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

Hundanol A. Kebede †

July 12, 2022

Abstract

I study to what extent farm household's production decisions are dictated by their consumption preferences – widely known as the separability hypothesis–and explore how this is related to market integration. My empirical approach is derived from a theoretical insight that if household production decision is independent of its consumption preferences, the household's tastes for different crops should not affect household land allocation across the crops, and the extent to which the crop tastes affect land allocation depends on the level of trade costs. I implement this test using a very rich household panel data on production and consumption from Ethiopia, which coincide with a period of massive rural road construction. I estimate household crop tastes from their preference function and show that these tastes significantly affect household land allocation across crops. This effect significantly decreases with proximity to markets and with improvement in market integration due to construction of new rural roads.

Keywords: Agricultural household models, Farm households, Market integration, Rural development, Rural roads, Separability, Trade costs.
JEL Codes: F14, H54, O12, O13 Q12, Q13, R12

*I am indebted to Kerem Cosar, Sheetal Sekhri, John McLaren and James Harrigan for their guidance in writing this paper. I would like to thank Jonathan Colmer and Sandip Sukhtankar for their helpful comments. This paper has also benefited from comments received from participants in UVA's Trade and Development workshops, WADES, PAC DEv, and MWIEDC conferences

†Assistant Professor, Department of Economics, Southern Illinois University Carbondale. Email: hundanol.kebede@siu.edu

1 Introduction

Poor agricultural sector productivity in low-income countries has been traditionally attributed to limited use of improved techniques (see for instance Suri, 2011). More recently, disparity in farm sizes and misallocation across farms have been suggested to explain substantial fraction of the agricultural productivity gap between developed and developing nations (Adamopoulos and Restuccia, 2014, Gollin and Udry, 2021). Another potentially important factor is that for farmers in developing countries, allocation of resources might be constrained by their own consumption needs, which is certainly not the case for farmers in advanced countries.

That is, distortion in farmers' allocation of plots across different crops to meet their own consumption needs could lead to inefficient utilization of resources such as land and lower agricultural productivity.[1] Absent these constraints, 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.[2] In this paper, I study to what extent smallholder farmers' production decisions are dictated by their own consumption preferences, instead of market forces, and the effects of improvement in market integration on the link between household consumption preferences and their production choices.

My empirical approach 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 affected by its own consumption 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. Thus, a decrease in

[1] In the context of Gollin and Udry (2021), this distortion could lead to dispersion of productivity within plots operated by a single farmer. Gollin and Udry (2021) find that as high as 70% of productivity dispersion is across plots operated by a single owner. But, they attribute this to measurement error and unobserved heterogeneity because they assume that there is no allocative inefficiency across plots of farm operated by a single owner.

[2] For instance, consider a situation where a farm household has to allocate land across several crops which constitutes its essential consumption bundles. If the farmer has to rely on its own production for its consumption needs due to some market failures, such as very high trade friction, then the farmer's allocation of plots across these crops is likely to be a function of, among others, the share of these crops in the farmer's consumption bundle and the productivity of the plots in these crops. This will lead to lower agricultural productivity compared to the situation where allocation of plots across these crops is dictated by market forces and land productivity.

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

I 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 to what extent household land allocation across crops is dictated by their crops tastes estimated in the first-stage, and explore how this effect varies across households of varying level of market access. I find that households' taste for a crop significantly affect the fraction of land allocated to the crop. 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 strong effect of household crop tastes on land allocation across crops has crucial implications for within-household resource misallocation and agricultural productivity.

Recent studies explore how misallocation of inputs across farm households in developing countries could explain the massive agricultural productivity gap between advanced and developing countries (Adamopoulos and Restuccia 2014, Gollin and Rogerson 2014, Adamopoulos et al. 2017, Restuccia and Rogerson 2017, Chen et al. 2017, Shenoy 2017, Gollin and Udry 2021, and Foster and Rosenzweig 2022). With the exception of Gollin and Udry (2021), most of these studies focus on factor market failures and misallocation of factors across farms in creating productivity dispersion across farms, which leads to lower aggregate productivity. Gollin and Udry (2021) document large within-farm productivity dispersion across plots, and attribute this to unobserved heterogeneity, seasonal shocks and measurement error. The results in the current paper provide an alternative explanation for within-farm productivity dispersion by showing how failure in the product (crop) market, due to high trade costs, could distort household resource allocation. That is, when household production decision is constrained by their consumption choices due to high trade costs, land and labor allocations are suboptimal (compared to when production decision is dictated by market forces) and there would be within-farm dispersion in productivity, which leads to lower agricultural productivity.

This paper is closely related to studies that empirically test the recursiveness hypothesis in farm household decisions (Benjamin 1992, LaFave and Thomas 2016, Dillon and Barrett 2017, and Dillon et al. 2019). The recursiveness hypothesis states that if markets are complete, household decisions can be modeled as recursive (Singh et al. 1986, Taylor and Adelman 2003). 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. These studies test recursiveness using the relationship between household on-farm labor demand and the household's demographic characteristics. If 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. More recently, LaFave et al. (2020) suggest a new consumption based test for recursiveness. 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.

There are at least two drawbacks to the labor market based tests of recursiveness. First, a positive correlation between a household's on-farm labor demand and its demographic characteristics is not necessarily evidence of market malfunctions. It is also consistent with a situation where a functioning labor market exists but unobserved shocks to household farm productivity and wealth may drive both labor demand and household demographic composition.[3] Second, the measurement of on-farm labor demand in the context of poor, uneducated and self-employing farm households has been a substantial challenge for the classic tests of recursivenness (LaFave et al., 2020). While the adoption of GPS tools in agricultural surveys has significantly reduced measurement errors in areas of farmlands (Carletto et al., 2013), the measurement of labor is still based on the household head's recall-based interviews on how many hours each member of the household worked on each plot from land preparation to harvesting.[4] The burden to recall and hence, the measurement error may increase with the household size and the number of plots operated.[5] This is a blow to the classic tests because household size is the main independent variable in their tests.

[3] For instance, some farm households may receive information about better agricultural techniques via development assistants (DAs), access to information technology (such as radio and TVs), or exposure to formal education of some household members, which may affect both labor demand and demographic characteristics of the household.

[4] However, GPS tools still cannot address the heterogeneity in quality of land, but no measure of labor (even if the farmers are able to keep their own daily record of labor input) would capture how much each member of household worked hard (labor quality) either.

[5] This is very similar to Beegle et al. (2012), who show that the measurement error in the case of consumption recall gets worse with the household size and number of consumption items included.

This paper suggests a product-market-based test of recursiveness. The advantage is that because product/crop markets have a known physical location, one can obtain a measure of households' proximity to crop markets and explore how recursiveness varies across households with different levels of access to markets. Such exercise is difficult to come by for labor market because there is no physical location for labor markets. Moreover, in the current approach, recursiveness is considered as a continuum, instead of dichotomous, concept measured by the strength of the correlation between household tastes and production choices. This is conducive to explore how the validity of recursiveness hypothesis varies across households of different characteristics such as farm size and access to market.

This paper also contributes to the literature on the development impact of rural roads. Asher and Novosad (2020) exploit strict implementation rule of India's massive rural road expansion project to identify the program's causal effect using fuzzy regression discontinuity design and 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 (2021) 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 (2020) 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 separation 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 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).[6]

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.[7] 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, a proxy for transport costs.

Rainfall and agro-climatic data: I use FAO/GAEZ agro-climatically attainable yield for low/intermediate input use to construct villages' crop suitability, which is used in the separability test and to test how road affects the relationship between local comparative advantage and local prices. Unfortunately the GAEZ data doesn't include some of the most widely grown crops in Ethiopia such as *Te* . 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

[6]The consumption information is based on a seven-day recall of basic consumption items, which are predominantly crops. Beegle et al. (2012) use experimental study in Tanzania to show that consumption survey design similar to the one used in the current data gives remarkably accurate statistics when compared to the experimental data.

[7]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.

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 62,000kms 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. Thus the survey round 2013 will be used as pre-road period and the survey round 2015 as post-road period. The same conclusions are reached if we use the survey round 2011 as pre-road period instead of 2013 survey.

2.2 Identification issues

One objective of this paper is to explore the link between market access and recursiveness using a large-scale rural road expansion under the URRAP. There are three challenges to identify the causal effects of the URRAP on the recursiveness: 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.[8] 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

[8] 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 the 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.

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 the 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 the 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 the ArcGIS tools to calculate average slope and elevation of each village.

To address the heterogeneity in treatment intensity and spillover effects, I use market 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 ensuing theoretical 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 in rural Ethiopia 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.[9]

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).[10] 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 a crop that they did not produce compared to rural households, and (2) households in small towns are more likely to consume vegetables and relatively more expensive cereals such as *Te* 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). 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

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

[10]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.

are connected to the rest of the country via all-weather roads.

Figure 2 presents the distribution of the share of consumption from purchase for rural households in each of the survey rounds. The analysis is restricted to rural households only because households in small towns are slightly different as highlighted in table 2. The figure presents the kernel densities of the share of consumption from purchase in villages that did not receive the URRAP roads and in villages that received the URRP roads. For both sets of villages, we see that there is significant shift towards consumption from purchase in 2015 (the period after the URRAP) compared to 2011 and 2013 (the period before the URRAP). However, comparing the two sets of villages, we see that the shift is stronger in villages that received roads under URRAP.[11] This is consistent with the hypothesis that these villages are direct beneficiaries from the road construction and the resulting improvement in market integration. The modest amount of shift towards consumption from purchase in the non-URRAP villages could be attributed to spillover benefits of roads to villages that are not directly connected by the URRAP but have seen their proximity to road improved via the connected villages.

An equally important issue is whether the road expansion had led to higher expenditure share on non-food items including manufactured goods, education, health, etc. To explore this, I calculate the share of non-food expenditure in household annual consumption and present the distribution in figure 3. The top panel presents the the kernel densities for non-URRAP villages while the bottom panel presents the kernel densities for households in URRAP villages. In both panels, the share of non-food expenditure shifts to the right in the years 2013 and 2015, compared to the year 2011, particularly in non-URRAP villages. However, there is no clear shift in 2015 compared to 2013 in both non-URRAP and URRAP villages. Overall, this figure implies that there is no significant change in the share of non-food expenditure over the survey rounds in both non-URRAP and URRAP villages. Because significant increase in the share of non-food expenditure is usually associated with income growth, the fact that we are not observing significant shift in expenditure towards non-food items suggests that there is no significant increase in real income of the households across the periods. Using the same survey, Kebede (2021) shows that there is no significant change in household real consumption expenditures across the survey rounds, once one takes into account the effects of local fluctuations in rainfall and local food price inflation.

