The gains from market integration:
The welfare effects of new rural roads in Ethiopia*

Hundanol A. Kebede †

Job Market Paper

February 26, 2021

Abstract

Developing countries, particularly those in sub-Saharan Africa, often face a trade-off in investing in highways that interconnect cities vs. rural roads that connect rural villages to markets. Prioritization of either roads requires estimation of the welfare gain from each. In this paper, I estimate the welfare gain from rural road construction. To understand how rural roads affect welfare in agricultural villages, I develop a Ricardian trade model that features multiple crops, multiple villages, and heterogeneous land quality. The model gives a number of sharp predictions on how roads affect welfare, which I take to a very rich micro data on agricultural production and crop prices from Ethiopia. I use a large-scale rural road expansion program called Universal Rural Road Access Program (URRAP) as a source of variation to trade costs. To address endogeneity of road placement, I use counterfactual road network predicted from land gradient and location of rivers and lakes to construct instrumental variable. I estimate that the road expansion resulted in about 13% increase in real agricultural income, on average. I show that this increase is attributed to the mechanisms suggested in the Ricardian trade model: the prices of villages' comparative advantage crops increase and villages allocate more fraction of land towards these crops following decreases in trade costs. The size of the welfare gain varies across villages depending on, e.g., the crop composition of village consumption vis-a-vis production – cash-crop villages gain more than cereal producing villages because the latter face increases in their consumption costs while the former experience the exact opposite.

Keywords: Market Integration, Ricardian Trade Models, Rural Development, Rural Roads, Trade Costs. JEL Codes: F11, H54, O13, O18, Q12, R12, R42

*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. This paper has also benefited from comments received from participants in UVA's Trade and Development workshops, and WADES DC. conference.

†DECRG, WB. Email: hundanolkebede@virginia.edu or hkebede1@worldbank.org

1 Introduction

Roads play key role in economic development by facilitating movement of goods, people and ideas across locations (Ali et al., 2015). Unfortunately, developing countries often face a trade-off in investing in highways and railroads that interconnect cities vs. rural roads that connect rural villages to nearby towns. These two categories of roads differ in the sectors and demographies they favor. Highways and railroads primarily serve urban population and the manufacturing sector while the rural roads favor rural population (agrarians) and the agricultural sector. As such, this trade-off can be considered as the trade-off between urban-focused vs rural focused development policies. Because lack of resources compels prioritization of either highways or rural roads in most developing countries (particularly sub-Saharan Africa), it is essential to quantify the welfare gains from each type of roads to make informed policy choices.

Recent trade literature extensively documents 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 USA. However, the welfare effects of rural roads that connect rural villages to nearby markets is least understood. Few studies in the development literature document reduced-form evidences on the effects of rural roads on different outcomes of interest such migration and agricultural productivity (e.g., Asher and Novosad (2019), Shamdasani (2018), and Gebresilasse (2018)). These reduced-form evidences lack theoretical framework to pin down the detail mechanisms through which roads affect the outcomes of interest.

In this paper, I empirically analyze the welfare effects of construction of new rural roads. To understand the mechanisms through which the construction of new roads affects welfare in rural economies, I develop a multi-crop multi-location 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 its village in different crops. Villages also engage in costly trade among each other, very similar to how countries trade in workhorse trade models such as Eaton and Kortum (2002).

The model provides a number of sharp predictions on the mechanisms through which improvement in road infrastructure affects village welfare. First, decreases in trade costs lead to increases in the relative prices of the villages' comparative advantage crops (CA-crops). Second, decreases in trade costs lead to reallocation of farmland from a village's comparative disadvantage crops (CD-crops) towards the village's CA-crops. Third, the size of welfare gain 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 cash-crops should benefit more from decreases in trade costs than a village that has CA in cereals because the latter faces significant increase in the relative costs of its consumption basket following decreases in trade costs. I use the model structure to derive a sufficient statistic for the welfare effect of new roads to quantify how much of the total welfare gain from roads estimated from a reduced-form regression is explained by these mechanisms suggested in the model.

The predictions of the model can be tested easily 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 in 2010. Second, to alter this fact, the Ethiopian government launched a large-scale rural road expansion project called Universal Rural Road Access Program (URRAP), which planned to connect all rural villages by all-weather roads in just five years between 2011 and 2015.[1] The program led to 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 covers over 2,000 nationally representative rural villages (locally named as *Kebele*s, which are the lowest administrative units) and all crops. About of half of these villages got road connection under URRAP. 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 location of rivers and lakes). I use this counterfactual road network to construct an instrumental variable for the actual road network.

At the core of my empirical exercises is defining a village's comparative advantage crop(s). This requires information on yield estimate for each crop in each village. My agricultural survey

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

data includes village level crop yield estimated by trained enumerators.[2][3] Given the yield estimates, I define a village's comparative advantage crops using the following procedure. First, I calculate a village's yield relative to national average for each crop. Next, I rank crops, within each village, based on their yield relative to national average. I define crops in the top 20% of ranking based on relative yield as my baseline comparative advantage crops. I relax this baseline threshold to top 30%, top 40%, etc. to see sensitivity of my results.

There are three main results in this paper. First, using a reduced-form estimation, I show that the road expansion resulted in about 13% increase in welfare (real agricultural income per hectare), on average. Second, the road expansion led to increases in the relative prices of CA-crops by about 2% and increases in the fraction of land allocated to CA-crops by about 4%, for a village with average increase in market access, for a village with average increase in market access. Finally, using the sufficient statistic derived from the model, I show that the estimated welfare gain from the reduced-form is attributed to the mechanisms suggested in the trade model (reallocation of farmland towards CA-crops and higher prices for CA-crops). These results are robust to an alternative methods of addressing endogeneity of road placement, a matching-based DID strategy. The result is also robust to different set of crops considered: all crops, non-tree crops only, or crops for which FAO-GAEZ yield measure is available.

This paper's novel contribution is that it uses micro data on agricultural production and crop prices to provide an in-depth analysis of the underlying mechanisms through which roads affect village welfare, particularly how the new roads change the relative prices of crops and lead to reallocation of farmlands across these crops. Moreover, the paper studies the welfare effect of low-cost gravel roads connecting rural villages (that were previously inaccessible by modern means of transport system) to the nearest market centers. Previous studies on the welfare effect of roads focused on trunk roads, high-ways, and railroads which connect different regions of a country, play different roles in an economy, and are much more expensive to build.

An emerging literature studies the gains from intra-national market integration, particularly in the agricultural sector, using many-location many-good Ricardian trade models with heterogeneous factors of production (Costinot and Donaldson 2016, Donaldson 2018, Sotelo 2020, Allen and Atkin 2018, and Adamopoulos 2018). The most closely related papers to the current paper

[2]The enumerators use a method called *crop cut* where they take a sample of plots (each with area of 4 square meter) and conduct crop cut to obtain estimate of yield.

[3]As a robustness check, I use FAO-GAEZ data on agro-climatically attainable yield of crops in each village. This data uses a number of agro-ecological, soil and climatic factors, and sophisticated agronomic models to provide yield estimate at 5 arc-minute resolution.

are Adamopoulos (2018), Donaldson (2018), and Sotelo (2020). Adamopoulos (2018) finds 13.6% increase in aggregate agricultural yield following road expansion and upgrading that reduced trade costs between Ethiopian districts and location of national grain market centers. The main difference with the current paper is that the 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) develops a multi-sector multi-region Ricardian model in which land is treated as homogeneous within a region to study the gains from the railway expansion in colonial India. Sotelo (2020) introduces heterogeneous land quality to study how falling trade costs due to (counter-factual) paving of roads increases agricultural productivity and welfare in Peru. The current paper builds on these two papers for the theoretical part and contribute 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 broad literature in development economics on how rural roads improve livelihood of households in developing countries (Gebresilasse 2018, Shamdasani 2018, Shrestha 2018, and Asher and Novosad 2019). Asher and Novosad (2019) exploit 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 fuzzy regression discontinuity design. They find 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 relies on proxies, instead of direct measures, for agricultural outcomes due to lack of data at fine geographic level. The current paper uses large household-level agricultural survey and price surveys at detailed geography to construct real agricultural income and consumption. Shamdasani (2018) studies the effect of large road-building program in India and finds that remote farmers who got access to road diversified their crop portfolio by starting to produce non-cereal hybrids, adopted complementary inputs and improved technologies, and hired more labor. Gebresilasse (2018) studies how rural roads complement with agricultural extension program, a program that trains farmers on how to use best agricultural practices and technology adoption, to increase farm productivity in Ethiopia. Shrestha (2018) finds that a 1% decrease in distance to roads due to expansion of highways resulted 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 give some descriptive statistics that motivate the model presented in Section 3. Sections 4 takes the key predictions of the model to the data. Section 5 presents estimation of key model parameters and the welfare gain. Section 6 concludes the paper.

