Are There Too Many Farms in the World?
Labor-Market Transaction Costs, Machine Capacities and Optimal Farm Size

Andrew D. Foster
Brown University and NBER

Mark R. Rosenzweig
Yale University and NBER

September 2017

This paper revisits the issue of the relationship between operation scale and productivity in agriculture. The research is motivated by three stylized facts characterizing world agriculture. First, farming in low-income countries is small scale while farming in developed countries is large scale. Figure 1 displays the proportions of operational holdings of farms that are below 10 acres across a sample of developed and developing countries for which reliable data are available on the size distribution of farms. As can be seen, only 10% or less of farms are below 10 acres in the United States and Canada, while for the three most populous low-income countries - China, India, and Indonesia - at least 80% of farms are below 10 acres. In major economies in Africa too, as seen in the figure, only a small proportion of farms are above 10 acres. Note that these figures may underestimate the extent of small-scale farming is in such countries to the extent that the landholdings of a farm are fragmented into spatially-separated plots.

The second stylized fact is that the productivity of developed-country agriculture is substantially higher than it is in low-income countries. For example, as shown in Figure 2, soybean yields are four time higher in the United States, where farm scale is high, than they are in Indonesia, India and the Philippines, where farms are small, and three times higher in Canada. This figure also illustrates, however, why output per acre is insufficient to gage productivity - China appears to be an outlier in that its yields are twice as high as those in the other three low-income countries in the figure, despite its similarity in operational scale. However, this is misleading, as the fertilizer-intensity in China, as seen in Figure 3, is 2.7 to 3.5 times higher than that in Indonesia, India, and the Philippines and 5 times higher than that in the United States.[1] Assessing productivity requires attention to input use and its cost.

An implication of any positive causal relationship between production scale and agricultural productivity implied by the differences in scale and productivity across countries is that there are too many farms in the world, especially in low-income countries. It implies that enlarging the size of farms via consolidation would increase overall agricultural output, with an accompanying substantial reduction in the amount of employment in agriculture. Another implication is that the large amount of labor devoted to farming observed in low-income countries should not solely be interpreted as a symptom of underdevelopment and poverty but also in part a cause. Understanding the role of scale economies in agriculture thus has major implications for not only the global supply of food but also

[1]The evidence suggests that fertilizer is over-used in China, with the marginal return on a dollar of fertilizer less than 7 cents (Huang *et al.*, 2008). The reasons for the high fertilizer intensity in China are the large fertilizer subsidies and governmental resource assessments based on crop yields rather than net returns.

for how best to improve overall per-capita incomes in countries where incomes are especially low.

Of course, differences between developed-country and developing-country agriculture are due to more than scale. The best evidence on scale relevant for a low-income country would come from a single country, based on farms in the same institutional environment, the same markets and facing the same technology frontier. When farm scale and farm productivity are examined within a country, however, we get the third stylized fact: there is an almost universal inverse relationship between farm or plot size and productivity within developing countries, over the span of plot and farm sizes observed in those countries, while increasing returns to scale are observed among the larger farms in developed countries (e.g. Paul *et al.*, 2004).

Most of the literature documenting the inverse relationship in low-income countries is based on data from Asia and Latin America (e.g, Schultz, 1964; Hayami and Otsuka, 1993; Binswanger *et al.*, 1995; Hazell, 2011; Vollrath, 2007; Kagin *et al.*, 2015). Recent studies of newly-available representative survey data from a number of African countries, used to describe the African landholding distributions in Figure 1, confirm that this inverse relationship also exists there (Larson *et al.*, 2013; Carletto *et al.*, 2013). Figure 4 provides a typical pattern, based on one of these data sets, the Nigeria General Household Panel of 2015-16, which describes a representative sample of rural households for whom there are accurate (GPS) measures of farm size. Based on per-acre yields, the very smallest farms are substantially more productive. And, as noted, the span of farm sizes is quite limited. The existing descriptive evidence on scale and farm productivity from data describing farming in low-income countries thus does not support the notion there are too many farms.

There is a large literature focused on low-income countries that has also attempted to address the puzzle of why the smallest farms are most productive, without little consensus. There is general agreement that the inverse relationship is not spurious - specifically, not due to a correlation of land quality and farm size (e.g., Carter, 1984; Bhalla and Roy, 1988; Benjamin, 1995; Barrett *et al.*, 2010) and/or measurement error that is correlated with scale (e.g., Ali and Deininger, 2014; Larson *et al.*, 2013; Carletto *et al.*, 2013). However, a general shortcoming of this literature is that it may be addressing the wrong puzzle. Given the global pattern of farm productivity, the puzzle that requires explanation is why there is a U-shape relationship between farm productivity and scale - why the smallest farms, which dominate low-income countries, are more productive than somewhat less small farms there and why in the developed world the large-scale farms are not only more productive on average, but productivity increases with scale.

Seen from this global perspective, some of explanations for the inverse relationship observed in low-income countries are at best incomplete. For example, the idea that farms exclusively managed and worked on by owner-operators and their families, which characterizes the smallest farms, have an advantage, because of superior incentives, lower supervision costs, and lower unit-labor costs (Yotopoulos and Lau, 1973; Carter and Wiebe, 1990; Binswanger-Mkhize, *et al.*, 2009; Hazel *et al.*, 2010;) while true, cannot explain why corporate farms, which are large scale, are even more productive. A common finding in this literature is that the smallest farms use labor more intensively than larger farms, which would generate higher per-acre output but not necessarily higher productivity accounting for input costs, but the reasons for this are not settled.

One other salient difference between low-income and high-income country agriculture is the degree to which mechanized implements are used. And there is evidence that larger farms and farms that become larger are more likely to be mechanized within countries (Zaibet and Dunn, 1998; Foster and Rosenzweig, 2011; Hornbeck and Naidu, 2014). However, in contexts in which all farms are mechanized, such as in developed countries, the mere use of mechanized equipment cannot by itself explain why larger farms are more productive than smaller farms.[2]

This paper seeks to explain then the U-shaped relationship between farm productivity and farm scale - both the initial fall in productivity as farm size increases from its lowest levels and the continuous upward trajectory as scale increases after a threshold. The explanation focuses on two factors: transaction costs in the labor market and economies of scale in machinery capacity. Transaction costs in the labor market are especially important in agriculture because agricultural operations are sequential and intermittent - labor is thus principally hired on a daily basis, with employers and workers seeking matches at high frequency. Moreover, the amount of work needed on a given day may vary, so there is daily variation in worker hours. We show that the existence of fixed transaction costs, to the extent they are born by farmers, makes farmers at the margin at which hiring labor would be productive on net (all family labor fully utilized) reluctant to hire labor. And, if labor is hired at all, average unit labor costs will vary by operational scale because larger scale entails more intensive use of labor. The result is a U-shape in which the smallest farms are most efficient in their use of labor, slightly larger farms least efficient and larger farms as efficient as the smallest

[2]Some studies have suggested that access to capital and a greater ability to insure against risk may explain why larger farms may be more productive than smaller farms. However, we show that the U-shaped relationship between scale and productivity holds across plots for the same farmer, which effectively holds constant the farmer's ability to take risk, finance capital and make better allocative decisions. We thus abstract from these considerations, but this does not imply they are not important determinants of agricultural productivity in low-income countries.

farms because the share of transaction costs in total labor expenditures are smallest.[34]

The existence of fixed labor market transaction cost can explain a U-shape, but it cannot alone account for the higher productivity of larger farms compared to the smallest farms, positive scale economies that continue at higher scales. For this, we focus on scale economies in machinery capacity. There is ample evidence that agricultural machinery saves on labor costs (Hornbeck and Naidu, 2014; Davis, 2016) and that mechanization is more likely on larger farms (Foster and Rosenzweig, 2011). If there is a minimum farm scale at which mechanized equipment can be used, then only larger farms can exploit this substitution to avoid the additional costs of hiring labor.[5] But, again, a single threshold cannot explain the continuing rise in productivity with scale. We show that to explain the upper tail of the U requires there be economies of scale in the capacity of machines, their ability to accomplish tasks at lower costs at greater operational scales. There are two conditions that must be met: effective machine capacity can only be increased at larger scale and the pricing of capacity must be non-linear. We address the question of whether these conditions are met within a low-income country.

We are able to examine the role of transactions costs and machine capacity scale economies as major factors accounting for the U-shape relationship between scale and productivity within a low-income country because of the existence of unique data from the India ICRISAT VLS panel survey. One key advantage of the ICRISAT survey is that the sampling scheme differs from almost all household surveys, which almost always seek to achieve household representativeness. The ICRISAT survey sampling frame is instead based on landholding size. As a consequence of this sampling frame, larger farms are over-sampled and we are able to examine both small and larger

---

[3] Allen (1988) shows that one of the reasons that larger farms were more productive than smaller farms in 18th century England, when mechanization was not a major factor, was that larger farms could hire labor crews. Hiring a worker team saves on hiring costs compared with hiring workers individually, but is only cost-effective for larger-scale operations.

[4] Foster and Rosenzweig (2011) highlight the additional supervision costs associated with using hired labor as an explanation for labor under-utilization at farm scales above the smallest. But supervision costs cannot explain why above a threshold larger farms become more efficient in the absence of labor-substituting mechanization since it is not likely such costs diminish as the amount of hired work increases.

[5] The idea that there are physical constraints associated with the size of plots inhibiting mechanization is well known. Bivar (2010), in her study of the French government- and union-led agricultural consolidation program initiated in the early 1950's - motivated by the potential productivity-enhancing effects of mechanization - cites documents written by the French Agriculture ministry that suggest for a tractor to be able to turn around, a minimum plot size of 1.5-2 hectares is required. Of course, there are more mechanized options today that require a smaller minimum scale, but these may have reduced performance, which is one of the key hypotheses we test here.

farms in a common environment. The data set thus contains the missing link between low-income country agriculture and developed-country agriculture - because of the over-sampling the sample of farms exhibits the U-shape that characterizes the global relationship between agricultural productivity and scale. Representative household surveys in low-income countries do not contain any large farms. As indicated in Figure 1, there are few farms even above 10 acres in such environments. The U-shape is simply not visible in low-income country rural data sets because of survey design.

To our knowledge, there are only two prior studies based on low-income country farm data that finds evidence of a U-shape. Kimhi (2006), using data on maize producers in Zambia, shows that dis-economies of scale characterize farms below 7.4 acres, which account for 84% of all farms, but that productivity rises with scale above that threshold. Muyanga and Jayne (2016), recognizing the representativeness sampling problem in existing data sets, obtain data from a dedicated survey of medium-sized farms and a representative sample of small farmers in Kenya that reside in the same villages. They also find the inverse relationship in the representative sample containing mostly small farms, but positive scale economies for the larger farms (25-124 acres), measuring productivity both in per-acre net returns, taking into account all input costs, and per-acre output. However, neither of these studies provides evidence on the mechanisms behind the U-shape.

There are measurement issues in most existing data sets as well that have made it difficult to identify the mechanisms that underlies scale economies that we focus on here. In many if not all low-income country data sets agricultural labor time is measured in days rather than hours. While time wages are generally paid on a daily basis for most agricultural operations, the true unit cost of labor time will be masked if there is variation in per day hours. The ICRISAT data record labor time use in hours and days. The data indicate that not only is there substantial variation in the average hours per day workers provide, but the amount of daily hours within an agricultural operation differs by operation scale.

Based both on the wage schedules provided by farm employers and the daily wages and hours reported by farm workers, we show that the average hourly wage decreases with the number of hours worked, consistent with the existence of a daily fixed cost of employment. We also are able to document that smaller farms (but not the smallest) on average employ more low-hour hired labor across all of their operations than do larger farms. We show that as a consequence, the average hourly wage, inclusive of the imputed cost of family labor, increases and then decreases with farm scale. Consistent with this, we also find that for the same plot across time, when the amount of work increases due to more rainfall, on a smaller plot a higher average wage is paid while on a larger plot average wages are lowered.

Another major deficiency of existing data describing farming in low-income countries even where mechanized equipment is used is that there is little or no information on the capacity of farm equipment. The best surveys provide a detailed inventory of owned equipment by type (thresher, tractor, sprayer) and value, but little or no information on power or capability (e.g., horsepower, bushels processed per time unit). Thus, any empirical evidence on scale economies obtained from such data must assume that within machine types machinery capability is homogeneous - an eight-row harvester and a four-row harvester are not distinguished, even though their capabilities and suitabilities to different production scales are likely quite different. Most data sets also do not provide information on the time use and the rental price of equipment, by type or capacity. Thus it has not been possible to measure scale economies in farming due to economies of scale in machine capacity that could underlie the positively-sloped upper segment of the U.

The ICRISAT data too do not provide direct information on the power or capacities of the equipment that is used by the farmers. Tractors, for example, are not distinguished by horsepower or speed or towing ability. However, we show how it is possible to identify the varying capacities of one major type of equipment - sprayers - using the information provided on the amount of material sprayed and the time use of sprayers. This enables us to estimate an effective capacity function relating capacity- material sprayed per hour - to scale and to estimate the capacity pricing schedule.[6] We find that, consistent with sprayer capacity scale economies, larger farms do less spraying per acre and use higher-capacity and more expensive sprayers and we estimate that the implicit rental price of capacity declines as capacity increases. Based on our structural estimates we are able to identify the "optimal" scale of operation based on the sprayer scale economies - the scale at which additional increases in scale would lower productivity, but below which operational scale is too low and thus excessively labor-intensive, at least with respect to the control of weeds and insects.

In section 2, we describe the data and show that profits per acre exhibit a U-shape with respect to both farm size and plot scale that is not due to relationships between size and soil quality or size and measurement error or to differing characteristics (ability, liquidity) of farmers. We also show that the average daily work hours of a worker varies substantially and that the hour-based wage

[6]Our measure of sprayer capacity is identical to that employed by sellers of sprayers. We are thus able to compare our estimated capacity pricing schedule to those provided by sprayer vendors in the India and the United States.

schedules are consistent with the existence of labor-market transaction costs. We also provide statistics on worker turnover and one component of fixed costs, worker distance to workplace that support the assumption of the empirical importance of fixed costs. In section 3, we set out the model, starting with a labor-only model in a market with transaction costs and then allowing the possibility of using machinery that is heterogeneous in capacity and substitutable for labor. We show that both models can replicate the U-shape, but with capacity scale economies the farm scale economies persist beyond a threshold land size depending on the shapes of the effective capacity and capacity pricing functions.

Section 4 provides evidence that the marginal return to profits is U-shaped in farm or plot scale, and describes evidence on the non-linear relationships between scale and the use of low-hour hired inputs and average hourly labor and bullock prices, and on the non-linear effects of rainfall on average input costs by plot size. All of these findings are consistent with labor-market transaction costs playing an important role in the decline in productivity with scale at low scale. In section 5, we examine equipment use and productivity, focusing on power sprayers. We show that descriptively, consistent with capacity scale economies, per-acre sprayer hours and weeding hours decline and the rental price of sprayer hours and sprayer capacity rises with farm scale after a threshold size. Our structural estimates of the capacity price schedule for sprayers, based on the model, rejects the hypothesis that capacity pricing is linear in capacity and thus is indicative of capacity scale economies in India. Indeed, we find that the price parameter we estimate from the ICRISAT data is similar to that characterizing the price schedule for power sprayers that are available throughout India. However, both estimates imply a steeper slope for capacity pricing than we see in US pricing schedules for power sprayers. The estimates suggest that a reduction in the sprayer price slope parameter from that faced by ICRISAT farmers to that associated with US power sprayers, for a farm at the median in the ICRISAT villages, would increase the capacity of the sprayer chosen by 5%, increase sprayer hours by 23.3% and reduce weeding hours by 3%.

Finally, our model-based estimates imply that effective capacity for power sprayers reaches its maximum at a farm size of 25 acres. The fact that this limit to scale economies is at the extreme right tail of the distribution of farm sizes in the ICRISAT setting (based on census landholding data) does not imply that the optimal farm size in India is limited to 25 acres (even if all mechanized inputs were characterized by the same parameters). Rather it is also consistent with a local equilibrium trap. If the market availability of scale-dependent technologies depends on the existing scale of operation, the largest farms have no incentive to increase scale because there are no readily available machines with effective capacity beyond their own operation size. The puzzle remains, however, why if scale economies are positive up to 25 acres we do not see a move towards the consolidation of land. Our findings suggest that this is an important topic of future research.[7]

1. The Data

a. *Sampling and information content*

We begin by describing the data we use. We do this for two reasons. The first is that we need to show the phenomenon that requires explanation, namely a non-constant relationship between farm productivity and scale that is not simply due to measurement error or omitted land quality. The second reason is that the data provide unique information on plot and farm location, prices and input characteristics, which motivate (and permit) the new explorations of the underlying causes of non-linear agricultural scale economies.

Our principal data source is the six latest rounds of the India ICRISAT VLS panel survey, covering the agricultural years 2009-2014. The survey has two components - a census of all households in 18 villages in five states - Andhra Pradesh, Gujurat, Karnataka, Maharasthra, and Madhya Pradesh - and a panel survey of the households in those villages, which includes 819 farmers. A key advantage of the ICRISAT survey is that the sampling differs from almost all household surveys, which seek to achieve household representativeness, because the sampling frame is based on landholding size. In particular, the survey contains in equal numbers landless households, small-farm households, medium-farm households, and large farm households. As a consequence of this sampling frame, we are able to examine both small and larger farms in a common environment, unlike in most surveys of farm households in countries with similar landholding distributions, in which most households own small plots.

The ICRISAT data are unique in other ways that are critical for identifying the underlying mechanisms of scale economies. First, there is information on input quantities and prices by type of input, by farm operation and by individual plot collected approximately every three weeks.[8] The

[7] Note that the persistence and ubiquity of small farms in low-income countries does not necessarily imply that our findings on scale economies must be incorrect. The output gains and increases in mechanization from the mandatory post-war governmental land consolidation scheme in France, described by Bivar (2010), from an initial distribution of land characterized by very small scales of operation that had persisted for centuries, suggests instead a market failure. Discovering the source of the land market failure, and remedies, thus may have high payoffs.

[8] The size of the basic unit of operation, the plot, is not a choice variable - the size of a given named plot does not vary from year to year. Similarly, farm sizes are stable. There is little change in the number of plots owned by a farmer over the full span of the panel, from 2009-2014 - only 5.8% of plots were bought or

high-frequency input information is thus likely to be more accurate than that found in almost all surveys, which collect information once or at best twice in an agricultural season. Second, there is information on market input prices for workers, machinery, and animal traction collected at the village level, in addition to that elicited from the households survey, by work time. Third, importantly for identifying the role of mechanization in scale economies, there is information enabling the measurement of the power and capacities of machines. Fourth, there is information on how plots were acquired, e.g., inheritance or purchases, including information on the dates of inheritance, along with information on all other assets of the household.

b. *Descriptive information on scale and farm productivity*.

Figure 5 displays from the village Census and from the surveyed households in 2014 the cumulative distribution of farms by total owned (agricultural-use) landholdings along with the sample-household distribution of plot sizes. The figure shows that the full population (census) land distribution is similar to that of most low-income countries - 92% of land-owning households have less than 10 acres. Because of the sampling scheme, however, we observe detailed information on farms above 10 acres in the household sample - in contrast to the population distribution, households with more than 10 acres of landholdings constitute almost 40% of the sample.

The oversampling of larger farms is key to understanding the global relationship between farm productivity and farm scale. The sampling scheme provides the missing link between developed-country large-scale farming and low-income country small-farm agriculture within the context of a single low-income country. This is because we are able to observe both the decline in profitability by scale, characteristic of low-income countries, and its rise with scale, characteristic of developed countries, in the same setting with comparable data across farms. Figure 6 displays the lowess-smoothed relationship between average real (1999 rupees) profits per acre in the main growing season (*Kharif*) and owned total landholdings for the full data set (2009-2014). As can be seen, as in most low-income countries, there is a monotonic decline in per-acre profitability with acreage below 10 acres. But then there is a monotonic increase, as is observed in developed countries.

Using the detailed information of the data set, we can rule out two reasons for the U-shape

sold, and the main reasons for any land turnover were inheritance or family transfer. Almost all plots therefore are inherited (0.74% of all plot observations involved a purchase of land). The 2014 Census data indicate the leasing market is only somewhat thicker than the land sales market, with 8.4% of landowners leasing out and 11.5% leasing in land.

that have been suggested in the literature, which has focused on the decline in productivity with scale below 10 acres: (a) measurement error in land size that is correlated with size and (b) land quality that is correlated with land size. We can also rule out credit constraints and farmer ability as the sole determinants of the U-shape relationship. With respect to measurement, we exploit the fact that we have two independent, and different, elicitations of farm size in 2009. Each household was visited and administered a Census questionnaire asking for the total number of acres of owned unirrigated and owned irrigated land. At a separate time in the same crop year, the panel households were asked to provide the acreage of each individual plot they owned. Thus, there are two sets of information on acreage for each panel household collected at different times. We summed the individual plot acreages to obtain total farm size from the survey and summed the reported total acreages of irrigated and unirrigated land to obtain total farm size from the Census elicitation. We then used the latter as an instrument for the former to estimate the effect of farm size on profits.

Table 1 reports the OLS and IV profit estimates. The specifications also include dummy variables for each village, capturing the influences of weather and prices at the village level. As can be seen, the IV and OLS estimates are economically identical. However, the fundamental issue is not how much measurement error there is in acreage reports, but whether measurement error leads to bias that is correlated with farm size. For example, even if measurement error is on average small, it may be proportionately large for small plot sizes, which would create a larger negative bias in the relationship between size and per-acre profitability at small compared with big farms. To assess this possibility, we obtained the coefficient of land size by land size, using the locally-weighted functional coefficient model (LWFCM),[9] obtained again using OLS and IV. Figure 7 displays the two land coefficients by land size, indicating that measurement error is small at all acreages. Measurement error does appear to increase with farm size above ten acres, suggesting that the upward component of the U-shape for profits per acre displayed in Figure 6 may actually be slightly understated due to measurement error. Measurement error, however, is clearly not the major cause of the downward slope in profitability below 10 acres.

The U-shape also survives control for land quality. The survey provides, at the plot level, multiple measures of plot characteristics. These include soil depth, distance of the plot from the household residence, four categories of soil fertility, six levels of soil degradation, and 11 soil types. Table 2 reports the relationship between real output and real *Kharif* profits and farm size for the full

[9] See Cai *et al.*, 2006. The specification we use is locally linear in profits and farm size.

sample. In the first column for each dependent variable, the specification only includes village/year fixed effects. The second column estimate for each is from a specification additionally including all the plot characteristics. The difference in estimates is negligible, despite the fact that the set of 24 variables associated with the plot characteristics are highly statistically significant for both profits and gross output value.[10]

To assess whether credit constraints associated with farmer characteristics, such as total wealth or ability, are responsible for the U-shape in the third and fourth columns for each dependent variable we use information on output and profits at the plot level. The third column for each reports the within-farmer-year estimate of plot size; the fourth column reports the within-farmer-year estimate of the plot size effect including the soil characteristics of the plots. Again, the estimates are virtually identical. The within-farmer-year plot size estimates are lower than the cross-farmer farm size estimates, suggesting that farmer ability or wealth may play some role in the relationship between farm size and productivity. [11]

The key issue is whether variation in soil characteristics and farmer characteristics by acreage are solely responsible for the U-shape. Figure 8 displays three plots: the relationship between per-acre real profits and farm size repeated from Figure 6; that relationship estimated using LWFCM from a specification including all of the soil characteristics in which the coefficients for farm size and the soil characteristics can vary non-parametrically with farm size; and the LWFCM-estimated relationship obtained solely from cross-plot variation within a farm and including the soil characteristics. All three plots display the U-shape, with the LWFCM-estimated curves obtained at mean soil and plot characteristics. Controls for soil quality evidently lower the observed profitability of the smallest farms (soil quality and size are negatively related among small farms) but have little effect on the upward slope. The within-farm estimates suggest that the U-shape also characterizes plots within farmers.[12] Thus, neither variation in farmer wealth or farmer ability or heterogeneity in

[10]Below, we will provide plot fixed-effects estimates of operational scale effects by plot size that control completely for land quality.

[11]The ICRISAT survey farms are not especially fragmented. 43% have only one plot, and 74% have two or less plots, with larger farm having more plots (correlation = 0.6). The correlation between average plot size and total farm size is 0.7. We show below that the size of plots as well as farm size matter in determining scale economies.

