Constraints to Tanzanian Agricultural Development: Input Use in Households Under Non-Separability

Jeffrey Dickinson*

Wednesday 13$^{\text{th}}$ January, 2021

Abstract

This paper analyzes labor market failures in Tanzania utilizing panel data and incorporating population estimates inferred from daytime satellite images. I utilize those images and empirical tests to provide insight into the potential sources of labor market disarray and I find population density and the presence of heterogeneity in labor quality contribute to agricultural labor allocation decisions. I am able to dissect the preparatory and harvest periods separately with respect to demand for family and hired labor. This dichotomy is critical since agricultural labor markets can 'bind' in the harvest period but not in the preparatory period; additionally, labor output of hired and family labor is directly observable in the harvest period while it is more costly to supervise in the preparatory period.

JEL Codes J1, J43, O12, Q10, Q12, Q13, Q16

*author contact: jeffrey.dickinson@graduateinstitute.ch, Department of International Economics, Graduate Institute of International and Development Studies, CP 1672 - CH-1211 Genève 1, Suisse. I am deeply grateful to my supervisors Nicolas Berman and Jean-Louis Arcand for their extensive support and comments. All remaining errors are my own.

1 Introduction

In this paper I analyze determinants of labor demand on smallholder farms in Tanzania utilizing the framework of the separation hypothesis, which was first examined empirically by Benjamin (1992). The 'separation' hypothesis states that household productive activities and consumption may be estimated separately as long as production does not rely on household-level parameters. In our context, this means that in order to estimate consumption and production separately farm labor decisions must not be linked to household characteristics.

Although there are a number of papers in the academic literature utilizing the separation hypothesis framework, the existing literature does not provide evidence for policy interventions, and several key papers in the literature focus on Indonesia (Benjamin, 1992; Grimard, 2000; LaFave and Thomas, 2016). Prior separation analysis often uses cross-sectional data which does not allow the researcher to control for individual-specific, unobserved, time-invariant heterogeneity. Controlling for these factors is critical in villages with different climates, ethnic composition, and cultural norms. It is also crucial to control for household-specific preferences for labor and leisure since those may be linked with household characteristics and labor supplied to the farm. To the best of my knowledge there is also no analysis in the separation literature that utilizes panel data and divides labor into the preparatory period and the harvest period. This matters because separation, or market failure, could be seasonal, with labor rationing occurring in the harvest season when labor markets are more likely to bind and which would be especially problematic if farmers incur losses due to unharvested crops.

The principal motivations for analyzing agricultural labor market failures in Tanzania are policy-based: if we have insight into *why* markets might not be functioning we can better target policies to fix those areas. If differences in the marginal product of family and hired labor are causing total household size to affect labor demand, the separation test will fail.

To provide further identification of potential drivers of labor market separation I incorporate data which are usually unavailable for low-income countries: high-resolution data on population changes inferred from daytime satellite imagery using algorithms. This allows me to provide insight into one potential driver of labor market dysfunction: a lack of liquidity of labor.

New data sources in the form of satellite-inferred population data bolster the story of what may be driving labor market failures.

One example of a potential policy recommendation to rectify illiquid harvest labor markets would be to provide subsidized transportation between dense urban areas and rural areas in harvest times, or to facilitate cross-border movements of laborers in those periods.

Another principal motivation is that Central Indonesia, the location of much of the previous separation analysis, is fairly homogenous with respect to agro-ecological zones and population density, while the geographic breadth and variety of Tanzania is enormous. Tanzania covers many agro-ecological zones, shares borders with 8 countries, and it has heterogeneity in geophysical characteristics such as mountains, vast savannah, dense forests in the western part of the country, and lakes in the northwest and southwest part of the country. Additionally, the stark differences between Tanzania and Indonesia provide fertile ground for discussing policy interventions.

Last, the high-quality nature of the Tanzanian Living Standards Measurement Survey (LSMS) farm and plot-level data including soil, slope controls, and other plot variables as well as variables measuring managerial human capital allows for robust identification. This is important as omitting managerial controls can plague similar estimations in the form of omitted variable bias, in particular when there are 'brain drain' effects in low-income countries. The panel dimension of the dataset allows the utilization of household and village-wave specific effects which control for heterogeneity among households and between villages in different years.

To summarize again, this paper conducts the first panel data based tests for labor market separation in East Africa. Using panel data for separation tests allows the researcher to control for time-invariant unobserved heterogeneity. In the case of household farm labor demand, this is significant because household's tastes for farm labor may affect labor demand. This is the first separation analysis that divides labor into harvest and preparatory periods, which allows me to analyze and test for season-specific separation again providing even more insight into potential drivers of labor-market dysfunction. This step in the analysis is significant because labor constraints may bind in the harvest season but not in the slack preparatory period meaning separation could be season-specific.

2 Theoretical Background

2.1 Separation of Production and Consumption Activities

Research into rural agriculture is a popular subject in the development microeconomics applied and theoretical literature. This is because much of the developed world is still characterized as living in rural or subsistence farms, and some themes from farm-household dynamics apply in both developed and developing countries such as the well-known stylized fact that agricultural labor markets "tighten" during the harvest period. There is a vein of the development economics literature which explores the market imperfections which affect these types of households. This paper will draw from the literature on market inefficiencies faced by rural agricultural households, with a specific focus on input markets including labor and manure. As argued in De Janvry and Sadoulet (2006), as well as in Thorbecke (1993), market failures are often the defining characteristics of rural markets, and thus analysis that does not consider households without an understanding of the constraints which these households face.

As argued in Benjamin (1992) and Card et al. (1987), market prices and wages should function as indicators if markets are complete and efficient. If this type of signalling mechanism is in operation it should lead to a detectable 'separation' between household productive and consumption activities. Benjamin (1992) theorizes that there may be three principal sources of breakdowns in the labor market that lead to non-separation: (1) a binding constraint on off-farm employment, (2) labor rationing, (3) and differences in the returns to on-farm and off-farm employment.

With respect to the first source, very few Tanzanian households are engaged in salaried work outside of the primary sector. Tanzania has a low labor force participation rate (less than 60% for the LSMS sample) and relatively high official unemployment. Over 35% of households in the LSMS sample have an unemployed adult member. Households in our sample clearly face an environment with limited outside opportunities.

One empirical test I employ addresses the question of labor rationing and whether households are reliant on family labor for household farm activities. If households still rely on family labor holding constant a household's taste for agricultural labor, this indicates separation which is, at least in part, driven by labor rationing. This rationing could be seasonal: labor markets can be slack in the preparatory period when few salaried jobs are available and farming tasks can be more easily divided over days. At harvest time, the work usually needs to happen within a relatively tight window and at a moment when many farm households have an increase in demand for high-quality labor. In that case that the separation test rejects this may be evidence of labor rationing.

For the third potential driver of separation I employ two separate tests for labor homogeneity to confirm that differential marginal products of family and hired labor exist. If the marginal product of labor in a production function is increasing in the share of family members, this indicates that family members are more productive than hired labor in terms of output. Like the labor demand analysis, these ratios are tested separately for the preparatory and harvest periods.

Although the principal analysis of separation in this paper is with respect to labor markets, I also consider fertilizer factor markets as a source of potential separation between household-level variables and plot-level decisions. Last, I check whether a similar story exists for fertilizers. If total fertilizer use relies on household characteristics, such as the number of animals in the household herd.

2.1.1 Labor Demand Equations

With respect to the estimations of plot-level labor demand, the dependent variables are the log number of total family labor days, and the log of total hired labor days. The regressions take the following form for plot i in household h:

$$L_{iht}^{FAM,HIRED} = \beta N_{iht} + \delta X_{iht} + \alpha z_h + \eta_{dt} + \epsilon_{hvt} \tag{1}$$

where N is vector of household characteristics, and X is a vector of other plot characteristics. In some equations subscripts for time and village are omitted for legibility. The regression includes a set of district-wave dummies η_{vt} and a set of household dummies z_h.

2.2 Labor Heterogeneity

Whether labor hired from the marketplace is comparable to family labor is an important question: If there exists a quality or skill differential between hired and family labor, this could contribute to labor markets being observably separable since farmers may be more reluctant to hire less-skilled workers from the marketplace, or they may be constrained by unmotivated family laborers. Put differently this means that if household members provide higher quality labor on the farm or vice-versa, it is possible we would observe a correspondence between household characteristics and labor use. The literature on statistical tests analyzing the homogeneity of labor provides two types of tests, one of them is a "Bardhan-Frisvold"-type test, and another is a "Deolalikar-Vijverberg"-type test. The first test, the Bardhan-Frisvold test, estimates a Cobb-Douglas production function but assumes that the marginal products differ between family and hired labor. This method avoids an explicit estimation of the elasticity of substitution *between* hired and family labor, whereas the Deolalikar-Vijverberg test, using a simultaneous estimation procedure to estimate a labor services function, seems, *a priori*, less restrictive in the way it permits substitution between family and hired labor, zero-labor inputs, and higher-order terms. As a result of the comprehensive nature of the LSMS dataset, I am also able to consider a differential between wages received by hired-in workers on the household farms and the wages received by household workers who work off-farm. In the next sections the theoretical background of each test is discussed.

