How does recall bias in farm labor impact separability tests?

Bailey Peterson-Wilhelm and Benjamin Schwab*

September 2022

Abstract

In the agricultural household literature, empirical tests of separability between production and consumption decisions commonly exploit theoretical predictions of household labor allocation. Many of these studies rely on data that asks respondents to recall labor usage over the entire growing season. Two recent field experiments in Tanzania and Ghana show that such labor use data, collected at the end of the growing season, is a systematically unreliable measure of actual labor allocation. In this study, we examine how inaccurate measures of labor use influence the reliability of market failure tests based on separability. In Ghana we find no statistical evidence that recall bias influences the reliability of the separability test. In Tanzania we find that recall bias increases the probability that such tests fail to reject separability. Thus, we find partial evidence that classic tests based on typical household survey labor data may erroneously conclude that markets are adequately functioning.

Keywords: Separability, Recall bias, Ghana, Tanzania, Labor

Acknowledgement The authors acknowledge funding from National Science Foundation grant #1828571.

*Schwab: Associate Professor, Department of Agricultural Economics, Kansas State University, benschwab@ksu.edu; Peterson-Wilhelm: PhD Student, Department of Agricultural Economics, Kansas State University, bailey27@ksu.edu

1 Introduction

The classic agricultural household model demonstrates the manner in which farm production decisions are nested within the household's optimal consumption problem. The joint structure implies several threats to the assumption that agricultural production decisions are made separately from an agricultural household's consumption decision. The separation assumption requires environments with complete markets. When a household cannot access certain input or output markets for example, then separation is unlikely to hold. Benjamin (1992) devised the classic empirical test of separability by focusing on the household labor allocation decision. If the size and nature of the family household labor endowment statistically influences its total labor demand for agricultural production, separability does not hold (Benjamin, 1992). Recent work expanding the separability test continues to rely on measures of labor use (Dillon and Barrett, 2017; Dillon et al., 2019).

A common practice among these studies is to use data collected via a survey conducted at the end of the growing season. These agricultural modules are the foundation of often used data sets like the World Bank Living Standards Measurement Study and Integrated Surveys on Agriculture (LSMS-ISA). However, two recent field experiments, one in Tanzania and one in Ghana, have revealed systematic errors in such 'one shot' survey based labor measures. These experiments find disparities in labor measures obtained via end-of-season surveys and intensive weekly surveys.

In Ghana, Gaddis et al. (2021) randomized the data collection procedure by assigning some agricultural households to weekly data collection and others to the traditional end of growing season method. They find significant recall bias in the end of season method driven by 'listing bias', or systematic failure to report on more marginal plots. Such bias tends to inflate per unit measures of labor productivity, as total household labor is underestimated due to the omission of certain plots and laborers.

The Gaddis et al. (2021) results complement a similar study by Arthi et al. (2018), which compares weekly surveys to two different end of season methods in Tanzania. The Tanzanian experiment similarly finds substantial recall bias. As in Ghana, omission of plots appears to play a substantial role in the discrepancies between intensive and end-of-season measures of labor use.

Both experiments identify consistent sources of mismeasurement of household labor supply from end-of-season recall. Households fail to recall some plots, and thus do not account for labor on them. They also exaggerate the quantity of labor on the plots they do list. However, the implications for household level aggregates of these competing sources of error differ between the two settings.

In Tanzania, the two sources of bias essentially even out, as Arthi et al. (2018) notes that household level aggregation "appears to cancel the competing biases arising from overreporting at the intensive margin and underreporting at the extensive margin."

In contrast, in Ghana, the depressing impact of listing bias dominates. As Gaddis et al. (2021) note,"listing bias dominates other forms of bias at the plot level and household level, even though recall households omit relatively marginal plots and household workers." Overall, then, it is unclear how further aggregating household labor demand will be impacted by these sources of mismeasurement.

While both Gaddis et al. (2021) and Arthi et al. (2018) focus on the implications of labor mismeasurement for understanding smallholder agricultural productivity, neither considers the potential influence of recall bias for empirical separability tests. Conceptually, the direction and magnitude of the effects are ambiguous.

Listing bias arising from a failure to list household members may reduce a true link to overall labor demand, thus causing an underestimation of the key test statistic utilized to reject the null hypothesis of separable labor allocation decisions. However, if marginal plots are relatively less likely to utilize household labor even if separability holds, traditional estimates may overstate the degree of non-separability.