[12]Interestingly, Assunção and Braido (2007) found, controlling for farmer fixed effects, an inverse relationship between output per acre and plot size but no uptick in productivity for larger plots using the initial ICRISAT data from the period 1975-1984. In that period the availability of mechanized implements was substantially lower than in the period covered by the latest rounds of the survey, 25 years later.

plot or soil characteristics explains the U-shape association between per-acre profitability and scale.[13]

c. *Fixed costs of labor hiring.*

In the next section we will set out a model to explain the U-shaped pattern of farm and plot efficiency. We will focus on transaction costs in the labor market and scale economies in machinery capacity. We do this because these aspects of the farming environment are evident in the data. With respect to identifying a fixed cost component to hired labor, we use information from the price schedules for hired labor and labor plus bullock pairs by farm operation. Information on daily wages paid was obtained monthly from one informant from each of the three classes of farmers in the 2010 and 2011 rounds of the survey by farm operation according to the number of hours worked in the day. The reports are obtained at the beginning of each month over the full year.

The first salient feature of the data from the wage schedules is that a large fraction of workers paid daily wages work less than eight hours in a day. That is, many workers are hired for less than a full day. In the wage schedule reports, 31% of the daily wage reports for hired males were for workers who worked less than eight hours; for bullock pairs and driver, over 58% of daily wages paid were for work that was less than eight hours. This is in accord with the survey data on off-farm employment reported by respondents. In the 2014 round, for example, 30.6% of respondents working off farm for wages in agriculture operations during the peak *Kharif* season reported that their average working hours were less than eight (29.8% reported that average hours were six hours or less per day).

We computed hourly wages based on the monthly wage schedules and then regressed the log of the hourly wage for the two categories of hired inputs on whether or not the work done was for the full eight hours, with a full set of dummy variables for farm operation. Any hourly wage difference by daily hours hired could be due to low-wage operations occurring in slack periods with little work. The operation fixed effects ensure that this is not the case. The within-operation log wage estimates are reported in Table 3, where it can be seen that while the daily wage is higher the more hours a worker works in a day, farmers pay an hourly premium for low-hour work - workers who work eight hours are paid a statistically-significant 33% less per hour than lower-hour workers;

[13]The U-shape is also not due solely due to different crop choices by plot size. Using only plots devoted to cotton, one also observes the U-shape pattern of per-acre profits and plot size, as displayed in Appendix Figure A. Cotton is the second largest cash crop in the ICRISAT sample, with 17% of all plots devoted to cotton. 20% of plots are devoted to soybeans, but soybeans are not grown on the very small plots that dominate the sample, so it is not possible to identify a U-shape relationship with that crop, but per-acre profitability on soybean plots rises with plot size above 10 acres.

a hired bullock pair and driver working eight hours is paid over 20% less per hour than his part-time counterpart. The hourly premiums paid to low-hour workers is consistent with the existence of fixed transaction costs for hiring workers compensated in part by employers. We will use the model to infer how the existence of these fixed costs affect labor-intensity and profitability by land size.

Is it plausible there are significant fixed cost components in labor costs, as implied by the labor price schedules? First, most agricultural laborers are hired on a daily basis. There are few formal longer-term contracts, so each worker must be matched with a farmer who is seeking workers for a given day's task. Second, farms are also spatially separated from where workers and farmers in the village reside, so travel costs are not trivial.[14] The ICRISAT data provide the distance of each plot from the farmers' home (in the village center). The median distance is one kilometer.

The distance of plots to residences in the sample understates the average distance a worker must travel to get to an employer because a significant proportion of workers residing in a village work for a farmer located outside the village. The Yale EGC-CMF Tamil Nadu Panel Survey contains a representative sample of rural households in 200 villages in the Indian state of Tamil Nadu in 2011. In this sample, 21.3% of farmers located in the villages who employed any agricultural laborers reported hiring laborers from outside the village. Consistent with this, 23.6% of the survey respondents who worked for wages in agriculture reported working for a farmer located outside the village. Among those traveling to a farm outside the village by foot or bicycle (63.8%), the average distance to the non-village farm was two kilometers. The median distance to a non-village farm for those traveling by bus (26.5%) was 8 kilometers. Finally, the median number of individual farmers that an agricultural laborer worked for in total during the main growing season (*kharif*) was seven. If at least some of these turnover/search and travel costs are born by farmers, this will be manifested in hourly wage schedules that resemble those we see in the data.

d. *Non-linear pricing of machinery capacity.*

A second key component of our model is the existence of non-linear pricing in the cost schedules for farm equipment by machine capacity. The ICRISAT data, like most data sets, has only an incomplete description of farm equipment. The size or horsepower of tractors owned or used, for example, are not provided. The inventory of owned equipment in the 2011 round does provide, however, information on the price and horsepower of electric motors and submersible pumps. Figure 9 displays the relationship from the inventory data between the cost per horsepower and total

[14] Typically, farmers and workers reside in a village center with farm plots surrounding the residential area.

horsepower for both sets of equipment. The figures clearly exhibit scale economies in pricing by machine power. The information on pricing, is insufficient, however, to demonstrate the importance of scale economies in equipment as an underlying mechanisms for farming scale economies, as we will show, because what is also required is the technical relationship between horsepower (in this case) and the ability of the equipment to accomplish agricultural tasks. We will demonstrate that there is sufficient information on sprayers, a key piece of farm equipment, from the input data files to pin down the relationships between machine power, machine capacity and price that is necessary for identifying the link between equipment and farming scale economies.

2. Model

a. *Labor-only model with market transaction costs.*

We develop a model with the goal of understanding the mechanisms that underlie the U-shaped profitability and efficiency variation by farm and plot size. We focus on transaction costs in the labor market and scale economies associated with farm machinery. We thus exclude consideration of constraints on input use that arise from imperfections in credit or insurance markets or relationships between farm size and farmer competence. Such constraints cannot explain variation in the profitability among plots for a given farmer, which we have seen mimic the profitability patterns observed across farmers sorted by total landholdings.[15]

Although agricultural production takes place in stages, to fix ideas we initially focus on a one-stage agricultural production function. We assume that agricultural production is described by a constant returns to scale production function g that consists of two inputs: land (a) and plant nutrients (e). The amount of nutrients applied is itself described by a production process. For example, the application of fertilizer requires labor time. Removing weeds, which reduces competition for nutrients by the plants, can be accomplished using labor for pulling weeds, and/or by spraying, using labor and a sprayer. We initially assume that the process of nutrient production uses only labor, focusing on transaction costs in the labor market. We think that this model is by itself applicable to settings such as in sub-Saharan Africa where landholdings are especially small and mechanization is thus infeasible. We will subsequently generalize the model to include a (heterogeneous) mechanized input that is substitutable with labor, as is the case in settings where

[15] In an appendix we explore directly the role of wealth differences and farmer ability in explaining the U-shape and show they do not account for the U-shape.

farm scale is above some threshold.[16]

The farm is endowed with total family labor time of l and can use family l_f or hired l_h labor time to produce nutrients e. Thus total output is

$$g(a,e) \tag{1}$$

To highlight the role of labor market transaction costs we assume that family and hired labor time are perfect substitutes, and the nutrient production function is e = $l_f + l_h$, which is also CRS. Workers entering the labor market for off-farm work in a given day face a fixed entry cost per day as a result of transaction costs and/or travel (in effect we define a production stage as work done on a particular day). This labor-market entry cost, f, may vary across households. As a consequence, in equilibrium, farmers wishing to employ workers for just a few hours must at least partly compensate these workers for this fixed cost. We model this compensation as having a fixed and variable component, consistent with observed wage schedules, so that the cost of hiring a worker for l_{h1} hours is

$$w(l_h) = \mathbb{I}\left(l_h > 0\right) w_0 + w_1 l_h\ . \tag{2}$$

With $w_0 = f$ workers are fully compensated for the fixed costs of off-farm work.

Profits are defined as

$$\pi(a,l_h,l_f) = ag(l_h + l_f) - w(l_h) - w_1 l_f \tag{3}$$

Costing out hired labor for the purpose of computing profits empirically is straightforward because total payments to hired labor $w(l_h)$ are readily observed. Here, we cost out family labor using only the variable component of wages. This is correct under two conditions, as we discuss below: (a) if the family is engaged in the external labor market regardless of the on farm labor supply or (b) if transactions costs are fully compensated by the labor market.[17]

We assume that the farmer, with fixed endowments of family labor time and land, maximizes

[16] Our model can identify that threshold, given estimates of the farm production technology, price schedules and available machine capacities.

[17] In principle one might also want to include the uncompensated component of the fixed cost of entry (e.g., $f - w_0$) in the profits calculation; however, such an approach would be problematic when comparing profits in the model to those in the data as, in contrast to hired labor costs, this cost cannot be directly observed. In practice, the model with fully compensated fixed costs seems to give the best match between the theory and data.

profits plus labor income minus any fixed costs of entry into the labor market for any given land area a

(4) $$\mathfrak{L} = \pi(a, l_h, l_f) + \mathbb{I}(l_o > 0)(w_0 - f) + w_1 l_o$$

subject to the constraint

(5) $$l_o + l_h = l$$

where l_o is off-farm work time.

The optimal labor allocations, and thus profitability, by area will depend importantly on entry costs. In general there will be three regimes, depending on land size and the magnitude of labor market entry cost.[18] There are two critical values of a that divide land size into regimes. For land areas a>0 with $a < a^*$ farmers will supply off-farm labor. For land areas $a^* < a < a^{**}$, farmers will operate in autarchy, and for land areas $a > a^{**}$ farmers will hire workers. In the first regime, at the lowest land sizes, family members work both on farm and off farm, as long as income working both on and off-farm exceeds the income from on-farm work only. The upper-bound critical value a* is where farmers do not hire workers and are just indifferent between entering the labor market and not.

(6) $$g(a^*, l) = g(a^*, l_f^{\,*}) + (w_0 - f) + w_1(l - l_f^{\,*}) \ ,$$

In that regime and at the regime upper bound for a, the marginal value product of family labor is equal to w_1:

(7) $$f_l(a^*, l_f^{\,*}) = w_1 \ .$$

The second regime is where farm size is sufficiently large so that the profitability of employing all family labor on farm exceeds that from employing any family labor off farm but no hired labor is employed on farm. This regime is not infinitesimal because of the existence of the labor market entry cost that must be paid by the farmer when hiring labor. This will make farmers reluctant to hire workers until land area reaches some threshold. Thus, in this regime, starting at threshold land size a*, given the fixity of family labor, the marginal product of labor declines as land

[18] A fourth regime in which farmers are both working off farm and hiring in workers is not feasible as long as $f > w_0$. The farmer will replace hired work with additional family on-farm work that has the same opportunity cost at the margin until either hired work is zero and the fixed cost of hiring is saved or off-farm work is zero and the fixed cost of off-farm work is saved.

size increases and hence profitability per acre also falls. The opportunity cost of (family) labor remains at the variable component of the wage w_1. This will continue until land size = a**, where farmers are just indifferent between hiring workers, and defraying worker entry costs, and working only with family labor

$$f(a^{**}, l) = f(a^{**}, l + l_h^{**}) - w_0 - w_1^{**} l_h \ . \tag{8}$$

The threshold land size at which any hired workers are employed is higher the larger is the transaction cost w_0.

At a** workers are hired and the marginal product of labor rises to the market wage

$$f_l(a^{**}, l + l_h^{**}) = w_1 \ . \tag{9}$$

However, average labor costs rise at a** because of the necessity of paying transaction costs, which are a large component of labor costs when hired workers are employed at low hours. Then, as land area increases above a**, average labor costs fall, as the fixed component becomes a smaller share of labor costs, and profitability per acre rises ultimately reaching that for the smallest-acreage farms.

To show that the model is capable of replicating the U-shape between profitability per-acre and farm scale, we simulate the model, assuming a Cobb-Douglas production function with a land share (a) equal to a half and with w_0=2, w_1=1/2, l=2 and f=2. Figure 10 displays a U-shaped pattern between per-acre profitability and farm size. The three regimes are evident : on the smallest farms, where family members are working both on and off farm, changes in acreage have no effect on profits per acre. At 2.5 acres, farms become autarchic with respect to labor, and profitability per acre declines as land size increases because family labor time is fixed, thus lowering productivity. At 11.8 acres in the simulation the farm begins to hire workers so average profitability starts to rise and continues to rise up to the initial value observed for the smallest farms as transaction costs become an infinitesimal component of total costs.

Figure 11 plots the pattern of labor costs per acre by farm size from simulations of the same model. As for per-acre profits in Figure 6, labor per acreage is initially flat with respect to acres because the marginal return of on-farm labor is fixed by the marginal return to labor given that these farmers are all working in the market. As acreage rises, however, they eventually leave the off-farm labor market and apply their labor to their own farm. Because of the fixed costs of hiring workers, they do not immediately add workers via the labor market. Consequently, farm labor stays constant as acreage rises and total labor costs per acre falls. At some point profitability falls so low that a farmer is willing to take on hired workers and so labor expenditures then rise discontinuously. They then fall as acreage rises and the fixed cost is distributed over more hours of hired work.

Finally, Figure 12 traces out the relationship between the average price of labor and farm size, showing that larger farms, due to the compensation of hired labor for transaction costs, pay higher average prices for labor than do the smallest farms. We will test for these relationships below.

One limitation of the above model is that production takes place in a single stage. With multiple stages, the fixed costs of hiring workers (and other inputs) will accrue in each stage. To capture this idea we augment the production function to two stages

(10) $$g(a, l_{f1} + l_{h1}, l_{f2} + l_{h2}) \,.$$

The resulting average profits (red) and marginal profits (blue) with respect to acreage are plotted in Figure 13.[19] The marginal profit line now rises in two steps. In the first step the farmer transitions from autarchy in both stages to hiring in the more intensive stage and autarchy in the less intensive stage. In the second step the farmer transitions to hiring in both stages. As should be evident, as the number of stages increases this curve will rise smoothly with acreage. Note further that the average profitability now falls and rises smoothly with acreage.

b. *Adding heterogeneous machinery*.

Labor market transaction costs evidently are sufficient to explain why profitability per acre initially falls as scale increases, with profitability then rising, as in our data, if any labor is hired. It can explain why small farms tend to be autarchic and why in African countries that have very low-scale farms and infrequently employ hired labor, per-acre profitability declines with farm scale and never rises. But the existence of such labor-market costs cannot explain why above some threshold farm size average profitability continues to rise above the initial per-acre profitability of the smallest farms, which is what we observe globally and in our data. Indeed, as seen in Figure 10, the labor-only model implies that at the highest farm sizes profitability per acre would never exceed that for the smallest (regime-1) farms. The smallest African farms would be the most proftitable on earth!

We now relax the assumption that the only input used is labor time and allow for the use of farm machinery to accomplish farm tasks. We allow machine time and labor time to be substitutes and emphasize the key feature of machinery, heterogeneity in machine capacity. We define capacity, consistent with definitions used for most farm equipment, as the amount of processed acreage a machine can accomplish per unit of time (e.g., acres covered per hour by irrigation or insecticide,

[19]To capture the difference is labor intensities, we consider a Cobb-Douglas with factors shares for *a*, e_1, and e_2 of ½, 1/3 and 1/6, respectively.

acres of corn per hour harvested). Unlike for *manual* labor, where individual heterogeneity in productivity per unit of time (within gender) is relatively low (and unrewarded in the market where time wages dominate (Foster and Rosenzweig, 1996), farm equipment devoted to specific tasks vary significantly in capacity and command different prices associated with capacity.[20] Thus we need to distinguish machine time and machine capacity, with the farmer choosing both machine capacity, based on acreage and capacity prices, and how much time to employ the machine.

To capture these ideas we redefine the nutrient production function as

(11) $$e(l,q,m) = (\omega_l(\xi l)^\delta + \omega_m((1-\frac{q}{\phi(a)})qm)^\delta)^{1/\delta},$$

where q is machine capacity and m is the number of units of time the machine is employed. The parameter ξ captures output per hour of labor time and the parameter δ captures the extent of substitutability between labor and machines.[21] Farmers can choose among machines of different capacities q. However, *effective* machine capacity depends on farm size. To model this, we define the function $\phi(a)$ with $\phi'(a) > 0$ capturing the loss associated with using a large capacity machine on a small plot, so that effective capacity is (1- $q/\phi(a)$) q. For example, a sprayer that can cover a radius of z yards would be cost ineffective on farms where the radii of farmed area are significantly less than z yards, assuming that machine prices rises with capacity. Similarly, it is not cost effective to rent an 8-row harvester for land that has four rows of crops.

We also allow for economies of scale in farm machinery capacity, such that the rental cost per unit of time x_m for a machine increases at a decreasing rate with machine capacity:

(12) $$x_m = p_q q^\nu,$$

where $0<\nu<1$ and assuming $q > 1$. Finally, we assume that operation of the machinery requires θ of family labor per hour of machine operation so that the hourly cost of a machine inclusive of labor is

[20]Farmers - that is the decision-makers - may differ importantly in capability relevant for making allocative decisions. We assume in the model, as is traditional, that all allocative decisions are correct, given technology and prices. We provide in an Appendix tests of whether farmer ability is correlated with owned landholdings. As noted, the within-farmer plot-specific relationships between profits and acreage indicate that any such a correlation is not solely responsible for the profitability patterns observed across landholdings of different size

[21]Some labor will be complementary with machine use, the labor used to actually run the machines, as we specify below.

$pq^{\nu} + w\theta$.

Profits, augmented to include the use of farm machinery, are

(13) $$\pi(a, l_h, l_f, q, m) = g(a, e(l_h + l_f, q, m)) - w(l_h) - wl_f - (w\theta + p_q q^{\nu})m\,.$$

In this model machine capacity is determined only by acreage, the parameter ν, and the ratio $\theta w/p$. The overall productivity of nutrients and the substitution of labor and machines in the nutrient production function affect the hours of machinery use but not machine capacity. In particular, *q,* solves

(14) $$\frac{\left(\phi(a) - 2q\right)\theta w}{p} + \left(1-\nu\right)q^{\nu}\phi\left(a\right) - (2-\nu)q^{\nu+1} = 0.$$

While a closed form solution for (14) is not generally available, it is evident that optimal capacity depends on the relationship between machine cost and capacity (ν) and on the effective capacity of machines by area $\phi(a)$, as well as the different prices.

Scale economies in farm production associated with machinery thus require both that effective machine capacity depends on acreage and that there are economies of scale in machine capacity. If there were no cost advantage to using higher-capacity machines, even large farmers would use the smallest capacity machine. And if there were only a cost advantage to using larger-capacity machines but no relationship between effective capacity and area given actual capacity we would not observe small machines being employed on small plots. We will obtain estimates of ν from actual price lists and from our survey data. These indicate non-linear pricing of capacity. The ν estimates will also allow us to identify non-parametrically the $\phi(a)$ function, as shown below.

The model implies that not only will the use of machinery increase with farm scale and with rising labor costs but so will the machine capacity chosen by the farmer. Implicitly differentiating (14), we get

$$\frac{dq}{da} = \frac{q^2\phi'(a)}{\left(\phi(a) - q\right)\left(\phi(a)\nu - 2q\nu - \phi(a) + 4q\right)} > 0,$$

which is positive as long as $\phi'(a) > 0$[22], and

[22] This expression must be positive as the first-order condition for q implies $q<\varphi(a)/2<\varphi(a)$ and the second-order condition requires $\left(\phi(a)\nu - 2q\nu - \phi(a) + 4q\right) > 0$.

$$\frac{dq}{dw} = \frac{\theta q\left(\phi(a)-2q\right)^2}{pq^{\nu}\nu\left(\phi(a)-q\right)\left(\phi(a)\nu-2q\nu-\phi(a)+4q\right)} > 0 .$$

The mechanism here is that at higher wages for the machine operator one wants to use fewer hours of machines per unit of nutrient added and this is only possible if the machine is higher capacity. On the other hand, an increase in the baseline cost of machinery tends to lower machinery capacity because it affects both the cost of capacity and the cost of machinery

$$\frac{dq}{dp} = -\frac{q\left(\phi(a)-2q\right)^2 w\theta q^{-\nu}}{p^2\nu\left(\phi(a)-q\right)\left(\phi(a)\nu-2q\nu-\phi(a)+4q\right)} < 0 ,$$

and a reduction in scale economies of will similarly lower capacity at any given scale of operation:

$$\frac{dq}{d\nu} = -\frac{q\left(\phi(a)-2q\right)^2 w\theta q^{-\nu}\ln(q)}{\nu\, p\left(\phi(a)-q\right)\left(\phi(a)\nu-2q\nu-\phi(a)+4q\right)} - \frac{q\left(\phi(a)-2q\right)}{\nu\left(\phi(a)\nu-2q\nu-\phi(a)+4q\right)} < 0 .$$

As noted, the determination of machine capacity *q* is, in the context of the model, independent of the return to nutrients; it simply involves minimizing the cost of producing a unit of nutrients using a machine, given land area. On the other hand the optimal number of hours the machine of capacity *q* is employed depends on nutrient use and on the cost of labor used to provide nutrients not associated with machine use. For example, mechanized sprayers may be used to spread herbicide and thus control competition for nutrients by preventing weed growth. Alternatively, labor may be used to remove weeds once they have grown. The determination of optimal machine use is thus considerably more complicated than the choice of optimal capacity, and analytical derivatives cannot in general be signed. We therefore proceed by calibrating the model to key parameters in the data and then use the calibrated model to examine how machine and labor use respond to underlying costs.

The full profit function, incorporating (11) and (13) is

$$\pi(a,l_h,l_f,q,m) = p_o a^{\alpha}(\omega(\xi l)^{\delta} + (1-\omega)(q(1-\frac{q}{\phi(a)})m)^{\delta})^{(1-\alpha)/\delta} - wl - (w\theta + p_q q^{\nu})m$$

With these basic results in place, we now consider how economies of scale in machinery capacity affect the relationship between average profitability and acreage, by simulating the model incorporating machinery, adding the parameters , δ=0.65, ν=0.32, $\theta = 1$, p_0=1.4, ξ=1, and $\phi(a) = a^2 / 40$ to the initial labor-only model.[23] Figure 14 displays the plots derived from the two models - the initial model with only labor, from Figure 10, and the model with labor and machinery. Both figures display the U-shaped patter for average-profitability. The key differences are that in the model that permits machine use and incorporates economies of scale in machine capacity, the decline in per-acre profits halts at an earlier point in the land distribution and per-acre profitability eventually rises above that implied by the labor-only model. Thus at the highest farm size, profits per acre exceed those of the smallest farms, replicating what we see across when comparing small-farm agriculture and large-farm agriculture across countries.[24]

Figure 15 displays expenditures per acre for labor in the labor-only model, from Figure 12, and for labor and machinery per-acre labor expenditures in the model with both labor and machinery. Labor expenditures decline with acreage at the lower farm sizes in both models. However, as can be seen, starting at around 9 acres machines are used to augment family labor. With a high substitutability assumed in the simulation between labor and machine time, the use of machinery occurs before the threshold acreage for hiring labor in the no-machinery model and thus no hired labor is employed in the model incorporating machine use. Thus labor expenditures per acre continue to fall as acreage increases. In addition, however, as scale increases farmers employ higher-capacity machines. This lowers the cost of producing nutrients, given ν<1. The net effect on expenditures is thus ambiguous because, depending on the factor elasticity of demand, the farmer may choose to increase nutrients per acre. At the parameter values chosen, machine expenditures per acre rise with acreage but at a decreasing rate.[25]

---

[23] The parameters of the machinery price function are from our estimates based on the data, as described below.

[24] Given our simplifying assumption in the simulation that labor and machinery are perfect substitutes the increase in effective machine capacity with acreage, given ν<1, and consequent decrease in the need for labor means that the farmers never exit autarchy. Average profitability thus rises continuously beyond the threshold at which hired labor is used in the model without machinery solely due to machine capacity economies.

[25] Finally, the patterns of labor and machinery expenditures by land size have implications for how yields and profits per acre differ as operational scale increases. Simulation results from the model permitting machine use indicate that both per-acre yields and profits display the U-shape but they differ in slope size and the location of the nadir. The change in the gap between per-acre profits and output value initially reflects how labor expenses change as farm size increases. As the farm moves to autarchy with increasing scale both labor expenses per acre and yields per acre drop but the gap between the two also declines, reflecting the fall

3. Identifying Scale Dis-economies due to Labor Market Transaction Costs

a. Production function and profit function methods.

