prices of CA crops are significantly larger in table A.3 compared with those in tables A.2 and 4. This is partly driven by differences in the composition of crops across the specifications. Finally, I use a matching-based DID estimation to address the endogeneity of road placement. The results are reported in table A.7 and are very similar to those in the main specification in table 4.

Overall, these results indicate that the relative prices of CA crops have increased significantly following increases in market access.

Next, I explore the effect of increases in market access on the village price index, which is computed using equation 5 and the elasticity of substitution estimated in section 4. The results in table 5 show that the village price index did not change much. This is not surprising given that the village price index is composed of the prices of CA crops and CD crops, which move in opposite directions when market access improves.

4.3.2 New roads and the reallocation of farmland

Another key prediction of the theoretical model is that villages reallocate more farmland toward crops whose relative prices increase as trade costs decrease. To test this, I estimate the following regression at the village level:

$$\begin{aligned}\eta_{vt}^{k} &= \alpha_1 \ln\text{MA}_{vt} + \alpha_2 CA_v^k + \alpha_3 (\ln\text{MA}_{vt} * CA_v^k) + \beta \mathbf{X} \\ &\quad + \gamma^k + \gamma_v + \gamma_t + \varepsilon_{vt}^k \end{aligned} \tag{14}$$

where η_{vt}^{k} denotes the share of land allocated to crop k in village v in year t, lnMA is log market access, CA_v^k is a dummy variable indicating whether crop k is among the village's CA crops, and X is a vector of village characteristics such as the preroad construction population density and rainfall. I also include crop fixed effects to account for the mean variation in land intensity across crops. The regression includes village fixed effects to account for time-invariant village characteristics that may confound our results and year fixed effects to account for any year specific factor shared across villages. I estimate using both OLS and IV strategies using predicted market access as an IV. The analysis is based on the same 25 nontree crops.

Table 6 presents the results. Panel A shows that the fraction of land allocated to CA crops increases significantly for villages that have seen an increase in their market access and that this effect becomes weaker as we relax the cutoff for the CA crop definition. The estimate in the first column shows that the fraction of land allocated to CA crops increased by approximately 2.5% for a village with an average increase in market access of 45%.[22] Columns 2 - 4 of Panel A clearly show that the estimated increase in fraction of land reallocated to CA crops decreases steadily as we relax the definition of CA crops to the top 30%, top 40%, etc. Panel B reports the IV results. Clearly, the IV results are very similar to the OLS result in Panel A. In particular, the estimates for the interaction term lnMA $* CA_v^k$ are identical to those in Panel A. This is because the regression is basically triple difference and that identification comes from variation across crops (CA crops vs. CD crops). Hence, the bias from the endogeneity of market access is minimal, and the OLS and IV results are very similar. Overall, the results in table 6 show that road expansion led to small reallocation of farmland towards villages' CA-crops.

I explore robustness exercises similar to the previous subsection. First, I include all 45 crops reported in AgSS data for which complete information is available. Table A.4 shows the estimation result. The estimated increases in the land share of CA crops based on Panel A and Panel B are 2.8% and 3.2%, respectively. Overall, the results in table A.4 are comparable to those in table 6, implying that the results are not sensitive to the selection of crops.

In the second robustness exercise, I use a matching-based DID estimation strategy (see Appendix E) to address potential bias in the selection of villages for the URRAP program. The results are reported in table A.8, and are very similar to their counterparts

[22]To obtain this percentage change, take the partial derivative of equation 14 with respect to lnMA to obtain -0.004+0.013*CA crop. For CA-crop=1, this equals 0.009. Multiply this by the average change in market access of 0.45, and to obtain the percentage change, divide it by land share of CA crops in 2012 given in the next to last row of table 6 and multiply by 100: $\frac{0.009*0.45}{0.159} * 100 = 2.5\%$.

in table 6. This implies that our results are robust to alternative methods used to account for potential selection bias.

5 Heterogeneity in welfare gains across villages

5.1 Cereal vs. cashcrop villages

One of the key testable implications from the model is that the welfare gain from roads is heterogeneous and depends on the fraction of land allocated to different crops in a village vis-a-vis the expenditure share of these crops in village consumption. In particular, villages that specialize in cereals should gain less from roads because they experience an increase in the prices of their consumption baskets, while noncereal-producing villages experience the opposite. That is, while both the cereal and noncereal (cashcrop) villages obtain better prices for their CA crops following improvement in market access, they experience different outcomes on the cost of their consumption baskets. Cereal villages experience increases in the costs of their consumption baskets because cereals (which account for 23% of CPI in Ethiopia, Durevall et al. (2013)) become more expensive locally. Cashcrop villages experience the opposite because cereals can now be imported at relatively cheaper prices and an increase in the prices of cashcrops has little effect on consumption costs because cashcrops have an insignificant weight in consumption baskets.

Table 7 presents the empirical evidence supporting this result. I interact the fraction of village farmland allocated to cereals in 2011 with the treatment and post dummies (in the first column) and with the market access measure (in Columns 2 and 3). To facilitate interpretation, the land share of cereal is standardized. The table shows that the welfare gain from roads is decreasing in the fraction of land allocated cereals; i.e., villages that allocate a higher fraction of land to cereal gain significantly less than those that allocate a lower fraction of land to cereals. The estimates in Column 1 imply that villages with a one-standard-deviation higher fraction of farmland allocated to cereals, compared with the average village, gain 11% lower in welfare compared with the average welfare gain of 17.5%. The estimates based on market access approach imply less dramatic heterogeneity between cereal and cash-crop villages in terms of welfare gain. For instance, the IV specification in Column 3 implies a 13% welfare gain for a village with an average increase in market access and average cereal share of land. Villages with two standard deviations higher in the share of land allocated to cereals, compared with the average, experience a slightly smaller welfare gain of 11%.

5.2 Remoteness and welfare gain from new roads

An important policy question is how the gain from infrastructure expansion, such as road connection, is distributed among villages of different characteristics. Would remote villages gain significantly more or less than villages that are located near roads or population centers? Theoretically, the answer is ambiguous. On the one hand, nonremote villages gain less from road expansion because they would face competition from remote villages that may now have improved access to market centers. On the other hand, the decrease in trade costs may not be large enough for remote villages to engage in trade with distant markets but are significant enough for villages near markets to engage in trade. In this subsection, I use villages' distance from the nearest town (population center with more than 20,000 population), distance to major trunk roads, which are considered the main trade routes (trunk roads include interstate roads, roads connecting the Zone capitals to Addis Ababa and roads that connect to neighboring countries), and distance to pre-URRAP road networks as a measure of remoteness.