Further complicating predictions, neither Gaddis et al. (2021) nor Arthi et al. (2018) describe how recall bias impacts the measurement of *hired* labor. Because empirical separability tests are based on total labor demand (i.e. hired and household labor), hired labor decisions are critical. Indeed, assessing the 'missingness' of the labor factor market is a key goal in the empirical literature.

Measurement error in hired labor may not follow the same pattern as household labor. The necessity of making explicit payments, for example, may increase the salience of hired labor allocations and reduce the types of cognitive and accounting biases that plague end-of-season surveys. Further, the propensity to hire labor may not be uniform across all plots, particularly the 'marginal' plots that Gaddis et al. (2021) and Arthi et al. (2018) find most likely to be omitted.

In this paper, we utilize the data from the original Gaddis et al. (2021) and Arthi et al. (2018) in Ghana and Tanzania to determine how recall and listing bias have impacted the substantial empirical literature on separability tests.

For the first time, we assess how hired labor is impacted by the forms of recall and listing bias identified in the original paper. The experimental settings in Ghana, where most households hire labor, and Tanzania, where hired labor is rare, provide contrasting settings to assess how mismeasurement impacts hiring data.

Including hired labor allows us to determine the influence of recall error on total household labor demand, not only household labor supply. We then use that data to estimate the classic labor separability tests using the Ghana and Tanzania data, and determine how recall bias impacts the validity of the tests common in the literature.

In Ghana, where hired labor is more common, we find little evidence that hired labor suffers from recall bias at either the household aggregate level or on a per acre basis. However, total household labor demand remains significantly underestimated by recall households as a result of the listing bias from household labor supply. In Tanzania, where hired labor is more rare, recall households significantly underestimate the amount of hired labor used.

In both Ghana and Tanzania, we strongly reject the null of separability for both the end-of-season and weekly data collection arms. Despite the large discrepancy between aggregate labor demand as measured via the two methods in Ghana, we find recall bias has no affect on the empirical test of separability. In contrast, in Tanzania, we find that the test statistic obtained from end-of-season data collection is underestimated relative to the weekly visit group.

Thus, in one context, recall bias increases the probability that separability tests fail to reject non-separability and that tests based on such data are more likely to erroneously conclude that markets are adequately functioning. The finding that labor measurement problems may lead to underrejection of separation confirms speculation in LaFave and Thomas (2016). Unlike LaFave and Thomas (2016), however, our analysis is able to specifically account for the role of measurement and the importance of listing bias.

2 Research Design and Data

Our data are drawn from the original data measurement experiments carried out by Gaddis et al. (2021) and Arthi et al. (2018) in Ghana and Tanzania, respectively. We discuss each experiment in the following sections.

2.1 Ghana Experiment

The data used for analysis was collected in the Ashanti and Brong Ahafo regions of Ghana during the 2015-2016 agricultural growing season. The 720 surveyed agricultural households from 20 villages were then randomly assigned to one of three data measurement groups: traditional end of season data collection (control), weekly visits by enumerators, and weekly phone interviews. Gaddis et al. (2021) excludes the phone group from the main analysis because of non-response issues. For completeness, we include the data from this group here, though we caution that non-response concerns may be significant. All groups were given an in-person baseline survey before the growing season began. The weekly visits group was surveyed weekly by an in-person enumerator. The weekly phone group was given a similar weekly survey by phone call on a phone provided to households. At the end of the growing season, all groups were given an endline in-person survey.

To maintain consistency, the data was cleaned similarly to Gaddis et al. (2021). Households in all treatment arms that did not complete the endline survey were removed from the sample. Households in both treatment groups that did not complete at least two thirds, or 16, of the weekly surveys were dropped to avoid introducing recall bias into the treatment groups. Child labor is not the focus of this study, so labor of children under 10 years old at the baseline survey was dropped. Household and hired labor hours across all treatment arms were winsorized at the top one percent at the person-plot level. Households that reported using zero labor hours over the season were removed from the sample in addition to households that did not report having any agricultural land.