2 Data

2.1 Sources

Agricultural production data: I primarily use the Agricultural Sample Survey (AgSS), which is the largest annual agricultural survey in the country covering over 40,000 farm households in about 2200 villages. While this dataset goes back as far as 1995, villages were resampled every year until 2010 which makes tracking a village overtime difficult. Staring from 2010, Central Statistical Agency (CSA) kept the sample of villages fixed but took a random sample of about 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-year starting from the year 2009/2010, CSA also gathered crop utilization information, i.e., the fraction of crop production used for own 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 is an exceptionally detailed panel data of about 4,000 nationally representative farm households for the years2011, 2013 and 2015. The main advantage of the ESS dataset is it includes consumption information disaggregated by crops.[4] A big caveat of this data set is that it covers households in only about 330 villages.

Price data: The main price data is 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 produces.[5] This data covers over 500 representative villages which can be tracked over period since 2010. I also use the Retail Price Survey (RPS), which is a monthly

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

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

survey of prices of almost all crops and non-agricultural 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 agro-climatic data: I use FAO/GAEZ agro-climatically attainable yield for low/intermediate input use to construct villages' crop suitability, as a robustness check to yield measures in AgSS data. This data covers about 19 crops. However, it misses some of the endemic crops that are widely grown in Ethiopia such as Enset and Teff. As a result, I use this data only for robustness exercise. The rainfall data comes from Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS), which provides rainfall dataset starting from 1981. CHIRPS incorporates 0.05° resolution satellite imagery with station data to create gridded rainfall time series for trend analysis and seasonal drought monitoring. It is widely used to monitor drought in East Africa (Funk et al., 2015).

Road data: I use administrative data on the entire road-network in the country. This data includes 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 URRAP as a source of variation to villages' access to road/market. Over the period 2011-2015, the Ethiopian government gave exclusive focus to the URRAP and constructed over 62,413 kms of new all-weather roads connecting village centers to the nearest road or the nearest town, which ever is shorter. Figure 1 shows map of the road network before and after 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 1000 square-km from 44.4 in 2010 to 100.4 in 2015 (Ethiopian Road Authority, 2016). Though the URRAP was launched in 2011, very few roads were commenced in the year 2011, which is officially considered as capacity building year. Almost all the rural roads constructed under this program were started and completed between 2013-2015.

2.2 Identification of the effect of roads

There are two main issues that have to be dealt with to identify the causal effect of roads on village outcomes. The first is heterogeneity in treatment intensity and potential spillover effects.

That is, villages that get connected to a dense network may gain more from the road than those that get connected to sparse network. Moreover, road connection in a village may have spillover effects to 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 market via the connected village. As a result, non-connected villages may not serve as a control group in identification of the effects of road connection. 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 budget, officials might prioritize villages that would gain more from the road when deciding which villages to connect.[6]

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

I address the potential endogeneity of road placement using Instrumental Variable (IV) estimation strategy. To construct the IV, I first obtain road network predicted based solely on construction costs. The road construction cost is a function of land gradient and location of rivers and lakes. Next, I construct market access measure for each village based on the predicted road network – *predicted* market access measure. I use this *predicted* market access measure as an instrument for the *actual* market access measure. See Appendix B for the details on construction of the IV. 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.

Table 1 reports 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 to villages that are not directly connected by 42%. Finally, the last column reports the first stage regression result. Market access measure constructed from the *predicted* road

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

network is strongly correlated with the market access measure constructed based the *actual* road network. The estimate in the 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-stat in all specifications is above about 1000.

To check 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 non-treated villages based on their pre-URRAP observable characteristics that might be relevant for selection of villages for URRAP. I then conduct DID estimation based on these matched sample of treated and non-treated villages. Combining matching with DID strategy is a powerful approach to address the selection problem. The matching step enables me to compare treated villages with non-treated villages that have similar observed characteristics and hence similar treatment probability. The DID strategy on these matched samples helps me to washout unobserved time-invariant village characteristics that may confound the treatment effect. See Appendix E for details.

2.3 Evidences on improvement in market integration

In this subsection, I present some evidences on how the road construction under URRAP improved market integration among rural villages.

Farmers face considerable barrier to trade: ESS data includes direct questions about transport costs. I use this survey to obtain estimation of transport costs. The ad valorem trade cost (transport cost per value of transaction) on vehicle 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 import, although in my data the distance traveled is just 12 kilometers.

URRAP decreased trade costs: The main objective of URRAP roads was to integrate rural villages to market centers (Ethiopian Road Authority, 2016). If URRAP roads really integrated rural villages to local market centers, we would see the price gap between the rural villages and the market centers decreasing for villages that got road connection relative to villages that did not get roads. I test this by looking at the difference in crop prices between zone capitals and the villages within the zones using the 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 crop, v is village, z is zone capital, m is month, t is year, $Post$ equals zero for all month-years before URRAP and one for all month-years after URRAP; $URRAP_v$ is a dummy variable representing whether a village got URRAP road between 2012 and 2015; and γ^k_m is crop-month fixed effect which captures possible seasonality of crop prices.

The result is reported in Table 2. It shows that road connection significantly decreased the urban-rural price gap. The first column pools all 56 crops for which data is available on both urban and rural prices. It shows that trade cost, as proxied by the ratio of urban to rural prices, decreased by about 3% for villages that got road connection, relative to villages that did not get road connection. In column 2, the estimation is restricted to perishable products, vegetables and fruits. The estimated decrease in trade cost for these products is more than twice the estimate for all crops: trade cost for vegetables and fruits decreased by about 8%. This is not surprising because trading such products is difficult when there is no road passable by vehicle connecting a village to the urban center due to their perishability. In the last column, the sample is restricted to observations in which urban prices are higher than rural prices, which is what one would expect if villages are net exporters of crops to urban centers.[7] The gap between these two prices are plausibly capturing trade costs, which decreased by about 2.4%.

URRAP decreases the correlation between local prices and yields: One good indicator of an integrated market is that local prices are less sensitive to local supply. Under autarky, prices are relatively lower (higher) for the goods in which a region has a comparative advantage (disadvantage). Market integration weakens this inverse relationship between local prices and local comparative advantage. I run the following generalized difference-in-differences regression to investigate this:

$$\begin{aligned} lnP^k_{vt} = {} & \alpha_1 lnA^k_v + \alpha_2(Post_t * URRAP_v) + \alpha_3(lnA^k_v * Post_t * URRAP_v) \\ & + \gamma_v + \gamma_k + \gamma_t + \varepsilon^k_{vt} \end{aligned}$$

[7]Note that about 80% of observations (67,147 out of 82,944) conform with this expectation.

where P^k_{vt} is price of crop k in village v, A^k_v is a village's productivity in crop k which is proxied by GAEZ potential yield for the crop.

The result is presented in table 3. Panel A uses binary treatment dummy. The first column shows that there is a negative relationship between local prices of a crop and local yield of the crops. Column 3 shows that this negative relationship is significantly weakened when a village gets road connection. The elasticity of village price to village yield is 2.7% for a village with no road connection and a road connection decreases this estimate to 1.8%.[8] Panel B of table 3 reports the corresponding estimation result using market access measure. The result clearly shows that in villages that see an increase in their market access, the negative correlation between crop price and yield becomes significantly weaker.

Informed by these evidences on the effect of roads on market integration, in the next section, I develop a multi-sector multi-village Ricardian trade model to analyze the mechanisms through which road constructions affect village welfare.

3 Theoretical framework

The model builds on Donaldson (2018) and Sotelo (2020). Consider an economy composed of V villages indexed by $v = 1, ..., V$, each village represented by a representative household. A village derives utility from consumption of K crops indexed by $k = 1, ..., K$ that can be potentially produced or purchased. Each crop k comes with finite varieties indexed by $j \in \Delta_k$, where Δ_k is the set of varieties in crop k.

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

$$U_v = \prod_k (q^k_v)^{\mu^k_n}, \quad \text{where}$$
$$q^k_v = \left(\int_{\Delta_k} c^k_v(j)^\sigma \right)^{\frac{1}{\sigma}}$$

where U_v is utility in village v, and $c^k_v(j)$ is the quantity of variety j of crop k consumed by the village, μ^k_r is the share of expenditure on crop k in zone n such that $\sum_k \mu^k_n = 1$, and σ is elasticity of substitution between varieties within a crop.

[8] Alternatively, a positive α_3 would imply that road connectivity increases the prices of crops in which a village has a comparative advantage.