2.2.1 Bardhan-Frisvold Type Tests

A notable paper on farm productivity and the returns to scale, written by Bardhan (1973) also analyzes the heterogenous contributions of different labor types using a Cobb-Douglas production function. Bardhan's paper, using Indian agricultural data from farm management surveys, finds that family and hired labor are *not* substitutable in West Godavari and Thanjavur districts, but for the remaining districts in the sample the author cannot reject homogeneity of labor. Frisvold (1994) explores labor heterogeneity, again using Indian household survey data. The author's primary motivations are to explore supervisory costs and how they affect farm activity, but he also examines the question of labor heterogeneity. Using a similar specification to Bardhan (1973), Frisvold (1994) rejects labor homogeneity and finds that family supervision labor augments hired labor. Bardhan (1973) and Frisvold (1994) both estimate a production function similar to:

$$q_i = \alpha_0 + \alpha_1 A_i + \alpha_2 V_i + \alpha_3 L_i + \theta RATIO_i + u_i \tag{2}$$

where L is total labor (F+H) labor services functions of the following form:

$$E = (F + H) \left[\frac{F}{F + H} \right]^{\gamma} \tag{3}$$

in log form:

$$\log(E) = \log(F + H) + \gamma \cdot \log \left(\frac{F}{F + H} \right) \tag{4}$$

where E represents effective total labor. In expression 2, A is area planted, V is the value of manure, fertilizers, and feed, and $RATIO$ is the ratio of family/total labor, the principal variable of interest. In this case it is possible to divide labor into the harvest and preparatory periods which is motivated by the differences in the types of labor required in the preparatory versus the harvest period and the observable nature of labor output in the harvest period. This allows tests on two separate ratios, the ratio of family to hired labor in the preparatory period and the ratio of family to hired harvest labor. γ, which is represented by θ in equation 2, can be estimated by OLS. If family and hired labor are perfect substitutes, I can test $\gamma = 0$. The point of this exercise, again, is to shed some light on potential sources of separation; if family labor and hired labor are of differing qualities this provides direct evidence for one potential source of separation. This also means the household would be more reliant on family labor and it validates the choice of estimating the marginal products of family labor and hired labor separately.

2.2.2 Deolalikar and Vijverberg tests

Deolalikar et al. (1987) use a generalized quadratic labor services function to test for the effects of labor heterogeneity using Indian and Malaysian data. Importantly they separate two aspects: (a) perfect substitutability, and (b) a quality differential between family and hired labor. They outline in their article the implications for labor heterogeneity between hired and family labor; if the two are substitutes, the authors argue, and family members migrate away from the village farms, this will raise the wages of hired labor. They add that in the case that the two are not at all substitutable, an out-migration of family labor could actually decrease demand for hired labor. This makes sense particularly when hired labor markets are illiquid or incomplete. In contrast with previous labor heterogeneity literature where most farms hire in some labor, in Tanzania a smaller percentage of farms sampled in the LSMS hire-in labor (43%) and the average total number of hired days per acre is quite low at around 2.5 hired-labor-days per acre. Deolalikar et al. (1987) reject perfect substitutability between family and hired labor in both India and Malaysia. They also find that hired labor is more efficient in terms of output than family labor using the ratio of marginal productivities.

Based on the specification in the original paper, a Cobb-Douglas functional form is estimated, and a generalized quadratic functional form is used to characterize the labor services function. The reason for using the quadratic form nested in a Cobb-Douglas is that, by contrast, in Cobb-Douglas the marginal product of all inputs goes to infinity as as the input goes to zero. Using the quadratic form will allow for slightly more flexibility than Benjamin (1992), as I would like to consider explicitly the nature of the substitution of hired and family labor.

$$Y = C + \beta_1 L + \beta_2 A + \Sigma_i \beta_i X_i + \varepsilon \tag{5}$$

in the above equation, Y is output, and labor services L, A represents services from land, and X_i= quantity of input i. Continuing in the format of Deolalikar et al. (1987), I assume that labor services are produced using family labor and hired labor by the generalized quadratic function:

$$L = \alpha_1 L_f + (1 - \alpha_1) L_h + \delta_{11} L_f^2 + \delta_{22} L_h^2 + \delta_{12} L_h \cdot L_f \tag{6}$$

This form is flexible enough to allow various elasticities of substitution between family and hired labor (Deolalikar et al., 1987). In order for equation (4) to be concave, equation (5) must also be concave, a necessary condition is that δ_{11} and δ_{22} are not positive. Furthermore, α_1 and $(1 - \alpha_1)$ must be positive. Following Deolalikar et al. (1987), the appropriate test is then an likelihood ratio (LR) test that $\delta_{11} = \delta_{22} = \delta_{12} = 0$, which is a direct test of the hypothesis of perfect substitutability between labor types. If the two types of labor are equivalent, $\alpha_1 = 0.5$, then equations (4) and (5) simplify to a standard Cobb-Douglas form:

$$L = \alpha_1 L_f + (1 - \alpha_1) L_h \tag{7}$$

Note that, in the case where $\delta_{11} = \delta_{22} = \delta_{12} = 0$, and we are in a Cobb-Douglas universe, the marginal product of labor is given in full by $\beta_1 \cdot \alpha_1 L_f$ and $\beta_1 \cdot (1 - \alpha_1) L_h$ and we have:

$$Y = C + \beta_1 \alpha_1 L_f + \beta_1 (1 - \alpha_1) L_h + \beta_2 A + \Sigma_i \beta_i X_i + \varepsilon \tag{8}$$

2.2.3 Fertilizer Factor Allocation

A final strategy I employ in understanding Tanzanian agricultural households is to analyze intensity of input use in the form of organic fertilizer. Organic fertilizer is much more abundant and accessible in Tanzania than chemical fertilizers, as organic fertilizer is simply an output from livestock kept by many farms. Similar to Gavian and Fafchamps (1996), I regress organic fertilizer use per acre on household and plot characteristics. Organic fertilizer is considered a short term investment since its benefits may last longer than one cropping season (Gavian and Fafchamps, 1996). If markets for organic fertilizer inputs are functioning and complete, returns to fertilizer should be equalized across all plots conditional on plot characteristics, crop choice, and weather. Although organic fertilizer is too bulky to transport, at least in the West African context overnight paddocking contracts have been documented. Gavian and Fafchamps (1996) find that land holdings per household member negatively influenced organic fertilizer use per hectare, and that organic fertilizer use was largely determined by the size of the livestock holdings of the household.

3 The Setting and the Data

Tanzania as a country is well-suited for agricultural production, and farming makes up a substantial portion of the activity of low-income households: 37% of men in the survey worked on their own farm last week, and 39% for women. Tanzania straddles several agro-ecological zones; in the north around Lake Victoria and in the south-western part of Tanzania there are cool