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

[11] Kolmogorov–Smirnov test rejects equality of the distributions for the URRAP and non-URRAP villages for the year 2015 at a p-value 0.001.

is consumed within the household and only 13% is marketed. However, there is significant variation across crops.

Positive relationship between land and expenditure shares of crops: I use ESS data and focus on 19 crops for which complete information is available on both production and consumption. I calculate the share of each crop in household consumption expenditure[12] and the share of cultivated land allocated to the production of each crop. Cultivated land includes land owned and rented by the farmer. Inter-cropping poses a challenge in calculating the fraction of land allocated to each crop. Fortunately, the survey carefully accounts for inter-cropping (which accounts about 20% of the plots and 11% of the cultivated area in a given year). For plots where multiple crops are inter-cropped, the survey includes the fraction of the plot that is covered by each crop. I use this information to calculate the fraction of land allocated to each crop.

I run the following regression:

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

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. A significant positive effect of expenditure share of crop on land allocation across crops within a household is suggestive evidence against separability. That is, the null of separability holds if $\hat{\beta}_1 = 0$. Under autarky, near perfect correlation between household land and expenditure shares of crops is expected, $\hat{\beta}_1 \approx 1$. I run the regressions for each of the survey rounds separately to show how the estimated β_1 changed over time.

Table 4 reports the results. Column 1 reports the estimated relationship between the land and expenditure share for the three rounds of survey. To make the table concise, instead of running the regression for each year, the regression was run only once by including year dummies and interacting the year dummies with the expenditure share of crops. For the sake of brevity table 4 reports only the coefficient of the expenditure share of crop (the variable of interest) interacted with year dummies. The estimated coefficient is 0.47 for the year 2011, which slightly increases to 0.52 for the year 2013 before it decreases significantly to 0.24 for the year 2015. Panel B shows that both of these changes are statistically significant at 1% significance level. Though the coefficient in 2015 is significantly lower, it still is statistically

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

significantly different from zero, implying failure to reject separability in 2015.

In column 2, I run analogous regressions where I use data on *plot level* labor use (both planting and harvesting hours of labor) and 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. For plots with inter-cropping, I assume that the fraction of labor allocated to each crop is proportional to the fraction of the area of the plot allocated to each crop. Information on the latter is included in the survey. I then 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 shares of crops slightly increase from 0.45 to 0.51 between 2011 and 2013 before it significantly decreases to 0.24 for the year 2015. Again, both of these changes are statistically significant.

Note that, despite significant measurement concern of farm labor compared to farm land (because the former is measured based on household head's recall of how many hours each member of household worked on each plot throughout the season and the later is measured using GPS tools by trained enumerator), the coefficients of the expenditure share in columns 1 and 2 of table 4 are very similar. One potential reason for this is that the way the variables are constructed (share of land/labor allocated to each crop) coupled with the fact that the regressions exploit within-household variation across crops may go a long way to minimize the concern of measurement error. For example, if a household mismeasures the hours of labor allocated to each crop and the measurement error is proportional to the true measure of labor, it could be less consequential if our variable of interest is the *share* of labor allocated to each crop.

To sum up, these statistically significant effect of expenditure share of crops on household land and labor allocation across crops strongly suggest that household resource allocation is at least partially dictated by their consumption preferences. This suggests potentially large resource misallocation and loss in agricultural productivity because household production decision is constrained by their own consumption needs instead of being directed by market forces.

New roads decrease the effect of expenditure share on land allocation: Before I explore the effects of URRAP roads on the effect of expenditure shares of crops on land/labor allocation, I provide evidence on whether the URRAP roads indeed decreased trade costs and improved market integration. I use two indicators of market integration: rural-urban price gap for crops and the correlation between local yield and local prices of crops (see appendix B for details). In appendix table A.1, 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. In appendix table A.2, 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 effect of expenditure share of crops on land/labor allocation. 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, respectively.[13] 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 using the DID strategy discussed above:

$$\begin{aligned}\eta_{hvt}^{k} = \beta_0 + \beta_1 s_{hvt}^{k} + \beta_2 (Post_t \quad URRAP_v) + \beta_3 (s_{hvt}^{k} \quad Post_t \quad URRAP_v) \\ + \delta Z_{vt} + \gamma_t^{k} + \gamma_v + \varepsilon_{hvt}^{k}\end{aligned} \tag{2}$$

where $Post_t \quad URRAP_v$ 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 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. The first two columns present DID estimation results based on households in all villages whereas the last two columns restrict estimation to households in matched sample of treated and non-treated villages (note the decrease in the number of observations). 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 effect of expenditure shares of crops on household land/labor allocation across 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 0.17, compared to households in villages that were not directly exposed to the program. This is a large effect, roughly about one-third of the baseline correlation in 2013.

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

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

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, whether household's production decisions are independent of their consumption preferences – the separability hypothesis. In doing so, I borrow tools from the Ricardian trade models. Particularly, I build on Sotelo (2020).

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 from market.

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_{it}^k is the quantity of crop k consumed by household i in year t, and μ_i^k is the household taste for crop k, which is assumed to be fixed over the short to medium period (see, for instance, Atkin (2013)). 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 in year t, and Π_{it} is household farm profit in year t. In what follows, I drop the year subscript unless it is necessary.

Production: I follow Sotelo (2020) 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 ω Ω_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.[14] 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 $\tau_{vv}^k = 1, \quad k$, and impose the standard assumption of triangle inequality in trade costs, $\tau_{vv}^k \times \tau_{v\,v}^k \geq \tau_{vv}^k, \quad k$.

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 $\tau_{vv}^k << \infty, \quad k$. Suppose trade costs are such that farmers can buy and sell any crop at a prevailing market price. Assuming perfect competition, no arbitrage condition implies that for any two villages v and v, equilibrium crop prices satisfy $p_{vv}^k = \tau_{vv}^k p_{vv}^k$ where p_{vv}^k is price in village v of crop k originating from village v, and p_{vv}^k is price in village v of crop k originating from the same village v.

[14] 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.

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

$$\eta_i^k = h\Big(p_v^k, \mathbf{p}_v^{\mathbf{l}}, A_i^k, \mathbf{A}_i^{\mathbf{l}}; \theta, \boldsymbol{\alpha}^k, \boldsymbol{\alpha}^{\mathbf{l}}\Big) \tag{4}$$

where $\mathbf{l} = 1, ..., K = k$. Assuming that land is the only input and technology is constant returns to scale, we can obtain a closed form solution for the land allocation rule using the Fréchet distribution $\eta_i^k = \frac{(p_v^k A_i^k)^\theta}{\sum_l (p_v^l A_i^l)^\theta}$. This implies that the quantity of crops produced and revenue from each crop are given, respectively, by:

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

$$R_i^k = Y\Big(p_v^k, \mathbf{p}_v^{\mathbf{l}}, \mathbf{r}_v, A_i^k, \mathbf{A}_i^{\mathbf{l}}, L_i; \theta, \boldsymbol{\alpha}^k, \boldsymbol{\alpha}^{\mathbf{l}}\Big) \tag{6}$$

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

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

$$q_i{}^k = C\Big(\mu_i^k, \mu_i^{\mathbf{l}}, p_v^k, \mathbf{p}_v^{\mathbf{l}}, \Pi_i\Big) \tag{7}$$

Equations 4-7 show that: (i) household land allocation across crops and quantities of crops produced are independent of crop tastes μ_i^k, (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.

If there is failure in crop market, the above result no longer holds. To show this, we see the extreme case of autarky (for some crops) below.

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 $\tilde{p}_i^k$. The household maximizes its utility function subject to the budget constraint in equation 16 and the technology function, the solution of which can be written as:

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

where $l = 1, ..., K = k$. Assuming CES utility function, linear production function in land and using the Fréchet distribution, we obtain the following closed-form solution for the share of land allocated to each crop in autarky : $\tilde{\eta}_i^k = \frac{(\mu_i^k)^{-\frac{\sigma}{\sigma-1}}(\tilde{p}_i^k)^{\frac{\sigma}{\sigma-1}}}{\sum_k (\mu_i^k)^{-\frac{\sigma}{\sigma-1}}(\tilde{p}_i^k)^{\frac{-\sigma}{\sigma-1}}}$, where $0 < \sigma < 1$.[15]

In equation 8, the fraction of land allocated to crop k depends on the household taste for the crops; that is, the household production decision is not independent of its consumption preferences. This is a key result from which the recursiveness test is derived in this paper. In Appendix C, I show how this conclusion holds when we have labor as an additional input, regardless of whether labor market is missing or complete.

3.2 Trade costs and recursiveness

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

Equation 8 suggests that one can test recursiveness empirically by looking at whether household allocation of land across different crops is independent of the household's tastes for the crops, conditional on market prices of the crops and suitability of land for growing the crops. To implement this test, we need to first estimate household tastes for crops from their preference function.

[15] See Appendix C for the derivation.

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:

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

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

$$s_{ivt}^k = \mu_{iv}^k + \sum_k \gamma^{kk} \ln p_{vt}^k + \beta_k \ln \frac{m_{ivt}}{P_{vt}} \tag{10}$$

where $\gamma^{kk} = \frac{1}{2}(\gamma^{kk} + \gamma^{kk})$, m_i is household nominal expenditure on food, P_v is village price index, and $\frac{m_{it}}{P_{vt}}$ is real expenditure. Having real expenditure in this specification is important to allow for non-homotheticity and to account for the potential effect of change in income on change in consumption patterns. Following Deaton and Muellbauer (1980) and Atkin (2013), I use Stone index for village price index, $\ln P_{vt} = \sum_k \bar{s}_v^k \ln p_{vt}^k$, where $\bar{s}_v^k$ is the average expenditure share of crop k in village v.

Household crop tastes μ_{iv}^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.6 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, conditional on taste differences, across rural households in Ethiopia that is approximated by AIDS function.

I estimate the following equation to identify household crop tastes:

$$s_{ivt}^k = \mu_{iv}^k + \sum_k \gamma^{kk} \ln p_{vt}^k + \beta^k \ln \frac{m_{ivt}}{P_{vt}} + \alpha_y \ln y_{ivt}^k + \alpha_n N_{ivt} + \delta_t + \varepsilon_{ivt}^k \tag{11}$$

where N denotes household size and other demographic characteristics, δ_t is year fixed effects and ε_{ivt}^k is the error term. $\ln y_{ivt}^k$ is household-level measure of crop yield, and it is included as a proxy for shadow price of consumption of a crop that may vary at a household level.[16] 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. Following Atkin (2013), I address this concern by instrumenting village prices by prices in the nearest villages. Prices are spatially correlated over shorter distances mainly due to trade. The exclusion restriction assumption behind this IV strategy is that a household's expenditure share on crop k in village v, s^k_{ivt}, is affected by the price of crop k in village v , $p^k_{iv\ t}$ (where v is nearest village to village v), only through the effect of $p^k_{iv\ t}$ on p^k_{ivt}. This is quite reasonable assumption to make and is similar to Hausman-type instrumental variables (Hausman, 1994).