Technology: Similar to Sotelo (2020) and Allen and Atkin (2018), 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.[9] 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 $\int_{\Omega_v} \omega d\omega = L_v$. Each of the plot is potentially different in how well it is suited to growing different crops, which I denote as $z_v^{k(j)}(\omega)$. 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:

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

where $y_v^{k(j)}(\omega)$ is quantity of variety j of crop k per unit of plot.

A representative farmer in each village draws $z_v^{k(j)}(\omega)$ independently for each plot-crop variety from a Fréchet distribution with the following cumulative distribution function:

$$F_v^k(z) = Pr(Z_v^k < z) = \exp(-(A_v^k)^\theta z^{-\theta})$$

where A_v^k is the location parameter for the distribution of crop-suitability of land across the set of plots in village v, Ω_v. 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 agro-climatic conditions that are 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 others such as cereals. I assume that $\tau_{vv}^k = 1, \forall k$, and impose the standard assumption of triangle inequality in 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 price in village v' of variety j of crop k originating from village v, $p_{vv}^{k(j)}$ is price in village v of variety j of crop k originating from the same village v.

[9] 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.

Distribution of prices: Let r_v is the rental rate of plots in village v, which is determined in equilibrium. The unit cost of production of variety j of crop k in village v is $c_v^{k(j)} = \frac{r_v}{Z_v^{k(j)}}$, which is stochastic because it is a function of stochastic productivity $Z_v^{k(j)}$. As a result, the price at which village v supplies variety j of crop k to village v', $p_{vv'}^{k(j)} = \frac{r_v}{Z_v^{k(j)}}\tau_{vv'}^k$, is stochastic.

Using the distribution of $Z_v^{k(j)}$, we obtain the following distribution of the prices of crop k that village v' is offered by another village v:

$$G_{vv'}^k(p) = 1 - \exp\big(- (A_v^k)^\theta (r_v \tau_{vv'}^k)^{-\theta} p^\theta \big)$$

Because villages shop across all other villages for lowest prices for each variety, the distribution of the price of crop k that is actually paid in village v' is the distribution of the lowest prices across all other villages and is given by:

$$\begin{aligned} G_{v'}^k(p) &= 1 - \prod_{v=1}^{V} (1 - G_{vv'}^k(p)) \\ &= 1 - \exp\big(- p^\theta \sum_{v=1}^{V} (A_v^k)^\theta (r_v \tau_{vv'}^k)^{-\theta} \big) \end{aligned} \tag{1}$$

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 to all other potential villages trading with v') is:

$$\begin{aligned} \pi_{vv'}^k &= \Pr\Big[p_{vv'}^{k(j)} \leq \min_n \{p_{nv'}^{k(j)}\}, \quad \text{for some} \quad j \in \Delta_k\Big] \\ &= \frac{(A_v^k)^\theta (r_v \tau_{vv'}^k)^{-\theta}}{\sum_v (A_v^k)^\theta (r_v \tau_{vv'}^k)^{-\theta}} \end{aligned}$$

which is increasing in the average productivity of village v in crop k, A_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

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

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

Following Donaldson (2018), we can obtain the expected value of the price distribution:

$$E[p_v^{k(j)}] \equiv p_v^k = \Gamma\Big(\sum_{v'=1}^{I} (A_{v'}^k)^\theta (r_{v'}\tau_{v'v}^k)^{-\theta}\Big)^{\frac{-1}{\theta}} \tag{2}$$

where Γ is a gamma function $\Gamma(1+1/\theta)$. I assume that farmers in village v make their production and consumption decisions based on this expected price. Given the CES preferences, the price index faced by village v is given by:

$$P_v = \prod_k \left(p_v^k\right)^{\mu_n^k} \tag{3}$$

where p_v^k is given by equation 2.

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 $p_v^{k(j)}$ and the suitability of the village land for various crops. Revenue maximization implies that each plot of land is allocated to a crop that yields the highest return:

$$R_v(\omega) = \max_{k(j)}\{p_v^{k(j)} z_v^{k(j)}(\omega)\} \tag{4}$$

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

$$\eta_v^{k(j)} = \frac{(p_v^{k(j)} A_v^k)^\theta}{(\Phi_v)^\theta}, \quad \text{where} \quad \Phi_v = \left(\sum_{l,i} (p_v^{l(i)} A_v^l)^\theta\right)^{\frac{1}{\theta}} \tag{5}$$

where $\eta_v^{k(j)}$ is the fraction of land in village v allocated to variety j of crop k. It increases with the price of the variety and the average productivity of the the village in the crop, relative to all other crops. Given equation 5, one can obtain the fraction of land allocated to each crop in a village by aggregating the fraction of land allocated to varieties within the crop:

$$\eta_v^k = \int_{j\in\Delta_k} \eta_v^{k(j)} dj = \frac{A_v^{k\theta}}{(\Phi_v)^\theta}\int_{j\in\Delta_k} (p_v^{k(j)})^\theta dj \propto \frac{(p_v^k A_v^k)^\theta}{\Phi_v^\theta} \tag{6}$$

where p_v^k in equation 2 is used to obtain the last equality.

The following proposition summarizes the key 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 more specialization of villages.*

Proof. The elasticity of land share of a crop in a village to the village's productivity is given as $\frac{d\ln\eta_v^{k(j)}}{d\ln A_v^k} = \theta(1-\eta_v^{k(j)})$. Differentiating with respect to trade costs, we obtain $\frac{d^2\ln\eta_v^{k(j)}}{d\tau_{vv'}^k d\ln A_v^k} = -\theta\frac{d\eta_v^{k(j)}}{dp_v^{k(j)}}\frac{dp_v^{k(j)}}{d\tau_{vv'}^k}$. The term $\frac{d\eta_v^{k(j)}}{dp_v^{k(j)}}$ is always positive (see equation 5). Consider two villages v and v' and suppose 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 $p_v^{k(j)} = \tau_{vv'}^k$ if k is CD-crop in village v (i.e., $p_v^k > p_{v'}^k = 1$) or $p_v^{k(j)} = \frac{1}{\tau_{vv'}^k}$ if k is CA-crop in village v (i.e., $p_v^{k(j)} < p_{v'}^{k(j)} = 1$). Thus the term $\frac{dp_v^{k(j)}}{d\tau_{vv'}^k}$ has a positive sign if crop k is a CD-crop in the village and a negative sign otherwise.[10] 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 cheaper 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.

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

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

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

Suppose crop k is the crop that maximizes revenue from plot ω so that optimal revenue from plot ω is given by $R_v(\omega) \equiv p_v^{k(j)} y_v^k(\omega) = p_v^{k(j)} z_v^{k(j)}(\omega)$. 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 term multiplied by a non-stochastic price $p_v^{k(j)}$. Moreover, given the assumption of competitive land rental market,

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

rental rate per plot is equal to revenue per plot. Thus, the conditional distribution of rental rate per plot is the same as the conditional distribution of revenue per plot (note from equation 4 that $r_v(\omega)|\omega \in \Omega_v^k = p_v^{k(j)} z_v^{k(j)}(\omega)$ which has a Fréchet distribution with parameter Φ_v).

A sufficient statistic for the welfare gain: Because land is the only factor of production in the model, average real rental rate per plot, which is also equal to 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 average revenue per plot, and P_v is the village price index. Given data on quantities of each crop produced in a village, village prices, and model parameters needed to construct village price index, one can construct village real revenue $\frac{R_v}{P_v}$ and compare its changes over time in villages that get new road connection against villages whose road status did not change in a difference-in-differences strategy. However, this DID estimate does not tell us how much of the change in welfare due to roads is explained by the mechanism suggested in this model. The simplicity of the current trade model allows me to derive a sufficient statistic for the welfare gain from road. This sufficient statistic is crucial to shed light on how much of the welfare gain from road obtained in the reduced-form estimation is attributed to the mechanism that our trade model captures.

Recall that the probability that village v is the cheapest supplier of any variety of crop k to village v' is equal to $\pi_{vv'}^k = \Gamma^{-1}(A_v^k)^\theta (r_v \tau_{vv'}^k)^{-\theta}(p_{v'}^k)^\theta$. Evaluating this expression at $v = v'$ and solving for p_v^k we obtain $p_v^k = r_v (A_v^k)^{-1} {\pi_{vv}^k}^{1/\theta}\Gamma^{1/\theta}$. Plugging this into the price index we obtain:

$$\ln\Lambda_v \equiv \ln\frac{r_v}{P_v} = \sum_k \frac{\mu_n^k}{\theta}\ln A_v^k - \sum_k \frac{\mu_n^k}{\theta}\ln\pi_{vv}^k + \text{constant} \tag{7}$$

where Λ_v denotes welfare. This expression tells us that welfare is given by the sum of two terms (up to a constant factor). The first term $\sum_k \frac{\mu_n^k}{\theta} A_v^k$ is composed of the the preference parameters μ_n^k, and the productivity parameters θ and A_v^k. The second term is composed of the the preference parameters μ_n^k, the productivity parameter θ, and *home share of expenditure* π_{vv}^k.