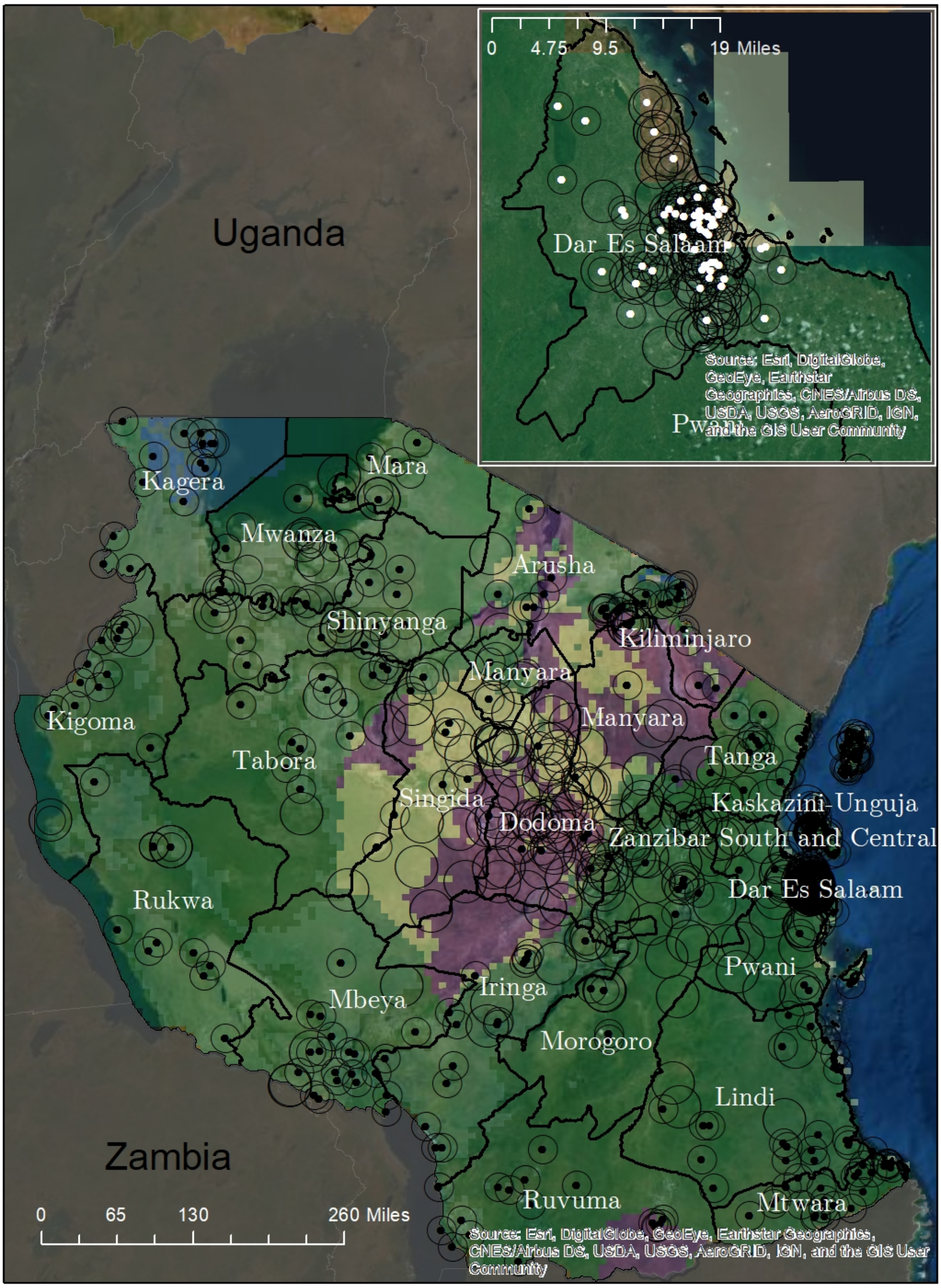

Figure 1: Survey Households Overlayed on Tanzanian Agricultural Ecological Zones; Source: Tanzania LSMS, IFPRI Raster Data; http://www.IFPRI.org
Small black dot: Wave 1 Household, Black dot with one circle: Wave 1 and Wave 2 Household
Black circle with no dot: Wave 2 and Wave 3 households; Largest diameter circle is households sampled in Wave 3

Long Rainy Season		Short Rainy Season	
Variable	Acres Planted	Variable	Acres Planted
Maize	19701	Maize	4211
Paddy	4346	Beans	1528
Beans	4161	Groundnut (Peanut)	579
Groundnut (Peanut)	3791	Sweet Potatoes	453
Sorghum	2503	Paddy	434
Cotton	2128	Cotton	416
Sweet Potatoes	2036	Cowpeas	259
Sunflower	1738	Green Gram	217
Cowpeas	1409	Sorghum	203
Pigeon Pea	1361	Cocoyams	113
Sesame	923		
Green gram	892		
Tobacco	692		
Bulrush MIllet	645		
Chickpeas	548		
Bambara Nuts	496		
Cassava	334		
Cocoyams	298		
Finger millet	282		
Pumpkins	267		
Irish Potatoes	179		
Kiwi	161		
Tomatoes	148		
Cashewnut	102		

Table 1: Area Planted by Crop in Tanzania During the Survey Period

sub-humid tropic climates. Much of the southern and eastern as well as south-eastern parts of Tanzania are warm sub-humid tropical climate, while a large central swath of Tanzania is characterized by a warm and cool semi-arid tropical climate. The dataset used in this paper is nationally representative, meaning all of these zones are included in the analysis. A map of enumeration areas (villages) used in Tanzania LSMS waves 1-3 can be seen in Figure 1. This is an important dimension of heterogeneity within the data, and it is one of the reasons for the inclusion of specific types of fixed effects. Principal crops grown in Tanzania can be seen in table 1. They include maize, rice, sweet potatoes, cassava, and sorghum among others with a higher share of farms growing peanuts in the second, shorter cropping season. According to survey data, agriculture and livestock make up a substantial part of Tanzanian economic activity. Those workers outside of the agricultural business are engaged in teaching, civil service, or natural resource/extractive industries.

The primary data used are from the World Bank's Living Standards Measurement Survey (LSMS) instrument from Tanzania, which includes a substantial agricultural component captured over four waves from 2008-2015. All waves of data are freely available from several sources including the World Bank website and the website of the Tanzanian National Bureau of Statistics. Data were collected on basic household demographic characteristics, and the questionnaire included modules on labor, consumption, assets, and anthropometric data for household members. Agricultural data were recorded separately, but at the same sitting for the two agricultural seasons experienced in some parts of Tanzania. For the two separate seasons, locally referred to as the 'short rainy' season and the 'long rainy' season, plot inputs and are recorded as one observation per year, though outputs are recorded separately and summed across seasons for our analysis.

An important feature of this dataset is that records kept at the plot level are highly detailed. Included are information on plot ownership, seed type and purchases, fertilizer use, which household member manages the plot, as well as which family members provide labor on the plot and whether or not any hired labor was used. Descriptive statistics for household demographic characteristics as well as farm assets and other characteristics can be found in table 12 in the appendix.

Wave 1 of the survey was collected from September 2008 and the bulk of interviews were completed by September of the following year. The sample contains 3,265 households, including

Table 2: Household Summary Statistics

VARIABLES	N	mean	sd	p50	min	max
Age of HH Head	6,447	48	16	46	16	108
HH Head Years of Educ.	6,447	4.46	3.88	4.2	0	22
Gender of HH Head (1=Man)	6,447	0.22	0.40	0	0	1
# Children in HH	6,447	2.03	1.75	2	0	23.2
# Adults in HH	6,447	3.09	1.85	2.82	0	22.8
# Seniors in HH	6,447	0.26	0.52	0	0	3
Agrictultural Wage (in 1000's of TSH)	6,447	1293	863.23	1040	0.5	3380
HH Experienced a Death	6,447	0.09	0.25	0	0	1
Total Animal Units	6,447	4.28	29.6	0.04	0	527
Total Value of Farm Assets	6,447	5183	78760	160.07	0	3338000
Total Value of HH (non-farm) Assets	6,447	12030	135600	1912	0	5909000
Population Density (persons per km sq)	6,402	955.9	4,870	40	0	77,028

*in 1000's of Tanzanian Shillings

5,126 plots held by 2,284 farm-households households. Wave 2 was collected from October 2010 with the majority of interviews completed by September 2011. The second wave sample contains 3,924 households, and 2,630 farm-households with 3,829 planted plots. Collection for wave 3 began in October of 2012 with interviews nearly complete by the end of October 2013. The 3rd wave of the survey is expanded, and includes 5,010 households with 3,300 farm-households including a total of 4,934 plots. The fourth wave of the survey sampled the same villages, but replaced the households in the sample, and for the fourth wave the data were collected from October 2014 through August 2015. It includes 3,352 households and data on 4,291 plots. The sample used in the in the estimations in this paper is restricted to only those households which own and operate a farm, though other non-farm households were also sampled. In the full sample of farm households, including all waves, there are 4,356 farm-households in waves 1-3, and 2,093 new farm-households in wave 4 for a total of 6,447 households. Descriptive statistics for key household-level variables can be found in table 2.

Descriptive Table 3 shows descriptive statistics of both family and hired labor use at the plot level. Labor is split into planting, weeding, and harvesting periods, though in the analysis planting+weeding activities are summed to simplify and because this is supported by the literature on observability of agricultural activities. Family labor use is much higher than hired labor use on average. Average hired labor use in both the preparatory and harvest periods appears to be very stable across all waves.

Table 2 contains household-level descriptive statistics of the farm households in the sample.

