Market Integration and Separability of Production and Consumption Decisions in Farm Households*

Hundanol A. Kebede †

University of Virginia

November 2, 2020

Abstract

Separability – whether farm household's production and consumption decisions can be separately analyzed – has been a subject of significant policy and academic debate. The existing empirical tests looked at the link between on-farm labor demand and household demographic characteristics to examine whether separability holds. One problem with this approach is that on-farm labor demand is likely to be poorly measured in the context of self-employing agricultural households. In this paper, I suggest an alternative and more straightforward empirical test of separability and analyze how it is linked to market integration/trade costs. The new test looks at the link between household land allocation across different crops and the household's tastes for the crops estimated from their preference functions. I find that households' tastes for crops significantly affect their land allocation across crops. I also show that this effect significantly decreases with improvement in market integration due to construction of new rural roads.

Keywords: Agricultural household models, Farm households, Rural development, Rural roads, Separability, Trade costs.
JEL Codes: F11, H54, O12, 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, WADES, PAC DEv, and MWIEDC conferences

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

1 Introduction

Agricultural sector in developing countries is dominated by smallholder farmers, who produce not only for the market but also for their own consumption needs; and use both purchased inputs such as fertilizers and non-purchased inputs such as family labor. Whether households' production decision is separable from their consumption preferences, and whether these two decisions can be analyzed separately/recursively has been a subject of debate among policy practitioners and academicians (Singh et al. 1986). There are a number of reasons why separability is an important subject.

First, separability determines how households respond to policy interventions, such as those that target increase in agricultural exports. For instance, governments' attempts to boost the production and export of cash crops is unlikely to materialize if households had to self-produce most of the crops they need for consumption – because that would make reallocation of land to cash crops more difficult. Second, separability has implications for efficient utilization of resources such as land and labor. For example, if separability holds, the household's crop production choices would be dictated by market prices and productivity of their land in the potential crops and the household would obtain higher return from its resources, compared to the situation where the household's consumption preferences dictate production decisions. Third, whether household production decisions can be analyzed independent of their consumption decisions is important in academic and policy research. Agricultural Household Models (AHMs) in which production and consumption decisions are made jointly lack tractability, which limits their widespread use in empirical research.

The vast majority of empirical studies invoke the assumption of complete markets to obtain a tractable AHMs (Singh et al. 1986, Taylor and Adelman 2003). Under this assumption, household decisions can be modeled as *recursive*. In the first stage, the household makes its production decisions independent of its consumption preferences, and in the second stage, the household makes its utility maximization/consumption decisions given its farm profits from the first stage. For instance, if a farmer can always buy a rice at a fixed market price and can hire labor at a fixed market wage, the quantity of rice produced by the farmer would be independent of how much rice it wants to consume and the amount of labor used to produce rice would be independent of the amount of family labor supply. There are at least two issues with the assumption of complete market. First, while the assumption significantly simplifies modeling of the farm household behaviour, it is unlikely to reflect the reality under which the farmers operate. For instance, land (rental) markets are very limited in many countries and functional labor markets too are far from reality.[1] Second, even in the case of more functioning markets, such as crop markets, high trade costs due to poor infrastructure makes these markets too thin for the farmers to rely on. As a consequence, farmers may choose to self-produce most of the crops they need for consumption, instead of specializing in few crops and sourcing their consumption from markets. The bottom line is that the mere existence of rudimentary markets may not guarantee recursiveness in farm production and consumption decisions. Hence, separability has been a subject of empirical test. Existing empirical tests look at the link between on-farm labor demand and household demographic characteristics, where a positive correlation is considered as evidence for rejection of separability.

In this paper, I develop a new approach to test recursiveness and investigate how it varies across households with varying access to markets. My empirical test is derived from a simple model of household decisions on crop production and consumption in an environment where a household can engage in costly trade. On the consumption side a household maximizes utility by choosing how much of different crops to consume given its tastes for different crops, its income, and local prices. On the production side, the household decides how to allocate its limited land across potential crops given productivity of its land in the crops and local crop prices. If the household does not face significant trade barriers, its production decision is separable from its consumption preferences. Hence, the household's land allocation across crops should not be correlated with its tastes for these crops. Otherwise, the household land allocation across crops will be dictated by the household's tastes – and the extent to which tastes dictate crop production choices depends on the level of trade costs the household faces. A decrease in trade costs due

[1] Scattered settlements and lack of transport infrastructure, typical in rural areas of developing countries, would make rural labor markets too thin to be reliable source of labor input.

to road construction would thus weaken the link between consumption preference and production choices by improving households' opportunities to trade.

I empirically implement this test using a very rich panel data from Ethiopia on household production and consumption disaggregated by crops. I use a large-scale rural road construction project called Universal Rural Road Access Program (URRAP) as a source of variation to the household's market access/trade costs. I first estimate household crop tastes from a preference structure represented by Almost Ideal Demand Systems (AIDS) (Deaton and Muellbauer, 1980), where a household's taste for a crop is inferred from shifts in its expenditure function conditional on prices of all crops, the household's real total expenditure and demographic characteristics. I then test the separability hypothesis by regressing land allocation across crops on the estimated crop tastes, and explore how this correlation varies across households of varying level of market access. I find that households' crop tastes significantly affect the fraction of land allocated to the crop, which implies rejection of the separability hypothesis. Moreover, the effect of tastes on land allocation is stronger for households that reside further from market centers and roads, and improvements in market access due to large-scale rural road construction project leads to significant decreases in the correlation between household land allocation and tastes.

The empirical approach suggested here has several advantages over the previous studies testing separability. First, previous studies test separability by using the correlation between household on-farm labor demand and the household demographic characteristics. However, in the context of farm households, who are predominantly self-employing, on-farm labor demand is likely to be poorly measured.[2] On the contrary, in most agricultural surveys, including the survey used in this paper, land area is measured using GPS tools by well trained enumerators. Second, the current approach makes the link between separability and market integration straightforward. Given information on physical location of households and their nearest market centers, one can obtain a

[2]Collection of such data requires a farm household to recall how much each member of the household worked on-farm thought the agricultural season – which includes land preparation, planting, weeding, and harvesting periods.

measure of households' proximity to crop markets and explore how separability varies across households with varying level of proximity to markets. Such exercise is difficult to come by for labor markets because there is no physical location for labor markets. Third, the separability test suggested in this paper is based on crop markets, which are relatively more developed market in rural areas, particularly compared to labor markets. Hence, finding evidence of market malfunction in this context would imply a strong case against the separability hypothesis.

This paper is closely related to studies that empirically test separability. The seminal paper by Benjamin (1992) tests separability using the relationship between household on-farm labor demand and the household's demographic characteristics. The basic idea is as follows. If markets are complete and farm household's production decisions are independent of the household's preferences, a household's on-farm labor demand should be independent of its demographic composition, such as the number of active age persons in the household. Using data from rural Indonesia, Benjamin (1992) runs a regression of on-farm labor demand on different demographic characteristics and fails to reject the separability hypothesis. However, his crossection data did not allow him to address a number of confounding factors. Using (better) panel data from the same country, LaFave and Thomas (2016) conduct a followup study to Benjamin (1992) in which they run similar regressions to the latter but use panel data specification. They strongly reject separability, contrasting Benjamin (1992).

