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

Hundanol A. Kebede †

Job Market Paper

University of Virginia

January 6, 2020

Abstract

This paper estimates the welfare gains from market integration due to new rural road construction. Using a large-scale rural road expansion project and a detailed agricultural dataset from Ethiopia, I provide evidence that rural road construction increased market integration and household consumption. I then develop a multi-sector multi-location Ricardian model of trade with heterogeneous land quality to identify the mechanisms through which roads improve household welfare. I use the model structure to derive a sufficient statistic for the effects of roads on household welfare that can be inferred from household land allocation and model parameters alone, without data on trade flows between households. I estimate a reduced-form effect of the roads on household welfare of about 13%. Using the sufficient statistic derived from the model, I show that about half of the reduced-form welfare effect is attributed to new trading opportunities. A notable feature of my model is the prediction that increased road connectivity would cause household production decisions to be dictated by market prices rather than consumption preferences – thus providing a novel test for *separability*. This separability between household production decisions and consumption preferences leads to efficiency gains because land is reallocated to crops that have higher productivity and higher prices. I provide robust evidence that support these predictions.First, I find a significant correlation between household land and budget allocation across crops, and that this correlation decreases over time for households that reside in villages that get road connectivity, relative to households in villages that did not get roads. Second, I show that land is reallocated to crops in which villages have a comparative advantage when they gain access to new roads.

Keywords: Ricardian Trade Models, Rural Development, Rural Roads, Separability, 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.

†Department of Economics, University of Virginia. Email: hundanolkebede@virginia.edu

1 Introduction

A common feature of rural areas in developing countries is that they lack all-weather roads interconnecting the villages or connecting the villages to market centers in towns and cities. This makes the villages isolated economies operating in a near-autarky environment. As a result, product markets are too thin for farmers to rely on, and farmers adopt subsistence farming where they self-produce most of the crops they need for consumption, instead of specializing in few crops and buying the rest from market.[1] This implies significant efficiency loss in land utilization because land would be used for less suited crops. In this paper I study how new road construction leads to better market integration and opportunities to trade, and consequently result in better allocation of land and higher welfare.[2]

In particular, I emphasize a key effect of roads in near-autarky environment: increased road connectivity would cause household production decisions to be dictated by market prices rather than consumption preferences – a phenomenon known as *separability* in development economics. When trade costs are high and opportunities to trade are limited, household resource allocation is primarily dictated by their consumption needs rather than market forces – land allocation across various crops would be determined by household consumption preferences for these crops rather than crop prices and crop-suitability of the land. A decrease in trade cost would break this link between consumption preference and resource allocation by improving households' opportunities to trade. This leads to reallocation of land according to comparative advantage and gains from trade.

I use a massive rural road construction project in Ethiopia under the Universal Rural Road Access Program (URRAP) that took place between 2011 and 2015, a rich household level data on agricultural production and consumption, and a novel price data at different levels of market chain to study the effect of road on household welfare. I start by providing a series of reduced-form evidences on the effects of roads on market integration, village specialization, and household consumption using a difference-in-differences strategy where villages that were not covered by URRAP are used as a control group. First, I use the gap between crop prices in the urban centers (Zone capitals) and farm gate prices in the villages within the zone as a proxy for

[1] A manifestation of this is that a farm household in a developing country usually grows several crops, even though the average land holding of a household is less than two hectares. In Ethiopia for instance, the median land holding is about a hectare and a median farmers grows 5 crops.

[2] In this paper, I refer to *subsistence farming* as a farming system in which household crop production is primarily for own consumption and a *market based farming* as one in which households produce crops for market and rely on market to buy their consumption needs.

trade cost and show that the URRAP roads decreased trade costs for the connected villages by about 3%, compared to villages that were not covered by the program.[3] Moreover, I show that the trade cost decreased more for perishable crops (vegetables, 8%) than non-perishable crops (cereals, 2%). Second, I use village level data on prices and FAO-GAEZ data on agro-climatically attainable yield of crops to show that road connectivity weakens inverse relationship between local prices and local productivity. I show that in villages that got new road connection, the elasticity of local prices to local productivity decreased by one-third. These two evidences suggest that the roads indeed improved market integration. Third, I use FAO-GAEZ data on agro-climatically attainable yield to identify each village's comparative advantage crops and show that villages that get new road reallocated more land to crops in which they have comparative advantage. The fraction of land allocated to crops in the highest quartiles of relative yield increased by about 12% in villages that got new roads, compared to villages that did not get roads. Fourth, I use panel data on households to show that the URRAP roads increased household real consumption expenditure by about 13%.

Next, I develop a multi-sector multi-location Ricardian model of trade with land heterogeneity to understand the mechanisms through which the roads affect household welfare and to quantify the gains from the road. In the model, farmers engage in costly trade in crops with each other in a Ricardian world, very similar to how countries trade in Eaton and Kortum (2002). Moreover, farmers make decision on how to allocate their available land across potential crops given the crop prices, local productivity of land in the crops, and the local tastes for these crops. Two crucial results emerge from the model.

First, the model gives a sharp prediction that households' resource allocation rules are quite different under autarky (or near-autarky) and under lower trade cost environments. Under autarky, household land allocation across different crops is dictated by the household's *taste* for the crops and the productivity of land in the crops whereas under separability land allocation is dictated by crop *prices* and productivity of land in the crops. I use this result to design a novel test of separability, and how it is related to trade costs. The key insight is that if separability holds, there should not be any correlation between the budget share of a crop and the land share of a crop conditional on the crop price. In contrast, under autarky the land and budget share

[3] Ethiopia follows a federal system where the the highest level of subnational administration is called *Regions*. Regions are subdivided into *Zones*, which are subdivided into *Woredas* (sometimes referred to as districts), which are further subdivided into *Kebeles*. *Kebele* is the lowest administrative unit, and rural kebeles are usually referred to as *villages or peasant associations*. Currently there are 9 regions, 70 zones, about 600 woredas and about 16,000 kebeles.

of crops will be correlated because they are both driven by crop tastes and crop-specific land productivity. Using this test I strongly reject both full separability and full autarky. I estimate a correlation of about 0.45 between the land and budget shares of crops for the year 2011 before the onset of URRAP project, and that this estimate decreases to 0.20 in 2015. Next, I exploit a large-scale rural road construction project as a source of variation to show that in villages that got new roads, the correlation between land and budget shares of crops decreased by about one-fourth of the baseline correlation, compared to villages that did not get URRAP roads. This suggests that URRAP significantly improved separability of resource allocation from household consumption preferences. Moreover, I show that the link between land allocation and budget share of a crop is strongly related to households' proximity to market: households living further away from market centers or all-weather roads have significantly higher correlation between land and budget share of crops.

Second, I use the model structure to derive a sufficient statistic for the welfare effect of roads which resembles the sufficient statistic for welfare gain from trade in workhorse trade models, e.g., Arkolakis et al. (2012), with a straightforward modification to account for geographic variation in both taste for crops and productivity of land in these crops. Interestingly, however, no trade flow data is required to construct the sufficient statistic – it can be inferred from households' land allocations across crops and the model parameters. I use the estimated model parameters together with household level data on land allocation to construct the sufficient statistic. I use this sufficient statistic to shed some light on to what extent the mechanisms in the model explain the reduced-form effect of roads on household welfare. If the road affected household real revenue (which is the measure of welfare in the model) through the mechanisms suggested in the model (i.e., through improving trading opportunities), then the reduced-form effect of road on real revenue should disappear (or at least significantly decrease) once the sufficient statistic is included as a regressor. The result lends a strong support to this hypothesis. I find a reduced-form effect of roads on real revenue of about 13%. However, when the sufficient statistic is included on the right-hand side, the coefficient of road decreases to 5.8%, suggesting that about half of reduced form effect of roads on real revenue can be attributed to improvement in trade opportunities.[4]

This paper's novel contribution is that it studies the welfare effect of low-cost gravel roads

[4]The other half is attributed to non-trade mechanisms such as improved access to agricultural inputs and advisory services, improved access to health facilities, time savings from trips to market centers, etc.

connecting villages that were previously inaccessible by any modern means of transport system using detailed household level production, consumption, and price data. Previous studies on the welfare effect of roads focused on paved roads, high-ways, and rail roads, which are much more expensive to build. Moreover, these studies were based on regionally aggregated production and trade data or proxy measures for welfare, compared to household-level production and consumption data utilized in the current paper.

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. Among the influential papers are Donaldson (2018), Sotelo (2018), Costinot and Donaldson (2016), Allen and Atkin (2018), and Adamopoulos (2018). 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 (2018) develops a model of heterogeneous land quality to study how falling trade costs due to (counter-factual) paving of roads increases agricultural productivity and welfare in Peru. Costinot and Donaldson (2016) find substantial long run gains from economic integration among US agricultural markets between 1880-1997. Allen and Atkin (2018), show that falling trade costs increased farmer's revenue volatility thereby causing farmers to shift production towards crops with less risky yields. These studies implicitly impose separability, thereby ruling out the limiting case of subsistence farming which emerges in a very high trade cost environments. I contribute to this literature by providing evidence on how falling trade costs in a near autarky environment improve resource allocation by weakening the link between household's production decisions and their consumption preferences, and hence facilitating a move away from subsistence farming.

The most closely related paper to the current one is Adamopoulos (2018) which 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 with district capitals, instead of a decrease in trade cost between district capitals and Addis Ababa or other major urban centers.

This paper brings together the development economics literature on separability with the international trade literature on gains from reduced trade costs and market integration. I suggest a novel and clean approach to test separability using crop production and consumption. Earlier

tests of separability are exclusively based on labor demand – they test whether a household's choice of farm labor is independent of the household's demographic characteristics (Benjamin, 1992; LaFave and Thomas, 2016). A major drawback of this approach is that hours of work on farm is arguably poorly measured in the context of self-employed agricultural households. My approach is *less* prone to such measurement problem: area of land allocated to a crop is measured by GPS and consumption of crops are measured in standard units (Kilograms) by well trained enumerators. Moreover, my paper complements this literature by showing how market integration and falling trade costs increase separability.

This paper also relates to literature on how rural roads improve livelihood of households in developing countries. 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, this 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 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. I contribute this literature by providing evidence on the effect of roads on household consumption, a more direct measure of household well-being. The current paper uses a Ricardian trade model to quantify the welfare gains from road connectivity and how much of the gain is attributed to improvement in trading opportunities. Moreover, this paper provides the first evidence on how roads induce separability of production decision from consumption preferences, which is a key step in transformation from subsistence farming to a market based farming.

The rest of the paper is organized as follows. In section 2 I present the data and give some

descriptive statistics. Section 3 provides a series of reduced form evidences on the effect of the URRAP roads on measures of market integration, household specialization and welfare. Section 4 presents the model. Sections 5 takes the key predictions of the model to the data. Section 6 quantifies household welfare gains from URRAP roads and how much of this gain is attributed to improvement in trade opportunities. Section 7 concludes the paper.

2 Data

2.1 Sources

Agricultural production and consumption data: I use two data sources for agricultural production. The first is Ethiopian Socioeconomic Survey (ESS), which is an exceptionally detailed panel data of about 4,000 nationally representative farm households for the years 2011, 2013 and 2015. The data includes farm household's production, consumption and market participation information. The second dataset is the Agricultural Sample Survey (AgSS), which is the largest annual agricultural survey in the country covering over 40,000 farm households in about 2000 villages. While this dataset goes back as far as 1995, villages were resampled every year until 2010. 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 detail 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.

The main advantage of the ESS dataset is its richness as it includes both production and consumption information. That is, I observe a household's production of each crop as well as consumption of each crop disaggregated by source (whether is comes from own production or purchase).[5] I use this data to estimate some of the model parameters and to test separability between production decisions and consumption preferences. A big caveat of this data set is that it covers households in only about 300 villages. The main advantage of the AgSS dataset over the ESS dataset is its wide geographic coverage, larger sample size, and reliable estimation of

[5]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.

quantity of production. Thus, I mainly use this dataset to analyze the welfare effect of road connectivity of villages.

Price data: The price data comes from three different sources. The first 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.[6] The second is the Retail Price Survey (RPS), which is a monthly 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. Both these datasets are collected by CSA and go back to at least 1996. Importantly, the agricultural products covered in both datasets overlap almost fully. For the years before 2010, CSA resamples villages for AgPPS so that following a village price over period is difficult. Starting from 2010, however, CSA keeps the sample villages fixed. I use the AgPPS and RPS data mainly to test how road connectivity of a village affected the price gap between the village and the zonal administrative centers.

AgPPS and ESS do not fully overlap in terms of geography or crops or both. I use the price survey in ESS for testing separability and to estimate the model parameters. Unfortunately ESS's price survey is not exhaustive in its coverage of crops. I overcome this problem by using the sample of households who report a positive purchases/sales of crops to construct village level unit values of crops in the cases where AgPPS prices are missing.

Rainfall and agro-climatic data: I use FAO/GAEZ agro-climatically attainable yield for low/intermediate input use to construct villages' crop suitability, which is used in the separability test and to test how road affects the relationship between local comparative advantage and local prices. Unfortunately the GAEZ data doesn't include some of the most widely grown crops in Ethiopia such as *Teff*. For such crops, I use the AgSS data to construct village level suitability of land to the crops from the average yield in the villages over the period 2010-2016. The high correlation between yield estimates provided by GAEZ and AgSS for the sample of crops that exist in both data ensures that this approach gives a remarkably credible estimate of land-suitability.

The rainfall data comes from Climate Hazards Group InfraRed Precipitation with Station

[6] 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.

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 the massive 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 district capital, which ever is shorter. Figure 1 shows map of the road network before and after URRAP.

The main objective of this project 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 years 2011 and 2012, which are officially considered as capacity building years. Almost all the rural roads constructed under this program were started and completed between 2013-2015. This enables me to use the 2011 and 2013 round of ESS to test the parallel trend assumption when evaluating the impact of URRAP.