Note that, zero expenditure shares for some households in some crops are common in the data. However, the AIDS satisfies homogeneity and symmetry when all households have a positive consumption of all crops. I thus follow Deaton (1997) and Atkin (2013), and interpret equation 11 as the conditional expenditure share averaging over households with zero and non-zero expenditure shares. Under this assumption, zero expenditure shares pose no problem in our estimation of tastes. If the household has zero expenditure share on some crops, given the vector of market prices and other controls, it implies that the household has lower taste for these crops compared to other crops with positive expenditure shares.

As a robustness check, I also estimate a specification where households within a village share the same crop tastes, i.e., $\mu^k_{iv} = \mu^k_v, \ \ i \quad v$. 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 shown below, the empirical result strongly supports the conjecture that households within a village largely share similar crop tastes – about 81% of variation in household crop tastes comes from across village variation.

4.2 Testing separability

Once I obtain estimates of household crop tastes, I test the separability hypothesis by looking at whether the fraction of land allocated to different crops is independent of the household crop tastes, conditional on village crop prices and yields (agro-climatic

[16] One concern in estimation of tastes in equation 11 is that some households might face a shadow price that is different from the market price. For instance, if a household can produce a crop more cheaply than the market price, the household would face lower consumption price for that crop than the market price. This would bias the taste estimates in the same way as measurement error in prices. To partially address this concern, I control for household level yield estimates in equation 11, where this yield estimate is replaced with zero for a crop that the household does not produce.

suitability of village for each crop). I estimate the following regression:

$$\eta_{ivt}^k = \beta_0 + \beta_1\mu_i^k + \beta_2\text{lnp}_{vt}^k + \beta_3\text{ln}y_v^k + \beta_4^k\text{lnRainfall}_{vt} + \beta_5\text{ln}\frac{m_{it}}{P_{vt}} + \gamma^k + \gamma_t + \gamma_i + \epsilon_{ivt}^k \tag{12}$$

where η_{ivt}^k is the fraction of land allocated to crop k. I include the household's real expenditure $\text{ln}\frac{m_{it}}{P_{vt}}$ to account for the potential effect of income changes on the household's production patterns. 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 should 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:

$$\begin{aligned}\eta_{ivt}^k = \beta_0 + \beta_1\mu_i^k + \beta_2\text{MA}_{vt} + \beta_3(\mu_i^k \times \text{MA}_{vt}) + \beta_4\text{lnp}_{vt}^k + \beta_5\text{ln}y_v^k \\ + \beta_6^k\text{lnRainfall}_{vt} + \beta_7\text{ln}\frac{m_{it}}{P_{vt}} + \gamma^k + \gamma_t + \gamma_i + \epsilon_{ivt}^k\end{aligned} \tag{13}$$

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 for those villages that got direct road connectivity under the program (see appendix A for the detailed procedure followed in constructing MA measures). As in equation 12, I include the household's real expenditure $\text{ln}\frac{m_{it}}{P_{vt}}$ to account for the potential income effect of change in road infrastructure on the household's production patterns. 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 taste estimates; the correlation between the taste estimates obtained from these approaches is about 0.96. Overall, the IV passes the under-identification and weak identification tests remarkably and borderline passes the weak instrument test with first-stage F-statistics of about 10. See table A.7 for the first-stage diagnosis tests. 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. The taste parameters estimated at household and village levels have a correlation of about 0.90. Table A.8 reports the weighted mean (where the weights are sampling weights of households) of the estimated household crop tastes (from the IV strategy) across the Ethiopian regions. Looking into the geographic variation in estimated tastes reveals that the estimated taste parameters strongly align with some notable local cultural foods. For instance, Maize is used to make the staple food called porridge in Gambella, Enset is the main ingredient in traditional Kocho food in SNNP, and Sorghum is used to make most of traditional foods in rural Dire Dawa and Harari areas. The estimated taste parameters clearly show that (see table A.8). Interestingly, the taste parameters do not necessarily reflect production patterns across regions. For instance, chat consumption is very popular in Somali region, but it is predominantly imported from a neighboring region of Oromia.

Testing separability: Next, I explore how the estimated taste parameters affect 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. For brevity, instead of running the regression for each year, I run only one regression by including year dummies and interacting the year dummies with taste, and I report only the coefficient of taste for each year. In Table 6 and all the subsequent regressions where the independent variable includes taste parameters, the standard errors are calculated using the method of bootstrapping clustered at household level to account for the fact that the taste parameters are estimated.

The first columns of table 6 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 recursiveness hypothesis. The coefficient of taste slightly varies across years. Focusing on the results based on IV taste in the second column, we observe that the coefficient of taste slightly increases from 0.573 in 2011 to 0.595 in 2013, before it slightly decreases to 0.583 in 2015. The increase in the coefficient of taste between 2011 and 2013 is statistically significant but decline between 2013 and 2015 is not (see panel B of Table 6). Below, we will see that the effect of taste on land allocation decreases more significantly in villages that experienced larger increases in their market access due to the road expansion.

A dimension of heterogeneity that is not captured in the theoretical model is that farm households with different levels of landholding may face different degree of constraints in crop markets and thus act differently. While the data used in this paper covers predominantly smallholder farmers, there is still significant variation in the size of land cultivated across households. For instance, the median farm size in the sample is just under a hectare while the 99th percentile rank is about 6 hectares. I explore how the correlation between the land share of crops and tastes varies across farmers of different size by interacting taste with the land size and its square in equation 12. The results are shown in Table A.3 and Figure A.1. The results in Table A.3 show a positive coefficient for the interaction of taste with the land size and a negative coefficient for interaction of taste with the square of the land size, both statistically significant at 1%. To present a clear picture of this non-linearity, I first calculate the correlation between tastes and land share of crops for each household and plot the marginal effects from the Kernel regression of this correlation on household land size. The plot of the marginal effects in Figure A.1 shows that the effect of land size on the correlation between tastes and land share of crops is $\cap-$shaped; smaller and larger farmers tend to have weaker correlation between the land share of crops and the crop tastes compared to medium-sized farmers. However, looking more closely into the data shows that the smaller and larger farmers differ in a crucial manner. Smaller farmers grow fewer crops and are more likely to consume a crop that they do not produce, perhaps because they do not have enough land to grow many crops. Larger farmers tend to grow more number of crops, but their allocation of land across crops is not as strongly aligned with their consumption tastes as it is for medium-sized farmers, implying that larger farmers somewhat act as commercial farmers.

Overall, the significant effect of household taste on the fraction of land allocated to different crops, conditional on market prices and yield, implies that household resource allocation is distorted. That is, household resource allocation is constrained by their own consumption preferences, which leads to sub-optimal allocation compared to the situation where household production choices are made purely based on market prices and productivity of their land. This result has important implication for studies that estimate misallocation in agriculture. For instance, Gollin and Udry (2021) implicitly assume that households' production decision is independent of their own consumption preferences. Consequently, they interpret the observed within-farm dispersion in productivity (which accounts for about three-quarters of the overall productivity dispersion) as measurement error and unobserved heterogeneity, because they assume that there is no allocative inefficiency across plots operated by a single farmer. But, if farmers' allocation of plots across different crops is dictated by the farmer's tastes for these crops, one can have dispersion in measured productivity across plots operated by a single farmer.

5.2 Separability and proximity to market

Before turning to the effect of road expansion under URRAP, I first explore how the effect of taste on land allocation 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 measure 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 effect of taste on land allocation, one would expect that this effect would be stronger for household that live further from towns or roads. Table 7 reports the results. Panel A shows that the effect of taste on land allocation significantly increases with distance to nearest population center. Using the result in the second column and the range of log distance to population center of about 6, the correlation between land allocation and taste ranges from about 0.33 for the nearest to 0.69 for the furthest household to population center.[17] Using the OLS taste estimates in the first column gives similar conclusion. Panel B reports similar results 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

[17] This result uses the fact that the furthest household from the nearest population center has distance of 6 log units and the nearest household has distance of zero log units.

towns.[18] Overall, the results in table 7 clearly indicate that lack of access to market is one of the major driving factors in the correlation between household production decision and their tastes. However, access to market does not fully explain why production decision is correlated with tastes in the sense that the correlation between land allocation and taste is positive and economically significant even for households that leave near market centers (towns) and roads.

5.3 The effect of URRAP on separability

I now 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 DID estimation. Figure 5 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 effect of tastes on land allocation. The first two columns show that the correlation between the land share of crops and crop tastes decreases by 0.035 - 0.04 for villages that got direct road connection under URRAP compared to the control villages, depending on the taste estimates used in the regression. Similarly, columns 3 and 4 show that one standard deviation increase in market access leads to 0.03 - 0.034 decrease in the correlation between the land share of crop and the crop tastes. That is, if we compare a village with average increase in market access against a village with maximum increase (three time the standard deviation), the correlation between land allocation and taste decreases significantly more by 0.075 - 0.09 in villages with the maximum increase in market access. 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 slightly larger in the first two columns.

Overall, the results in tables 9 and 10 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

[18] 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.

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.

However, note that recursiveness is still rejected even in the villages that get road connection (for villages that experience the largest increase in market access) in the sense that the correlation between land allocation and tastes did not decrease to zero in these villages. That is, the road expansion only loosens the link between household tastes and production decisions – it does not reduce the correlation between land allocation and taste to zero. This might suggest that trade costs did not decrease enough to achieve full separability of production decision from consumption tastes, or that trade costs are not the only reasons why production is dependent on consumption tastes, or both. One may also 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.

5.4 Comparison with Benjamin-style tests of separability

In this subsection, I compare the results obtained above with Benjamin-style tests of recursiveness. Appendix D presents details about the Benjamin-style tests using data for Ethiopia and following the specifications in LaFave and Thomas (2016). According to this approach, if 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 results in table A.4 clearly show that household demographic characteristics significantly influence the household's on-farm labor demand, implying that the assumption of complete markets is not warranted in the context of rural Ethiopia.[19] Furthermore, table A.5 shows that proximity to markets and roads, and the massive expansion of roads under URRAP have important effects on the correlation between on-farm labor demand and household demographic characteristics. The correlation between on-farm labor demand and number of prime age male in the household increases with distance from towns, and decreases in villages that see improvement in market access due to URRAP.