Table 3: Labor Use on Farms in Sample

Labor Type	Pct Used	Mean total labor-days	Median total labor-days	Pct. Hired	Mean hired labor-days	Median hired labor-days	Mean Total Labor-Days per Acre	Median Total Labor-Days per Acre	Mean Total Labor-Days per Hect	Median Total Labor-Days per Hect	Mean Hired Labor-Days per Acre	Median Hired Labor-Days per Acre	Mean Wage	Median Wage
Planting	0.99	121.9	57.5	0.34	2.1	0	47.2	21.6	116.5	53.3	2.1	0	6,646	2,857
		(847.6)			20.0		62.9		155.4		5.0		9,043	
Weeding	0.88	94.13	8	0.28	8.1	0	28.1	9.1	69.4	22.6	1.1	0	12,126	4,000
		999.7			73.3		46.0		113.6		3.2		16,276	
Harvesting	0.96	43.07	14	0.19	6.2	0	16.8	8.9	38.6	19.8	0.6	0	3,604	2,500
		314.5			73.8		20.4		48.9		1.4		3,153	

This table only includes households that farm. All non-farm households have been dropped. Household heads are 48 years old on average, and they average about 4 years of education which is equivalent to a primary school education. Households include 2 children and 3 adults on average.

3.1 LandScan Data

LandScan gridded population data is a set of gridded population estimates, available on an annual basis, with a fine resolution allowing analysis at a more dis-aggregated level. The benefit of using these data are that they allow us to capture fluctuations in population that might be otherwise difficult to observe, and where we can be mostly sure the measurement error of NTL is orthogonal to our other controls. In Tanzania, I am not aware of any data covering the entirety of the country on an annualized basis to measure population. As such the LandScan data will be a great benefit, and they will also help us identify the effects of population changes on household economic activities. These data originate from the OakRidge National Laboratory (ORNL), which is a research institution funded by the US Department of Energy, and managed in partnership with the University of Tennessee. The estimates are generated by an algorithm that takes as its primary inputs high resolution, proprietary daytime imagery (Rose and Bright, 2014). The following brief description comes from the ORNL-LandScan documentation, "the modeling process uses sub-national level census counts for each country and primary geospatial input or ancillary datasets, including land cover, roads, slope, urban areas, village locations, and high resolution imagery analysis; all of which are key indicators of population distribution."[1]

[1]https://landscan.ornl.gov/documentation/#inputData

VARIABLES	ag wage	ag wage
hired wage	0.0265	
	(0.0406)	
harv wage		0.0150
		(0.0341)
Observations	740	740
R-squared	0.004	0.004
Wave FE	yes	yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4: Regressions of Wage Differentials

4 Results

4.1 Tests for Labor Heterogeneity

4.1.1 Wage Differential

The dataset has wage data from both contract workers hired in to work on the farm and from the labor module on wages paid to family members who work on other farms or in agricultural sector jobs. Family agricultural wages were scaled to a daily wage, and then standardized by removing the most extreme values before being collapsed to the village level median wage. The same process was applied to the wages of hired-in labor. The raw data are also processed to remove extreme values, and the data are collapsed to their median values. I then ran the simple regression of family wages on hired wages. If there is no relationship, the coefficient on hired wages should be equal to zero. Looking at Table 4, we can see a normal linear regression of family wage on hired wage reveals no relationship which is significant at standard levels.

4.1.2 Bardhan-Frisvold Test for Labor Homogeneity

The next test I run to examine the relationship between hired and family labor is based on those used in papers by authors Pranab Bardhan (1973) and George Frisvold (1994), but adapted to include indicator variables for irrigation status and land tenancy. This test estimates a Cobb-Douglass production function where the dependent variable is output regressed on inputs and other controls; the unit of observation is the individual plot since we have labor allocations for both family and hired labor at the plot level. As mentioned the preparatory period and harvest period labor are divided into separate categories for analysis which means we can check if the ratio is significant in the harvest period and preparatory period separately or if the two effects are equal. Household fixed effects control for household-specific tastes for farm labor. Other controls include the log of the plot area, indicator variables for soil quality, soil type, irrigation status, whether the plot is rented by the cultivator, and an indicator variable in case the plot was planted to hybrid or other improved varieties are also included. The log of kg of organic fertilizer use is also incorporated into the control variables as well as the log of total plot expenditure. The following expression can be estimated by ordinary least squares (OLS) where L_{prep} and L_{harv} are the log of total preparatory and harvest labor days :

$$\ln q_i = \ln \alpha_0 + \alpha_3 \ln L_i^{prep} + \alpha_4 \ln L_i^{harv} + \theta_1 \ln RATIO_i^{prep} + \theta_2 \ln RATIO_i^{harv} + \delta Z_i + u_i \quad (9)$$

	Observations	Mean	Std. Dev.	Median	Min	Max
Total Prep Labor	21,837	53.76	62.57	35	0	948
Total Harvest Labor	21,837	22.06	35.55	10	0	540
Ratio_{Prep}	21,837	1.76	8.46	0	0	335
Ratio_{Harv}	21,837	0.14	1.95	0	0	133
Plot Size (in Acres)	21,837	2.79	8.62	1.1	0	600
Soil Type	19,783	2.05	0.68	2	1	4
Soil Quality	20,923	1.63	0.59	2	1	3
Irrigated	21,837	0.02	0.14	0	0	1
Organic Fertilizer Used (in KG)	21,837	91.89	656.9	0	0	32,000
Plot is Rented	21,837	0.04	0.19	0	0	1
Improved Seeds Used	21,837	0.33	0.47	0	0	1
Total Plot Expend.*	21,837	43.29	185.75	0	0	7608
Plot Has Tree Crops	21,837	0.35	0.48	0	0	1

*1000's of Inflation-Adjusted Tanzanian Shillings

Table 5: Descriptive Stats of Variables Used in Bardhan-Frisvold Tests for Labor Homogeneity

Descriptive statistics for the variables used in the regressions can be found in Table 5, and the results can be found in Table 6. The terms $RATIO_{prep}$ and $RATIO_{harv}$ are the main variables of interest derived from expression 9. These ratios represent the expression defined earlier in (2) and (3), and the coefficient of these ratios corresponds to the expression $\theta = \alpha_3 \gamma$, where $L_j = (F + 1)/L$. Therefore a test of $\theta = 0$ is a test for the substitutability of labor. Columns 1 and 2 of Table 6 are

Column 1 in table 6 represents the estimation of (9) without soil control variables for which there are some missing values. In columns 2 and 3 the same model is fit, this time incorporating soil controls which reduces the sample by about 2000 observations. Recall that

Table 6: Bardhan-Frisvold Tests for Labor Homogeneity

VARIABLES	(1) Quant. Harvested	(2) Quant. Harvested	(3) Quant. Harvested
Plot Size	0.0707**	0.138***	0.219***
	(0.0330)	(0.0351)	(0.0312)
Total Prep. Labor	0.389***	0.310***	0.310***
	(0.0206)	(0.0238)	(0.0187)
Total Harv. Labor	0.372***	0.357***	0.382***
	(0.0211)	(0.0214)	(0.0220)
Total Plot Expend.	0.0548***	0.0515***	0.0520***
	(0.00448)	(0.00445)	(0.00409)
$Ratio_{prep}$	-0.0952***	-0.0883***	-0.101***
	(0.0241)	(0.0237)	(0.0223)
$Ratio_{harv}$	0.381***	0.369***	0.445***
	(0.0603)	(0.0592)	(0.0619)
Improved Seeds Used	0.869***	0.748***	0.697***
	(0.0426)	(0.0438)	(0.0414)
Organic Fert. Used	0.0281***	0.0403***	0.0508***
	(0.00999)	(0.00990)	(0.00704)
Irrigated	0.417***	0.349***	0.628***
	(0.136)	(0.130)	(0.105)
Plot is Rented	0.00886	0.0215	0.0724
	(0.0931)	(0.0940)	(0.0702)
Plot has Trees	1.621***	1.465***	1.380***
	(0.0390)	(0.0399)	(0.0477)
Observations	21,826	19,783	19,783
Soil Type Controls	no	yes	yes
Soil Quality Controls	no	yes	yes
Household FE	yes	yes	no
Wave×Village FE	yes	yes	yes
Number of Households	6,447	5,768	

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

the main coefficients of interest are the ones associated with the ratios from expression 9: family/total prep labor and family/total harvest labor θ_1 and θ_2 which corresponds to Ratio_{prep} and Ratio_{harv}. Across all three columns the null hypothesis that $\text{Ratio}_{prep} = 0$ and $\text{Ratio}_{harv} = 0$ is strongly rejected. The negative coefficient on the ratio in preparatory period suggests that a higher fraction of family labor in this period sharply reduces the marginal product of preparatory labor. Focusing on the Ratio_{harv}, the coefficient is now much larger and positive meaning that an increase in the share of family labor over total harvest labor leads to an increase in the marginal product of harvest labor. These regressions indicate that in the preparatory labor period, when supervisory costs are high for both family and hired labor, a larger share of family labor leads to lower overall productivity of labor. Additionally, in the harvest period when supervisory costs are low, an increase in family labor share leads to a higher marginal product of harvest labor. This, I argue, is evidence that differentials in productivity between hired and harvest labor could be one source of separation.