Concern of endogeneity of road placement: One big concern when evaluating the effect of infrastructure such as roads is that of potential endogeneity – villages get road connection not randomly but rather are selected based on some demographic, social and economic factors. Even though the project was fully funded by the federal government, the federal government completely decentralized the implementation of URRAP 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. Unfortunately there was no official guideline as to which villages should be selected for a given year. However, informal talks to officials at Ethiopian Roads Authority (ERA) suggests that the main factors determining whether a village would be selected for URRAP in a particular year, in their order of importance, 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. Hence, construction of URRAP roads that connects a village center to preexisting road should start from preexisting road network and progress all the way towards the center of the village. As a result, villages that are closer to the preexisting road network would by default get priority. Population density is relevant both for political consideration and the project's labor input requirement. Most of the labor input for the URRAP roads are contributed by local residents, about three-quarters of which is a free labor. Finally, terrain and landscape significantly affects the road construction costs. Villages that require many bridge constructions or those with highly rugged terrain are usually less favorable due to cost considerations.

I address endogeniety concern by controlling for these factors in all my regressions. Moreover, because these factors are largely time invariant in the short-run (two - three years specifically), using a village fixed effect would substantially reduce the concern of endogeneity. Thus all my regressions concerning the impact of road include either village fixed effects, or the above mentioned control variables.

2.2 Descriptive statistics

In this section I present some descriptive statistics about farm households in rural Ethiopia to guide the theoretical framework and empirical analysis.

Farmers face considerable barrier to trade: These barriers to trade are both physical and pecuniary. Table 1 shows the modes of transport used by farmers to get to market to sell their produce. The most frequently used mode of transport are *on foot* and *pack animals*, together accounting for more than 85% of transaction cases. Vehicle transport accounts for just 2.34% in 2011, and increases to 5.69% in 2015. Though vehicle transport is the least frequently used, it accounts for about one-third of the volume of transaction by value and quantity. The ad valorem trade cost (transport cost per value of transaction) on vehicle is very high (the median is 6.49 % in 2011 which decreases to 3% in 2015). The size of this cost is comparable to international trade costs estimated by Hummels (2007) for US and New Zealand import, although here the distance traveled is just few kilometers. Perhaps the low share of vehicle transport is attributed to farmers choosing not to use this option due to its higher pecuniary cost. The last row of table 1 shows inflation adjusted median transport fare from a village to district capital decreases from

0.7 Birr/km to 0.523Birr/km between 2011 and 2015.[7]

Households are less likely to consume a crop that they do not produce: Table 2 shows the fraction of households who have reported a positive amount of consumption of a crop and the fraction who consumed a positive amount of a crop but did not produce the crop (consumed from purchase).[8] The first two columns report the statistics for a sub-sample of households from small towns (a population of less than 10,000) while the last two columns are for rural households. There is a clear distinction between small town and rural households: (1) households in small towns are more likely to consume vegetables and relatively more expensive cereals such as *Teff* compared to their counterparts in the rural areas (on the contrary, rural households are more likely to consume cheaper cereals such maize, sorghum and millet compared to their urban counterparts), and (2) households in small towns are more likely to consume a crop that they did not produce compared to rural households. For example, about 59% of rural households report consumption of maize while only 23% consumed from purchase (in other words only 40% (23/59) of the households who consumed maize purchased the maize, the rest consumed from own production). On the contrary, in small towns, most of those who consumed a crop did not produce the crop.

While the difference between households in small towns and those in rural areas could in part be driven by income gaps and by the fact that households in small town are more likely to engage in non-farm activities (though over 75% of the sample households in small town and 94% of household in rural villages did not have any non-farm income), a significant part might be attributed to better access to markets in small towns. In small towns there are more frequent markets, there are shops, the markets are larger since most of the surrounding villages transact with them and more importantly, the towns are connected to the rest of the country via all-weather roads.

Most of crop production is consumed within the household: Table 3 reports crop utilization within a household. On average, about 71% of all crop production is consumed within the household and only 13% is marketed.

[7] Ethiopia's currency is called Birr. One USD is sold for about 17 Birr in 2011.

[8] ESS asks households how much of each crops they consumed over the seven days before the interview day, disaggregated into from purchase, and from own production.

3 The impacts of URRAP: DID estimation

In this section, I provide a series of reduced-form evidences on the effect of URRAP on measures of market integration, specialization, and household welfare. I use the generalized difference-in-differences estimation strategy to identify the causal effect of village road connectivity.

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 whether this was achieved 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:

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

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; $Road_v$ 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 4. 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.[9] The gap between these two prices are plausibly capturing trade costs, which decrease by about 2.4%.

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

URRAP decreases the correlation between local prices and yields: One key 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:

$$lnP_{vt}^{k} = \alpha_1 lnA_v^k + \alpha_2(Post_t * Road_v) + \alpha_3(lnA_v^k * Post_t * Road_v) + \gamma_v + \gamma_k + \gamma_t + \varepsilon_{vt}^k$$

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

The result is presented in Table 5. We see that there is a negative relationship between local prices of a crop and local comparative advantage, and 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 1.7%.[10]

URRAP and reallocation of land: I investigate whether villages respond to road connectivity by allocating more land to crops in which they have comparative advantage. In order to identify a village's comparative advantage crop(s), I use GAEZ data on agro-climatic attainable yield for 20 crops. I define a village's comparative advantage crops as those crops that are in the top 30% in terms of relative productivity compared to national average. That is, let A_v^k denotes yield in kilograms per hectare in crop k for village v, and $\frac{A_v^k}{A^k}$ is the village yield relative to national average A^k for crop k. First, I define a village's absolute advantage crops (AA-crops) as those crops with $\frac{A_v^k}{A^k} > 1$. To define a village's comparative advantage crops (CA-crops), for each village I rank crops according to their relative productivity $\frac{A_v^k}{A^k}$ and use crops in the top 30% as baseline cutoff for CA-crops. I extend this baseline cutoff to top 40%, top 50% and top 60% to see how the results change. One issue is many villages grow only a handful of the 20 crops for which GAEZ data is available, and using the above procedure would end up classifying all or most of the crops grown in a village as CA-crops in villages that grow few crops.[11] For

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

[11] About 10% of the villages grow four or less crops and the maximum number of crops grown in a given village is 15, out of the 20 GAEZ crops.

instance, if a village grows only 6 of the 20 crops for which GAEZ data is available and we define CA crops as top 30% in relative yield, in most cases all the 6 crops grown in the village are defined as CA crops and as a result we would not discern any land reallocation because we are considering 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-road years and rank these crops in their relative productivity $\frac{A_v^k}{A^k}$. In this approach, for a village that grows only 6 of the 20 GAEZ crops, the CA-crops are the top 2 crops in $\frac{A_v^k}{A^k}$ ranking within the village.

I estimate the following regression at a village level:

$$\eta_{vt}^k = \alpha_1(Post_t * Road_v) + \alpha_2 CA_v^k + \alpha_3(Post_t * Road_v * CA_v^k) + \beta\mathbf{X} + \delta CA_v^k * t + \gamma_v + \gamma_t + \varepsilon_{vt}^k$$

where η_{vt}^k denotes the share of land allocated to crop k in village v in year t. CA_v^k is dummy variable indicating whether crop k is among the village's comparative advantage crops, and X is a vector of village characteristics such as the population density and rainfall. I also include the interaction of time trend with CA-crop dummy to capture any secular trend in land allocation to CA-crops over time. The regression includes village fixed effect to address the potential endogeneity of road placement and year fixed effect to account for any year specific factor shared across villages.

Table 6 presents the results. The result clearly shows that villages that got road connection are allocating more land to their CA-crops. The first column shows that area of land allocated to the top 30% crops (in relative yield rankings) increased by about 11.2% (0.022/0.196), relative to the land share of these crops in 2011, following road connection. The second, third and fourth columns show that the effect of road significantly decreases and loses its statistical significance as we extend the cutoff for the definition of CA-crops, which supports the result in the first column. The last column sees if villages reallocate more land to their AA-crops, i.e., the crops in which their productivity is higher than the national average. The result shows that about 10% more land is reallocated to these crops following road connectivity. I estimate similar regressions by restricting the sample to villages that grow at least five crops. Table A2 in the appendix reports the results. Land allocated to CA-crops (top 30% crops) increase by about 19.6% relative to the land share of these crops in the year 2011.

In the appendix, I also estimate an alternative specification in which crops are grouped into quartiles, within a village, based on their relative yield, $\frac{A_v^k}{A^k}$. The 1st quartile includes crops with

lowest relative yield and the 4th quartile includes crops with the highest relative yield. The result is reported in table A1 in the appendix. Clearly, land is reallocated to crops in the 4th quartile following road connectivity – the land share of the crops in the 4th quartile increases by 11.4% relative to the average land share of a crop in the year 2011, which strengthens the above result that land is reallocated to crops in the top 30% in relative yield.

URRAP increased household consumption: I investigate whether households in the villages that got road connection under URRAP have seen an improvement in their well–being using ESS panel data on household expenditure on food and non-food items for the years 2011/12 and 2015/16. I use nominal adult-equivalent expenditures and local CES price indices with an elasticity of substitution equal to 1.27 to convert the nominal expenditures to real expenditures.[12] I estimate the following difference-in-differences regression:

$$lnC_{hvt} = \alpha_1 Post_t + \alpha_2(Post_t * Road_v) + \gamma_v + \varepsilon_{hvt}$$

where C_{hvt} nominal or real adult-equivalent expenditure on food and/or non-food items for household h in village v and year t.

Table 7 presents the result. Both nominal and real adult-equivalent household consumption expenditures increased significantly for households in villages that got new road connection, relative to households in villages that did not get road connection under URRAP. Also, both total expenditure and expenditure on food increased by almost equivalent amounts, 12-14%, consistent with the fact that on average, more than three quarters of spending are on food.

Informed by these reduced-form empirical results, I develop a multi-sector Ricardian trade model with heterogeneous land in the next section to further analyze how road constructions affect household welfare.

4 Theoretical framework

The model builds on Donaldson (2018) and Sotelo (2018). In particular, I emphasize the effects of roads where the pre-existing environment is autarky or near-autarky. In such an environment, I show that roads potentially have a fundamental effect on how households allocate resources.

Consider an economy populated by I farmers indexed by $i = 1, ..., I$. I use the same index

[12] Section 6 presents in detail the estimation of the substitution elasticity between crops from my own data.

to refer to a farmer's geographic location as well. There are K homogeneous crops indexed by $k = 1, ..., K$ that can be potentially produced and/or consumed by the farmers.

Preferences: A farm household spends all its income on crops and its preference over different crops is given by

$$C_i = \Big(\sum_k (a_i^k)^{\frac{1}{\sigma}} (c_i^k)^{\frac{\sigma-1}{\sigma}} \Big)^{\frac{\sigma}{\sigma-1}}$$

where σ is elasticity of substitution between crops, c_i^k is the quantity of crop k consumed by farmer i, and a_i^k is crop demand shifter or *taste*. For estimation, households living in the same region, say d, are assumed to share the same taste, i.e., $a_i^k = a_{i'}^k, \forall i, i' \in d$.

Technology: Similar to Sotelo (2018) 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.[13] Each farmer owns L_i amount of land, which is divided into a continuum of plots of size one indexed by $\omega \in \Omega_i$, where Ω_i is the set of plots owned by farmer i such that $\int_{\Omega_i} \omega d\omega = L_i$. Each of the plot is different in how well it is suited to growing different crops which I denote as $z_i^k(\omega)$. 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:

$$q_i^k(\omega) = z_i^k(\omega)$$

where $q_i^k(\omega)$ is yield per unit of plot.

The farmer draws $z_i^k(\omega)$ independently for each plot-crop from a Fréchet distribution with the following cumulative distribution function:

$$F_i^k(z) = Pr(Z_i^k < z) = \exp(-(A_i^k)^\theta z^{-\theta})$$

where A_i^k is the location parameter for the distribution of crop-suitability of land across the set of plots owned by farmer i, Ω_i. A_i^k can be interpreted as the average productivity of farmer i's plots for crop k. For farmers who live in agro-climatic conditions that are impossible to produce crop k, A_i^k is set to zero. θ is the degree of homogeneity in the set of plots owned by a farmer,

[13] 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.

and it is constant across farmers and crops.

Trade and equilibrium prices: Farmers operate in a perfectly competitive crop market. Farmers are geographically separated and there is an iceberg trade cost τ_{ij}^k between farmers i and j in crop k. Motivated by the result in section 3, the trade cost is 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 $\tau_{ii}^k = 1, \forall k$, and impose standard assumptions that $\tau_{ij}^k \times \tau_{jn}^k \geq \tau_{in}^k, \forall k$.

Let r_i is the rental rate of plots owned by farmer i, which is determined in equilibrium. The unit cost of production $c_i^k = \frac{r_i}{Z_i^k}$ is stochastic because it is a function of stochastic productivity Z_i^k. As a result, the price at which farmer i supplies crop k to farmer j, $P_{ij}^k = \frac{r_i}{Z_i^k}\tau_{ij}^k$, is stochastic.

Using the distribution of Z_i^k, we obtain the following distribution of the prices of crop k that farmer j is offered by another farmer i:

$$G_{ij}^k(p) = 1 - \exp\big(- (A_i^k)^\theta (r_i \tau_{ij}^k)^{-\theta} p^\theta \big)$$

Because crops supplied by different farmers are homogeneous, farmer j buys each crop k from any farmer that supplies the crop at the lowest price. Thus, the distribution of the price of crop k that is actually paid by farmer j is the distribution of the lowest prices across all other farmers and is given by:

$$\begin{aligned} G_j^k(p) &= 1 - \prod_{i=1}^{I} (1 - G_{ij}^k(p)) \\ &= 1 - \exp\big(- p^\theta \sum_{i=1}^{I} (A_i^k)^\theta (r_i \tau_{ij}^k)^{-\theta} \big) \end{aligned} \tag{1}$$

The expected value of this price distribution is given by

$$p_j^k = \Gamma\big(\sum_{i=1}^{I} (A_i^k)^\theta (r_i \tau_{ij}^k)^{-\theta} \big)^{\frac{-1}{\theta}} \tag{2}$$

where Γ is a gamma function $\Gamma(1 + 1/\theta)$. In the empirical analysis, I assume that the prices collected by the statistical office are equal to p_j^k. Given the CES preferences, the price index

faced by farmer j is given by:

$$P_j = \Big(\sum_k a_j^k p_j^{k\,1-\sigma}\Big)^{\frac{1}{1-\sigma}} \tag{3}$$

where p_j^k is given by (2).