Though, in principle, any missing markets (e.g., credit markets, labor markets, insurance markets, etc) could be consistent with the significant correlation between on-farm labor demand and household demographic characteristics, one can argue that unobserved shocks to household agricultural productivity and wealth or exposure to knowledge about both labor-saving agricultural technique and family planning

[19] Dillon et al. (2019) find similar results for multiple countries in sub-Saharan Africa, including Ethiopia.

through trained development assistants (DAs) and/or information technology (such as radio and TVs) could lead to changes in both household labor demand and family composition.

Crop markets are relatively well developed and almost all farmers in the Ethiopian data engage in crop markets both as buyers and sellers of different crops. In view of this fact, it is relatively surprising to find that household's allocation of land across different crops is strongly influenced by the household's own tastes of these crops. This suggests that the existence of crop markets is not sufficient to make household's crop production decision independent from their consumption tastes. The evidence presented above show that this is at least partially attributed to very poor road infrastructure and the resulting high trade costs in the areas. Households who are located closer to roads or market centers tend to be relatively more reliant on markets for their consumption needs and thus have weaker correlation between production decision and consumption tastes.

5.5 Implications of failure of recursiveness

Failure of recursiveness has crucial implications about within-household resource misallocation and agricultural productivity.[20] It implies that agricultural inputs, such as land and labor, are allocated across different crops based on the household's own consumption needs, instead of market forces such as market prices and suitability of land to different crops. This results in dispersion in productivity across plots operated by the same household, and lower agricultural productivity at a household level. This is analogous to how dispersion in revenue productivity across firms within an industry implies lower industry and aggregate productivity in Hsieh and Klenow (2009), except that, here, failure in the crop market results in within-household (within-'firm') dispersion in productivity across plots. This margin of productivity dispersion and misallocation is different from the productivity dispersion and resource misallocation across farm households widely documented in the literature (see for, instance, Adamopoulos and Restuccia 2014, Adamopoulos et al. 2017, Chen et al. 2017, Shenoy 2017, Gollin and Udry 2021, and Foster and Rosenzweig 2022). Future studies that quantify the magnitude of productivity loss from within-farm resource misallocation, and how these factors vary with improvements in market integration would be informative.

[20]See Jones et al. (2022) for a recent evidence on how labor market failures affects households' adoption of irrigation across different plots in Rwanda.

6 Conclusions

Whether farm household's production and consumption decisions can be sequentially 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, and this measurement error is likely to be correlated with household demographic characteristics, such the number of active-age adults in the household.

In this paper, I suggest alternative test to investigate to what extent farm household's production decisions are dictated by their consumption preference and explore how this is related to market integration. My empirical approach 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.

The finding that household land allocation across crops is strongly dictated by its own consumption preferences suggests that there is likely massive distortion in the allocation of land and labor within a farm household. That is, households may achieve higher living standard if their resource allocation is made based on market forces and unconstrained by their own consumption preference. Market integration via improvement in road infrastructure may improve efficient allocation of land and labor by households by loosening the influence of consumption preference on production decisions.