Table 8 reports the estimation results. To facilitate interpretation, the distance measures are standardized so that the coefficients are interpreted as the effect of one standard deviation increase in distance relative to average. Panel A reports the OLS estimates based on the market access approach, and Panel B reports the corresponding IV results. Across all three columns, remoteness is associated with lower welfare gain from road expansion. Focusing on the IV results in Panel B, Column 1 implies that villages that are one standard deviation further from the population center (compared with a village with average distance) gain about 50% less in welfare (compared to a village with average distance). Column 2 implies that villages that are one standard deviation further from trunk roads (compared with a village with average distance) gain 34% less welfare. Column 3 implies that villages that are one standard deviation further from pre-URRAP roads (compared with a village with average distance) gain approximately 21% less than the welfare gain by a village with average distance. In summary, remote villages gain much less welfare than do nonremote villages, suggesting that authorities may also have an economic reason to prioritize nonremote villages in terms road infrastructure, in addition to the obvious technical reason of connecting nearby villages first to facilitate access to remote villages.

6 Conclusions

In this paper, I estimate the welfare gain from massive rural road expansion in Ethiopia. To explore the key mechanisms through which roads affect village welfare, I develop a Ricardian trade model with multicrop multilocation feature where land productivity is allowed to vary both within and across villages. On the demand side, each village maximizes utility by choosing optimal quantities of crops to consume, given prices. On the production side, villages decide how to allocate their limited farmland across potential crops, given local prices and local crop productivity. A village also engages in costly trade with other villages.

The model gives sharp predictions about the effects of decreases in trade costs on village land allocation across crops and village prices of crops. However, the model implies that the size of the welfare gain from roads could vary across villages. While a decrease in trade cost leads to increases in village income by inducing reallocation of crops toward village CA crops and enables farmers to receive better prices for these crops, it also leads to an increase in the cost of consumption baskets for the villagers. The size of the net welfare gain depends on the strength of these two contrasting forces, which in turn depends on composition of the villages' CA crops vis-á-vis their consumption baskets.

I directly test the model's predictions using micro data on agricultural production, crop prices, and geospatial data on the entire road network before and after the road expansion program. To address the potential endogeneity of road placements, I use a counterfactual road network predicted solely based on cost considerations (land gradient and location of rivers and lakes) to construct an instrumental variable for the actual road network.

I estimate a total welfare gain of 13% from the road expansion between 2012 and 2016 for a village with an average increase in market access and show that this is attributed to the mechanisms suggested in the trade model. That is, the road expansion led to (i) significant increases in the relative prices of CA crops and (ii) the reallocation of farmland towards a village's CA crops. These mechanisms account for the lions share of the estimated welfare gain.

Future studies should undertake a detailed analysis of the effect of decreases in trade costs on the incentives to invest in improved agricultural techniques, such as the adoption of chemical fertilizers and irrigation systems. Theoretically, the returns on investment in these technologies improve as market access of villages improves. Additionally, improvements in road conditions enhance the accessibility of imported agricultural inputs.

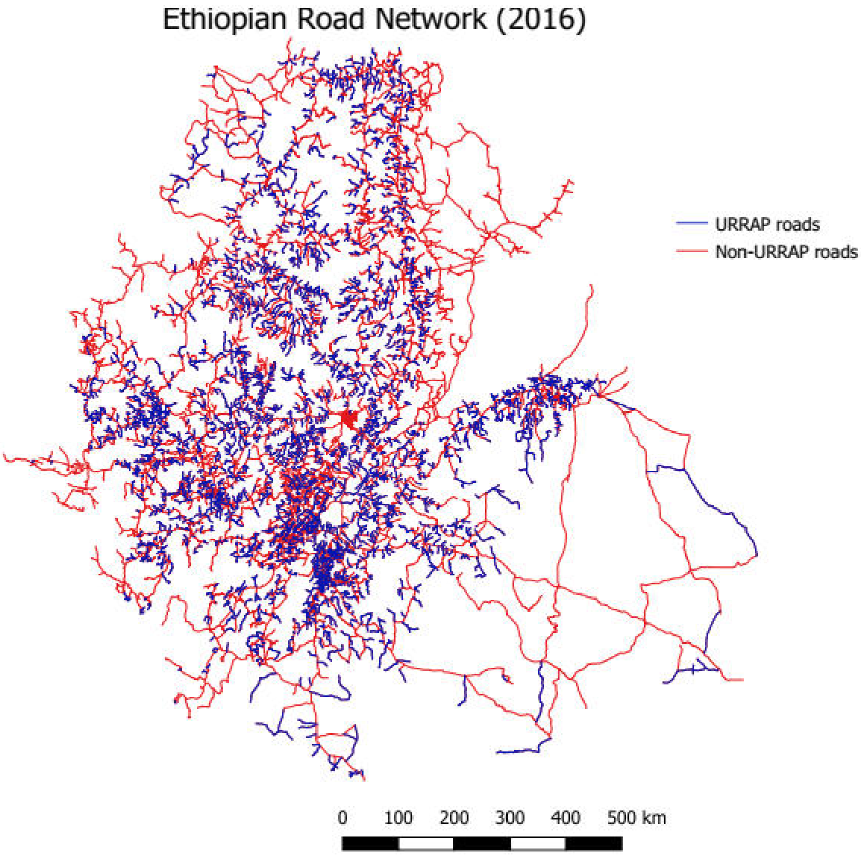


Figure 1: Rural road expansion under the URRAP

Figure 2: Completed URRAP roads (pictures were taken by Oromia Roads Authority).