The probability that farmer i is the cheapest supplier of crop k to farmer j (the probability that farmer i's productivity draw adjusted for trade costs and rental rates is the highest compared to all other potential farmers trading with farmer j) is:

$$\begin{aligned}\pi_{ij}^k &= \Pr\Big[P_{ij}^k \leq \min_n\{P_{nj}^k\}\Big] \\ &= \frac{(A_i^k)^\theta (r_i \tau_{ij}^k)^{-\theta}}{\sum_i (A_i^k)^\theta (r_i \tau_{ij}^k)^{-\theta}}\end{aligned}$$

which is increasing in the average productivity of farmer $i's$ plots in crop k, A_i^k and decreasing in the trade cost, τ_{ij}^k and the rental rate of farmer i's plot r_i relative to other farmers.

The probability that a farmer will be the cheapest supplier of a crop to *itself* is

$$\begin{aligned}\pi_{ii}^k &= \Pr[P_{ii}^k \leq \min_{\{n\neq i\}}\{P_{ni}^k\}] \\ &= \frac{(A_i^k)^\theta r_i^{-\theta}}{\sum_n (A_n^k)^\theta (r_n \tau_{ni}^k)^{-\theta}}\end{aligned}$$

Farmer i is more likely to self-produce crop k if the farmer is more productive in the crop relative to other farmers and/or the higher the trade cost between farmer i and other farmers. Note also that $\frac{\pi_{ii}^k}{\pi_{ii}^{k'}} = \big(\frac{A_i^k}{A_i^{k'}}\big)^\theta$, that is, the probability that a farmer would be the cheapest supplier of crop k to itself, relative to crop k', is related to the farmer's productivity in crop k relative to crop k' and the trade elasticity.

Equilibrium rental rate: It can be shown that the $r_j(\omega)|\omega \in \Omega_j^k$ has a Fréchet distribution with parameter Φ_j (see Appendix A1 for the derivation). The distribution of rental rate of plots owned by a farmer have the same parameter for all crops. That is the expected value of the rental rate of farmer j's plots is given by Φ_j irrespective of which crop is planted.

Equilibrium land allocation: *Case 1: separability ($\tau_{ij}^k < \infty, \forall k$)*

When $\tau_{ij}^k < \infty, \forall k$, the farmer's problem is standard: the farmer first decides how to allocate its land across different crops given market prices of crops and the productivity of its land in the

crops, and given the revenue from production, the farmer then maximizes utility by consuming optimal quantities of each crop given the crop prices and the farmer's tastes for the different crops. The farmer's problem is called *separable* because the farmer's production decision is independent of the farmer's consumption preferences. Importantly, this requires trade costs to be less than infinity.

The farmer decides how to allocate its land across different crops given prices p_j^k and the suitability of the land for various crops. Revenue maximization implies that each plot of land is allocated to a crop that yields the highest return:

$$r_j(\omega) = \max_k \{p_j^k z_j^k(\omega)\} \tag{4}$$

where $r_j(\omega)$ is rental income/revenue from plot ω. Together with the Fréchet distribution this implies the following land allocation:

$$\eta_j^k = \frac{(p_j^k A_j^k)^\theta}{(\Phi_j)^\theta}, \quad \text{where} \quad \Phi_j = \Big(\sum_{l=1}^{K} (p_j^l A_j^l)^\theta\Big)^{\frac{1}{\theta}} \tag{5}$$

where η_j^k is the fraction of farmer j's land allocated to crop k. Thus the fraction of land allocated to crop k increases with the price of the crop and the average productivity of the household's land in the crop, relative to other crops. The elasticity of land share of a crop to the productivity of land is given as $\frac{d\ln\eta_j^k}{d\ln A_j^k} = \theta(1-\eta_j^k)$, which is higher the higher the value of θ. Intuitively, when land is more homogeneous (higher θ) small difference in land productivity translates to larger difference in land share of the crop. The elasticity of land share to crop price is also given by the same expression.

Most importantly, optimal land allocation across crops is independent of the household's consumption preferences for different crops when $\tau_{ij}^k < \infty, \forall k$.

Case 2: autarky ($\tau_{ij}^k \longrightarrow \infty, \forall k$): Under autarky the household makes production and consumption decisions jointly. The household's problem is how to allocate its limited land across different crops in such a way that the household's utility is maximized subject to the land constraint.

Because there is no trade, land allocation is based on some implicit crop price $\tilde{p}_j^k$, which is in equilibrium equal to the household's marginal utility from the crop. The household then solves the land allocation problem in analogous way as in the separable model, except that here we have an implicit price for each crop as opposed to a market price. The household allocates land

across crops in such a way that the rent (or value) of each parcel ω is maximized:

$$\tilde{r}_j(\omega) = \max_k \{\tilde{p}_j^k z_j^k(\omega)\}$$

where $\tilde{p}_j^k = C_j^{1/(\sigma-1)}(a_d^k)^{1/\sigma}(c_j^{k*})^{-1/\sigma}$ is the implicit price of crop k.

Following the same procedure as in the separable model, we obtain the following optimal allocation of land:

$$\tilde{\eta}_i^k = \frac{(\tilde{p}_i^k A_i^k)^\theta}{(\tilde{\Phi}_i)^\theta}, \quad \text{where} \quad \tilde{\Phi}_i = \Big(\sum_{l=1}^{K}(\tilde{p}_i^l A_i^l)^\theta\Big)^{\frac{1}{\theta}}$$

Imposing the autarky restriction that the household has to consume all the amount produced of each crop, $q_i^k = c_i^{k*}, \forall k$, gives the following land allocation:

$$\tilde{\eta}_i^k = \frac{\Big(a_d^{k\,\frac{1}{\theta+\sigma-1}} A_i^{k\,\frac{\sigma-1}{\theta+\sigma-1}}\Big)^\theta}{(\tilde{\Phi}_i)^\theta}, \quad \text{where} \quad \tilde{\Phi}_i = \Big(\sum_{l=1}^{K}\Big(a_d^{l\,\frac{1}{\theta+\sigma-1}} A_i^{l\,\frac{\sigma-1}{\theta+\sigma-1}}\Big)^\theta\Big)^{\frac{1}{\theta}} \tag{6}$$

Two differences from land allocation under separability in equation 5 are worth discussing. First, as opposed to the case of separability, the fraction of land allocated to a crop under autarky is influenced by the household's taste for the crop relative to other crops, a_d^k. Second, the elasticity of land share of a crop to the productivity of land A_i^k is $\frac{d\ln\tilde{\eta}_j^k}{d\ln A_j^k} = \frac{\theta(\sigma-1)}{\theta+\sigma-1}(1-\tilde{\eta}_j^k)$, which is smaller than the case of separability. That is, under autarky land allocation is less sensitive to crop specific productivity of the land. This is because the higher the productivity of land in crop k, the lower is the amount of land required to produce the crop for consumption need since the household equalizes the marginal utility of consumption across crops. This implies that land is utilized less efficiently under autarky.

5 Testing the model's implications

One of the prediction of the above model is that when trade costs are high, a farm household's production decisions (particularly allocation of lands across different crops) is dictated by the household's consumption preferences for the crops, given the productivity of land. This follows from the observation that as $\tau_{ij}^k \longrightarrow \infty, \forall k$ the household becomes the cheapest supplier of all the crops to itself. Under this case the farmer's land allocation across crops is dictated by its

tastes for the crops. As $\tau_{ij}^k \longrightarrow 1, \forall k$ the household specializes in a crop in which its productivity is higher compared to other farm households and relies on the market to consume other crops. In this case the household's production decision is separable from its consumption preferences.

The budget share of a crop under separability and autarky are given, respectively, by:

$$s_i^k = \frac{a_i^k (p_i^k)^{1-\sigma}}{\sum_l a_i^l (p_i^l)^{1-\sigma}}, \quad \text{and} \quad \tilde{s}_i^k = \frac{\left((a_d^k)^{\frac{1}{\theta+\sigma-1}} (A_i^k)^{\frac{\sigma-1}{\theta+\sigma-1}} \right)^\theta}{(\tilde{\Phi}_i)^\theta} \tag{7}$$

The following proposition summarizes the relationship between budget and land allocation under separability and autarky given in equations 5-7, and how change in trade costs affect the correlation.

Proposition 1. *Under the assumption that local taste shifters are orthogonal to local crop productivities, (i)-Under autarky, the land and budget shares of a crop are perfectly correlated while under separability there should not be correlation between the two; and (ii)- road connectivity reduces the correlation between the land and budget share of crops by decreasing trade costs.*

In this section I test these two prediction of the model.

5.1 The separability test

The above discussion suggests that one can test for both separability and autarky in the following regression:

$$\eta_{hvt}^k = \beta_0 + \beta_1 s_{hvt}^k + \gamma_h + \gamma_t + \varepsilon_{hvt}^k \tag{8}$$

where η and s are land and budget share of crop, h is household, v is a village, k is crop, and t is year. Separability requires that $\beta_1 = 0$, that is, there is no significant correlation between land and budget share of crops within household. Thus, finding a positive correlation between land share of a crop and expenditure share of a crop, i.e., $\beta_1 > 0$ is evidence against separability. Autarky requires $\beta_1 = 1$, a perfect correlation between household land and budget allocation across crops.

Threats to the test The threats to the validity of this test come from confounding factors that would make the land share of a crop and the expenditure share of a crop correlated under separability. That is if η_{hi}^k and s_{hi}^k are correlated for any other reason than consumption

preferences dictating production choices, the test will be invalid. I address the known concerns below.

A. Geographic variation in taste for crop: One big concern is the violation of the orthogonality assumption between taste and productivity; locations that have a higher taste for a crop could be those that are naturally more suitable to grow the crops. Atkin (2013) argues that tastes for foods are formed over generations and that the cause of regional taste difference is spatial difference in the crop specific productivity of soil in the past. If a_i^k is systematically correlated with A_i^k (crop-suitability of locations) as suggested in Atkin (2013), then we would have $\hat{\beta}_1 \neq 0$ even if the household's production decision is independent of their consumption choices. For example, a region that is well suited to grow wheat is more likely to be populated by people who have developed a taste for wheat (meals made of wheat) over generations. As a result, we do not know if households in this region are allocating more land to wheat due to their endogenously formed higher taste for wheat or due to profitability (high yield) of growing wheat in the region. I use FAO/GAEZ crop-suitability index as a control variable to address this concern. FAO/GAEZ crop-suitability index gives each location's suitability to grow each crop based a number of factors related to climate and soil fertility.

B. Non-homothetic preferences: If preferences are non-homothetic, expenditure share of a crop will depend on the household income. Consider the Stone-Geary preferences as an example: $U_i = \sum_k \beta^k ln(c_i^k - \gamma^k)$ which implies expenditure share of a crop of $s_i^k = \frac{p_i^k \gamma^k}{m} + \frac{\beta^k (m - \sum_k p_i^k \gamma^k)}{m}$, where m is income and γ^k is the subsistence level of consumption of crop k. The above test will be invalid if: (1) households' production choices are also dependent on household income/wealth, e.g., wealthier households might take higher risk or might allocate resource in a more profit maximizing way because their wealth would give them a leverage to experiment different resource allocation; (2) m is correlated with A_i^k, which is likely to be the case because higher yield translate into higher income; and (3) γ^k varies across households and is correlated with A_i^k. This could happen, for instance, if the household tries to maintain minimum subsistence level of few crops and these crops are chosen based on crop-specific yield of land for food security reason. The last two concerns are accounted for by including the FAO/GAEZ yield measure as a control variable. To account for the concern in (1), I use household land size as a proxy for wealth. I

also use household farm and non-farm income and find that the estimates are robust.[14]

The final estimation equation that addresses the above threats is as follows:

$$\eta^k_{hvt} = \beta_0 + \beta_1 s^k_{hvt} + \beta_2 \ln A^k_d + \beta_3 \ln L_{ht} + \gamma_h + \gamma^k + \gamma_t + \varepsilon^k_{hvt} \tag{9}$$

where A^k_d is FAO/GAEZ district-level crop suitability, L_{ht} is household's cultivated land size (a measure of wealth or income), γ_h is household fixed effect, γ^k is crop fixed effect, and γ_t is year fixed effect.

5.2 Estimation and result

I use ESS data and focus on 20 crops for which complete information is available on both production and consumption. Some issues arise during estimation which are worth mentioning here.

First, it is common to see some households neither producing nor consuming some of these crops. As a result both the land share of a crop and the budget share of crop would be zero. Since this would mechanically inflate the correlation between land share and budget share of crops, I include crop fixed effects in addition to household fixed effects. As a sensitivity check, I also report results based on cases where a household either produces or consumes a crop. This is conservative in the sense that the estimated correlation will be smaller than other sensible options, but it still allows me to exploit the extensive margin variation (how often a household consumes a crop they do not grow and how often they grow a crop they do not consume). Thus, in this case the estimate is interpreted as the correlation between land and budget share of crops, conditional on the household either consuming or producing a positive amount of the crops.[15]

I run the regressions for each year separately to show how the estimated correlation changed across years. Even though I have data for the years 2011, 2013, and 2015, I focus on the years 2011 and 2015 which are before and after the massive rural road expansion under URRAP Phase-I. I use the 2013 data for falsification test because while the fraction of roads completed before 2013 was small, there was a wave of new roads that opened up in 2013.