Standard methods used in the literature to identify the existence of scale economies have been to estimate production functions or profit functions. The model helps clarify the circumstances under which scale dis-economies associated with transaction costs in the labor market can be detected by these two methods. By construction the model is constant returns to scale in land and nutrients. Scale economies thus do not operate through the technology but through (a) the cost of labor (b) the selection of machine capacity. For illustration we assume a Cobb-Douglas form for (1). We first consider the labor-only model incorporating labor market fixed costs as above, and where $e = l_e = l_h + l_f$, so output is

(15) $$y = a^{\alpha} l_e^{1-\alpha} .$$

When the production function is estimated directly, say by regressing the log of output on log of land and labor time,

(16) $$\ln y = \beta_0 + \beta_1 \ln a + \beta_2 \ln l_e .$$

α is identified. The standard test for scale economies is then whether $\beta_1 + \beta_2 = 1$. This equality holds of course under the assumed production function as

(17) $$\frac{\partial \ln y}{\partial \ln a} = \beta_1 = 1 - \beta_2 = 1 - \frac{\partial \ln y}{\partial \ln l_e} = \alpha .$$

Thus, estimating the production function will correctly identify that there are not scale economies in the production function. However, such an approach will miss the presence of scale economies more generally. The reason is that the fixed costs of labor affect the cost of inputs and thus the level of inputs but not the relationship between inputs and outputs.

We now examine how the sources of scale economies in the model are manifested in profit function estimates in the labor-only model with labor-market transaction costs, again specifying the production technology as Cobb-Douglas with nutrients linear in labor time. The first implication of the model is that the existence of transaction costs in the labor market imply that estimates of the profit function area parameter will differ by regime; that is the parameters of the profit function will be a function of land area. Thus, unlike when directly estimating the production function, scale dis-

in per-acre labor expenses when the farm is relying only on its fixed family labor endowment in autarchy. Then as the farmer begins to use machines the gap between per-acre yields and per-acre profits widens as the farmer is increasingly able to take advantage of higher-capacity machinery, saving on labor expenses and exploiting scale economies in machine capacity.

economies associated with input transaction costs can be detected.

In the smallest-area, off-farm labor regime, profits are

$$\pi = a^{\alpha} l_f^{1-\alpha} - w_1 l_f \tag{27}$$

Using the fact that the marginal product of family labor must equal w_1 it follows that the regression

$$\ln \pi = \beta_0 + \beta_1 \ln a \tag{28}$$

yields

$$\beta_1 = 1 , \tag{29}$$

which (correctly) implies that profits exhibit constant returns to scale in this regime. In the autarchy regime, at higher land area levels, however, labor is constrained at the family endowment l. The same regression yields

$$\beta_1 = 1/(1 + \frac{(1-\alpha)(a/l)^{\alpha} - w_1}{\alpha(a/l)^{\alpha}}) < 1 \tag{30}$$

where the numerator in the fraction is the positive amount by which the marginal product of labor exceeds the variable component of the wage, w_1. Thus, in the autarchy regime there are decreasing returns to scale, the marginal effect of land area in this range decreases with area. Finally, in the regime in which hired labor is employed on farm and transaction costs diminish proportionally with scale

$$\beta_1 = 1/(1 - \frac{w_0}{\alpha a^{\alpha} l_e^{1-\alpha}}) > 1 , \tag{31}$$

and there are increasing returns - β_1 rises with land area.

In sum, as acreage increases in the labor-only model, the *marginal* effect of land area on profits is initially constant, then negative, and then positive. Testing for the existence of transaction costs in the labor market, if labor market transaction costs are the source of scale economies, thus entails a search for this pattern of profit-function β_1 coefficients by land area. If farm machinery is not employed at low acreages, as our model simulations imply, the profit-function approach, but not estimates from a production function, will correctly identify the scale dis-economies associated with labor market transaction costs - that is, the decrease in profitability with acreage.

The uptick in the marginal effect of land on profitability beyond a land area threshold can be explained by both the presence of labor market transaction costs and the employment of farm machinery that is characterized by scale economies in capacity. This is because estimates of the effect of land size on farm profits by land area will also reflect machine capacity scale economies and pricing. To see this we incorporate into the Cobb-Douglas production function machine scale economies as in the model with $\theta = 0$ so that $e = mq(1 - q/\phi(a))$. Profits optimized with respect to machine capacity, in this special case, are therefore

(32) $$\pi = a^{\alpha}(m\phi(a)\frac{1-\nu}{(2-\nu)^2})^{1-\alpha} - (p(\phi(a)\frac{1-\nu}{2-\nu})^{\nu})m$$

Using the fact that the first-order condition for machine hours must be satisfied, a regression of ln profits on ln area

(33) $$\ln \pi = \beta_0 + \beta_1 \ln a$$

yields the coefficient

(34) $$\beta_1 = 1 + \frac{aq(1-\alpha)\phi'(a)}{\alpha\phi(a)(\phi(a)-q)}.$$

The profit function parameter is again a function of land area, and is greater than 1 if $\phi'(a) > 0$.[26]

We first test for varying β_1 coefficients in the profit function by land size. As noted, an initial decline in the β_1 coefficient as farm scale increases is consistent with the existence of labor market transaction costs. We estimated *a locally*-linear profit equation using LWFCM of the following form:

$$y_{ijt} = \beta_0(a_{ij}) + \beta_1(a_{ij})a_{ij} + \Sigma\, \beta_n(a_{ij})X_{ijt} + \delta_{jt}(a_{ij}) + \eta_{ijt}(a_{ij})$$

where the y_{ijt} are total profits over the *kharif* season for a farmer *i* in village *j* in year *t*; the X_{ijn} are soil characteristics, the δ_{jtk} are village/time fixed effects (capturing village-level time-varying input prices and weather); and the η_{ijt} are time-varying land specific iid errors. Figure 16 plots the coefficient β_1 and its 95% confidence interval by farm size. The figure corresponds to the implications of the model in which there are decreasing returns to scale at smaller farm sizes, as larger farms first use sub-optimal amounts of labor and then employ more-expensive low-hour hired labor. Of course, the subsequent rise in β_1 - increasing returns - is consistent both with a fall in average labor costs due to the proportionate decline in the fixed cost component of hired labor expenditures and with the exploitation of scale economies associated with farm machinery.

The model incorporating fixed transaction costs in the labor market implies that the smallest

[26]Similarly, estimation of a production function identifies the existence of scale economies if they are due to machine capacity pricing and capacity economies of scale. For example, if the total rental cost of machinery, $x_m = pq^{\nu}m$, is used as a measure of the capital input, as is common in estimating agricultural production functions, the regression of ln y = $\beta_0 + \beta_1 \ln a + \beta_2 \ln x_m$ yields $\beta_1 = \alpha + a(1-\nu)(1-\alpha)[\varphi'(a)/\varphi(a)]$ and $\beta_2 = 1 - \alpha$. Thus, the sum of the β's exceeds 1 if $\phi'(a) > 0$ and $\nu < 1$.

farms are efficient producers because they spend within each operation some of their time to off-farm work and thus allocate the optimal amount of family labor time to own production, priced at the marginal cost of labor. As land size increases less time is spent off farm until all family labor time is devoted to own farm production. We should therefore observe days of off-farm work by family members to decline with land size. This is what we see in Figure 17, which plots the lowess-smoothed relationship between average days per month male and female members of land-owning households spend in off-farm work. For the smallest farms, prime-age males spend on average almost 16 days of the month working off farm; at 10 acres this drops to only 6.5 days.[27]

If at higher acreages farmers are reluctant to hire labor because of transaction costs then total own farm labor costs (family + hired) *per acre* should also fall with land size. The first and second columns of Table 4 report village/year fixed effects estimates of the effect of farm size on the log of labor costs per acre without and with, respectively, controls for land quality. Both estimates indicate that as farm size increases the labor-land ratio declines - for every one acre increase in farm size there is a four percent fall in per-acre labor costs. The estimates are similar when we control comprehensively for all farm-level characteristics (wealth, ability) by using farmer/year fixed effects, reported in columns there and four of the table. - as plot size increases, per-acre labor costs decrease by about four percent. However, the fall in per-acre labor costs could be simply due to the substitution of machinery for labor. We need to show that transaction costs play a significant role in the decline of labor intensity at lower acreages.

b. Direct tests of the role of labor market transaction costs in determining scale economies.

To test more directly that a mechanism for the U-shape in the marginal effect of land size on profits is due to changes in unit labor costs by land size, we first plotted the relationship between the real average hourly wage and farm size. In the ICRISAT data, family labor is priced at the marginal or eight-hour wage (as if fixed costs were fully born by the employer), while hired labor is priced at the wage actually paid. Since the latter will be higher per-hour for low-hour hired labor according to the wage schedule, we should see that moving from the smallest farms to the largest, the average hourly wage first rises, as farms initially employ only family labor and then employ low-hour hired labor. At some threshold, the average wage paid falls as less low-hour labor is used. This is what we see in Figure 18.

To further test that the marginal land size effect on unit labor costs differs by land size, we

[27]There is no relationship between farm size and total days worked in the month. Own farm production time concomitantly rises with farm size

estimated the relationships between the fraction of operations in the *kharif* season that employ low–hour (hours<=6) daily hired male labor, hired tractor services, and hired bullock pair services and the corresponding average hourly wages for each. We allow the marginal effect of land size to differ by land size by employing a quadratic in land. The estimates, which include village/year fixed effects and the plot characteristics, are reported in Table 5. For all three factors we see that the fraction of low-hour operations declines with farm size, and for both hired male labor and hired bullock pairs so do average wages. Thus, these estimates account for the rise in profitability per acre, and fall in unit-costs, above some threshold due to the declining use of high-cost hired labor and hired bullock pairs. The exception is for tractors, for which the effect is statistically insignificant but positive. This may reflect the fact that on larger farms more expensive tractors with more capacity are hired, an issue we will discuss below.

There are three limitations to the estimates in Table 5. First, there may be incomplete control for land characteristics, which may be correlated with land size and with input use. Second, the model and the labor cost figure suggests that a quadratic specification will not fully capture the change in the marginal effect of labor demand with farm size. Third, farm acreage is positively correlated with farmer wealth, so the acreage effects will in part reflect wealth effects. To remedy these limitations, we exploit the plot-specific panel feature of the data and intertemporal rainfall variation to estimate using plot fixed effects the effects of rainfall on plot-specific input usage and average input costs by plot size.[28]

For most levels of rainfall in the semi-arid tropics in which the ICRISAT farmers are located, increases in rainfall increase input productivity and thus should increase input use. The exceptions are inputs that are employed in the planting stage, which principally occurs before the major component of the rainfall realization is known. Tractor use is mostly confined in the sample to planting-stage operations (tillage, plowing). Thus we use tractor employment as a placebo - rainfall should neither affect tractor hours nor the average per-hour rental price of tractors. On the other hand, for small plots higher levels of rainfall will increase average wages if the additional

[28]Plot size and not just total farm size will matter for input costs as long as operations differ across plots in the same time period. To gage the synchronicity of plot operations, we exploited the daily calendar of operation start dates in the survey. We used these to compute the standard deviation of the start date of each operations across plots for farmers with two or more plots and the standard deviation of the distribution of average operation-specific start dates across farmers in the same village. Note that if in all operations the days of initiation were the same, the average standard deviations would be zero. Appendix Table 1A in the appendix reports these results, which show that the average standard deviations in operation start dates across plots for the same farmer are significantly different from zero and almost as large as those characterizing the synchronicity of operations across farmers..

rainfall induces the hiring of low-hour post-planting labor while for the larger plots, increases in rainfall induces a shift from low- to normal-hour operations and average input costs decline.

We first establish that rainfall does indeed increase plot-level productivity and affects the demand for inputs. In the first column of Table 6 plot fixed-effects estimates of rainfall and rainfall squared on *kharif*-season profits from each plot are reported. As expected, increases in rainfall increase profits. And, in the second column, the estimates indicate that increases in rainfall also increase the number of hours of hired labor employment and hired bullock pairs, but the latter effect is only statistically significant at the .07 level (one-tailed test), consistent with bullocks being primarily used in the early stages of the production cycle. The effect of rainfall on tractor hours, as expected, is not statistically significant by conventional standards and is economically insignificant as well. In parallel, an increase in rainfall decreases both average hired male labor and bullock rental costs, but has no effect on the hourly cost of tractors.

Having found that variation in rainfall on a given plot affects its profitability, the number hired labor hours, and per-hour hired labor costs on average, we then estimated the effects of rainfall on the fraction of operations on the plot that employ low-hour hired male labor and the average wage paid by plot size, using LWFCM. The plot fixed-effect estimates of the effects of rainfall at mean rainfall by plot size on low-hour labor use, and the associated 95% confidence interval, are reported in Figure 19. The figure is consistent with the shifting of regimes of labor employment in the model - at small plot sizes, increases in rainfall statistically significantly increase hired low-hour labor use while for larger plots low-hour labor operations are statistically significantly reduced when rainfall increases.[29] And in Figure 20, among the larger plots increases in rainfall statistically significantly reduce average hourly hired labor costs.[30]

4. Identifying Equipment Scale Economies as a Source Farm Scale Economies: the Case of Sprayers.

Figure 6 indicates that beyond a threshold plot size there are positive scale economies that continue up through the largest land size. Our model implies that if the reduction in the importance of input transaction costs were the only source of scale economies, or if machinery was employed by larger farms only to completely eliminate the use of hired labor, then larger farms would be no more

[29] Because, as noted, plot size and farm size, and thus farmer wealth, are positively correlated in the data, the rainfall coefficients at small plot sizes may be underestimated due to credit-market or liquidity constraints on the ability of small farmers to employ additional hired labor. More relaxed liquidity or credit constraints for larger farmers, however, cannot explain the negative effect of rainfall on per-unit labor costs.

[30] The effects of rainfall on per-acre profits does not vary by plot size over the full range of plot sizes.

productive than the smallest farms unless there are also machinery-specific economies of scale, which is not what we observe in our sample, or across the world. In this section we adduce indirect and direct evidence on machine scale economies, inclusive of estimation of capacity pricing schedules and the parameters of the effective capacity function, the fundamental components of such scale economies.

As in most data sets describing farming where mechanized equipment is used, we see that larger farms are more likely to be using mechanized farm equipment, as shown in Figure 21 for tractors and sprayers. There has been scant evidence, however, on the relationship between machine capacity and scale. The data on usage of machinery, as seen in Figure 22, is suggestive of the rise in machine capacity with farm scale - while average hours of equipment use per acre first increases with scale, above 12 acres per-acre use of both types of equipment *declines* with farm size. This decline in machine use on a per-acre basis as farm size increases among larger farms is consistent with machine capacity scale economies. However, these patterns are not directly informative about whether machine capacity actually increases with farm size, whether there are scale economies in capacity due to non-linear capacity pricing, or at what scale, if any, capacity scale economies dissipate completely. To address these issues we need a measure of machine capacity.

The capacities of machines used by farmers are rarely, if ever, available in data sets based on household surveys from low-income countries. The ICRISAT data set, however, permits the computation of capacity for one type of equipment - sprayers. This is because there is information on the amount of material sprayed - weedicide and insecticide - as well as information on hours of sprayer usage by plot and operation. These data can thus be used to compute capacity - amount sprayed per hour. Sprayer capacity is typically given in spray rates for a given nozzle size - material volume per time unit. The relevance of this measure for farming scale is that flow rates translate directly into area sprayed per hour, given a target amount of material per area.

Another advantage of sprayer technology is that we can exploit the information on input use by operation to directly measure the labor savings from spraying. This is because an important alternative to spraying for protecting plant nutrients is weeding, which is typically done manually. The data provide the hourly rental rate for the sprayer used and labor usage for both spraying and weeding operations. Figure 23 displays the relationships between per-acre expenditures on sprayers and on the labor used in spraying and weeding. As can be seen, as farm size increases farmers are using more expensive, and thus, as we will see, higher-capacity sprayers;[31] and weeding labor costs per -acre plummet while per-acre labor time used in spraying only slightly increases with scale.

We focus on sprayer technology so that we can for the first time obtain direct evidence from data in a low-income setting on how capacity heterogeneity in equipment contributes to economies of scale in agriculture that persists above a farm (plot) size threshold, conditional on machine use. This, of course, does not mean that sprayer technology is the sole source of economies of scale due to mechanization. But spraying weedicide and insecticide is an important operation. Spraying labor costs alone account for 13.6% of total input costs in the *Kharif* season. And as can be seen in Figure 22 there are more hours of use of sprayers at every land size than hours of use for tractors, the next most used machinery.

The data on asset ownership from the last round of the ICRISAT survey indicates that ICRISAT farmers use two types of sprayers - manual sprayers, whose median cost in 2014 is 700 rupees, and power sprayers, whose median cost is 2700 rupees. Of the 10% of farmers who own a sprayer, 25% own a power sprayer. Table 7 reports, based on these data, regressions of tractor ownership, ownership of any sprayer, and, conditional on owning a sprayer, ownership of a power sprayer on land size and land wealth (based on the farmer's own assessment of the rental value of the land). As can be seen scale, not just wealth, matters for equipment ownership: net of wealth, farmers with more land area are more likely to own a tractor and sprayer; and they are more likely to own a power sprayer if they own any sprayer.[32]

Even among power sprayers, however, there is heterogeneity in capacity. Table 8 provides information taken from the web site of an Indian purveyor of power sprayers (KrisanKraft) that provides power sprayer prices by precisely the measure of capacity we can construct from the ICRISAT data - the amount a sprayer can broadcast in litres per hour. A notable feature of the listing is that there are substantial differences in sprayer capacities - the spray rate of the highest-capacity power sprayer is over 13 times that of the lowest-cost model. More importantly, the posted price schedule exhibits economies of scale in sprayer capacity, the sprayer price per unit of capacity

[31]The per-hour price of the sprayer used increases monotonically with acreage. As noted, the average hourly wage falls with acreage at medium farm sizes and is constant at larger farm sizes.

[32]The market value of land fully captures land quality in a well-functioning market. However, the effect of scale conditional on the value of land may under-estimate scale economies because the value of land may reflect in part the existence of scale economies, with per-acre rental prices varying positively with land size. Over all acreages we find that the reported rental prices of individual plots do indeed rise with total acreage.

decreases as capacity increases.[33]

The survey data on input usage suggest that farmers are exploiting economies of scale in spraying. Table 9 reports village/year fixed effects estimates of the effects of land size on the use of any sprayer, on weeding hours per acre, on sprayer hours per acre, on the log of the price of the sprayer used, and on the sprayer flow rate (capacity) net of the effects of the land quality variables. These estimates indicate that net of year-village effects, larger landowners use pricier and higher-capacity sprayers and larger landowners spend less time per acre in *both* spraying and weeding compared with smaller farmers.

To test directly for scale economies in spraying and the limits, if any, to sprayer scale economies we use the structure of the model, simultaneously estimating the effective capacity function $\phi(a)$ and the key parameter of the price function ν from the information on the capacity and per-hour rental prices of the sprayers used by the ICRISAT farmers. The challenge for estimation is that we observe only the capacity of the sprayer that is used by the farmer: capacity and the per-capacity price are choice variables. As a consequence we use GMM using land area and land area squared as instruments.

Equation (14) solves for the optimal choice of *q* and embeds within it the effective capacity function and the capacity pricing parameter ν. It also contains, however an additive term that includes the village wage rate *w* and the base price *p* of the capacity pricing schedule, which may be endogenously determined. We thus rearrange (14) and difference across randomly selected pairs of households *i* and *i'* in each village *j* to eliminate *w* and *p* in (14). Taking the log (12), we then get moment conditions of the following form

$$(14)\qquad E\left(\frac{(1-\nu)q_{ij}{}^{\nu}\phi\left(a_{ij}\right)-(2-\nu)q_{ij}{}^{\nu+1}}{\left(\phi\left(a_{ij}\right)-2q_{ij}\right)}-\frac{(1-\nu)q_{i'j}{}^{\nu}\phi\left(a_{i'j}\right)-(2-\nu)q_{i'j}{}^{\nu+1}}{\left(\phi\left(a_{i'j}\right)-2q_{i'j}\right)}\middle| a_{ij},a_{i'j}\right)=0 \ .$$

$$(14)\qquad E\left(\ln(x_{ij})-\nu\ln(q_{ij})-\ln(x_{i'j})-\nu\ln(q_{i'j})=0\middle| a_{ij},a_{i'j}\right)=0 \ .$$

We parameterize $\phi(a)=b_0+b_1a+b_2a^2$ and employ GMM using land area and land area squared as instruments to and the b_k.

[33] We will formally test below whether the sprayer capacity ν parameter is less than one in the KrisanKraft price schedule and in the ICRISAT data based on the sprayers rented by the farmers.

If b_1>0 and b_2<0, we can identify a maximum farm scale at which further increases in acreage could not exploit *existing* equipment scale economies. This establishes the land size at which per-acre profits are maximized - the technologically-determined optimal land size at which further increases in scale would not increase productivity but below which, given that b_1>0, farms are less productive net of costs. It is important to note that we can only identify the farm scale upper bound, if any, based on machinery that is actually available to the ICRISAT farmers. Machinery scale economies for farm sizes beyond the maximum scale of the farms in the population, if any, cannot be estimated because machinery for such farm sizes would not be marketed - they would not be available for rent or purchase. Farmers (and policy-makers) would thus have limited knowledge of how expansion of scale beyond that in the population would reap benefits via machine scale economies.

More importantly, finding a maximum located at the upper tail of the actual land distribution would suggest an equilibrium trap - no single farmer would attempt to expand land size beyond this truly local maximum because there are no *available* machines that could be used to exploit the increase in scale. However, if there were a land consolidation so that a sufficient number of farms were above this threshold, there might be enough demand for higher-capacity machinery to support a market for them. There would then be enhanced farm efficiency at larger scale than is found in the setting.

Table 10 reports the GMM estimates of the effective capacity and pricing function parameters and their robust standard errors. All parameters are precisely estimated. We can reject the hypothesis that ν = 1 and thus that there are no scale economies arising from the cost of higher capacity machines. We can also compare our estimate of ν based on the sprayers used by the ICRISAT farmers to that characterizing the price schedule for the four power sprayers sold by KrisanKraft, listed in Table 8, and to that for four power sprayers offered in the United States, as surveyed in Stiles and Stark (2016). The estimated ν's are reported in Table 11. As can be seen, our GMM estimate of ν for the sprayers used by ICRISAT farmers is comparable to that for the KrisanKraft sprayers that are sold across India. However the estimated India sprayer ν is more than double that for the sprayers sold in the United States - economies of scale are evidently substantially greater for the power sprayers available in the United States.

The estimates of the b_k indicate that $\phi'(a) > 0$. Thus, smaller farms are less cost-effective than larger farms, given that ν is substantially less than one. The estimates also indicate that there is a land scale at which effective capacity reaches a maximum. The point estimate of the maximum is 24.5

acres, with a 95% confidence interval of ± 3.6 acres. The 2014 Census of all households in the 20 ICRISAT VLS villages indicates that only 1.1% of households owning land have total landholdings above even the estimated lower bound of the maximum (20.9 acres). As expected, there are essentially no farms that could exploit further scale economies, given the sprayers that are available. It does not suggest, as noted, that larger farms than are observed in the ICRISAT area would not be more productive; rather it is consistent with an equilibrium trap in which none of the largest farmers has an incentive to expand given the available sprayers in India. Of course, that most farms are below this maximum, conditional on the local availability of machinery, implies that there are other barriers to land consolidation, resulting in an excess number of farmers.

How does variation in v affect the choice of machine capacity q and machine and labor usage m and l? To gage the magnitudes of these effects using our estimates, we calibrate the additional parameters of the model, enhancing the production technology to allow for n non-plant-protection factors whose normalized input price is set to one:

$$\pi(a,n,l,q,m) = p_o a^{\alpha} n^{\beta} (\omega(\xi l)^{\delta} + (1-\omega)(q(1-\frac{q}{\phi(a)})m)^{\delta})^{(1-\alpha-\beta)/\delta} - n - wl - (w\theta + p_q q^{v})m$$

For the rest of the parameters of the model, we set $\alpha = 0.27$, $\beta = 0.65$,, and $\omega = 0.83$ by calculating factor shares from average weeding cost, spraying cost, other input costs and the value of output. We set $\theta = 1$ (one worker per machine), as it is the mode based on the data on machine and labor hours. The values of $\xi = 0.3$ and $\delta = 0.65$ are set to match the level and change with respect to scale of average weeding and spraying costs. The base-cost of sprayer capacity $p_q = 3.7$ is determined based on the average sprayer rental cost net of capacity using the estimate of v. The value $p_0 = 33.8$ matches average profits given other information. The wage $w = 21$ is set to the average wage for a full 8-hour day.

