ARE THERE TOO MANY FARMS IN THE WORLD? LABOR MARKET TRANSACTION COSTS, MACHINE CAPACITIES, AND OPTIMAL FARM SIZE

by

Andrew D. Foster and Mark R. Rosenzweig

COWLES FOUNDATION PAPER NO. 1773

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

2022

http://cowles.yale.edu/

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

Andrew D. Foster

Brown University and National Bureau of Economic Research

Mark R. Rosenzweig

Yale University and National Bureau of Economic Research

We show that labor market transaction costs explain why the smallest farms are more efficient than slightly larger farms in most low-income countries and that increases in machine capacity with operational scale result in the globally observed rising upper tail of productivity. We find evidence consistent with these mechanisms using Indian data, and we show that if all Indian farms were at the minimum scale required to maximize the return on land, the number of farms would be reduced by 82% and income per farm worker would rise by 68%.

I. Introduction

> The most dominant as well as the most intractable feature of our agrarian economy is the small size of the holding occupied by the vast majority of the cultivators. No effective solution of the problem of improved production and the crushing burden of poverty can be found until we devise a system in which the unit of agricultural

Data are provided as supplementary material online. This paper was edited by Ali Hortaçsu.

Electronically published January 6, 2022

Journal of Political Economy, volume 130, number 3, March 2022.

https://doi.org/10.1086/717890

> organisation will not ordinarily be below the minimum unit. (United Provinces Zamindari Abolition Committee 1948, 501)

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.

The second stylized fact is that the agricultural productivity of developed country agriculture is substantially higher than it is in low-income countries. The Economic Research Service of the US Department of Agriculture estimates, for example, that 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. It

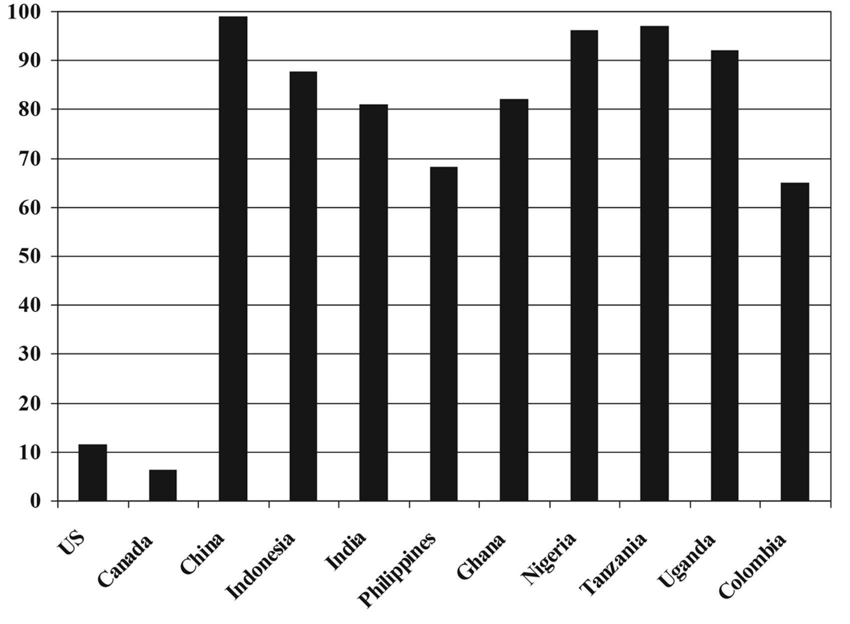

Fig. 1.—Percent of households with operational landholdings below 10 acres by country.

also indicates that in India, productivity in rice and wheat are one-third that of the best producers, located in large-farm countries (USDA 2016).

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 poverty and employment in agriculture. This was the conclusion reached by the United Provinces Zamindari Committee, charged in 1946 with recommending the redistribution of the large landholdings of the Zamindari in the United Provinces of India, in its 1948 report. It suggested a specific minimal scale of operation of 10 acres to boost productivity, based on the principal mode of motive power in India at the time—a bullock pair.

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, in the same markets, and facing the same technology frontier. When farm scale and farm productivity are examined within a country, however, we observe 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 continuous increasing returns to scale are observed among the larger farms in developed countries (e.g., Paul et al. 2004).

While most of the literature documenting the inverse relationship in low-income countries is based on data from India, the Philippines, and Latin America (e.g., Schultz 1964; Hayami and Otsuka 1993; Binswanger, Deininger, and Feder 1995; Vollrath 2007; Hazell 2011; Kagin, Taylor, and Yúnez-Naude 2015), recently available representative data sets reveal that this same inverse relationship, based on well-measured land areas, holds almost universally. Figure 2 displays the relationships between farm area and output per acre in China, Nigeria, Mexico, and Bangladesh.[1] In all four of these data sets based on per-acre yields, the very smallest farms are substantially more productive. Moreover, as can also be seen, the span of farm sizes, based on representative data, 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, though it does suggest that land is misallocated, given heterogeneity in productivity by size.

There is general agreement in the literature that the inverse relationship is not spurious—and, specifically, not due to a correlation of land quality

[1] The data sets used to create these graphs are the Integrated Agricultural Productivity Project 2013 (Bangladesh), the China Living Standards Survey 1995–97, the Mexico Family Life Survey 2002, and the Living Standards Measurement Study Integrated Surveys on Agriculture General Household Survey Panel for 2015–16 (Nigeria).

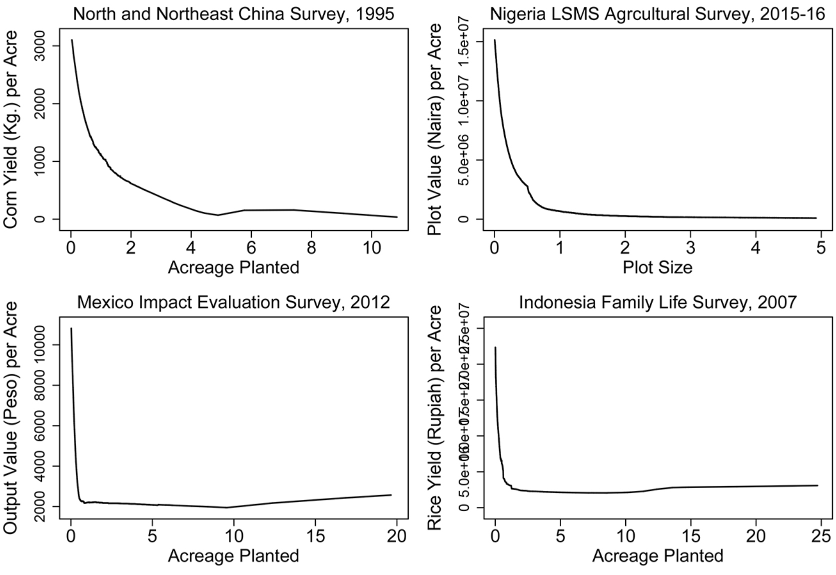

FIG. 2.—Lowess-smoothed relationship of yield and acreage planted by country.

and farm size (e.g., Carter 1984; Bhalla and Roy 1988; Benjamin 1995; Barrett, Bellemare, and Hou 2010) and/or measurement error that is correlated with scale (e.g., Larson et al. 2013; Carletto, Savastano, and Zezza 2013; Ali and Deininger 2014). However, a general shortcoming of this large 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-shaped 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 productivity increases with scale.[2]

Another feature of the agricultural sector of low-income countries, which has received less attention, is the existence of farming operations in which no labor is hired and no family members work off the farm—the operation/farm is autarchic with respect to the labor market. There are only a few survey data sets that provide information on both the composition of the labor force engaged in farming and the off-farm labor supply of family members that permit the empirical identification of autarchic farming. Of

[2] Seen from this global perspective, some of the 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, Bourguignon, and van den Brink 2009; Hazell et al. 2010), while true, cannot explain why corporate farms, which are large scale, are even more productive.

these, the survey data from Nigeria indicate that 36.2% of planting operations are autarchic, the data from China indicate that 17% of farms operated all operations over the entire agricultural under autarchy, and the data from India that we use in this research indicate that 34% of all operations are autarchic. Moreover, all three data sets indicate that autarchic farming is not concentrated among the smallest farms but among intermediate-sized farms, as depicted in figure 3, which shows by country the fraction of autarchic farms or operations by land size.

In this paper, we seek to explain the U-shaped relationship between farm productivity and farm scale and the incidence and patterns of autarchic farming with a model that incorporates labor market transaction costs and scale economies in machine capacities. We provide tests of the implications of the model and quantify a parameterized version based on estimated and calibrated structural parameters to show that the model is capable of yielding the U-shaped productivity relationship, even if the production technology is constant returns to scale. We then use the estimates to identify the magnitudes of labor transaction costs to estimate the optimal farm scale given existing machine technology in India and to carry out a counterfactual land consolidation, embedding the model in an equilibrium framework in which all farms are at the optimal scale. The result of the counterfactual is both an increase in total output from the same total cultivated land as well as a substantial increase in output per laborer. The exercise thus enables us

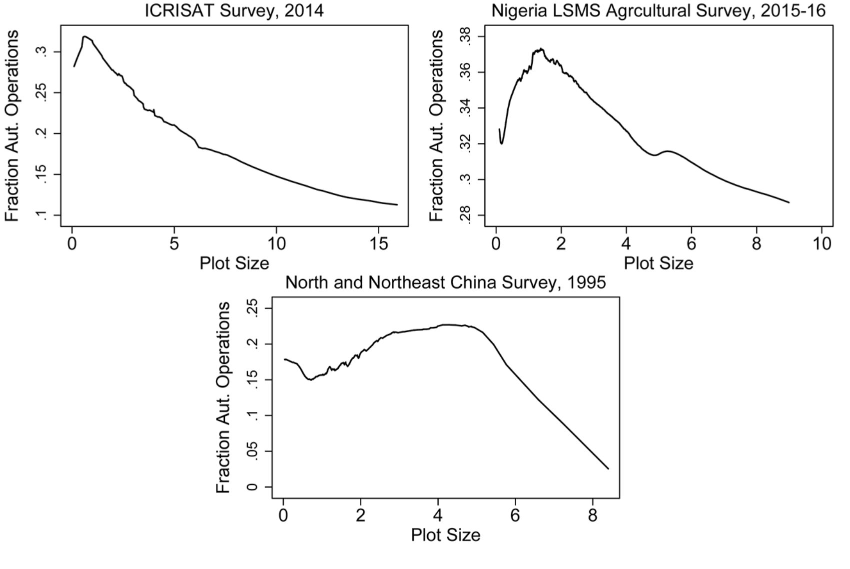

Fig. 3.—Lowess-smoothed relationship of fraction of autarchic operations and plot size by country.

to quantify the number of surplus farms and farm laborers in India and the loss in incomes associated with the existing distribution of landholdings.

We highlight transaction costs in the labor market because they are especially important in agriculture. Agricultural operations, at any scale or level of technology, are sequential and intermittent, and their timing is based on unpredictable weather events—labor is thus principally hired on a daily basis. 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 that 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. Moreover, if labor is hired at all, average unit labor costs will vary by operational scale because larger scale entails more intensive use of labor.[3] The result is a U shape in which the smallest farms are most efficient in their use of labor, slightly larger farms are least efficient, and larger farms are as efficient as the smallest farms because the share of transaction costs in total labor expenditures are smallest. We also show that fixed transaction costs can explain the high share of operations that are autarchic and their absence among the smallest farms.

To account for the increasingly higher productivity of larger farms, we focus on scale economies in machinery capacity.[4] 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). But these facts cannot explain the continuing rise in productivity with scale. We show that the upper tail of the U shape can be explained if two conditions are met: effective machine capacity is increasing with scale and the pricing of capacity is nonlinear.