Table 8 reports the results. The first two columns in Table 8 are based on all observations,

[14] Because land and consumption shares sum up to 1 within a household, using household fixed effect is redundant, hence has no effect on the estimate.

[15] I also estimate similar regressions conditioning on: (1) the household consuming the crop; (2) the household producing the crop; and (3) the household both producing and consuming the crop. While the size of the estimates vary moderately, all approaches yield qualitatively similar results.

including data points in which the household neither consumes nor produces some of the crops. The first column reports the result for 2011. It shows a positive and precisely estimated effect of the budget share of a crop on the land share of the crop equal to 0.471. Column 2 reports the result for 2015 and we see that the estimate drops dramatically to 0.198. Columns 3-4 are based on data points in which a household either consumes or produces a crop, i.e., crops which a household neither produces nor consumes are dropped from the household's observations. We see that the results are qualitatively similar: the correlation between land and budget share of crops decreases from 0.356 in 2011 to 0.07 in 2015.

Table 8 shows that households' production decisions are partially dictated by their consumption preferences. If separability holds, we should not see such robust and statistically significant effect of consumption preferences on production decisions. Hence, these results can be taken as strong rejection of the separability assumption. Table 8 also shows that the influence of consumption preferences on production decisions dramatically drops between 2011 and 2015. The results in table 8 are robust to restricting the estimation to cases where: (1) the household consumes the crop, (2) the household produces a crop, and (3) the household both produces and consumes a crop. As shown in the subsequent sections, the massive rural road expansion project under URRAP is the main reason behind the dramatic drop in the correlation between land and budget shares of crops.

Robustness: using labor allocation. As a main robustness check of the above results, I use data on *plot level* labor use (both planting and harvesting hours of labor) which I convert to *crop level* labor use given the information on which crops covered a plot during a given year. Given this, I calculate the labor share of crop in exactly analogous way to the land share of crop. I then redo all the above regressions using the labor share of a crop as the dependent variable. Table A3 in the appendix presents the results. The results looks very similar to the one we obtained using the land allocation. The correlation between the labor and budget share of crops significantly decreases from 0.456 to 0.191 between the years 2011 and 2015.

Robustness: the classic test of separability. The next robustness check exploits the richness of the ESS data to test separability following the classic approach introduced by Benjamin (1992). This approach tests separability using the relationship between household on-farm labor demand and the household's demographic characteristics. The basic idea is as

follows.[16] If labor market is complete and farm household's production decisions are independent of the household's preferences, household's on-farm labor demand should be independent of the household's demographic composition, such as the number of active age persons in the household.

The critical challenge in testing separability in this approach is that unobserved factors may affect both the household demographic composition and the household's farm labor demand. For example, household's land holding and/or the quality of the land may affect both household labor demand and household size (which is likely to be endogenously chosen based on wealth/land holding). While household land holding is reported in many surveys, accounting for land quality is often quite difficult. Another example includes shocks (such as weather shock) that effect both farm labor demand and household size through migration of family members. Drought decreases farm labor demand and may also lead some of the household members to migrate to cities for non-farm employment. Household specific shocks such as death and giving birth affect both labor demand and household demography.

Equipped with a panel data and a significant geographic variation in my sample households, I mitigate most of these problems using fixed effects. Time invariant household characteristics such as land size/quality are subsumed into household fixed effects. Shocks that uniformly affect households at a given location are accounted for by location-year fixed effects. The effect of household specific shocks that are likely to be correlated with household labor demand and demographic characteristics are addressed by restricting estimation to sub-samples with constant household size across the sample period.

I run similar specifications as Benjamin (1992) and LaFave and Thomas (2016) to compare my results with theirs. In my data, labor is measured in hours of work, and I observe hours spent on *planting* and *harvesting* separately. I report results for *total* labor demand (harvesting *plus* planting hours), and separately for planting and harvesting labor. Table B1 reports the estimation results.

The result implies an unambiguous rejection of separability – household demographic composition significantly affects household labor demand. See Appendix B for a detailed explanation of the results.

To sum up, this classic test strongly rejects separability and is consistent with the new test suggested in this paper. However, there are important differences in the two approaches. While

[16] I refer interested readers to Benjamin (1992) and LaFave and Thomas (2016) for detailed discussions on the theoretical frameworks underlying this approach.

any market incompleteness can lead to rejection of separability in the classic test, the test can be considered as a direct test of *missing* labor markets. On the other hand, the approach suggested in the current paper can be considered as a direct test of *missing* crop markets. In this sense, the two approached also complement each other. Also important is that the classic separability test relies on recall based data on labor input. Given the fact that most of the farm households are self-employing, the reliability of such data is questionable. The method suggested in the current paper is *less* prone to such problem: land area is measured by trained enumerators using GPS tools and quantities of crops are measured in local metrics and then converted to standard metrics.

5.3 Separability and proximity to market centers

Next I investigate how separability is related to proximity to market or trade costs. The model suggests that the correlation between the land and budget share of a crop should increase with trade costs. I use distance to nearest population centers (towns with above 20,000 population) and distance to all-weather roads as proxy for trade costs. The first is time invariant while the latter decreases for households residing in villages that obtained new roads under URRAP. Towns are market centers for the surrounding villages. They are also connected with the rest of the country by all-weather roads. Hence proximity to towns is important for market access. The same is true for proximity to all-weather roads. I investigate whether the correlation between land and budget shares increases with these measures of trade costs in the following regression:

$$\eta^k_{hvt} = \beta_0 + \beta_1 s^k_{hvt} + \beta_2 DIST_{hvt} + \beta_3 (s^k_{hvt} * DIST_{hvt}) + \beta_4 \ln A^k_{dt} + \delta Z_{vt} + \gamma_t + \varepsilon^k_{hvt} \qquad (10)$$

where $DIST_{hvt}$ is a measure of distance, which is a dummy variable or log distance depending specification. A positive and statistically significant β_3 implies that non-separability becomes stronger as trade costs increase. Z is a vector of village-level controls.

Table 9 reports the results. The first column shows that households who live above median distance (30km) from population centers have significantly higher, by about 0.101, correlation between land share and budget share of crops. The second column shows the similar result using distance to all weather roads: households with above median (10km) distance from all-weather roads have 0.117 higher correlation between land and budget share of crops. The third and fourth columns use continuous measures of distances and tell similar story. In the third column

we see that one log-unit increase in distance from population centers is associated with 0.08 increase in the correlation between land and budget shares, while the fourth column shows that one log-unit from all-weather roads increases the correlation by 0.052.

These results clearly indicate that trade costs are at the center of non-separability of production choices from consumption preferences. This is consistent with the model's prediction that households facing higher trade costs are more likely to be the cheapest supplier of a given crop to themselves, thereby resorting to subsistence farming where the household produces almost all of its consumption needs.

Table A4 uses household labor allocation to estimate similar regressions. The results are both quantitatively and qualitatively similar to table 9.

5.4 URRAP and separability: generalized DID estimation

One of the main goals of this paper is to investigate to what extent roads facilitate a move from subsistence farming to market oriented one by separating household production decision from their consumption preferences. I investigate whether the road construction under URRAP is responsible for the significant drop in the correlation between land/labor allocation and budget share of crops over the 2011-2015 period reported in the previous subsections. Even though URRAP was launched in 2011, very few roads were constructed before 2013. Thus the 2013 data can be used to test the parallel trend assumption – whether the correlation between land and budget share of crop evolved similarly between treated villages (those which got new URRAP road) and non-treated villages (those which didn't get URRAP road). Figure 5 shows how the estimated correlation between land and consumption shares evolves over time. The figure shows that between 2011 and 2013, the period with very little road construction, the correlation for both the treated and non-treated villages did not change significantly. But between 2013 and 2015, we see significant decrease in the correlation for treated villages, while no such significant change occurred in the non-treated villages.

To test this formally I employ the following generalized DID framework:

$$\begin{aligned}\eta^k_{hvt} = \beta_0 + \beta_1 s^k_{hvt} + \beta_2(\text{Road}_v * s^k_{hvt}) + \beta_3(\text{Post}_t * s^k_{hvt}) \\ + \beta_4(\text{Road}_v * \text{Post}_t * s^k_{hvt}) + \delta Z_{vt} + \gamma_v + \gamma_t + \varepsilon^k_{hvt}\end{aligned} \tag{11}$$

where Road_v is a dummy variable indicating whether village v got road connectivity under

URRAP and $Post_t$ is a dummy variable which equals zero for 2011 and one for 2015, and Z_{vt} includes the vector of control variables in the baseline separability test. Z_{vt} also includes population density and village distance to the baseline road network to address the endogeneity concern of road placement in the regressions without village fixed effects. I also include year fixed effects, and crop fixed effects in regressions that include household-crop data points with zero land and budget share. In this specification β_1 captures the correlation between land and budget share in 2011 for non-treated villages, β_2 captures any difference in the *level* of correlation between land and budget share between treated and non-treated villages, and β_3 captures the general trend in the correlation between land and budget share over time. The main parameter of interest is β_4, which captures the causal effect of road connectivity under the assumption that assignment of roads is not endogenous to the the correlation between land and budget shares of crops, conditional on population density, village distance to the baseline network, or village fixed effect.

Table 10 presents the results. Note that the first two columns do not include village fixed effects but instead include population density and village distance to the baseline network to account for potential endogeneity of roads. The last two columns include village fixed effects. Columns 1 and 3 are based on data points in which the household either consume or produce a crop while column 2 and 4 condition on a household consuming the crop. The results clearly show that road construction under URRAP caused a significant decline in the correlation between land and budget shares of crops. Households in villages that got road connection between 2011-2015 under the URRAP program have seen a decrease in the correlation between land and budget shares by about 0.11-0.12, depending on the specification, compared to households in villages that were not directly exposed to the program. This is a large effect, roughly about 25% decline in the correlation relative to the baseline estimates in table 8 of about 0.47 or 0.36 in 2011. In particular, the result is not sensitive to the alternative ways used to address the potential endogeneity of road placements.

Robustness: As a robustness check, I run similar regressions but using labor share of crops as dependent variable. The result is reported in Table A5 in the appendix. This table shows a stronger effect of road on separability compared to the result based on land allocation: the correlation between labor and budget shares decreases by about 0.135-0.154, depending on the specifications.

Overall, these results are strong evidence that the URRAP roads reduced the link between household's production decision and consumption preferences. That is, the roads tend to move households from a subsistence farming towards a market-based one by creating opportunities to trade thereby enhancing a more efficient allocation of resources.

6 Quantifying the welfare gains from URRAP

The model suggests that a farmer's welfare measure is given by real revenue $W_j = \frac{r_j}{P_j}$ where r_j is revenue and P_j is the price index. While closed form solution for the effect of trade cost on W_j cannot be obtained due to complex general equilibrium interaction of prices across locations, it is intuitive that road connectivity increases a farmer's welfare. First, when a village is connected to any other village farmers in the village obtain higher prices, net of trade cost, for their comparative advantage crop which directly increases their revenue per unit of land r_j. Second, road connectivity decreases the prices of consumption bundles P_j since the farmer can now potentially source each crop from another farmer who can offer it at a cheaper price. Hence, the effect on welfare of a decrease in trade cost following road connectivity is positive.

6.1 A sufficient statistic for the welfare gain

A sufficient statistic for the welfare gain from road connectivity can be derived from the model. This sufficient statistic is crucial to shed light on whether the welfare gain from the road connectivity is attributed to the mechanisms suggested in the model. Recall that the probability that farmer i is the cheapest supplier of crop k to farmer j is equal to $\pi_{ij}^k = (A_i^k)^\theta (r_i \tau_{ij}^k)^{-\theta} (p_j^k)^\theta$. Evaluating this expression at $i = j$ and solving for p_j^k we obtain $p_j^k = r_j (A_j^k)^{-1} {\pi_{jj}^k}^{1/\theta}$. Plugging this into CES price index we obtain $P_j = r_j \big[\sum_k a_j^k \big((A_j^k)^{-1} (\pi_{jj}^k)^{\frac{1}{\theta}} \big)^{1-\sigma} \big]^{\frac{1}{1-\sigma}}$ which implies:

$$\mathcal{W}_j \equiv \ln(r_j/P_j) = \frac{1}{\sigma - 1} \ln \sum_k a_j^k A_j^{k\sigma-1} {\pi_{jj}^k}^{\frac{1-\sigma}{\theta}} \tag{12}$$

This expression is similar to the sufficient statistic expressions for welfare gain from a decrease in trade cost given in Arkolakis et al. (2012), except that the expenditure share on own product π_{jj}^k is weighted by local tastes for and productivity of crops. These modifications directly follow from: (1) locations are assumed to have different tastes for crops, and (2) the productivity distribution is assumed to vary across locations for each crop (compared to the same technology parameters assumed for all goods in Arkolakis et al. (2012)). Differentiating this expression with

respect to τ_{ij} we obtain the following

$$\frac{d\mathcal{W}_j}{\mathcal{W}_j} = \frac{-1}{\theta}\sum_k \alpha_j^k \frac{d\pi_{jj}^k}{\pi_{jj}^k}$$

where $\alpha_j^k = \frac{a_j^k A_j^{k\sigma-1}}{\sum_k a_j^k A_j^{k\sigma-1}}$. That is, a decrease in trade cost increases real revenue since $\frac{d\pi_{jj}}{d\tau_{ij}} \geq 0$, for some $i \neq j$.

The main result in this model that enables me to obtain the sufficient statistic of the welfare gain from roads without observing any trade flows is Proposition 2.