More recently, LaFave et al. (2020) suggest a new consumption based test for separability. The central idea of the test is that if household production and consumption decisions are recursive, input prices affect household demand for goods only through their effect on profits. This implies that the ratio of the effects of two inputs on demand for a good is equal across all goods. They implement this test using demand estimations and Wald tests of non-linear coefficient restriction. The downside of this approach is that the test lacks power to discern separability, particularly when the number of farm inputs are many. Moreover, in general equilibrium, input prices should not affect demand for goods once the goods' prices are controlled for (because the output prices are themselves determined by the input prices) unless the households are the owners of the inputs.[3]

This paper is also related to the literature on the development impact of rural roads. 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. Shamdasani (2018) studies the effect of a 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 an agricultural extension program that trains farmers on how to use best agricultural practices and technology adoption in Ethiopia. Shrestha (2018) finds that a 1% decrease in distance to roads due to expansion of highways resulted in a 0.1–0.25% increase in the value of agricultural land in Nepal. I contribute to this literature by providing evidence on another potential channel through which rural roads affect resource allocation and welfare, which is increased separability of household production and consumption decisions.

The rest of the paper is organized as follows. In section 2, I present the data and a series of descriptive evidences motivating the theoretical and empirical methods. Section 3 presents the theoretical model and section 4 discusses the empirical implementation of the theoretical model. Sections 5 presents the results. Section 6 concludes the paper.

2 Data

2.1 Sources

Agricultural production and consumption data: I use the 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

[3] This perhaps explains why most of the input coefficients in their demand estimation are either statistically insignificant or enter with signs that are inconsistent with their theory.

farm household's production, consumption and market participation, disaggregated by crops. The main advantage of the ESS dataset is its richness as it includes both production and consumption information, land and labor utilization, and a number of household demographic and geographic 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).[4]

Price data: The price data comes from three different sources. I construct village level prices of crops by combining two 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.[5] In villages that are not covered by AgPPS, I use ESS's price survey. 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.

I also use 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. RPS dataset covers over 100 urban centers across all administrative zones of the country. Both AgPPS and RPS are collected by CSA and go back to at least 1996. Importantly, the agricultural products covered in both datasets overlap almost fully. I use RPS, together with village prices constructed using the above procedure, to explore how rural road expansion affected urban-rural price gaps.

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

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

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

widely grown crops in Ethiopia such as *Teff*. For such crops, I use the Agricultural Sample Sample Survey (AgSS) data to construct village level suitability of land to the crops from the average yield in the villages over the period 2010-2013. 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 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 large-scale rural road expansion under URRAP as a source of variation to villages' access to road/market. Over the period 2012-2015, the Ethiopian government gave exclusive focus to the URRAP and constructed over 62,413kms 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 the first-round of this program were completed between 2013-2015.

2.2 Identification issues

One objective of this paper is to explore the link between market access and separability using the large-scale rural road expansion project under URRAP. There are three challenges to identify the causal effects of URRAP on the separability: selection bias, heterogeneity in treatment intensity, and spillover effects of road connectivity. Selection bias is a concern because villages are selected for the URRAP based on some demographic, geographic, social, and economic factors.[6] Villages that get connected to a dense network may gain more from the road than those that get connected to sparse network, implying heterogeneity in treatment intensity. Spillover effects is a concern because when a village is connected to the preexisting road network or to the nearest urban center, all its neighbors which did not get direct connection would also have improved access to market via the connected village. This would lead to underestimation of the causal effect of URRAP on separability.

I address the potential selection bias by using a matching-based Difference-in-Differences (DID) strategy where I first obtain a matched sample of treated and non-treated villages based on their observable characteristics that might be relevant for selection of villages for URRAP and then conduct DID estimation based on these matched sample of treated and non-treated villages. Combining matching with DID strategy is a powerful approach to address the selection problem. The matching step enables me to compare treated villages with non-treated villages that have similar observed characteristics and hence similar treatment probability. The DID strategy on these matched samples helps me to washout unobserved time-invariant village characteristics that may confound the treatment effect. I identify the following village characteristics for matching treated and non-treated villages in consultation with officials at Ethiopian Roads Authority (ERA): distance to nearest town, distance to preexisting road network, population size, average slope of land in the village, average elevation of the village, and average rainfall over 1990-2010 period. I use Digital Elevation Model (DEM) data and ArcGIS tools to calculate average slope and elevation of each village.

To address the heterogeneity in treatment intensity and spillover effects, I use market

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

access approach (Donaldson and Hornbeck, 2016) which captures treatment benefits from both direct and indirect connectivity, and accounts for the density of the network to which a village is connected. The market access measure is derived from a general equilibrium trade model and calculated using the entire road network and the distribution of population across Ethiopian villages. See Appendix A for details in the construction of market access measure. The constructed market access measure increases both for villages that are directly connected and those that are not by 47%, on average, but it increases more for the directly connected villages by about 40%.

2.3 Descriptive statistics

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

Large barriers to trade: Farmers face large trade barriers. These barriers 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 % and the mean is about 11%). 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]

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

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 frequent and larger crop markets because towns serve as hubs where most of the surrounding villages transact. Also, 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. However, there is significant

[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.

variation across crops.

Positive correlation between land and expenditure shares of crops: I use ESS data and focus on 20 crops for which complete information is available on both production and consumption. I calculate the share of each crop in household consumption expenditure[9] and the share of land allocated to the production of each crop. I run the following regression:

$$\eta^k_{hvt} = \beta_0 + \beta_1 s^k_{hvt} + \beta_2 p^k_{vt} + \beta_3 y^k_v + \gamma^k_t + \gamma_h + \varepsilon^k_{hvt} \tag{1}$$

where η and s are the land and expenditure share of crop, p is price, y is the GAEZ yield/productivity estimate which measures agro-climatic suitability of a village in each crop , h is household, v is a village, k is crop, and t is year. It is crucial to control for prices and yield in this regression because both production and consumption decisions are functions of these variables, directly or indirectly. Any significant positive correlation between the land and expenditure share of crop with in a household is suggestive evidence against separability. Under autarky, near perfect correlation between household land and expenditure shares of crops is expected. I run the regressions for each of the survey rounds separately to show how the estimated correlation changed over time.

Table 4 reports the results. Panel A reports the estimated correlations between the land and expenditure share for the three rounds of survey. The estimated correlation is 0.47 for the year 2011, which slightly increases to 0.53 for the year 2013 before it decreases significantly to 0.20 for the year 2015. We fail to reject the hypothesis that the correlations for the years 2011 and 2013 are equal, but similar hypotheses between 2015 and 2011 or 2013 are rejected at 1% significance level.

In panel B, I run analogous regressions where 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 about 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

[9] Household consumption from own production is valued at village level prices.

redo all the above regressions using the labor share of a crop as the dependent variable. The results looks very similar to the one we obtained using the land allocation. The correlation between the labor and expenditure share of crops slightly increase from 0.45 to 0.51 between 2011 and 2013 before it significantly decreases to 0.19 for the year 2015. To sum up, these statistically significant correlations strongly suggest that household resource allocation is at least partially dictated by their consumption preferences.

New roads decrease the correlation between land and expenditure share of crops: Before I explore the effects of URRAP roads on the correlation between the land and expenditure shares of crops, I provide evidence on whether the URRAP roads indeed decreased trade costs and improved market integration. In appendix B, I show that URRAP roads significantly decreased the urban-rural price gaps for crops, particularly for perishable crops such as vegetables. Strong negative correlation between local prices and local yield is an indicator of significant trade barrier. I show that, the construction of roads significantly weakened the inverse relationship between local prices and local productivity of crops. These evidences imply that the URRAP roads have indeed improved market integration of rural areas.