We are able to examine and quantify the role of transactions costs and machine capacity scale economies as major factors accounting for the U-shaped relationship between scale and productivity within a low-income country because of the existence of the unique data from the India International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) Village Level Studies (VLS) panel survey. One key advantage of the ICRISAT

[3] Allen (1988) shows that one of the reasons that larger farms were more productive than smaller farms in eighteenth-century England, when mechanization was not a major factor, was that larger farms could hire labor crews. If multiple workers are needed at the same time, then hiring a worker team can reduce per-worker search costs for the employer. We test for (but do not find) evidence of this form of scale.

[4] 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 (Rosenzweig and Binswanger 1993). 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 that they are not important determinants of agricultural productivity in low-income countries.

survey relative to many other representative surveys is that larger farms are oversampled. The data set thus contains the missing link between low-income country agriculture and developed country agriculture—because of the oversampling, we are able to observe the U shape that characterizes the global relationship between agricultural productivity and scale in a common environment. Population-representative household surveys in low-income countries contain few if 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.[5]

A characteristic of most existing data sets that has made it difficult to identify the mechanisms that underlie the scale economies that we focus on here is that 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 hours per day. The ICRISAT data record labor time use in hours and days. The data indicate not only that there is substantial variation in the average hours per day workers provide but also that the amount of daily hours within an agricultural operation differs by operation scale.

On the basis of the daily wages and hours information in the ICRISAT data, we can estimate the magnitude of the fixed components of daily wages paid by employers, which we find makes up over 50% of the daily wage paid to a full-time (8-hour) male wage worker. 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 result, the average hourly wage, inclusive of the imputed cost of family labor, increases and then decreases with farm scale.

Another deficiency of existing data describing farming 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 (e.g., thresher, tractor, sprayer) and value, but they provide little or no information on power or capability (e.g., horsepower, bushels processed per time unit). 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 shape.

[5] To our knowledge, there are only two prior studies based on low-income country farm data that find evidence of a U shape. Kimhi (2006), using data on maize producers in Zambia, shows that diseconomies 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 (2019), 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 as both per-acre output and per-acre net returns.

The ICRISAT data also do not provide direct information on the power or capacities of the equipment that is used by the farmers. 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. 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. On the basis of our structural estimates, we are able to identify the optimal scale of operation, conditional on existing wage rates and available sprayer technology, based on the sprayer scale economies—at 24 acres. This compares with the existing mean farm size in India of just over 3 acres, and is 2.4 times the minimum optimal size derived, on the basis of bullock technology, in the Zamindari report.

Incorporating our structural estimates of sprayer technology and the fixed-cost specification of wage schedules, we calibrate the model by fitting its predictions to two moments of the data, stratified by area. On the basis of the calibrated parameters, we are able to reproduce the U shape in profitability even when the production technology exhibits no scale economies. We are also able to show that the marginal product of labor in autarchic production is on average 40% higher than the hourly marginal wage rate, consistent with the underutilization of labor in those operations.

Our counterfactual land consolidation simulation—in which all farms are cultivating at the optimal scale and in which we allow for labor exit from the agricultural sector so that wage rates are endogenously determined—indicates that there are 7.7 times too many farms in India.[6] With all farms at optimal scale, output per acre is increased by 42% and output per worker by 68%. The principal sources of the gains are the elimination of labor misallocation due to the elimination of autarchic farming and the exploitation of machine scale economies. The new land distribution, resulting in an 87% decrease in farms, is also characterized by a reduction in the total labor force but by only 16%. These results indicate that there is thus simultaneously an overall surplus of labor and an underutilization of the existing labor force in agriculture.

In section II, 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 robust to soil quality, crop choice, and farmer characteristics. We also present the evidence on and estimates of labor market transaction costs and nonlinear machine pricing by capacity. In section III, we set out the model, and in

[6] The Zamindari Commission concluded that consolidating landholdings to meet their minimum size of 10 acres would entail a reduction in farms of 66%.

section IV, we carry out tests of the model based on the nonlinear relationships between land size, production labor intensity, and average hourly labor costs. We also present the structural estimates of the sprayer capacity pricing schedule. Section V describes the calibration of the full model and presents the estimated parameters. In section VI, we assess the fit of the model to the data and discuss the estimates of the key unobservables, including the marginal products of labor on small, autarchic, and large farms and the entry costs faced by workers in market participation. Section VII contains the counterfactual land consolidation simulation based on the calibrated model, and section VIII contains a summary of our findings and the potential implications of an endogenous consolidation of landholding from the existing distribution that might arise if legal and institutional barriers to land transactions were eliminated.

II. The Data

A. *Sampling and Information Content*

Our principal data source is the six latest rounds of the India ICRISAT VLS panel survey, covering the agricultural years 2009–14. 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 other household surveys, which seek to achieve household representativeness, because the sampling frame is based on landholding size.[7] 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. First, there is information on input quantities and prices by type of input, farm operation, and individual plot collected approximately every 3 weeks.[8] The high-frequency

[7] Exceptions are Muyanga and Jayne (2019), which oversampled larger farms in Kenya, and the ARIS-REDS surveys (Foster and Rosenzweig 2011), which oversampled large farms in 1967.

[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 (2009–14); only 5.8% of plots were bought or 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 means of land acquisition, including inheritance, do not differ by plot size. The 2014 census data indicate that 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.

input information is thus likely to be more accurate than that found in almost all other 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, and importantly for identifying the role of mechanization in scale economies, there is information enabling the measurement of the power and capacities of machines.

B. *Descriptive Information on Scale and Farm Productivity*

Figure 4 displays from the ICRISAT 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 landowning 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 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

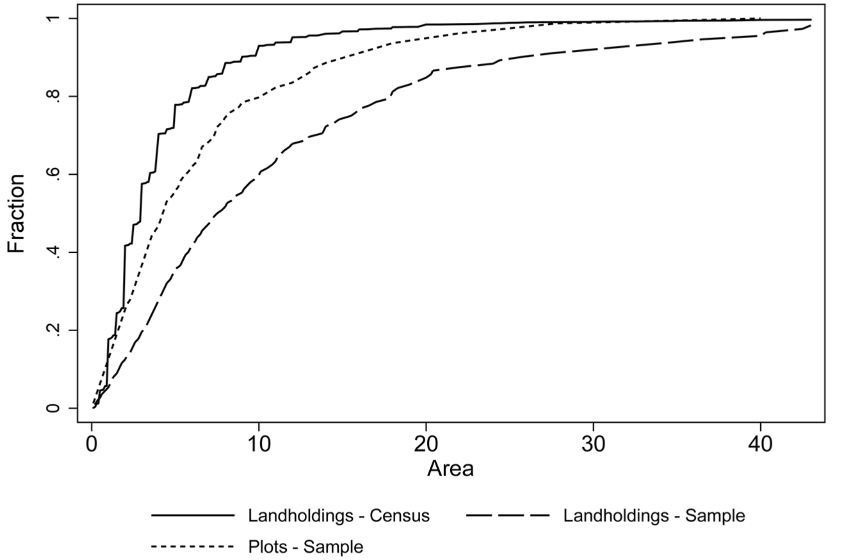

FIG. 4.—Cumulative distributions of owned total land and land plots (acres).

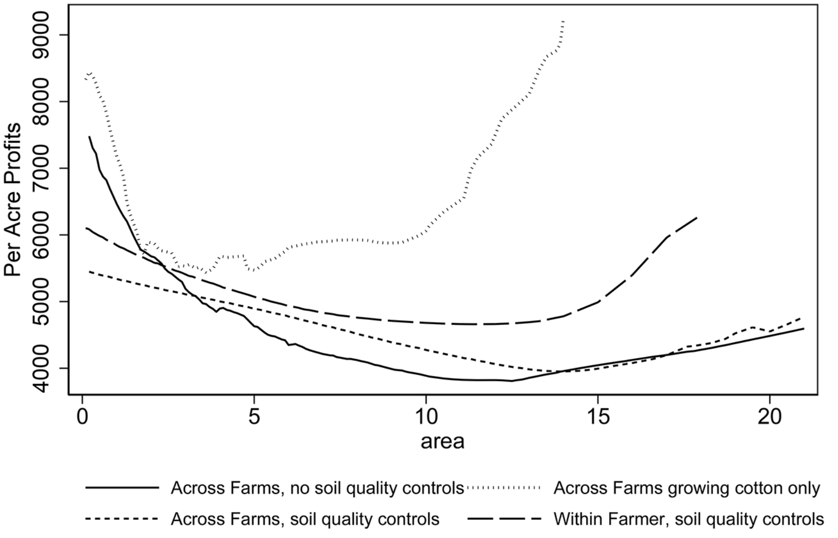

FIG. 5.—Real profits per acre by owned area: roles of plot quality and farmer characteristics (ICRISAT VLS 2009–14).

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.

The solid line in figure 5 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 (2009–14). 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.[9] Using the detailed information of the data set, we can rule out three reasons for the U shape that have been suggested in the literature, which has focused on the decline in productivity with scale below 10 acres. First, the dotted line in figure 5 shows that differences in crops grown by farm size do not account for the U shape; even among farmers growing one crop—cotton—the relationship displays the U shape. Second, the U shape is robust to controls for the 24 land quality variables available in the data set—the relationship estimated using the locally weighted functional coefficient model (LWFCM) from a specification including all of the soil characteristics, in which the coefficients for farm size and the soil characteristics can vary nonparametrically with farm size, is depicted by the short-dashed line in figure 5.[10] Finally, using the same method and the same soil quality controls but also including farmer

[9] This is also true for output per acre as well.

[10] See Cai et al. (2006). The specification we use is locally linear in profits and farm size.

fixed effects, we see that within the same farm plot size and per-acre profitability displays the U shape, as shown by the long-dashed line in figure 5. Thus, variation in farmer wealth or farmer ability or heterogeneity in plot or soil characteristics does not explain the U-shaped association between per-acre profitability and scale.

Similarly, variation in the cost of durable physical inputs, such as seeds or fertilizer, does not explain the U shape, as these costs would not vary systematically within farmer. However, the price of applying those inputs may vary within farmer. For example, as we will discuss in detail below, if the costs of family and hired labor differ and family labor is fixed on any given day, then the marginal cost of labor may be higher on one day than another on the basis of the quantity of labor demanded. If operations were coordinated so that labor is always applied to two different plots on the same day, then the marginal cost of labor would be the same for the two plots across the course of the season. But this turns out not to be the case. On the basis of the dates of operation initiation, we find 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. Evidently, on net the farmers face diseconomies of coordinating operations across their different plots.

C. *Fixed Costs of Labor Hiring*

We will set out a model to explain the U-shaped pattern of farm and plot efficiency as well as the existence of autarchic operations based in part on transaction costs in the labor market. That there is a fixed cost component to hired labor is evident both in the data from price schedules and from the survey information. The first salient feature of the data from the daily wage and hour schedules, provided in 2010 and 2011 monthly for each farm operation by one informant in each land class, is that a large fraction of workers that are paid daily wages work less than 8 hours in a day. Figure 6 displays the distribution of hours worked in the day for hired male workers and, for comparison, bullock pairs and drivers in the kharif season for 2010 and 2011 combined. As can be seen, many workers are hired for less than a full day: 31% of the daily wage reports for hired males were for workers who worked less than 8 hours; for bullock pairs and driver, over 58% of daily wages paid were for work that was less than 8 hours. This is in accord with the survey data on off-farm employment reported by respondents. In the 2014 round, for example, 44.4% of respondents working off farm for wages in agriculture operations during the peak kharif season reported that their average working hours were less than 8 hours.