Aggregate labor demand for each household was constructed by combining household labor and hired labor. The consideration of hired labor is an extension beyond the scope of Gaddis et al. (2021). For both treatment groups, household labor includes labor hours reported at the baseline, weekly hours, and hours reported at the endline. In the baseline survey households were asked to report hours worked by person and plot prior to the initial baseline survey. The weekly survey asked the number of hours worked by person and plot since the last survey. Finally, the endline survey asked the number of hours worked by person and plot since the last weekly visit. Hired labor also includes baseline, weekly, and endline hours reported; however, the questions were asked on the plot level rather than the person-plot level. Control group household labor includes only the hours reported in the endline, where the households were asked to report the number of days spent on each of four activities over the entire growing season by person and plot and the typical day length by activity. Similarly, hired labor for the control group was collected by asking total days spent on each activity by plot for the entire season.

2.2 Tanzania Experiment

The Tanzania data used for this analysis was collected during the 2014 rainy season in the rural Mara Region and were drawn from Arthi et al. (2018). The sample consists of 854 agricultural households spanning 18 enumeration areas, or villages. Within each village the households were randomized to one of four survey arms: weekly visit, weekly phone, traditional recall, and alternative recall. The weekly visit and weekly phone groups were given an in-person baseline survey, weekly surveys, and an in-person endline survey. The recall groups was only surveyed once at the end of the growing season. For the traditional recall group, the labor module asked households the total number of days spent on each of four agricultural tasks through out the growing season and the typical numbers of hours spent per day for each task, which is identical to the labor module used previously in Tanzania and very closely mirrors the labor module for the Ghana recall group. The labor module for the alternative recall group instead asks the total weeks worked over the entire growing season, typical number of days worked per week, and typical hours worked per day, without specifying labor activity. To maintain consistency, the same Ghana cleaning rules were applied to the Tanzania data. In order for the Tanzania treatment groups to complete at least two-thirds of the surveys, households must have completed 19 weekly surveys.

Aggregate labor demand includes both household agricultural labor and hired agricultural labor. Arthi et al. (2018) did not include hired labor in their analysis, so the affect of recall bias in hired labor has been previously unexplored. The revisit and phone treatment groups include household and hired labor reported at the baseline survey, weekly surveys, and endline survey. The traditional and alternative recall groups only include hired and household labor reported in the endline survey.

2.3 Summary Statistics

Summary statistics describing the labor data are presented in Table 1. Household labor is averaged on per worker, per active plot, per acre, and per household levels for all treatment arms in Ghana and Tanzania. Hired labor and aggregate labor demand are averaged on the per active plot, per acre, and per household level for all groups in Ghana and Tanzania. For tests of differences in means, in Ghana the recall group is the reference group and in Tanzania the reference group is the traditional recall group.

In Ghana, the dominant effect of listing bias–manifested in the recall group omitting more marginal plots–is apparent in the treatment arm differences for household labor across the levels of aggregation. At the per person level, the visit group is not statistically different from the recall group. However, at every other level the visit group (and phone group) report statistically more household labor hours than the recall group. The average number of plots for visit and phone groups are also statistically larger than the recall group. The average number of acres per household is smaller, but not significantly so, reflecting the fact that the omitted plots tend to be smaller and more marginal. The recall group's lagging household-level totals indicate that by forgetting to list marginal workers and their labor, in aggregate the recall group is under-reporting household labor.

In contrast, in Tanzania, visit and phone groups are statistically different from the traditional recall group in granular measures but the difference dissipates as the labor hours are aggregated. The two treatment groups also report significantly more acres and significantly more plots compared to the two control groups. The statistical difference at the per person and per plot level combined with the large difference in plots and acreage, could indicate that the traditional recall group is reporting an accurate number of household labor hours in total but are incorrectly allocating the labor hours. Since the number of plots is lower for the recall group, they are attributing household labor completed on unlisted plots to the plots that are listed. It is unclear whether the downward bias (from failing to list plots) evenly cancels out the upward bias (from overreporting labor hours on listed plots) due to chance or from some heuristic process common to farmers outside of this setting.

The impact of recall bias on hired labor also differs in both countries. As hired labor requires payment, it is possible that respondents are able to more accurately recall hired labor hours than household labor. In Ghana, where hired labor is common, there is little evidence that hired labor is affected by recall bias. At the per-plot, only the phone group, which suffers from potentially non-random non-response concerns, differs meaningfully from measures of hired labor. For all other levels of aggregation, hired labor measures are strikingly similar across treatment arms.

In Tanzania, however, recall bias appears to result in an underestimation of hired labor at the household level. On a per-household basis, the visit group reported nearly 50 percent more hired labor than the traditional recall group, with the difference significant at the 10 percent level. However, the low utilization of hired labor in Tanzania means that the discrepancy accounts for only 2 percent of total labor demand.