Table 1: Market access measure and first-stage regression

	Dependent variable: log actual market access		
	(1)	(2)	(3)
Post	0.456***	0.252***	-0.130***
	(0.009)	(0.006)	(0.022)
URRAP			
Post*URRAP		0.420***	
		(0.015)	
LogPredictedMarketAccess			0.560***
			(0.024)
N	4042	4042	4042
R^2	0.937	0.954	0.957

Notes: Robust standard errors are in parentheses. All regressions include village fixed effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2: URRAP road access and trade costs

	Dependent variable: $\lvert logPrice_{zone} - logPrice_{village} \rvert$	
	All crops	Vegetables and fruits
Post * URRAP	-0.022*	-0.047
	(0.013)	(0.035)
N	82944	24468
R^2	0.458	0.358
Mean of dep var	0.35	0.67

Notes: Standard errors are clustered at the village level. This table is based on AgPPS and RPS datasets, which are monthly crop data at village and urban centers, respectively. The regression includes 422 villages, 57 urban centers, and 56 crops. All regressions include village, crop-month, and year fixed effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Welfare gains from new rural roads

	Binary treatment	Market access approach	
	OLS	OLS	IV
	Panel A: Dep var. is real agricultural income per hectare		
Post*URRAP	0.202***		
	(0.044)		
LogMarketAccess		0.114**	0.295***
		(0.051)	(0.094)
N	4042	4042	4042
R^2	0.850	0.849	0.002
	Panel B: The dependent variable $-\sum_k \widehat{\frac{\mu_n^k}{\gamma} \ln \pi_{vvt}^k}$		
Post*URRAP	0.202***		
	(0.044)		
LogMarketAccess		0.114**	0.296***
		(0.051)	(0.094)
N	4042	4042	4042
R^2	0.851	0.850	0.002

Notes: Bootstrap standard errors are clustered at the village levels and in parentheses. All regressions include year and village fixed effects, and log rainfall. The agricultural production data come from AgSS. The price data come from AgPPS. The estimation is based on 25 nontree crops for which full information on prices and quantities of production is available. Column 1 uses a binary treatment dummy. Columns 2 and 3 use a market access measure constructed from the entire road network before and after the program, and pre-program spatial population distribution. In Column 3, a market access measure constructed from *predicted* road network using land gradient and location of rivers and lakes is used as an instrument for market access measure constructed from *actual* road network. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: New road construction and prices of CA crops

	Definition of CA crops:			
	Top 20%	Top 30%	Top 40%	Top 50%
	Panel A: Market access approach – OLS			
CA crop	-0.303**	-0.305**	-0.293**	-0.306**
	(0.125)	(0.122)	(0.134)	(0.149)
LogMarketAccess	0.001	-0.003	-0.005	-0.008
	(0.025)	(0.024)	(0.024)	(0.025)
LogMarketAccess * CA-crop	0.029**	0.028**	0.026*	0.026*
	(0.012)	(0.012)	(0.013)	(0.015)
N	31737	31737	31737	31737
adj. R^2	0.800	0.800	0.801	0.801
	Panel B: Market access approach – IV			
CA crop	-0.369***	-0.318**	-0.253*	-0.234
	(0.134)	(0.131)	(0.133)	(0.146)
LogMarketAccess	-0.028	-0.030	-0.030	-0.028
	(0.051)	(0.050)	(0.050)	(0.050)
LogMarketAccess * CA-crop	0.036***	0.029**	0.022	0.018
	(0.013)	(0.013)	(0.013)	(0.015)
N	31737	31737	31737	31737
adj. R^2	0.728	0.728	0.728	0.729

Notes: Standard errors are clustered at the village level. The dependent variable is the log village crop price. The analysis includes 25 major nontree crops and over 450 nationally representative rural villages. Each column represents different cutoffs for the definition of CA crops. In Column 1, CA crops are crops in the top 20% of the within village ranking of crops based on yield relative to national average. In Column 2, CA crops are in the top 30% in the ranking, and so forth. All regressions include village, crop-month and year fixed effects. Panel A reports OLS estimation results using a market access measure constructed from the entire road network before and after the program, and pre-program spatial population distribution. Panel B uses a market access measure constructed from *predicted* road network using land gradient and location of rivers and lakes as an instrument for market access measure constructed from *actual* road network. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: New road construction and village price index

	Binary treatment	Market access – OLS	Market access – IV
Post*URRAP	0.009		
	(0.010)		
LogMarketAccess		-0.000	0.006
		(0.012)	(0.020)
N	4148	4148	4148
R^2	0.948	0.948	0.054

Notes: Robust standard errors are in parentheses. The regression includes 450 villages. The price index is computed from the 25 major nontree crops in the main analysis and using elasticity of substitution estimated in section 4. All regressions include village and month fixed effects.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6: Road construction and reallocation of farmland towards CA crops

	Definition of CA crops:			
	Top 20%	Top 30%	Top 40%	Top 50%
	Panel A: Market access approach – OLS			
CA crop	-0.080**	-0.054*	-0.015	-0.007
	(0.034)	(0.030)	(0.028)	(0.027)
LogMarketAccess	-0.004**	-0.005**	-0.005**	-0.005**
	(0.002)	(0.002)	(0.002)	(0.002)
CA crop*LogMarketAccess	0.013***	0.011***	0.007**	0.007**
	(0.003)	(0.003)	(0.003)	(0.003)
N	34839	34839	34839	34839
adj. R^2	0.340	0.348	0.353	0.352
	Panel B: Market access approach – IV			
CA crop	-0.082**	-0.054*	-0.014	-0.009
	(0.033)	(0.029)	(0.027)	(0.026)
LogMarketAccess	-0.005	-0.005*	-0.005	-0.005
	(0.003)	(0.003)	(0.003)	(0.003)
CA crop*LogMarketAccess	0.013***	0.011***	0.007***	0.007**
	(0.003)	(0.003)	(0.003)	(0.003)
N	34815	34815	34815	34815
adj. R^2	0.278	0.286	0.292	0.291
Mean land share (CA crops) in 2012	0.159	0.162	0.160	0.153
Mean land share (all) in 2012	0.118	0.118	0.118	0.118

Notes: Standard errors are clustered at the village level. These regressions include 25 major nontree crops. Each column represents different cutoffs for the definition of CA crops. In column 1, CA crops are crops in the top 20% of within village ranking of crops based on yield relative to national average. In Column 2, CA-crops are the top 30% in the ranking, and so forth. All regressions include village, crop and year fixed effects. Panel A reports the OLS result using a market access measure constructed from the entire road network before and after the program, and preprogram spatial population distribution. Panel B uses a market access measure constructed from *predicted* road network using land gradient and location of rivers and lakes as an instrument for market access measure constructed from *actual* road network. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: Welfare gain: cereal vs. noncereal villages