The second feature of the data on wages and hours is that hourly wages differ by the amount of time worked. We computed hourly wages based on the monthly wage schedules and then regressed the log of the hourly wage

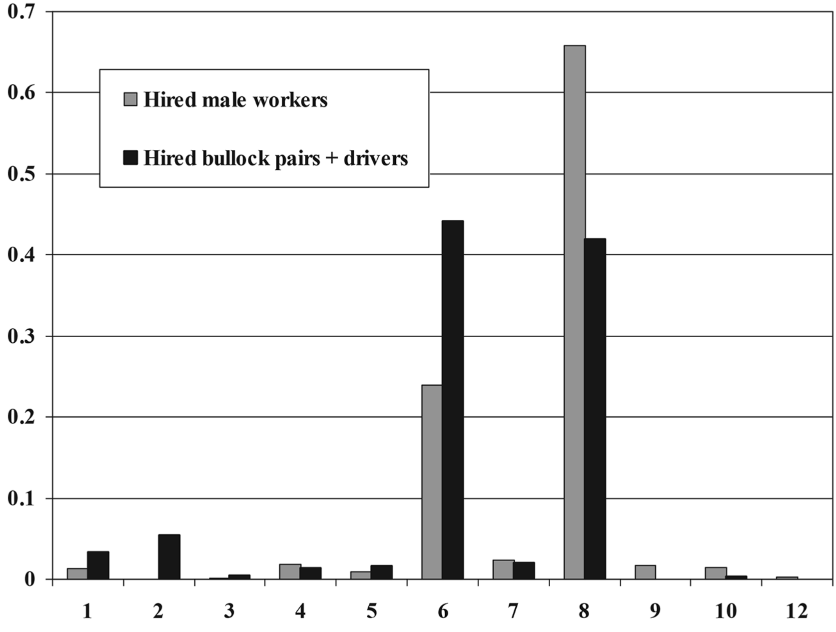

FIG. 6.—Input supply: distribution of average hours worked per day for wages (kharif season) by hired input.

for the two categories of hired inputs on whether the work done was for the full 8 hours. We include a full set of dummy variables for farm operation to ensure that any hourly wage difference by daily hours hired is not merely due to low-wage operations occurring in slack periods with little work. The within-operation log wage estimates from the wage schedule data are reported in column 1 of table 1, where it can be seen that farmers pay an hourly premium for low-hour work: workers who work 8 hours are paid a statistically significant 33% less per hour than lower-hour workers; a hired bullock pair and driver working 8 hours is paid over 22% less per hour than his part-time counterpart.

Data on wages paid and hours of work for hired workers as reported by farmers from the 2014 survey transaction files conform to these patterns. The survey data do not directly report how many workers were hired on a given task, but they do report expenditure and hours for each task by demographic group (men, women, and children). We therefore focus on tasks in which there were ≤12 hours of hired male labor, and we assume that anything in the 8–12-hour range represents full-time work by one worker and that <8 hours refers to part-time work by one worker.[11] Because the

[11] Note that if there are multiple part-time workers in the 8–12-hour category, that would bias down an estimate of the part-time premium.

TABLE 1
Operation Fixed Effects Estimates: Percent Difference in Hourly Wage Rates Paid for 8 Hours versus Less than 8 Hours of Work by Input and Data Source

	Hired Male Labor		Hired Bullock Pair + Driver		Sprayer
	2010, 2011 Monthly Price Schedules (1)	2014 Input Survey (2)	2010, 2011 Monthly Price Schedules (3)	2014 Input Survey (4)	2009–14 Input Surveys[a] (5)
Worked 8 hours per day versus <8 hours	−33.2	−34.7[b]	−22.3	−30.0[b]	−13.2[b]
	(3.14)	(11.9)	(4.54)	(8.42)	(13.1)
Log capacity	. . .	. . .	. . .	. . .	.626[b]
					(.128)
Mean wage (Rs)	22.1	34.7	78.7	114.2	15.9
	(9.34)	(19.3)	(39.6)	(35.0)	(23.3)
Percent working <8 hours	30.7	44.4	58.4	61.0	19.3
Observations	729	3,387	450	1,240	1,201

Note.—Standard errors (in parentheses) are clustered at the village/year level. Hourly wage rate = daily wage/hours worked. Sprayer capacity = material sprayed per hour of use.
[a] Specification also includes village/year fixed effects.
[b] Fixed effects instrumental variables estimate; first stage includes log of owned area and all land quality characteristics.

equilibrium hours chosen by a farmer will itself be a function of the wage schedule, we instrument hours with the characteristics of the farmer's plot, which determine labor demand but should not, net of hours hired, affect the wage in a competitive market.

The estimates in columns 2 and 4 in table 1 are comparable to those from the monthly price schedules. In particular, there is a 34.7% discount in the hourly wage rate for full-time male hired workers and a 30% discount for full-time bullock pairs with a driver. In column 5, we use the same approach for one form of machinery, sprayers, for which we can control for an important sprayer quality—spray capacity, as discussed in detail below. The result, which controls for sprayer capacity, suggests that the scale economies in labor hiring do not also extend to machinery—there is no statistically significant difference in the hourly rental payment for sprayers by hours of use.

The estimates in table 1 are consistent with the existence of a fixed cost associated with hiring a worker for any amount of time and a fixed hourly wage. We can use the distribution of hours from figure 6 and the wage estimates of table 1 to construct an estimate of the fixed cost. Let the expenditure function for hiring one laborer for 1 day who works l_h hours be

$$w_h(l_h) = w_0 + w_1 l_h, \tag{1}$$

where w_0 is the fixed hiring cost and w_1 is the marginal hourly wage. Table 1 provides the wage discount for the average hourly wage in full-time work compared with that for part-time work (<8 hours). We need the distribution of hours worked by part-time/full-time to compute the fixed costs because, as seen in (1), the fixed cost is a different share of the total wage for different levels of part-time work. With f_i^p denoting the fraction of part-time workers working l_{hi}^p hours, the average wage for part-time work given in (1) is

$$\bar{w}^p = \sum_{i=1}^{7}\left(\frac{w_0 + w_1 l_{hi}^p}{l_{hi}^p}\right) f_i^p,$$

and, similarly, the average wage for full-time work is

$$\bar{w}^f = \sum_{l=8}^{12}\left(\frac{w_0 + w_1 l_{hi}^f}{l_{hi}^f}\right) f_i^f.$$

Table 1 tells us that $\bar{w}^p = 1.347\bar{w}^f$. Substituting and using the marginal hourly wage rate of $w_1 = 21$ from the wage schedules to calculate w_0 yields an estimate of the fixed cost of 178 Rs per worker.[12]

Is it plausible that there are significant fixed cost components in labor costs, as implied by the labor price schedules and the transactions survey data? First, most agricultural laborers are hired on a daily basis. Farm operations are episodic and sequential, and operation timing is stochastic, depending on weather. Farm scale is too small for full-time work, and contracting in advance is difficult, given the vagaries of weather. Thus, each worker must be matched with a farmer who is seeking workers for a given day's task on a given day. Perhaps most importantly, there are important transaction costs in the daily hiring of workers that arise from the fact that farms are spatially separated from where workers and farmers in the village reside; travel costs are thus not trivial. The ICRISAT data provide the distance of each plot from the farmers' home (in the village center). The median distance is 1 kilometer.[13] 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. Moreover, these fixed costs of hiring paid by farmers may differ by land size, which will further affect the

[12] We calculate $w_1 = 21$ by determining the average change in daily earning of an extra hour of work when working more than 8 hours.

[13] 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 Economic Growth Center–Centre for Microfinance 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, 23.6% of the survey respondents who worked for wages in agriculture reported working for a farmer located outside the village, and 21.3% of farmers who employed any agricultural laborers reported hiring laborers from outside the village. Among those workers traveling to a farm outside the village by foot or bicycle (63.8%), the average distance to the nonvillage farm was 2 kilometers. The median distance to a nonvillage farm for those traveling by bus (26.5%) was 8 kilometers.

relationship between output and scale. For example, farmers cultivating larger plots may be willing to pay more up front to attract workers for a given operation, since the inability to hire would be more costly than it would be on smaller plots.

In table 2, we report instrumental variables estimates with fixed effects for operation of the hourly wage discounts for full-day work differentiated by both plot size and plot distance for male hired workers and, again, for sprayers from the input transactions data. In column 1, we see that there is a statistically significant positive relationship between the wage discount and plot size. Inclusion of plot distance in the specification, as seen in column 2, does not eliminate the plot size gradient but indicates that, indeed, fixed costs of hiring rise with distance—for every kilometer the plot is located away from the homestead (village center), there is a 14% drop in the hourly wage associated with full-day work. Columns 3 and 4 confirm that the rental price of sprayers of given capacity does not exhibit scale economies associated with hours of employment for any plots differentiated by size and/or distance.

III. Model

A. *Constant Returns to Scale Production and Economies of Scale in Pricing*

In this section, we set out the model that integrates both the fixed cost of labor and heterogeneous capacity in machines. A central feature of both

TABLE 2
Fixed Effects Instrumental Variables Estimates: Percentage Hourly Cost Discount by Plot Area and Plot Distance from Homestead for Male Hired Workers and Rented Sprayers
(Dependent Variable: Log Hourly Wage/Rental Price)

	Hired Male Labor (2014)		Sprayer (2009–14)	
Input	(1)	(2)	(3)	(4)
Worked 8 hours per day versus <8 hours	−12.6	−6.21	−7.59	−2.02
	(12.7)	(12.6)	(14.4)	(2.20)
Worked 8 hours per day × plot area	−31.2	−32.0	−4.86	−6.28
	(6.84)	(6.98)	(4.91)	(5.04)
Worked 8 hours per day × plot distance from homestead	...	−13.5	...	−11.1
		(2.78)		(12.1)
Log capacity	...	...	.582	.423
			(.135)	(.139)
Operation fixed effects	Yes	Yes	No	No
Village/year fixed effects	No	No	Yes	Yes
Observations	3,387		1,201	

Note.—First stage includes owned area and all land quality characteristics. Standard errors (in parentheses) are clustered at the village/year level. Sprayer capacity = material sprayed per hour of use.

is that they induce scale economies not through the structure of the production, as is common in the literature, but via the costs of inputs. Therefore, to highlight their role, 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):

$$\psi g(a, e), \tag{2}$$

where ψ is a total factor productivity parameter, which we also assume is scale invariant. Thus, any scale economies in the model must come from input costs.

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, can be achieved using labor alone and/or by spraying, using labor and a sprayer.[14] Workers entering the labor market for off-farm work on a given day face a fixed entry cost per day f as a result of transaction costs and/or travel (in effect, we define production for work done on a single day). 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 by paying a fixed fee w_0.

The estimates in table 1 pertain to the hiring of one worker for different hours of work. For a given operation, however, a farmer may need more than one worker, given a limit on the maximum hours any worker is willing to work. With labor compensation for one worker having a fixed and variable component, as depicted in equation (1), the cost of hiring l_h hours of work on a given day when each worker works only up to l_{mx} hours is

$$w_h(l_h) = \mathrm{ceil}\left(\frac{l_h}{l_{mx}}\right) w_0 + w_1 l_h, \tag{3}$$

where ceil() is the ceiling function. This structure says, for example, that a farmer must pay $w_0 + 4w_1$ for 4 hours of work but $2w_0 + 11w_1$ for 11 hours of work if 11 hours exceeds the maximum l_{mx} in a day that any worker will work. Equation (3) also characterizes the wage income for off-farm work by family members. Consequently, if we account for the entry cost, f, the opportunity cost of applying l_f units of family labor to the farm is

$$w_f(l_f) = \mathrm{floor}\left(\frac{l_f}{l_{mx}}\right)(w_0 - f) + w_1 l_f \tag{4}$$

if workers are fully compensated for the fixed costs of off-farm work.