Next, I estimate the effects of URRAP roads on the correlation between land and expenditure share of crops. Even though URRAP was launched in 2011, almost all the roads were completed between 2013 and 2015. Thus, I use 2013 and 2015 as pre- and post-program periods.[10] Table 4 shows that the correlation between land and expenditure shares decreases significantly between 2013 and 2015. I estimate how much of this decline is attributed to the URRAP roads in the DID framework:

$$\begin{aligned} \eta^k_{hvt} = \beta_0 + \beta_1 s^k_{hvt} + \beta_2(Post_t * URRAP_v) + \beta_3(s^k_{hvt} * Post_t * URRAP_v) \\ + \delta Z_{vt} + \gamma^k_t + \gamma_v + \varepsilon^k_{hvt} \end{aligned} \quad (2)$$

where $Post_t * URRAP_v$ is a dummy variable indicating whether village v got new road connectivity under URRAP, which equals zero in 2013 and equals one in 2015 for villages

[10] Using 2011 as a pre-program period gives very similar result.

that get new roads, and Z_{vt} includes the vector of control variables in equation 1. The main parameter of interest is β_3, 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.

Table 5 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. The last two columns include village fixed effects. Columns 1 and 3 use the land share of a crop as dependent variable while column 2 and 4 use labor share of a crop. The results clearly show that road construction under URRAP caused a significant decline in the correlation between land/labor and expenditure shares of crops. Households in villages that got road connection between 2013-2015 have seen a decrease in the correlation between land and expenditure shares by about 0.16, compared to households in villages that were not directly exposed to the program. This is a large effect, roughly about 25% of the baseline correlation in 2011.

Overall, the above results suggest that household production and consumption decisions are likely made jointly. Moreover, the link between production and consumption decisions seems to be significantly influenced by the level of underlying market integration. Below, I build on these evidences to suggest a formal framework to test whether household production decision is dictated by the household's consumption preferences and how improvements in market access affects the link between the two.

3 Theoretical framework

Informed by the above facts, in this section I develop a theoretical framework to test the separability hypothesis. In doing so, I borrow tools from the Ricardian trade models. Particularly, I build on Eaton and Kortum (2002), Donaldson (2018), and Sotelo (2018).

Consider an economy constituting villages $v = 1, ..., V$. Each village is populated by I households indexed by $i = 1, ..., I_v$. The household derives utility from consumption of K homogeneous crops indexed by $k = 1, ..., K$ that can be potentially produced or purchased.

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

$$U_{ivt} = f\Big(\mu_i^k; q_{it}^k\Big)$$

where $f(.)$ is a common utility function across households in the country, q_i^k is the quantity of crop k consumed by household i, and μ_i^k is the household taste for crop k, which is assumed to be fixed over the short to medium period. The household crop tastes act as pure demand shifters. The household maximizes this utility subject to the following budget constraint:

$$\sum_k p_{vt}^k q_{it}^k \leq \Pi_{it} \tag{3}$$

where p_{vt}^k is village level crop price, and Π_{it} is household farm income/profit.

Production: I follow Sotelo (2018) to describe the farmer's production problem. 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:

$$y_i^k(\omega) = g\Big(z_i^k(\omega), \mathbf{x}_i(\omega); \boldsymbol{\alpha}^k\Big)$$

where $y_i^k(\omega)$ is the quantity of crop, $\mathbf{x}_i(\omega)$ is the amounts of vector of variable inputs (such as labor and fertilizer) used on the plot, and $\boldsymbol{\alpha}^k$ denotes parameters.

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 a farmer, Ω_i. It can be interpreted as the average productivity of farmer i's land in crop k, as determined by agro-climatic conditions of the village, and soil, slope, and other characteristics of the farmer's plots. θ is the degree of homogeneity in the set of plots owned by a farmer, and it is constant across crops.

Farmers are geographically separated and there is an iceberg trade cost of $\tau_{vv'}^k \geq 1$ between farmers in villages v and v' in crop k.[11] Motivated by the result in appendix B, which shows that spatial price variation differs across crops, trade costs are assumed to vary across crops to reflect that some crops, such as vegetables, are more costly to trade (e.g., perishable) than others such as cereals. I assume that $\tau_{vv}^k = 1, \forall k$, and impose the standard assumption of triangle inequality in trade costs, $\tau_{vv'}^k \times \tau_{v'v''}^k \geq \tau_{vv''}^k, \forall k$.

3.1 Two extreme cases

To motivate the separability test, it suffices to consider the farmers' problem under two extreme cases so that we can characterize which of the two cases closely matches the farmer's observed choices. The first is the case where farmers are allowed to trade with each other paying reasonable trade costs, and the second is the case where trade costs are too high for the farmers to engage in trade. I discuss how we can generalize from these two extreme cases and provide a general proof of the link between separability and trade costs in appendix C.

Case-I: Separability $\tau_{vv'}^k << \infty, \forall k$. Suppose trade costs are such that farmers can buy and sell any crop at a prevailing market price. Assuming perfect competition, no

[11] For simplicity, I assume that within village trade costs between farmers are negligible. The median village has area of about $25km^2$. While distance is not a big impediment to trade within village, the fact that farmers within a village share similar agro-climatic condition implies that there is less room for crop trade within a village compared to across villages.

arbitrage condition implies that for any two villages v and v', equilibrium crop prices satisfy $p^k_{vv'} = \tau^k_{vv'} p^k_{vv}$ where $p^k_{vv'}$ is price in village v' of crop k originating from village v, and p^k_{vv} is price in village v of crop k originating from the same village v.

Under this case, the farmer takes local crop prices p^k_v and a vector of local input prices $\mathbf{r}_v$ as given, and allocates land across crops. The fraction of household land allocated to crop k is given by:

$$\eta^k_i = h\Big(p^k_v, \mathbf{p}^{\mathbf{l}}_v, \mathbf{r}_v, A^k_i, \mathbf{A}^{\mathbf{l}}_i; \theta, \boldsymbol{\alpha}^k, \boldsymbol{\alpha}^{\mathbf{l}}\Big) \tag{4}$$

where $\mathbf{l} = 1, ..., K \neq k$. This implies that the quantity of crops produced and revenue from each crop are given, respectively, by:

$$y^k_i = \mathcal{Y}\Big(p^k_v, \mathbf{p}^{\mathbf{l}}_v, \mathbf{r}_v, A^k_i, \mathbf{A}^{\mathbf{l}}_i, L_i; \theta, \boldsymbol{\alpha}^k, \boldsymbol{\alpha}^{\mathbf{l}}\Big), \quad \text{and} \tag{5}$$

$$R^k_i = \mathcal{Y}\Big(p^k_v, \mathbf{p}^{\mathbf{l}}_v, \mathbf{r}_v, A^k_i, \mathbf{A}^{\mathbf{l}}_i, L_i; \theta, \boldsymbol{\alpha}^k, \boldsymbol{\alpha}^{\mathbf{l}}\Big) \tag{6}$$

where $\mathbf{l} = 1, ..., K \neq k$.

Given farm profit $\Pi_i = \sum_k R^k_i - \mathbf{r}_v.\mathbf{x}_i$, the farmer then maximizes utility subject to the budget constraint. The optimal quantities of each crop are given by:

$$q^{*k}_i = \mathcal{C}\Big(\mu^k_i, \mu^{\mathbf{l}}_i, p^k_v, \mathbf{p}^{\mathbf{l}}_v, \Pi_i\Big) \tag{7}$$

Equations 4-7 show that: (i) household land allocation across crops and quantities of crops produced are independent of crop tastes μ^k_i, (ii) tastes affect household demand for crops but not production decisions, and (iii) household production decisions affect household demand only through its effect on farm profits. These imply that household decision is recursive: the household first makes production decision to maximize its farm profits given local crop prices, inputs prices and productivity, and in the second stage the household chooses optimal quantities of crops to consume given local crop prices, tastes, and farm profit.