	Dep. var. is log real revenue per hectare ln$W$		
	Binary treatment	Market access: OLS	Market access: IV
Post*URRAP	0.175***		
	(0.064)		
Post*URRAP*CerealShare	-0.111**		
	(0.046)		
LogMarketAccess		0.150*	0.535***
		(0.083)	(0.143)
LogMarketAccess*CerealShare		-0.008***	-0.008***
		(0.003)	(0.003)
N	4042	4042	4042
adj. R^2	0.749	0.749	-0.890

Notes: Bootstrap standard errors are clustered at the village level and in parentheses. The first column reports results based on binary treatment dummy. The second column reports the OLS results based on a market access approach. The third column reports the IV results where market access from *predicted* road network is used as an IV for market access measure based on *actual* road network. All regressions include year and village fixed effects. The estimation is based on 25 nontree crops for which full information on prices and quantities of production is available. $CerealShare$ is the share of farmland allocated to cereal crops in a village in the year 2011. Cereal crops include Barley, Wheat, Maize, Teff, Sorghum, Millet, and Enset. Noncereal crops include all vegetables, legumes and cashcrops which are predominantly produced for market. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 8: Remoteness and welfare gain from new roads

	Dep. var. is log real revenue per hectare ln$W$		
	Distance to town	Distance to trunk roads	Distance to pre-URRAP roads
	Panel A: Market access approach – OLS		
LogMarketAccess	0.105*	0.294***	0.260***
	(0.061)	(0.065)	(0.064)
LogMarketAccess * Distance	-0.339***	-0.285***	-0.170***
	(0.052)	(0.044)	(0.036)
N	4042	4042	4042
adj. R^2	0.596	0.596	0.592
	Panel B: Market access approach – IV		
LogMarketAccess	0.014	0.414***	0.367***
	(0.112)	(0.121)	(0.122)
LogMarketAccess * Distance	-0.499***	-0.336***	-0.214***
	(0.071)	(0.060)	(0.048)
N	4042	4042	4042

Notes: Bootstrap standard errors are clustered at the village level are in parentheses. Distance measure is standardized so that the coefficients are interpreted as the effect of one standard deviation increase in distance. The first column uses distance to the nearest town with a population above 20,000. The second column uses distance to the nearest trunk road (interstate roads and roads connecting zone capitals to the center). The third column uses distance to a preexisting road network. All regressions include year and village fixed effects. The estimation is based on 25 nontree crops for which full information on prices and quantities of production is available. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendices

A Construction of market access measure

I follow Donaldson and Hornbeck (2016) to construct a market access measure for each village derived from general equilibrium trade models. I use the spatial distribution of the population before the road program and pre- and postprogram the entire road network of the country to construct the market access measure.

$$MarketAccess_{ot} = \sum_{d} \tau_{odt}^{-\theta} Population_d \tag{15}$$

where $Population_d$ is the destination village population from the 2007 census (before the onset of the URRAP program). The use of the pre-URRAP population distribution is necessary because the population distribution is likely to respond to improvements in road infrastructure. θ is the inverse land heterogeneity parameter (which governs trade elasticity), which I estimate in section 4.

τ_{odt} is the freight costs of transporting one ton of cargo from origin village o to destination village d along the least cost path, before $(t = 0)$ and after $(t = 1)$ the construction of URRAP roads.[23] I use the following procedure to estimate τ_{odt} for each year. First, I construct a link from each village centroid to the nearest available road in year t. Next, I use data on costs of moving weight (in USD per ton-kilometer) for five different road quality levels: asphalt, major gravel, cobbled road, minor gravel, and earth road. Because there are no similar cost estimates along the link roads, I scale up the costs along the earth road by the factor of $\frac{Cost\ along\ earth\ road}{Cost\ along\ minor\ gravel}$ to obtain an estimate of cost along the links.[24] After assigning each road type (including the links) with the estimated costs in USD per ton-kilometer, I use the Network Analysis tool in ArcGIS to calculate the costs (in USD) of moving a ton of weight from origin o to destination d along the least cost path, in each year. I use these estimates as τ_{odt}. As seen in equation 15, a change to

[23] Alternatively, I use travel time along the least (time) cost path, instead of freight costs. The market access measures are strongly correlated (correlation of 0.92).

[24] I show that the results are robust to using alternative scales that are half or twice of the baseline scale $\frac{Cost\ along\ earth\ road}{Cost\ along\ minor\ gravel}$.

a village's market access comes only from changes in τ_{odt}, which in turn comes from the construction of new roads.

B Construction of the instrumental variable

I use two pieces of information to construct the predicted road network. The first is the regional budget constraint, which is the total road length planned to be constructed in 2010 in each of the seven Ethiopian regions covered under the URRAP.[25] The second is a cost raster. I combine land gradient, location of rivers, and location of lakes data to obtain the cost raster, which gives me the costs of road construction in every 30-meter-by-30-meter grid cell. Given the cost raster and regional budget constraint, I use the following algorithm to obtain the predicted road network. First, I start by connecting villages that are adjacent to the pre-URRAP road network. Next, I connect villages that neighbor the first group to either the pre-URRAP road or to the roads constructed in step-1 (whichever option is less costly). I iterate this procedure until the regional budget is exhausted. For two of the seven regions (Oromia and Harari), the regional budget constraint is not binding. That is, I was able to connect all villages in those regions before exhausting the respective regional budgets. This is because the planned road lengths take into account several factors other than just construction costs and as a result, tend to be longer than the predicted roads based only on construction costs.

Once I obtain the predicted road network, I follow exactly the same procedure as in Appendix A to construct the market access measure based on the predicted roads. I assume the freight cost on all the predicted roads is equal to that of minor gravel roads (almost all the URRAP roads are minor gravel roads).

[25] While the initial URRAP plan included all nine Ethiopian regions, two regions (Afar and Somali) were later dropped from the project later.