[14] The nutrition interpretation of e is relevant for the analysis of scale in plant protection, which will be the focus of the first component of our analysis. However, in order to incorporate harvest work by workers and machinery in the analysis of overall worker demand, we will subsequently expand the definition of e to include any form of inputs applied on the farm.

Our production process allows labor and machinery to be substitutable in the production of nutrients and for heterogeneous machinery. Unlike for *manual* labor, where individual heterogeneity in productivity per unit of time (within gender) is relatively low and is in any case unrewarded in the market where time wages dominate (Foster and Rosenzweig 1996), farm equipment devoted to specific tasks varies significantly in capacity and commands different prices associated with capacity.[15] 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. 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, acres of corn per hour harvested).

Thus, we define the nutrient production function as

$$e(l, q, m) = \left(\omega l^{\delta} + \left(\left(1 - \frac{q}{\phi(a)}\right) qm\right)^{\delta}\right)^{1/\delta}, \tag{5}$$

where q is machine capacity and m is the number of units of time the machine is employed. The parameter ω captures the relative productivity of workers and machines, and δ captures the extent of substitutability between labor and machines.[16]

Equation (5) embodies a relationship between machine capacity and *effective* machine capacity, which depends on farm size. We thus 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. Similarly, it is not cost effective to rent an eight-row harvester for land that has four rows of crops.

Economies of scale in farm machinery capacity may arise from the pricing of machine capacity. In particular, we allow the rental cost per unit of time x_m for a machine to increase at a decreasing rate with machine capacity:

$$x_m = p_m q^{v}, \tag{6}$$

[15] Farmers—i.e., 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. As noted, the within-farmer plot-specific relationships between profits and acreage indicate that any such correlation between farmer ability and farm size is not solely responsible for the profitability patterns across farms of different size.

[16] Some labor will be complementary with machine use, the labor used to actually run the machines.

if $0 < \nu < 1$ and assuming $q > 1$.[17] Finally, we assume that operation of the machinery requires θ units of family labor per hour of machine operation. Thus, we subtract θm from the quantity of labor used to get the labor component of the nutrient production function.

Thus, profits are

$$\pi(a, l_h, l_f, q, m) = \psi g(a, e(l_h + l_f - \theta m, q, m)) - w_h(l_h) - w_f(l_f) - p_m q^{\nu} m. \tag{7}$$

Farmers maximize (6) subject to their family labor constraint that $l_f \leq l_T$, where l_T is the family labor endowment in hours.

B. *Scale Economies and Labor Market Transaction Costs*

Before we take the model to the data, it is helpful to understand the different ways that the cost structures for hiring labor and machinery contribute to scale economies. We begin with the labor market entry costs. Setting aside machine use and thus machine scale economies and assuming $w_o = f$, profits are[18]

$$\pi(a, l_h, l_f) = \psi g(a, l_h + l_f) - w(l_h) - w_1 l_f. \tag{8}$$

The existence of the labor market transaction costs gives rise to three regimes that depend on land size a differentiated by the allocations of family to farm and nonfarm work and the use of on-farm hired labor.[19] In the first regime, at the lowest land sizes, family members work both on farm and off farm, as long as income from working both on farm and off farm exceeds the income from on-farm work only. No workers are hired, given the transaction costs, and thus the upper bound critical value a^* for this regime is where farmers do not hire workers and are just indifferent between entering the labor market and not.

In this regime, with only labor used as an input and at the regime upper bound for a, the marginal value product of family labor is equal to w_1:

$$\psi g_e(a^*, e(l_T))e_l(l_T) = w_1. \tag{9}$$

Thus, if all farms were in this small-scale regime, labor would be allocated efficiently across farms.

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 is

[17] Note that we have assumed that there are no machine fixed hiring costs, consistent with the estimates in tables 1 and 2. The existence of such fixed costs would further increase machine scale economies.

[18] When we calibrate the model, we allow fixed costs f to deviate from w_0.

[19] A fourth regime in which farmers are both working off farm and hiring in workers on the same day would not arise when $w_0 - f = 0$.

the autarchic regime, which arises because of the existence of the labor market entry cost that must be paid by the farmer when hiring labor. This hiring cost will make farmers reluctant to hire workers until land area reaches some higher threshold.

In the autarchic regime, starting at threshold land size a^*, given the fixity of family labor, the marginal product of labor exceeds the market wage and rises as land size increases. Profitability per acre falls with land size as the discrepancy between the (marginal) opportunity cost of labor w_1 and its marginal product rises—increasingly too little labor is used per acre as long as farm profitability without hired labor exceeds profitability employing farm labor and paying the added transaction costs. The fall in profitability per acre will continue until land size equals a^{**}, where farmers are just indifferent between hiring a worker for sufficient hours l_h^{**} that the marginal product of labor equals the marginal wage and defraying worker entry costs by working with only family labor. If we assume at most that one outside worker is hired, a^{**} and l_h^{**} are defined by the pair of equations

$$\psi g_e(a^{**}, e(l_T + l_h^{**}))e_l(l_T + l_h^{**}) = w_1, \tag{10}$$

$$\psi g(a^{**}, e(l_T)) = \psi g(a^{**}, e(l_T + l_h^{**})) - w_0 - w_1 l_h^{**}. \tag{11}$$

Thus, the existence of transaction costs in the labor market is capable of explaining both the decline in profitability per acre with land size and the existence of autarchic farming at small but not the smallest land sizes—they both are manifestations of labor transaction costs. Moreover, the existence of at least some autarchic farms/operations means that the agricultural labor force is underutilized on average and misallocated across farms, with labor marginal products differing on the basis of land size and the size of the family labor force.

At a^{**}, a worker is hired and the marginal product of labor falls to the market wage. 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 as long as the hired worker is less than full-time. If, converse to our assumption in (3), the fixed cost were paid only on the first hired worker (labor hired in teams), then profits per acre would rise asymptotically, reaching that for the smallest-acreage farms. Otherwise, there would be a new set of regimes associated with hiring the second and further workers. The existence of hiring costs can thus explain a partial upturn in profitability per acre. But even where one worker is part time so the marginal cost of additional work is w_1, the farm incurs the fixed cost for each worker; consequently, profits per acre must be strictly less than those that are achieved on the smallest farms in the absence of another form of scale economies.

C. Scale Economies and Machinery

Economies of scale in machinery can lead to additional scale economies by land size, which are governed by machine capacity and its pricing. 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

$$(\phi(a) - 2q)\frac{\theta w_1}{p_m} + (1-\nu)q^{\nu}\phi(a) - (2-\nu)q^{\nu+1} = 0. \qquad (12)$$

While a closed-form solution for (12) is not generally available, it is evident that optimal capacity is determined only by acreage, the parameter ν that determines the relative cost of higher capacity machines, the ratio $\theta w_1/p_m$, and the function $\phi(a)$.[20]

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. The cost advantage of using larger capacity comes from two sources in the context of our model: the nonlinearity of machine pricing schedules, as parameterized by ν, and the necessity of using a machine operator. The data indicate that farmers use one worker hour per machine hour ($\theta = 1$) to operate the machine regardless of capacity.[21] Thus, even if $\nu = 1$, the combined hour labor and machine rental cost of a machine that has twice the capacity will be less than twice as much.

The model implies that not only will the use of machinery increase with farm scale and with rising labor costs but also so will the machine capacity chosen by the farmer. Implicitly differentiating (12), we get

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

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

[20] We assume that labor utilization is inframarginal rather than at the cusp where additional labor hours would require hiring an additional worker.

[21] In 2014, for all 850 sprayer operations, average hours of labor use was 5.84 and that for the sprayers was 5.37.

[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 $(\phi(a)\nu - 2q\nu - \phi(a) + 4q) > 0$.

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

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 possible only if the machine is higher capacity. On the other hand, a reduction in rental pricing scale economies ν will lower machine capacity:

$$\begin{aligned}\frac{dq}{d\nu} = & -\frac{q(\phi(a) - 2q)^2 w\theta q^{-\nu}\ln(q)}{\nu p_m(\phi(a) - q)(\phi(a)\nu - 2q\nu - \phi(a) + 4q)} \\ & -\frac{q(\phi(a) - 2q)}{\nu(\phi(a)\nu - 2q\nu - \phi(a) + 4q)} \\ < 0. &\end{aligned}$$

While the determination of machine capacity q is, in the context of the model, independent of the return to nutrients, 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. 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.

IV. Testing Model Implications

A. *Unit Labor Costs and Scale*

To test more directly that the U shape in the marginal effect of land size on profits arises from changes in unit labor costs by land size, we first plotted the relationship between the real average hourly wage paid, which incorporates fixed costs, and farm size. In the ICRISAT data, family labor is priced at the marginal 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 7.

To further test that the marginal land size effect on unit labor costs differs by land size, we estimated the relationships between the fraction of operations in the kharif season that employ low-hour (≤6 hours) 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 3. For all three production factors, we see that

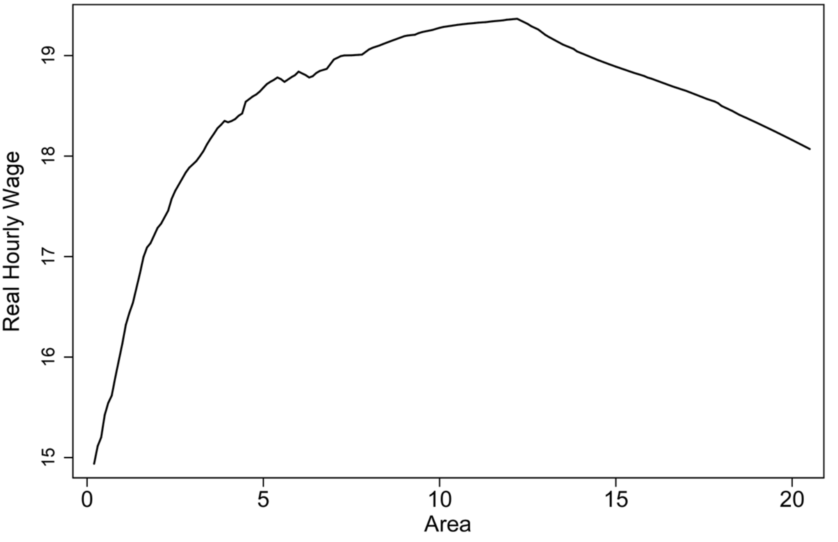

FIG. 7.—Average hourly wage paid for male farm labor by farm size (ICRISAT VLS 2009–14).

the fraction of low-hour operations declines with farm size, and for both hired male labor and hired bullock pairs, so do average wages across the range of farm sizes, as captured by the quadratic specification. Thus, these estimates account for the rise in profitability per acre—and fall in unit costs—above some threshold because of the declining use of high-cost hired labor and hired bullock pairs. The exception is for tractors, for which the

TABLE 3
PLOT SIZE AND FRACTION OF OPERATIONS THAT EMPLOY HIRED INPUTS AT LOW (≤6) DAILY HOURS AND AVERAGE HOURLY WAGE PAID BY INPUT TYPE (Kharif Seasons, 2009–14)

	Fraction of Operations <6 Hours/Day			Average Hourly Wage		
	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^2 × 10^{-3}	.450	.449	.555	8.29	18.3	29.3
	(.112)	(.0682)	(.117)	(3.23)	(32.4)	(10.9)
Village/year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
25 plot and household characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6,777	6,777	6,777	6,777	6,777	6,777

NOTE.—Standard errors (in parentheses) are clustered at the village/year level.

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 3. First, there may be incomplete control for land characteristics, which may be correlated with land size and input use. Second, the model and the labor cost figure suggest 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 may 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. The plot fixed effects absorb both any differences in plot quality and any differences in farmer permanent wealth.