Our summary statistics confirm and expand upon the findings of Gaddis et al. (2021) and Arthi et al. (2018). While listing bias and recall bias clearly influence the accuracy of labor modules, the question of how these biases influence common tests of separability remains unanswered.

3 Econometric Specification

Using the survey data on both family and hired labor, we calculate total household labor demand for the agricultural season among all the treatment groups. We then replicate the classic separability tests, as used in Dillon and Barrett (2017), where log total labor demand is regressed on the log of household size. We include measures of treatment status and its interaction with the log of household size.

Our basic regression specification takes the following form:

$$(1)\quad ln(labor_i) = \beta_0 + \beta_1 ln(HHSize_i) + \beta_2 Treat_i + \beta_3 (Treat_i * HHSize_i) + \beta_4 ln(acres_i) + u_i$$

In the standard analysis, if $\hat{\beta_1} > 0$, the null of separability is rejected. In (1), that would be the equivalent of rejecting the null for the end-of-season data collection.

To determine if the weekly visit treatment influences the standard separability test, we focus on β_3. Estimating $\hat{\beta}_3 > 0$ implies that the relationship between household labor demand and household labor composition is underestimated in typical surveys, and that separability tests are susceptible to falsely under-rejecting the null of separability because of labor recall bias. However, $\hat{\beta}_3 < 0$ would imply the opposite, and that the literature showing strong rejections of separability across, for example, Sub-Saharan Africa (SSA), could potentially be an artifact of non-classical measurement error in labor data.

LaFave and Thomas (2016) argue that separability tests based solely on cross-sectional data potentially suffer from endogeneity concerns. Such concerns threaten consistency in estimating $\hat{\beta}_1$. However, identification of β_3 relies solely on the random allocation of measurement treatment (i.e. recall length).

4 Results

4.1 Ghana Results

We first analyze the impact of recall bias on separability tests in Ghana by pooling the phone and visit treatment groups (Table 2). Separability is rejected regardless of model specification. Household size significantly increases labor demand (p <0.01). Other household characteristics, such as household family structure, significantly contribute to labor demand (p <0.01), providing further evidence that separability does not hold. Treatment is positive and significant in both the base and extended model (p <0.01), which is consistent with the conclusion from Table 1 that both visit and phone treatment groups use more labor than the recall group.

The interaction of treatment with household size is close to zero and not significant at traditional significance levels in both model specifications. The null effect of the treatment household size interaction indicates that despite the evidence of substantial recall bias at the household level in traditional labor measurement methods, that bias did not affect the separability test in this context.

Since there was some evidence that plot listings might be susceptible to bias, we considered the treatment effect on acreage in Table 2. Across all of the model specifications acreage significantly contributes to labor demand ($p < 0.01$). Unsurprisingly, as area increases, the demand for labor also increases. The treatment area interaction is also statistically significant ($p < 0.01$); however, it is negative. The negative interaction term indicates that as plots are more accurately accounted for, in this case through additional opportunities to add plots to the plot roster, the impact of acreage on labor demand is reduced by roughly half.

The differences in labor demand across treatment groups prompted us to look closer at the differential impact of each treatment. Table 3 considers the base regression for each treatment independently and then again with separate indicator variables and interaction terms for each treatment.

We find that the results are largely unchanged from the pooled results presented in Table 2. The coefficient on the log of household size remains large, positive, and statistically significant ($p < 0.01$). The treatment effect is still positive and statistically significant ($p < 0.01$) for both treatments in each of the specifications, though the phone treatment effect exceeds the visit treatment effect. The larger phone treatment effect is consistent with the summary statistics in Table 1, where the phone group reported higher labor demand. Once again, each of the treatment household size interaction terms are insignificant. In Ghana, recall and listing bias do not significantly effect the reliability of separability tests.

4.2 Tanzania Results

As in Ghana, we first pool the phone and visit treatment to estimate equation (1). For the pooled results in Tanzania, we omit the alternative recall group altogether, so the pooled weekly visit treatments are estimated against the traditional recall counterfactual only. The resulting estimates in Table 4 include the base and extended model for Tanzania.

Across all specifications, separability is rejected. The household size coefficient is large, positive and significant (p <0.01). The rejection of separability is further indicated, as other household characteristics also significantly contribute to labor demand. In Tanzania, a male head of household increases labor demand (p <0.1) and the household structure also significantly influences labor demand (p <0.05).