In the model, the productivity parameter (the location parameter of Frechet distribution) A_v^k is time invariant. Thus, if the change in trade costs have any effect on real income, this effect should be mediated through change in the *aggregate* home share of expenditure in a village $\sum_k \frac{\mu_n^k}{\theta}\pi_{vv}^k$. That is, $\sum_k \frac{\mu_n^k}{\theta}\pi_{vv}^k$ is a sufficient statistics for the welfare effects of roads. And the effect of change in trade cost on real income exactly equals the effect of trade cost on the aggregate home share of expenditure in a village (with the opposite sign).

However, village's productivity is likely to change over time following decreases in trade costs. For instance, decreases in trade costs improve village's access to agricultural inputs such as chemical fertilizers or the improvement in market access following lower trade costs might increase profitability of adopting improved inputs. Thus, in my empirical exercises, I allow the productivity distribution to shift over time through change in the location parameter A_v^k. Under this relaxed assumption $\sum_k \frac{\mu_n^k}{\theta} \pi_{vv}^k$ is no longer a sufficient statistics because decreases in trade costs could also increase welfare by shifting the distribution of the village productivity, i.e., change in A_v^k. Importantly, equation 7 has feature that allows us to decompose the total welfare gain into these two channels.

The key challenge is that I do not observe trade flows between villages and the home share of expenditure π_{vv}^k. However, given values of the parameters μ_n^k and θ, and data on A_v^k and real income, we can infer the aggregate home share of expenditure in a village $\sum_k \frac{\mu_n^k}{\theta} \pi_{vv}^k$. And that is all we need to quantify how much of the welfare gain from decreases in trade costs is explained by our trade mechanism. That is, we can decompose the *total effect* on real income of decrease in trade costs into two: (i) the effect that is attributed to the trade mechanism, the *trade channel* and (ii) the effect that is attributed to change in yield (say due to adoption of modern agricultural inputs), *yield channel.* Given the aggregate home share at a village level, we can estimate the effect of road connection on this measure to obtain the trade channel. The difference between the total effect on real income and the estimated trade channel gives us the *yield channel.*

Proposition 2. *The size of the welfare gain from road 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:

$$d\ln W_v = \sum_k (\eta_v^k - \mu_n^k) d\ln p_v^k \tag{8}$$

This equation implies that a decrease in trade costs of CA-crop increases village welfare by a larger magnitude if $\eta_v^k >> \mu_n^k$. A testable implication of this proposition is that a village that specializes in cereals gains less from road connection compared to a village that specializes in cash-crops. 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 the villages that specialize in non-cereals experience the exact opposite.

4 Estimation of parameters and the welfare gain

4.1 Estimation of model parameters

In order to construct the empirical measure of sufficient statistic for welfare gain from roads, 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.

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 habit in the country.

Estimation of θ: I follow Sotelo (2020) in the estimation of θ. I rely on AgSS village level crop yield estimate that is constructed based on a random sample of crop cut. To purge out the noise in yield estimate and fluctuations due to whether conditions, I take the average across all years (2010-2016) to obtain a time invariant measure of yield for a crop in a village. I assume that the true crop productivity in a village A_v^k is related 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 a random noise, and δ^k is crop-specific constant. Plugging this for A_v^k in equation 5 and taking logs gives:

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

The empirical counterpart of this is:

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

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 around 1.7, but smaller than that of Donaldson (2018) who reports a mean of about 7.5 across the 17 crops in his data. However, Donaldson (2018) estimates θ from a gravity equation using data on trade flows between regions in colonial India.

4.2 Estimation of the welfare gain

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 $R_{vt} = \sum_k p^k_{vt} y^k_{vt}$. However, my price data (AgPPS data) does not cover all the villages for which agricultural production data is available (AgSS villages). I impute missing prices as follows. For each village in AgSS data that got road connection under URRAP, I find the nearest village in AgPPS data that also got road under URRAP. Similarly, for each village in AgSS data that did not get road under URRAP, I find the nearest village in the AgPPS data that did not get road under URRAP. I use these imputed prices to calculate village revenue. The price index is also calculated using the same data and the expenditure share of crops estimated in the previous subsection: $P_{vt} = \prod_k \left(p^k_{vt}\right)^{\mu^k_n}$.

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

$$\ln W_{vt} = \alpha_1 (Post_t * URRAP_v) + \alpha_2 Post_t + \delta \mathbf{X} + \gamma_v + \varepsilon_{vt} \tag{9}$$

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

$$\ln W_{vt} = \gamma_1 \ln \mathrm{MA}_{vt} + \delta \mathbf{X} + \gamma_v + \gamma_t + \varepsilon_{vt} \tag{10}$$

where the IV estimation uses market access measure constructed using the counterfactual road network predicted from land gradient and location of rivers and lakes.

Table 7 reports the estimation results. Column 1 reports results for binary treatment approach while column 2 and column 3 report OLS and IV results using market access approach. Column 1 shows that real agricultural income per a hectare of land (a measure of welfare) increased by 20% for treated villages compared to non-treated 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 elaticities imply 5% and 13% increase in real agricultural income, respectively for the OLS and IV estimation.[11]

In table A.8, I explore robustness of the above results to alternative way of addressing the endogeneity of roads – matching-based DID strategy. Column 1 reports the results for binary treatment while column 2 reports the results for market access approach. These estimates are very close to their counterparts in 7, implying that our results are robust to alternative ways of addressing potential endogeneity of road placement.

Overall, the results in tables 7 and A.8 show that the URRAP roads have led to significant increases in village welfare. In the next subsections, I explore the mechanisms through which the road connections increased real agricultural income.

4.3 Mechanisms

The theoretical results in section 3 suggest that decreases in trade costs would lead to increases in the relative prices of a village's CA-crops and reallocation of farmland towards these crops. I test this directly. In order to identify a village's CA-crop(s), I primarily use village level crop yield estimates provided in AgSS. This yield estimate is conducted by trained enumerators using a method called *crop cut* where they take sample plot of area of 4 square meters and conduct crop cut to obtain yield estimate. 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. As 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 agro-climatically attainable yield of crops in each village. This data uses a number of agro-ecological, soil and climatic factors, and sophisticated agronomic models to provide yield estimate 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 as that is more likely to reflect the reality in Ethiopia. While the FAO-GAEZ data overcomes the above two problems with the yield estimates provided in AgSS data, it covers only a partial list of crops produced in Ethiopia. In particular, it misses

[11]This value is obtained by multiplying the estimated elasticity by average increase in market access: 0.114*0.45 and 0.295*0.45.

some of 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 about 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 comparative advantage crops using the following procedure. First, I calculate a village's yield relative to national average for each crop, $\frac{A_v^k}{A^k}$. Next, I rank crops, within each village, based on their yield relative to national average. I define crops in the top 20% of ranking as my baseline comparative advantage crops. I relax this baseline threshold to top 30%, top 40%, etc. to see sensitivity of my results. One issue is 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.[12] For instance, if a village grows only 5 of, say, 25 crops in the data and we define CA crops as top 20% in relative yield, all the 5 crops grown in the village would be classified as CA crops. As a result, we would 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 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.

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 CA-crops increase (or the relative prices of the village's CD-crops decrease). This prediction of the model 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:

$$\begin{aligned} \ln p_{vmt}^k = {} & \alpha_1(Post_t * URRAP_v) + \alpha_2 CA_v^k + \alpha_3(Post_t * URRAP_v * CA_v^k) \\ & + \gamma^k + \gamma_v + \gamma_{mt} + \varepsilon_{vt}^k \end{aligned}$$

where $\ln p_{vmt}^k$ denotes the log price of crop k in village v in month m of year t. CA_v^k is dummy variable indicating whether crop k is among the village's CA-crops. I include crop-month fixed

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

effects to account for seasonal fluctuation in crop prices. I also run similar regression using log market access instead of binary treatment dummy:

$$\begin{aligned}\ln p_{vmt}^{k} &= \alpha_1 \ln\text{MA}_{vt} + \alpha_2 CA_v^k + \alpha_3(\ln\text{MA}_{vt} * CA_v^k) + \beta\mathbf{X} \\ &\quad + \gamma^k + \gamma_v + \gamma_{mt} + \varepsilon_{vt}^k\end{aligned}$$

I use the AgPPS data for the same 25 major non-tree crops in the main analysis. This data covers about 450 villages (after cleaning for missing information). The results are reported in table 5. While the estimates in the Panel A based on binary treatment are statistically insignificant, those in Panel B and Panel C are statistically and economically significant. Focusing on the results in Panel C, 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 1 log unit increase in market access, implying 1.62% increase in prices of CA-crops (relative to CD-crops) for a village with average (45%) increase in market access.[13]. However, this estimate significantly drops as we relax the cut-off for the definition of CA-crops, eventually becoming statistically insignificant in columns 3 and 4 of Panel C.