For most levels of rainfall in the semiarid 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 affect neither 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 hourly wages if the additional rainfall induces the hiring of low-hour postplanting labor, while for the larger plots in the same rainfall area, increases in rainfall induce 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 column 1 of table 4, plot and year fixed effects estimates of rainfall and rainfall squared on kharif season profits from each plot are reported. As expected, increases in rainfall increase profits. In column 2, 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 statistically significant only 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 the costs of both average hired male labor and bullock rentals 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 of 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 and year fixed effects estimates

TABLE 4
PLOT FIXED EFFECTS ESTIMATES: EFFECTS OF KHARIF SEASON RAINFALL ON PROFITS, HOURS EMPLOYED, AND AVERAGE HOURLY WAGE RATES BY INPUT TYPE (Kharif Seasons, 2009–14)

		HOURS EMPLOYED			AVERAGE HOURLY WAGE		
	PROFITS (1)	Hired Male Labor (2)	Hired Tractor (3)	Hired Bullock Pair (4)	Hired Male Labor (5)	Hired Tractor (6)	Hired Bullock Pair (7)
Rainfall (mm)	38.1	.182	.00362	.0347	−.0158	.0130	−.0593
	(17.1)	(.0701)	(.00316)	(.0248)	(.00672)	(.0601)	(.0355)
Rainfall^2 × 10^{-3}	−21.2	−.107	−.00214	−.0500	.00778	−.0132	.0757
	(8.59)	(.0377)	(.00161)	(.0268)	(.00398)	(.0282)	(.0331)
Year and plot fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
H_0: rainfall and rainfall^2 = 0, $F_{2,n}$	3.09	4.18	.99	1.97	3.47	.28	3.02
p	.0504	.0183	.3742	.1452	.0352	.7589	.0538
Observations	5,291	3,987	4,016	2,523	3,987	4,016	2,523

NOTE.—Standard errors (in parentheses) are clustered at the village/year level.

of the effects of rainfall at mean rainfall by plot size on low-hour labor use (and the associated 95% confidence intervals) are reported in figure 8. The figure is consistent with the shifting of regimes of labor employment in the model—at small plot sizes, increases in rainfall statistically significantly

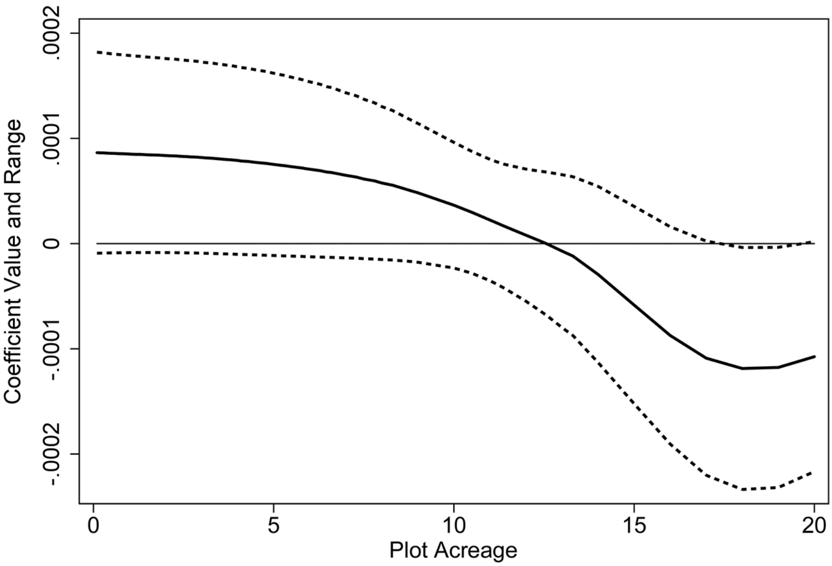

FIG. 8.—Plot fixed effect estimates: effect of rainfall on low-hour labor use (with 95% confidence interval) by plot size.

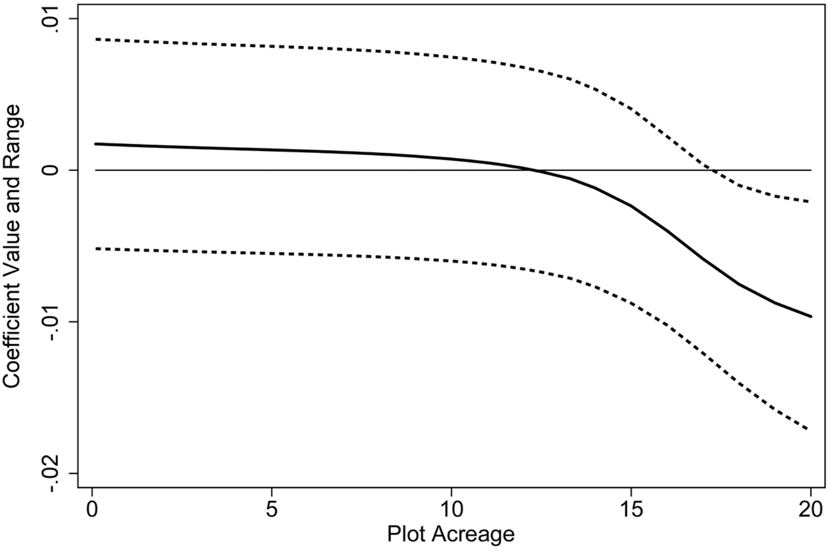

FIG. 9.—Plot fixed effect estimates: effect of rainfall on average male wage (with 95% confidence interval) by plot size.

increase hired low-hour labor use, while for larger plots, low-hour labor operations are statistically significantly reduced when rainfall increases.[23] Also, in figure 9, among the larger plots, increases in rainfall statistically significantly reduce average hourly hired labor costs.[24]

B. *Identifying Equipment Scale Economies as a Source of Farm Scale Economies: The Case of Sprayers*

There has been scant evidence in the literature on the relationship between machine capacity and scale. The ICRISAT transaction data on usage of machinery (as seen in fig. 10) are 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

[23] 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 because of 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.

[24] The effects of rainfall on per-acre profits do not vary by scale over the full range of plot sizes.

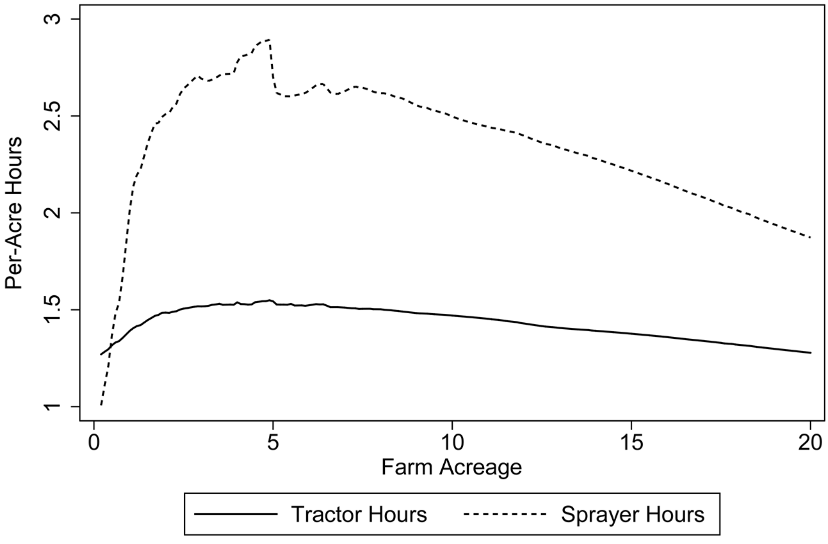

FIG. 10.—Per-acre equipment hours for tractors and sprayers by farm size.

whether machine capacity actually increases with farm size, whether there are scale economies in capacity because of nonlinear 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 is no exception. However, the detailed information on materials purchased and on hours of machine use by machine type in the ICRISAT transaction modules 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.

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. 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. Of course, this focus on sprayers 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. Moreover, as can be seen in figure 10, there are more hours of sprayer use at every land size than hours of tractor use, the next most used machinery.

Power sprayers are heterogeneous in capacity. Figure 11 provides information taken from the website of an Indian purveyor of power sprayers (KisanKraft) 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 liters per hour. In this price list, 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 significantly declines as capacity increases, as the ratio of the highest-capacity machine price is only seven times that of the lowest.

The survey data on input usage suggest that farmers are exploiting economies of scale in spraying. Table 5 reports village/year fixed effects estimates of the effects of land size on the use of any sprayer, weeding hours per acre, sprayer hours per acre, log of the price of the sprayer used, and 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 that larger landowners spend less time per acre in both spraying and weeding compared with smaller farmers.

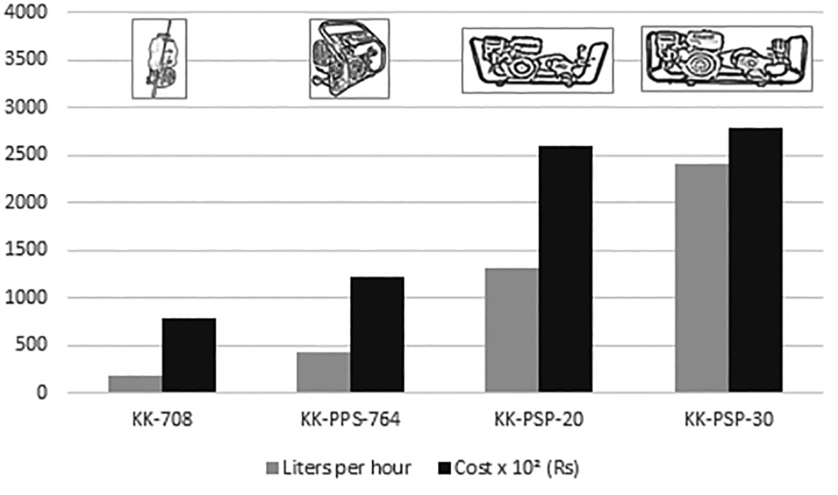

Fig. 11.—Cost and capacities of Indian KisanKraft power sprayers (2017).

TABLE 5

Estimates of 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

	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	.006197	−.5631	−.4063	.01335	.01360
	(.0009879)	(.1286)	(.0853)	(.00669)	(.00667)
All land characteristics	Yes	Yes	Yes	Yes	Yes
Village/year fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	3,374	3,374	1,219	1,219	1,219

Note.—Standard errors (in parentheses) are clustered at the village/year level. OLS = ordinary least squares.

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 capacity and thus the per-capacity price paid are choice variables. A positive shock to equipment prices, for example, will lower a farmer's selected capacity, leading to a negative bias in the estimated parameters of the cost function. As a consequence, we employ the generalized method of moments (GMM) using land area and land area squared as instruments.

Equation (12) implicitly 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 (11) and difference across randomly selected pairs of households i and i' in each village j to eliminate w_1 and p_m. In combination with the differenced log of (6), we then have moment conditions of the following form:

$$E\left(\frac{(1-\nu)q_{ij}^{\nu}\phi(a_{ij})-(2-\nu)q_{ij}^{\nu+1}}{(\phi(a_{ij})-2q_{ij})}-\frac{(1-\nu)q_{i'j}^{\nu}\phi(a_{i'j})-(2-\nu)q_{i'j}^{\nu+1}}{(\phi(a_{i'j})-2q_{i'j})}\,\middle|\,a_{ij},a_{i'j}\right)=0, \tag{13}$$

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

We parameterize $\phi(a)=b_0+b_1a+b_2a^2$ and employ GMM using land area and land area squared as instruments to estimate ν and the b_k. The errors over which these expectations are taken would by variation in the effective price of the machinery, say, due to demand at the time one is undertaking a task, which we have assumed to be uncorrelated with differences in farm area within a village.

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.