Case-II: Autarky $\tau_{vv'}^{k} \longrightarrow \infty$, for some k. Under this case, there is no market for some crops and hence, no market prices which the farmer takes as given. Instead, the farmer's decision is based on shadow prices which are functions of the household tastes and other household characteristics: $\tilde{p}_i^k = \mathcal{P}\Big(\mu_i^k, \mu_i^{\mathbf{l}}, \mathbf{r}_v, A_i^k, \mathbf{A}_i^{\mathbf{l}}, L_i, U_i; \theta, \boldsymbol{\alpha}^k, \boldsymbol{\alpha}^{\mathbf{l}}\Big)$ where $\mathbf{l} = 1, ..., K \neq k$.

The household makes production decision given the shadow prices and productivity distribution parameters. Plugging this in the land allocation we have the following land allocation rule under autarky.

$$\tilde{\eta}_i^k = h\Big(\tilde{p}_i^k, \tilde{\mathbf{p}}_i^{\mathbf{l}}, \tilde{\mathbf{r}}_v, A_v^k, \mathbf{A}_v^{\mathbf{l}}; \theta, \boldsymbol{\alpha}^k, \boldsymbol{\alpha}^{\mathbf{l}}\Big) \tag{8}$$

where $\mathbf{l} = 1, ..., K \neq k$.

In equation 8, the fraction of land allocated to crop k depends on the household taste for the crops via the shadow prices. That is, household production decision is not independent of its consumption preferences. This is a key result from which the separability test is derived in this paper.

3.2 Trade costs and separability

Here, I describe the intuition for generalizing the link between trade costs and separability, postponing formal proof to appendix C. To make the generalization clear, consider the case where, due to lack of transport, some goods are non-tradable. Perishable vegetables are good examples in rural areas of developing countries. Because the households have to rely on self-production for these high trade cost crops, the separability assumption no longer holds. The fraction of land allocated to such crops would be dictated by the households' tastes for these crops. In general, the probability that a farmer is the cheapest supplier of any given crop to itself, compared to all other farmers in the country, increases with trade costs. On the other hand, the probability that a farmer is the cheapest supplier of a given crop to any other farmer decreases with trade costs. These two probabilities determine household land allocation rule as a function of trade costs, tastes, prices and productivity for any level of trade costs. Given this land allocation rule, one can obtain different comparative statics. In appendix C, I show that as trade costs increase, the correlation between the fraction of land allocated to a crop and crop tastes increases, $\frac{\partial^2 \psi_i^k}{\partial \tau_{ij}^k \partial \mu_i^k} \geq 0$ where ψ_i^k is the fraction of land allocated to crop k for arbitrary trade costs.

4 Empirical methodology

4.1 Estimating household crop tastes

I follow Atkin (2013) to estimate household tastes for crops. Suppose household preference for crops is represented by the following expenditure function corresponding to Almost Ideal Demand System (AIDS) (Deaton and Muellbauer, 1980) where the coefficients of the first-order price terms are allowed to vary across households to allow for taste variations:

$$\ln e(u, \mathbf{p}_{vt}; \Theta) = \mu_0 + \sum_k \mu_i^k \ln p_{vt}^k + \frac{1}{2} \sum_k \sum_{k'} \gamma^{*kk'} \ln p_{vt}^k \ln p_{vt}^{k'} + u\beta_o \prod_k {p_{vt}^k}^{\beta_k} \tag{9}$$

where t represents years. Applying Shefard's Lemma and replacing u by indirect utility function gives the following expression for the expenditure share of crops:

$$s_{it}^k = \mu_i^k + \sum_{k'} \gamma^{kk'} \ln p_{vt}^{k'} + \beta_k \ln \frac{m_{it}}{P_{vt}} \tag{10}$$

where $\gamma^{kk'} = \frac{1}{2}(\gamma^{kk'*} + \gamma^{k'k*})$, m_i is household nominal expenditure on food, P_v is village price index, and $\frac{m_i}{P_v}$ is real expenditure. Following Deaton and Muellbauer (1980) and Atkin (2013), I use Stone index for for village price index, $\ln P_v = \sum_k \bar{s}_v^k \ln p_v^k$ where $\bar{s}_v^k$ is the average expenditure share of crop k in village v.

Household crop tastes μ_i^k are thus demand shifters, conditional on prices and total real expenditure of the household. The key assumption here is that tastes for crops do not change over short period of time. Atkin (2013) shows that regional tastes are indeed stable over time due to habit formation.

Atkin (2013) discusses two necessary conditions for identification of tastes in a similar equation to 10. The first is the existence of temporary and supply driven price variation within village. In my setup, this condition is satisfied by price variation due to rainfall fluctuations. See table A.4 in appendix E for evidence of price volatility in response to rainfall fluctuations. The second condition, which is assumed to hold, is the existence of a common preference structure across rural households in Ethiopia, conditional on taste differences, and that AIDS function approximates this preferences reasonably well.

I estimate the following equation to identify household crop tastes:

$$s_{it}^k = \mu_i^k + \sum_{k'} \gamma^{kk'} \ln p_{vt}^{k'} + \beta^k \ln \frac{m_{it}}{P_{vt}} + N_{it} + \delta_t + \varepsilon_{it}^k \tag{11}$$

where N denotes household size and other demographic characteristics, δ_t is year fixed effects and ε_{it}^k is the error term. Estimating equation 11 using OLS might be problematic because unobserved factors correlated with both village prices and household idiosyncratic tastes could bias the estimated price coefficients and the taste parameters. Similar to Atkin (2013), I address this concern by instrumenting village prices by prices in the nearest villages (Hausman, 1994).

As a robustness check, I also estimate a specification where households within a village share the same crop tastes, i.e., $\mu_i^k = \mu_v^k, \forall i \in v$. The motivation for this is that village sizes are small (a median village in my sample has an area of about $25km^2$) and village population share the same culture including ethno-linguistic culture, and perhaps also the same food culture. As is shown below, the empirical result also strongly supports this conjecture – about 81% of variation in household crop tastes comes from across village variation. The taste parameters estimated following the two approaches have a correlation of about 0.90.

4.2 Testing separability

Once I obtain estimates of household crop tastes, I test the separability hypothesis by looking at whether household land allocation across crops is independent of the household crop tastes, conditional on village crop prices and yields (agro-climatic suitability of village for each crop). I estimate the following regression:

$$\eta_{ivt}^k = \beta_0 + \beta_1\mu_i^k + \beta_2 \text{ln}p_{vt}^k + \beta_3 \text{lnY}_v^k + \beta_4^k \text{lnRainfall}_{vt} + \gamma_t^k + \gamma_v + \epsilon_{ivt}^k \tag{12}$$

where η_{ivt}^k is the fraction of land allocated to crop k. Recursiveness requires that $\beta_1 = 0$, that is, there is no significant correlation between household land allocation across crops and the household crop tastes. On the other hand, a positive and statistically significant β_1 is evidence against recursiveness. The higher β_1, the closer the village economy is to an autarky.