Proposition 2. *The probability that a household is the cheapest supplier of crop k to itself π_{jj}^k is proportional to the fraction of the household land allocated to the crop, i.e., $\pi_{jj}^k = \kappa\eta_j^k$.*

Proof. The proof is straight forward. Recall that the probability that farmer i is the cheapest supplier of crop k to farmer j is given by $\pi_{ij}^k = \kappa(A_i^k)^\theta(r_i\tau_{ij}^k)^{-\theta}(p_j^k)^\theta$. Evaluating this expression at $i = j$ we obtain $\pi_{jj}^k = (A_j^k)^\theta(r_j)^{-\theta}(p_j^k)^\theta$, which can be rearranged to give $\pi_{jj}^k = \kappa\frac{(A_j^k p_j^k)^\theta}{r_j^\theta} = \kappa\frac{(A_j^k p_j^k)^\theta}{\sum_l (A_j^l p_j^l)^\theta} = \kappa\eta_j^k$, where the second equality uses the expression for the average rent from the distribution of equilibrium rent, $r_j = \Phi_j$. $\kappa = \Gamma(1+1/\theta)^{-\theta}$, where $\Gamma(.)$ is a gamma function. □

In order to construct the sufficient statistic measure of welfare gain, we need to obtain estimates for the parameters of the model: the elasticity of substitution σ, the measure of land homogeneity within a farmer's plots of land θ, and the local tastes of crops a_j^k.

Estimation of σ and a_i^k: I use the expression for the budget share of a crop to estimate the σ and the taste parameters, a_i^k. It is important to note that estimation of taste parameters at household or village level is impossible because the price varies at village level.[17] As a result, I estimate tastes at a higher level of geographic aggregation, regions. Regions in Ethiopia are relatively ethnically and geographically homogeneous.[18] Hence, it is fairly reasonable to assume that households within a region share similar taste for a crop. Thus, I let a_i^k to be a_d^k within a region d. Taking log of the budget share equation 7:

$$\ln s_i^k = \ln a_d^k - (\sigma-1)\ln p_i^k - \ln\sum_l a_d^l (p_i^l)^{1-\sigma}$$

[17] One need to have a long panel to identify the taste parameters at village level. In that case, village-crop fixed effects could be used as a measure of village crop taste.

[18] Ethiopia practices a federal government system where the federal states, commonly known as regions, are formed mainly based on ethnolinguistic and geographic criteria. There are 9 regional states in the country, and 7 in my dataset.

To estimate (a_d^k, σ), I replace the price index, itself is a function of these parameters, with household's total food expenditure E_{it} to obtain the following estimation equation:

$$\ln s_{it}^k = \ln a_d^k - (\sigma - 1)\ln p_{it}^k + \delta E_{jt} + \varepsilon_{it}^k \tag{13}$$

Unfortunately there is a well known endogeneity concern because unobserved factors could be correlated with both prices and the budget share of crops or measurement error in price could attenuate the coefficient of price. Also GAEZ yield measure does not satisfy the exclusion restriction assumption because it is likely to be correlated to the a_d^k (Atkin (2013)). Because of this, I use distance to road to instrument for price. Distance to road is a valid instrument because: (1) it is unlikely to be correlated to taste, and (2) it is correlated with prices since remote locations have lower prices. I estimation equation 12 using log distance to instrument for price and obtain the coefficient of price of -0.27, with standard error of 0.064. This implies $\hat{\sigma} = 1.27$. The first-stage F-stat is 38.4.[19] This estimate is inline with other studies estimating the elasticity of substitution across food varieties in low income countries. For instance, Behrman and Deolalikar (1989) find $\hat{\sigma} \approx 1.2$. But other studies such as Sotelo (2018) find larger estimate of about 2.3. In my data, restricting the estimation to the set of cereals only gives slightly higher (≈ 2) median estimate. In the subsequent analysis, I use the value of 1.27 for elasticity of substitution. Once I obtain the estimate for σ, I estimate the taste parameters $\hat{a}_d^k$ as the median value (for each crop-region) of the residual from equation 12, transformed by taking exponents.

Estimation of θ I follow Sotelo (2018) in the estimation of θ. Because GAEZ does not provide potential yield estimate for many of the crops in my data, 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 four years (2012-2016) to obtain a time invariant measure of yield for a crop in a village. I assume that the average productivity of a farmer's land is related to the AgSS's village yield measure in the following equation:

$$A_i^k = \delta^k Y_v^k \exp(-u_i^k)$$

[19] I also estimate 12 for each region separately. I obtain $\hat{\sigma} \in [1.05, 2.1]$ with a median 1.3, for six of the seven regions. For one region, I obtain 0.85, which is not statistically different from 1.

where A_i^k is the average productivity of a farmer's land in crop k, Y_v^k is the village yield, $\exp(-u_i^k)$ is a random noise, and δ^k is crop-specific constant. Plugging this for A_i^k in the land share equation and taking logs gives:

$$\ln(P_i^k Y_i^k) = \frac{1}{\theta}\eta_i^k + \ln\Phi_i - \ln\delta^k + u_i^k$$

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. Notice that because the left hand side varies at village level (because both price and yield vary at village level), I aggregate the land share of a crop at village level as well. Thus the estimated θ is essentially a measure of within village land homogeneity.

I obtain a value of $\hat{\theta} = 2.7$ for productivity heterogeneity, which is larger than the estimate of Sotelo (2018) 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.

6.2 Results

I use these estimated parameters along with data on η_j^k and A_i^k to construct the sufficient statistic for welfare gain and village level price index. Because A_j^k (average productivity in crop k across plots owned farmer j) is not directly observable, I use Y_v^k as a proxy for A_j^k.

To estimate the welfare gain from road connectivity I run the following difference-in-differences regression:

$$ln\hat{W}_{hvt} = \alpha_1 Post_t + \alpha_2(Post_t * Road_v) + \delta\mathbf{X} + \gamma_v + \varepsilon_{hvt} \tag{14}$$

where $\hat{W}_{hvt} = \frac{\sum_k p_{vt}^k \times q_{hvt}^k}{\hat{P}_{vt}}$ is real agricultural income of household h, q_{hvt}^k is quantity of crop produced by the household, P_{vt} is village price index calculated using the estimates for regional crop tastes and the elasticity of substitution using the CES price index expression $\hat{P}_{vt} = (\sum_k \hat{a}_d^k (p_{vt}^k)^{1-\hat{\sigma}})^{\frac{1}{1-\hat{\sigma}}}$.

Finally, in order to investigate to what extent the mechanism suggested in the model explains the welfare gain from the roads, I re-run equation 14 controlling for the sufficient statistic for

welfare gains $\hat{\mathcal{W}}_{ht}$ imputed from the model:

$$ln\hat{W}_{hvt} = \alpha_1 Post_t + \alpha_2(Post_t * Road_v) + \beta_1 \hat{\mathcal{W}}_{hvt} + \delta \mathbf{X} + \gamma_v + \varepsilon_{hvt} \tag{15}$$

If roads affected real income *only* through the mechanism suggested in the model, then α_2 in equation 15 should decrease to zero and the coefficient of $\hat{\mathcal{W}}_{ht}$, β_1 should be close enough to one. Thus we can attribute the difference in α_2 in equations 14 and 15 to the effect of road connectivity that is channeled through trade.

Table 11 reports the results for equations 14 and 15. Column 1 excludes household characteristics (age, gender and education of the household head, and household size) whereas columns 2-4 include these household characteristics. The result in column 2 shows that, conditional on household characteristics, road connectivity increased household real agricultural income by about 13% between 2012 and 2015. In column 3, I regress the sufficient static measure of the welfare gain on road connectivity. Households in villages that got URRAP roads have seen about 13.4% in this sufficient statistic compared to those in villages that did not get URRAP roads. In column 4, I run the regression in equation 15. Column 4 is simply the regression in column 2 but including the sufficient static measure as a regressor. Comparison of columns 2 and 4 indeed shows that including the sufficient statistic decreased the estimated coefficient of road connectivity by about a half (from 12.8% to about 5.8%) and significantly increased the R-squared. This suggests that about 7% (12.8%-5.8%) of the increase in real agricultural revenue due to roads is attributed to the improvement in market integration and trading opportunities created by the road connectivity. This result strongly supports the evidences presented in section 3 that URRAP roads decreased urban-rural price gaps (a proxy for trading cost) and increased specialization in comparative advantage crops.

The remaining effects of 5.8%, i.e., α_2 in equation 15 captures the effects of road that are channeled through other mechanisms than trade.These may include, among other things, better access to health facilities, modern agricultural inputs such as fertilizer and extension services, etc that are made possible by road connectivity. In particular, access to improved agricultural inputs and extension services are shown to be critical to agricultural productivity of small holder farmers.

6.3 Heterogeneous effects of URRAP roads

Figures 6 and 7 plot the histogram and kernel density of the log real agricultural income and the sufficient statistic for welfare effect of road for the years before and after URRAP roads. The distributions for both quantities have shifted to the right as shown in both the histogram and the kernel density. In fact, the two-sample Kolmogorov–Smirnov test for equality of the distributions of log real revenue before and after URRAP is rejected with a p-value of 0.00, and similarly for the sufficient statistic measure of welfare. The aggregate welfare (log weighted real agricultural income, where the weights are sampling weights of households) increases by 12.6% between 2012 and 2015. Similarly, the weighted average of the sufficient statistic measure of welfare increases by 5.5% between 2012 and 2015. Even though there is sizable aggregate welfare gain and the distribution of log real agricultural income shifts to the right overall, there is considerable heterogeneity across villages and among farm households within villages.

One potential source of heterogeneity in gains from roads among households within a village is households' land holding status. Atkin (2013) argues that landless laborers are losers from decreases in trade costs because their region's comparative advantage crops, for which they have developed a taste over generations, become more expensive. As shown in section 3, prices go up in a village for comparative advantage crops following road connectivity. For land owning households, this has two opposing effects on their real income. On one hand, their nominal income goes up because the prices for their comparative advantage crops increased and more land is reallocated to these crops (see section 3). On the other hand, the increase in the prices of comparative advantage crops hurts the farmers because they are likely to have better taste for these crops (these crops have higher weights in the farmers' consumption bundle). The net effect depends on the strength of these two forces. For the landless households or households with small land size (who are likely to be net purchasers of most crops, including comparative advantage crops) the nominal income gain is likely to be nil but they are hurt by the increase in prices of local crops. Hence, their welfare is likely to decline.

In table 12, I divide farmers according to the size of their land holdings. In the first two columns, I interact a dummy variable for whether a farmer's land holding is above or below median with the road access. The result shows that farmers with above median land holding gain more by about 6%. In the last two columns, I divide farmers into quartiles of land holding size and interact the dummy for quartiles of land holding with road access. The result shows that farmers in the highest quartile of land holding gain more by about 9.6%. In summary, the

results in 12 show that most of the gains from new roads are captured by farmers with above median land holding, particularly those in the highest quartile of land holding. This shows that roads might also have important redistribution effect and might lead to increased inequality.

Another source of heterogeneity to consider is whether welfare gains from URRAP project is different across villages based the villages' prior road status. Some of the villages that got road connectivity by URRAP project had prior road connections of two types: (1) highway trunk roads that cross the villages, and (2) old subpar community roads that are particularly not passable by vehicle during rainy seasons. Regardless of variation in such prior access, however, all villages were equally considered for the URRAP project with the rationale that the villages were not benefiting enough from the highways crossing them because of a lack of access ramps, or because old community roads were substantially subpar.

In table 13, I divide the villages into two groups: those that had the above mentioned roads and those that did not, and I identify the welfare gains from URRAP separately for each group by comparing villages that got URRAP and those which did not in each group. The first two columns report the welfare effect of URRAP roads in the villages that had no prior roads while the last two columns estimate the welfare gains for the villages that had some prior roads. The results clearly show that the welfare gain from URRAP is significantly higher in villages that had no pre-existing roads, compared to villages that had some prior roads or the average welfare gain reported in table 11. In fact, the welfare gain from URRAP is statistically insignificant in villages that had pre-existing roads.

7 Conclusions and future works

In this paper I study how roads improve market integration and alter households resource allocation using a massive rural road expansion project in Ethiopia as a source of variation. The paper shows that, high trade costs can distort resource allocation by forcing households to resort to subsistence farming, instead of allocating resources based on market forces. The paper also shows that, falling trade costs due to construction of new roads reduces this distortion by breaking the link between household resource allocation and consumption.

Another key result in this paper is that, in the context of multi-sector Ricardian trade model with heterogeneous land, one can derive the sufficient statistic for the welfare effects of road from household's land allocation only, i.e., without any information on trade flows. The paper

uses this result to quantify how much of a welfare gain from roads estimated in the reduced form is attributed to improvement in trade costs. Out of 13% welfare gain from roads, about half is attributed to improvement in trade opportunities following the construction of roads. While the paper is silent on what other mechanisms explain the remaining half, several other mechanisms through which roads can improve welfare can be imagined. In particular, future studies may explore how road connectivity might affect accessibility of health and agricultural inputs, such as fertilizer and extension services, which are critical for agricultural productivity. Anecdotal evidences show that construction of rural roads has significantly improved maternal and child health by facilitating faster access to health stations using modern transport.

Currently, I am working on calculating change in market access of each village due to the URRAP project. This requires estimation of travel time along the least-cost path from a village to all the potential destinations. Compared to the binary treatment dummy, the market access approach enables me to account for the general equilibrium spillovers in road connectivity. If such spillovers are significant, the current estimate for the welfare gain from road would be biased downwards.