Table 12 contains the calibrated elasticities and their computed standard errors, based on the sprayer GMM error covariance matrix for farms at the median (3 acres) of the farm size distribution in the ICRISAT villages.[34] The first three rows display the elasticities with respect to the change in

[34] Let $K_1 =< n,m,l,q >$ denote the vector of endogenous variables, $K_2 =< a,\beta,\omega,\theta,\xi,\delta,w,p_o,p_q >$ the vector of calibrated parameters, $K_3 =< v,b_0,b_1,b_2 >$, the vector of estimated parameters, and $K_4 < K_2,K_3>$ the combined vector of exogenous variables. Then an analytic expression for the matrix of implicit derivatives is

$$\frac{dK_1}{dK_4'} = -\left[\frac{d\pi}{dK_1 dK_1'}\right]^{-1} \frac{d\pi}{dK_1 dK_4'} = \Gamma(K_1,K_2,K_3)$$

the ν . The estimates suggest that a reduction in the sprayer price slope parameter ν from that faced by ICRISAT farmers to that characterizing the price schedule for US power sprayers would increase the capacity of the sprayer chosen by 5%, increase sprayer hours by 23.3% and reduce weeding hours by 3% even for these very small farms.

The estimates also indicate how changes in wage rates affect plant protection inputs, with endogenous sprayer capacity choice. In particular, a doubling of the wage reduces the usage of sprayers by 76%, reflecting the fact that each hour of sprayer use is assumed to require one hour of labor, but reduces weeding hours by 130%. The disproportionate decline in weeding hours reflects the substitution of sprayer capacity for labor, as sprayer capacity increases by 3% in response to the wage change. Finally, sprayer capacity choice is considerably more sensitive to scale than to wage rates: a doubling of farm size (in this case to only 6 acres) would increase the capacity of sprayers used by 30% and increase sprayer use by 140%. The increase in scale, which pushes up the capacity and use of sprayers, results in a decline in the per-acre number of weeding hours by 10%, reflecting again the substitution between mechanized spraying and labor-intensive weeding in protecting plant nutrients.

5. Conclusion

Much of the recent literature on economic development in agriculture has focused on the adoption of new technology. The thrust of the argument is that productivity differences across countries are importantly driven by the fact that farmers in low income countries do not have access to the same types of seeds and inputs that are available in other countries. One of the salient differences in agricultural sectors between countries of very different productivities, however, the scale of operations, has been relatively ignored.[35] The evidence on persistent differences in productivity within countries associated with scale cannot be easily understood to be a consequence

---

We use the delta method to compute an estimate of the standard errors of these derivatives. If Ω is the variance-covariance of the estimates of K_3 then for any particular element γ_{ij} of Γ:

$$Var(\gamma_{ij}) = \left(\frac{d\gamma_{ij}}{dK_1'}\frac{dK_1}{dK_3'} + \frac{d\gamma_{ij}}{dK_3'}\right)\Omega\left(\frac{d\gamma_{ij}}{dK_1'}\frac{dK_1}{dK_3'} + \frac{d\gamma_{ij}}{dK_3'}\right)'$$

where dK_1 / dK_3' is just a sub-matrix of dK_1 / dK_4'. This expression is evaluated at the estimated values of K_3 and K_2 and the calibrated parameters. Note that because the analytic expressions for the implicit derivatives depend both on the estimated parameters and the endogenous variables, the variance of the estimated derivative depends both directly on the variance-covariance of the estimated parameters and indirectly through the effects of this variance-covariance matrix on the estimated optimal endogenous variables.

[35] Seed technologies across low-income and developed-country agriculture are in fact not that different. Most of the cotton grown by the ICRISAT farmers is modern BT cotton, with cotton being the second most cultivated crop by area planted.

of differential lack of access to technologies given evidence on the transmission of technologies between farmers over time. If farmers within an area have equal access and markets for inputs work reasonably well, marginal products should be equated across farmers and thus it is unclear why there should be any differences in productivity across farms. Moreover, despite the global differences in farm productivity indicating higher productivity among larger farms, almost all of the evidence from within low-income countries suggests that agricultural productivity and scale are inversely related over the full distribution of farm sizes observed in those countries. Understanding what lies at the source of these differentials by scale may thus be indicative of important barriers to development in rural areas of low-income countries.

In this paper we examined unique data from India that allows us to look at agricultural operations among a wider distribution of farm scales than is typically observed in low-income countries because of the over-sampling of larger farms. We found a distinctive U-shaped pattern in which both small and large farmers are more productive, in terms of both yield and profitability, than intermediate sized farmers. This pattern replicates within one rural setting what is observed across countries, with productivity decreasing in scale for smaller farms and increasing in scale for larger farms. We showed that these productivity patterns by scale are not attributable to differences in measured aspects of land quality. We also established that measurement error in area, that might lead to high profitability per acre on small farms, is small and does not importantly explain the observed patters. In addition we found that the U-shape pattern is observed across plots for the same farmer, thus ruling out the importance of credit access as the main explanation of higher profitability among large farmers.

We proposed two alternative mechanisms that drive productivity differences by scale and can account together for the U-shape patter by scale across countries of the world and in India. We first considered the role of fixed costs in the hiring of labor and other inputs. We provided evidence that in fact many workers work for less than a full day, that the hourly price of a workers is higher when workers are used for part of day, and that intermediate sized farmers are most likely to employ workers part time. A similar pattern is observed for rented equipment. The implication of this pattern is that small farmers will be relatively efficient because the shadow price of family worker time is set by the outside market, as most smallholders work part-time off farm. As farm size grows however, the farm moves to autarchy and is reluctant to take on hired workers for just a few hours per operation. This leads to lower yields and profitability per acre. Eventually, this strategy proves costly and there is discrete jump upward in total work per acre due to the hiring of non-family workers.

Transaction costs in hired inputs can explain the U-shaped relationship between size and productivity, but cannot explain the continuous rise in profitability by scale beyond a threshold of farm size. The second mechanism we focus on to explain this component of the U-shape is the adoption of machine technology that is differentially adapted to farm size. The idea is that higher capacity machines do more work per hour and that the cost of these higher machines does not increase proportionately with capacity. However, large machines cannot be used at full capacity on small farms or plots. As a result large farmers use more productive machines and small farmers, if they use machines at all, use less productive machines that are also less cost-effective. This leads to an increase in yield and profitability as farm size increases.

We focused in particular on power sprayers because we are able to measure capacity based on time use and the amount that is sprayed. We showed that sprayers fit well with our model - optimal sprayer capacity increases with farm size and the shadow price of sprayer capacity increases less than proportionally as capacity increases. The mechanisms of hired-input transaction costs and economies of scale in machinery together allowed us to replicate the patterns of agricultural profitability and input costs per acre by farm size that are observed in our data and across countries of the world. Based on our parameter estimates, we found in particular that in our sample, economies of scale peter out at the very top of the land distribution (25 acres), suggesting that average farm size in India is, along this one dimension of technology that we can precisely estimate, too small - there are too many farms. We also find that the pricing of power sprayer capacity in India exhibits less scale economies than in the United States. Indian farmers as a consequence use less sprayer capacity and employ more weeding labor than they would if they has access to US sprayer technology. However, this difference in labor use and associated loss in productivity due to differing technologies is substantially less than the productivity differences across small and large farms in India.

Our results thus suggest that there are barriers to agricultural productivity growth that may not be easily overcome through individual investment decisions or cross-country technology transfer. First, on the labor cost side, the fixed costs of the hiring represent real costs associated with travel to farms and searching for work that may be particularly acute in relative sparsely-populated areas and/or areas where a small fraction of total land is arable. These geographic features can largely be taken as exogenous. Moreover, fluctuating labor demand is intrinsic to most agricultural production processes. Labor in agriculture does not substitute well across time because certain tasks must be carried out a specific times in the growing season. Farmers needing to execute a task cannot simply wait until they have enough undone tasks accumulated to hire a worker for a full day. Nor are multi-day contracts that reduce search costs likely to be a panacea as demand for labor is episodic. There is the potential to aggregate tasks by expanding acreage, but since different tasks will have different levels of labor demand, the optimal farm size for one stage will not be the same as the optimal farm size for other tasks.

Second, the ability to make profitable use of larger farms depends critically on access to and the pricing of larger-capacity machines. Again the episodic demand is likely to play a role. A single large farmer set among smaller farmers is unlikely to be able to justify the purchase of a large machine on his own. If there were multiple large farmers than one farmer might be the owner and then rent it to other farmers. Indeed there is currently an active market for tractors in many Indian villages. Our key finding that scale economies for sprayers have a limit that is precisely at the top of the land-size distribution suggest an equilibrium trap. While we see the adoption of larger-scale power sprayers, equipment such that as available to large Brazilian soy farmers is simply unavailable in most parts of India because there are no farms of that scale to demand them. In short, without a simultaneous expansion of large farms it is difficult to see how one expands the scale of available technologies, but it is difficult to see how this expansion would take place if such technologies were not available.

Finally, even barring cultural, economic, or legal barriers to the buying or selling of land, the U-shaped profitability curve complicates the process of transition from profitable small-holder agriculture to (more) profitable large-scale agriculture that makes use of hired labor. Land is likely to be accumulated through a series of individual transactions. In a world of one-acre farmers one would need twenty-five separate transactions among contiguous neighbors to get to an efficiently sized farms based on our sprayer exercise. This number of transactions would be complicated even in a relatively competitive market, but the need for contiguous plots raises important issues such as hold-up problems that will fully extract the rents that would otherwise accrue to a farmer who puts together a large farm.

In short, despite evidence of the potential gains to profitability of large farms, small farms are likely to be the dominant force of production in low income countries for the foreseeable future without external intervention. In India, the higher population density and increased availability of equipment increases the potential to move to large farms relative to Sub-Saharan Africa, but it seems unlikely that there will be a large-scale transition absent an expansion in high-paying employment opportunities that draw workers out of the agricultural sector, thus raising wages and encouraging farmers to expand machine usage. In other areas, such as Brazil we have seen a substantial transition to large-scale agricultural production in many areas. This reflects the availability of large-scale machines, the conversion of land that has previously been used for low intensity activities such as grazing, and the development of an effective series of highly profitable export markets.

In summary, the documented differences in productivity by farm size that have received substantial attention within countries over the years are not only indicative of underlying inequality in rural areas. They are also a barometer for the efficacy of markets in the allocation of technologies, workers, and ultimately land to their most productive usage. Imperfections in these markets importantly underlie the differences in profitability within villages and the identification and alleviation of these imperfections are ultimately required for the agricultural sector in low-income rural areas to reach its full potential.

References

Ali, D. A, and K. Deininger. 2014. "Is there a Farm-Size Productivity Relationship in African Agriculture? Evidence from Rwanda." World Bank Policy Research Paper 6770, World Bank, Washington, DC.

Allen, Robert C. 1988. "The Growth of Labor Productivity in Early Modern English Agriculture." *Explorations in Economic History* 25, 117-146.

Assunção, Juliano J., and Luis H. B. Braido. 2007. "Testing Household-Specific Explanations for the Inverse Productivity Relationship." *American Journal of Agricultural Economics* 89 (4): 980–90.

Barrett, C. B., M. F. Bellemare, and J. Y. Hou. 2010. "Reconsidering Conventional Explanations of the Inverse Productivity-Size Relationship." *World Development* 38 (1): 88-97.

Benjamin, D. 1995. "Can Unobserved Land Quality Explain the Inverse Productivity Relationship?" *Journal of Development Economics* 46 (1): 51-84.

Bhalla, S. S., and P. Roy. 1988. "Misspecification in Farm Productivity Analysis: The Role of Land Quality." *Oxford Economic Papers* 40 55-73.

Binswanger-Mkhize, H.P, Camille Bourguignon, and Rogier van den Brink. 2009. *Agricultural Land Redistribution: Toward Greater Consensus*. Washington D.C.: World Bank.

Binswanger, H.P., K. Deininger, and G. Feder. 1995. "Power Distortions Revolt and Reform in Agricultural Land Relations." In J. Behrman, and T.N. Srinivasan, ed. *Handbook of Development Economics,* Volume III. Amsterdam, the Netherlands: Elsevier Science B.V.

Bivar, Venus. 2010. The Ground Beneath Their Feet: Agricultural Industrialization and the Remapping of Rural France, 1945-1976. Department of History, University of Chicago Dissertation.

Cai, Zongwu, Mitali Das, Huaiyu Xiong, and Xizhi Wu. 2006. "Functional Coefficient Instrumental Variables Models." *Journal of Econometrics* 133 (1): 207-241.

Carletto, C., S. Savastano, and A. Zezza. 2013. "Fact or Artifact: The Impact of Measurement Errors on the Farm Size Productivity Relationship." *Journal of Development Studies* 103: 254-261.

Carter, Michael. R. 1984. "Identification of the Inverse Relationship between Farm Size and Productivity: an empirical Analysis of Peasant Agricultural Production." *Oxford Economic Papers* 36: 131-45.

Carter, M. R., and K. Wiebe. 1990. Access to Capital and its Impact on Agrarian Structure and Productivity in Kenya. *American Journal of Agricultural Economics,* 72: 1146-1150. 41

Collier, P., and S. Dercon. 2014. "African Agriculture in 50 Years: Smallholders in a Rapidly Changing World?" *World Development* 63: 92-101.

Davis, C. Austin. 2016. "Why Did Sugarcane Growers Suddenly Adopt Existing Technology?" University of Michigan.

Foster, A., and M. R. Rosenzweig. 1996. "Comparative Advantage, Information and the Allocation of Workers to Tasks: Evidence from an Agricultural Labour Market." *The Review of Economic Studies* 63(3): 347-374.

Foster, A., and M. R. Rosenzweig. 2011. *Are Indian Farms too Small? Mechanization, Agency Cost, and Farm Efficiency*. New Haven CT: Economic Growth Center, Yale University.

Hayami, Yujiro, and Keijuro Otsuka. 1993. *The Economics of Contract Choice: An Agrarian Perspective.* Oxford: Oxford University Press.

Hazell, Peter. 2011. "Five Big Questions about Five Hundred Million Small Farms." Paper presented at the IFAD Conference on New Directions for Smallholder Agriculture, IFAD Headquarters, Rome, Italy, 24-25 January, 2011.

Hazell, P., C. Poulton, S. Wiggins, and A. Dorward. 2010. "The Future of Small Farms: Trajectories and Policy Priorities." *World Development* 38(10): 1349-1361.

Hornbeck, Richard and Suresh Naidu. 2014. "When the Levee Breaks: Black Migration and Economic Development in the American South." *American Economic Review* 104(3): 963–990.

Huang, Jikan; Ruifa Hu; Jianmin Cao; and Scott Rozelle. 2008. "Training Programs and In-the-field Guidance to Reduce China's Overuse of Fertilizer Without Hurting Profitability." *Journal of Soil and Water Conservation* 63 (5): 165A-175A.

Jayne, T.S., J. Chamberlin, and Derek D. Headey. 2014. "Land pressures, the evolution of farming systems, and development strategies in Africa: A synthesis." *Food Policy*, 48: 1-17.

Kagin, J., E. Taylor, and A. Yúnez-Naude. 2015. Inverse Productivity or Inverse Efficiency? Evidence from Mexico, *Journal of Development Studies* DOI:10.1080/00220388.2015.1041515.

Kimhi, A. 2006. "Plot Size and Maize Productivity in Zambia: Is there an Inverse Relationship?" *Agricultural Economics* 35(1): 1-9.

Larson, D., K. Otsuka, T. Matsumoto, and T. Kilic. 2013. "Should African Rural Development Strategies Depend on Smallholder Farms? "An Exploration of the Inverse Productivity Hypothesis." *Agricultural Economics* 45: 1-13.

Mazumdar, D. 1965. "Size of Farm and Productivity: A Problem of Indian Peasant Agriculture." *Economica* 32(126): 161-173.

Muyanga, Milu and T.S. Jayne. 2016. "Is Small Still Beautiful? The Farm Size-productivity Relationship Revisited in Kenya."

Catherine Morrison Paul, Catherine Morrison; Catherine Paul; Richard Nehring; David Banker, and Agapi Somwaru. 2004. "Scale Economies and Efficiency in U.S. Agriculture: Are Traditional Farms History?" *Journal of Productivity Analysis* 22(3): 185-205 .

Schultz, T.W. 1964. *Transforming Traditional Agriculture*. New Haven, CT: Yale University. Press.

Stiles, Scott and C. Robert Stark, jr. 2016. "Estimating Farm Machinery Costs." *FSA21*, Division of Agriculture, Research and Extension, University of Arkansas.

Vollrath, D. 2007. "Land Distribution and International Agricultural Productivity." *American*

Journal of Agricultural Economics 89(1): 202-216.

Yotopoulos, P.A., and L.J. Lau. 1973. "A Test for Relative Economic Efficiency: Some Further Results." *Am. Econ. Rev.* 63(1): 214–223.

Zaibet, L.T., and E.G. Dunn. 1998. "Land Tenure, Farm Size, and Rural Market Participation in Developing Countries: The Case of the Tunisian Olive Sector." *Economic Development and Cultural Change* 46(4), 831-848.

Zhang, B., and A.C. Carter. 1997. "Reforms, the Weather, and Productivity Growth in China's Grain Sector." *American Journal of Agricultural Economics,* 79(4): 1266-

Figure 1. Percent of Households with Operational Landholdings Below 10 Acres, by Country

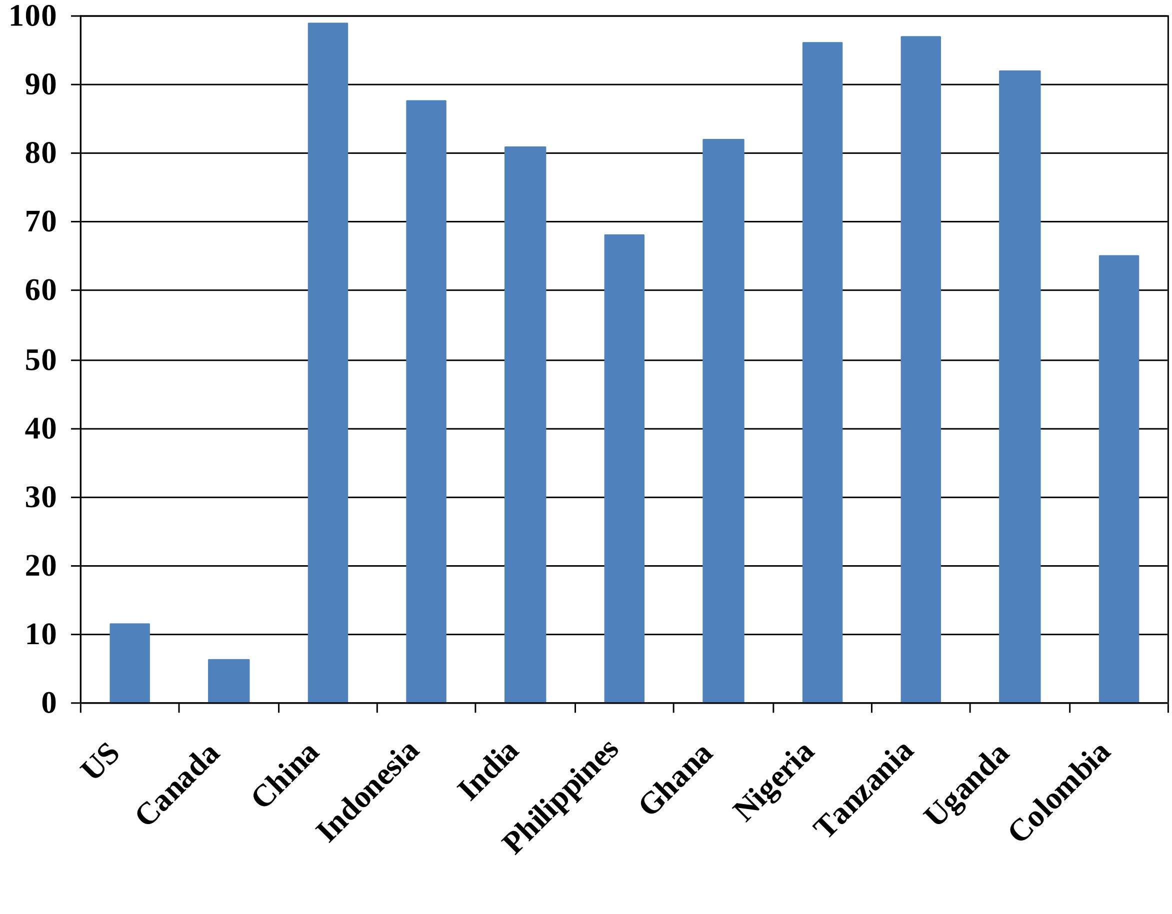


Figure 2. Soybean Yields (Metric Tons per Hectare) in 2016, by Country (Source: USDA, 2016)

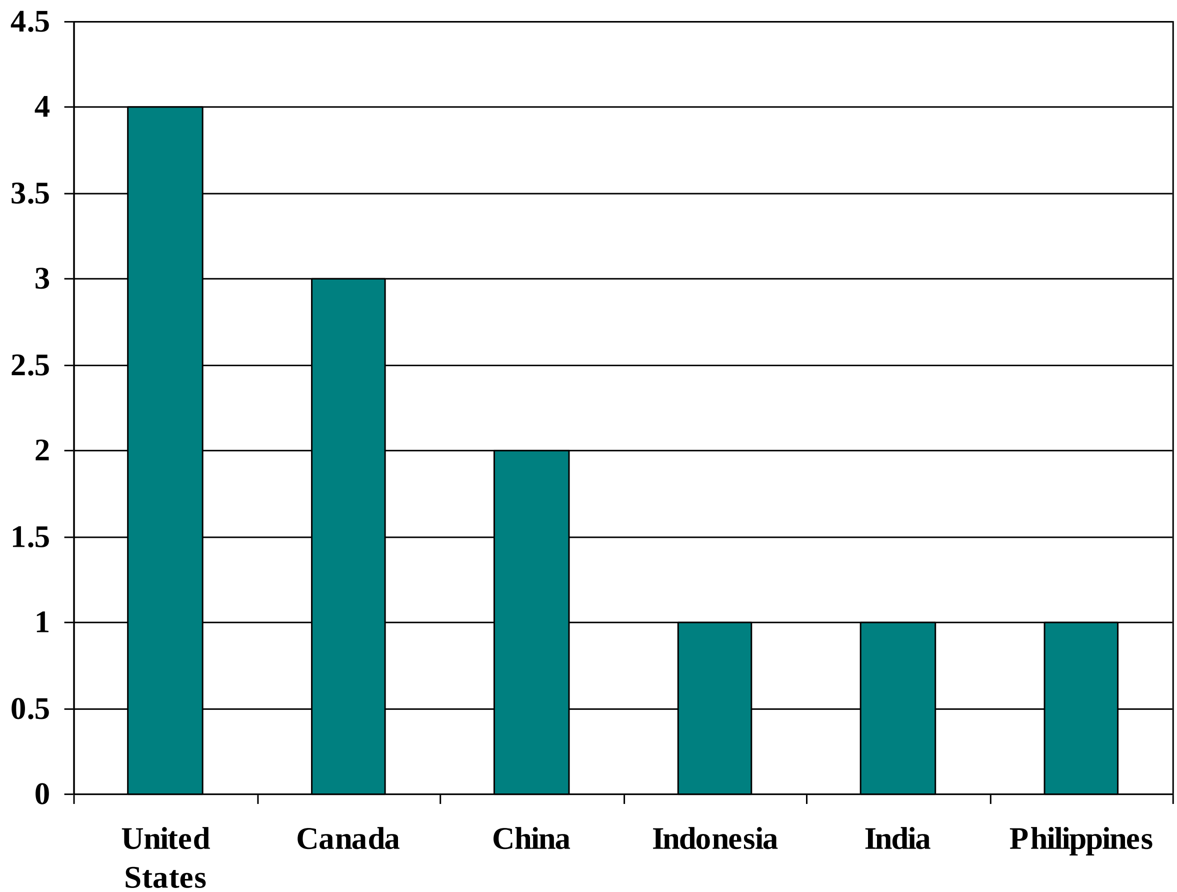


Figure 3. Fertilizer Intensity (Kilograms per Hectare) in 2014, by Country (Source: World Bank, 2016)