The role of market access: Next, I explore how infrastructure and market integration affect the link between household production and consumption choices. The theoretical model implies that decreases in trade costs lead to a decrease in the correlation between the land share of crops and crop tastes (see appendix C for a formal proof). I run the following regression:

$$\begin{aligned}\eta_{ivt}^k = {} & \beta_0 + \beta_1\mu_i^k + \beta_2 \text{MA}_{vt} + \beta_3(\mu_i^k \times \text{MA}_{vt}) + \beta_4 \text{ln}p_{vt}^k + \beta_5 \text{ln}Y_v^k \\ & + \beta_6^k \text{lnRainfall}_{vt} + \gamma_t^k + \gamma_v + \epsilon_{ivt}^k\end{aligned} \tag{13}$$

where MA is a measure of village market access derived from general equilibrium trade models (Donaldson and Hornbeck, 2016). The market access (MA) is calculated using data on (i) the entire road network in Ethiopia, (ii) the spatial distribution of population across the country and (iii) the freight costs of transporting one ton of cargo from origin village to destination village along the least cost path, before and after the construction of URRAP roads, and trade elasticity parameter. The large-scale rural road expansion between 2013 and 2015 led to significant decreases in freight costs, increasing MA for all villages, particularly those villages that got direct road connectivity under the program (see appendix A for the detailed procedure followed in constructing MA measures). In equation 13, a negative and statistically significant β_3 would imply that market integration plays important role in weakening the link between household production and consumption choices.

5 Results

5.1 Estimating tastes and the separability test

The taste estimates: It is worth mentioning few points about the estimated taste parameters. First, both OLS and IV estimation of equation 11 give very similar results. The correlation between the taste estimates obtained from these approaches is about 0.96. Second, the estimated crop tastes show significant variation across households. However, most of the variation comes from across village variations – on average, 81% of the variation in tastes comes from across villages. Third, because of small within village variation in tastes for crops, estimating tastes at village level gives very similar result to household level taste estimates when both OLS and IV estimation is used. Overall, the IV passes the under-identification and weak identification tests remarkably and borderline passes the weak instrument test with average first-stage F-statistics of about 10.

Testing separability: Next, I explore how the estimated taste parameters correlate with household land allocation. The theoretical results in section 3 suggest that, if household production decisions are independent of their consumption preferences, a household taste for a crop should not affect the fraction of land the household allocates to the crop. Table 6 reports the results for estimation of equation 12. I allow the coefficient of taste to vary across years in order to see whether the estimated coefficient changes over time. The first columns uses OLS taste estimates while the second column uses IV taste estimates. Across all rows and columns, we observe that tastes significantly affect household land allocation, implying rejection of the separability/recursiveness hypothesis. The coefficient of taste slightly increases from 0.68 to 0.70 between 2011 and 2013, before it declines to about 0.66 in 2015. However, these coefficients are not statistically different from each other.

5.2 Separability and proximity to market

Before turning to the effect of road expansion under URRAP, I first explore how the correlation between land allocation and tastes varies across households with varying proximity to population centers (towns with above 20,000 population) and to all-weather roads. Towns serve as hubs and market centers for the surrounding villages. Also, most villages access the rest of the country via the nearest towns. Hence, proximity to towns is important for market access. Similarly, proximity to all-weather roads improve the village's access to the rest of the country. I use distances to nearest population centers and nearest roads to measure proximity. The first is time invariant while the latter decreases for households residing in villages that obtained new roads under URRAP.

If lack of access to market is a driving factor for the observed correlation between land allocation and tastes, one would expect that the correlation would be stronger for household that live further from towns or roads. Table 7 reports the results. Panel A shows that the correlation between land allocation and the taste significantly increases with distance to nearest population center. Using the result in the first column and the range of log distance to population center of about 6, the correlation between land allocation and taste ranges from about 0.46 for the nearest to 0.81 for the furthest household to population center. Panel B reports similar result using distance to nearest road. The correlation between land allocation and tastes increases significantly with distance from road, even though distance to road has weaker effect compared to distance to distance to towns.[12]

5.3 The effect of URRAP on separability

Finally, I explore the effect of road expansion under URRAP on the correlation between land allocation and tastes. As mentioned in section 2, I use a matching-based DID estimation strategy to minimize selection bias. That is, I first obtain a matched sample of treated and non-treated villages based on a set of village characteristics before conducting

[12] This is partly because there is less variation in households' distances from road compared to distances from population center. The standardized coefficients are similar across the two variables.

DID estimation. Figure 3 shows the histogram of propensity score by treatment status and table 8 reports the balancing of the matching variables. I also report DID estimation results without matching for comparison, but my discussions will be based on the results from matching-based DID estimation.

Table 9 reports the matching-based DID estimation results for equation 13. The first two columns use binary treatment while the last two columns use a continuous market access measure. To facilitate interpretation, market access measure is standardized. Across all columns, we see that road connections under URRAP led to significant decreases in the correlation between land allocation and tastes. The first two columns show that the correlation between the land share of crops and crop tastes decreases by 0.054-0.056 for villages that got direct road connection under URRAP compared to the control villages. Columns 3 and 4 show that one standard deviation increase in market access leads to 0.071-0.077 decrease in the correlation between the land share of crop and the crop tastes.

Overall, the results in table 9 clearly show that improvement in access to market due to URRAP has led to decreases in the correlation between land allocation and tastes. That is, road connection under URRAP has led to more separability between household production decision and consumption preferences. Moreover, the estimated decrease in correlation between land allocation and tastes is significant considering the fact that the time span after the roads were completed is is too short for the village economy to adjust fully to the expansion of infrastructure. One would expect that the correlation would decrease more in the long term because the infrastructure expansion would lead to over time improvement in transport options and the thickness of local crop markets, which would significantly alter household land allocation rule.

Table 10 reports the result for DID estimation without matching for comparison. The results in this table look similar to those in table 9, except that the estimated effects of URRAP are smaller and statistically insignificant in the first two columns.

6 Conclusions

In this paper, I suggest a new approach to test separability/recursiveness between household production and consumption decisions, and explore how it is related to market integration. My empirical test is derived from on a simple theoretical insight that if household production decision is independent of its consumption preferences, the household's tastes for different crops would not affect household land allocation across crops. The theoretical model also suggests that the extent to which tastes affect household land allocation across crops depends on the level trade costs the households are facing.

I implement this test using a very rich household panel data from Ethiopia. The dataset includes household production and consumption information disaggregated by crops and coincides with period of large-scale rural road expansion. I first estimate household crop tastes from Almost Ideal Demand Systems (AIDS) where household taste for a crop is inferred from shifts in expenditure share of a crop conditional on prices of all crops, household real total expenditure, and household demographic characteristics. Next, I conduct the separability test by regressing the land share of crop on the estimated crop tastes and find that the separability hypothesis is strongly rejected. I also show that the correlation between land allocation and tastes is stronger for households that reside further from market centers and roads. Finally, I explore the effect of a large-scale rural road expansion on the correlation between land allocation and tastes, and find that improvement in market access due to the road expansion led to significant decreases in the correlation between land allocation and tastes.