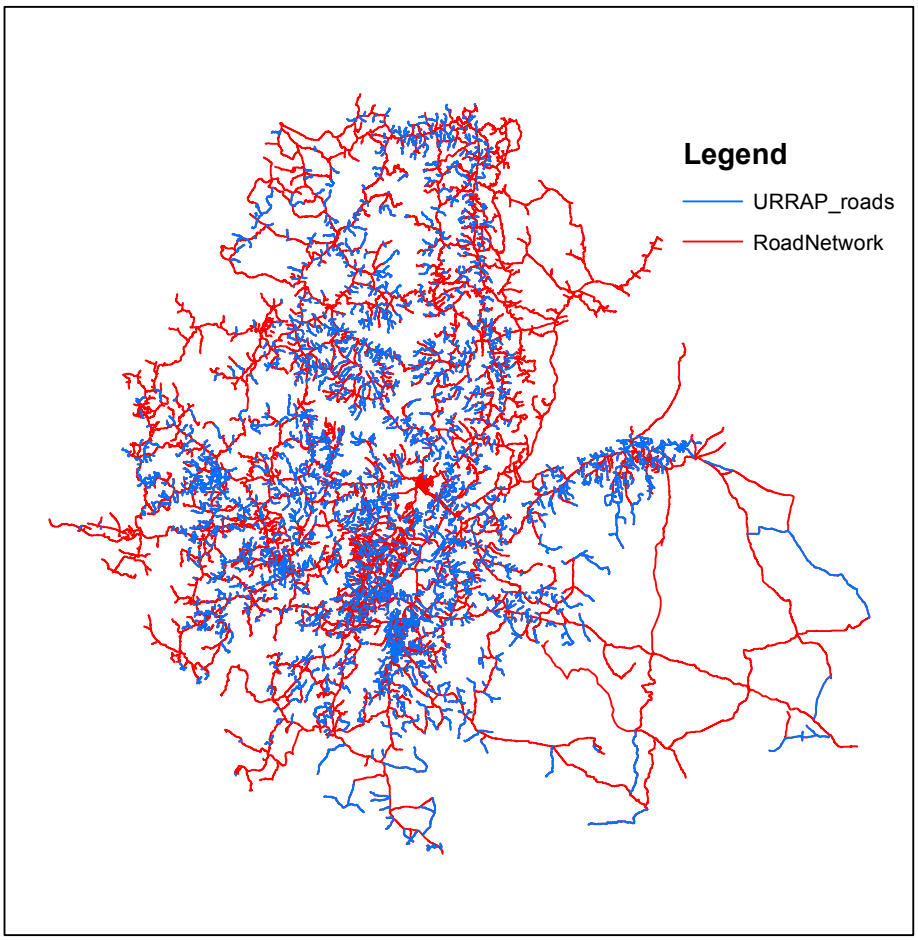

Figure 1: Rural road expansion under URRAP

Figure 2: Construction sites of selected URRAP roads (pictures are taken from Oromia Roads Authority).

[d]

Figure 3: Completed URRAP roads (pictures are taken from Oromia Roads Authority).

Figure 4: URRAP roads in use (pictures are taken from Oromia Roads Authority and other websites).

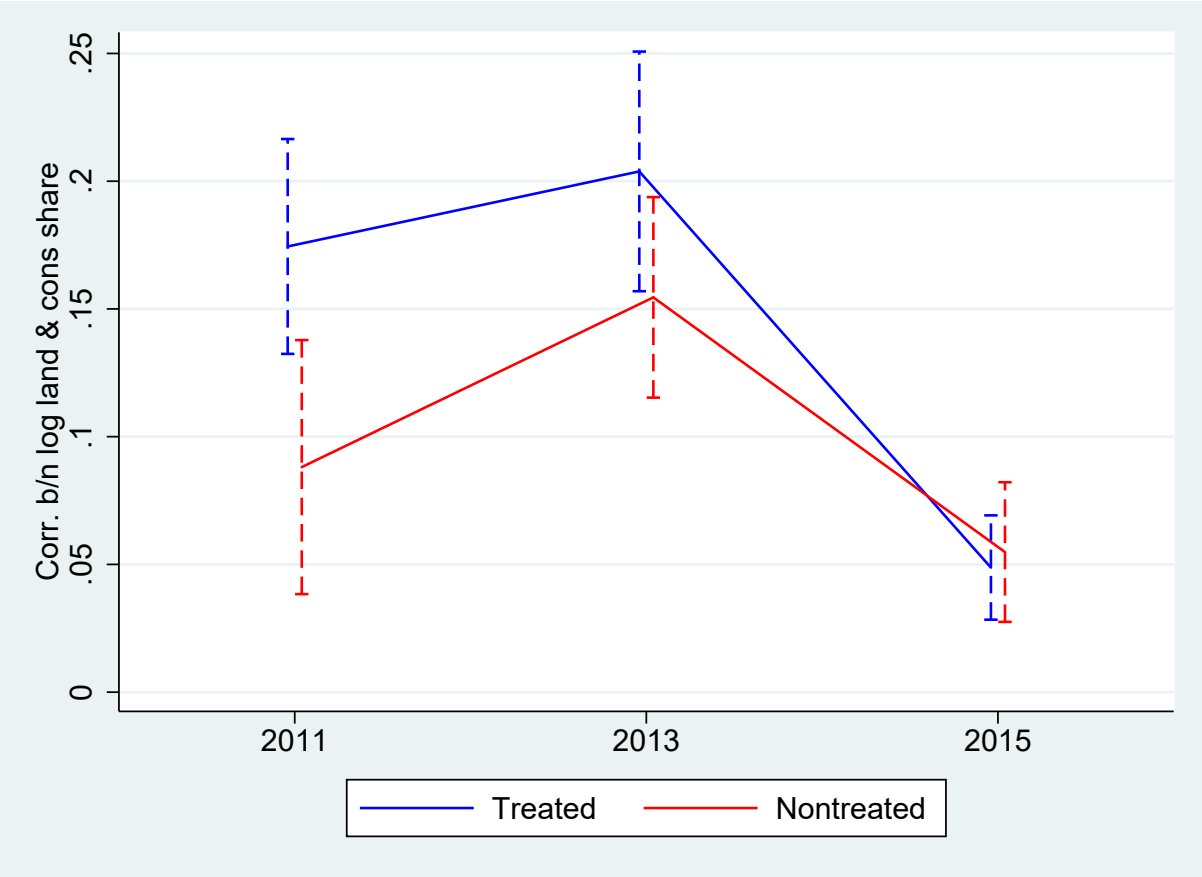

Figure 5: URRAP and the correlation between land and budget share of crops

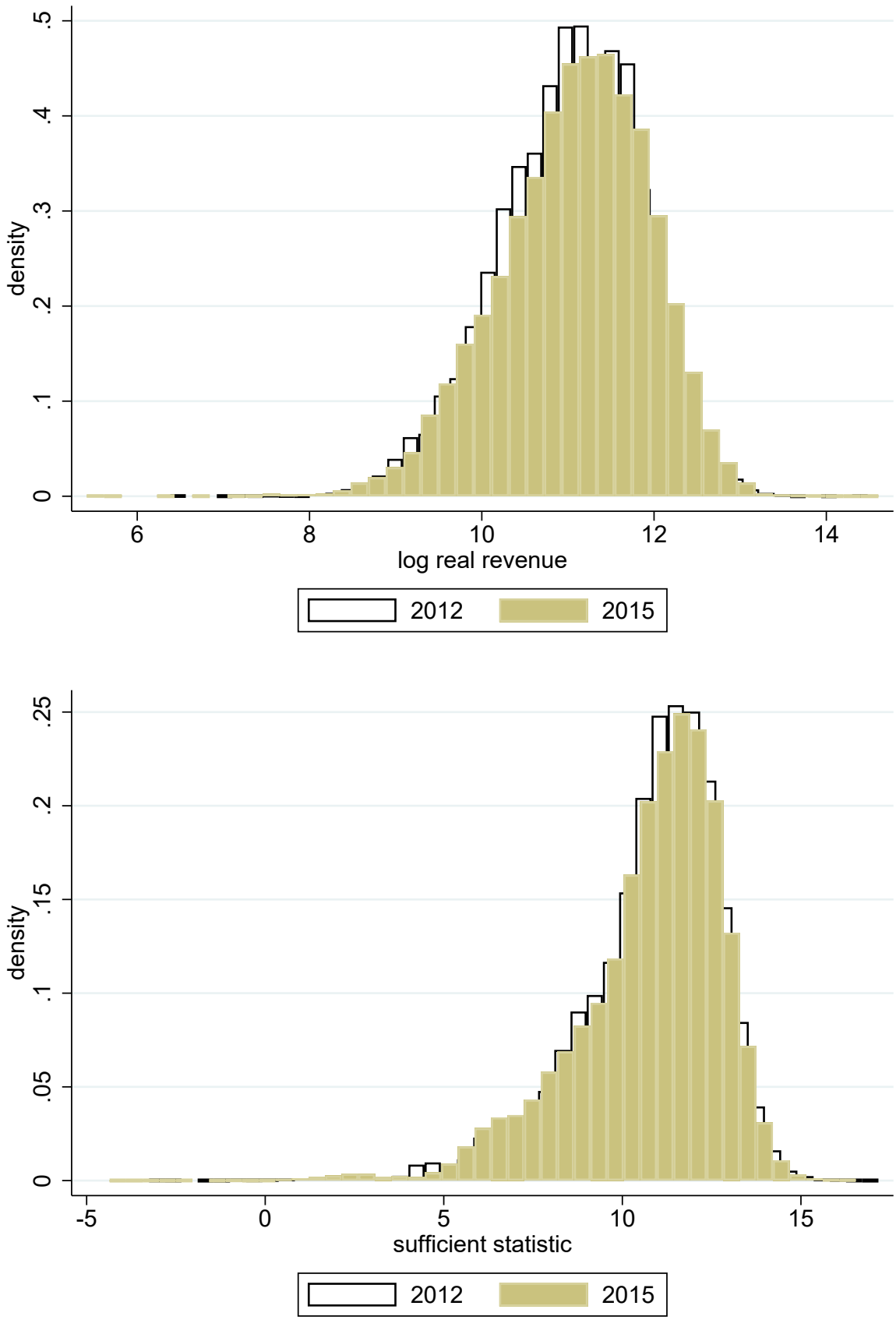

Figure 6: Histogram of log real agricultural income and the sufficient statistic measure $\hat{\mathcal{W}}_{ht}$ in the years 2012 and 2015.

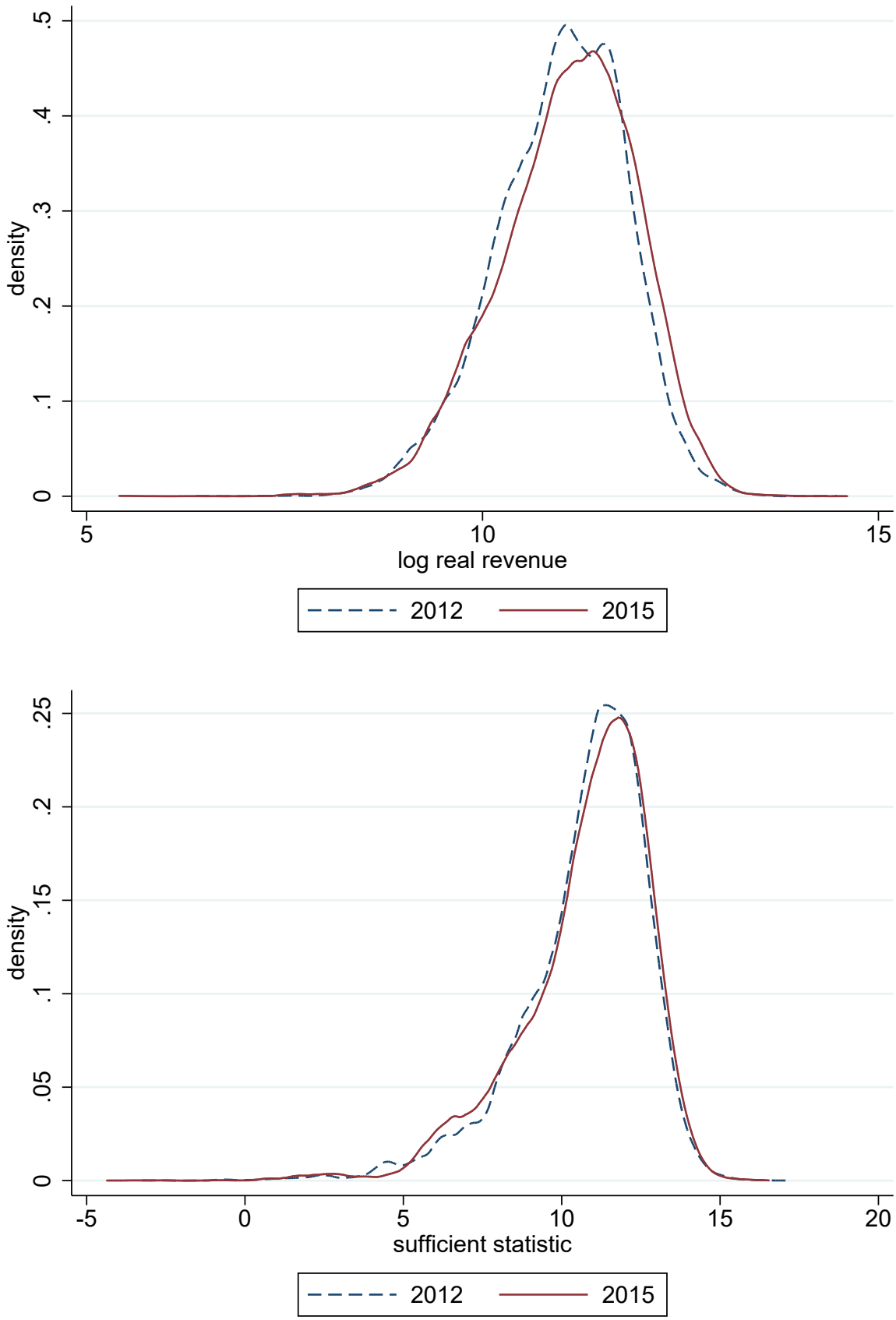

Figure 7: Kernel density of log real agricultural income and the sufficient statistic measure $\hat{\mathcal{W}}_{ht}$ in the years 2012 and 2015.

Table 1: Transport modes to market

Mode of transport	2011	2013	2015
On Foot	43.6	49.9	41.9
Pack Animals	45.8	41.80	43.9
Own Bicycle or Oxcart	6.74	1.76	4.78
Vehicle	2.34	2.53	5.69
Others	1.43	3.9	3.62
Ad valorem trade cost vehicle (mean)	11.37	6	6.4
Ad valorem trade cost vehicle (median)	6.49	4.3	3
distance to all-weather road (median KM)	10	10	8.5
distance to population centers (median KM)	30	30	30
distance to district (woreda) town (median KM)	17	17	17
distance to nearest weekly market (median KM)	12	8	8
Median transport fare to district capital (real Birr/KM)	0.7	0.597	0.523

Notes: This table is based on ESS data.