In Tanzania the treatment indicator variable is negative and significant across specifications (p <0.05). In contrast to the Ghana results, the treatment interaction with household size is positive and significant (p <0.05). Given an average household size of 6.55, the independent effect of treatment is near zero. The near zero total effect of treatment is consistent with the summary statistics from Table 1, which reveal little difference in labor demand between treatment and control at the aggregated household level. The positive and significant interaction between treatment and household size indicates that the impact of household size on labor demand is an underestimate. Thus, recall bias makes it more likely to falsely fail to reject separability.

The positive and significant effect of acreage on labor demand (p <0.01) is consistent with the Ghana results. As expected, a larger agricultural area would require additional labor. Differing from Ghana, the interaction of acreage and treatment is insignificant in Tanzania. The insignificance of the interaction term is consistent with the hypothesis that households were not omitting labor, but instead mis-attributing labor from unlisted plots to listed plots.

In Table 5, the treatments are again considered separately. Treatment and the treatment interaction are insignificant for the phone group. However, the weekly visit treatment negatively and significantly affects labor demand, a similar finding to that in Table 4. Crucially, the interaction term between the weekly visit treatment and household size is significant and positive (p <0.01). The positive effect of the interaction term indicates that the effect of household size is even larger when recall bias is reduced. When the visit group is considered alone, reducing recall bias increases the effect of household size on labor demand by roughly

40 percent. The effect is even larger, nearly doubling the effect of household size on labor demand, when the full sample is used and each arm is estimated with a separate indicator variable.

When we consider the alternative baseline group as an additional treatment group rather than a control group, we see it behaves similarly to the visit treatment group. The treatment coefficient is negative and significant (p <0.1) and the interaction between treatment and household size is positive and significant (p <0.05) and of similar magnitude. The treatment model in Table 5 also highlights the heterogeneous impact of the different treatment types. The insignificance of the phone group paired with the significance of the visit group suggests the phone treatment does not prevent recall bias.

The estimates in Table 5 indicate that when recall bias is reduced in Tanzania, separability is more strongly rejected. The presence of some significant relationship between the reduction of bias and separability indicates that, at least in Tanzania, recall bias might lead to under-rejection in tests of separability in the literature.

4.3 Robustness Checks

In order to test the robustness of our results, we add a series of additional controls to our existing preferred model in a step-wise procedure. The preferred model in Table 2 & 4 included village level fixed effects, captured using binary enumeration area variables. The first additional set of controls include soil quality and soil type. Next controls for household education were included. For the presented specification, maximum household education consists of binary variables for the highest level of education by any household member. The final set of controls are measurements of off-farm employment. Off-farm employment is measured by number of non-agricultural enterprises, average monthly income of household non-agricultural enterprises, and the share of the household that is employed off-farm.

The results from the step-wise addition of control variables are presented in Table 6 for the full sample and Table 7 by treatment arm. Results are robust to addition of these additional variables. In both Ghana and Tanzania for the full data set and treatment arm subsets, the point estimates for both the measurement treatment effect estimate (β_3) and primary separability test (β_1) remain large unchanged as control variables are included.

The stability of the estimates to household education and agricultural covariates suggests that the lack of panel data, which is commonly used to control for idiosyncratic household labor demand shifters, is unlikely to be driving the results or adversely affecting the identification of our main variable of interest, β_3.

5 Concluding Remarks

The evaluation of the complete markets assumption in rural settings has long relied on labor use data collected via surveys of agricultural households. These surveys typically occur once, after harvest, and ask respondents to recall data over the previous growing season. A recent experimental literature has documented that this 'one-shot' method leads to mismeasurement. Generally, compared to those surveyed frequently, households do not list all the laborers or plots that use household labor, and overstate the amount of labor expended on those plots they do list.

The experimental literature on labor measurement has focused primarily on the implications of survey recall bias for measures of agricultural productivity. As a result, these studies analyze household labor supply, not total labor demand, which includes hired labor.

In this paper, we reanalyze data from labor measurement field experiments in Ghana and Tanzania. We examine the consequences of labor mismeasurement for estimates of total labor demand, which includes hired labor, and for empirical test of separability. Such separability tests estimate the relationship between household composition and total labor demand, and are thus highly dependent on household labor data.