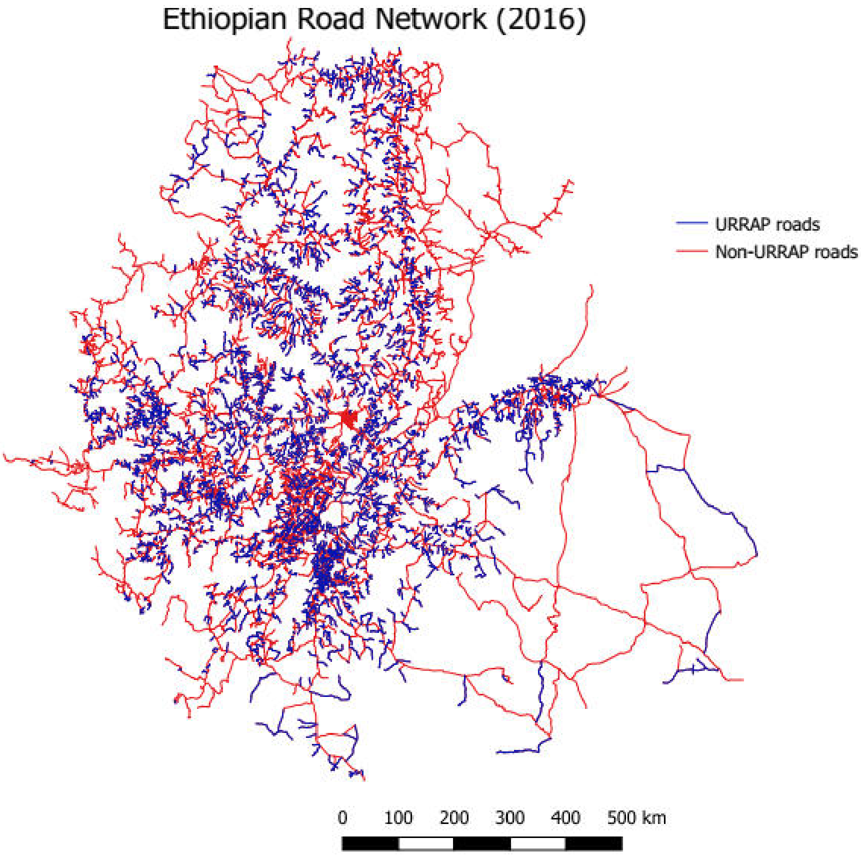

Figure 1: Rural road expansion under URRAP

Figure 2: Completed URRAP roads (pictures are taken from Oromia Roads Authority).

Figure 3: Common support of propensity score matching

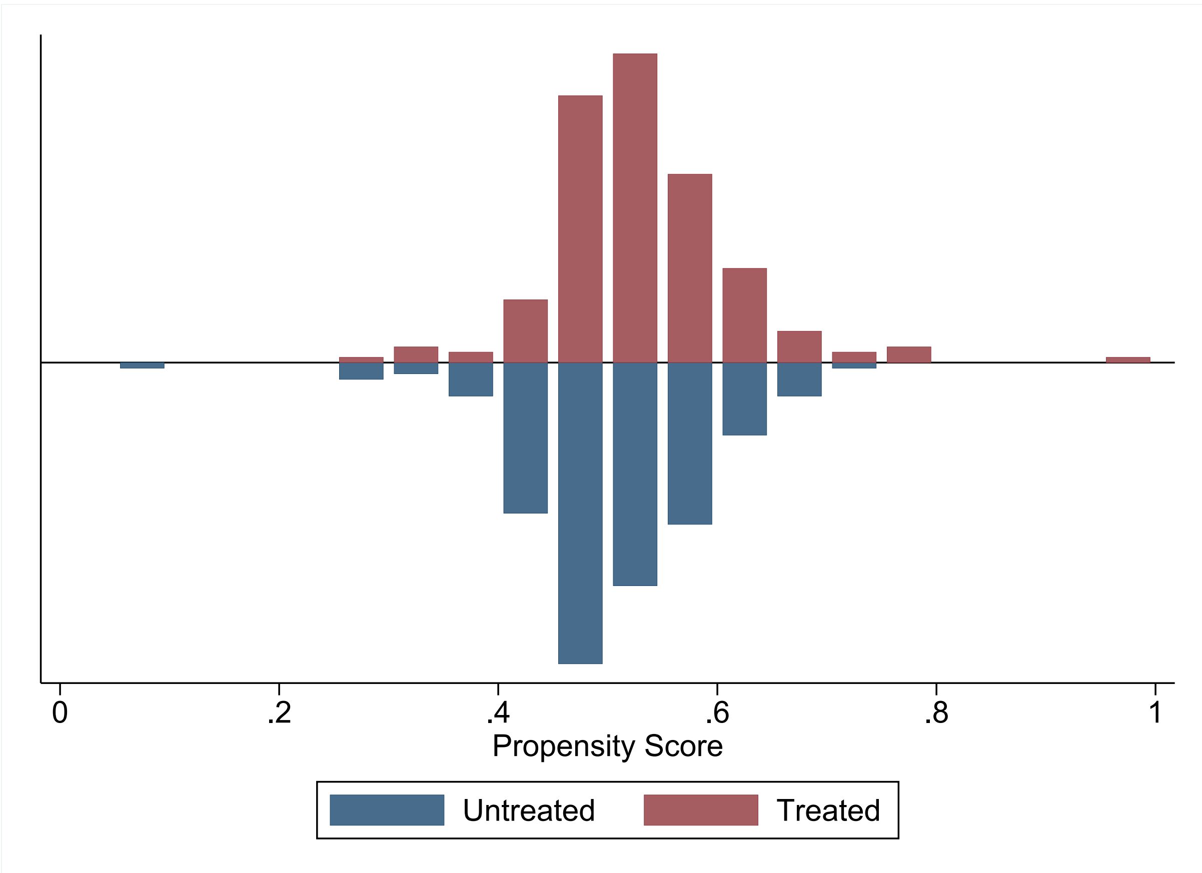

Table 1: Transport modes to market, proximity to market, and trade costs

	2011	2015
Transport mode		
On Foot	43.6	41.9
Pack Animals	45.8	43.9
Own Bicycle or Oxcart	6.74	4.78
Vehicle	2.34	5.69
Others	1.43	3.62
Proximity to market		
Distance to all-weather road (median KM)	10	8.5
Distance to population centers (median KM)	30	30
Distance to district (woreda) town (median KM)	17	17
Distance to nearest weekly market (median KM)	12	8
Trade Costs		
Ad valorem trade cost vehicle (mean)	11.37	6.4
Ad valorem trade cost vehicle (median)	6.49	3
Median transport fare to district capital (real Birr/KM)	0.7	0.523

Notes: This table is based on households who report market participation in 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	
	Consumed	Consumed& Not produced	Consumed	Consumed &Not produced
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: Correlation between household production and consumption decisions

	2011	2013	2015
	Panel A: Land share of crops		
Expenditure Share	0.468***	0.530***	0.200***
	(0.025)	(0.028)	(0.021)
N	71067	71130	69932
R^2	0.284	0.331	0.177
	Panel B: Labor share of crops		
Expenditure Share	0.453***	0.513***	0.193***
	(0.025)	(0.030)	(0.021)
N	71067	71130	69932
R^2	0.292	0.328	0.182

Notes: Standard errors are clustered at village level. In panel A the dependent variable is the share of household land allocated to each crop while in panel B it is the share of labor allocated to each crop. All regressions include the control variables of village crop prices and yields, household fixed effects, and crop fixed effects. Observations are weighted by the household sampling weight. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5: URRAP roads and the correlation between production and consumption decisions

	No village FE		With Village FE	
	Land	Labor	Land	Labor
Expenditure Share	0.372***	0.361***	0.370***	0.359***
	(0.026)	(0.027)	(0.026)	(0.027)
Post*URRAP	0.008***	0.008***	0.008***	0.008***
	(0.002)	(0.002)	(0.002)	(0.002)
Expenditure Share*Post*URRAP	-0.165***	-0.162***	-0.164***	-0.161***
	(0.032)	(0.032)	(0.032)	(0.032)
N	133372	133370	135461	135458
R^2	0.214	0.217	0.217	0.220

Note: Standard errors are clustered at village level. In columns 1 and 2 I include log distance to population centers and log distance to roads in 2011 (before the onset of URRAP). In columns 3 and 4, I include village fixed effects. All regressions include crop and year fixed effect. Observations are weighted by the household sampling weight.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6: The separability test

	OLS Taste	IV Taste
Taste*2011	0.666***	0.677***
	(0.030)	(0.030)
Taste*2013	0.693***	0.702***
	(0.032)	(0.032)
Taste*2015	0.659***	0.655***
	(0.034)	(0.034)
N	153293	153293
R^2	0.327	0.330