C Derivation of the conditional distribution of productivity and rental rate

Because the distribution of the rental rate of a plot depends on the distribution of the productivity of land, we need to first derive the distribution of land productivity conditional on the land being used for crop k, i.e., $\mathcal{P}\Big(z_v^k(\omega)|\omega \in \Omega_v^k\Big)$, which I denote as $G_v^k(t)$. This derivation of the conditional distribution of land quality is similar to Sotelo (2020):

$$\begin{aligned}
G_v^k(t) &= \mathcal{P}\Big[z_v^k(\omega) < t | p_v^k z_v^k(\omega) = \max_{l\in\Delta_l} p_v^l z_v^l(\omega)\Big] \\
&= \frac{\mathcal{P}\Big[z_v^k(\omega) < t \wedge p_v^k z_v^k(\omega) = \max_{l\in\Delta_l} p_v^l z_v^l(\omega)\Big]}{\mathcal{P}\Big[p_v^k z_v^k(\omega) = \max_{l\in\Delta_l} p_v^l z_v^l(\omega)\Big]} \\
&= \frac{1}{\eta_v^k}\mathcal{P}\Big[z_v^k(\omega) < t \wedge p_v^l z_v^l(\omega) < p_v^k z_v^k(\omega), \quad \forall l \neq k\Big] \\
&= \frac{1}{\eta_v^k}\mathcal{P}\Big[\frac{p_v^l}{p_v^k} z_v^l(\omega) < z_v^k(\omega) < t, \quad \forall l \neq k\Big] \\
&= \frac{1}{\eta_v^k}\int_0^t \Pi_{l\neq k}\mathcal{P}\Big[\frac{p_v^l}{p_v^k} z_v^l(\omega) < z\Big] f_v^k(z)dz
\end{aligned}$$

Using the distribution of $z_v^k(\omega)$, we obtain the following:

$$\begin{aligned}
G_v^k(t) &= \frac{1}{\eta_v^k}\int_0^t \Pi_{l\neq k}\exp\Bigg((-A_v^l)^\theta(\frac{p_v^k}{p_v^l}z)^{-\theta}\Bigg) \quad \theta(A_v^k)^\theta z^{-\theta-1}\exp\Bigg(-(A_v^k)^\theta z^{-\theta}\Bigg)dz \\
&= \frac{1}{\eta_v^k}\int_0^t \exp\Bigg(-\Big(p_v^k z\Big)^{-\theta}\sum_{l\neq k}\Big(A_v^l p_v^l\Big)^\theta\Bigg)\exp\Bigg(-(A_v^k)^\theta z^{-\theta}\Bigg) \quad \theta(A_v^k)^\theta z^{-\theta-1}dz \\
&= \frac{1}{\eta_v^k}\int_0^t \exp\Bigg(-(p_v^k z)^{-\theta}\sum_{l}\Big(A_v^l p_v^l\Big)^\theta\Bigg) \quad \theta(A_v^k)^\theta z^{-\theta-1}dz \\
&= \int_0^t \exp\Bigg(-(p_v^k z)^{-\theta}\Phi_v^\theta\Bigg) \quad \theta\Phi_v^\theta(p_v^k)^{-\theta}z^{-\theta-1}dz, \quad \text{where} \quad \Phi_v = \Bigg(\sum_l\Big(A_v^l p_v^l\Big)^\theta\Bigg)^{\frac{1}{\theta}} \\
&= \exp\Bigg(-\Big(\frac{\Phi_v}{p_v^k}\Big)^\theta t^{-\theta}\Bigg)
\end{aligned}$$

Thus, the distribution of productivity of the set plots in village v which are covered by crop k is a Fréchet with the parameters $\frac{\Phi_v}{p_v^k}$ and θ. Note that the average productivity of land covered with a crop decreases with the crop price. Intuitively, an increasing amount of land is allocated to a crop with a higher price, which leads to a decrease in the average quality of land allocated to the crop.

Recall that the rental rate on plot ω, conditional on ω being used for crop k, is given by $r(\omega) = p_v^k z_v^k(\omega)$. Thus, the conditional distribution of rental rate $r(\omega)|\omega \in \Omega_v^k$ is Fréchet with parameters Φ_v and θ. That is, the rental rate of plots covered with different crops has the same distribution regardless of which crops are planted. This result follows from the property of the Fréchet distribution and the fact that $r(\omega)$ is homogeneous of degree one in crop prices.

D Appendix Tables

Table A.1: Welfare gains from new rural roads: restricting estimation to only villages that have both agricultural and price data

	Binary treatment	Market access approach	
	OLS	OLS	IV
	Panel A: Dep var. is real agricultural income per hectare		
Post*URRAP	0.214*		
	(0.114)		
LogMarketAccess		0.144	0.533*
		(0.130)	(0.285)
N	752	752	752
R^2	0.815	0.814	-0.011
	Panel B: Dependent variable $-\sum_k \widehat{\frac{\mu_n^k}{\gamma} \ln \pi_{vvt}^k}$		
Post*URRAP	0.214*		
	(0.115)		
LogMarketAccess		0.145	0.534*
		(0.130)	(0.286)
N	752	752	752
R^2	0.816	0.814	-0.011

Notes: Bootstrap standard errors clustered at village levels are in parentheses. All regressions include year and village fixed effects, and log rainfall. The agricultural production data comes from AgSS. The price data come from AgPPS. The estimation is based on 25 nontree crops for which full information on prices and quantities of production is available. Column 1 uses a binary treatment dummy. Columns 2 and 3 use a market access measure constructed from the entire road network before and after the program, and preprogram spatial population distribution. In Column 3, a market access measure constructed from *predicted* road network using land gradient and location of rivers and lakes is used as an instrument for market access measure constructed from *actual* road network. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.2: New road construction and prices of CA crops: all crops

	Definition of CA crops:			
	Top 20%	Top 30%	Top 40%	Top 50%
	Panel A: Market access approach – OLS			
CA crop	-0.389***	-0.234**	-0.249**	-0.188
	(0.129)	(0.113)	(0.109)	(0.119)
LogMarketAccess	0.009	0.010	0.006	0.007
	(0.026)	(0.026)	(0.027)	(0.028)
LogMarketAccess * CA crop	0.036***	0.020*	0.021*	0.015
	(0.013)	(0.011)	(0.011)	(0.012)
N	41930	41930	41930	41930
adj. R^2	0.798	0.799	0.799	0.799
	Panel B: Market access approach – IV			
CA crop	-0.398***	-0.260**	-0.267**	-0.156
	(0.139)	(0.122)	(0.118)	(0.125)
LogMarketAccess	-0.040	-0.040	-0.043	-0.039
	(0.050)	(0.050)	(0.050)	(0.051)
LogMarketAccess * CA crop	0.037***	0.023*	0.023*	0.012
	(0.014)	(0.012)	(0.012)	(0.013)
N	41928	41928	41928	41928
adj. R^2	0.740	0.740	0.741	0.740