Figure 1: Rural road expansion under URRAP

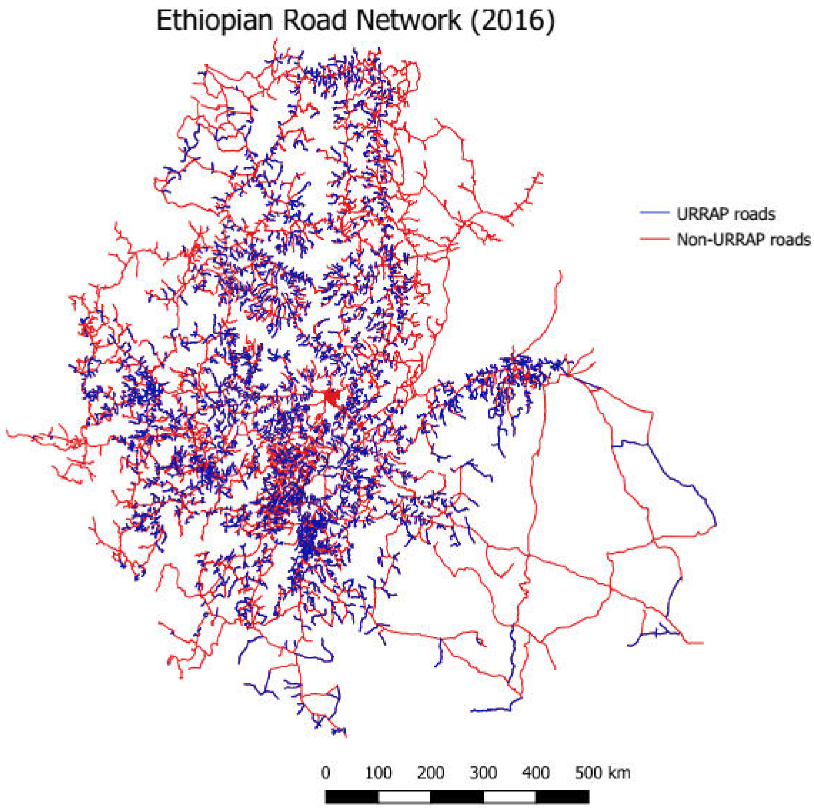


Figure 2: Share of consumption from purchase

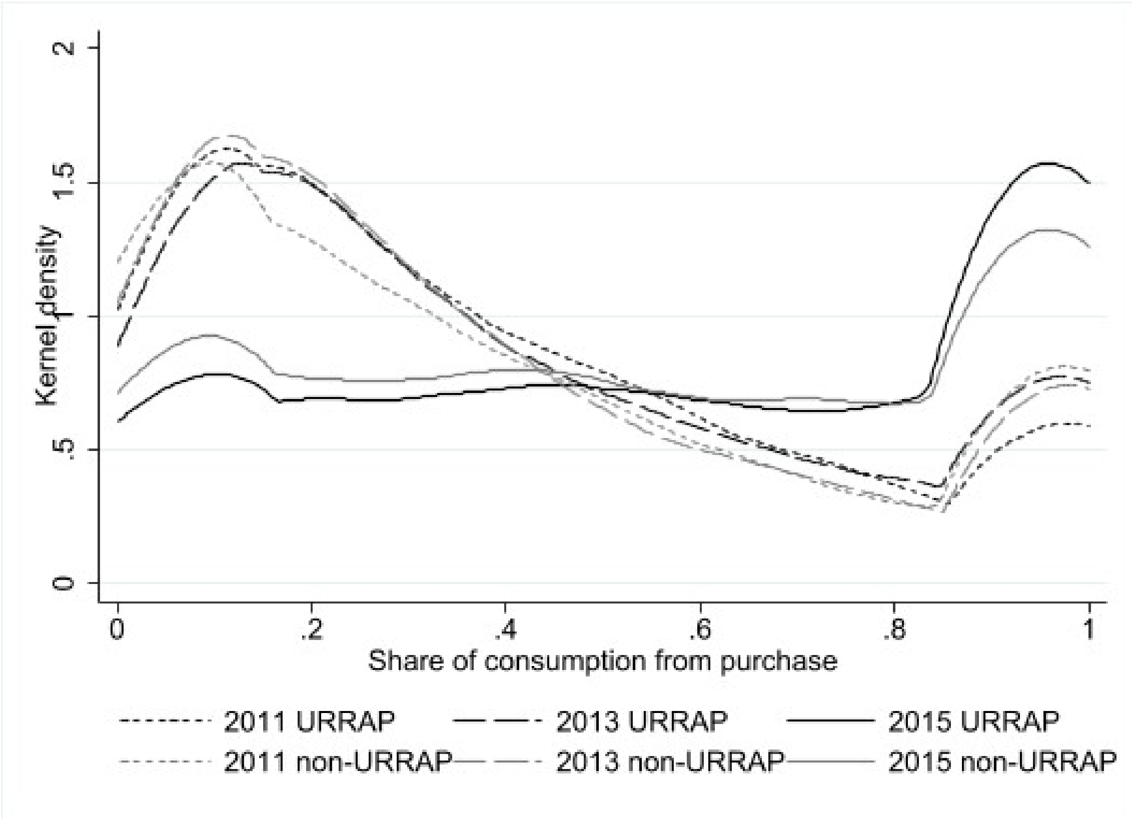


Figure 3: Share of non-food expenditure in total consumption

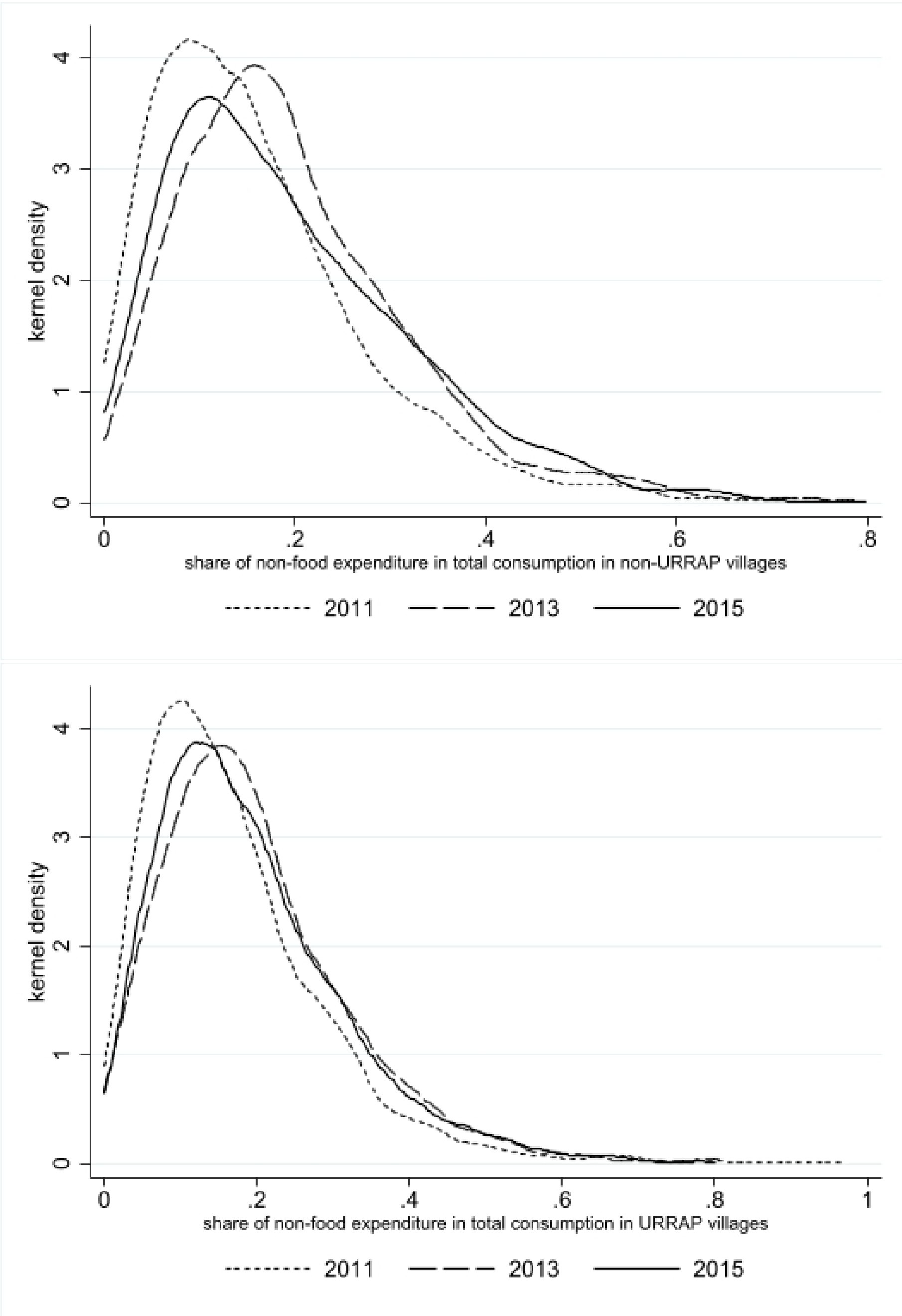


Figure 4: Completed URRAP roads (pictures are taken from Oromia Roads Authority).

Figure 5: Common support of propensity score matching

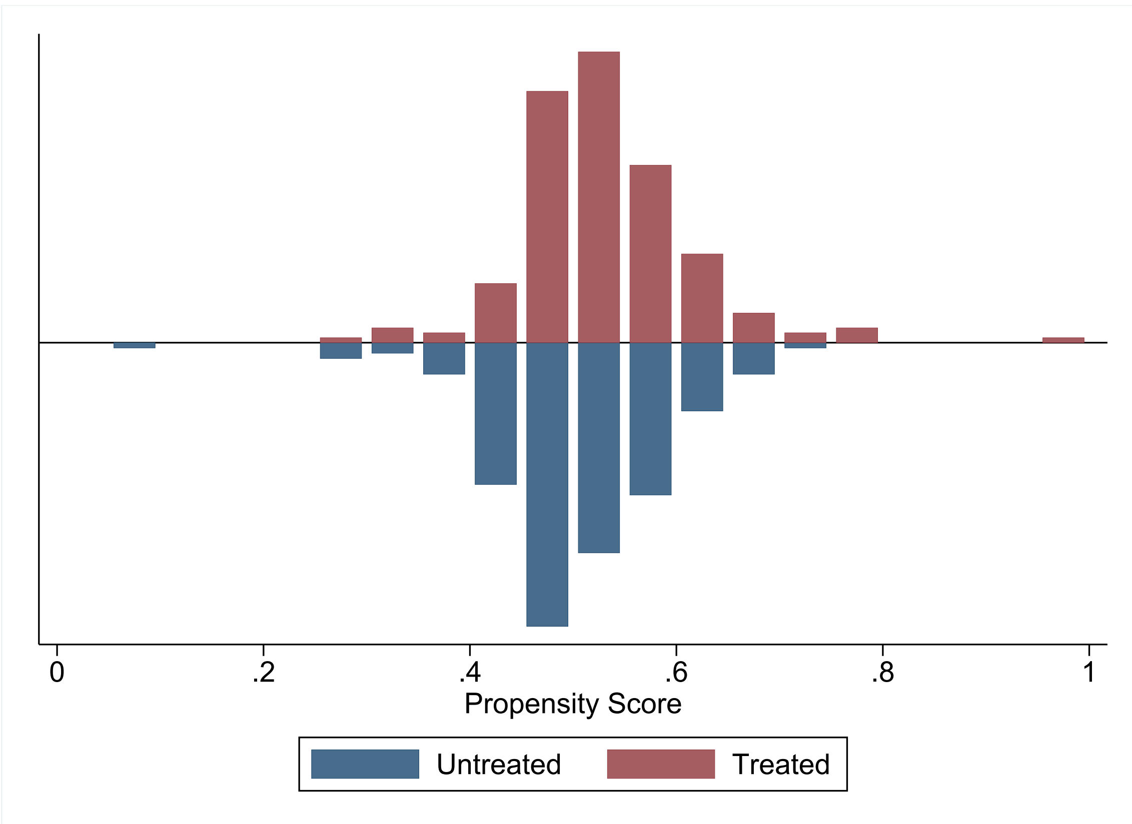


Table 1: Transport modes to market, proximity to market, and trade costs

	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. The columns do not necessarily add up to 100 because there are other crop utilization including animal feed, use for in-kind wage and other payments, gifts, lost crops, etc.

Table 4: Correlation between household production and consumption decisions

	Land allocation	Labor allocation
	Panel A: Correlation	
Expenditure Share*2011	0.469***	0.454***
	(0.022)	(0.021)
Expenditure Share*2013	0.520***	0.508***
	(0.022)	(0.022)
Expenditure Share*2015	0.239***	0.235***
	(0.016)	(0.016)
N	158173	158169
R^2	0.286	0.288
	Panel B: Testing equality of coefficients	
Expenditure Share*2011=Expenditure Share*2013	8.47	9.78
	(0.004)	(0.002)
Expenditure Share*2011=Expenditure Share*2015	112.6	107.8
	(0.000)	(0.000)
Expenditure Share*2013=Expenditure Share*2015	207.16	202.92
	(0.000)	(0.000)
Expenditure Share*2011=Expenditure Share*2013 =Expenditure Share*2015	104.83	102.86
	(0.000)	(0.000)

Notes: Standard errors clustered at household level in are parentheses in Panel A. P-values are in parentheses in Panel B. The dependent variable is the share of land allocated to different crops. In column 1, the dependent variable is the share of household land allocated to each crop, while in column 2 it is the share of labor allocated to each crop. Expenditure share is the share of household expenditure allocated to each crop. All regressions include control variables of village crop prices and yields (interacted with survey year dummies), household fixed effects, year fixed effects and crop fixed effects interacted with rainfall measure. Panel B reports test statistics and p-values for testing the equality of the correlation across years. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

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

	DID estimation		Matching based DID estimation	
	Land	Labor	Land	Labor
Expenditure Share	0.404	0.396	0.421	0.417
	(0.020)	(0.021)	(0.023)	(0.024)
Post*URRAP	0.007	0.008	0.007	0.008
	(0.001)	(0.001)	(0.002)	(0.002)
Expenditure Share*Post*URRAP	-0.164	-0.165	-0.169	-0.177
	(0.025)	(0.025)	(0.028)	(0.028)
N	107558	107554	81902	81900
R^2	0.268	0.268	0.264	0.265

Note: Standard errors are clustered at household level. The dependent variable is the share of land allocated to different crops. In columns 1 and 2, I pool households in all rural villages whereas in columns 3 and 4, I restrict estimation to households in matched sample of villages using the matching criteria for treated and non-treated villages discussed in section 2. All regressions include control variables of village crop prices and yields (interacted with survey year dummies), household fixed effects, year fixed effects and crop fixed effects interacted with rainfall measure. $p < 0.10$, $p < 0.05$, $p < 0.01$

Table 6: The separability test

	OLS Taste	IV Taste
	Panel A: Testing separability	
Taste*2011	0.571	0.573
	(0.013)	(0.013)
Taste*2013	0.591	0.595
	(0.012)	(0.012)
Taste*2015	0.585	0.583
	(0.014)	(0.014)
N	153293	153293
R^2	0.298	0.300
	Panel B: Testing equality of the coefficients of taste across years	
Taste*2011=Taste*2013	4.62	6.10
	(0.031)	(0.014)
Taste*2011=Taste*2015	2.04	0.95
	(0.154)	(0.33)
Taste*2013=Taste*2015	0.36	1.9
	(0.546)	(0.17)
Taste*2011=Taste*2013 =Taste*2015	4.67	6.36
	(0.097)	(0.042)

Note: Bootstrap standard errors with 500 replications clustered at household level are in parentheses in Panel A. P-values are in parenthesis in Panel B. The dependent variable is the share of land allocated to different crops. Column 1 uses crop tastes estimated without instrumenting for prices. Column 2 uses crop tastes estimated by instrumenting for prices using prices from neighboring village. All regressions include the control variables of village crop prices and yields (interacted with survey year dummies), household fixed effects, year fixed effects and crop fixed effects interacted with rainfall measure. The second panel reports Chi-square test statistics together with the p-values for the test of equality of the coefficient of taste across the years. $p < 0.10$, $p < 0.05$, $p < 0.01$

Table 7: Separability and proximity to market

	Panel A: Distance to Population center	
	OLS Taste	IV Taste
Taste	0.242	0.328
	(0.040)	(0.039)
Log Dist. to Pop. Center	-0.001	-0.003
	(0.001)	(0.001)
Taste*Log Dist. to Pop. Center	0.083	0.062
	(0.009)	(0.009)
N	153293	153293
R^2	0.301	0.301
	Panel B: Distance to Road	
Taste	0.483	0.485
	(0.019)	(0.019)
Log Dist. to Road	-0.000	-0.001
	(0.000)	(0.000)
Taste*Log Dist. to Road	0.035	0.035
	(0.005)	(0.005)
N	153293	153293
R^2	0.300	0.301

Note: Bootstrap standard errors with 500 replications clustered at household level are in parentheses. The dependent variable is the share of land allocated to different crops. All regressions include the control variables of village crop prices and yields (interacted with survey year dummies), household fixed effects, year fixed effects and crop fixed effects interacted with rainfall measure. $p < 0.10$, $p < 0.05$, $p < 0.01$

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	0.627	0.627	0.626	0.628
	(0.015)	(0.015)	(0.015)	(0.015)
Post*URRAP	-0.001	-0.000		
	(0.000)	(0.000)		
Taste*Post*URRAP	-0.040	-0.035		
	(0.013)	(0.013)		
Market Access			-0.001	0.000
			(0.000)	(0.001)
Taste*Market Access			-0.034	-0.030
			(0.014)	(0.014)
N	77872	77872	77109	77109
R^2	0.297	0.298	0.296	0.298