Note: Standard errors are clustered at village level. All regressions include the following control variables: log village prices, log village yields, and log rainfall with crop specific coefficients. All regressions include village, crop, and year fixed effect. Observations are weighted by the household sampling weight.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7: Separability and proximity to market

	Panel A: Distance to Population center	
	OLS Taste	IV Taste
Taste	0.454***	0.504***
	(0.112)	(0.104)
Log Dist. to Pop. Center	-0.000	-0.002**
	(0.000)	(0.001)
Taste*Log Dist. to Pop. Center	0.059**	0.047*
	(0.028)	(0.026)
N	153293	153293
R^2	0.322	0.324
	Panel B: Distance to Road	
Taste	0.613***	0.616***
	(0.052)	(0.051)
Log Dist. to Road	-0.000	-0.001*
	(0.000)	(0.001)
Taste*Log Dist. to Road	0.029*	0.029*
	(0.016)	(0.015)
N	153293	153293
R^2	0.321	0.324

Note: Standard errors are clustered at village level. All regressions include the following control variables: log village prices, log village yields, and log rainfall with crop specific coefficients. All regressions include village, crop, and year fixed effect. Observations are weighted by the household sampling weight.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 8: Balancing of variables for Average Treatment Effect on Treated (ATT)

	Treated	Control	% bias	t-stat	p-value
Population	5992.6	5983.3	0.2	0.23	0.822
Distance to nearest asphalt road	39.284	39.333	-0.1	-0.12	0.906
Distance to Woreda town	17.017	17.017	0.0	0.00	1.000
Distance to nearest major town	63.647	63.626	0.0	0.05	0.962
Distance to the nearest weekly market	7.3438	7.3438	0.0	0.00	1.000
Land slope	2.6623	2.6639	-0.1	-0.12	0.905
Fraction of land covered by forest	14.342	14.429	-0.6	-0.65	0.516
Average rainfall (1990-2010)	1149.8	1150.5	-0.1	-0.16	0.870

Notes: Population and rainfall correspond to the period before URRAP. Land slope is categorical variable with Flat=1, Slightly Slopeing=2, Moderately Sloping=3, Seeply sloping=4, and Hilly=5.

Table 9: The effects of URRAP on separability: Matching-based DID estimation

	Binary treatment		Market access approach	
	OLS Taste	IV Taste	OLS Taste	IV Taste
Taste	0.735***	0.738***	0.728***	0.730***
	(0.035)	(0.035)	(0.034)	(0.034)
Post*URRAP	-0.002**	0.000		
	(0.001)	(0.001)		
Taste*Post*URRAP	-0.056*	-0.054*		
	(0.029)	(0.029)		
Market Access			-0.000	0.002*
			(0.001)	(0.001)
Taste*Market Access			-0.077***	-0.071***
			(0.026)	(0.023)
N	77872	77712	77109	76949
R^2	0.343	0.345	0.345	0.347

Note: Standard errors are clustered at village level. Market access measure is standardised so that the coefficient can be interpreted as the effect of one standard deviation increase in market access. All regressions include the following control variables: log village prices, log village yields, and log rainfall with crop specific coefficients. All regressions include village, crop, and year fixed effect. Observations are weighted by the household sampling weight.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 10: The effects of URRAP on separability; DID estimation

	Binary treatment		Market access approach	
	OLS Taste	IV Taste	OLS Taste	IV Taste
Taste	0.680***	0.684***	0.685***	0.688***
	(0.035)	(0.035)	(0.034)	(0.034)
Post*URRAP	-0.001*	0.000		
	(0.001)	(0.001)		
Taste*Post*URRAP	-0.043	-0.038		
	(0.030)	(0.029)		
Market Access			-0.000	0.002
			(0.001)	(0.001)
Taste*Market Access			-0.068**	-0.063**
			(0.028)	(0.025)
N	102552	102552	98029	98029
R^2	0.366	0.368	0.370	0.372

Note: Standard errors are clustered at village level. Market access measure is standardised so that the coefficient can be interpreted as the effect of one standard deviation increase in market access. All regressions include the following control variables: log village prices, log village yields, and log rainfall with crop specific coefficients. All regressions include village, crop, and year fixed effect. Observations are weighted by the household sampling weight.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendices

A Construction of market access measure

The major concerns in identifying the effects of road connectivity based on a binary treatment dummy include: (i) heterogeneity in treatment intensity across villages that get connected to sparse network and those that get connected to dense network, and (ii) the potential spillover effects of the roads to villages that are not directly connected. When a given village is connected to the pre-existing road network or to the nearest urban center, all its neighboring villages which are not directly connected also have improved access to market via the connected village. As a result, non-connected villages may not serve as control groups in identification of the effects of road connection. Both these concerns can be addressed by using a treatment measure that takes into account change in *market access* from both direct and indirect connectivity, and the density of the network to which a village gets connected. I use market access measure derived from general equilibrium trade models (see Donaldson and Hornbeck (2016)) that are calculated using the entire road network and the distribution of population across Ethiopian villages:

$$MarketAccess_{ot} = \sum_{d} \tau_{odt}^{-\theta} Population_d \tag{14}$$

where $Population_d$ is destination village population from the 2007 census (before the onset of the URRAP program). Using pre-URRAP population distribution is necessary because population distribution is likely to respond to improvement in road infrastructure. θ is trade elasticity parameter which I estimate as discussed below.

τ_{odt} is the freight costs of transporting one ton of cargo from origin village o to destination village d along the least cost path, before ($t = 0$) and after ($t = 1$) the construction of URRAP roads. I use the following procedure to estimate τ_{odt} for each year. First, I construct a link from each village centroid to the nearest available road in year t. Next, I use data on costs of moving weight (in USD per ton-kilometer) for five different road quality levels: asphalt, major gravel, cobbled road, minor gravel, and earth road. Because there is no similar cost estimates along the link roads, I scale up the costs along earth road by the factor of $\frac{\textit{Cost along earth road}}{\textit{Cost along minor gravel}}$ to obtain estimate of cost along the links.[13] After assigning each road type (including the links) with the estimated costs in USD per ton-kilometer, I use ArcGIS tools to calculate the costs (in USD) of moving a ton of weight from origin o to destination d along the least cost path, in each year. I use these estimates as τ_{odt}. As can be seen in equation 14, a change to a village's market access comes only from changes in τ_{odt}.

Estimation of θ: I follow Sotelo (2018) in the estimation of θ. I assume that the average productivity of a farmer's land is related to the GAEZ yield measure Y_v^k in the following equation:[14]

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

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[15] $\eta_i^k = \frac{(p_i^k A_i^k)^\theta}{(\Phi_i)^\theta}$, where $\Phi_i = \left(\sum_{l=1}^{K}(p_i^l A_i^l)^\theta\right)^{\frac{1}{\theta}}$, and taking logs gives:

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

The empirical counterpart of this is:

$$\ln\left(P_v^k Y_v^k\right) = \frac{1}{\theta}\eta_v^k + \gamma_v + \gamma^k + u_v^k$$

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

[14] Some of the crops in my sample do not have GAEZ productivity estimate. For these crops, 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 (2010-2013) to obtain a time invariant measure of yield for a crop in a village. I make sure that this approach gives reasonable village productivity estimate from comparison of GAEZ and AgSS village productivity measures for those crops that are covered in both datasets.

[15] I drop the variable input prices from the land share equation, as they do not affect the estimation equation to obtain θ (they are subsumed in village fixed effects).

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.