4.1.3 Deolalikar and Vijverberg Generalized Quadratic NLLS Estimates

In the case of this dataset I chose to estimate preparatory labor (any labor that occurs pre-harvest including planting, weeding, and fertilizing activity) and harvest labor separately. This is in contrast with the original authors who estimate all farm labor together, with the only distinction being between family and hired labor. The first test is a likelihood ratio test of the model from equation 5: $\delta_{11} = \delta_{22} = \delta_{12} = 0$. The test for the preparatory labor period rejects with $\lambda_3 = 28672.14$, and $\lambda_3 = 12929.81$ which are both significant at the .1% level. This means that in both the harvest period and the preparatory labor period I can reject perfect substitutability between hired labor and family labor. I perform the likelihood ratio tests then for the perfect substitutability of labor from equation 6, $\delta_{11} = \delta_{22} = \delta_{12} = 0; \alpha_1 = 0.5$ and in both cases, homogeneity of labor is rejected: $\lambda_4 = 52043.73$ for the preparatory period, and $\lambda_4 = 42101.23$.

Next I present the full results from the nonlinear least squares estimates of the parameters in expression (4). The estimates for α_1 are 0.424 for the preparatory season, and $\alpha_1 = 0.461$ in the harvest labor season. This indicates that family labor increases to be more productive during the harvest labor period, and that the ratio of the marginal productivities $(\alpha_1/1 - \alpha_1)$ is larger in the harvest season, 0.74 (prep) compared to 0.86 (harv). This ratio being closer

Table 7: Deolalikar-Vijverberg Test - NLLS Estimates

VARIABLES	b0	α_1	δ_{11}	δ_{22}	δ_{12}
Prep	13.22***	0.424***	0.00326***	0.000859***	0.0141***
	(0.172)	(0.00131)	(0.000107)	(1.24e-05)	(0.000136)
Harvest	3.013***	0.461***	-0.00153***	0.000781***	0.0236***
	(0.0582)	(0.00150)	(0.000174)	(2.27e-05)	(0.000281)
Observations			25,467		
R-squared			0.809		

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

to unity indicates higher/greater substitutability. This is slightly lower than but comparable to 0.78 for Malaysia, and quite far off from the estimated 0.32 for Matar Taluka (India) in Deolalikar et al. (1987).

As we can see, the preceding exercise has indicated that hired and family are not perfect substitutes, neither in the preparatory period, nor in the harvest period. The harvest period estimates indicate that the marginal product of family labor is positive but decreasing, significant at the 0.1% level. The coefficient on the interaction term is also positive, which could be interpreted as signifying that increased supervision costs improve the performance of hired labor.

4.2 Family and Hired Labor Demand Estimates

Based on the results of the earlier analysis, family labor and hired labor are considered separately here, as are pre-harvest (preparatory) labor and harvest labor.

Columns 1-2 of table 9 contain within-household estimates of preparatory and harvest labor demand, while columns 3-4 contain within-household estimates of hired labor demand. All columns containing estimates include district-wave fixed effect dummy variables, which control for elements such as district-specific weather and price shocks, as well as controls for soil type and for the slope or gradient of the plot. Variables above the line in table 9 are household-level controls which are thus the principal variables of interest while all other control variables vary at the plot level. Although a few household-level variables have statistical significance, the variable that provides the most striking rejection of separation which is the number of adults per household. This variable being a key driver of the rejection of household separation is consistent with the previous separation literature on labor markets. The effect of an additional adult member in the household is a statistically significant increase in the amount of labor provided in the preparatory period. With respect to the hired labor demand in the preparatory and harvest periods in columns 3 and 4 an increase in the number of adults in the household results in a statistically significant decrease in the amount of hired labor demanded, though the effect relatively small in magnitude especially relative to size of the shift in the family labor demand estimates. The fact that these variables are statistically significant constitutes a firm rejection of labor market separation. Looking at the plot-level variables we see labor is increasing in plot area, the use of organic fertilizer, and total plot expenditure.

4.3 Fertilizer Factor Allocation Regressions

Fertilizer in the form of manure from animals and livestock is considered a very important investment for farmland. In Tanzania, besides for labor it is probably the most important input the farmers have easy access to. For these reasons I examine also the use of manure as an input. Fertilizer regressions represent the following estimated model:

$$M_{ih} = \beta N_{ih} + \delta X_h + \eta_{jt} + \zeta_{hjt} \tag{10}$$

where M_{ih} the dependent variable is the log of fertilizer per acre applied to plot i in household h. N_{ih} and, X_h are vectors of plot characteristics at the plot and household level. Dummy variables for household (within-transform) and village-wave fixed effects included.

Results from the regression of the log of fertilizer per acre on plot and household control variables are shown in Table 9. Descriptive statistics are again found in the appendix. All models are identical to the earlier within-household fixed effects model. Column 1 contains controls for the total number of animals the household owns, and column 2 includes the value of animal holdings as an indicator of herd size. As the animal units variable is more likely to be correlated with fertilizer use (often livestock is left overnight on the field for the purposes of fertilizing), this offers the advantage of representing the value of the stock while hopefully being less endogenous if livestock prices are more or less exogenous.

First and most important the number of adults in the household is a positive and statistically

Table 8: Plot-level Family Labor Demand

	Family		Hired	
	(1)	(2)	(3)	(4)
VARIABLES	Prep. Labor	Harvest Labor	Prep. Labor	Harvest Labor
# Children	0.00126	0.00259	0.00363	-0.0102
	(0.0150)	(0.0174)	(0.0142)	(0.0114)
# Adult members	0.0788***	0.0842***	-0.0264**	-0.0250**
	(0.0136)	(0.0164)	(0.0130)	(0.0100)
# Senior members	0.102*	0.0211	0.0174	-0.0235
	(0.0521)	(0.0660)	(0.0501)	(0.0441)
Persons per km^2	-0.0244	-0.0360	0.0576*	-0.00926
	(0.0325)	(0.0363)	(0.0309)	(0.0230)
Total HH Non-farm assets	-0.169*	-0.0915	0.0466	0.0649
	(0.0867)	(0.119)	(0.0674)	(0.0559)
Total farm assets	-0.0384	0.0181	0.0475	-0.0127
	(0.0352)	(0.0392)	(0.0324)	(0.0245)
Animal units	0.190**	0.211**	-0.0399	0.0583
	(0.0792)	(0.0954)	(0.0730)	(0.0582)
HH head's age	-2.641***	-1.284**	0.0709	-0.306
	(0.564)	(0.609)	(0.557)	(0.407)
HH head's educ	0.0273	-0.117***	-0.00161	0.00400
	(0.0376)	(0.0425)	(0.0328)	(0.0260)
HH Head's Gender	0.296***	0.187*	-0.0478	-0.0112
	(0.0855)	(0.1000)	(0.0690)	(0.0534)
HH experienced a death	-0.0954	0.00498	0.0341	0.0244
	(0.0909)	(0.0980)	(0.0806)	(0.0601)
Local agricultural wage	-0.184	-0.629**	-0.228	0.0692
	(0.252)	(0.254)	(0.166)	(0.163)
Plot mgs. all women	-0.109*	-0.207***	0.0409	0.00642
	(0.0622)	(0.0673)	(0.0481)	(0.0365)
Plot mgs. mixed gender	-0.223***	-0.241**	-0.00990	-0.0109
	(0.0823)	(0.0954)	(0.0672)	(0.0556)
Plot mgrs. avg. educ.	0.0503**	0.0693**	0.0440**	0.00405
	(0.0248)	(0.0277)	(0.0195)	(0.0167)
Plot mgrs. avg. age	0.833***	0.611***	0.0752***	0.105***
	(0.0429)	(0.0420)	(0.0288)	(0.0239)
Plot mgrs. avg. BMI	-0.0467**	-0.0643***	-0.0209	-0.0203
	(0.0230)	(0.0244)	(0.0171)	(0.0152)
Plot mgr. is HH head	-0.242***	-0.212***	0.0482	-0.0721
	(0.0746)	(0.0752)	(0.0572)	(0.0473)
Soil & Plot Slope Controls	yes	yes	yes	yes
Household FE	yes	yes	yes	yes
District-year FE	yes	yes	yes	yes
Observations	19,780	19,780	19,780	19,780
Number of Households	5,768	5,768	5,768	5,768

Cluster-Robust standard errors in parentheses; clustered at household level
*** p<0.01, ** p<0.05, * p<0.1