Table 6 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 $\nu = 1$ 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 with that characterizing the price schedule for the four power sprayers sold by KisanKraft (listed in fig. 11) and with 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 7. Our GMM estimate of ν for the sprayers used by ICRISAT farmers is comparable to that for the KisanKraft sprayers that are sold across India. However, while the estimated India sprayer ν is more than double that for the sprayers sold in the United States, we cannot reject the hypothesis statistically that they are the same. Technical scale economies in equipment across the United States and India, at least for sprayer technology, cannot explain the differences in their agricultural productivity

TABLE 6
GMM ESTIMATES OF EFFECTIVE CAPACITY FUNCTION $\varphi(a)$ AND PRICE PARAMETER υ

Coefficient	Point Estimate		Robust Standard Error
υ	.316		.124
b_0	5.58		.0375
b_1	.933		.0343
b_2	−.0190		.00211
H_0: $\upsilon < 1$, χ^2_1		30.4	
p		.0000	
Maximum land size (acres) $= \varphi(a)' =$ $-b_1/(2 \times b_2) = 0$	24.5		1.84
Observations		617	

NOTE.—Instruments are owned land area and land area².

TABLE 7
Estimates of Sprayer ν by Source

	India		United States
	ICRISAT Survey (2009–14)	KisanKraft Price List (2016)	Stiles and Stark (2016)
Estimation procedure	IV[a]	OLS	OLS
ν	.5802	.5209	.1458
	(.1200)	(.0605)	(.0789)
H_0: $\nu = 1$, $F_{1,x}$	$\chi^2 = 12.2$	$F_{1,2} = 62.8$	$F_{1,2} = 117.1$
p	.0005	.0156	.0084
Observations	1,219	4	4
Village/year fixed effects	Yes	No	No

Note.—Standard errors (in parentheses) are clustered at the village/year level. IV = instrumental variables; OLS = ordinary least squares.
[a] First stage includes log of owned area and all land quality characteristics.

The estimates of the b_k indicate that $\phi'(a) > 0$ over the relevant range. Thus, smaller farms are less cost effective than larger farms, given that ν is substantially less than 1. 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. The estimated peak maximum land size does not suggest that larger farms than are observed in the ICRISAT area would not be more productive; rather, it is consistent with an equilibrium 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. We will discuss these in the conclusion in section VIII.

V. Model Calibration

In this section, we calibrate the full model, using our estimates of the sprayer pricing functions and other moments from the ICRISAT data. There are three principal aims: first, to ascertain whether our model structure using empirically validated and plausible parameters reproduces the observed U shape in per-acre profitability and other moments of the data; second, to calculate the optimal size of farms using existing machinery technology and pricing in India; and third, to carry out a counterfactual in which we move from the existing distribution of farm sizes to one in which all farms are at the estimated optimum size, conditional on existing machine technology in India. Calibration of the model also enables us to estimate separately hiring fixed costs w_0 and the cost of labor market entry f, because they are identified from multiple moments of the data, and also to estimate the marginal product of labor under autarchy, in which the family labor marginal product, which exceeds the marginal market wage, is not directly observed.

As noted, to fix ideas we used a one-operation, 1-day farm model and focused on sprayers to identify machine scale economies because it is one of the few cases in which machine capacity can be measured explicitly. However, in order to realistically calibrate the model to the relationships between land scale, profits, output, and labor that describe multiple operations with a variety of machinery, we need to specify how many days of work there are for all operations and how the demand for farm labor, at the aggregate level, translates into the demand for workers at different scales and prices of labor.

We proceed as follows. First, we augment the model to allow for multiple days of inputs. Second, because we do not have information on machine capacity for all machinery, we assume that the basic structure of scale economies in machinery and employment are the same for each operation and that the contribution of each operation to overall production is similar. As a result, the model can be collapsed into a single decision that is then replicated on each day in which the farmer works on his farm. The model will thus depict a representative operation. We work with the plot-level data to identify the number of such operations and days per operation over the course of the season by plot size. The numerical model solves for the farmer's optimal input use, given parameters and plot area. The model is nested in a method of moments estimator that relates key features of the data to the predictions from the farmer's problem. The target moments are the profitability per acre and output per worker by total acreage. We assess the fit of the model to a series of nontargeted moments, including labor and machine utilization by farm size. The particulars of the calibration exercise are given in section A of the online appendix.

We use the estimated coefficients for $\phi(a)$ and ν and the estimate of the marginal wage based on the wage and hours data of $w_1 = 21$. We establish from the data that on average we have three potential agricultural workers per household who work on average 8 hours per day so we set full-time work hours l_{mx} to 8 hours and the daily family labor endowment l_T to $3 \times 8 = 24$ hours. The vector of remaining parameters that are calibrated is $[\alpha, \psi, w_0, f, \delta, \omega, p_m, n_0]$. We solve the model on the basis of the assumption that farmers maximize profits conditional on farm area and family size, minimizing over the sum of squared distance between the predicted values for profits per acre and output per worker and those values in the data at 20 different levels of area.

Although the seven parameters are being calibrated jointly, it is possible to get some sense of the identification of specific parameters. We assume, as is apparent in the data, that machine use and hired labor are minimal over the regime in which family members are working off farm and incorporate the smoothed return to off-farm work. Then, given the assumed production function, l, and w_1, the parameters ψ, p_m, α_i, $w_0 - f$, n_0 can be identified from profits per unit area and output per worker on very small farms and those just at the cusp of autarchy (see sec. B in the online appendix). The size of the fixed cost is then determined, given these parameters, as well as the minimum of the profits-per-acre curve, which corresponds to the point at which the farm first hires workers. The elasticity of substitution is identified from how the changes in the productivity of machines (as dictated by the capacity schedule) translate into changes in demand for workers as acreage increases. The price of machinery and relative worker productivity come more broadly from the overall pattern of profits and output per worker on larger farms, where machinery is used, relative to smaller farms, where it is not.

The calibrated structural parameters values are

$$[\alpha = 0.381, \psi = 484, \omega = 0.880, f = 65.5,$$
$$\delta = 0.934, p_m = 20.8, w_0 = 156, n_0 = 0.761].$$

Of primary substantive interest are the fixed cost of hiring workers, the incidence of these costs, and the substitutability of labor and machines. The latter is quite high. The estimated coefficient on δ is consistent with an elasticity of substitution of 13.9. The fixed cost paid by farmers for a hired laborer w_0 is 159 for the mean size 3-acre farm, consistent with our calculation of w_0 based on the regression estimates across all farms and the distribution of observed hours for a given worker. Moreover, consistent with the wage and hours estimates for hiring a single worker from the transaction data in table 2, w_0 is increasing in acreage. It is worth emphasizing that the sources of variation used for identifying these two estimates are quite different: there was nothing, for example, in the matched moments from the structural analysis that distinguishes between full- and part-time work or family and hired workers.[25]

VI. Model Fit and Implications

Figures 12 and 13 present the fit of the data and model to the two targeted series. In figure 12, profits per acre, the model captures both the initial

[25] Interestingly, the estimated fixed cost f incurred by workers in the village who work in the village of Rs 65.5 (or just over 3 hours at the variable component of the wage), which cannot be observed in the data, is less than the fixed cost component of compensation. Because the fixed cost paid through the wage will reflect the opportunity cost of the last worker hired to meet demand and this worker may be from outside the village, local workers who will have to travel less far may be more than compensated for their lower costs of village employment.

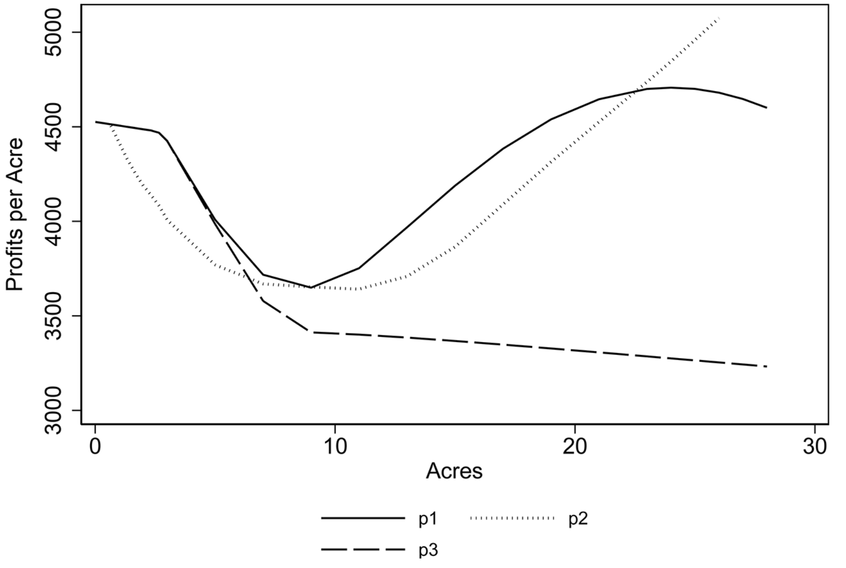

Fig. 12.—Model fit: profits per acre and land size.

decline and the subsequent rise in profits per acre, with both constant technical scale economies and scale-invariant total factor productivity (ψ). As might be expected given the lack of heterogeneity in the model, the data change a bit more sharply than does the model. Variance in, for example, labor productivity in different stages would cause the minimum in the model to move to higher or lower acreage. But the minimum in the model for our representative task corresponds well to the middle of the lower plateau in the data. Similarly, the model does not capture the extreme right of the data. For larger farms, the pattern predicted by the model is governed by the quadratic capacity cure $\phi(a)$ that is estimated from the sprayer data. In practice, one might expect the curve to reach a maximum and then flatten out. But we cannot test this idea, given that we have so few farms of this size or greater.

Output per worker (fig. 13) also fits well. There is a strong upward trend predicted by the model as well as in the data, with the largest farms exhibiting output per worker that is almost twice that of the smaller farms. There is also a distinctive peak in the model at around 9 acres, which would be just below the point at which the farmer first decides to hire workers. While, as with profits per acre, the maximum output per worker in the data is much smoother, it rises, relative to trend, at essentially the same point as the one predicted by the model. Again, the fit seems to diverge most among the largest farms after the capacity curve reaches its maximum.

We did not directly target the total number of worker hours. While this measure is implied by profits per acre and output per worker when no

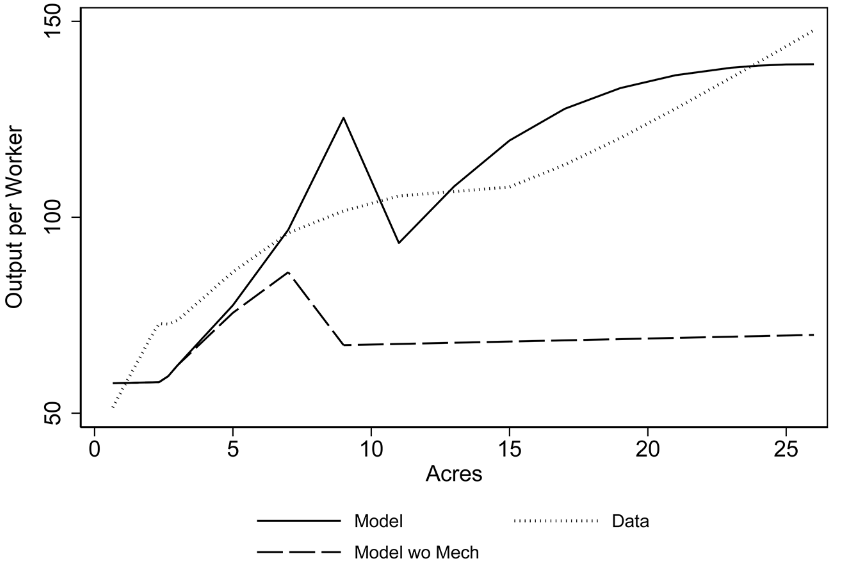

FIG. 13.—Model fit: output per worker and land size.

machinery is employed, such as in the lowest farm sizes, the calculation of hours at the upper end depends importantly on the structure of the model and implications on machine scale derived from the sprayer data. Figure 14 shows the resulting fit, which also looks quite good. The overall shape is similar, and while the overall estimates from the model are systematically lower, the difference is only about 10% for both small (1 acre) and large (24 acres) farms.[26]

We have already noted that the calibrated model estimate of the fixed component of the wage paid by farmers to workers matches well with that estimated from the wage schedules and that the unobserved fixed cost facing workers is actually below on average the fixed compensation component. The model can also be used to predict the fraction of farmers in autarchy in the labor market, something that, as noted above, is a common feature of farms in the developing world and that can be approximated only in the data. The model predicts that 33.8% of farms are autarchic, given the distribution of farm size in the ICRISAT data, which is comparable to the lower bound estimate of 34% of all operations, population weighted,

[26] The model does less well in predicting labor per acre for very small farms. Our smoothing assumption implies that the marginal product of labor is constant (or slightly rising, given that the fixed cost rises with acreage) over the range in which the family works. As noted, this assumption corresponds to how costs for family work are cost out in the computation of profits. To capture this pattern, the model would have to be constructed such that labor market opportunity cost of on-farm work for small farms (where one may need just 4–30 hours over the whole season) is close to zero, as might be the case if work is done largely outside of normal working hours and thus has a low opportunity cost.