Notes: Standard errors are clustered at the village level. The dependent variable is the log village crop price.The analysis includes 45 crops and over 450 nationally representative rural villages. Each column represents different cutoffs for the definition of CA crops. In Column 1, CA crops are crops in the top 20% of within village ranking of crops based on yield relative to national average. In Column 2, CA crops are in the top 30% in the ranking, and so forth. All regressions include village, crop-month and year fixed effects. Panel A reports OLS results using market access measure constructed from the entire road network before and after the program, and preprogram spatial population distribution. Panel B uses market access measure constructed from *predicted* road network using land gradient and location of rivers and lakes as an instrument for market access measure constructed from *actual* road network. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.3: Road construction and prices of CA crops: based on GAEZ yield measure

	Definition of CA crops:			
	Top 20%	Top 30%	Top 40%	Top 50%
	Panel A: Market access approach – OLS			
CA crop	-0.591***	-0.506***	-0.415**	-0.294
	(0.179)	(0.195)	(0.199)	(0.196)
LogMarketAccess	0.014	0.010	0.012	0.015
	(0.032)	(0.032)	(0.032)	(0.033)
LogMarketAccess * CA crop	0.053***	0.045**	0.036*	0.025
	(0.018)	(0.020)	(0.020)	(0.020)
N	21428	21428	21428	21428
adj. R^2	0.761	0.760	0.760	0.760
	Panel B: Market access approach – IV			
CA crop	-1.018***	-0.879***	-0.771***	-0.665***
	(0.186)	(0.204)	(0.212)	(0.211)
LogMarketAccess	-0.038	-0.041	-0.038	-0.043
	(0.058)	(0.057)	(0.058)	(0.059)
LogMarketAccess * CA crop	0.096***	0.083***	0.072***	0.063***
	(0.019)	(0.021)	(0.021)	(0.021)
N	21423	21423	21423	21423
adj. R^2	0.686	0.686	0.686	0.685

Notes: Standard errors are clustered at the village level. The dependent variable is the log price of crops. This regression analysis includes 19 crops for which GAEZ yield measure is available. Each column represents different cutoffs for definition of CA crops. In Column 1, CA crops are crops in the top 20% of within village ranking of crops based on yield relative to national average. In Column 2, CA crops are in the top 30% in the ranking, and so forth. All regressions include village, year and crop-month fixed effects. Panel A reports the OLS result using a market access measure constructed from the entire road network before and after the program, and preprogram spatial population distribution. Panel B uses a market access measure constructed from *predicted* road network using the land gradient and location of rivers and lakes as an instrument for market access measure constructed from *actual* road network. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.4: Road construction and reallocation of farmland towards CA crops: all crops

	Definition of CA crops:			
	Top 20%	Top 30%	Top 40%	Top 50%
	Panel B: Market access approach – OLS			
CA crop	-0.060***	-0.041**	-0.036**	-0.018
	(0.020)	(0.017)	(0.016)	(0.015)
LogMarketAccess	-0.003**	-0.003***	-0.004***	-0.004***
	(0.001)	(0.001)	(0.001)	(0.001)
CA crop*LogMarketAccess	0.009***	0.007***	0.007***	0.005***
	(0.002)	(0.002)	(0.002)	(0.002)
N	56211	56211	56211	56211
adj. R^2	0.359	0.364	0.366	0.365
	Panel C: Market access approach – IV			
CA crop	-0.068***	-0.045***	-0.041**	-0.024
	(0.020)	(0.017)	(0.016)	(0.015)
LogMarketAccess	-0.002	-0.003	-0.004**	-0.003*
	(0.002)	(0.002)	(0.002)	(0.002)
CA crop*LogMarketAccess	0.009***	0.008***	0.007***	0.006***
	(0.002)	(0.002)	(0.002)	(0.002)
N	56211	56211	56211	56211
adj. R^2	0.306	0.312	0.314	0.313
Mean land share (CA-crops) in 2012	0.097	0.101	0.101	0.096
Mean land share (all) in 2012	0.073	0.073	0.073	0.073

Notes: Standard errors are clustered at the village level. The analysis includes 45 crops. Each column represents different cutoffs for definition of CA crops. In column 1, CA crops are crops in the top 20% of within village ranking of crops based on yield relative to national average. In column 2, CA-crops are in the top 30% in the ranking, and so forth. All regressions include village, crop and year fixed effects. Panel A reports OLS result using market access measure constructed from the entire road network before and after the program, and pre-program spatial population distribution. Panel B uses market access measure constructed from *predicted* road network using land gradient and location of rivers and lakes as an instrument for market access measure constructed from *actual* road network. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