Table 8: Plot-level Family Labor Demand

	Family		Hired	
	(1)	(2)	(3)	(4)
VARIABLES	Prep. Labor	Harvest Labor	Prep. Labor	Harvest Labor
Plot area in acres^2	0.395***	0.290***	0.215***	0.0843***
	(0.0236)	(0.0253)	(0.0220)	(0.0171)
Total plot expenditure	0.0188***	0.0201***	0.116***	0.0512***
	(0.00232)	(0.00256)	(0.00261)	(0.00196)
Collective plot	0.424***	0.296***	0.0177	0.00919
	(0.0786)	(0.0913)	(0.0642)	(0.0534)
Plot is rented in	0.120**	0.0794	-0.395***	-0.138***
	(0.0481)	(0.0568)	(0.0605)	(0.0489)
Plot is irrigated	0.0794	0.294***	-0.0597	0.0650
	(0.0731)	(0.0819)	(0.0782)	(0.0636)
Kg. of organic fert.	0.0352***	0.0202***	-0.0220***	-0.0105**
	(0.00458)	(0.00560)	(0.00519)	(0.00509)
Plot is intercropped	0.209***	0.0741***	-0.00613	-0.0205
	(0.0197)	(0.0226)	(0.0187)	(0.0155)
Plot uses improved seeds	0.119***	-0.00333	-0.170***	-0.0490***
	(0.0250)	(0.0285)	(0.0227)	(0.0190)
Plot dist. to household (in miles)	0.0219***	0.00885***	0.0151***	0.0113***
	(0.00278)	(0.00296)	(0.00272)	(0.00206)
Area of all other plots	0.00247	0.00282*	-0.00139	-0.00243*
	(0.00185)	(0.00145)	(0.00143)	(0.00132)
Plot value	0.0470***	0.0642***	0.0153*	0.0337***
	(0.0100)	(0.0109)	(0.00810)	(0.00606)
Value all other plots	-0.0112***	-0.00665*	0.00986***	0.00117
	(0.00306)	(0.00343)	(0.00307)	(0.00244)
Soil & Plot Slope Controls	yes	yes	yes	yes
Household FE	yes	yes	yes	yes
District-year FE	yes	yes	yes	yes
Observations	19,780	19,780	19,780	19,780
Number of Households	5,768	5,768	5,768	5,768

Cluster-Robust standard errors in parentheses; clustered at household level

*** p<0.01, ** p<0.05, * p<0.1

significant determinant of the amount of fertilizer applied to the plot. The effect size is small meaning a 1% increase in adult household members results in a .01% increase in fertilizer. The fact that household size is a significant determinant of fertilizer use constitutes a rejection of the separation hypothesis.

Organic fertilizer use is decreasing in area planted, as well as decreasing in area planted to all other plots indicating that organic fertilizer may be "stretched too thin." Organic fertilizer use is also increasing in plot expenditure, though the effect is very small. Rented plots receive less fertilizer, and irrigated plots receive much less fertilizer as well. The fact that the coefficient of rented plots is statistically significant and negative mirrors the results of Gavian and Fafchamps (1996), who find that tenure status affects manuring in Niger since organic fertilizer is a short-term investment whose benefits may last beyond a single agricultural season. The size of the household's herd is also a highly significant determinant of the use of organic fertilizer, which reflects the primary findings in Gavian and Fafchamps (1996), who also find that application of manure is determined by the amount of livestock in a household's herd. In column 2, the animal portfolio variable is also statistically significant, meaning there is another indication of separation when using a different indicator for the size of the households animal holdings. The remainder of the plot-level variables provide results which are largely consistent with expectations.

4.4 Robustness Checks

Due to evidence of recall bias in data collection, some of which came from Tanzania itself, I have included a robustness check that adds dummies for the month in which the survey interview was conducted (Beegle et al., 2012). These dummies are also included in all subsequent robustness checks unless otherwise noted. These results are excluded for brevity, but the results remain largely unchanged, though the interview-month dummies are statistically significant in some cases.

4.4.1 Check 1 - Endogenous HH Size

According to a paper by Grimard (2000), endogeneity of household demographics and composition to agricultural decisions is a significant concern in the context of Cote d'Ivoire, where large kinship networks facilitate the movement of family members to and from regions in need

Table 9: Fertilizer Factor Allocation Regressions

VARIABLES	(1) Kg. Fert. per Acre	(2) Kg. Fert. per Acre
# Children	-0.00454	-0.000437
	(0.00303)	(0.00304)
# Adults	0.00846***	0.0134***
	(0.00318)	(0.00324)
# Seniors	-0.00468	-0.000167
	(0.0112)	(0.0113)
Persons per sq. km	0.00328	0.00298
	(0.00248)	(0.00253)
Age HH Head	-0.0242	-0.0160
	(0.0207)	(0.0210)
Educ HH Head	0.00419	0.00555
	(0.00611)	(0.00617)
Gender HH Head	-0.0103	-0.00643
	(0.0153)	(0.0156)
HH exp a death	-0.0170	-0.0178
	(0.0190)	(0.0192)
HH Assets	0.00836***	0.0104***
	(0.00308)	(0.00312)
Farm Assets	0.00928**	0.0123**
	(0.00467)	(0.00525)
Animal Units	0.207***	
	(0.0157)	
Animal Holdings		0.00658***
		(0.000910)
Observations	19,780	19,780
Number of Households	5,768	5,768
Household FE	yes	yes
District-wave FE	yes	yes
Soil and Slope Controls	yes	yes

Cluster-Robust standard errors in parentheses; s.e. clustered at household level
*** p<0.01, ** p<0.05, * p<0.1

Table 9: Plot-level Fertilizer Factor Allocation

VARIABLES	(1) Kg. Fert. per Acre	(2) Kg. Fert. per Acre
Area Planted	-0.0388***	-0.0297***
	(0.00737)	(0.00742)
Plot Exp.	0.00604***	0.00597***
	(0.00100)	(0.00101)
Collective	-0.0232	-0.0298
	(0.0230)	(0.0236)
Plot is Rented	-0.123***	-0.128***
	(0.0247)	(0.0249)
Irrigated	0.334***	0.334***
	(0.0555)	(0.0555)
Intercropped Plot	0.0764***	0.0696***
	(0.0102)	(0.0102)
Improved Seeds Used	0.0171	0.0166
	(0.0111)	(0.0111)
Dist. to HH	-0.0195***	-0.0198***
	(0.00113)	(0.00114)
Area Planted OP	-0.0296***	-0.0214***
	(0.00549)	(0.00549)
Plot Value	0.0154***	0.0164***
	(0.00332)	(0.00336)
Plot Mgr is Head	0.0441**	0.0431*
	(0.0219)	(0.0223)
All Female	0.0259	0.0147
	(0.0177)	(0.0179)
Mixed Gender Mgr.	0.0283	0.0307
	(0.0241)	(0.0247)
Educ Mgr.	0.00109	-0.00204
	(0.00836)	(0.00842)
Age of Mgr.	0.00156	0.00274
	(0.0114)	(0.0115)
BMI of Mgr.	-0.00155	-0.000375
	(0.00555)	(0.00564)
Avg Age Prep Labor	0.0114*	0.0114*
	(0.00598)	(0.00605)
Avg BMI Prep Labor	0.00278	0.00300
	(0.00340)	(0.00350)
Avg Educ Prep Labor	0.0153***	0.0158***
	(0.00588)	(0.00597)
Observations	19,780	19,780
Number of Household	5,768	5,768
Household FE	yes	yes
District-wave FE	yes	yes
Soil and Slope Controls	yes	yes

Cluster-Robust standard errors in parentheses; s.e. clustered at household level
*** p<0.01, ** p<0.05, * p<0.1

of agricultural labor. In Tanzania, by contrast, the large distances make this type of movement, I argue, much less of a concern. Nevertheless, this question can be analyzed using a robustness check.

For this robustness check, to combat concerns with respect to the endogeneity of household members to household labor demand, I exclude all labor which was carried out on the plot by household members who have recently joined the household as a measure to control against endogeneity of household composition to agricultural labor decisions. Based on the survey questionnaire it is possible to identify which household members have joined the household in the past year and for what reason they have moved. In this robustness check, all labor contributions by survey participants who reported moving in the last year due to acquiring agricultural land or for work purposes are excluded. The test in this case still strongly rejects labor market completeness and the results can be found in Table 11.

4.4.2 Check 2 - Farm Size Check

The third robustness check, found in Table 15, evaluates whether farms of different sizes have different demands for labor. Farms are broken into quintiles based on the area under control by each farm. The smallest quintile of farms are less than a football field, the largest quantile farms are over ten football fields in size. All tests still reject labor market completeness, although households in the largest quintile of farms appear to be the most constrained in their labor use.

5 Conclusion

This paper uses high-quality panel data from Tanzania to examine labor market inefficiencies. I first check for differences in the productivity of family and hired labor. Using two tests I find that hired labor is more efficient than family labor, though in the harvest season the differential in productivities between hired and family labor decreases according to the Deolalikar-Vijverberg test. This result is important because differentials between family and hired labor are considered to be an potential source of labor market inefficiency.