We find that while per plot and per acre measures of hired labor are not generally affected by mismeasurement, recall bias led households in Tanzania, but not Ghana, to understate the total amount of hired labor employed during the growing season.

In both Tanzania and Ghana, the null hypothesis of separability is strongly rejected regardless of data collection frequency. In Ghana, recall bias has no impact on the reliability of the separability test. However, in Tanzania, recall bias causes an underestimation of the separation test statistics. Thus, we demonstrate the possibility that the long empirical literature relying on separation tests may suffer from under-rejection bias, and thus overestimates market completeness.

TABLE 1: SUMMARY STATISTICS	GHANA			TANZANIA			
Group/ Season-wide hours	Visit	Phone	Recall	Visit	Phone	Alternative Recall	Traditional Recall
PER PERSON							
HOUSEHOLD LABOR	337.8	477.8***	333.6	200.1***	223.9***	357.4***	286.8
	(10.9)	(13.4)	(18.9)	(6.5)	(7.3)	(16.0)	(10.9)
NUMBER PERSON-PLOTS	846	722	630	900	928	651	608
PER PLOT							
HOUSEHOLD LABOR	435.9**	571.8***	374.0	182.9***	220.5***	444.1***	345.2
	(18.4)	(19.9)	(21.0)	(7.4)	(9.7)	(22.5)	(14.3)
HIRED LABOR	74.0	30.4***	85.0	9.4*	17.3	9.5	12.6
	(5.4)	(3.5)	(13.3)	(1.0)	(4.3)	(1.5)	(1.9)
TOTAL LABOR	509.9	602.2***	458.9	192.2***	237.8***	453.6***	357.8
	(20.9)	(21.1)	(25.1)	(7.7)	(12.0)	(22.5)	(14.2)
NUMBER OF PLOTS	641	610	562	985	942	524	505
PER ACRE							
HOUSEHOLD LABOR	216.4***	263.9***	134.6	335.5**	399.8	588.0*	432.5
	(11.6)	(14.0)	(11.4)	(28.7)	(39.7)	(72.3)	(36.8)
HIRED LABOR	28.3	25.9	28.7	13.2	15.1	10.5	12.9
	(2.6)	(2.046)	(4.8)	(1.9)	(4.9)	(2.1)	(2.6)
TOTAL LABOR	244.8***	289.8***	163.3	348.7**	414.9	598.5*	445.4
	(12.2)	(14.6)	(12.4)	(28.8)	(40.2)	(72.2)	(36.6)
AVERAGE ACRES	8.4	8.5	8.0	4.0***	3.9***	2.5	2.5
	(0.5)	(0.5)	(0.4)	(0.3)	(0.30)	(0.2)	(0.2)
AVERAGE PLOTS PER HOUSEHOLD	3.0***	3.0***	2.6	5.2****	4.9***	2.8	2.8
	(0.1)	(0.1)	(0.1)	(0.2)	(0.2)	(0.1)	(0.1)
NUMBER OF HOUSEHOLDS	225	224	227	212	213	218	212
PER HOUSEHOLD							
HOUSEHOLD LABOR	1313.4***	1641.7***	925.8	843.9	959.6**	1067.4**	785.3
	(63.7)	(71.6)	(86.4)	(48.4)	(57.2)	(103.8)	(58.0)
HIRED LABOR	211.5	168.2	210.3	46.1*	68.7*	23.6	31.0
	(19.5)	(13.7)	(37.8)	(5.9)	(18.8)	(4.8)	(5.6)
TOTAL LABOR	1524.9***	1809.9***	1136.2	889.9	1028.3**	1091.0**	816.3
	(73.5)	(75.3)	(95.7)	(50.3)	(65.0)	(103.9)	(58.20)
NUMBER OF HOUSEHOLDS	225	224	227	212	213	218	212

*, **, *** denotes statistical difference from Recall group at 1, 5, 10% levels respectively

Table 2: Ghana Separability Tests

	(1)	(2)	(3)
log(hh size)	0.599 (0.072)***	0.554 (0.115)***	0.488 (0.121)***
log(acre)	0.449 (0.044)***	0.434 (0.039)***	0.586 (0.079)***
Head Gender (male=1)	-0.055 (0.097)	-0.115 (0.086)	-0.107 (0.084)
Share of HH: Adult Female	0.842 (0.211)***	0.783 (0.188)***	0.759 (0.189)***
Share of HH: Adult Male	0.852 (0.187)***	0.758 (0.170)***	0.732 (0.169)***
Treated		0.647 (0.187)***	0.937 (0.216)***
TreatedXlog(hh size)		0.002 (0.113)	0.083 (0.120)
TreatedXlog(acre)			-0.225 (0.084)***
R^2	0.41	0.52	0.53
N	676	676	676

*, **, *** denotes statistical significance at 1, 5, 10% levels respectively

Note: Robust standard errors are reported in the parentheses.