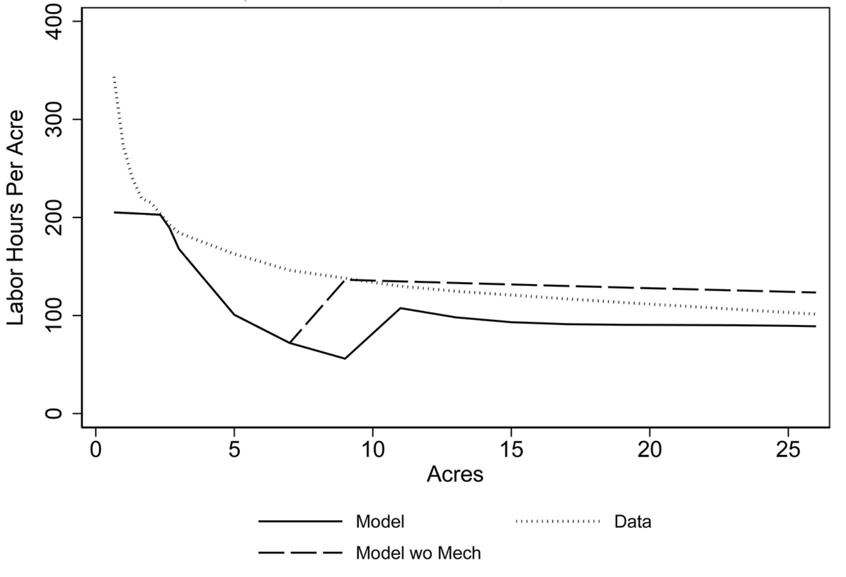

FIG. 14.—Model fit: labor hours per acre and land size.

based on direct calculations from the survey data using input and labor supply data.

We can also use our calibrated parameters to estimate the marginal product of an extra hour of labor by farm size. The former is not observed for autarchic farms that do not hire labor and in which family members do not work off farm. The marginal opportunity cost of employing an extra hour of a hired laborer also is not directly observed on all the other farms, as such costs must take into account fixed costs. Estimates of labor marginal products by farm size are important, as variation in marginal products is indicative of inefficiency in the labor allocation across farms.

The marginal product of the *inframarginal* hour is the same for the smallest and largest farms and corresponds to the marginal hourly wage $w_1 = 21$ Rs, as observed in the data. The marginal product of an additional hour of work on autarchic farms depends, however, on the number of family workers, on acreage, and on all of the parameters describing the farm technology. For the average family size of three workers, the calibrated parameter estimates indicate that the maximum marginal product of workers on the over one-third of farms that are autarchic is 53.8 Rs, farms that are on threshold of hiring, and the estimated mean marginal product is 43.9 Rs. This is more than double the marginal product of an inframarginal hour on the nonautarchic farms. These figures indicate that, given the large fraction of farms/operations that are autarchic, agricultural labor is not only misallocated across farms but is significantly underutilized.

As noted, among farms employing multiple family workers but no hired workers, the net compensation for the costs of wage labor market entry also can matter at some margins. Moreover, for the largest farms that can benefit from hiring more than one worker in an operation, fixed costs also matter. We find that among farms with family members working off farm, the per-hour average marginal product of adding one worker per day is Rs 30.7, still 30% below the average marginal product of labor on autarchic farms. This figure corresponds to the variable component of the wage w_1 plus the hourly component of the difference between the fixed cost of working off farm and the payment made to the worker by the employer. For the largest farms (24 acres), the per-hour marginal product of adding one worker per day is Rs 43.1 on average, reflecting our empirical finding that the fixed component of wage payments w_0 is higher on the larger farms.

The model also allows us to determine how per-acre profitability, output per worker, and employment per acre would vary by land size in the absence of machinery but retaining labor market transaction costs. Figure 12, which shows how profits per acre evolve with land size, indicates that the entire upswing in profits per area for large acreages is due to machinery. The lack of an increase in per-acre profitability by scale without machines among the larger farms is the result of the fact that on these farms we found that labor fixed costs rise with acreage—there does not appear to be team hiring to avoid worker transaction costs.

Figure 13 shows that machines play an important role in expanding output per worker. Output per worker in the presence of machines increases with acreage, only leveling off at the point where capacity is maximized. But output per worker without machines is essentially flat after the farm enters the regime in which it is hiring workers. In terms of either profitability per acre or output per worker, there is little benefit to expanding acreage per farm if one cannot take advantage of the scale economies inherent in machinery.

Finally, figure 14 illustrates the point that workers and machines are important substitutes. Absent machinery, the number of workers employed per acre is about 50% higher than it is when farmers can use available machine technology in the regime in which workers are hired. For very small farms where no machinery is employed, even when machines are available, the two curves coincide, as is the case for profits per acre and output per worker.

VII. Redistributing Land So That All Farms Are Conditionally Optimal Sized

In this section, we use the calibrated model to carry out a counterfactual in which we shift the existing distribution of farms to one in which all farms are at the optimal size conditional on the existing machine technology in

India. Shifting to a uniform farm size that also fully exploits the existing scale economies in machinery eliminates the misallocation of labor across farms and reduces the underutilization of labor due to labor market transaction costs, thus maximizing the return on land.

As is evident from the labor hours per acre graph, an expansion in land size reduces workers per area at the given current wage. To gauge what happens to the wage rate in equilibrium would require the construction and calibration of a full general equilibrium model incorporating a nontradeable nonagricultural sector, capital flows, and attention to domestic and international consumer demand for agricultural and nonagricultural goods. Instead of constructing such a model, we take two benchmark cases. In the first, we assume that the reduction in the demand for agricultural workers that results from the land consolidation is offset by the consequent increase in incomes that raise demand for nonagricultural products and/or increases in foreign capital attracted by the freed-up labor available for nonagricultural work. Demand for employment in the nonagricultural sector thus rises, such that there is no decline in the equilibrium wage. We call this the Lewis scenario, based on the key implication of Lewis' surplus labor model (Lewis 1954).

A second benchmark, which would represent an upper bound on the decline in worker wages, assumes that there is no increase in demand for workers anywhere in the economy after the land consolidation, and thus with a downward-sloping demand curve for labor, the wage declines. For this scenario, in order to calculate the optimum farm size conditional on the existing machine technology of India, we need to take into account nonagricultural labor demand. This is because optimum farm size depends on the market wage rate, which will depend on the total number of workers who stay in the agricultural sector, given the changes in agricultural worker demand.

There are three steps. First, for any existing wage, we calculate the farm size that would be optimal. We do this by calculating the acreage using the model parameters that maximizes for any given wage the profitability per acre. Second, we translate the reduction in total hours associated with the change in farm size that we observe in the data into reduction in the number of workers over the season. If farmers used the same number of workers every day, then this would be straightforward. To make this adjustment, we turned again to the ICRISAT sample and constructed population estimates of total hours (188,101) and of the size of the agricultural work force (1,023 workers) using the sample weights. Assuming that the ratio of workers to worker hours is independent of the distribution of farm sizes (the number of agricultural operations is independent of scale), we use the projected total hours from our model to compute the number of agricultural workers who must be available over the agricultural season for any given land size distribution.

Finally, when the urban or industrial demand elasticity is not infinite and the labor demand curve is not shifted by income effects and/or capital inflows, we need to calculate the decline in the rural wage that would be necessary to fully employ the workers who exit from the agricultural sector so as to equate supply and demand in the economy. To do this, we used the labor demand elasticity estimate from the literature of −0.4 and assumed that the rural wage decline to maintain equilibrium would be proportionate.[27] We obtained from the 2011 India census an estimate of the number of farms in India and the number of urban workers. This then allows us to translate our estimates from the ICRISAT data into national statistics. Combining the wage elasticity estimate with the size of the urban work force, we can then determine a fixed point where the rural wage and urban wage change proportionately for any given distribution of land.[28]

Column 1 of table 8 reports the existing characteristics of Indian agriculture, including the mean farm size, number of farms, and size of the agricultural work force based on the 2011 census data and our calibrated estimates for output, per-acre profits, and machine use. Columns 2 and 3 report, respectively, the Lewis surplus labor equilibrium and the equilibrium in which there is urban wage decline that results from the shift to a farm distribution in which all farms are at the calibrated optimal size. The key difference between the two counterfactual equilibria is that, by construction, there is no decline in the agricultural wage rate in the Lewis equilibrium, while in the equilibrium with no increase in overall labor demand, there is a 8.6% decline in the wage. However, optimal farm size (24 acres) and the reduction in number of farms by 87% is virtually the same, and there is a substantial increase in total output and output per worker in both counterfactuals. We discuss the wage-inelastic equilibrium results, noting any differences from the Lewis equilibrium.

The most notable change from expanding the size of farms to the technology-conditional optimum of 24 acres, from the average size of 3.1 acres, is that there is an expansion of total output by 42% and, because of the substantial reduction in the size of the agricultural labor force by 16%, an increase in per-worker output of over 68%, a change that is only 4% lower than the increase in the Lewis equilibrium in which wages stay

[27] Table B2 in Lichter, Peichl, and Siegloch (2015) contains a meta-analysis regression coefficient for India of −0.398. In Goldar (2009), the most recent estimates (1996–2003) for the short-run and long-run elasticities are −0.309 and −0.480, respectively.

[28] We could assume that rural and urban wages must be equal before and after, but because our measure of urban labor demand is in the form of an elasticity, it is sufficient to assume that the ratio of the rural to urban wage stays constant, e.g., allowing for persistent differences in the cost of living.

TABLE 8
CALIBRATED EQUILIBRIUM EFFECTS OF MAKING ALL FARMS OF OPTIMAL SIZE

Scenario	Baseline (1)	Postreform (2)	Postreform (3)
Urban wage elasticity	. . .	0	−.4
Average farm size (acres)	3.13	24.0	24.1
Number of farms (millions)	95.2	12.4	12.4
Profits per acre (Rs)	4,276	4,705	4,845
Total agricultural output (trillion Rs)	2.71	3.71	3.85
Profit per worker (Rs)	6,302	9,634	9,225
Output per worker (thousand Rs)	14.5	25.4	24.4
Hourly wage (Rs)	21	21	19.2
Size of agricultural labor force (millions)	187	146	158
Work hours per farm	361	2,157	2,340
Machine hours per farm	58.5	163	162
Fraction of farms using machines	.213	1	1
Machine capacity index (mode)	4.49	7.74	7.72

constant.[29] There are thus both too many farms and too many persons engaged in the agricultural sector in terms of the return on land (profitability per acre) and in terms of total output, with the Indian rural work force as a whole substantially hurt by the small size of Indian farms as measured by output per worker. However, if there is not a Lewis equilibrium, landless laborers, who make up a substantial fraction of the agricultural labor force, are made slightly worse off because of the wage decline in the absence of any redistributions of the output gains. Absorption of the rural labor force into the industrial/service sector is thus key to achieving the gains in agriculture that do not penalize a substantial fraction of the population.