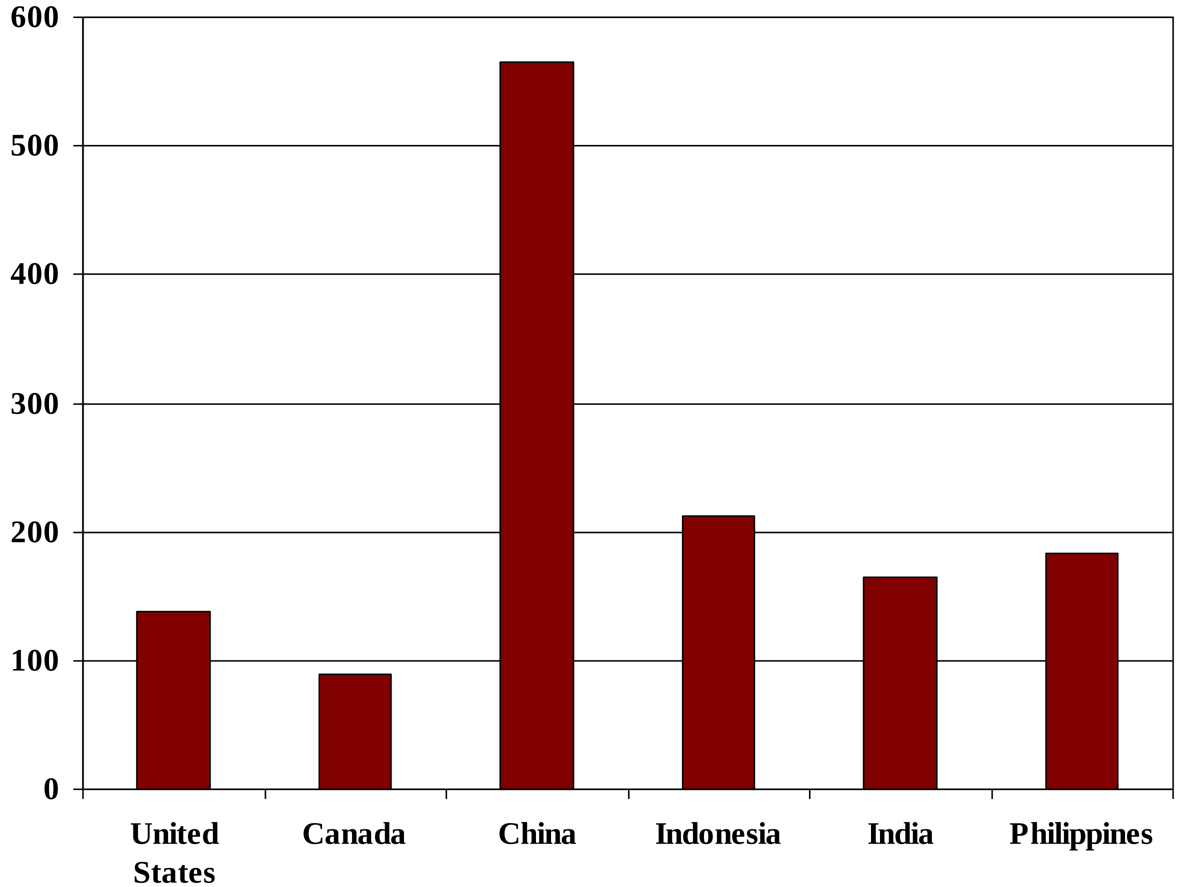


Figure 4. Relationship Between Plot Value per Acre (Nigerian Naira) and GPS-Measured Plot Size (Acres) (Nigeria General Household Panel, 2015-16)

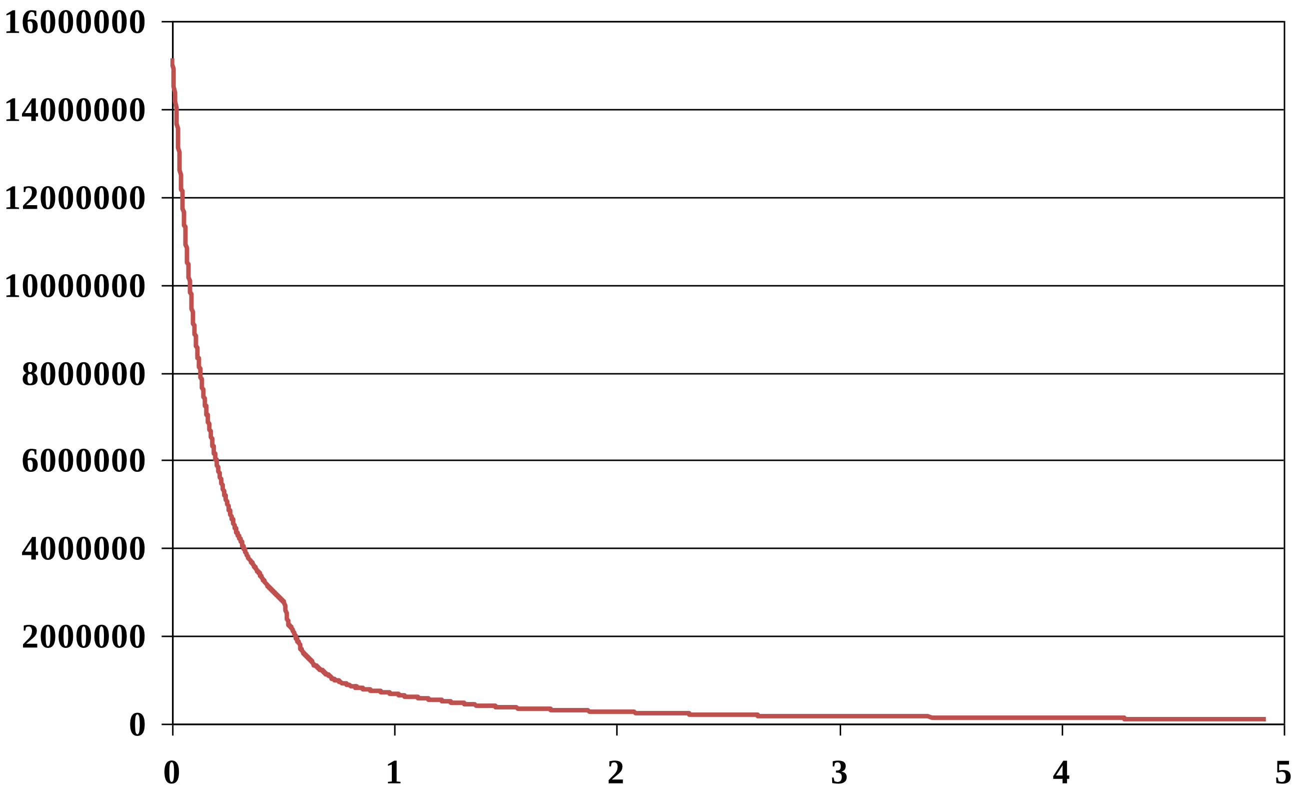


Figure 5. Cumulative Distributions of Owned Total Land and Land Plots (Acres), By Sample and Census: ICRISAT VLS 2014

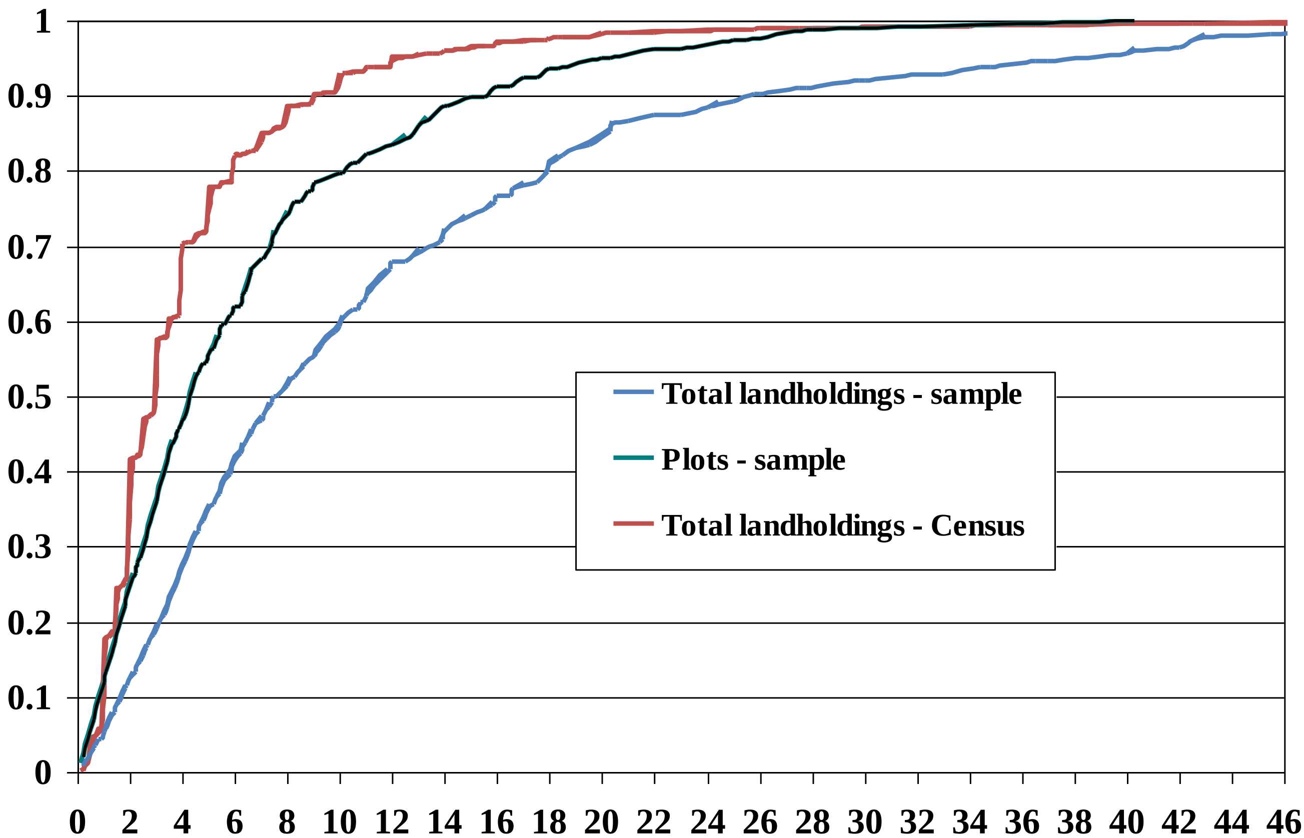


Figure 6. Relationship Between Real Average Profits per Acre and Farm Size (Acres) (ICRISAT VLS 2009-14)

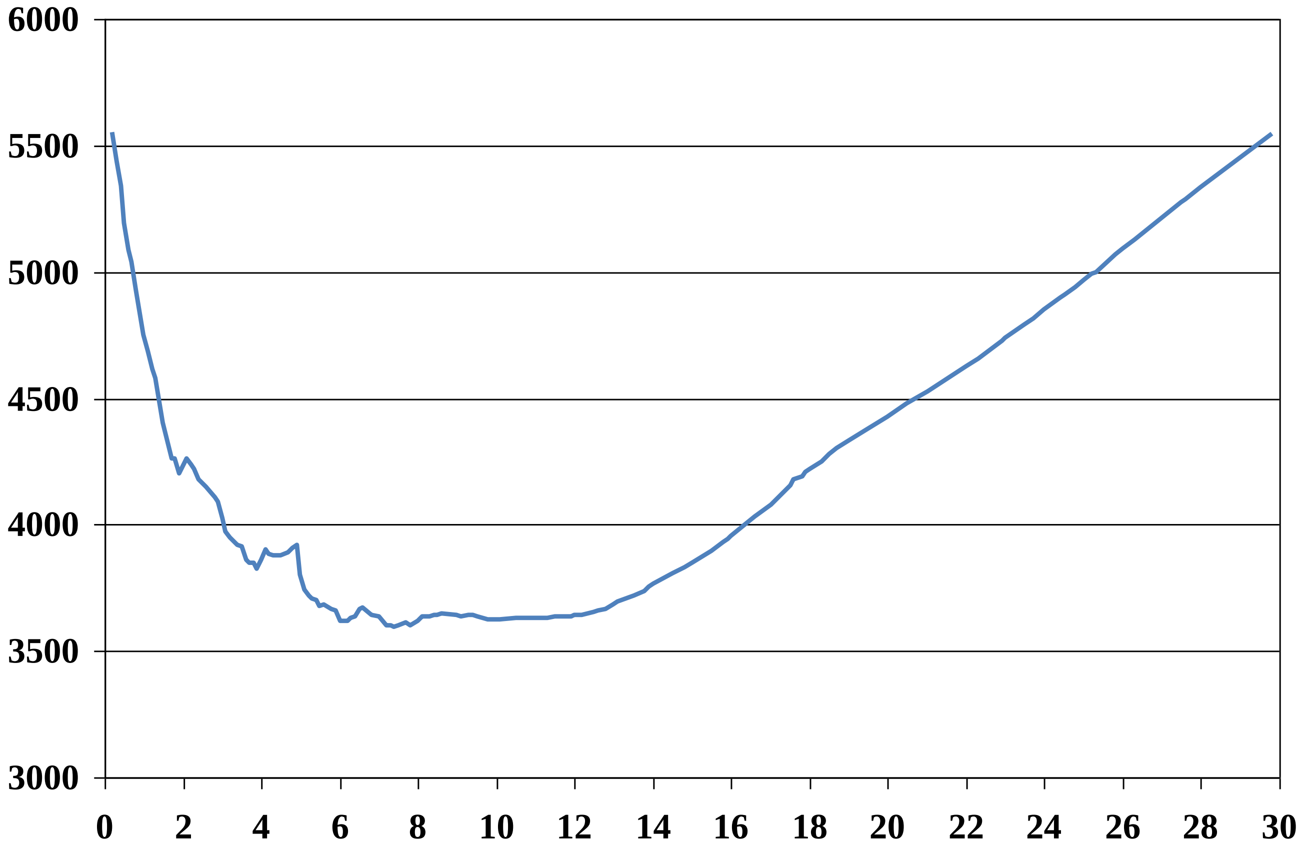


Figure 7. Measurement Error Effects: Profits and Owned Acreage Coefficient Point Estimates, by Farm Acreage and Estimation Procedure (ICRISAT VLS 2009)

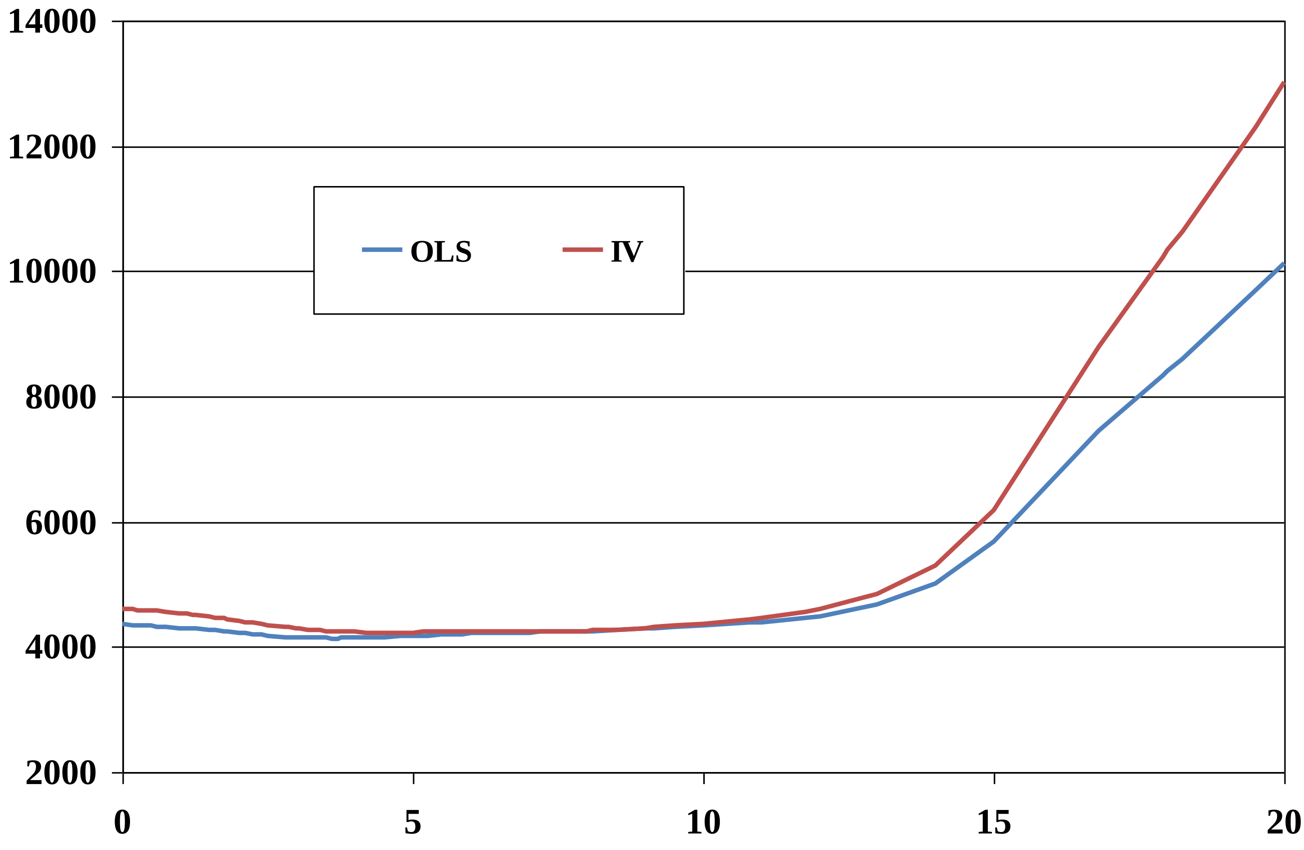


Figure 8. Real Profits per Acre by Owned Area: Roles of Plot Quality and Farmer Characteristics (ICRISAT VLS 2009-14)

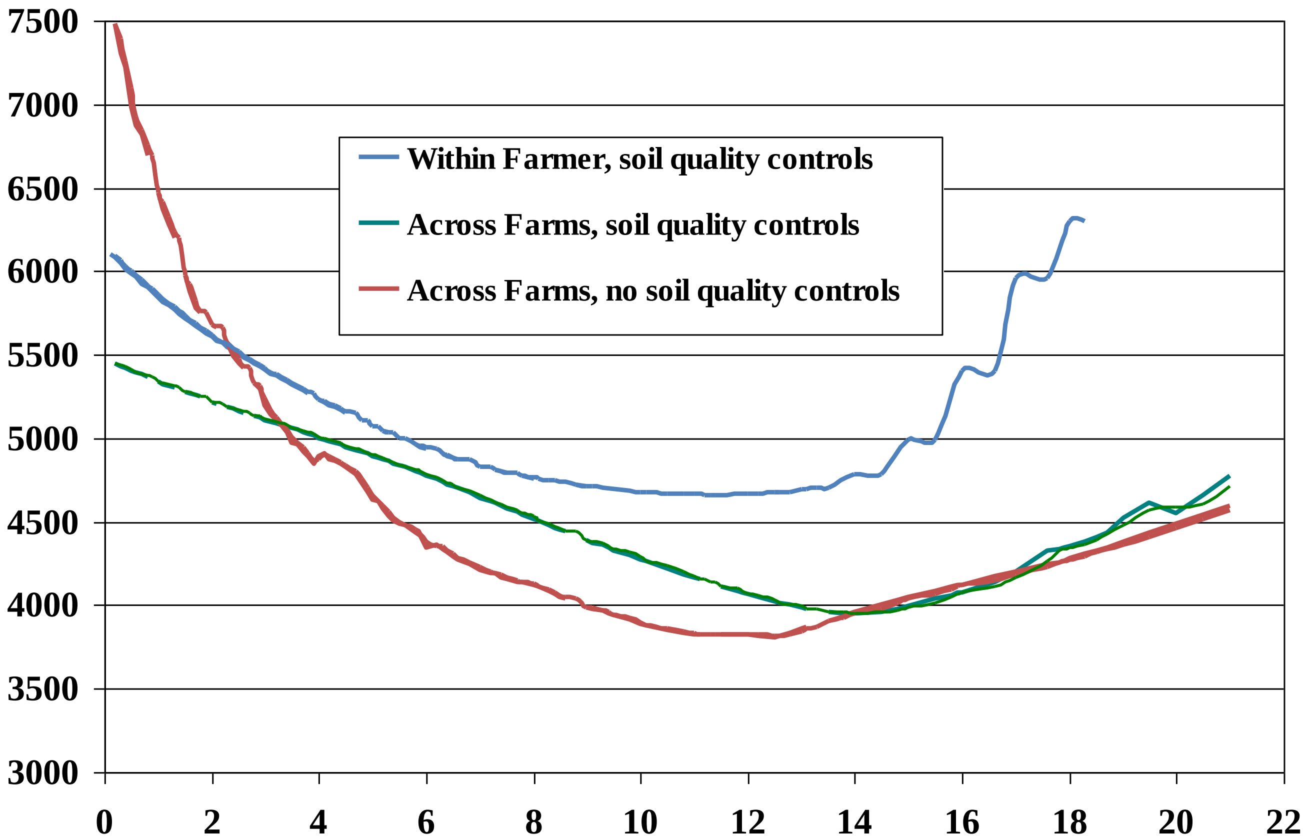


Figure 9. Cost per Horsepower for Electric Motors and Submersible Pumps by Horsepower (ICRISAT VLS Equipment Inventory, July 2011)

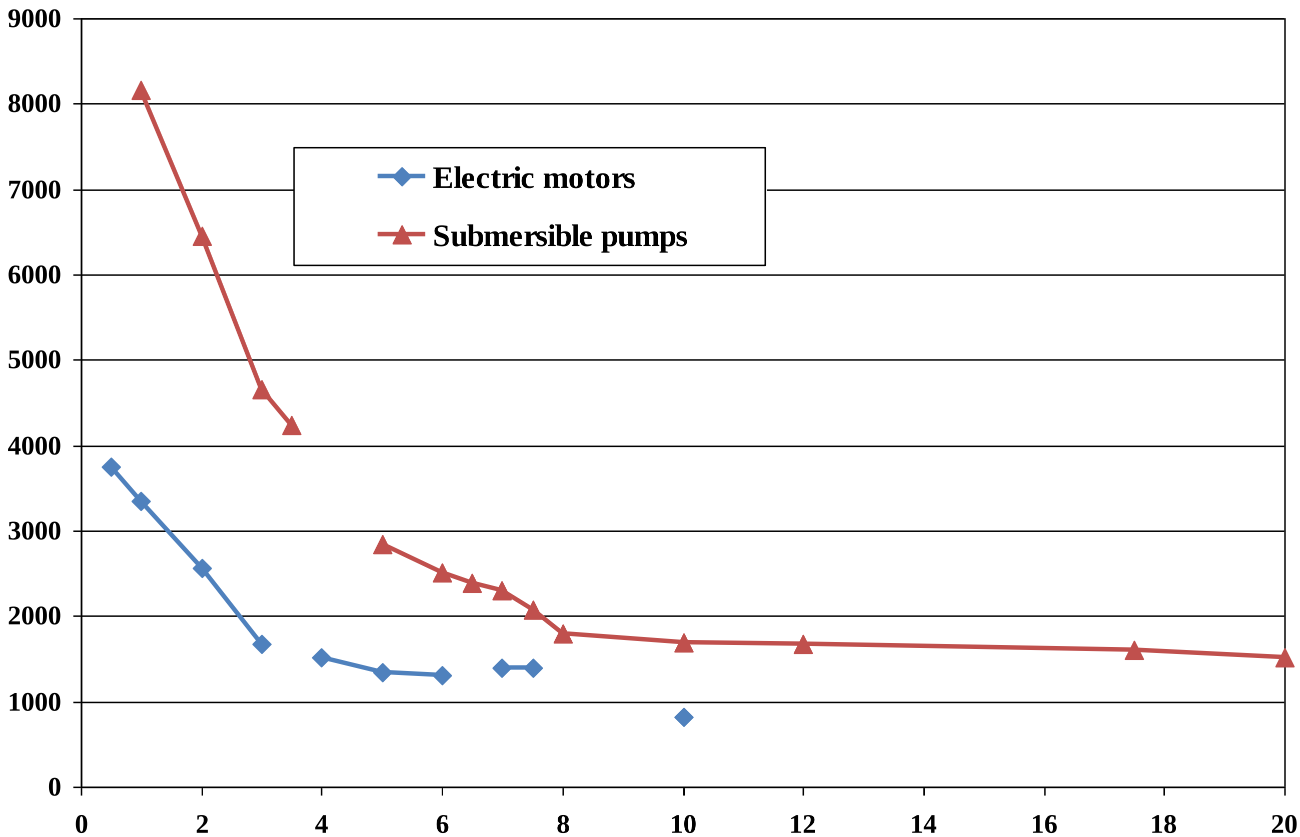


Figure 10: Profits per acre by farm-scale:
simulation based on the model with only labor