Table 3: Ghana Separability Tests- By Treatment Arm

	Revisit Only	Phone Only	By Treatment
log(hh size)	0.530 (0.127)***	0.540 (0.125)***	0.556 (0.116)***
Treated	0.537 (0.211)**	0.733 (0.197)***	
TreatedXlog(hh size)	-0.001 (0.126)	0.020 (0.118)	
log(acre)	0.474 (0.049)***	0.457 (0.057)***	0.432 (0.040)***
Head Gender (male=1)	-0.148 (0.112)	-0.165 (0.126)	-0.107 (0.086)
Share of HH: Adult Female	0.749 (0.258)***	0.668 (0.242)***	0.774 (0.187)***
Share of HH: Adult Male	0.707 (0.240)***	0.671 (0.217)***	0.758 (0.169)***
Visit			0.554 (0.212)***
VisitXlog(hh size)			-0.013 (0.126)
Phone			0.717 (0.194)***
PhoneXlog(hh size)			0.032 (0.117)
R^2	0.50	0.53	0.53
N	452	451	676

*, **, *** denotes statistical significance at 1, 5, 10% levels respectively

Note: Robust standard errors are reported in the parentheses.

Table 4: Tanzania Separability Tests

	(1)	(2)	(3)
log(hh size)	0.776	0.550	0.560
	(0.083)***	(0.139)***	(0.141)***
log(acre)	0.361	0.371	0.341
	(0.038)***	(0.039)***	(0.065)***
Head Gender	0.152	0.151	0.151
(male=1)	(0.088)*	(0.087)*	(0.087)*
Share of HH: Adult Female	0.527	0.464	0.452
	(0.221)**	(0.223)**	(0.224)**
Share of HH: Adult Male	0.317	0.384	0.379
	(0.237)	(0.229)*	(0.230)
Treated		-0.648	-0.656
(Weekly=1)		(0.271)**	(0.272)**
TreatedXlog(hh size)		0.349	0.334
		(0.146)**	(0.148)**
TreatedXlog(acre)			0.044
			(0.074)
R^2	0.39	0.40	0.40
N	574	574	574

*, **, *** denotes statistical significance at 1, 5, 10% levels respectively

Note: Robust standard errors are reported in the parentheses.

Table 5: Tanzania Separability Tests- By Treatment Arm

	Revisit Only	Phone Only	Treatments
log(hh size)	0.584	0.494	0.580
	(0.149)***	(0.148)***	(0.135)***
log(acre)	0.389	0.352	0.370
	(0.045)***	(0.051)***	(0.038)***
Treated	-0.910	-0.441	
(Weekly=1)	(0.291)***	(0.325)	
TreatedXlog(hh size)	0.443	0.282	
	(0.155)***	(0.174)	
Head Gender	0.148	0.308	0.219
(male=1)	(0.101)	(0.123)**	(0.083)***
Share of HH: Adult Female	0.659	0.461	0.635
	(0.285)**	(0.285)	(0.217)***
Share of HH: Adult Male	0.288	0.433	0.635
	(0.279)	(0.261)*	(0.217)***
Revisit			-1.032
			(0.306)***
RevisitXlog(hh size)			0.509
			(0.161)***
Phone			-0.407
			(0.328)
PhoneXlog(hh size)			0.258
			(0.174)
Alternative Control			-0.690
			(0.352)*
AltXlog(hh size)			0.435
			(0.192)**
R^2	0.40	0.39	0.38
N	385	377	770

*, **, *** denotes statistical significance at 1, 5, 10% levels respectively

Note: Robust standard errors are reported in the parentheses.