Note: Bootstrap standard errors with 500 replications clustered at household level are in parentheses. The dependent variable is the share of land allocated to different crops. All regressions include the control variables of village crop prices and yields, household fixed effects, year fixed effects and crop fixed effects interacted with rainfall measure. The market access measure is standardised so that the coefficient can be interpreted as the effect of one standard deviation increase in market access. See appendix A for detailed discussion on the construction of the market access measure. The drop in the number of observations in the last two columns is due to some villages missing data on calculated market access. $p < 0.10$, $p < 0.05$, $p < 0.01$

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

	Binary treatment		Market access approach	
	OLS Taste	IV Taste	OLS Taste	IV Taste
Taste	0.593	0.593	0.610	0.612
	(0.013)	(0.013)	(0.014)	(0.013)
Post*URRAP	-0.001	0.001		
	(0.000)	(0.000)		
Taste*Post*URRAP	-0.051	-0.043		
	(0.013)	(0.012)		
Market Access			-0.001	-0.000
			(0.000)	(0.000)
Taste*Market Access			-0.035	-0.030
			(0.013)	(0.013)
N	102552	102552	98029	98029
R^2	0.295	0.296	0.293	0.295

Note: Bootstrap standard errors with 500 replications clustered at household level are in parentheses. The dependent variable is the share of land allocated to different crops. All regressions include the control variables of village crop prices and yields, household fixed effects, year fixed effects and crop fixed effects interacted with rainfall measure. The market access measure is standardised so that the coefficient can be interpreted as the effect of one standard deviation increase in market access. See appendix A for detailed discussion on the construction of the market access measure. The drop in the number of observations in the last two columns is due to some villages missing data on calculated market access. $p < 0.10$, $p < 0.05$, $p < 0.01$

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 villages in Ethiopia:

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

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 is estimated in Kebede (2020) using data from rural Ethiopia. I use $\hat{\theta} = 2.7$, which is the estimated value in the preferred specification of Kebede (2020).

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 $\frac{Cost\ along\ earth\ road}{Cost\ along\ minor\ gravel}$ to obtain estimate of cost along the links.[21] 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

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

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}.

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:

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

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.

URRAP decreases the correlation between local prices and yields: Another 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

Table A.1: URRAP road access and trade costs

	Dependent Variable: log(Price in Zone Capital/Price in village)	
	All crops	Vegetables and Fruits
$Post_t \quad URRAP_v$	-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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

between local prices and local yield. I run the following generalized DID regression to investigate this:

$$\begin{aligned} lnP_{vt}^k &= \alpha_1 lnA_v^k + \alpha_2(Post_t \quad URRAP_v) + \alpha_3(lnA_v^k \quad Post_t \quad URRAP_v) \\ &\quad + \gamma_v + \gamma_k + \gamma_t + \varepsilon_{vt}^k \end{aligned}$$

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%.[22]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.

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

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

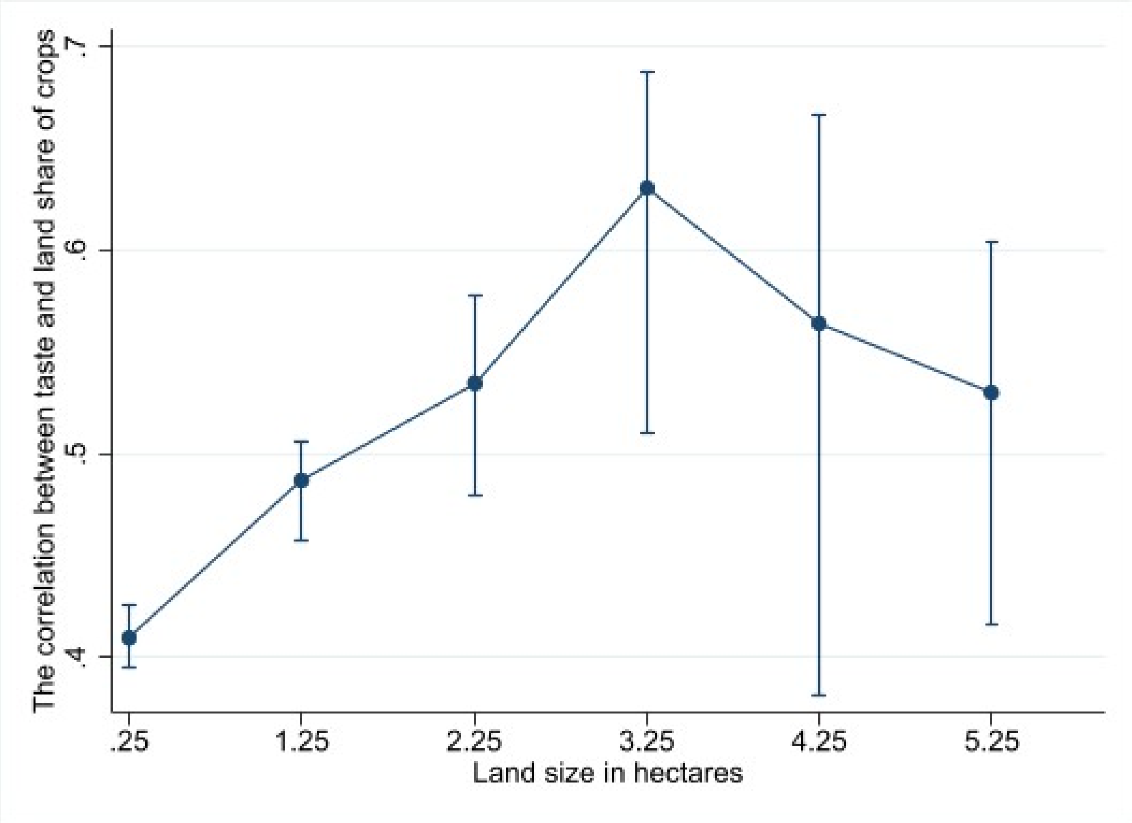
Notes: Standard errors are clustered at village level. The regression includes 277 villages, and 20 crops. All regressions include crop and year fixed effects, and log rainfall as a control. The last column includes village fixed effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.3: Heterogeneity in the effect of taste on land allocation

	OLS Taste	IV Taste
Taste	0.487	0.482
	(0.016)	(0.015)
Land size	-0.000	-0.006
	(0.000)	(0.001)
Taste*Landsize	0.176	0.187
	(0.016)	(0.016)
Landsize *Landsize	-0.000	0.001
	(0.000)	(0.000)
Taste*Landsize*Landsize	-0.032	-0.034
	(0.004)	(0.004)
N	151643	151643
R^2	0.303	0.305

Note: Bootstrap standard errors with 500 replications clustered at household level are in parentheses. The dependent variable is the share of land allocated to different crops. Column 1 uses crop tastes estimated without instrumenting for prices. Column 2 uses crop tastes estimated by instrumenting for prices using prices from neighboring villages. All regressions include the control variables of village crop prices and yields (interacted with survey year dummies), household fixed effects, year fixed effects and crop fixed effects interacted with rainfall measure. Land size is measured in hectares. The regression drops farm sizes higher than the 99th percentile. $p < 0.10$, $p < 0.05$, $p < 0.01$

Figure A.1: Heterogeneous effect of taste on land allocation across crops



Note: This figure plots the marginal effects from Kernel regression of the correlation between the land share of crops and crop tastes on household land size. The median land size is approximately 1 hectare. The vertical lines represent the confidence interval for the marginal effects at the corresponding land size. The confidence intervals get wider as land size increases because the number of households significantly drops with land size.

C Derivations and proofs

Derivation of land allocation rule under autarky: Suppose the household's preferences over different crops is given by the following CES function:

$$U_i = \Big[\sum_k (\mu_i^k q_i^k)^\sigma\Big]^{1/\sigma} \quad \text{where} \quad 0 < \sigma < 1 \tag{15}$$

where μ_i^k are crop tastes and q_i^k are quantities consumed.

The household budget constraint is given by:

$$\sum_k \tilde{p}_i^k q_i^k \leq \tilde{\Pi}_i \tag{16}$$

where $\tilde{p}_i^k$ is shadow price of crop k for household i, and $\tilde{\Pi}_i$ is shadow profit.

Assume a constant returns to scale production technology with land as the only input and each plot of land has a size of unity. Output from plot ω (assuming a plot is used for one crop at a time) is given by

$$y_i^k(\omega) = z_i^k(\omega) \tag{17}$$

where $y_i^k(\omega)$ is the quantity of crop, and $z_i^k(\omega)$ is the productivity of plot ω in crop k, which is randomly drawn from a Fréchet distribution as explained in the main text.

The household maximizes its utility function subject to the budget constraint and production technology. The Lagrangean function for this optimization problem is

$$L = \Big[\sum_k (\mu_i^k q_i^k)^\sigma\Big]^{1/\sigma} - \lambda\Big[\sum_k \tilde{p}_i^k q_i^k - \tilde{\Pi}_i\Big] \tag{18}$$

Solving this gives the following for optimal quantity demanded:

$$q_i^k = \frac{(\mu_i^k)^{-\frac{\sigma}{\sigma-1}} (\tilde{p}_i^k)^{\frac{1}{\sigma-1}}}{\sum_m (\mu_i^m)^{-\frac{\sigma}{\sigma-1}} (\tilde{p}_i^m)^{\frac{-\sigma}{\sigma-1}}} \tilde{\Pi}_i \tag{19}$$

Turning to the production side, the farmer allocates land across crops to maximize the return from each plot using the shadow prices as value of crops; i.e., each plot is allocated to the crop that gives the highest return:

$$r_i(\omega) = \max_k \{\tilde{p}_i^k z_i^k(\omega)\} \tag{20}$$

Using the Fréchet distribution, this implies the following closed-form solution for optimal land allocation from the supply side:

$$\tilde{\eta}_i^k = \frac{(\tilde{p}_i^k A_i^k)^\theta}{\tilde{\Phi}_i} \tag{21}$$

where $\tilde{\eta}_i^k$ is the fraction of land allocated to crop k and $\tilde{\Phi}_i = \left[\sum_m (\tilde{p}_i^m A_i^m)^\theta\right]^{1/\theta}$, which can be interpreted as the average return or revenue per plot of land. It can be shown that the optimal quantity of crop k produced (supplied), y_i^k, is given by:

$$y_i^k = \tilde{\eta}_i^k L_i \frac{\tilde{\Phi}_i}{\tilde{p}_i^k} \tag{22}$$

where L_i household land endowment and $\frac{\tilde{\Phi}_i}{\tilde{p}_i^k}$ is the average productivity of land allocated to crop k.[23]

Equilibrium in autarky requires that $q_i^k = y_i^k$. Thus, combining equations 19 and 22 to solve for $\tilde{\eta}_i^k$, we obtain:

$$\tilde{\eta}_i^k = \frac{(\mu_i^k)^{-\frac{\sigma}{\sigma-1}} (\tilde{p}_i^k)^{\frac{\sigma}{\sigma-1}}}{\sum_m (\mu_i^m)^{-\frac{\sigma}{\sigma-1}} (\tilde{p}_i^m)^{\frac{-\sigma}{\sigma-1}}} \frac{\tilde{\Pi}_i}{L_i \tilde{\Phi}_i} \tag{23}$$

Using the fact that $\sum_k \tilde{\eta}_i^k = 1$, we can rewrite the above as

$$\tilde{\eta}_i^k = \frac{\tilde{\eta}_i^k}{\sum_m \tilde{\eta}_i^m} = \frac{(\mu_i^k)^{-\frac{\sigma}{\sigma-1}} (\tilde{p}_i^k)^{\frac{\sigma}{\sigma-1}}}{\sum_m (\mu_i^m)^{-\frac{\sigma}{\sigma-1}} (\tilde{p}_i^m)^{\frac{\sigma}{\sigma-1}}} \tag{24}$$

This shows that under autarky, the fraction of land allocated to a crop is a function of the household's crop tastes and shadow prices of the crops.