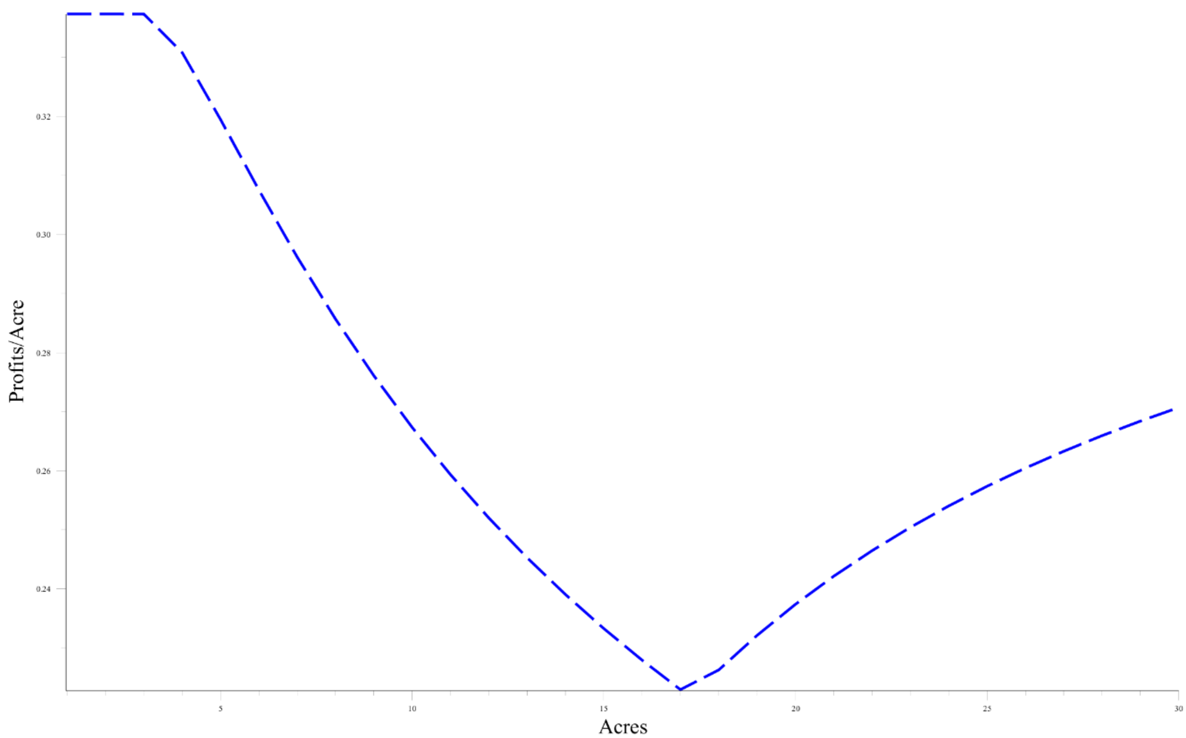


Figure 11: Input costs per acre by farm-scale:
simulation based on the model with only labor

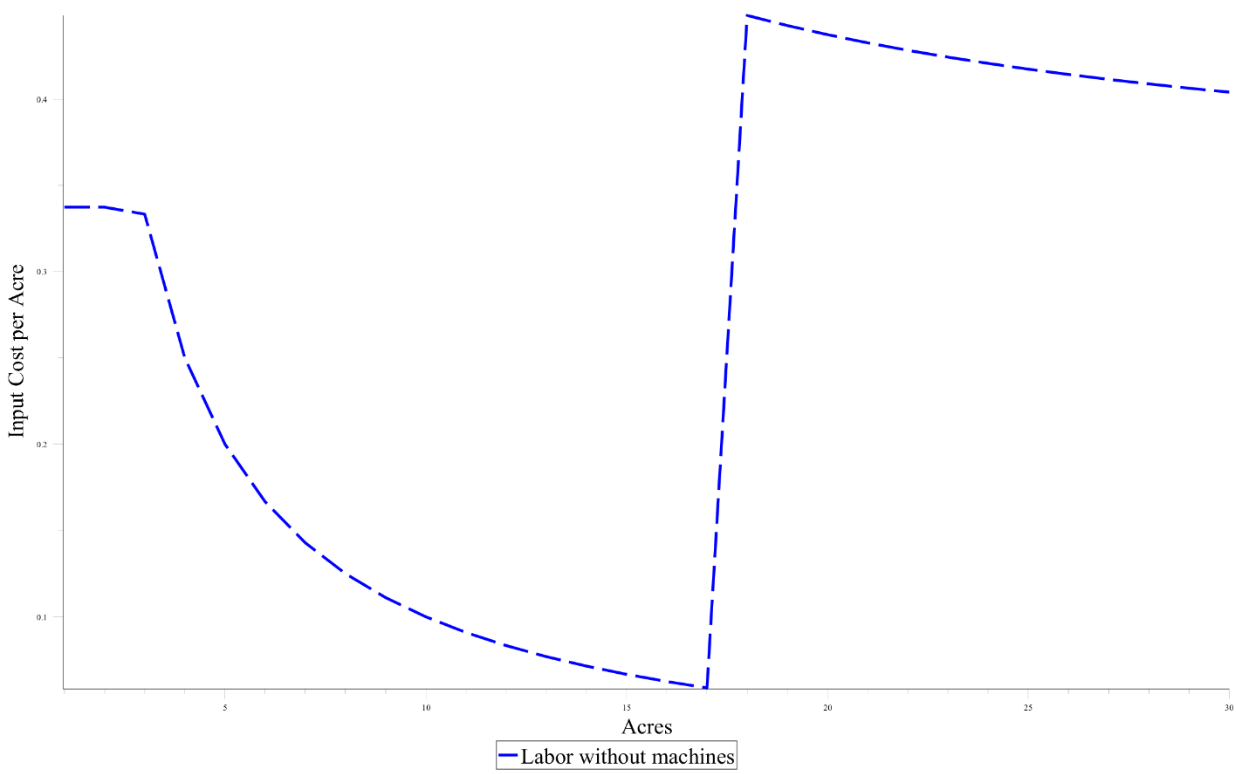


Figure 12: Average wage by farm-scale:
simulation based on the model with only labor

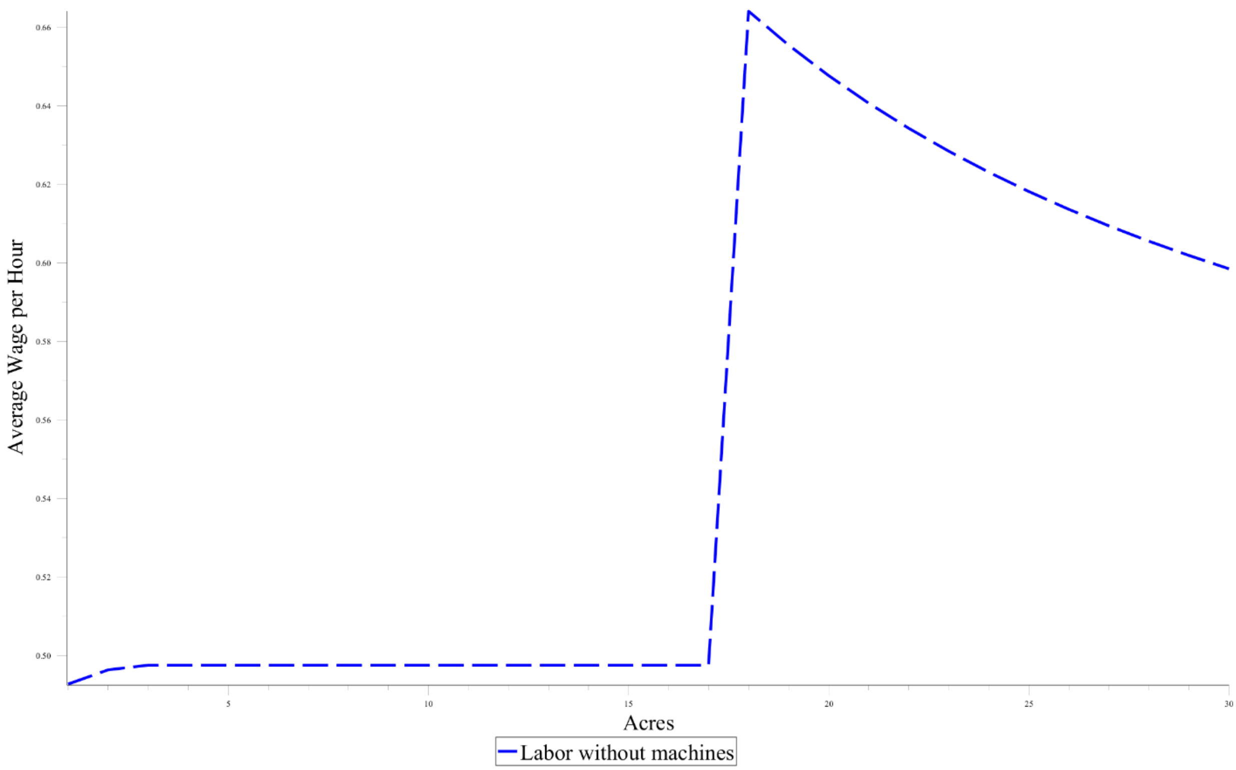


Figure 13: Profits per acre by farm-scale:
simulation based on the model with two labor stages

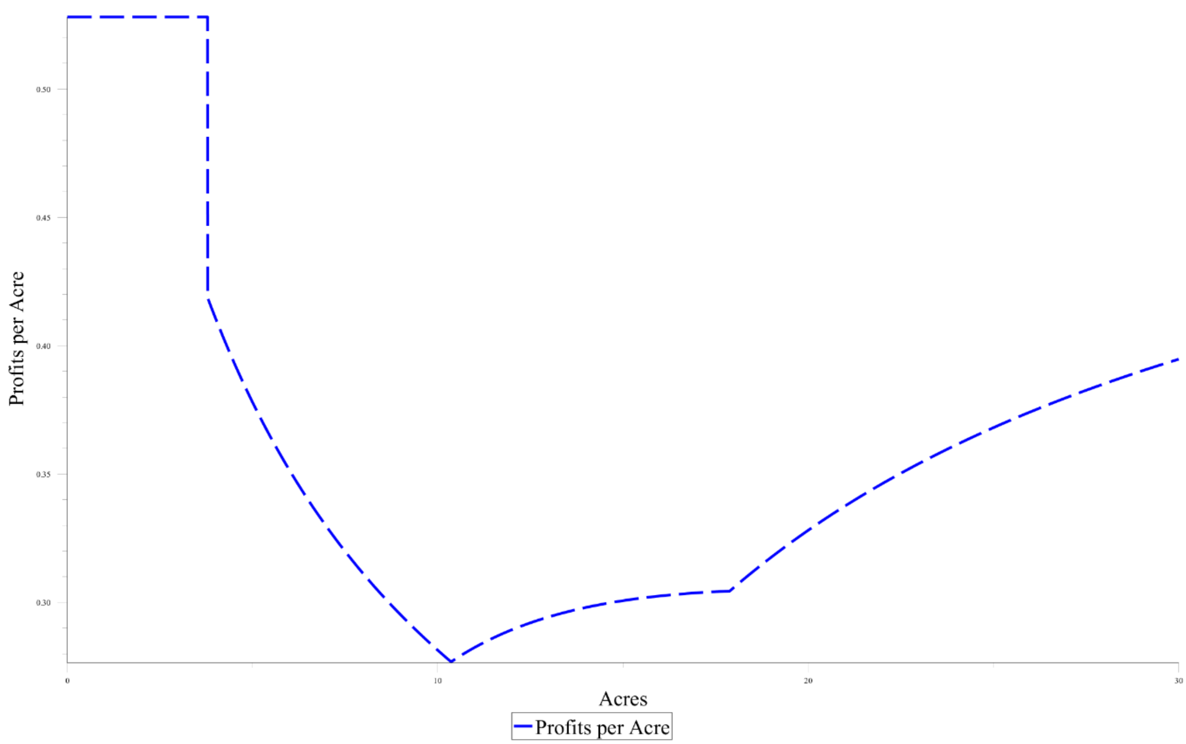


Figure 14: Profits per acre versus farm-scale:
simulation based on the model with and without machines

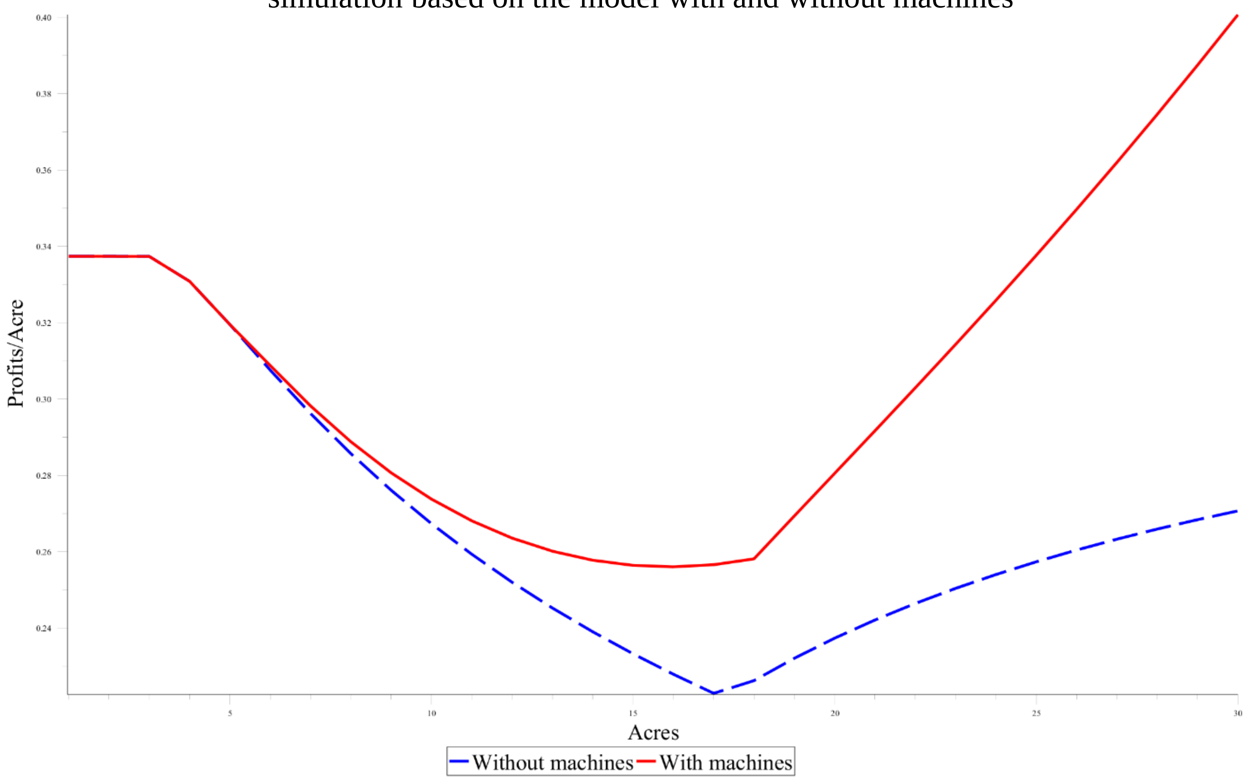


Figure 15: Labor and machine costs per acre versus farm-scale: simulation based on the model with and without machines

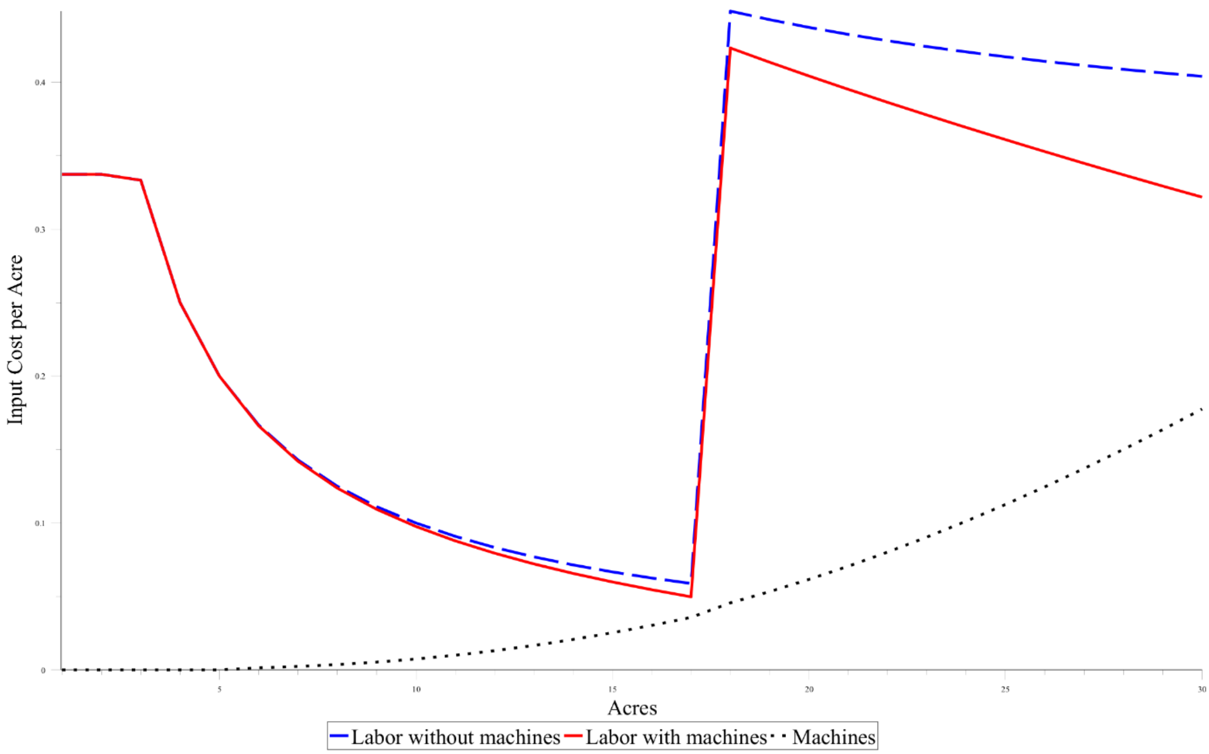


Figure 16. LWFCM Estimates of the Effects of Land Size on Profits with 95% CI, Net of Soil Quality and Time/Village Fixed Effects, by Farm Size (ICRISAT VLS, 2009-14)

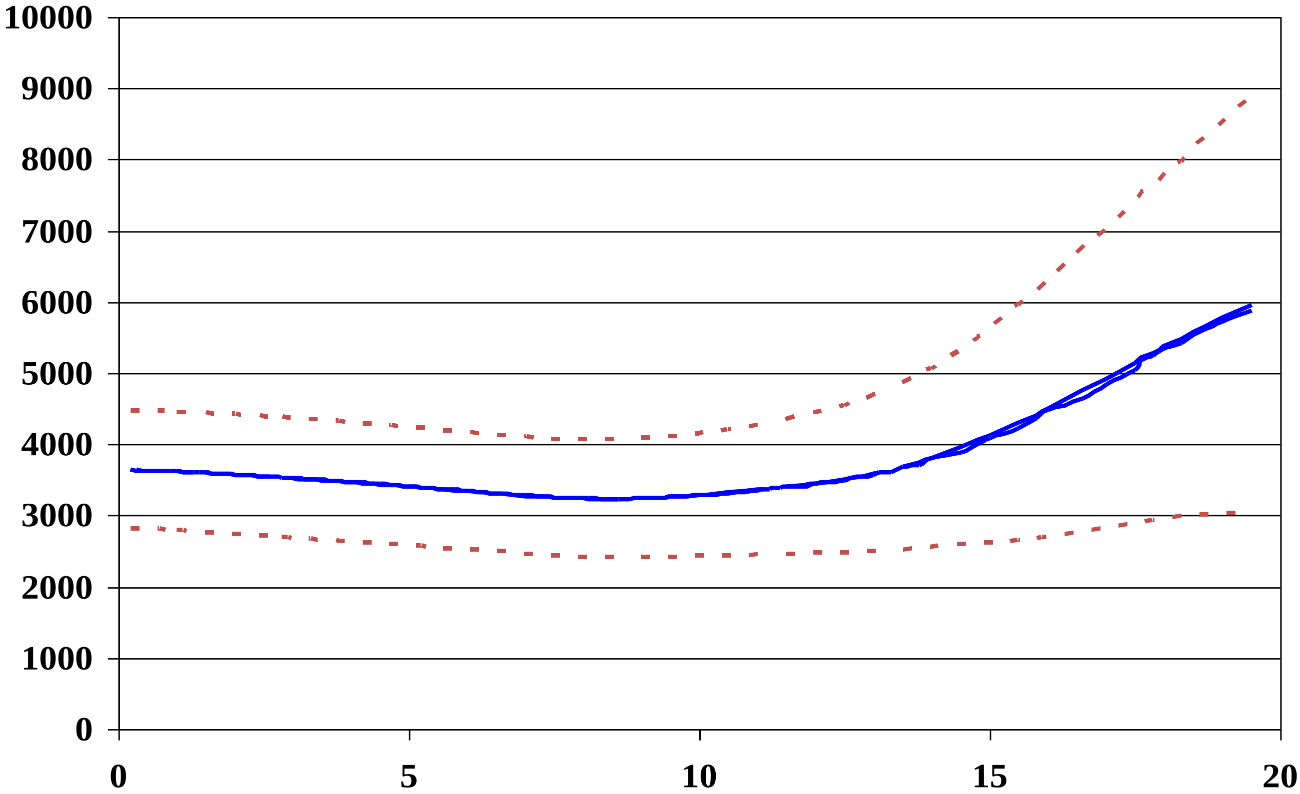


Figure 17. Average Days per Month Worked Off Farm, by Farm Size
Male and Female Workers Aged 21-59
(ICRISAT VLS 2014)

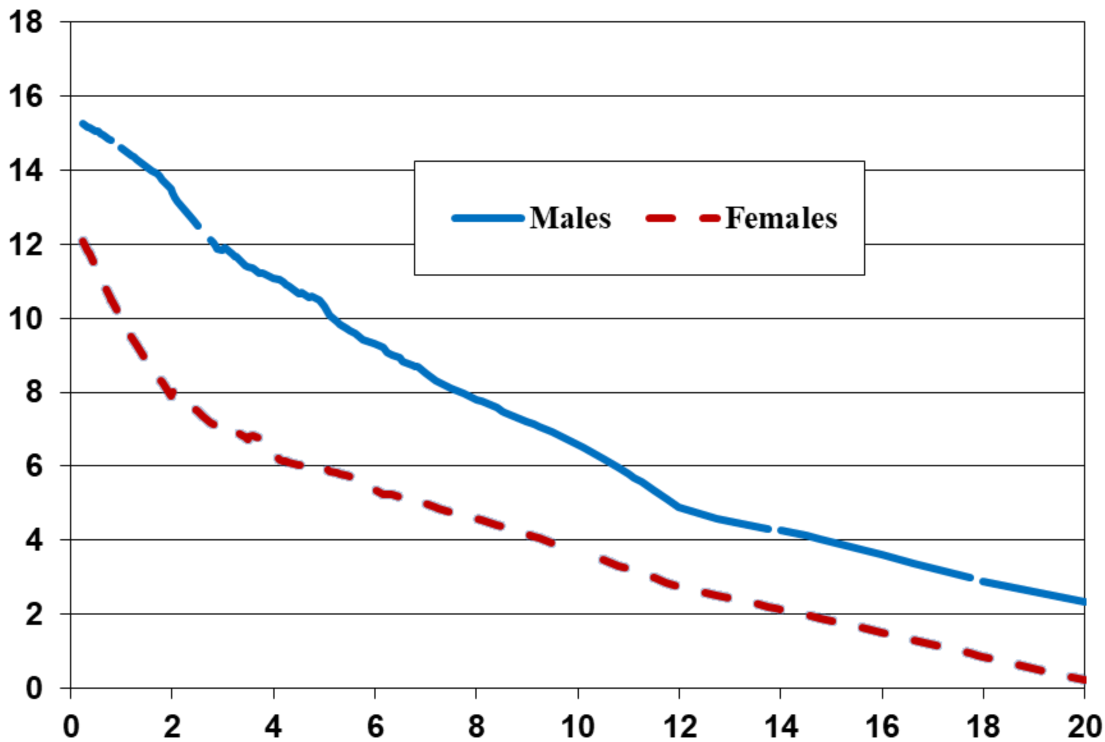


Figure 18. Average Hourly Wage Paid for Male Labor, by Farm Size (ICRISAT VLS 2009-14)