In all specifications my test rejects the completeness of labor markets, and confirms the non-separable nature of household production and consumption decisions. In all cases, increases in the number of working adults in the household results in increases in labor applied to the household farm, measured at the plot level. The same results hold for fertilizer, with fertilizer application being highly reliant on household characteristics such as the size of the household's herd and assets

References

Bardhan, P. K. (1973). Size, productivity, and returns to scale: An analysis of farm-level data in indian agriculture. *Journal of political Economy*, 81(6):1370–1386.

Beegle, K., Carletto, C., Himelein, K., et al. (2012). Reliability of recall in agricultural data. *Journal of Development Economics*, 98(1):34–41.

Benjamin, D. (1992). Household composition, labor markets, and labor demand: testing for separation in agricultural household models. *Econometrica: Journal of the Econometric Society*, pages 287–322.

Card, D. E. et al. (1987). *Supply and demand in the labor market*. Number 228. Industrial Relations Section, Princeton University.

De Janvry, A. and Sadoulet, E. (2006). Progress in the modeling of rural households' behavior under market failures. In *Poverty, inequality and development*, pages 155–181. Springer.

Deolalikar, A. B., Vijverberg, W. P., et al. (1987). A test of heterogeneity of family and hired labour in asian agriculture. *Oxford Bulletin of Economics and Statistics*, 49(3):291–305.

Frisvold, G. B. (1994). Does supervision matter? some hypothesis tests using indian farm-level data. *Journal of Development Economics*, 43(2):217–238.

Gavian, S. and Fafchamps, M. (1996). Land tenure and allocative efficiency in niger. *American Journal of Agricultural Economics*, 78(2):460–471.

Grimard, F. (2000). Rural labor markets, household composition, and rainfall in côte d'ivoire. *Review of Development Economics*, 4(1):70–86.

LaFave, D. and Thomas, D. (2016). Farms, families, and markets: New evidence on completeness of markets in agricultural settings. *Econometrica*, 84(5):1917–1960.

Rose, A. N. and Bright, E. A. (2014). The landscan global population distribution project: current state of the art and prospective innovation. *Oak Ridge, TN: Oak Ridge National Laboratory (ORNL).*

Thorbecke, E. (1993). Impact of state and civil institutions on the operation of rural market and nonmarket configurations. *World Development*, 21(4):591–605.

Variables	N	mean	median	sd	min	max
# of Children	19,782	2.15	2	1.90	0	26
# Adult Members	19,782	3.27	3	1.99	0	29
# Senior Members	19,782	0.27	0	0.55	0	3
Persons per sq. km*	19,782	1.40	1.51	0.55	0	2.51
Total HH Assets*	19,782	2.73	2.75	0.14	0	3.16
Total Farm Assets*	19,782	1.05	1.28	0.51	0	1.42
Total Animal Units*	19,782	0.23	0.05	0.31	0	1.09
Age of HH head*	19,782	1.58	1.58	0.0658	1.34	1.74
Years Educ. HH Head*	19,782	0.67	0.96	0.54	0	1.42
Gender of HH Head	19,782	0.14	0	0.27	0	0.69
HH Experienced a Death	19,782	0.05	0	0.16	0	0.53
Agricultural Wage*	19,782	2.69	2.69	0.08	1.98	2.78
Managers All Women	19,782	0.22	0	0.41	0	1
Managers Mixed-Gender	19,782	0.47	0	0.50	0	1
Avg. Education of Mgrs.*	19,782	2.31	2.83	0.93	0.69	3.85
Avg. Age of Mgrs.*	19,782	3.71	3.81	0.62	0.69	4.62
Avg. BMI of Mgrs.*	19,782	2.53	3.08	1.20	0.69	9.90
Plot Manager is Head	19,782	0.92	1	0.27	0	1
Plot Area (acres sq.)	19,782	0.95	0.74	0.70	0	6.40
Total Plot Expense*	19,782	5.42	7.78	5.24	0	15.84
Collectively Farmed Plot	19,782	0.49	0	0.5	0	1
Plot is Retned	19,782	0.04	0	0.19	0	1
Plot is Irrigated	19,782	0.02	0	0.14	0	1
Organic Fertilizer Used (kg)*	19,782	0.64	0	1.89	0	10.37
Plot is Intercropped	19,782	0.41	0	0.49	0	1
Plot Uses Improved Seeds	19,782	0.36	0	0.48	0	1
Distance to Household*	19,782	2.95	1.3	4.05	0	18
Area of All Other Plots*	19,782	4.41	1.89	13.11	0	620
Est. Value of Plot*	19,782	13.27	13.16	1.63	0	22.84
Value of All Other Plots*	19,782	11.54	13.46	5.44	0	21.96
Plot Soil Type	19,780	2.05	2	0.68	1	4
Plot Soil Quality	19,782	1.61	2	0.60	1	3
Slope of Plot	19,782	1.75	1	0.99	1	4
Family Harvest Labor*	19,782	2.36	2.40	1.34	0	6.29
Hired Harvest Labor*	19,782	0.33	0	0.86	0	5.29
Family Prep Labor*	19,782	3.39	3.61	1.33	0	6.78
Hired Prep Labor*	19,782	0.66	0	1.21	0	5.80

*variables are in log form

Soil type: 1=sandy; 2=loam; 3=clay; 4=other

Soil quality: 1=good; 2=average; 3=bad

Plot slope: 1=flat bottom; 2=flat top; 3=slightly sloped; 4=very steep

Table 10: **Summary Statistics of Plot-level Regression Variables**

Table 11: Robustness 2 - Endogenous HH Size Check

	Family		Hired	
	(1)	(2)	(3)	(4)
VARIABLES	Prep. Labor	Harvest Labor	Prep. Labor	Harvest Labor
# children (no migrants)	0.0622	0.0167	-0.0159	-0.0430
	(0.0569)	(0.0528)	(0.0379)	(0.0312)
# adult members (no migrants)	0.0216	0.161**	-0.0932**	-0.132***
	(0.0739)	(0.0629)	(0.0445)	(0.0354)
# senior members (no migrants)	-0.394***	-0.307***	-0.00561	-0.0457
	(0.119)	(0.102)	(0.0701)	(0.0601)
Persons per km^2	0.0124	-0.0180	0.0521*	-0.0109
	(0.0444)	(0.0410)	(0.0315)	(0.0233)
Non-farm HH Assets	-0.287**	-0.137	0.0517	0.0796
	(0.117)	(0.127)	(0.0668)	(0.0561)
Total Farm Assets	-0.0667	-0.00658	0.0446	-0.00965
	(0.0459)	(0.0443)	(0.0324)	(0.0241)
Animal Units	0.315***	0.241**	-0.0352	0.0584
	(0.107)	(0.105)	(0.0719)	(0.0574)
Age of HH Head	-2.160***	0.748	0.0773	-0.334
	(0.751)	(0.681)	(0.547)	(0.401)
Yrs. Educ of HH Head	0.0787*	-0.114***	-0.00495	0.00134
	(0.0466)	(0.0436)	(0.0328)	(0.0259)
Gender of HH Head	0.244**	0.166	-0.0457	-0.00718
	(0.105)	(0.103)	(0.0687)	(0.0527)
HH experienced a death	-0.0571	-0.0468	0.0298	0.0218
	(0.108)	(0.101)	(0.0802)	(0.0598)
Agricultural Wage	0.0179	-0.603**	-0.228	0.0330
	(0.322)	(0.261)	(0.167)	(0.165)
Plot mgrs. all women	-0.0540	-0.175***	0.0428	0.0110
	(0.0712)	(0.0657)	(0.0481)	(0.0362)
Plot mgrs. mixed gender	-0.197*	-0.286***	-0.0100	-0.00720
	(0.117)	(0.0963)	(0.0665)	(0.0545)
Avg. Mgr. Educ.	0.0585*	0.0543**	0.0456**	0.00692
	(0.0312)	(0.0276)	(0.0195)	(0.0166)
Avg. Mgr. Age	0.809***	0.179***	0.0737**	0.104***
	(0.0486)	(0.0377)	(0.0288)	(0.0237)
Avg. Mgr. BMI	-0.0188	0.0311	-0.0203	-0.0204
	(0.0276)	(0.0230)	(0.0171)	(0.0152)
Plot manager is head	-0.324***	-0.102	0.0500	-0.0735
	(0.0840)	(0.0706)	(0.0567)	(0.0465)
Observations	19,774	19,774	19,774	19,774
Number of Households	5,766	5,766	5,766	5,766
Soil & Plot Slope Controls	yes	yes	yes	yes
Household FE	yes	yes	yes	yes
District-year FE	yes	yes	yes	yes