Table 2: Fraction of households who consume a positive amount of a crop, and those who consume and do not produce

	Small towns		Rural villages	
	Cons.	Cons.& Not prod.	Cons.	Cons. &Not prod.
Teff	0.719	0.640	0.349	0.114
Maize	0.438	0.382	0.593	0.232
Wheat	0.442	0.390	0.401	0.202
Enset	0.145	0.092	0.184	0.057
Barley	0.177	0.145	0.198	0.049
Sorghum	0.326	0.276	0.462	0.127
Millet	0.049	0.042	0.112	0.023
Field pea	0.432	0.399	0.232	0.151
Lentils	0.356	0.351	0.134	0.110
Linseed	0.044	0.042	0.074	0.043
Haricot beans	0.095	0.084	0.179	0.079
Horse beans	0.466	0.433	0.401	0.242
Onions	0.878	0.872	0.710	0.683
Potatoes	0.586	0.573	0.285	0.231
Tomatoes	0.660	0.656	0.350	0.333
Banana	0.273	0.259	0.161	0.100
Coffee	0.773	0.736	0.709	0.557
Total	0.560	0.536	0.455	0.366

Notes: This table shows fraction of households consuming a given crop and the source (own production or purchase) of the consumption. I present the statistics for rural areas and small towns separately to emphasize the potential role of access to market. Small towns are towns with a population of below 10,000. For each location groups, the table reports the fraction of households who consumed a specific crop and the fraction that consumed the crop and not produced it (i.e., the fraction who consumed a crop from purchase). The statistics is an average across the years 2011, 2013 and 2015.

Table 3: Crop utilization by farm households

	Consumed	Kept for seed	marketed
Barley	68.18	19.07	7.58
Maize	80.46	7.11	8.56
Millet	78.29	10.17	5.61
Oats	66.72	19.14	9.83
Rice	81.64	14.07	4.29
Sorghum	80.44	8.81	6.50
Teff	58.66	13.34	22.46
Wheat	62.35	17.76	14.76
Mung bean	20.84	12.11	62.76
Cassava	50.00	35.00	15.00
Chick pea	69.82	14.90	12.01
Haricot beans	85.28	7.97	5.76
Horse beans	71.07	14.02	11.48
Lentils	37.98	20.05	40.65
Field pea	63.88	18.11	13.97
Vetch	60.28	16.99	18.88
Gibto	29.23	26.31	43.69
Soya beans	14.59	13.54	69.20
Red kidney beans	75.43	8.78	14.19
Total	70.80	12.20	12.74

Notes: This table shows crop utilization by households. The first column shows the percent of production consumed within the household. Column 2 shows the percent kept for seed (input for next planting season), and column 3 shows the percent sold.

Table 4: 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
Getting road access ($Post_t * Road_v$)	-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 5: Rural roads and the link between local prices and local yield: the dependent variable is crop-village level prices in 2011 and 2015.

	(1)	(2)	(3)
Log yield ($\ln A_v^k$)	-0.036***	-0.036***	-0.027***
	(0.003)	(0.003)	(0.003)
Getting road access ($Post_t * Road_v$)		0.023	-0.083***
		(0.019)	(0.023)
Log yield * Getting road access ($\ln A_v^k * Post_t * Road_v$)			0.009***
			(0.003)
N	119117	119117	119117
R^2	0.752	0.752	0.776

Notes: Standard errors are clustered at village level. The regression includes 427 villages, and 20 crops. All regressions include village, crop, and year fixed effects. The regression includes log rainfall as a control. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6: URRAP and land reallocation towards comparative advantage (CA) and absolute advantage (AA) crops: the dependent variable is land share of crop within a village

	GAEZ CA crop				GAEZ AA crop
	1	2	3	4	5
Road access	-0.011**	-0.011**	-0.007	-0.006	-0.012**
	(0.004)	(0.005)	(0.005)	(0.006)	(0.006)
Top 30%	0.069***				
	(0.005)				
Road access*Top 30%	0.022**				
	(0.010)				
Top 40%		0.073***			
		(0.004)			
Road access*Top 40%		0.018*			
		(0.009)			
Top 50%			0.071***		
			(0.004)		
Road access*Top 50%			0.008		
			(0.009)		
Top 60%				0.067***	
				(0.004)	
Road access*Top 60%				0.004	
				(0.008)	
AA crop					0.074***
					(0.004)
Road access*AA crop					0.017*
					(0.009)
N	25576	25576	25576	25576	25576
R^2	0.112	0.115	0.114	0.110	0.112
Mean land share (2011)	0.144	0.144	0.144	0.144	0.144
Mean land share of CA or AA-crops (2011)	0.196	0.188	0.180	0.170	0.169

Notes: Standard errors are clustered at village level. I use GAEZ yield measure to define a village's absolute and comparative advantage crops. Columns 1-4 are based on ranking of crops within a village based on the village's productivity in a crop relative to the national average of the crop. For instance, in column 1 a crop is considered as a CA crop if it is ranked in the top 30%, within a village, based on its productivity relative to the national average. In columns 2-4 the threshold is extended to the top 40%, 50% and 60% respectively. In column 5, a village's Absolute Advantage (AA) crops are defined as crops in which the village productivity is above national average. All regressions include year and village fixed effects, year × CA crop or year × AA crop interactions, log rainfall, and log of total area of land cultivated in each village. The estimation is based on 20 crops for which GAEZ yield measure is available: Bananas, Barley, Cabbage, Carrot, Chick Peas, Citrus, Coffee Field Peas, Maize, Millet, Onion, Potatoes, Rape Seed, Rice, Sorghum, Soya Beans, Sunflower, Sweet Potatoes, Tomatoes and Wheat. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7: URRAP and household consumption

	Log nominal consumption expenditure		Log real consumption expenditure	
	Total	Food	Total	Food
Road Access	0.122*	0.138*	0.130*	0.146*
	(0.069)	(0.076)	(0.074)	(0.081)
N	7971	7971	7971	7971
R^2	0.460	0.399	0.403	0.351

Note: This table is based on ESS data. Standard errors are clustered at village level. Observations are weighted by the household sampling weight. Household consumption expenditures are normalized by adult equivalent family size. Real expenditures are calculated using the price index obtained from the model. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 8: Basic test of separability: land allocation and budget share of crops (the dependent variable is *Land share of a crop*)

	All observations		Conditional on consumed or produced	
	2011	2015	2011	2015
Consumption Share	0.471***	0.198***	0.356***	0.070***
	(0.025)	(0.011)	(0.030)	(0.026)
lnYield	0.003***	0.004***	0.006***	0.004***
	(0.000)	(0.000)	(0.001)	(0.001)
N	71048	69933	24179	24399
R^2	0.280	0.171	0.179	0.114

Notes: Standard errors are clustered at village level. Observations are weighted by the household sampling weight. In the first two columns, household-crop is in the data even if the household neither produces nor consumes some of the crops. In the last two columns, household-crop is in the data if the household either produced or consumed the specific crop during the specific year. *Yield* is FAO/GAEZ measure of crop yield which varies at district level. All regressions include household fixed effects, and columns 1 and 2 also include crop fixed effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 9: Separability and proximity to markets and roads (the dependent variable is *Land share of a crop*)

	D=1 if distance is above median		D= log(distance)	
	to towns	to all-weather road	to towns	to all-weather road
Cons. Share	0.259***	0.246***	0.033	0.177***
	(0.018)	(0.020)	(0.041)	(0.029)
Distance	-0.006***	-0.006***	-0.005***	-0.003***
	(0.002)	(0.002)	(0.001)	(0.001)
Distance*Cons. Share	0.101***	0.117***	0.080***	0.052***
	(0.031)	(0.029)	(0.012)	(0.011)
N	265632	265632	265632	265632
R^2	0.208	0.209	0.151	0.212

Notes: Standard errors are clustered at village level. Observations are weighted by the household sampling weight. All regressions include the the control variables in table 8, household fixed effects, and year fixed effects. Towns are population centers with over 20,000 population in 2007 census. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 10: Separability and URRAP: the dependent variable is *Land share of a crop*

	No village FE		With Village FE	
	All observations	consumed or produced	All observations	consumed or produced
Cons. Share	0.370***	0.300***	0.372***	0.282***
	(0.032)	(0.035)	(0.032)	(0.033)
$\text{Road}_v * Cons.Share$	0.124***	0.125***	0.121***	0.136***
	(0.039)	(0.039)	(0.038)	(0.041)
Post*Cons. Share	-0.169***	-0.283***	-0.174***	-0.282***
	(0.045)	(0.049)	(0.044)	(0.048)
$\text{Road}_v * Post * Cons.Share$	-0.120**	-0.118**	-0.111**	-0.116**
	(0.055)	(0.052)	(0.052)	(0.052)
N	137893	47457	138312	47548
R^2	0.190	0.042	0.202	0.069

Notes: Standard errors are clustered at village level. Observations are weighted by the household sampling weight. In columns 2 and 4, household-crop is in the data if the household either produced or consumed the specific crop during the year. All regressions include the control variables in table 8 and year fixed effects. In columns 1 and 3, I also include crop fixed effects. In columns 1 and 2 I include log distance to population centers and log distance to roads in 2011 (before the onset of URRAP) to account for the potential endogeneity of road placement. In columns 3 and 4, I address for potential endogeneity of road placement by including village fixed effects.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 11: Welfare effects of URRAP

	lnRealRevenue		$\hat{\mathcal{W}}$	lnRealRevenue
	(1)	(2)	(3)	(4)
Road Access	0.118***	0.128***	0.134***	0.058*
	(0.045)	(0.043)	(0.046)	(0.033)
$\hat{\mathcal{W}}$				0.521***
				(0.011)
N	70257	70249	70249	70249
R^2	0.411	0.458	0.372	0.623

Notes: Standard errors are clustered at village level. Observations are weighted by the household sampling weight. $\hat{\mathcal{W}}$ is the sufficient statics for the welfare effect effect of road. All regressions include village and year fixed effects, and log rainfall. In column 1, I exclude household characteristics (age, gender, and education of the household head and household size). Column 2-4 include all these household characteristics. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 12: Welfare effects of URRAP: heterogeneity by land holding size

	Above or below median		Quartiles of land holding size	
	lnRealRevenue	$\hat{\mathcal{W}}$	lnRealRevenue	$\hat{\mathcal{W}}$
Road access	0.049	0.066	0.038	0.049
	(0.033)	(0.055)	(0.037)	(0.066)
Land above median*Road access	0.061***	0.094**		
	(0.019)	(0.041)		
2nd Quartile*Road access			0.024	0.025
			(0.023)	(0.055)
3rd Quartile*Road access			0.050*	0.100*
			(0.026)	(0.058)
4th Quartile*Road access			0.096***	0.100*
			(0.028)	(0.059)
N	70249	70248	70249	70248
R^2	0.943	0.429	0.943	0.469

Note: Standard errors are clustered at village level. Observations are weighted by the household sampling weight. $\hat{\mathcal{W}}$ is the sufficient statics for the welfare effect effect of road. All regressions include village and year fixed effects, and all the household and village characteristics listed in table 11. The first two columns also include a dummy for above or below median land holding while the last two columns include the dummies for the quartiles of land holding size.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 13: Welfare effects of URRAP: heterogeneity by prior road status

	Villages with no prior roads		Villages with some prior roads	
	lnRealRevenue	$\hat{\mathcal{W}}$	lnRealRevenue	$\hat{\mathcal{W}}$
Road Access	0.203***	0.191**	0.055	0.073
	(0.068)	(0.082)	(0.055)	(0.059)
N	31291	31291	38958	38957
R^2	0.514	0.421	0.464	0.351

Note: Standard errors are clustered at village level. Observations are weighted by the household sampling weight. $\hat{\mathcal{W}}$ is the sufficient statics for the welfare effect effect of road. All regressions include village and year fixed effects, and log rainfall. The first two columns estimate the welfare effects of URRAP for villages that had no prior roads while the last two columns estimate the welfare effects of URRAP roads in villages that had some prior roads (villages that were either crossed by the trunk roads or those that had subpar seasonal roads). All regressions include the control variables in table 11.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix A

A.1 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 productivity conditional on the land being used for crop k, i.e., $z_i^k(\omega)|\omega \in \Omega_i^k$, which I denote as $G_i^k(t)$. This derivation of conditional distribution of land quality is similar to Sotelo (2018):

$$
\begin{aligned}
G_i^k(t) &= \mathcal{P}\big[z_i^k(\omega) < t | p_i^k z_i^k(\omega) = \max_l p_i^l z_i^l(\omega)\big] \\
&= \frac{\mathcal{P}\big[z_i^k(\omega) < t \wedge p_i^k z_i^k(\omega) = \max_l p_i^l z_i^l(\omega)\big]}{\mathcal{P}\big[p_i^k z_i^k(\omega) = \max_l p_i^l z_i^l(\omega)\big]} \\
&= \frac{1}{\eta_i^k}\mathcal{P}\big[z_i^k(\omega) < t \wedge p_i^l z_i^l(\omega) < p_i^k z_i^k(\omega), \quad \forall l\big] \\
&= \frac{1}{\eta_i^k}\mathcal{P}\big[\frac{p_i^l}{p_i^k} z_i^l(\omega) < z_i^k(\omega) < t, \quad \forall l\big] \\
&= \frac{1}{\eta_i^k}\int_0^t \Pi_{l\neq k}\mathcal{P}\big[\frac{p_i^l}{p_i^k} z_i^l(\omega) < v\big] f_i^k(v)dv
\end{aligned}
$$

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

$$
\begin{aligned}
G_i^k(t) &= \frac{1}{\eta_i^k}\int_0^t \Pi_{l\neq k}\exp\big((-A_i^l)^\theta(\frac{p_i^k}{p_i^l}v)^{-\theta}\big) \quad \theta(A_i^k)^\theta v^{-\theta-1}\exp(-(A_i^k)^\theta v^{-\theta})dv \\
&= \frac{1}{\eta_i^k}\int_0^t \exp\big(-(p_i^k v)^{-\theta}\sum_{l\neq k}(A_i^l p_i^l)^\theta\big)\exp(-(A_i^k)^\theta v^{-\theta} \quad \theta(A_i^k)^\theta v^{-\theta-1})dv \\
&= \frac{1}{\eta_i^k}\int_0^t \exp\big(-(p_i^k v)^{-\theta}\sum_{l}(A_i^l p_i^l)^\theta\big) \quad \theta(A_i^k)^\theta v^{-\theta-1}dv \\
&= \int_0^t \exp\big(-(p_i^k v)^{-\theta}\Phi_i^\theta\big) \quad \theta\Phi_i^\theta(p_i^k)^{-\theta}v^{-\theta-1}dv \\
&= \exp\Big(-(\frac{\Phi_i}{p_i^k})^\theta t^{-\theta}\Big)
\end{aligned}
$$

Thus the distribution of the set of farmer i's plots which are covered by crop k is a Fréchet with the parameters $\frac{\Phi_i}{p_i^k}$ 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 crop k, is given by

$r(\omega) = \max_k\{p_i^k z_i^k(\omega)\}$. Thus the conditional distribution of rental rate $r(\omega)|\omega \in \Omega_i^k$ is Fréchet with parameters Φ_i 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.