Labor as additional input: Including other factors does not affect this result, as long as these other factors do not have prices that vary across crops. To see this, suppose that the production technology is given by

$$y_i^k(\omega) = (l_i^k(\omega))^{\alpha_k} (z_i^k(\omega))^{\gamma_k} \tag{25}$$

where $l_i^k(\omega)$ is labor input on plot ω, $z_i^k(\omega)$ is effective unit of land (note that all plots are of the same size of unity), and α_k and γ_k are output elasticities with respect to labor and land, respectively. We assume constant returns so that $\alpha_k + \gamma_k = 1$ Suppose that labor market is also non-existent (the case of complete labor market trivially follows the same procedure by just replacing the shadow wage by market wage) so that the farmer makes decision based on household-specific shadow wage $\tilde{w}_i$ and the rental rate per plot is denoted by $r_i(\omega)$. The unit/marginal cost function

[23] See Sotelo (2020) for derivation of these equilibrium quantities.

consistent with equation 25 is given by

$$c_i^k(\omega) = \frac{\tilde{w}_i^{\alpha_k} r_i(\omega)^{\gamma_k}}{z_i^k(\omega)^{\gamma_k}} \tag{26}$$

Let plot ω is used to grow crop k; i.e., k the crop that gives the highest return on plot ω. Under autarky, note that shadow prices are equal to the unit costs, $\tilde{p}_i^k = c_i^k$. Thus, conditional on plot ω being used to grow crop k, we have the following relationship between rental rate of a plot, and shadow prices and wages:

$$r_i(\omega) = \max_k \{ (\tilde{p}_i^k \tilde{w}_i^{-\alpha_k})^{\frac{1}{\gamma_k}} z_i^k(\omega) \} \tag{27}$$

Together with the Fréchet distribution, this equation can be used to solve for the fraction of land allocated to each crop from the supply side:

$$\hat{\eta}_i^k = \frac{\left(\tilde{w}_i^{-\alpha_k/\gamma_k} \tilde{p}_i^k A_i^k \right)^{\theta}}{\tilde{\Phi}_i^{\theta}} \tag{28}$$

where $\hat{\Phi}_i = \left[\sum_m \left(\tilde{w}_i^{-\alpha_m/\gamma_m} \tilde{p}_i^m A_i^m \right)^{\theta} \right]^{1/\theta}$. I used $\hat{\eta}_i^k$ and $\hat{\Phi}_i$ as notations for equilibrium values for the model with labor as an input to distinguish them from $\tilde{\eta}_i^k$ and $\tilde{\Phi}_i$, which are equilibrium values for the model without labor as an input. Note that these eequilibrium values reduce to their counterparts in equations 21-23 when the labor share of income is zero, $\alpha_k = 0, \quad k$.

The quantity of crop k produced is given by

$$y_i^k = \hat{\eta}_i^k L_i \frac{\hat{\Phi}_i}{\tilde{p}_i^k} \tag{29}$$

Imposing autarky equilibrium $y_i^k = q_i^k$ and using equation 19, we have the following

$$\hat{\eta}_i^k = \frac{(\mu_i^k)^{-\frac{\sigma}{\sigma-1}} (\tilde{p}_i^k)^{\frac{\sigma}{\sigma-1}}}{\sum_m (\mu_i^m)^{-\frac{\sigma}{\sigma-1}} (\tilde{p}_i^m)^{\frac{-\sigma}{\sigma-1}}} \frac{\hat{\Pi}_i}{L_i \hat{\Phi}_i} \tag{30}$$

Using the fact that $\sum_k \hat{\eta}_i^k = 1$, we have

$$\hat{\eta}_i^k = \frac{(\mu_i^k)^{-\frac{\sigma}{\sigma-1}} (\tilde{p}_i^k)^{\frac{\sigma}{\sigma-1}}}{\sum_m (\mu_i^m)^{-\frac{\sigma}{\sigma-1}} (\tilde{p}_i^m)^{\frac{\sigma}{\sigma-1}}} \tag{31}$$

This is identical to the model without labor as an input. The term involving shadow wage drops from the land share equation because shadow wages do not vary across crops.

The effect of trade costs on the link between household land allocation and tastes: In this section, I provide a general proof of the effect of trade costs on correlation between land share and taste. To do so, I use the model in section 3 and the assumption about the functional form of the productivity term Z_i^k. To simplify notation, I assume land is the only input and the production function is linear in land.[24]

Proof. Let π_{ij}^k is the probability that farmer i is the cheapest supplier of crop k to farmer j. Note that $\pi_{ij}^k \leq \pi_{ii}^k$ because consumption of own produce does not involve trade costs. That is, a farmer is more likely to be the cheapest supplier of a crop to itself than being the cheapest supplier 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:

$$\begin{aligned}\psi_i^k &= \sum_j [\pi_{ij}^k \eta_i^k + (1 - \pi_{ij}^k)\pi_{ii}^k \tilde{\eta}_i^k] \\ &= \eta_i^k \sum_j \pi_{ij}^k + \pi_{ii}^k \tilde{\eta}_i^k (N - \sum_j \pi_{ij}^k)\end{aligned}$$

where η_i^k and $\tilde{\eta}_i^k$ are given in section 3, and N is the number of farmers. Taking derivative with respect to crop taste μ_i^k gives

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

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:

$$\begin{aligned}\frac{\partial^2 \psi_i^k}{\partial \tau_{ij}^k \partial \mu_i^k} &= \frac{\partial \pi_{ii}^k}{\partial \tau_{ij}^k} (N - \sum_j \pi_{ij}^k) \frac{\partial \tilde{\eta}_i^k}{\partial \mu_i^k} + \pi_{ii}^k (N - \frac{\partial \sum_j \pi_{ij}^k}{\partial \tau_{ij}^k}) \frac{\partial \tilde{\eta}_i^k}{\partial \mu_i^k} \\ &+ \pi_{ii}^k (N - \sum_j \pi_{ij}^k) \frac{\partial^2 \tilde{\eta}_i^k}{\partial \tau_{ij}^k \partial \mu_i^k} \end{aligned} \tag{32}$$

The first two terms in equation 32 are positive. Also, note that the last term is also non-negative. Intuitively, under autarky the correlation between taste for a crop and the fraction of land allocated to the crop is non-decreasing in trade costs. Thus, $\frac{\partial^2 \psi_i^k}{\partial \tau_{ij}^k \partial \mu_i^k} > 0$. □

[24]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.

D An alternative test of separability

In this section, I exploit 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.[25] If markets are 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.

A key challenge that is difficult to address is that on-farm labor demand is likely to be measured with substantial error. The surveys use recall based interviews where the household head is asked to report how many hours each members of the household and employed labor worked on each plot of land from planting to harvesting. This is burdensome for any self-employed farm household that does not keep records, which is typical. Most importantly, the burden to memorize and hence the measurement error is likely to increase fast with the household's demographic characteristics, such as the number of active age members, and the number of plots the household farms. In a very similar setup, Beegle et al. (2012) show that the measurement error in recall based consumption surveys gets worse with the household size and the number

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

of consumption items included in the survey. Moreover, the weights that should be used for labor supply of women and children to obtain aggregate labor demand is not straightforward.

I run similar specifications as Benjamin (1992) and LaFave and Thomas (2016) to compare my results with theirs. Table A.4 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.

Finally, I investigate how the correlation between on-farm labor demand and household demographic characteristics varies across households with different market access. But, it is important to note that the roads are designed to connect villages to goods markets and input (such as chemical fertilizer) markets. Their effect on facilitating commuting within or even across rural villages is likely to be negligible. The Benjamin-style specification is not convenient to introduce the role of proximity to markets. First, labor demand is regressed on a vector of household demographic characteristics and it is not feasible to interact a measure of market access with each of the household demographic characteristics (doing so would lead to multicollinearity). In the current paper, I use only the number of prime-age male members of the household, which the most important variable in explaining on-farm labor demand, and drop the other household characteristics.[26] Second, crop markets have a known physical location which can be used to measure the households proximity to the market. Such exercise is difficult for labor markets.

Table A.5 reports results for the effect of proximity to market and road expansion under URRAP on the correlation between on-farm labor demand and the number of prime age male in the household. Columns 1 and 2 pool all the three rounds of survey to estimate how this correlation varies across households with different proximity to market centers or roads. Proximity to market is measured as household's distance to the nearest population center with above 20,000 population (towns), or distance to the nearest all weather road. Column 1 shows that, the correlation between on-farm labor demand and household labor supply significantly increases with distance from towns. In fact, the correlation between on-farm labor demand and household labor supply is zero for households very close to towns. This is likely because labor markets are relatively more developed near towns, i.e., households might have better access to hire market labor and/or obtain off-farm employment at market wage. Column 2 shows similar result using distance to road, though the coefficient is less precise.