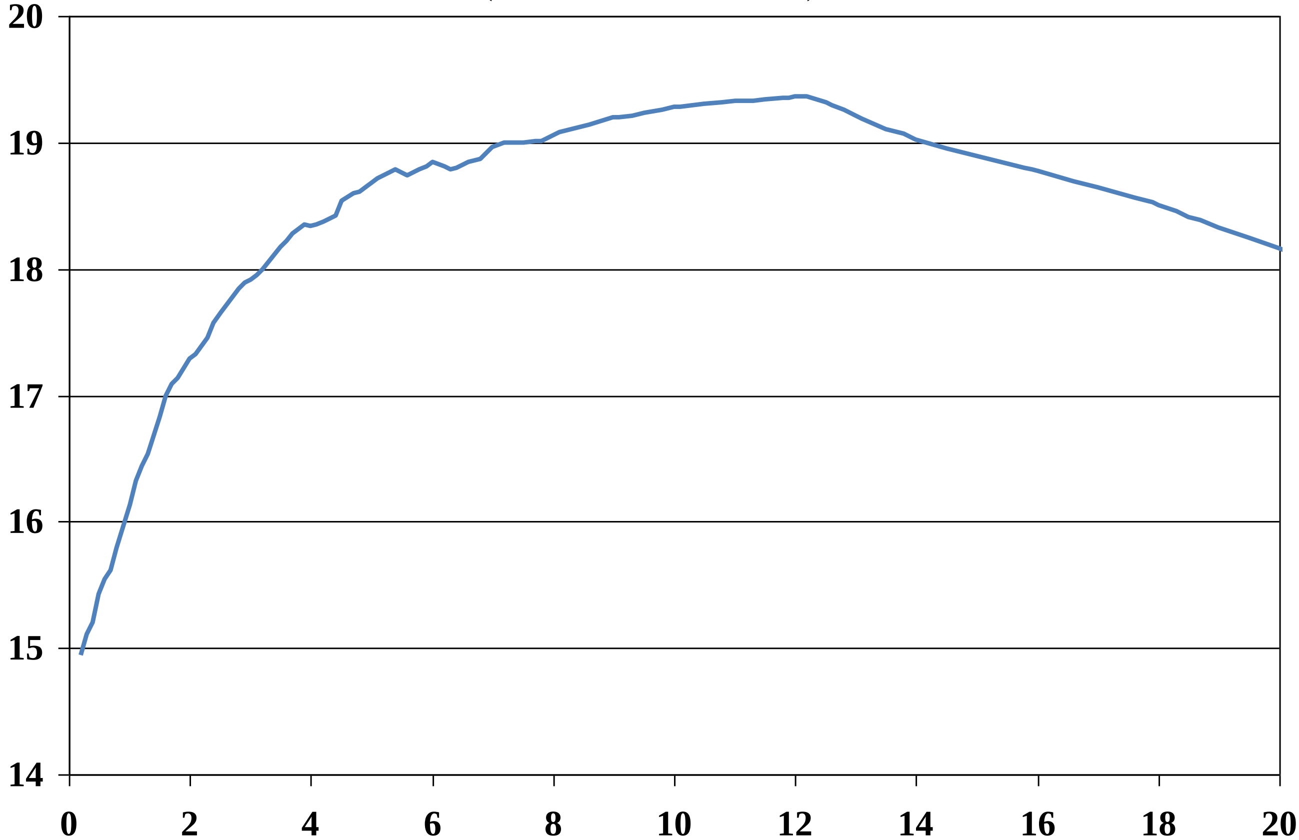


Figure 19. LWFCM Plot Fixed Effect Estimates:
The Effect of Rainfall on the Fraction of Operations Using Low-Hours Hired Male Labor, with 95% CI, by Plot Size (ICRISAT VLS 2009-14)

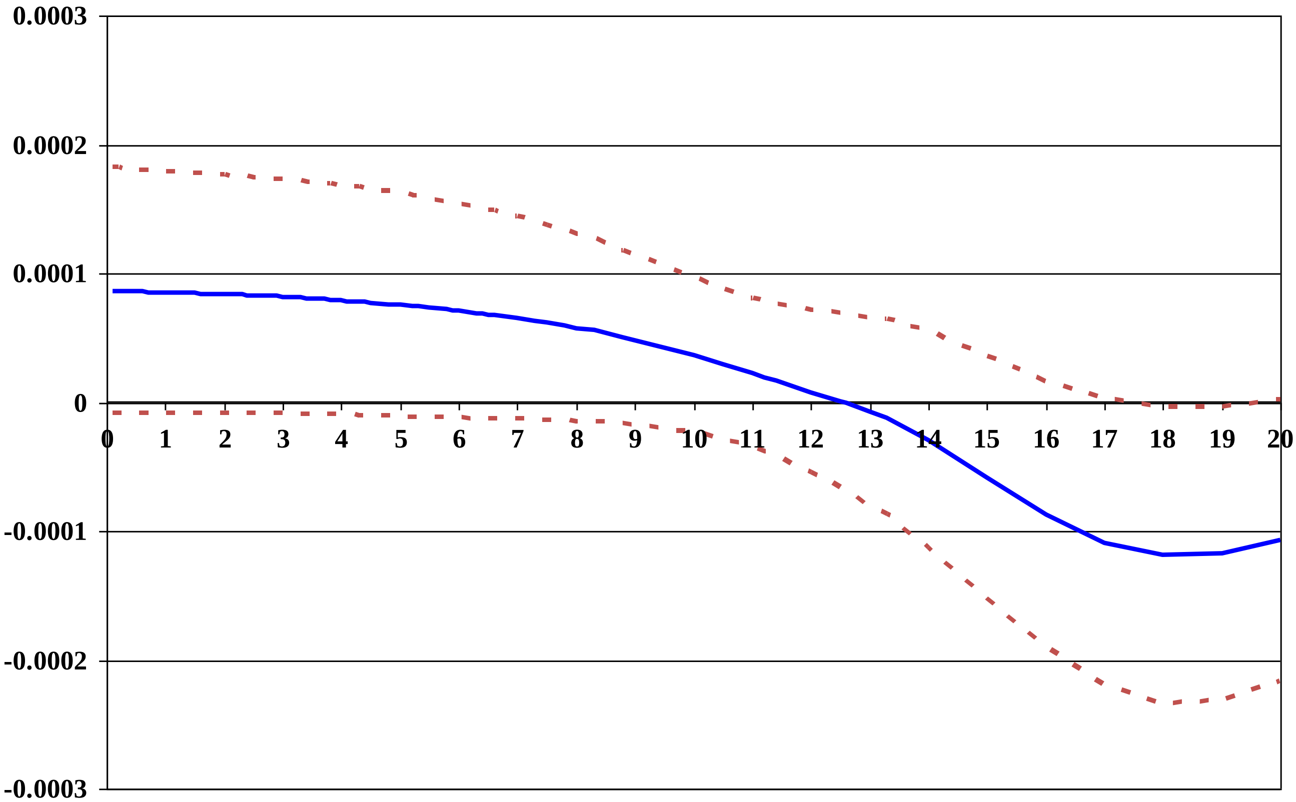


Figure 20. LWFCM Plot Fixed-Effect Estimates:
the Effect of Rainfall on the Average Male Wage, with 95% CI, by Plot Size
(ICRISAT 2009-14)

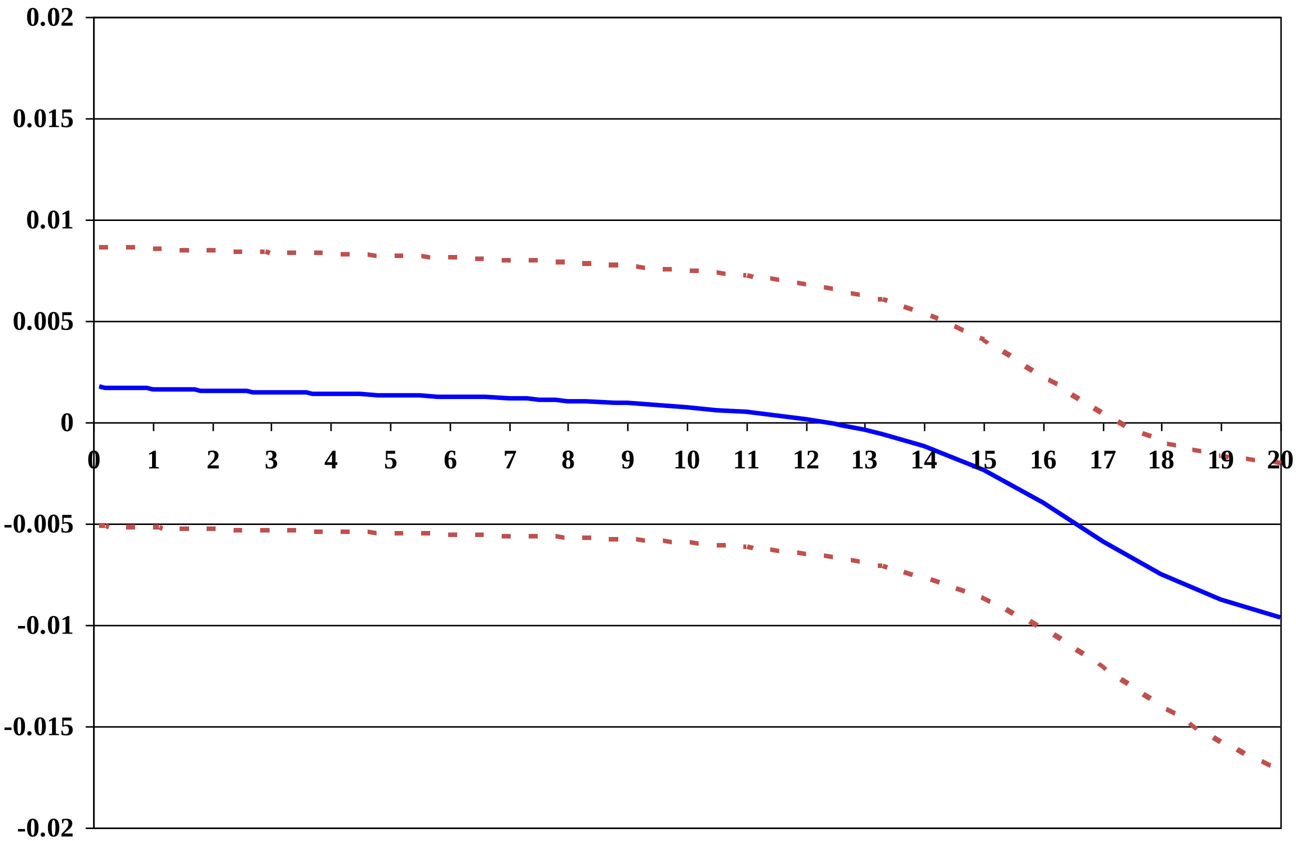


Figure 21. Fraction of Farms using Sprayers and Tractors, by Farm Size (ICRISAT VLS 2009-14)

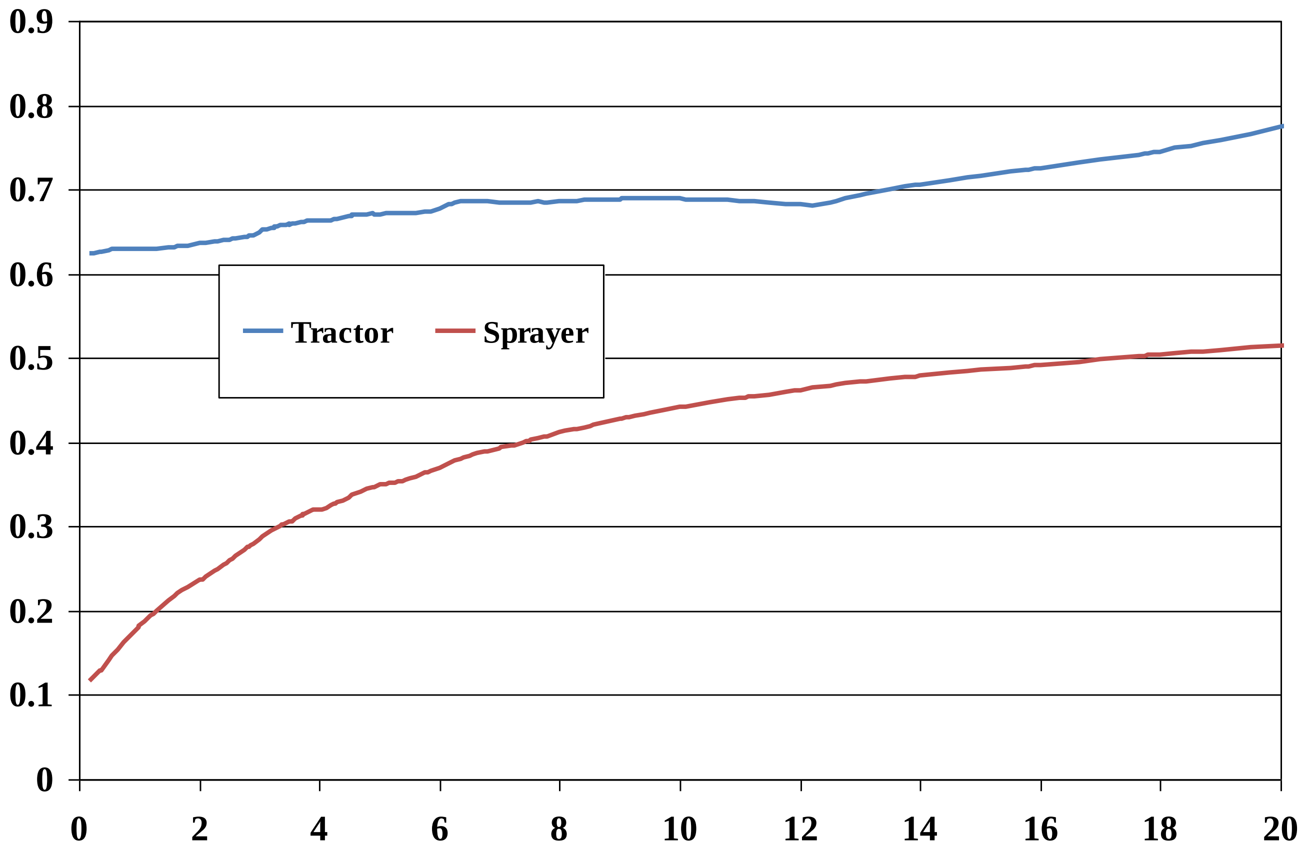


Figure 22. Per-Acre Equipment Hours for Tractors and Sprayers, by Farm Size (ICRISAT VLS 2009-14)

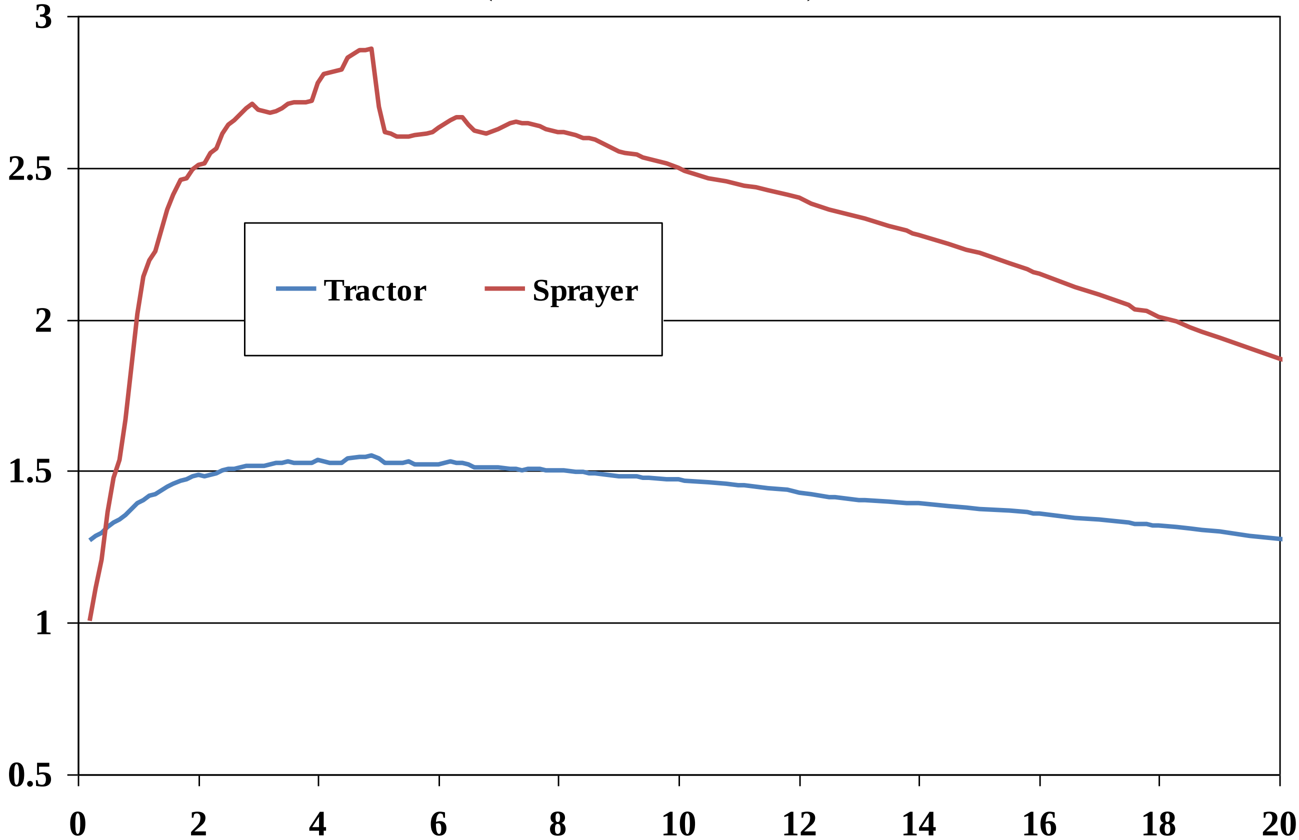
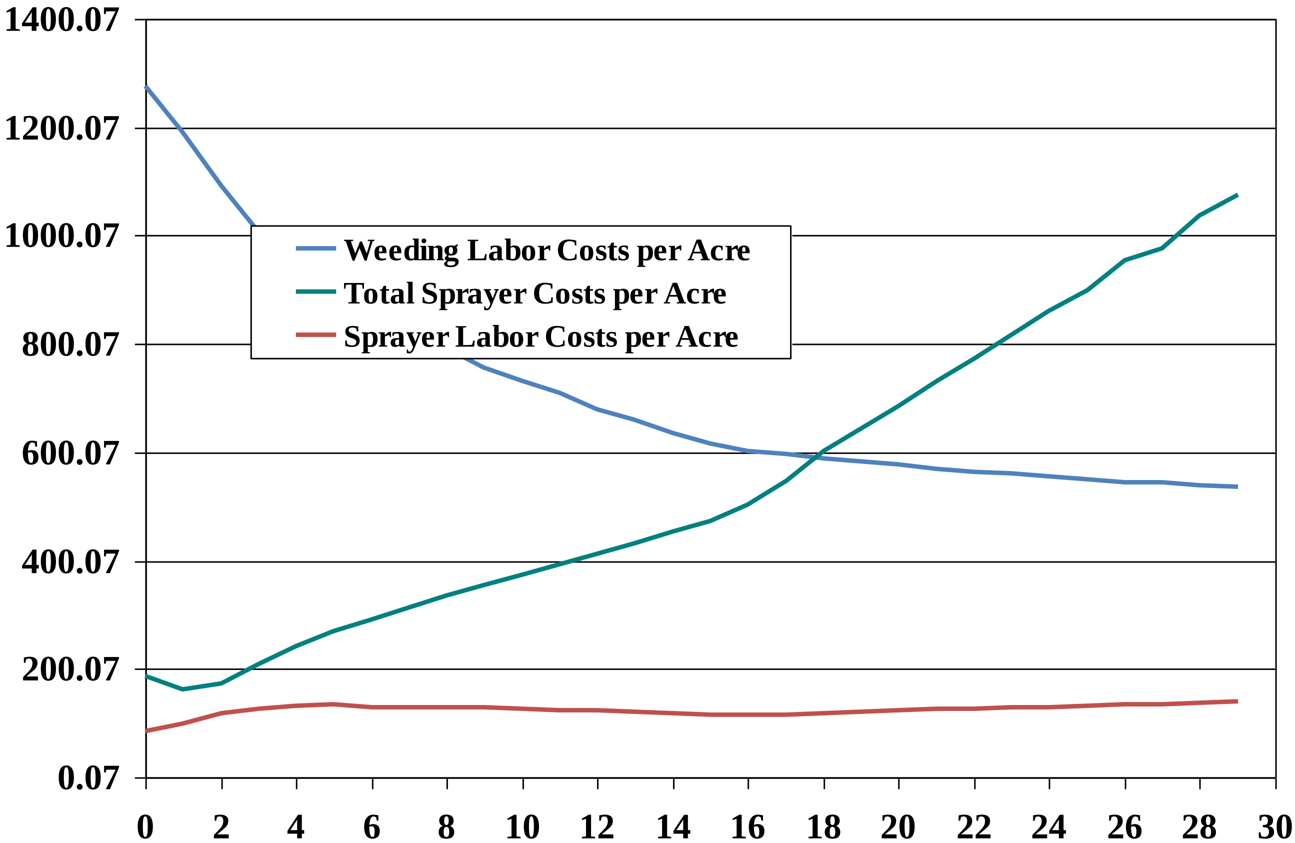
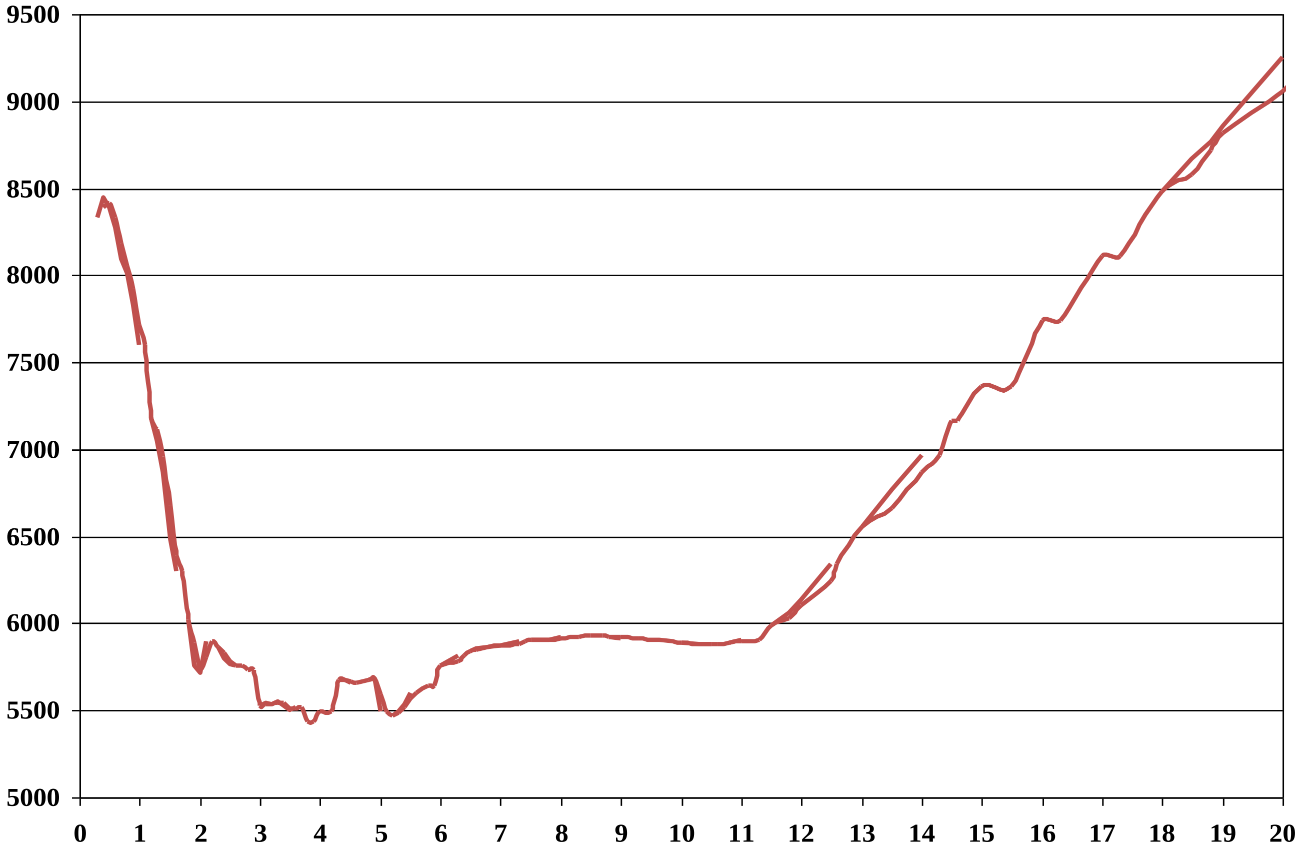