I explore a number of robustness exercises. First, I include all the 45 crops for which price data is available. The results are reported in table A.3. Clearly the results are very similar to those in table 5. Second, I use GAEZ data to define CA-crops. The result is reported in table A.4. The estimated increases in prices of CA-crops is significantly larger in table A.4 compared to tables A.3 and 5. This is partly driven by difference in the composition of crops across the specifications. Finally, I use the matching-based DID estimation to address endogeneity of road placement. Results are reported in table A.7 and they are very similar to those in the main specification, table 5.

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

Next, I explore the effect of increases in market access on village price index, which is computed using equation 3 and the elasticity of substitution estimated in section 4. The results in table 6 show that the village price index did not change much. This is not surprising given that village price index is a CES aggregate of prices of CA-crops and CD-crops, which move in the opposite direction when market access improves.

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

4.3.2 New roads and reallocation of farmland

I estimate the following regression at a village level:

$$\begin{aligned}\eta_{vt}^{k} &= \alpha_1(Post_t * URRAP_v) + \alpha_2 CA_v^k + \alpha_3(Post_t * Road_v * CA_v^k) + \beta\mathbf{X} \\ &+ \gamma^k + \gamma_v + \gamma_t + \varepsilon_{vt}^k\end{aligned}$$

where η_{vt}^{k} denotes the share of land allocated to crop k in village v in year t. $Post$ equals zero for the period before URRAP and one for the period after. $URRAP$ equals one for villages that got road under URRP and zero for other villages. CA_v^k is dummy variable indicating whether crop k is among the village's CA-crops, and X is a vector of village characteristics such as the population density and rainfall. I also include crop fixed effects to account for mean variation in land intensity across crops. The regression includes village fixed effect to account for time invariant village characteristics that may confound our result and year fixed effect to account for any year specific factor shared across villages.

To address the problems associated with binary treatment measure, I also estimate similar regression by using log market access derived from general equilibrium trade model.

$$\eta_{vt}^{k} = \alpha_1 \text{lnMA}_{vt} + \alpha_2 CA_v^k + \alpha_3(\text{lnMA}_{vt} * CA_v^k) + \beta\mathbf{X} \tag{11}$$

$$+ \gamma^k + \gamma_v + \gamma_t + \varepsilon_{vt}^k \tag{12}$$

where lnMA is log market access. I estimate using both OLS and IV strategies using predicted market access as an IV.

For the main analysis, I drop tree-crops and crops that are very rarely produced and have insignificant land share. The issue with tree-crops such as coffee, banana, orange, etc. is that farmers are unlikely to switch between these crops and others in the short and medium run due to their long gestation period. This brings down the number of crops in the analysis to 25. However, I show that my results are robust to including all crops.

Table 4 presents the results. Panel A uses binary treatment dummy. The first column shows that area of land allocated to the top 20% crops (in relative yield rankings) increased by about 3.8% (0.006/0.159), between 2012 and 2016, following road connection. As the definition of CA-crops is extended to top 30%, top 40% and top 50% in the next three columns, the estimated effect of road slightly decreases because as we relax the cutoff for definition of CA-crops we classify more and more crops as CA-crops even if the village is not strongly more productive in those crops relative to national average. Panel B shows similar results using market access measure. We see 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 CA-crop definition. The estimate in the first column shows that the fraction of land allocated to CA-crops increased by about 2.5% for a village with average increase in market access of 45%.[14] Columns 2-4 of Panel B clearly shows that the estimated increase in fraction of land reallocated to CA-crops decreases steadily as we relax the definition of CA-crops to top 30%, top 40%, etc. Panel C reports the IV result. Clearly the IV results are very similar to the OLS result in Panel B. In particular, the estimates for the interaction term $\ln\text{MA} * CA_v^k$ are identical to those in Panel B. This is because the regression is basically triple difference (Panel A makes this clear) and that identification comes from variation across crops (CA-crops vs CD-crops). Under such specification, endogeneity of MA is not a problem and OLS and IV results will be very similar. This is not the case in specifications that do not include triple interaction below. Overall, the results in table 4 show that the road expansion led to modest reallocation of farmland towards villages' CA-crops.

I explore robustness exercises similar to the previous subsection. First, I include all the 45 crops reported in AgSS data for which complete information is available. Table A.1 shows the estimation result. The implied increase in the share of land allocated CA-crops, based on estimates in Panel A, is 6.2%. The estimated increase in land share of CA-crops based on Panel B and Panel C are 2.8% and 3.2% respectively. Overall, the results in table A.1 are comparable to those those in table 4, implying that the results are not sensitive to the selection of crops.

In the second robustness exercise, I use FAO-GAEZ yield measure. This data includes only 19 crops. Some of these crops are rarely produced. Only 11 of the 19 crops are produced across many villages which significantly reduces the variation in the data and the precision of the estimates. Nevertheless the results in table A.2 are comparable to those in table 4 in magnitude.

In a third robustness exercise, I use a matching-based DID estimation strategy (see appendix E) to deal with potential bias in selection of villages for the URRAP program. That is, I first obtain a matched sample of treated and non-treated villages based on a set of observed village

[14] To obtain this percentage change, take partial derivative of equation 11 with respect to lnMA to obtain -0.004+0.013*CA-crop. For CA-crop=1, this equals 0.009. Multiply this by average change in market access of 0.45, and to obtain percentage change divide it by land share of CA-crops in 2012 given in the next to last row of table 4 and multiply by 100: $\frac{0.009*0.45}{0.159} * 100 = 2.5\%$.

characteristics that might affect selection of villages for the program before conducting DID estimation. The results are reported in table A.6. The estimates in both Panel A and Panel B are very similar to their counterparts in table 4. This implies that our results are robust to alternative methods used to account for potential selection bias.

4.4 Quantifying the welfare gain from trade

How much of the welfare gain estimated in the reduced-form estimation (table 7) is explained by the mechanism in our trade model, and how much is explained by increases in yield A_v^k? To answer this question, we rely on equation 7 which is derived from the theoretical model. Our objective is to infer the aggregate *home share of expenditure* in a village $\sum_k \frac{\mu_n^k}{\theta}\pi_{vvt}^k$, which enables us to quantify the effect of change in trade costs that is explained by the trade mechanisms in our model. Given the estimated values of parameter μ_n^k and θ, and data on A_{vt}^k and real agricultural income per hectare, we can infer from equation 7 the aggregate home share of expenditure:

$$-\sum_k \widehat{\frac{\mu_n^k}{\theta}\text{ln}\pi_{vvt}^k} = \widehat{\text{ln}\Lambda_{vt}} - \sum_k \widehat{\frac{\mu_n^k}{\theta}\text{ln}A_v^k} \tag{13}$$

Next, we estimate how much this measure changes in response to change market access:

$$-\sum_k \widehat{\frac{\mu_n^k}{\theta}\text{ln}\pi_{vvt}^k} = \alpha\text{lnMA}_{vt} + \delta\mathbf{X} + \gamma_v + \gamma_t + \varepsilon_{vt} \tag{14}$$

Note that in equation 7 the coefficient of $-\sum_k \widehat{\frac{\mu_n^k}{\theta}\text{ln}\pi_{vvt}^k}$ is one, implying that $\hat{\alpha}$ in equation 14 should be very close to the estimated effect of lnMA on real agricultural income in equation 10.

Table 8 reports the results for estimation of equation 14. Column 1 reports results using the binary treatment while columns 2 and 3 report OLS and IV results using market access approach. The results across all the three columns show statistically significant coefficients. Moreover, the sizes of the estimates in table 8 are very close to the estimated increase in real agricultural income in table 7. This implies that the increase in real agricultural income we estimated in table 7 is attributed to the mechanism that is suggested in our trade model, not by increases in yield or shift in productivity distribution.

5 Heterogeneity in the welfare gain across villages

5.1 Cereal vs. cash-crop 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 non-cereal producing villages experience the opposite. That is, while both the cereal and non-cereal (cash-crop) villages obtain better prices for their CA-crops following improvement in market access, they experience different outcomes on the cost of their consumption baskets. The 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. The cash-crop villages experience the opposite because cereals can now be imported at relatively cheaper prices and an increase in the prices of cash-crops has little effect on consumption costs because cash-crops have insignificant weight in consumption baskets.