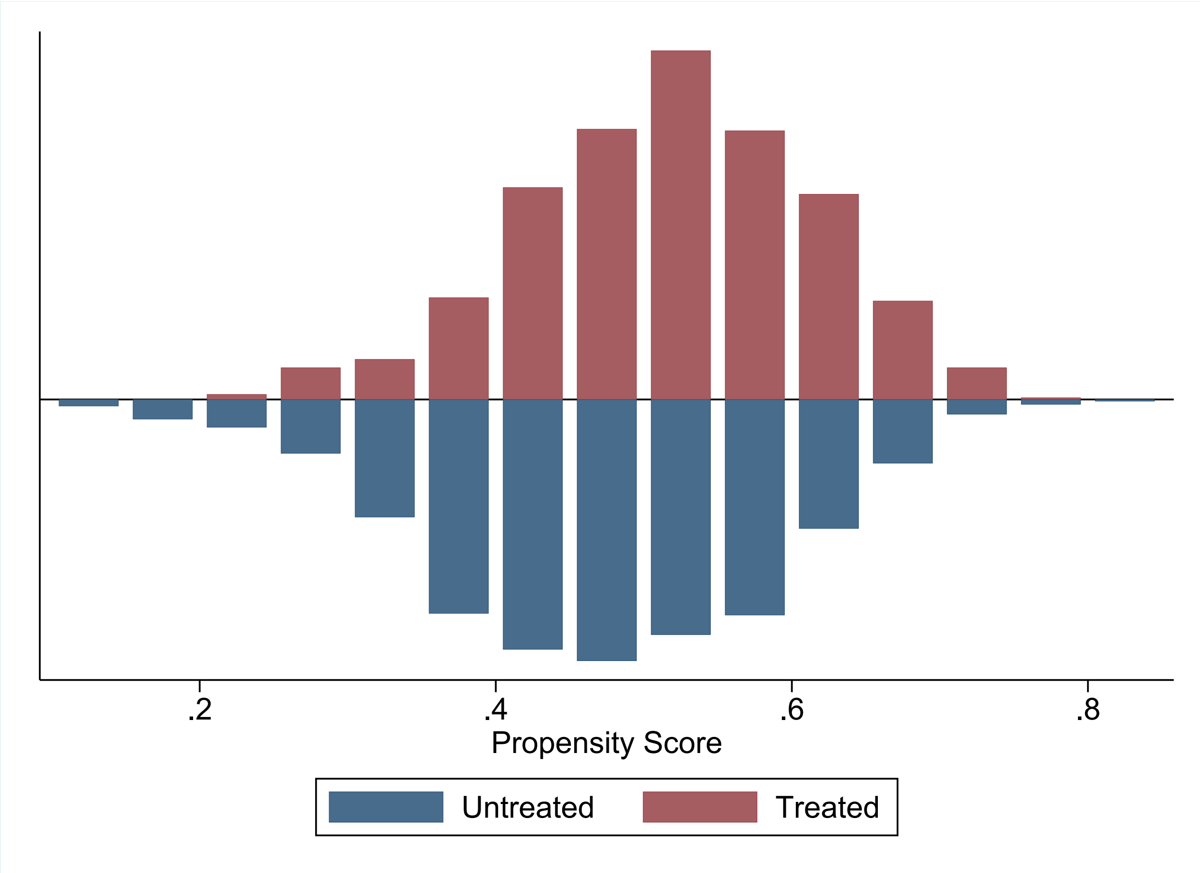
E Robustness: Matching-based difference-in-differences

As a further robustness exercise, I also use a matching-based difference-in-differences (MB-DID) strategy to address the endogeneity of road placement. That is, I first obtain a matched sample of treated and nontreated villages based on their observable characteristics that might be relevant for the selection of villages for URRAP. I then conduct a DID estimation using the matched sample. Combining matching with the DID strategy is a powerful approach to address the selection problem. The matching step enables me to compare treated villages with nontreated villages that have similar observed characteristics and, hence, have similar treatment probability. The DID strategy on these matched samples helps me to wash out unobserved time-invariant village characteristics that may confound the treatment effect.

To identify relevant village characteristics for matching treated and nontreated villages, I use information from officials at the Ethiopian Roads Authority (ERA), who suggest that the main factors determining whether a village would be selected for URRAP in a particular year are (1) the village's distance to the preexisting road network, (2) the population density of the village, and (3) the terrain and landscape of the village. Distance to preexisting roads is crucial because movement of machineries and other construction materials to the construction sites, by itself, requires roads that are passable by vehicles. Population density is relevant both for political considerations and the project's labor input requirements.[26] Finally, terrain and landscape significantly affect road construction costs. Villages that require many bridge constructions or cutting and digging of hills are usually less favorable due to cost considerations. Based on these insights from the officials, I use the following list of variables to match treated and nontreated villages: distance to nearest town, distance to preexisting road network, population size, average slope of land in the village, average elevation in the village, and average rainfall over the 1990-2010 period. The population and population density data used are from the 2007 census (preroad era). I use digital elevation model (DEM) data and ArcGIS tools to calculate average slope and elevation of each village.

[26]Most of the labor input for the URRAP roads are contributed by local residents, approximately three-quarters of which is free labor.

Figure A.1: Common support of propensity score matching



Given these village characteristics, I use the *gmatch* command in STATA (for its handiness in the panel data setting) to match treated and nontreated villages. For each treated village, the *gmatch* algorithm finds, nontreated village(s) that have the closest observed characteristics or propensity score. I conduct a DID estimation on these matched samples of treated and nontreated villages. Figure A.1 shows the histogram of propensity score by treatment status. The figure clearly shows that the region of common support is large as very few nontreated villages lie outside the common support. Table A.5 reports the balancing of the matching variables. All the t-statistics are insignificant and the bias percentage is small.

Table A.5: Balancing of variables for Average Treatment Effect on Treated (ATT)

	Treated	Control	% bias	t-stat	p-value
Distance to nearest town (meters)	10596	10417	2.2	0.57	0.571
Distance to nearest trunk road (meters)	4551.8	4280.5	6.2	1.47	0.143
Distance to preexisting road network (meters)	1320.5	1221.5	3.6	0.93	0.351
Population	5340.9	5458.1	-3.1	-0.63	0.528
Average slope (degrees)	9.9697	10.018	-0.9	-0.20	0.840
Average altitude (meters)	1946.2	1959.1	-2.3	-0.54	0.587
Rainfall (mm)	1183	1161.1	6.6	1.50	0.133

Notes: Population and rainfall correspond to the period before URRAP.

Table A.6: Welfare gains from new rural roads: matching-based DID estimation

	Binary treatment	Market access approach
Post*URRAP	0.128*	
	(0.064)	
LogMarketAccess		0.196**
		(0.081)
N	3268	3268
R^2	0.88	0.87

Notes: Robust standard errors are in parenthesis. All regressions include year and village fixed effects. The estimation is based on 25 nontree crops for which full information on prices and quantities of production is available. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.7: New road construction and prices of CA crops: matching-based DID estimation

	Definition of CA crops:			
	Top 20%	Top 30%	Top 40%	Top 50%
	Market access approach			
CA crop	-0.388**	-0.387***	-0.382**	-0.355*
	(0.151)	(0.141)	(0.161)	(0.181)
LogMarketAccess	-0.001	-0.006	-0.010	-0.011
	(0.026)	(0.025)	(0.026)	(0.026)
LogMarketAccess * CA crop	0.037**	0.036**	0.034**	0.031*
	(0.015)	(0.014)	(0.016)	(0.018)
N	26786	26786	26786	26786
adj. R^2	0.797	0.797	0.797	0.798

Notes: Standard errors are clustered at the village level. The dependent variable is the log village crop prices. The analysis includes 25 major nontree crops and over 450 nationally representative rural villages. Each column represents different cutoffs for the definition of CA crops. In column 1, CA crops are crops in the top 20% of within village ranking of crops based on yield relative to the national average. In column 2 CA-crops are top 30% in the ranking, and so forth. All regressions include village, crop-month and year fixed effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.8: Road construction and reallocation of farmland towards CA crops: matching-based DID estimation