Cluster-robust standard errors in parentheses; clustered at household level

*** p<0.01, ** p<0.05, * p<0.1

Table 11: Robustness 2 - Endogenous HH Size Check

	Family		Hired	
	(1)	(2)	(3)	(4)
VARIABLES	Prep. Labor	Harvest Labor	Prep. Labor	Harvest Labor
Plot area	0.336***	0.231***	0.217***	0.0861***
	(0.0284)	(0.0247)	(0.0220)	(0.0170)
Total Plot Expenditure	0.0175***	0.0136***	0.116***	0.0512***
	(0.00268)	(0.00250)	(0.00262)	(0.00196)
Collective Plot	0.424***	0.284***	0.0186	0.00846
	(0.113)	(0.0915)	(0.0635)	(0.0525)
Plot is Rented	0.0695	0.0569	-0.395***	-0.140***
	(0.0635)	(0.0597)	(0.0605)	(0.0490)
Plot is Irrigated	0.00561	0.211***	-0.0602	0.0615
	(0.0799)	(0.0786)	(0.0780)	(0.0640)
Kg. of Organic Fert.	0.0340***	0.0157***	-0.0220***	-0.0102**
	(0.00543)	(0.00558)	(0.00518)	(0.00509)
Plot is Intercropped	0.202***	0.0360*	-0.00730	-0.0222
	(0.0232)	(0.0218)	(0.0187)	(0.0155)
Improved Seeds Used	0.108***	-0.00150	-0.169***	-0.0492***
	(0.0300)	(0.0291)	(0.0227)	(0.0189)
Plot Dist. to Household	0.0218***	0.00724**	0.0151***	0.0115***
	(0.00306)	(0.00291)	(0.00271)	(0.00206)
Area Planted to Other Plots	-0.000166	0.00148	-0.00135	-0.00233*
	(0.00275)	(0.00139)	(0.00144)	(0.00131)
Plot Value	0.0393***	0.0590***	0.0142*	0.0329***
	(0.0108)	(0.0107)	(0.00809)	(0.00605)
Value All Other Plots	-0.00628	-0.00699*	0.00979***	0.000846
	(0.00412)	(0.00376)	(0.00306)	(0.00245)
Observations	19,774	19,774	19,774	19,774
Number of Households	5,766	5,766	5,766	5,766
Soil & Plot Slope Controls	yes	yes	yes	yes
Household FE	yes	yes	yes	yes
District-year FE	yes	yes	yes	yes

Cluster-robust standard errors in parentheses; clustered at household level
*** p<0.01, ** p<0.05, * p<0.1

	Family		Hired	
	(1)	(2)	(3)	(4)
VARIABLES	Prep. Labor	Harvest Labor	Prep. Labor	Harvest Labor
# children	0.00450	0.00451	0.00405	-0.0102
	(0.0150)	(0.0173)	(0.0142)	(0.0115)
# adult members	0.0805***	0.0828***	-0.0250*	-0.0247**
	(0.0135)	(0.0164)	(0.0130)	(0.00995)
# senior members	0.104**	0.0203	0.0154	-0.0259
	(0.0528)	(0.0658)	(0.0500)	(0.0445)
Persons per km^2	-0.0102	-0.0216	0.0507	-0.0116
	(0.0332)	(0.0371)	(0.0315)	(0.0234)
Non-farm HH Assets	-0.173**	-0.0990	0.0441	0.0694
	(0.0876)	(0.122)	(0.0675)	(0.0561)
Total Farm Assets	-0.0300	0.0256	0.0445	-0.0114
	(0.0349)	(0.0389)	(0.0325)	(0.0243)
Animal Units	0.192**	0.196**	-0.0264	0.0620
	(0.0794)	(0.0953)	(0.0725)	(0.0581)
Age of HH Head	-2.347***	-1.309**	0.166	-0.238
	(0.565)	(0.614)	(0.558)	(0.408)
Yrs. Educ of HH Head	0.0270	-0.118***	-0.00301	0.00240
	(0.0374)	(0.0423)	(0.0327)	(0.0258)
Gender of HH Head	0.283***	0.188*	-0.0488	-0.0121
	(0.0858)	(0.101)	(0.0691)	(0.0531)
HH experienced a death	-0.0870	-0.00207	0.0318	0.0242
	(0.0908)	(0.0974)	(0.0804)	(0.0594)
Agricultural Wage	-0.225	-0.621**	-0.226	0.0262
	(0.253)	(0.253)	(0.167)	(0.164)
Plot mgrs. all women	-0.117*	-0.211***	0.0404	0.00810
	(0.0618)	(0.0674)	(0.0481)	(0.0364)
Plot mgrs. mixed gender	-0.218***	-0.251***	-0.0109	-0.00482
	(0.0819)	(0.0957)	(0.0669)	(0.0550)
Avg. Mgr. Educ.	0.0509**	0.0696**	0.0442**	0.00600
	(0.0248)	(0.0276)	(0.0194)	(0.0167)
Avg. Mgr. Age	0.835***	0.611***	0.0741**	0.104***
	(0.0429)	(0.0419)	(0.0288)	(0.0238)
Avg. Mgr. BMI	-0.0439*	-0.0631***	-0.0196	-0.0184
	(0.0230)	(0.0243)	(0.0171)	(0.0152)
Plot manager is head	-0.250***	-0.216***	0.0503	-0.0740
	(0.0742)	(0.0751)	(0.0569)	(0.0469)
Observations	19,774	19,774	19,774	19,774
R-squared	0.278	0.165	0.330	0.165
Number of Households	5,766	5,766	5,766	5,766

Cluster-Robust standard errors in parentheses; clustered at household level
*** p<0.01, ** p<0.05, * p<0.1

Table 12: Robustness check 3: quintiles of farm size

	Family		Hired	
	(1)	(2)	(3)	(4)
VARIABLES	Prep. Labor	Harvest Labor	Prep. Labor	Harvest Labor
Plot Area in Acres^2	0.434***	0.294***	0.230***	0.0949***
	(0.0255)	(0.0272)	(0.0233)	(0.0184)
Total Plot Expenditure	0.0186***	0.0200***	0.116***	0.0512***
	(0.00232)	(0.00257)	(0.00261)	(0.00196)
Collective Plot	0.419***	0.302***	0.0177	0.00375
	(0.0783)	(0.0916)	(0.0639)	(0.0529)
Plot is rented in	0.116**	0.0827	-0.397***	-0.143***
	(0.0480)	(0.0569)	(0.0606)	(0.0490)
Plot is irrigated	0.0851	0.296***	-0.0578	0.0643
	(0.0732)	(0.0819)	(0.0776)	(0.0638)
Organic fertilizer used (in kg)	0.0349***	0.0197***	-0.0221***	-0.0103**
	(0.00458)	(0.00560)	(0.00517)	(0.00509)
Plot is intercropped	0.208***	0.0774***	-0.00721	-0.0220
	(0.0197)	(0.0225)	(0.0187)	(0.0155)
Plot uses improved seeds	0.116***	-0.00370	-0.170***	-0.0495***
	(0.0250)	(0.0284)	(0.0227)	(0.0189)
Plot distance to HH	0.0210***	0.00866***	0.0148***	0.0113***
	(0.00278)	(0.00296)	(0.00272)	(0.00207)
Area of all other plots	0.00446**	0.00306**	-0.000665	-0.00202
	(0.00182)	(0.00151)	(0.00145)	(0.00134)
Plot est. value	0.0467***	0.0647***	0.0138*	0.0326***
	(0.00995)	(0.0109)	(0.00808)	(0.00606)
Est. value of all other plots	-0.00650**	-0.00605	0.0113***	0.00225
	(0.00327)	(0.00375)	(0.00336)	(0.00262)
Second Farm Size Quintile	-0.0190	0.0327	-0.0257	-0.0109
	(0.0461)	(0.0533)	(0.0376)	(0.0299)
Third Farm Size Quintile	-0.0863	0.0438	-0.0212	-0.0308
	(0.0554)	(0.0622)	(0.0477)	(0.0363)
Fourth Farm Size Quintile	-0.178***	-0.0123	-0.0420	-0.0685
	(0.0653)	(0.0730)	(0.0577)	(0.0450)
Fifth Farm Size Quintile	-0.336***	-0.0152	-0.120	-0.0793
	(0.0820)	(0.0872)	(0.0752)	(0.0578)
Observations	19,774	19,774	19,774	19,774
R-squared	0.278	0.165	0.330	0.165
Number of Households	5,766	5,766	5,766	5,766

Cluster-Robust standard errors in parentheses; clustered at household level
*** p<0.01, ** p<0.05, * p<0.1

Table 12: Robustness check 3: quintiles of farm size