Table 6: Robustness Check – Stepwise Add Control Variables

	Ghana			Tanzania		
	[1]	[2]	[3]	[1]	[2]	[3]
log(HH size)	0.476***	0.497***	0.499***	0.542	0.512	0.525
	(0.122)	(0.123)	(0.120)	(0.142)***	(0.143)***	(0.145)***
log(Acre)	0.597***	0.604***	0.599***	0.349	0.346	0.338
	(0.079)	(0.080)	(0.077)	(0.063)***	(0.063)***	(0.063)***
Male Head	0.115	0.112	0.109	0.171	0.176	0.186
	(0.085)	(0.084)	(0.084)	(0.086)**	(0.085)**	(0.086)**
Share of HH: Adult Female	0.716***	0.742***	0.745***	0.449	0.408	0.417
	(0.187)	(0.190)	(0.189)	(0.222)**	(0.225)*	(0.223)*
Share of HH: Adult Male	0.720***	0.757***	0.831***	0.332	0.276	0.273
	(0.169)	(0.172)	(0.170)	(0.230)	(0.231)	(0.229)
Treated	0.926***	0.948***	0.974***	-0.688	-0.680	-0.670
	(0.219)	(0.218)	(0.219)	(0.277)**	(0.279)**	(0.289)**
TreatedXlog(hh size)	0.091	0.081	0.074	0.349	0.347	0.342
	(0.121)	(0.120)	(0.118)	(0.149)**	(0.150)**	(0.154)**
TreatedXlog(Acre)	-0.225***	-0.228***	-0.231***	0.033	0.034	0.046
	(0.084)	(0.084)	(0.082)	(0.072)	(0.071)	(0.072)
Fixed Effects:						
Village Level	Yes	Yes	Yes	Yes	Yes	Yes
Soil Quality & Type	Yes	Yes	Yes	Yes	Yes	Yes
Maximum HH Education		Yes	Yes		Yes	Yes
Off-Farm Income			Yes			Yes
R^2	0.53	0.53	0.54	0.41	0.41	0.42
N	676	676	676	573	573	573

Table 7: Robustness Check – Stepwise Add Control Variables By Treatment Arm

	Ghana			Tanzania		
	Visit Only	Phone Only	By Treatment	Revist	Phone	Treatments
log(HH size)	0.540***	0.527***	0.565***	0.507	0.574	0.560
	(0.129)	(0.124)	(0.117)	(0.145)***	(0.157)***	(0.138)***
Treated	0.622***	0.691***		-0.942	-0.426	
	(0.214)	(0.199)		(0.308)***	(0.335)	
TreatedXlog(HH size)	-0.042	0.041		0.457	0.260	
	(0.127)	(0.120)		(0.153)***	(0.175)	
log(Acre)	0.482***	0.453***	0.438***	0.394	0.331	0.359
	(0.049)	(0.056)	(0.040)	(0.043)***	(0.046)***	(0.036)***
Male Head	0.152	0.161	0.112	0.171	0.289	0.247
	(0.112)	(0.125)	(0.086)	(0.098)*	(0.123)**	(0.082)***
Share of HH: Adult Female	0.733***	0.607**	0.768***	0.513	0.566	0.613
	(0.259)	(0.239)	(0.187)	(0.266)*	(0.282)**	(0.215)***
Share of HH: Adult Male	0.826***	0.737***	0.855***	0.125	0.456	0.574
	(0.249)	(0.214)	(0.170)			(0.213)***
Visit			0.647***			-1.064
			(0.214)			(0.317)***
VisitXlog(HH size)			-0.056			0.520
			(0.127)			(0.161)***
Phone			0.687***			-0.382
			(0.196)			(0.333)
PhoneXlog(HH size)			0.048			0.235
			(0.119)			(0.175)
Alternative Control						-0.696
						(0.354)**
Alt Cntrl X log(HH size)						0.440
						(0.194)**
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.52	0.55	0.54	0.38	0.36	0.36
N	452	451	676	384	377	769

References

Arthi, V., K. Beegle, J. De Weerdt, and A. Palacios-López (2018, Jan). Not your average job: Measuring farm labor in tanzania. *Journal of Development Economics 130*, 160–172.

Benjamin, D. (1992, Mar). Household composition, labor markets, and labor demand: Testing for separation in agricultural household models. *Econometrica 60*(2), 287.

Dillon, B. and C. B. Barrett (2017, Feb). Agricultural factor markets in sub-saharan africa: An updated view with formal tests for market failure. *Food Policy 67*, 64–77.

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

Gaddis, I., G. Oseni, A. Palacios-Lopez, and J. Pieters (2021, Oct). Measuring farm labor: Survey experimental evidence from ghana. *The World Bank Economic Review 35*(3), 604–634.

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