Table 9 presents the empirical evidence supporting this result. I interact the fraction of village farmland allocated to cereals in the year 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 higher fraction of land to cereal gain significantly less than those that allocate lower fraction of land to cereals. The estimates in column 1 imply that villages with one standard deviation higher fraction of farm land allocated to cereals compared to the average gain 11% lower in welfare compared to 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 13% welfare gain for a village with average increase in market access and average cereal share of land. Villages with two standard deviation higher in the share of land allocated to cereals, compared to the average, experience 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 road connection, is distributed among villages of different characteristics. Would remote villages gain significantly more or less compared to villages that are located near roads or population centers? Theoretically, the answer is ambiguous. On one hand, non-remote 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 might not be large enough for remote villages to engage in trade with distant markets but significant enough for villages near markets to engage in trade. In this subsection, I use villages' distance from nearest town (population center with more than 20,000 population), distance to major trunk roads which are considered as main trade routes (trunk roads include inter-state roads, roads connecting the Zone capitals to Addis Ababa and roads that connect to neighboring countries), and distance to pre-URRAP road networks as measure of remoteness.

Table 10 reports the estimation result. 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 results using the binary treatment indicator, Panel B reports OLS estimates based on market access approach, and Panel C reports the corresponding IV results. Across all the three panels, remoteness is associated with lower welfare gain from road expansion. Focusing on the IV results in Panel C, column 1 implies that villages that are one standard deviation further from population center (compared to a village with average distance) gain about 15% less in welfare. Columns 2 implies that villages that are one standard deviation further from trunk roads (compared to a village with average distance) gain 11% less in welfare. Columns 3 implies that villages that are one standard deviation further from pre-URRAP roads (compared to a village with average distance) gain less than half of the welfare gain by a village with average distance. To sum up, remote villages gain much less in welfare compared to non-remote villages suggesting that authorities may also have economic reason to prioritize non-remote villages in road infrastructure, in addition to the obvious technical reason of connecting the near villages to facilitate access to the remote ones.

6 Conclusions

In this paper, I estimate the welfare gain from a massive rural road expansion in Ethiopia. I develop a Ricardian trade model with multi-crop multi-location feature where land productivity is allowed to vary both within and across villages to study the key mechanisms through which roads affect village welfare. On the demand side, a representative farmer in each village maximizes utility by choosing optimal quantities of crops to consume, given prices. On the production side, the representative farmer decides how to allocate its limited farmland across potential crops, given local prices and local productivity of crops. 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 the size of 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 towards village CA-crops and enables farmers receive better prices for these crops, it also leads to increase in the cost of consumption baskets for the villagers. The size of net welfare gain depends on the strength of these two contrasting forces, which in turn depends on the composition of the villages' CA-crops vis-a-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 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 15% from the road expansion between 2012-2016 for a village with 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 by 1.62% for a village with average increase in market access and (ii) reallocation of farmland towards a village's CA-crops – the fraction of land allocated to CA-crops increased by about 4% for a village with average increase in market access.

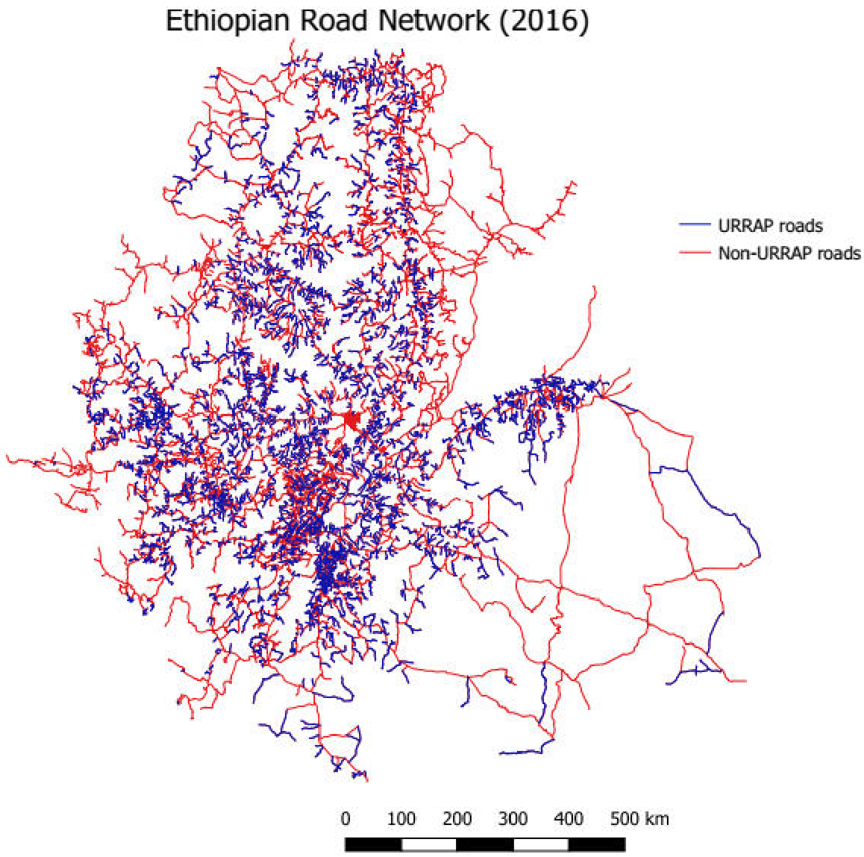


Figure 1: Rural road expansion under URRAP

Figure 2: Completed URRAP roads (pictures are taken from 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	4332	4332	4332
R^2	0.937	0.954	0.957

Notes: Robust standard errors in parenthesis. 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: log(Price in Zone Capital/Price in village)		
	All crops	Vegetables and Fruits	Cases where dep. var >0
Post*URRAP	-0.031**	-0.079*	-0.024*
	(0.016)	(0.044)	(0.013)
N	82944	24468	67147
R^2	0.378	0.360	0.493

Notes: Standard errors are clustered at village level. This table is based on AgPPS and RPS datasets. 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: Rural roads and the link between local prices and local yield: the dependent variable is crop-village level prices in 2012 and 2015.

	(1)	(2)	(3)
	Panel A: Binary Treatment		
LogYield	-0.036***		-0.027***
	(0.003)		(0.003)
Post*URRAP		0.017	-0.083***
		(0.023)	(0.023)
LogYield*Post*URRAP			0.009***
			(0.003)
N	59270	59270	59270
R^2	0.752	0.739	0.776
	Panel B: Market access approach		
LogYield	-0.036***		-0.099***
	(0.003)		(0.026)
LogMarketAccess		-0.026**	-0.043**
		(0.012)	(0.020)
LogYield*LogMarketAccess			0.006**
			(0.003)
N	59270	59270	59270
R^2	0.790	0.780	0.795

Notes: Standard errors are clustered at village level. The regression includes 277 villages, and 20 crops. All regressions include crop and year fixed effects, and log rainfall as a control. The last column includes village fixed effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Road construction and reallocation of farmland towards CA-crops

	Definition of CA-crops:			
	Top 20%	Top 30%	Top 40%	Top 50%
	Panel A: Binary treatment			
CA-crop	0.042***	0.050***	0.056***	0.055***
	(0.003)	(0.003)	(0.002)	(0.002)
Post*URRAP	-0.003**	-0.004**	-0.004**	-0.005**
	(0.002)	(0.002)	(0.002)	(0.002)
CA-crop*Post*URRAPs	0.009**	0.008**	0.006*	0.007**
	(0.005)	(0.004)	(0.004)	(0.003)
N	34839	34839	34839	34839
adj. R^2	0.340	0.347	0.353	0.352
	Panel B: 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 C: 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 village level. These regressions include 25 major non-tree 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 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. Panel A uses binary treatment dummy. Panel B uses market access measure constructed from the entire road network before and after the program, and pre-program spatial population distribution. Panel C 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 5: New road construction and prices of CA-crops

	Definition of CA-crops:			
	Top 20%	Top 30%	Top 40%	Top 50%
	Panel A: Binary treatment			
CA-crop	-0.023**	-0.029***	-0.038***	-0.048***
	(0.011)	(0.011)	(0.011)	(0.013)
Post*URRAP	0.013	0.022	0.023	0.029
	(0.019)	(0.020)	(0.021)	(0.023)
Post*URRAP*CA-crop	0.017	-0.009	-0.009	-0.016
	(0.022)	(0.020)	(0.021)	(0.021)
N	31737	31737	31737	31737
adj. R^2	0.800	0.800	0.800	0.801
	Panel B: 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 C: 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 village level. The dependent variable is log village crop price. The analysis includes 25 major non-tree crops and over 450 nationally representative rural villages. 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 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. Panel A uses binary treatment dummy. Panel B uses market access measure constructed from the entire road network before and after the program, and pre-program spatial population distribution. Panel C 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 6: 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 in parenthesis. The regression includes 450 villages. The price index is computed from the 25 major non-tree 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 7: The welfare gains from new rural roads