What are the sources of the gains? Farm efficiency, measured as profits per acre, of course increases at the optimal farm size by 13%. The key reason for the substantial rise in output is that the larger farms permit the use of higher-capacity machines. This lowers the marginal cost of providing plant nutrients to the farm and thus leads to a substantial increase in output. All the optimal-sized farms use machinery compared with only 21% of farms currently, with mode machine capacity growing by 72%. Because

[29] We are assuming that the increase in total agricultural output has a negligible effect on the output price. The postconsolidation 42% increase in India's output of, e.g., wheat (rice) would represent a 7% (12%) increase in the world output of wheat (rice). The most recent estimated price elasticities of wheat and rice in India (Kozicka et al. 2014) imply that it would result in at most 2% and 4% reductions in the world prices of wheat and rice, respectively, even if there were no increase in the demand for these staples in India because of the large postconsolidation increase in incomes.

machinery and workers are substitutable, this leads to a reduction in the workforce.

Perhaps what is surprising is that the workforce declines, while substantial, are not larger than they are given the elasticity of substitution of 13.9 between worker and machine use. The key reason that the workforce does not decline substantially due to higher machine use—machine hours per acre per day expands from almost no machine use at <7 acres to 3.8 hours per acre at 24 acres—is that labor is significantly underutilized in the current regime. This is because of the existence of fixed labor costs, which induce many farmers with fully utilized family workers to avoid hiring labor. Our calibrated model indicates that labor hours per acre per day on a 24-acre farm is 4.75 hours and that on a 1-acre farm, in which family labor works off farm, is 9.7 hours. But on 9-acre autarchic farms, the threshold acreage above which hired labor is employed, predicted labor hours per acre per day is only 2.7 hours. We noted above that 34% of plots in India are in autarchy, but since autarchic plots are larger than the mean plot, 52% of the acreage is currently farmed in autarchy. With all farms at 24 acres, no farms are in autarchy, and the underutilization of labor is eliminated.

VIII. Conclusion

In this paper, we used unique data from India that permit the examination of agricultural operations across a wider distribution of farm scales than is typically observed in low-income countries because of the oversampling 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 of a low-income country 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 or crops grown and found that the same U-shaped pattern is seen across plots for the same farmer, thus ruling out credit access as the main explanation of higher profitability among large farmers.

We proposed two complementary mechanisms that drive productivity differences by scale and can account together for the U-shaped pattern by scale across countries of the world and in India—fixed costs of hiring labor and machine scale economies. We provided evidence that many workers work for less than a full day; that, consistent with the existence of fixed transaction costs to hiring labor, the hourly price of a worker is higher when workers are used for part of day; that a large proportion of farms and agricultural operations are in autarchy, with neither family workers working off the farm nor any laborers hired; and that intermediate-sized farmers are most likely to employ workers part-time. The implication of this pattern is that small farmers will be relatively efficient in terms of labor utilization 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. The underutilization of labor leads to lower yields and profitability per acre. Eventually, this strategy proves costly and there is a discrete jump upward in total work per acre because of the hiring of non-family workers.

The second mechanism we focused on to explain the rising component of the U shape is the adoption of machine technology that is differentially adapted to farm size. Gains from the use of mechanized inputs that increase with scale come from the fact that higher-capacity machines do more work per hour and the cost of machine capacity does not increase proportionately with capacity. However, large machines cannot be used at full capacity on small farms or plots. This leads to an increase in yield and profitability as farm size increases. We showed empirically that power sprayers conform to the assumptions and predictions of our model—measured sprayer capacity increases with farm size, the shadow price of sprayer capacity increases less than proportionally as capacity increases, and sprayer usage per acre is lower on larger farms.

We calibrated our model—in which neither total factor productivity nor the production technology exhibits any scale economies—to show that labor market transaction costs and machine capacity scale economies are able by themselves to capture both the levels and the nonlinear patterns of aggregate measures of farm outputs and inputs by acreage in the VLS study areas of rural India. Calibration of the model also yields estimates of the transaction costs of hiring labor that are consistent with the wage and hours schedules in the data and reveal that for the 34% of farm operations (accounting for over 50% of total land in India) in autarchy, the marginal opportunity cost of labor (marginal labor product) is on average over 40% higher than it is on the smallest and largest farms. Thus, the existence of high labor market transaction costs leads to a substantial misallocation of labor across farms as well as the underutilization of rural labor on a large fraction of farms

To assess the cost of the existing structure of land and labor allocations in India in terms of total agricultural output and output per agricultural laborer, we used our calibrated model to carry out a counterfactual in which we eliminated all scale differences across farms and fixed farm size at the level that, by exploiting the scale economies of available machines in India, maximizes profitability per acre. We embedded the model in an equilibrium framework in which workers can migrate to urban areas, where labor demand is inelastic. The new calibrated equilibrium—in which the optimal farm size at the equilibrium wage rate and conditional on existing machine technology is 24 acres compared with a mean of just over 3 acres in India today—results in a reduction in the number of farms from over 95 million to 12.4 million; there are over 82 million excess farms in India.[30] Most importantly, the reduction in the number of farms, which eliminates the misallocation of labor across farms and permits exploitation of machine scale economies, results in a 42% increase in agricultural output and a 68% increase in output per worker. The large reduction in the number of farms, despite the high degree of evident substitutability between workers and machines, is, however, accompanied by only a 16% decline in the total agricultural labor force due to the current underutilization of labor.

Our findings also suggest that an endogenous evolution of the distribution and number of farms in India, or other low-income countries where farms are small, to resemble those attributes of farms in high-income countries—even in the absence of any cultural, economic, or legal barriers to the buying or selling of land—is unlikely given the U-shaped profitability curve. Any marginal increase in acreage for most of small farms, which are located on the declining portion of the output-per-acre curve, would reduce profitability per acre. As a consequence, the selling price of the marginal acre from a small seller would exceed the discounted lifetime flow of additional profits for a similarly sized buyer. Moving from the mean farm size of 3 acres to a more profitable farm on the rising segment of the productivity curve would thus require multiple transactions, given the generally small size of farms. The large 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.[31] Thus, despite the potential gains to farm productivity and output per worker of transitioning to an equilibrium of fewer but larger farms, small farms are likely to be the dominant force of production in low-income countries for the foreseeable future without external intervention or increases in the payoffs to workers in the nonagricultural sector.

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 no. 6770, World Bank, Washington, DC.

Allen, Robert C. 1988. "The Growth of Labor Productivity in Early Modern English Agriculture." *Explorations Econ. Hist.* 25:117–46.

[30] The optimality of 24 acres is conditional, as noted, on the machine technology available in India. There is machinery that exists for much larger farms in the world, e.g., sprayer drones. Such technology is not widely available in India because there are no farms of sufficient scale that could profitably exploit it. Machine technology available in India is endogenous to farm size.

[31] The same issues would appear applicable to farmers attempting to cooperate among themselves to exploit scale economies by operating their farms jointly—many farmers would have to coordinate production, not just one or two, and all would have to be from contiguous farms.

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?" *J. Development Econ.* 46 (1): 51–84.

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

Binswanger, H. P., K. Deininger, and G. Feder. 1995. "Power, Distortions, Revolt, and Reform in Agricultural Land Relations." In *Handbook of Development Economics*, vol. 3, edited by J. Behrman and T. N. Srinivasan, 2659–772. Amsterdam: Elsevier Science.

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

Cai, Zongwu, Mitali Das, Huaiyu Xiong, and Xizhi Wu. 2006. "Functional Coefficient Instrumental Variables Models." *J. Econometrics* 133 (1): 207–41.

Carletto, C., S. Savastano, and A. Zezza. 2013. "Fact or Artifact: The Impact of Measurement Errors on the Farm Size Productivity Relationship." *J. Development Studies* 103:254–61.

Carter, M. R. 1984. "Identification of the Inverse Relationship between Farm Size and Productivity: An Empirical Analysis of Peasant Agricultural Production." *Oxford Econ. 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 J. Agricultural Econ.* 72:1146–50.

Davis, C. Austin. 2016. "Why Did Sugarcane Growers Suddenly Adopt Existing Technology?" Working paper, Univ. Michigan.

Foster, A., and M. R. Rosenzweig. 1996. "Comparative Advantage, Information and the Allocation of Workers to Tasks: Evidence from an Agricultural Labour Market." *Rev. Econ. Studies* 63 (3): 347–74.

———. 2011. "Are Indian Farms Too Small? Mechanization, Agency Cost, and Farm Efficiency." Working paper, Econ. Growth Center, Yale Univ.

Goldar, Bishwanath. 2009. "Trade Liberalisation and Labour Demand Elasticity in Indian Manufacturing." *Econ. and Polit. Weekly* 44 (34): 51–57.

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

Hazell, P. 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, January 24–25, 2011.

Hazell, P., C. Poulton, S. Wiggins, and A. Dorward. 2010. "The Future of Small Farms: Trajectories and Policy Priorities." *World Development* 38 (10): 1349–61.

Hornbeck, Richard, and Suresh Naidu. 2014. "When the Levee Breaks: Black Migration and Economic Development in the American South." *A.E.R.* 104 (3): 963–90.

Kagin, J., E. Taylor, and A. Yúnez-Naude. 2015. "Inverse Productivity or Inverse Efficiency? Evidence from Mexico." *J. Development Studies* 52 (3): 396–411.

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

Kozicka, Marta, Matthias Kalkuhl, Shweta Saini, Jan Brockhaus. 2014. "Modeling Indian Wheat and Rice Sector Policies." Working paper.

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 Econ.* 45:355–67.

Lewis, W. Arthur. 1954. "Economic Development with Unlimited Supplies of Labor." *Manchester School* 22 (2): 139–91.

Lichter, Andreas, Andreas Peichl, and Sebastian Siegloch. 2015. "The Own-Wage Elasticity of Labor Demand: A Meta-Regression Analysis." *European Econ. Rev.* 80:94–119.

Muyanga, Milu, and T. S. Jayne. 2019. "Revisiting the Farm Size-Productivity Relationship Based on a Relatively Wide Range of Farm Sizes: Evidence from Kenya." *American J. Agricultural Econ.* 101 (4): 1140–63.

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

Rosenzweig, Mark, and Hans Binswanger. 1993. "Wealth, Weather Risk and the Composition and Profitability of Agricultural Investments." *Econ. J.* 103 (416): 56–78.

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

Stiles, Scott, and C. Robert Stark Jr. 2016. "Estimating Farm Machinery Costs." Working paper, Division Agriculture, Research, and Extension, Univ. Arkansas.

United Provinces Zamindari Abolition Committee. 1948. *Report of the United Provinces Zamindari Abolition Committee*, vol. 1. Allahabad: Superintendent Printing and Stationery.

Vollrath, D. 2007. "Land Distribution and International Agricultural Productivity." *American J. Agricultural Econ.* 89 (1): 202–16.

Yotopoulos, P. A., and L. J. Lau. 1973. "A Test for Relative Economic Efficiency: Some Further Results." *A.E.R.* 63 (1): 214–23.