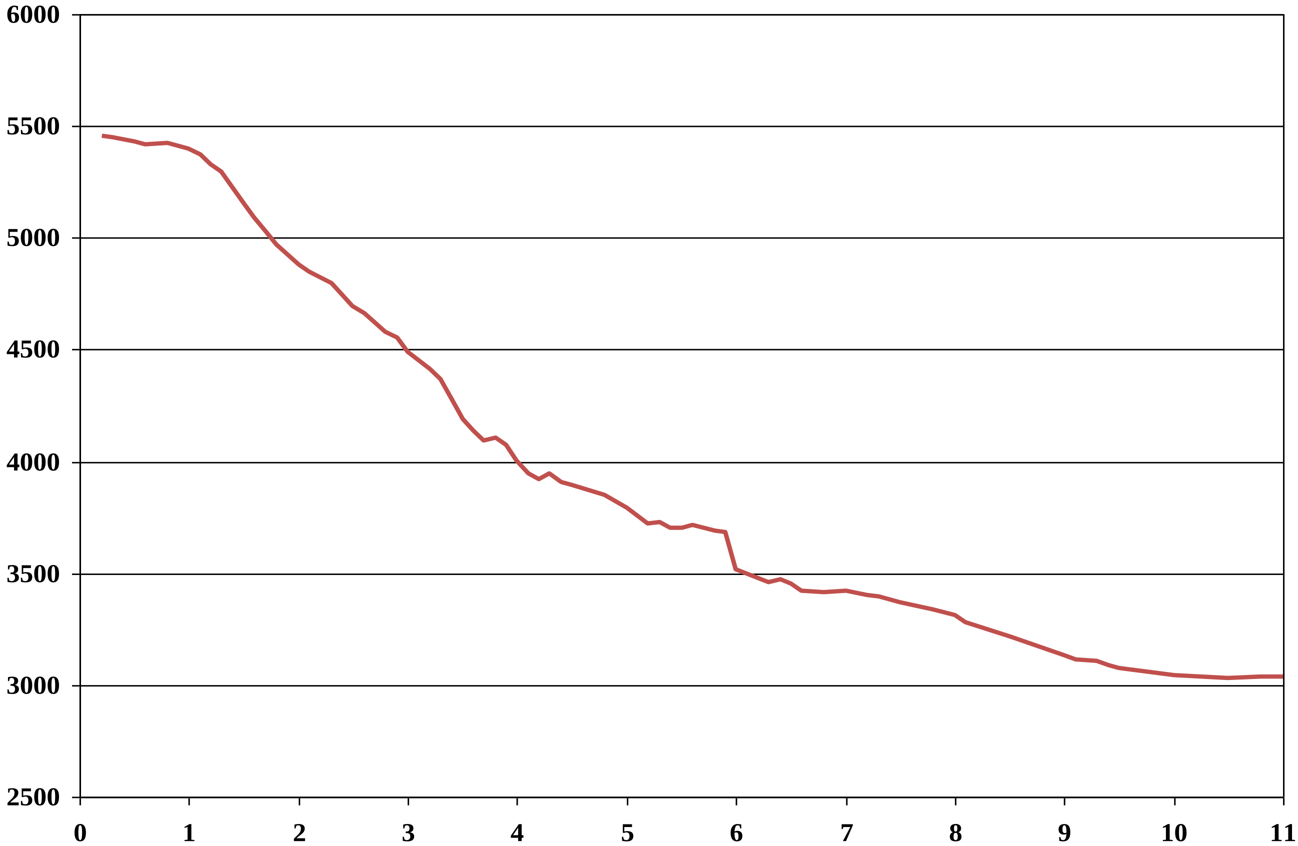
Figure 23. Real Weeding and Sprayer Labor Costs and Total Sprayer Costs per Acre, by Farm Size (ICRISAT VLS 2009-14)



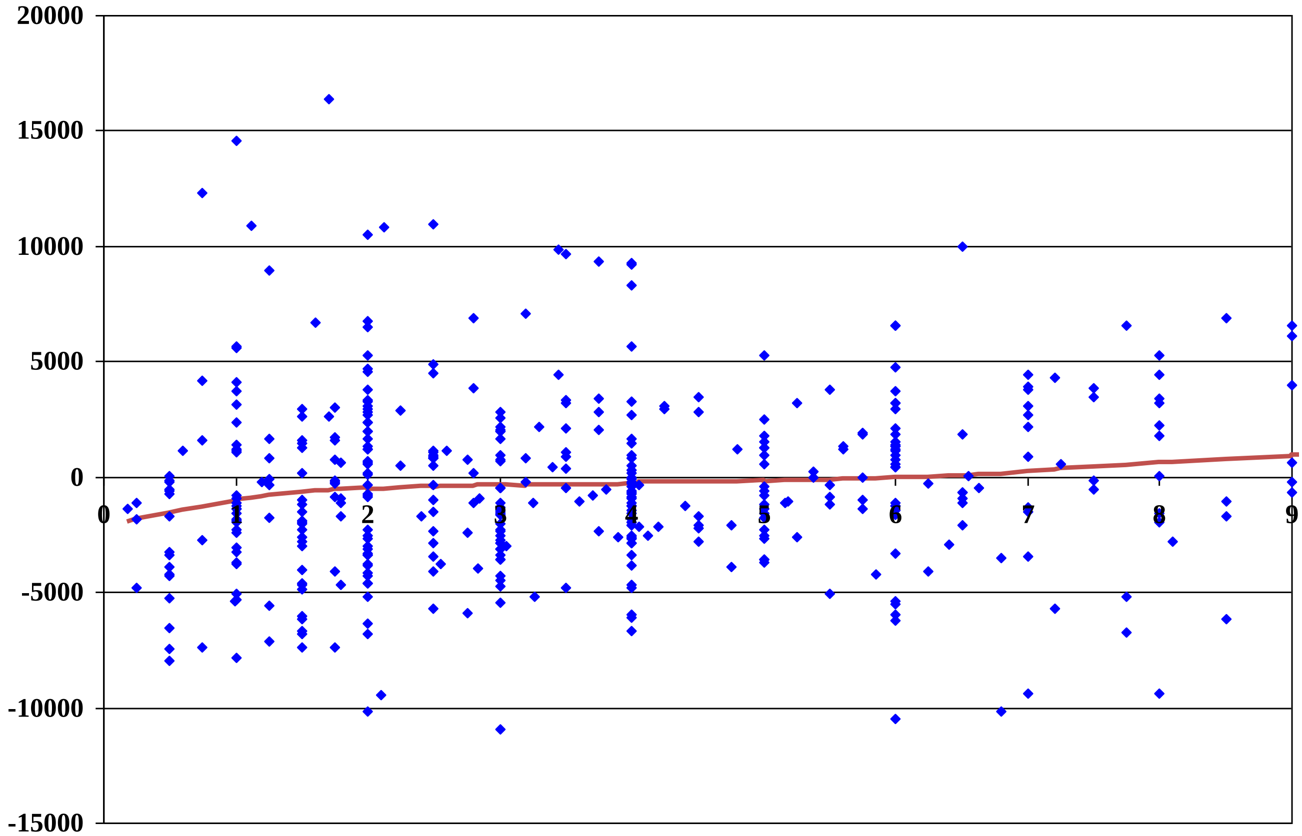
Appendix Figure A1. Lowess-Smoothed Relationship of Profits per Acre and Owned Plot Size, Rice Plots Only (ICRISAT Survey, 2009-14)



Appendix Figure A2. Lowess-Smoothed Relationship of Profits per Acre and Owned Area
(ICRISAT VLS Survey, 2009)



Appendix Figure A3. Average Per-Acre Profit Residuals and Farm Size, With Lowess-smoothed Average (ICRISAT VLS 2009)



Appendix Figure A4. LWFCM Estimates of Predicted Per-Acre Profit Residuals on Farm Size, with 95% CI, by Farm Size (ICRISAT VLS 2009)

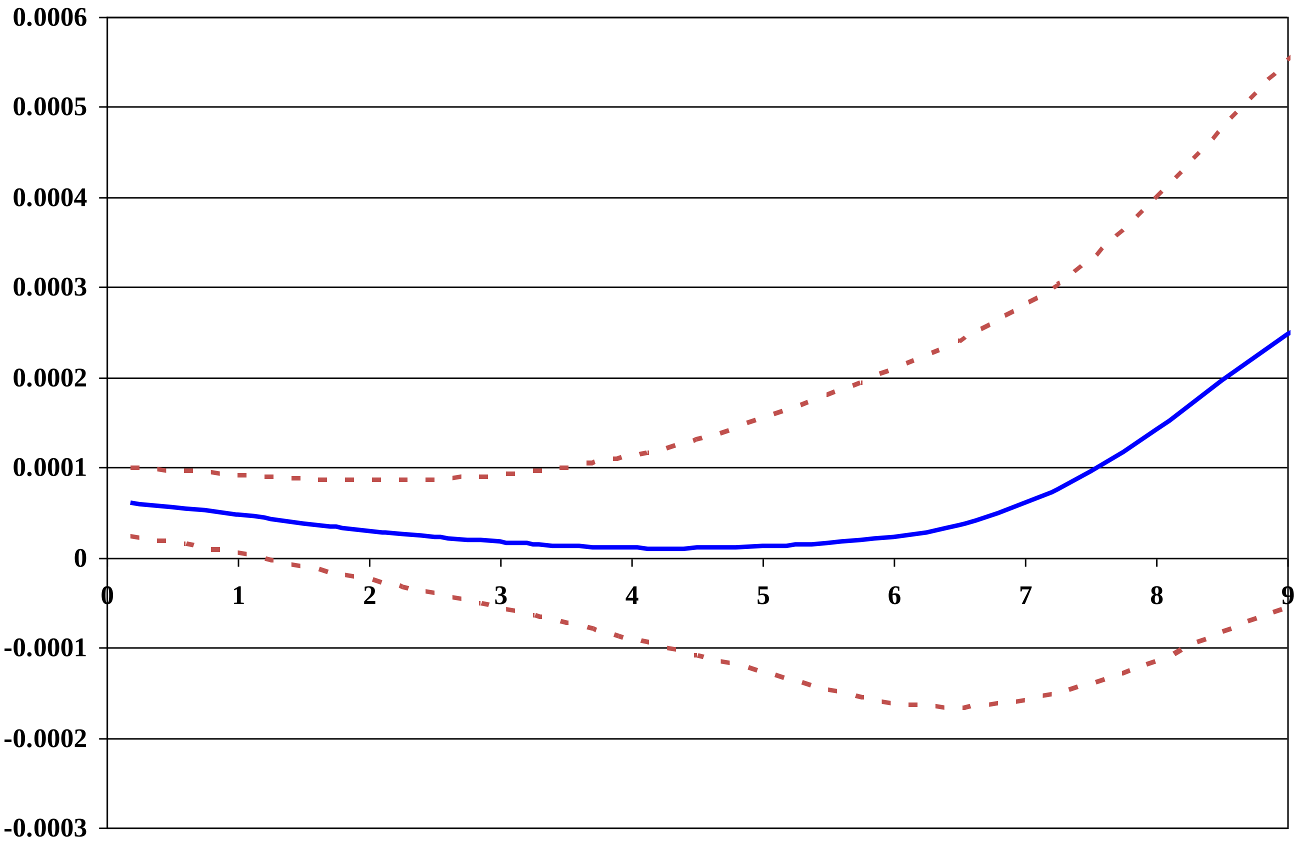


Table 1
Measurement Error Test: The Effect of Total Owned Area on Profits,
by Estimation Procedure, Using the Survey and Census Owned Land Self-reports in 2009

Variable	Profits		Owned Land - Survey
Estimation Procedure	OLS	IV	OLS
Total owned land - survey, sum of plot sizes	4324.2 (1305.8)	4433.5 (1328.1)	-
Owned land - Census, sum of irrigated and non-irrigated land	-	-	0.9637 (0.0128)
Village fixed effects	Y	Y	Y
N	306	306	306

Standard errors in parentheses clustered at the village level. Profits are for the *Kharif* season.

Table 2
Does Land Quality Affect Estimates of the Relationship Between Real Output Value and Profits and Area?
(*Kharif* Seasons 2009-14)

Variable	Real Output Value				Real Profits			
Farm size (acres)	9138.9 (676.3)	9053.4 (671.2)	-	-	3911.8 (501.4)	3871.9 (500.7)	-	-
Plot size (acres)	-	-	7390.6 (1463.8)	7548.6 (1506.0)	-	-	5841.8 (1474.0)	5939.5 (1521.7)
Village/year FE	Y	Y	-	-	N	Y	-	-
24 plot characteristics/categories	N	Y	N	Y	N	N	N	Y
Farmer/year fixed effects	N	N	Y	Y	N	N	Y	Y
H_0: Plot and household characteristics = 0 F(24,107) [p]	-	8.45 [.0000]	-	1.22 [.239]	-	3.61 [.0000]	-	1.22 [.245]
Number of observations	3,835	3,835	7,865	7,865	3,835	3,835	7,865	7,865

Standard errors in parentheses clustered at the village/year level.

Table 3
Operation Fixed Effects Estimates:
the Percentage Difference in Hourly Wage Rates Paid for Eight Hours
versus Less than Eight Hours of Work, by Input
(2010 and 2011 ICRISAT Monthly Price Schedules)

Variable	Hired Male Labor	Hired Bullock Pair + Driver
Worked eight hours in the day versus <8	-0.332 (0.0314)	-0.223 (0.0454)
Mean wage (rupees)	22.1 (9.34)	78.7 (39.6)
Percent working <8 hours	30.7	58.4
N	729	450

Standard errors in parentheses. Hourly wage rates constructed by dividing reported daily wages by the report of hours worked.

Table 4
Log Total Labor Costs per Acre and Area: Cross-Farm and Within-Farm Estimates
(*Kharif* Seasons 2009-14)

Variable	(1)	(2)	(3)	(4)
Farm size (acres)	-0.0576 (0.00750)	-0.0388 (0.00486)	-	-
Plot size (acres)	-	-	-0.0497 (0.0172)	-0.0437 (0.0168)
Village/year FE	Y	Y	-	-
24 plot characteristics/categories	N	Y	N	Y
Farmer/year fixed effects	N	N	Y	Y
H_0: Plot and household characteristics = 0 F(24,107) [p]	-	3.99 [.0000]	-	8.17 [.0000]
Number of observations	3,177	3,177	5,786	5.786

Standard errors in parentheses clustered at the village/year level.

Table 5
Plot Size and Fraction of Operations that Employ Hired Inputs at Low (<=6) Daily Hours and the Average Hourly Wage Paid, by Input Type (*Kharif* Seasons 2009-14)

Variable	Fraction of Operations <6 Hours/Day			Average Hourly Wage		
Input type	Hired Male Labor	Hired Tractor	Hired Bullock Pair	Hired Male Labor	Hired Tractor	Hired Bullock Pair
Plot size (acres)	-.0165	-.0197	-.0170	-.183	1.25	-.866
	(.00306)	(.00247)	(.00306)	(.0876)	(.769)	(.306)
Plot size squared x10^{-3}	.450	.449	.555	8.29	18.3	29.3
	(.112)	(.0682)	(.117)	(3.23)	(32.4)	(10.9)
Village/year FE	Y	Y	Y	Y	Y	Y
25 Plot and household characteristics	Y	Y	Y	Y	Y	Y
Number of observations	6,777	6,777	6,777	6,777	6,777	6,777

Standard errors in parentheses clustered at the village/year level.

Table 6
Plot Fixed Effects Estimates: The Effects of *Kharif*-Season Rainfall on Profits, Hours Employed and Average Hourly Wage Rates, by Input Type (*Kharif* Seasons, 2009-14)

Variable	Profits	Hours Employed			Average Hourly Wage		
Input type	-	Hired Male Labor	Hired Tractor	Hired Bullock Pair	Hired Male Labor	Hired Tractor	Hired Bullock Pair
Rainfall (mm)	38.1 (17.1)	.182 (.0701)	.00362 (.00316)	.0347 (.0248)	-.0158 (.00672)	.0130 (.0601)	-.0593 (.0355)
Rainfall squared x10^{-3}	-21.2 (8.59)	-.107 (.0377)	-.00214 (.00161)	-.0500 (.0268)	.00778 (.00398)	-.0132 (.0282)	.0757 (.0331)
Year and plot FE	Y	Y	Y	Y	Y	Y	Y
H_0: Rain and rain squared = 0 F(2,n) [p]	3.09 [.0504]	4.18 [.0183]	0.99 [.3742]	1.97 [.1452]	3.47 [.0352]	0.28 [.7589]	3.02 [.0538]
Number of observations	5,291	3,987	4,016	2,523	3,987	4,016	2,523

Standard errors in parentheses clustered at the village/year level.

Table 7
Farm Size, Wealth and Mechanization (Ownership): 2014 ICRISAT VLS Round

Variable	Owns a Tractor	Owns a Power Sprayer	
Sample	All Farmers	All Farmers	Farmers Who Own Any Sprayer
Total owned land (acres)	.0125 (.00415)	.0107 (.00474)	.0133 (.00494)
Total rental value of land (wealth) x 10^{-5}	.0506 (.0146)	.0512 (.0166)	.0273 (.0144)
Village FE	Y	Y	Y
Percent owning	3.5	10.3	24.8
Number of farmers	652	652	288

Standard errors in parentheses clustered at the village level. All specifications include the head's age and schooling.

Table 8
Cost and Capacities of Indian KrisanKraft Power Sprayers, 2017

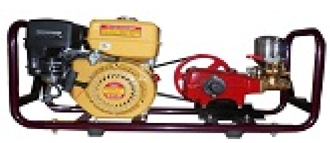
Power sprayer	Litres/Hour	Current Price (Rupees)
	180	7830
	420	12260
	1320	25900
	2400	27900

Table 9
Estimates of the Effects of Owned Land Size
on Sprayer Use, Weeding Hours per Acre, Sprayer Hours per Acre, Log Sprayer Price per hour, and Sprayer Flow Rate

Variable	Any sprayer use	Weeding hours per acre	Sprayer hours per acre	Sprayer log price per hour	Sprayer flow rate
Estimation procedure	OLS	OLS	OLS	OLS	OLS
Owned area	0.006197 (0.0009879)	-0.5631 (0.1286)	-0.4063 (0.0853)	0.01335 (0.00669)	0.01360 (0.00667)
All land characteristics	Y	Y	Y	Y	Y
Village/year fixed effects	Y	Y	Y	Y	Y
N	3,374	3,374	1,219	1,219	1,219

Standard errors in parentheses clustered at the village/year level.

Table 10
GMM Estimates of the Effective Capacity Function $\varphi(a)$ and Price Parameter υ

Coefficient	Point Estimate	Robust SE
υ	0.316	0.124
b_0	5.58	0.0375
b_1	0.933	0.0343
b_2	-0.0190	0.00211
H_0: $\upsilon < 1$, $\chi^2(1)$ [p]	30.4 [.0000]	
Maximum land size (acres) = $\varphi(a)' = -b_1/(2*b_2) = 0$	24.5	1.84
N	617	

Instruments: owned land area and land area squared.

Table 11
Estimates of Sprayer *v*, by Source

Country	India		United States
Source	ICRISAT Survey, 2009-2014	KrisanKraft Price List (2016)	Stiles and Stark (2016)
Estimation procedure	IV[a]	OLS	OLS
v	0.5802 (0.1200)	0.5209 (0.0605)	0.1458 (0.0789)
H_0: $v = 1$, F(1,x) [p]	χ^2=12.2 [0.0005]	F(1,2)= 62.8 [0.0156]	F(1,2)=117.1 [0.0084]
N	1,219	4	4
Village/year fixed effects	Y	N	N

[a]First-stage includes log of owned area and all land quality characteristics. Standard error clustered at the village/year level.

Table 12
Elasticities for Changes in Area, υ and Wage Rates on Sprayer Capacity (q), Sprayer Hours (m) and Weeding Labor Hours (l) for a Farm of Median Size (3 acres), from the Calibration and GMM Estimates

Coefficient	Point Estimate	Robust SE
dq/da	0.297	0.0124
dq/dυ	-0.0498	0.0728
dm/dυ	-0.233	0.113
dl/dυ	0.0299	0.130
dq/dw	0.0292	0.0399
dm/dw	-3.16	0.977
dl/dw	-3.77	0.112

Appendix Table A1
Mean Standard Deviations (Days) of First Operation Dates Across Plots
Within and Across Farmers, by Operation
(*Kharif* Season 2014)

Operation/sample	Cross-Plot within Farmer[a]	Cross-Farmer within Village
Sowing	9.55	14.8
	(14.7)	(10.9)
	N=713	N=946
Fertilizing	20.2	24.8
	(19.7)	(14.2)
	N=464	N=647
Weeding	16.6	19.3
	(15.9)	(9.8)
	N=737	N=984
Spraying	17.7	22.4
	(19.5)	(12.0)
	N=447	N=664
Threshing	13.7	18.0
	(13.5)	(8.31)
	N=468	(678
Harvesting	21.8	24.9
	(19.7)	(12.4)
	N=914	N=1,133

[a]Farmers with two or more plots. Standard errors in parentheses.

Appendix Table A2
Test of Farmer Ability Bias: Profit Function Estimates for Farm Size<11 Acres
(*Kharif* Season 2009)

Variable/Estimations Procedure	OLS	IV	IV
Owned cultivated land (acres)[a]	-725.7 (302.1)	-1284.9 (636.4)	-1254.5 (656.6)
Rental value of owned cultivated land[a]	0.1341 (0.0644)	0.2245 (0.0853)	0.2102 (0.1167)
Total wealth[a]	0.00596 (0.00298)	0.00763 (0.00624)	0.00720 (0.00667)
Total value of inheritance	-	-	0.000153 (0.000857)
Village FE	Y	Y	Y
Sargan overidentification test: $\chi^2(4)$ [p]	-	1.292 [.863]	1.270 [.736]
H_0: Owned land orthogonal to the error, Durbin-Wu-Hausman $\chi^2(1)$ [p]	-	0.4603 [.498]	-
H_0: Owned land, land value, and total wealth orthogonal to the error, Durbin-Wu-Hausman $\chi^2(3)$ [p]	-	1.408 [.704]	-
Number of observations	466	466	466

Standard errors in parentheses robust to error clustering and heteroscedasticity. [a]Endogenous variable. Identifying instruments include total inherited land, the real value of inherited land, the real value of nonland inherited wealth, average village rainfall since inheritance, the standard deviation of rainfall since inheritance, and the two rainfall variables interacted with inherited land area.

Appendix Table A3
First-Stage Profit Function Estimates for the Test of Farmer Ability Bias, Farm Size<11 Acres
(*Kharif* Season 2009)

Variable	Owned Cultivated Land	Rental Value of Owned Cultivated Land	Total Wealth
Total inherited land area	0.563 (0.2436)	2205 (1492)	13749 (26989)
Real value of inherited land (x 10^{-5})	0.0148 (0.0194)	743 (119)	-1840 (2147)
Real value of inherited nonland wealth (x 10^{-5})	0.291 (0.0603)	593 (369)	0.450 (0.0668)
Mean rainfall since year of inheritance (MR)	0.00587 (0.00242)	8.16 (14.8)	-129 (268)
Standard deviation of rainfall since year of inheritance (SDR)	-.0134 (0.00368)	-89.0 (22.5)	-387 (407)
MR x total inherited land area	-.00108 (0.000396)	-7.07 (2.43)	-13.2 (43.9)
SDR x total inherited land area	0.00336 (0.000668)	24.6 (4.09)	37.5 (74.0)
Village FE	Y	Y	Y
Partial R^2 of excluded instruments	0.338	0.291	0.142
H_0: Excluded instruments = 0, F(7,441) [p]	32.1 [.000]	25.9 [.000]	10.4 [.000]
Number of observations	466	466	466

Standard errors in parentheses robust to error clustering and heteroskedasticity.