	Binary treatment	Market access approach	
	OLS	OLS	IV
Post*URRAP	0.202***		
	(0.045)		
LogMarketAccess		0.114**	0.295***
		(0.053)	(0.096)
N	4148	4148	4148
R^2	0.850	0.849	0.002

Notes: Robust standard errors in parenthesis. All regressions include year and village fixed effects, and log rainfall. The estimation is based on 25 non-tree crops for which full information on prices and quantities of production is available. Column 1 uses binary treatment dummy. Columns 2 and 3 use market access measure constructed from the entire road network before and after the program, and pre-program spatial population distribution. In column 3, 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 8: The trade mechanism

	Binary treatment	Market access approach	
	OLS	OLS	IV
Post*URRAP	0.191***		
	(0.047)		
LogMarketAccess		0.108**	0.239**
		(0.054)	(0.102)
N	4148	4148	4148
R^2	0.885	0.884	0.014

Notes: Robust standard errors in parenthesis. The dependent variable is the measure of sufficient statistics $-\sum_k \widehat{\frac{\mu_n^k}{\theta}\ln\pi_{vvt}^k}$. All regressions include year and village fixed effects, and log rainfall. The estimation is based on 25 non-tree crops for which full information on prices and quantities of production is available. Column 1 uses binary treatment dummy. Columns 2 and 3 use market access measure constructed from the entire road network before and after the program, and pre-program spatial population distribution. In column 3, 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 9: Welfare gain: cereal vs non-cereal villages

	The dependent variable 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**		
	(0.045)		
LogMarketAccess		0.150*	0.535***
		(0.082)	(0.139)
LogMarketAccess*CerealShare		-0.008***	-0.008***
		(0.003)	(0.002)
N	4148	4148	4148
adj. R^2	0.770	0.771	0.18

Notes: Robust standard errors in parenthesis. The first column reports results based on binary treatment dummy. The second column reports OLS results based on market access approach. The third column reports 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 non-tree 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. Non-cereal crops include all vegetables, legumes and cash-crops which are predominantly produced for market. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 10: Remoteness and welfare gain from new roads

	The dependent variable is log real revenue per hectare ln$W$		
	Distance to town	Distance to trunk roads	Distance to pre-URRAP roads
	Panel A: Binary treatment		
Post*URRAP	0.091	0.113*	0.113*
	(0.062)	(0.062)	(0.062)
Post*URRAP*Distance	-0.145***	-0.061	-0.026
	(0.056)	(0.050)	(0.049)
N	4148	4148	4148
adj. R^2	0.747	0.746	0.746
	Panel B: Market access approach – OLS		
LogMarketAccess	0.090	0.217***	0.196**
	(0.080)	(0.080)	(0.080)
LogMarketAccess * Distance	-0.217***	-0.198***	-0.125**
	(0.082)	(0.061)	(0.057)
N	4148	4148	4148
adj. R^2	0.747	0.747	0.747
	Panel C: Market access approach – IV		
LogMarketAccess	0.127	0.409***	0.379***
	(0.135)	(0.142)	(0.145)
LogMarketAccess * Distance	-0.345***	-0.241***	-0.158**
	(0.103)	(0.076)	(0.068)
N	4148	4148	4148

Notes: Robust standard errors in parenthesis. 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 above 20,000 population. The second column uses distance to the nearest trunk road (inter-state roads and roads connecting zone capitals to the center). The third column uses distance to pre-existing road network. All regressions include year and village fixed effects. The estimation is based on 25 non-tree 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 market access measure for each village derived from general equilibrium trade models. I use spatial distribution of population before the road program, and pre- and post-program 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 destination village population from the 2007 census (before the onset of the URRAP program). Using pre-URRAP population distribution is necessary because population distribution is likely to respond to improvement 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.[15] 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 is no similar cost estimates along the link roads, I scale up the costs along earth road by the factor of $\frac{Cost\ along\ earth\ road}{Cost\ along\ minor\ gravel}$ to obtain estimate of cost along the links.[16] After assigning each road type (including the links) with the estimated costs in USD per ton-kilometer, I use 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 can be seen in equation 15, a change to a village's market access comes only from changes in τ_{odt}, which in turn comes from construction of new roads.

[15] 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).

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

B Construction of the instrumental variable

I use two main information to construct the predicted road network. The first is 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.[17] 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 30meter-by-30meter grid cell. Given the cost raster and regional budget constraint, I use the following algorithim 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 other factors than just construction costs, and as a result tend to be longer than the predicted roads based only on construction costs.

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

C Derivation of the conditional distribution of productivity and rental rate

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

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

conditional distribution of land quality is similar to Sotelo (2020):

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

Using the distribution of $z_i^{k(j)}(\omega)$:

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

Thus, the distribution of productivity of the set plots in village v which are covered by variety j of crop k is a Fréchet with the parameters $\frac{\Phi_v}{p_v^{k(j)}}$ and θ. Notice that the average productivity of land covered with a crop decreases with the crop price. Intuitively, more and more land is allocated to a crop with 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 variety j of crop k, is given by $r(\omega) = \max_{k(j)}\{p_v^{k(j)} z_v^{k(j)}(\omega)\}$. Thus the conditional distribution of rental rate $r(\omega)|\omega \in \Omega_v^{k(j)}$ is Fréchet with parameters Φ_v and θ. That is, the rental rate of plots covered with different crops have 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: Road construction and reallocation of farmland towards CA-crops: all crops

	Definition of CA-crops:			
	Top 20%	Top 30%	Top 40%	Top 50%
	Panel A: Binary treatment			
CA-crop	0.021***	0.027***	0.031***	0.030***
	(0.002)	(0.002)	(0.001)	(0.001)
Road Access	-0.002**	-0.004***	-0.003**	-0.003**
	(0.001)	(0.001)	(0.001)	(0.001)
CA-crop*Road Access	0.008***	0.009***	0.005**	0.004**
	(0.003)	(0.002)	(0.002)	(0.002)
N	56211	56211	56211	56211
adj. R^2	0.359	0.363	0.366	0.365
	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 village level. The analysis includes 45 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 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. Panel A uses binary treatment dummy. Panel B uses market access measure constructed from the entire road network before and after the program, and pre-program spatial population distribution. Panel C 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.2: Road construction and reallocation of farmland towards CA-crops: based on GAEZ yield measure

	Definition of CA-crops:			
	Top 20%	Top 30%	Top 40%	Top 50%
	Panel A: Binary treatment			
CA-crop	0.062***	0.071***	0.076***	0.077***
	(0.005)	(0.004)	(0.004)	(0.004)
Post*URRAP	-0.002	-0.000	0.001	-0.002
	(0.003)	(0.004)	(0.004)	(0.004)
CA-crop*Post*URRAP	0.015*	0.009	0.006	0.008
	(0.008)	(0.007)	(0.007)	(0.006)
N	22466	22466	22466	22466
adj. R^2	0.322	0.330	0.335	0.336
	Panel B: Market access approach – OLS			
CA-crop	0.012	0.054	0.082	0.099**
	(0.057)	(0.052)	(0.050)	(0.048)
LogMarketAccess	-0.002	0.001	0.003	0.003
	(0.004)	(0.004)	(0.004)	(0.005)
CA-crop*LogMarketAccess	0.006	0.002	-0.000	-0.002
	(0.006)	(0.005)	(0.005)	(0.005)
N	22466	22466	22466	22466
adj. R^2	0.322	0.330	0.335	0.336
	Panel C: Market access approach – IV			
CA-crop	-0.025	0.033	0.068	0.087*
	(0.058)	(0.052)	(0.050)	(0.048)
LogMarketAccess	-0.008	-0.004	-0.002	-0.002
	(0.007)	(0.007)	(0.007)	(0.007)
CA-crop*LogMarketAccess	0.009	0.004	0.001	-0.001
	(0.006)	(0.005)	(0.005)	(0.005)
N	22466	22466	22466	22466
adj. R^2	0.300	0.308	0.313	0.314
Mean land share (CA-crops) in 2012	0.198	0.207	0.209	0.2015
Mean land share (all) in 2012	0.176	0.176	0.176	0.176

Notes: Standard errors are clustered at village level. The analysis includes 19 GAEZ 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 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. Panel A uses binary treatment dummy. Panel B uses market access measure constructed from the entire road network before and after the program, and pre-program spatial population distribution. Panel C 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: New road construction and prices of CA-crops: all crops