A.2 Appendix Tables

Table A1: URRAP and land reallocation: quartiles of relative productivity - the dependent variable is land share of crop within a village

	1	2
Road access	-0.008	-0.005
	(0.005)	(0.005)
2nd Quartile	0.025***	0.027***
	(0.005)	(0.005)
3rd Quartile	0.073***	0.066***
	(0.006)	(0.006)
4th Quartile	0.094***	0.073***
	(0.006)	(0.005)
Road access*2nd Quartile	0.005	0.002
	(0.007)	(0.007)
Road access* 3rd Quartile	-0.005	-0.003
	(0.008)	(0.008)
Road access*4th Quartile	0.016**	0.018**
	(0.008)	(0.008)
N	25576	23037
R^2	0.117	0.045
Mean land share (2011)	0.14	0.13

Notes: Standard errors are clustered at village level. I use GAEZ yield measure to calculate each village's relative productivity in each crop compared to the national average yield for the crop. I then group crops into quartiles, within each village, based on the relative yields. The 1st quartile includes crops that have the lowest relative productivity within a village and the 4th quartile includes crops with the highest relative productivity. In the first column I include all villages. In the second column, estimation is restricted to villages that produce 5 or more crops. All regressions include village and year fixed effects. The estimation is based on 20 crops for which GAEZ yield measure is available listed in table 6.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A2: URRAP and land reallocation towards comparative advantage (CA) and absolute advantage (AA) crops – villages that produce four or more crops: the dependent variable is land share of crop within a village

	GAEZ CA-crops				GAEZ AA-crops
	1	2	3	4	5
Road access	-0.012***	-0.013***	-0.008*	-0.006	-0.010**
	(0.004)	(0.004)	(0.005)	(0.006)	(0.005)
Top 30 %	0.050***				
	(0.005)				
Road access*Top 30	0.032***				
	(0.010)				
Top 40%		0.055***			
		(0.004)			
Road access*Top 40		0.029***			
		(0.009)			
Top 50%			0.056***		
			(0.004)		
Road access*Top 50			0.016*		
			(0.008)		
Top 60%				0.057***	
				(0.004)	
Road access*Top 60				0.009	
				(0.008)	
AA crop					0.059***
					(0.004)
Road access*AA crop					0.018**
					(0.008)
N	23037	23037	23037	23037	23037
R^2	0.039	0.042	0.042	0.041	0.041
Mean land share	0.130	0.130	0.130	0.130	0.134
Mean land share of CA or AA crops	0.163	0.160	0.156	0.150	0.150

Notes: Standard errors are clustered at village level. I use GAEZ yield measure to define a village's absolute and comparative advantage crops. Columns 1-4 are based on ranking of crops within a village based on the village's productivity in a crop relative to the national average of the crop. For instance, in column 1 a crop is considered as a CA crop if it is ranked in the top 30%, within a village, based on its productivity relative to the national average. In columns 2-4 the threshold is extended to the top 40%, 50% and 60% respectively. In column 5, a village's Absolute Advantage (AA) crops are defined as crops in which the village productivity is above national average. All regressions include year and village fixed effects, year × CA crop or year × AA crop interactions, log rainfall, and log of total area of land cultivated in each village. The estimation is based on 20 crops for which GAEZ yield measure is available: Bananas, Barley, Cabbage, Carrot, Chick Peas, Citrus, Coffee Field Peas, Maize, Millet, Onion, Potatoes, Rape Seed, Rice, Sorghum, Soya Beans, Sunflower, Sweet Potatoes, Tomatoes and Wheat. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A3: Basic test of separability: labor allocation and budget share of crops (the dependent variable is *labor share of a crop*)

	All observations		Conditional on consumed or produced	
	2011	2015	2011	2015
Cons. Share	0.456***	0.191***	0.349***	0.075***
	(0.025)	(0.021)	(0.029)	(0.026)
lnYield	0.003***	0.004***	0.006***	0.004***
	(0.000)	(0.001)	(0.002)	(0.001)
N	71048	69932	24094	24306
R^2	0.288	0.175	0.183	0.114

Notes: Standard errors are clustered at village level. Observations are weighted by the household sampling weight. In columns 3 and 4, household-crop is in the data if the household either produced or consumed the specific crop during the year. *Yield* is FAO/GAEZ measure of crop yield which varies at district level. *Price* is farm-gate crop price which varies at village level. All regressions include household fixed effects. Columns 1 and 2 also include crop fixed effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A4: Separability and proximity to markets and roads (the dependent variable is *Labor share of a crop*)

	D=1 if distance is above median		D= log(distance)	
	to towns	to all-weather road	to towns	to all-weather road
Cons. Share	0.251***	0.240***	0.022	0.176***
	(0.019)	(0.022)	(0.037)	(0.031)
Distance	-0.001	0.002	0.001**	0.001*
	(0.002)	(0.002)	(0.001)	(0.001)
Distance*Cons. Share	0.092***	0.106***	0.072***	0.048***
	(0.031)	(0.029)	(0.011)	(0.011)
N	265629	265629	265629	265629
R^2	0.194	0.195	0.200	0.198

Notes: Standard errors are clustered at village level. Observations are weighted by the household sampling weight. All regressions include the the control variables in table A3, and year fixed effects. Towns are population centers with over 20,000 population in 2007 census. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A5: Separability and URRAP: labor allocation (the dependent variable is *labor share of a crop*)

	No village FE		With Village FE	
	All observations	Consumed or produced	All observations	Consumed or produced
Cons. Share	0.429***	0.286***	0.346***	0.266***
	(0.032)	(0.034)	(0.032)	(0.033)
$\text{Road}_v * Cons.Share$	0.149***	0.141***	0.146***	0.155***
	(0.038)	(0.038)	(0.038)	(0.039)
Post*Cons. Share	-0.219***	-0.261***	-0.151***	-0.261***
	(0.048)	(0.049)	(0.045)	(0.048)
$\text{Road}_v * Post * Cons.Share$	-0.154***	-0.138***	-0.140***	-0.135***
	(0.058)	(0.052)	(0.054)	(0.051)
N	138311	47375	138311	47375
R^2	0.119	0.042	0.200	0.069

Note: Standard errors are clustered at village level. Observations are weighted by the household sampling weight. In columns 2 and 4, household-crop is in the data if the household either produced or consumed the specific crop during the year. All regressions include the control variables in table A3, and year fixed effect. Columns 1 and 3 include crop fixed effects. In columns 1 and 2 I include log distance to population centers, and log distance to roads in 2011 (before the onset of URRAP) to account for the potential endogeneity of road placement. In columns 3 and 4, I address for potential endogeneity of road placement by including village fixed effects.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix B The classic separability test

In this section I explain the results for the classic test of separability introduced by Benjamin (1992). I run similar specifications as Benjamin (1992) and LaFave and Thomas (2016) to compare my results with theirs.

Table B1 reports the estimation results. In my data labor is measured in hours of work, and I observe hours spent on *planting* and *harvesting* separately. I report results for *total* labor demand (harvesting *plus* planting hours), and separately for planting and harvesting labor. The result shows an unambiguous rejection of separability – household demographic composition significantly affects household labor demand. This result is robust across specifications that include household fixed effects and those that do not, and across planting and harvesting labor. Panel A includes the effects of the number of *males* of different age groups. Higher number of males of any age group is positively associated with on-farm labor demand throughout the specifications, with the effect peaking at the age group 35-49 for the preferred specification (those with household fixed effects). Panel B reports the effect of number of females of different age groups on labor demand. Clearly the number of female members of a household is not significantly associated with farm labor demand regardless of their age groups. This is less of a surprise for those who are familiar with agriculture in least developed countries such as Ethiopia. Farming in these part of the world is extremely physical, and women participation is limited to less physical activities such as weeding. Also important is the traditional division of labor where men work in the fields and women stay at home taking care of children and household activities such as cooking and cleaning.

Panel C reports the joint significance test of the coefficients for different age and sex groups. Both the F-statistics and the p-values are reported. Consistent with the statistical significance of the individual coefficients we observe that the coefficients for male members of different age groups are jointly statistically significant across all the specifications while the coefficients for females is jointly statistically significant only in the specifications without the household fixed effects and in the labor demand for planting (women are more likely to take part in planting activities such as weeding). Overall, the demographic variables are jointly statistically significant as shown by the F-statistics and the p-values of *all* age and sex groups, and in particular the joint significance of the *prime-age* groups (ages 15-64).

Table B1: The effect of household composition on farm labor demand: labor demand is measured as log-hours

	Pooled		Household Fixed effect		
	Total	Total	Total	Harvesting	Planting
	(1)	(2)	(3)	(4)	(5)
A. Number of Males					
age0_14	0.349***	-	0.136***	0.064	0.144***
	(0.025)		(0.036)	(0.043)	(0.044)
age15_19	0.275***	0.483*	0.200***	0.176**	0.253***
	(0.048)	(0.248)	(0.059)	(0.069)	(0.066)
age20_34	0.561***	0.999***	0.300***	0.244***	0.338***
	(0.050)	(0.22)	(0.058)	(0.068)	(0.063)
age35_49	0.691***	1.205***	0.295***	0.261**	0.338***
	(0.075)	(0.311)	(0.095)	(0.103)	(0.102)
age50_64	0.840***	2.305***	0.182	0.242*	0.156
	(0.084)	(0.327)	(0.111)	(0.127)	(0.122)
age65_above	0.413***	0.977***	0.087*	0.087	0.098*
	(0.038)	(0.140)	(0.047)	(0.054)	(0.053)
B. Number of females					
age0_14	0.286***	-0.280	0.054	0.020	0.088**
	(0.027)	(0.173)	(0.038)	(0.044)	(0.043)
age15_19	0.118**	-0.218	0.014	0.011	0.043
	(0.052)	(0.249)	(0.054)	(0.058)	(0.061)
age20_34	0.033	-0.478*	0.017	0.032	0.041
	(0.061)	(0.266)	(0.065)	(0.070)	(0.072)
age35_49	0.188**	0.198	0.120	0.065	0.151
	(0.081)	(0.297)	(0.090)	(0.100)	(0.097)
age50_64	0.622***	0.984***	0.059	0.189	0.023
	(0.086)	(0.265)	(0.117)	(0.133)	(0.123)
age65_above	0.032	-0.168	-0.052	-0.031	-0.046
	(0.042)	(0.152)	(0.046)	(0.052)	(0.052)
Log household size		1.835***			
		(0.074)			
C. Joint tests of significance					
All groups	62.46***	14.54***	3.81***	1.78**	4.31***
	(0.000)	(0.000)	(0.000)	(0.046)	(0.000)
Males	82.37***	21.31***	5.80***	2.84***	6.31***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
Females	26.55***	4.85***	1.64	0.57	2.23**
	(0.000)	(0.000)	(0.132)	(0.753)	(0.037)
Prime age	50.13***	12.42***	4.01***	2.45**	4.88***
	(0.000)	(0.000)	(0.000)	(0.012)	(0.000)
N	10353	10349	10264	10264	10264
R^2	0.354	0.380	0.864	0.820	0.830

Standard errors are clustered at household level. All regressions include Zone-Year fixed effects. The first three columns use the sum of planting and harvesting labor as dependent variable. Column 2 uses household size and shares of age groups in the household as regressors (see Benjamin (1992), and LaFave and Thomas (2016)). Prime age is defined as ages 15-64.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

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.

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

Arkolakis, C., Costinot, A., and Rodríguez-Clare, A. (2012). New trade models, same old gains? *American Economic Review*, 102(1):94–130.

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

Atkin, D. (2013). Trade, tastes, and nutrition in india. *American Economic Review*, 103:1629–63.

Behrman, J. R. and Deolalikar, A. B. (1989). Is Variety the Spice of Life? Implications for Calorie Intake. *The Review of Economics and Statistics*, 71(4):666–672.

Benjamin, D. (1992). Household composition, labor markets, and labor demand: Testing for separation in agricultural household models. *Econometrica*, 60:287–322.

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.

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 .

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*.

Hummels, D. (2007). Transportation costs and international trade in the second era of globalization. *Journal of Economic Perspectives*, 21:131–154.

LaFave, D. and Thomas, D. (2016). Farms, families, and markets:new evidence on completeness of markets in agricultural settings. *Econometrica*, 84:1917–1960.

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. (2018). Domestic trade frictions and agriculture. *Journal of Political Economy*, Forthcoming.