Columns 3 and 4 estimate the effect of road expansion under URRAP using DID estimation strategy. Similar to my main analysis, I use 2013 as pre-program and 2015 as post-program periods. Column 3 uses binary treatment dummy while column 4 uses a continuous measure of market access (see appendix section A). The results based on binary treatment are statistically insignificant. However, the results based on market access approach is statistically significant. It shows that the correlation between log on-farm labor demand and number of prime age male decreases by 0.137 following one standard deviation increase in market access. This is economically meaningful – villages with three standard deviation higher market access relative to the mean have about 0.4 lower correlation between on-farm labor demand and the number of prime age male in the household. Columns 5 and 6 use matching based DID estimation strategy to address the potential endogeniety issue in selection of

[26] Dillon and Barrett (2017) experience similar issue when analysing how the correlation between labor demand and household demographic characteristics varies with proximity to markets or across different agro-ecological zones.

villages for URRAP program. The results are very similar to those based on DID estimation. Overall, the results in table A.5 show that proximity towns and roads plays important role for the correlation between household on-farm labor demand and household demographic characteristics, i.e., for the separability of household production decisions from their consumption preferences.

Table A.4: 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. $p < 0.10$, $p < 0.05$, $p < 0.01$

Table A.5: The effects of proximity to market and roads on the correlation between labor demand and household labor supply. The dependent variable is log hours of total on-farm labor demand

	Proximity to market		DID approach		Matching + DID	
	(1)	(2)	(3)	(4)	(5)	(6)
Prime age male	-0.093	0.433	0.584	0.616	0.598	0.638
	(0.243)	(0.106)	(0.047)	(0.044)	(0.054)	(0.051)
Log dist to pop center	-0.255					
	(0.250)					
Log dist to road		0.023				
		(0.113)				
URRP * Post			0.023		-0.027	
			(0.155)		(0.180)	
Market Access				0.101		-0.025
				(0.262)		(0.271)
Prime age male * log dist to pop center	0.166					
	(0.058)					
Prime age male * log dist to road		0.054				
		(0.034)				
Prime age male * URRP * Post			0.025		0.045	
			(0.076)		(0.083)	
Prime age male * Market access				-0.137		-0.134
				(0.046)		(0.045)
N	10263	10263	6858	6596	5210	5201
R^2	0.519	0.517	0.544	0.510	0.542	0.513

Notes: Standard errors are clustered at village level. All regressions include village and year fixed effects. The first two columns pool all three rounds of survey to estimate how the correlation between on farm labor demand and number of prime age male varies with the household's proximity to population centers (towns of over 20,000 population) and to all weather roads. Columns 3 and 4 use DID estimation to evaluate the effect of road expansion under URRAP using 2013 as pre-road and 2015 as post-road periods. Column 5 and 6 combine DID with matching to address the selection issue. Prime age is defined as ages 15-64. The market access measure is standardized, so that the coefficient is interpreted as the effect of one standard deviation increase in market access relative to the mean. $p < 0.10$, $p < 0.05$, $p < 0.01$

E Appendix Tables

Table A.6: 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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.7: First-stage statistics for IV estimation of taste parameters

	Household-level taste estimation		Village-level taste estimation	
	Test statistic	P-values	Test statistic	P-values
	Underidentification tests			
Anderson canonical correlation LM statistics	Chi-sq(1)=20876.16	0.0000	Chi-sq(1)=20859.91	0.0000
	Weak identification test			
Cragg-Donald Wald F statistic	794.78	.	1182.81	.
	Weak instrument robust inference			
Anderson-Rubin Wald test	F(19,133280)=9.65	0.000	F(19,201846)= 11.01	0.000
Anderson-Rubin Wald test	Chi(19)= 282.20	0.000	Chi-sq(19)= 215.71	0.000
Stock-Wright LM S statistic	Chi(19)=281.81	0.000	Chi-sq(19)= 215.48	0.000

Notes: The endogenous regressors in equation 11 are the vector of prices, i.e., there are 19 crop prices and hence 19 endogenous regressors. The instrumental variables are the prices of these crops in the nearby village The table reports the first-stage test statistics for estimation of taste at both household and village levels.

Table A.8: Crop taste parameters by region

Crop	Tigray	Afar	Amhara	Oromia	Somale	B. Gumuz	SNNP	Gambella	Harari	Dire Dawa
Bananas	0.009	0.020	0.008	0.012	0.011	0.021	0.014	0.015	0.011	0.011
Barley	0.128	0.093	0.112	0.124	0.086	0.081	0.104	0.069	0.091	0.098
Chat	-0.088	-0.018	-0.066	-0.010	0.160	-0.008	-0.056	-0.054	0.235	0.283
Chickpeas	-0.010	-0.028	0.001	-0.010	-0.026	-0.008	-0.012	-0.019	-0.022	-0.024
Coffee	-0.200	-0.195	-0.223	-0.209	-0.416	-0.247	-0.149	-0.169	-0.438	-0.528
Enset	0.074	0.091	0.078	0.096	0.082	0.081	0.221	0.084	0.090	0.099
Fieldpeas	-0.028	-0.046	-0.015	-0.030	-0.053	-0.003	-0.029	-0.033	-0.048	-0.037
Haricot Beans	-0.010	-0.011	-0.007	0.010	0.047	0.055	0.029	-0.004	-0.012	-0.013
Horse Beans	0.006	-0.024	0.034	0.010	-0.053	-0.011	-0.004	-0.023	-0.038	-0.035
Lentils	-0.065	-0.058	-0.067	-0.059	-0.084	-0.070	-0.062	-0.045	-0.090	-0.082
Linseed	-0.050	-0.062	-0.049	-0.048	-0.064	-0.054	-0.059	-0.063	-0.056	-0.067
Maize	0.001	0.131	0.012	0.062	0.143	0.050	0.117	0.232	0.042	0.027
Millet	0.028	-0.003	0.037	0.008	-0.009	0.091	-0.003	-0.006	-0.008	-0.001
Nueg	-0.007	-0.011	-0.007	-0.007	-0.011	-0.001	-0.011	-0.011	-0.012	-0.011
Onion	0.124	0.140	0.115	0.111	0.214	0.146	0.097	0.171	0.105	0.079
Potatoes	-0.033	-0.023	-0.006	-0.021	0.002	-0.018	-0.009	-0.032	0.006	-0.021
Sorghum	0.292	0.190	0.270	0.206	0.214	0.263	0.166	0.194	0.381	0.439
Teff	0.278	0.157	0.262	0.241	0.105	0.198	0.147	0.207	0.127	0.139
Wheat	0.176	0.263	0.132	0.141	0.232	0.068	0.104	0.076	0.214	0.263

This table reports the mean values of crop tastes for each region, based on household level estimation of t approach where village prices of crops are instrumented by crop prices in neighboring villages. Household sa is used to calculate the mean values.

References

Adamopoulos, T., Brandt, L., Leight, J., and Restuccia, D. (2017). Misallocation, Selection and Productivity: A Quantitative Analysis with Panel Data from China. Working Paper 23039, National Bureau of Economic Research.

Adamopoulos, T. and Restuccia, D. (2014). The Size Distribution of Farms and International Productivity Differences. *American Economic Review*, 104(6):1667–97.

Asher, S. and Novosad, P. (2020). Rural Roads and Local Economic Development. *American Economic Review*, 110(3):797–823.

Atkin, D. (2013). Trade, Tastes, and Nutrition in India. *American Economic Review*, 103:1629–63.

Beegle, K., De Weerdt, J., Friedman, J., and Gibson, J. (2012). Methods of household consumption measurement through surveys: Experimental results from Tanzania. *Journal of Development Economics*, 98(1):3–18. Symposium on Measurement and Survey Design.

Benjamin, D. (1992). Household Composition, Labor Markets, and Labor Demand: Testing for Separation in Agricultural Household Models. *Econometrica*, 60:287–322.

Carletto, C., Savastano, S., and Zezza, A. (2013). Fact or artifact: The impact of measurement errors on the farm size–productivity relationship. *Journal of Development Economics*, 103:254–261.

Chen, C., Restuccia, D., and Santaeulalia-Llopis, R. (2017). Land Misallocation and Productivity. Working Paper 23128, National Bureau of Economic Research.

Deaton, A. (1997). *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy*.

Deaton, A. S. and Muellbauer, J. N. J. (1980). American Economic Association An Almost Ideal Demand System. *The American Economic Review*, 70:312 – 326.

Dillon, B. and Barrett, C. B. (2017). Agricultural factor markets in Sub-Saharan Africa: An updated view with formal tests for market failure. *Food Policy*, 67:64–77. Agriculture in Africa – Telling Myths from Facts.

Dillon, B., Brummund, P., and Mwabu, G. (2019). Asymmetric non-separation and rural labor markets. *Journal of Development Economics*, 139:78 – 96.

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

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

Foster, A. and Rosenzweig, M. (2022). Are There Too Many Farms in the World? Labor Market Transaction Costs, Machine Capacities, and Optimal Farm Size. *Journal of Political Economy*, 130(3):636–680.

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*.

Gollin, D. and Rogerson, R. (2014). Productivity, transport costs and subsistence agriculture. *Journal of Development Economics*, 107:38–48.

Gollin, D. and Udry, C. (2021). Heterogeneity, Measurement Error, and Misallocation: Evidence from African Agriculture. *Journal of Political Economy*, 129(1):1–80.

Hausman, J. A. (1994). Valuation of New Goods under Perfect and Imperfect Competition. Working Paper 4970, National Bureau of Economic Research.

Hsieh, C.-T. and Klenow, P. J. (2009). Misallocation and Manufacturing TFP in China and India. *The Quarterly Journal of Economics*, 124:1403–1448.

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

Jones, M., Kondylis, F., Loeser, J., and Magruder, J. (2022). Factor market failures and the adoption of irrigation in rwanda. *American Economic Review*, 112(7):2316–52.

Kebede, H. (2020). The gains from market integration: The welfare effects of rural roads in Ethiopia.

Kebede, H. (2021). The Pass-Through of International Commodity Price Shocks to Producers' Welfare: Evidence from Ethiopian Coffee Farmers. *The World Bank Economic Review*, lhab020.

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. Working Paper 26636, National Bureau of Economic Research.

Restuccia, D. and Rogerson, R. (2017). The Causes and Costs of Misallocation. *Journal of Economic Perspectives*, 31(3):151–74.

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

Shenoy, A. (2017). Market failures and misallocation. *Journal of Development Economics*, 128:65–80.

Shrestha, S. A. (2020). Roads, Participation in Markets, and Benefits to Agricultural Households: Evidence from the Topography-Based Highway Network in Nepal. *Economic Development and Cultural Change*, 68(3):839–864.

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

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

Suri, T. (2011). Selection and Comparative Advantage in Technology Adoption. *Econometrica*, 79(1):159–209.

Taylor, J. E. and Adelman, I. (2003). Agricultural Household Models: Genesis, Evolution, and Extensions. *Review of Economics of the Household*, 1(1):33–58.