	Definition of CA-crops:			
	Top 20%	Top 30%	Top 40%	Top 50%
	Panel A: Binary treatment			
CA-crop	-0.040***	-0.041***	-0.046***	-0.038***
	(0.013)	(0.011)	(0.011)	(0.011)
Post* URRAP	0.028	0.025	0.027	0.044**
	(0.018)	(0.019)	(0.019)	(0.021)
Post*URRAP*CA-crop	0.041*	0.027	0.022	-0.009
	(0.021)	(0.018)	(0.017)	(0.018)
N	41930	41930	41930	41930
adj. R^2	0.798	0.785	0.799	0.799
	Panel B: 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 C: 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 village level. The dependent variable is log village crop price.The analysis includes 45 crops and over 450 nationally representative rural villages. 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 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. Panel A uses binary treatment dummy. Panel B uses market access measure constructed from the entire road network before and after the program, and pre-program spatial population distribution. Panel C 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.4: 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: Binary treatment			
CA-crop	-0.089***	-0.093***	-0.097***	-0.074***
	(0.018)	(0.018)	(0.018)	(0.017)
Post*URRAP	-0.006	-0.030	-0.043*	-0.046*
	(0.021)	(0.022)	(0.023)	(0.025)
CA-crop*Post*URRAP	0.090***	0.121***	0.128***	0.118***
	(0.022)	(0.022)	(0.025)	(0.026)
N	21428	21428	21428	21428
adj. R^2	0.761	0.762	0.762	0.761
	Panel B: 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 C: 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 village level. The dependent variable is log price of crop. This regression analysis includes 19 crops for which GAEZ yield measure is available. 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 national average. In column 2 CA-crops are top 30% in the ranking, and so forth. All regressions include village, year and crop-month fixed effects. Panel A uses binary treatment dummy. Panel B uses market access measure constructed from the entire road network before and after the program, and pre-program spatial population distribution. Panel C 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

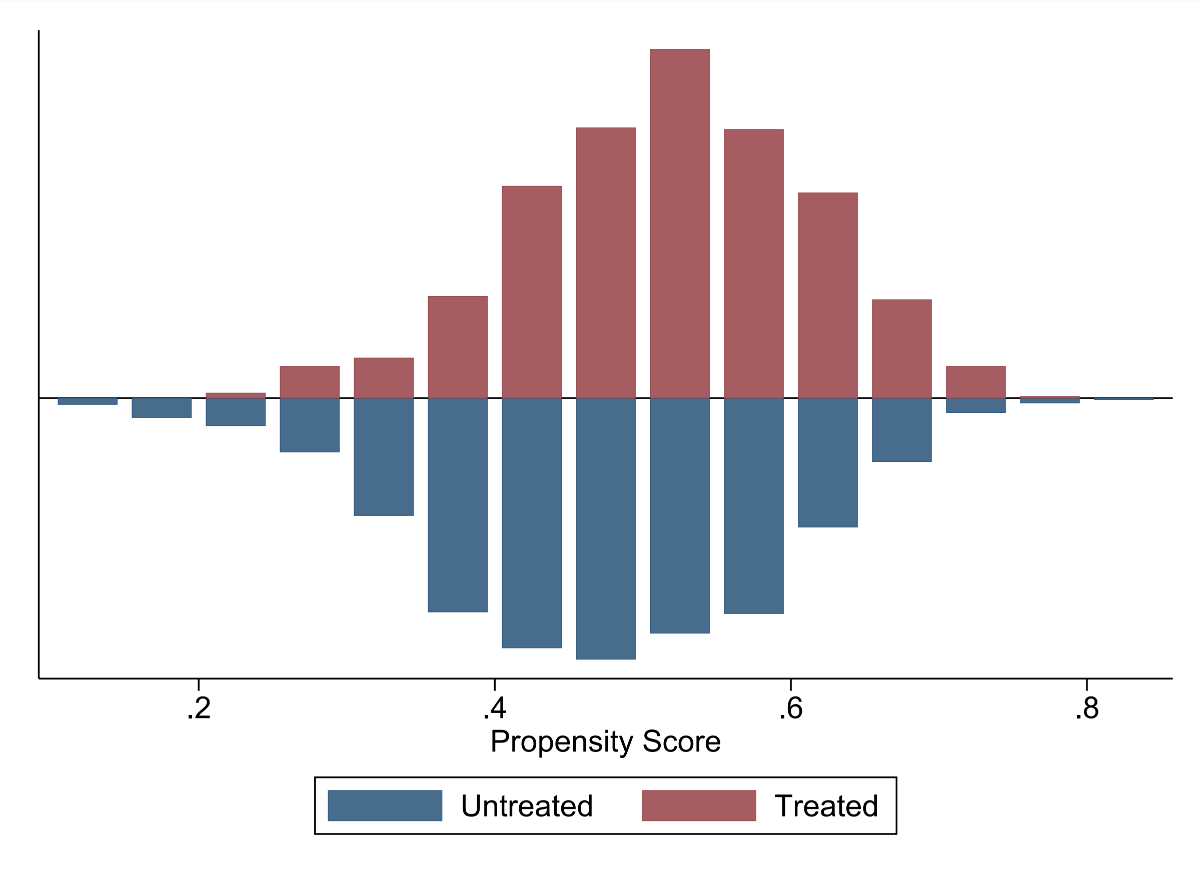
As a further robustness exercise, I also use a Matching-Based Difference-in-Differences (MB-DID) strategy to address endogeneity of road placement. That is, I first obtain a matched sample of treated and non-treated villages based on their observable characteristics that might be relevant for selection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

To identify relevant village characteristics for matching treated and non-treated villages, I use information from officials at Ethiopian Roads Authority (ERA). They 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 preexisting road network, (2) 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 consideration and the project's labor input requirements.[18] Finally, terrain and landscape significantly affects the 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 non-treated 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 1990-2010 period. I use Digital Elevation Model (DEM) data and ArcGIS tools to calculate average slope and elevation of each village.

Given these set of village characteristics,I use the *gmatch* command in STATA (for its handiness in panel data setting) to match treated and non-treated villages. For each treated village, the *gmatch* algorithm finds, non-treated village(s) that has/have the closest observed characteristics or propensity score. I conduct a DID estimation on these matched samples of treated and non-treated villages. Figure A.1 shows the histogram of propensity score by

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

Figure A.1: Common support of propensity score matching



treatment status. The figure clearly shows that the region of common support is large as very few non-treated 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: 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%
	Panel A: Binary treatment			
CA-crop	0.041***	0.049***	0.055***	0.056***
	(0.003)	(0.003)	(0.003)	(0.003)
Post*URRAP	-0.004**	-0.005**	-0.006***	-0.005**
	(0.002)	(0.002)	(0.002)	(0.002)
CA-crop*Post*URRAP	0.010**	0.010**	0.010**	0.007**
	(0.005)	(0.004)	(0.004)	(0.004)
N	28030	28030	28030	28030
adj. R^2	0.310	0.318	0.324	0.324
	Panel B: 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 village level. This regression analysis includes 25 major non-tree 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 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. Panel A uses binary treatment dummy. Panel B uses market access measure constructed from the entire road network before and after the program and pre-program spatial population distribution. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

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%
	Panel A: Binary treatment			
CA-crop	-0.028**	-0.033***	-0.040***	-0.047***
	(0.011)	(0.011)	(0.012)	(0.013)
Post*URRAP	0.009	0.019	0.022	0.032
	(0.020)	(0.021)	(0.023)	(0.025)
Post*URRAP*CA-crop	0.021	-0.008	-0.013	-0.026
	(0.024)	(0.021)	(0.022)	(0.023)
N	26786	26786	26786	26786
adj. R^2	0.797	0.797	0.797	0.797
	Panel B: 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 village level. The dependent variable is log village crop price. The analysis includes 25 major non-tree crops and over 450 nationally representative rural villages. 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 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. Panel A uses binary treatment dummy. Panel B uses market access measure constructed from the entire road network before and after the program and pre-program spatial population distribution. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.8: The 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 in parenthesis. All regressions include year and village fixed effects. The estimation is based on 25 non-tree crops for which full information on prices and quantities of production is available. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

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.

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*, .(.):.

Allen, T. and Atkin, D. (2018). Volatility and the gains from trade. *Econometrica*, Forthcoming.

Asher, S. and Novosad, P. (2019). Rural roads and local economic development. Dartmouth College.

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. (2016). How large are the gains from economic integration? theory and evidence from u.s. agriculture, 1880-1997. *Econometrica*, Forthcoming.

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. 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*.

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.

Shamdasani, Y. (2018). Rural road infrastructure & agricultural production: Evidence from india. page .

Shrestha, S. A. (2018). Roads, participation in markets, and benefits to agricultural households: Evidence from the topography-based highway network in nepal. page .

Sotelo, S. (2020). Domestic trade frictions and agriculture. *Journal of Political Economy*, 128(7):2690–2738.