Once I obtain the estimate for θ, I plug in equation 14 along with data on population and freight costs to obtain village level market access measures, before and after the URRAP program.

B URRAP roads and market integration

I use two measures of market integration to provide evidence on the effect of URRAP on market integration. The first measure is urban-rural price gap while the second measure is correlation between local prices and local yields for crops.

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:

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

where k denotes crop, v is village, z is zone capital, m is month, t is year, $Post$ equals zero for all month-years before URRAP and one for all month-years after URRAP; $URRAP_v$ is a dummy variable representing whether a village got URRAP road between 2013 and 2015; and γ^k_m is crop-month fixed effect which captures possible seasonality of crop prices.

The result is reported in Table A.1. It shows that road connection significantly decreased the urban-rural price gap. The first column pools all 56 crop varieties 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.

Table A.1: URRAP road access and trade costs

	Dependent Variable: log(Price in Zone Capital/Price in village)	
	All crops	Vegetables and Fruits
$Post_t * URRAP_v$	-0.031**	-0.079*
	(0.016)	(0.044)
N	82944	24468
R^2	0.378	0.360

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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

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 higher (lower) productivity. Market integration weakens this inverse relationship between local prices and local yield. I run the following generalized DID regression to investigate this:

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

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 A.2. We see that there is a negative relationship between local prices of a crop and local yield, 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%.[16]Panel B of table A.2 reports the corresponding estimation result using market access measure instead of binary treatment dummy. The result clearly shows that in villages that see an increase in their market access, the negative correlation between crop price and yield becomes significantly weaker.

C The link between trade costs and separability

In this section, I provide a general proof the effect of trade costs on correlation between land share and taste. To simplify notation, I assume land is the only factor and the production function is linear in land.[17]

Proof. Let r_v is the rental rate of a plot of land in village v, which is determined in equilibrium (see Sotelo 2018). The unit cost of production $c_i^k = \frac{r_v}{Z_i^k}$ 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, $P_{ij}^k = \frac{r_v}{Z_i^k}\tau_{ij}^k$, 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:

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

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

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

[17]None of the results in this section hinge on these assumptions. One can show that similar results hold if we use multiple inputs in production function.

Table A.2: Rural roads and the link between local prices and local yield: the dependent variable is village crop prices

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

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

across all other farmers and is given by:

$$\begin{aligned} G_j^k(p) &= 1 - \prod_{i=1}^{I}(1 - G_{ij}^k(p)) \\ &= 1 - \exp\Big(- p^\theta \sum_{i=1}^{I} (A_i^k)^\theta (r_v \tau_{ij}^k)^{-\theta}\Big) \end{aligned} \tag{15}$$

Given this probability distribution, we can derive 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) as:

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

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.[18]

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

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

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 that $\pi_{ij}^k \leq \pi_{ii}^k$ because of trade costs: a farmer is more likely to be the cheapest supplier of a crop to itself than being the cheapest supplier to any other farmer. Figure A.1 illustrates this.

[18] Note that, to save on notation I use r_i instead of r_v, even though rental rates are the same across farmers in the same village. This is without loss of generality because $r_i = r_j, \forall i, j \in v$.

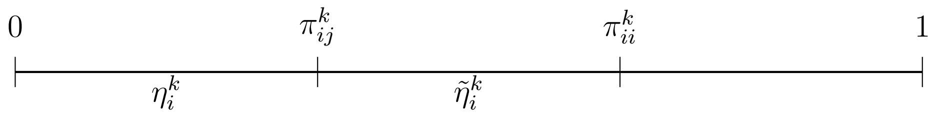

Figure A.1: This figure illustrates the probability of a farmer being a cheapest supplier of a crop to itself and to any other farmer.

Let ψ_i^k is the fraction of land allocated to crop k by farmer i. ψ_i^k is given by:

$$\psi_i^k = \pi_{ij}^k \eta_i^k + (\pi_{ii}^k - \pi_{ij}^k)\tilde{\eta}_i^k$$

where η_i^k and $\tilde{\eta}_i^k$ are given in section 3. Taking derivative with respect to crop taste μ_i^k gives

$$\frac{\partial \psi_i^k}{\partial \mu_i^k} = (\pi_{ii}^k - \pi_{ij}^k)\frac{\partial \tilde{\eta}_i^k}{\partial \mu_i^k}$$

which is positive given the expression for $\tilde{\eta}_i^k$. That is, more land is allocated to a crop for which the household has higher taste. Now, to show that the effect of taste on land share is stronger if the household's trade cost for crop k with any other farmer j is higher, we take derivative of the above equation with respect to τ_{ij}^k:

$$\frac{\partial^2 \psi_i^k}{\partial \tau_{ij}^k \partial \mu_i^k} = \frac{\partial(\pi_{ii}^k - \pi_{ij}^k)}{\partial \tau_{ij}^k}\frac{\partial \tilde{\eta}_i^k}{\partial \mu_i^k}$$

Notice that the first term is positive because $\frac{\partial \pi_{ii}^k}{\partial \tau_{ij}^k} > 0$ (a farmer is more likely to be its own cheapest supplier the higher is the trade cost) and $\frac{\partial \pi_{ii}^k}{\partial \tau_{ij}^k} < 0$ (a farmer is less likely to be the cheapest supplier to any other farmer the higher is the trade costs). As a result, $\frac{\partial^2 \psi_i^k}{\partial \tau_{ij}^k \partial \mu_i^k} > 0$, which completes the proof that the correlation between the land share of a crop and the household taste becomes stronger with increase in trade costs. □

D An alternative 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.[19] 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

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

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 A.3 reports the estimation results.

Table A.3 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).

The result implies an unambiguous rejection of separability – household demographic composition significantly affects household labor demand. This is consistent with the new test suggested in this paper. However, there are important differences in the two approaches. While any market incompleteness can lead to rejection of separability in the Benjamin's test, the test can be considered as a direct test of *missing* or *thin* labor markets. On the other hand, the approach suggested in the current paper can be considered as a direct test of *missing* or *thin* crop markets. In this sense, the two approached also complement each other. Also important is that the Benjamin's approach 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 since land area is measured by trained enumerators using GPS tools.

Table A.3: 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.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

E Appendix Tables

Table A.4: The effect of rainfall on village prices

	(1)	(2)	(3)
Log Rainfall	-0.087***	-0.087***	-0.087***
	(0.028)	(0.028)	(0.029)
Log GAEZ Yield		-0.025***	
		(0.003)	
Crop×YearFE	Yes	Yes	Yes
Village FE	Yes	Yes	.
Village ×CropFE	No	No	Yes
N	208324	208324	208324
R^2	0.809	0.813	0.920

Notes: Standard errors are clustered at village level. The regression includes 333 villages, and 20 crops. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

References

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.

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

Deaton, A. S. and Muellbauer, J. N. J. (1980). American economic association an almost ideal demand system. *The American Economic Review*, 70:312 – 326.

Donaldson, D. (2018). Railroads of the raj: Estimating the impact of transportation infrastructure. *American Economic Review*, 108(4-5):899–934.

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

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.

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*.

Hausman, J. A. (1994). Valuation of new goods under perfect and imperfect competition. Working Paper 4970, National Bureau of Economic Research.

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.

LaFave, D. R., Peet, E. D., and Thomas, D. (2020). Farm profits, prices and household behavior. (26636).

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 .

Singh, I., Squire, L., and Strauss, J. (1986). *Agricultural Household Models: Extensions, Applications, and Policy.*

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

Taylor, J. E. and Adelman, I. (2003). Agricultural household models: Genesis, evolution, and extensions. *Review of Economics of the Household*, 1(1):33–58.