	Definition of CA crops:			
	Top 20%	Top 30%	Top 40%	Top 50%
	Market access approach			
CA crop	-0.100**	-0.091**	-0.053	-0.014
	(0.040)	(0.035)	(0.033)	(0.031)
LogMarketAccess	-0.005***	-0.007***	-0.007***	-0.006**
	(0.002)	(0.002)	(0.002)	(0.003)
CA crop*LogMarketAccess	0.015***	0.014***	0.011***	0.007**
	(0.004)	(0.004)	(0.003)	(0.003)
N	28030	28030	28030	28030
adj. R^2	0.310	0.318	0.325	0.324
Mean land share (CA-crops) in 2012	0.159	0.162	0.160	0.153
Mean land share (all) in 2012	0.118	0.118	0.118	0.118

Notes: Standard errors are clustered at the village level. This regression analysis includes 25 major nontree crops. Each column represents different cutoff for definition of CA crops. In column 1, CA crops are crops in the top 20% of within village ranking of crops based on yield relative to the national average. In column 2 CA-crops are top 30% in the ranking, and so forth. All regressions include village, crop and year fixed effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

References

Adamopoulos, T. (2018). Spatial Integration, Agricultural Productivity, and Development: A Quantitative Analysis of Ethiopia's Road Expansion Program. ICG Working Paper F-32404-ETH-1.

Aggarwal, S. (2021). The long road to health: Healthcare utilization impacts of a road pavement policy in rural india. *Journal of Development Economics*, 151:102667.

Aggarwal, S., Giera, B., Jeong, D., Robinson, J., and Spearot, A. (2018). Market Access, Trade Costs, and Technology Adoption: Evidence from Northern Tanzania. Working Paper 25253, National Bureau of Economic Research.

Ali, R., Barra, A. F., Berg, C., Damania, R., Nash, J., and Russ, J. (2015). *Highways to Success or Byways to Waste: Estimating the Economic Benefits of Roads in Africa.* The World Bank.

Allen, T. and Arkolakis, C. (2014). Trade and the Topography of the Spatial Economy. *The Quarterly Journal of Economics*, 129(3):1085–1140.

Allen, T. and Arkolakis, C. (2020). The Welfare Effects of Transportation Infrastructure Improvement. *Review of Economic Studies*, page Forthcoming.

Allen, T. and Atkin, D. (2016). Volatility and the gains from trade. Working Paper 22276, National Bureau of Economic Research.

Arkolakis, C., Costinot, A., and Rodríguez-Clare, A. (2012). New Trade Models, Same Old Gains? *American Economic Review*, 102(1):94–130.

Asher, S. and Novosad, P. (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. Working Paper 21439, National Bureau of Economic Research.

Cosar, A. K. and Demir, B. (2016). Domestic road infrastructure and international trade: Evidence from Turkey. *Journal of Development Economics*, 118(C):232–244.

Costinot, A. and Donaldson, D. (2012). Ricardo's Theory of Comparative Advantage: Old Idea, New Evidence. *American Economic Review*, 102(3):453–58.

Costinot, A. and Donaldson, D. (2016). How Large Are the Gains from Economic Integration? Theory and Evidence from U.S. Agriculture, 1880-1997. *Econometrica*, Forthcoming.

Costinot, A., Donaldson, D., and Komunjer, I. (2012). What Goods Do Countries Trade? A Quantitative Exploration of Ricardo's Ideas. *Review of Economic Studies*, 79(2):581–608.

Dekle, R., Eaton, J., and Kortum, S. (2008). Global rebalancing with gravity: Measuring the burden of adjustment. *IMF Staff Papers*, 2008(005):A006.

Donaldson, D. (2018). Railroads of the Raj: Estimating the Impact of Transportation Infrastructure. *American Economic Review*, 108(4-5):899–934.

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

Durevall, D., Loening, J. L., and Ayalew Birru, Y. (2013). Inflation dynamics and food prices in Ethiopia. *Journal of Development Economics*, 104:89 – 106.

Eaton, J. and Kortum, S. (2002). Technology, Geography, and Trade. *Econometrica*, 70(5):1741–1779.

Ethiopian Road Authority, E. (2016). Road Sector Development Program: 19 Years Assessment. *Report*, page .

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

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., and Michaelsen, J. (2015). The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific Data*.

Gafarso, H. N. and Tufa, M. E. (2019). Effect of universal rural road access program (urrap) on rural households' sociocultural features: The case of jimma and bunno bedele zones, south western ethiopia. *Üniversitepark Bülten*.

Galle, S., Rodríguez-Clare, A., and Yi, M. (2021). Slicing the pie: Quantifying the aggregate and distributional effects of trade. *Review of Economic Studies*, Forthcoming.

Gebresilasse, M. (2018). Rural Roads, Agricultural Extension, and Productivity. *Job Market Paper*.

Ghani, E., Goswami, A. G., and Kerr, W. R. (2016). Highway to Success: The Impact of the Golden Quadrilateral Project for the Location and Performance of Indian Manufacturing. *The Economic Journal*, 126(591):317–357.

Hummels, D. (2007). Transportation Costs and International Trade in the Second Era of Globalization. *Journal of Economic Perspectives*, 21:131–154.

Nakamura, S., Bundervoet, T., and Nuru, M. (2020). Rural roads, poverty, and resilience: Evidence from ethiopia. *The Journal of Development Studies*, 56(10):1838–1855.

Nunn, N. and Qian, N. (2011). The Potato's Contribution to Population and Urbanization: Evidence from a Historical Experiment. *Quarterly Journal of Economics*, 126(2):593–650.

Shamdasani, Y. (2021). Rural road infrastructure agricultural production: Evidence from india. *Journal of Development Economics*, 152:102686.

Shrestha, S. A. (2020). Roads, participation in markets, and benefits to agricultural households: Evidence from the topography-based highway network in nepal. *Economic Development and Cultural Change*, 68(3):839–864.

Sotelo, S. (2020). Domestic Trade Frictions and Agriculture. *Journal of Political Economy*, 128(7):2690–2738.

Wondemu, K. A., Weiss, J., and Weiss, J. (2012). Rural roads and development: Evidence from ethiopia. *European Journal of Transport and Infrastructure Research*, 12(4).