Farm Size, Productivity, Mechanization and the Land Consolidation Program in Ha Tinh, Viet Nam

Nguyen, Ngoc Phuong Linh

Abstract

Despite being within the top 5 biggest world exporters of rice, Viet Nam's farming sector has been experiencing low productivity. A typical Vietnamese farm household owns multiple rice plots of very small sizes, which leads to high logistics costs and hinders large-scale production. In order to improve productivity, local governments have been implementing a land consolidation program by merging small plots together. However, this program has met several oppositions, mostly due to the widely-acclaimed inverse relationship between land size and productivity. To investigate this relationship, this study used a sample of 1204 rice plots from 464 households in Ha Tinh, Viet Nam. Its results show that while the relationship between *farm size* and *output (harvest)* is mixed, *farm size* does exhibit increasing return to scale on *profit*, which supports the rationale behind the program. Nonetheless, when analyzing regions that have gone through consolidation, the program's results are unclear. Several regional, cultural, and historical factors affecting the program's outcomes are discussed.

Literature review

There has been a large body of literature supporting the **inverse relationship** between agriculture productivity and farm size, which has been used as the main counter-argument against land consolidation programs. One of the earliest proponents of this relationship is Sen (1962), whose brief analysis opened up a lengthy debate on the validity of different productivity-size models. Until now, most papers agree with Sen, oftentimes using data from Africa and Asia. In sub-Saharan African countries such as Uganda and Tanzania, it has been found that gross output and farm business income per unit of land decrease as land size increases (Carletto et al., 2013; Gollin & Udry, 2017; Kilic et al., 2013). The scale of this negative relationship remains rather consistent when fixing for a wide array of parameters (Gollin & Udry, 2017). In Asian countries such as India (Khusro, 1964; Wang et al., 2015), Pakistan (Mahmood et al., 2014), Viet Nam (Liu et al., 2020), and China (Rada et al., 2015), the same inverse relationship was observed, with average output per unit of land and total factor of productivity decline as farm size becomes larger.

There have been quite a few attempts to explain this inverse relationship. One explanation involves labor intensity: it is argued that in smaller plots, labor tends to be employed more per land unit, leading to higher yields (Dyer, 1991; Rudra & Sen, 1980; Sen, 1962, 1964). In particular, more labor work often translates into higher attention to farming techniques, denser cultivation, as well as greater care from farm owners compared to hired labors in large farms. All these factors could then improve output per unit of land. Other explanations include small farms' greater access to credits (Berry & Cline, 1979, as cited by Dyer, 2013), lower yield risks (Srinivasan, 1972), lower price risks (Barrett, 1996), and better management quality (Savastano & Scandizzo, 2017).

Besides the highly acclaimed inverse relationship, several studies found other types of correlations. Savastano and Scandizzo (2017) found that in Ethiopia, farm size and productivity exhibit a "direct-inverse-direct" relationship, or an **inverted-U relationship**. This indicates that for smaller farmers, productivity increases with size but starts to decrease after reaching a certain threshold. The inverted U curve is also found in Southern China (Sheng et al., 2019), and in five Sub-Saharan African countries after controlling for a wide range of variables including soil quality, climate, market access, and topography (Scandizzo & Savastano, 2017). Another relationship is a **positive correlation** between farm size and profitability, which was found by Foster and Rosenzweig (2011) in studying Indian farms. A similar result is shown in Khan's study in Pakistan (1979) and Wang et al. in China (2015). Others found a **mixed relationship** (Dyer, 1991), or **no significant relationship** at all (Chen et al., 2011).

Although the inverse relationship is most prevalent in farm size studies, there is much criticism against it. Even Sen admitted that the inverse relationship is "not really something that has been proved beyond legitimate doubt of exacting statisticians" (1964, p. 323). The main criticisms are summarized below, which are also what this paper attempts to take into account for a more accurate analysis.

First, earlier works often show the inverse relationship through aggregated and/or averaged data for productivity. However, Dyer (2013), Sen (1964), and Rudra (1968a) argued that aggregated and averaged values are not reliable enough to prove the inverse relationship. They emphasized the importance of distributions and variations, as well as the fact that farms of the same size can have very different characteristics in quality, crop use, and labor intensity. For example, after breaking down aggregated data in India, Rudra (1968b) found that the inverse relationship only exists in aggregate data across villages and disappears when each village is analyzed separately.

Second, even in dis-aggregated data analysis, there are often important biases that are hard to be investigated. For example, land quality, which is a missing factor in many papers, can be endogenous to both farm size and productivity (Chen et al., 2011; Dyer, 2013; Eastwood et al., 2010; Gollin, 2019; Sen, 1964). It is possible that historically, more people might have settled down in regions with better land quality, and thus each received only a small plot of land. In contrast, fewer people might have settled in areas where land is not fertile, and thus each received relatively larger plots (Dyer, 2013). Therefore, the inverse relationship might have merely been a product of differences in quality and not in size.

Third, data credibility also raises much concern. Data come mostly from self-reported surveys, which are subjected to misinterpretation of survey questions (especially for farmers with a relatively lower level of literacy) or self-report biases. For instance, there could be a systematic bias in terms of harvest, where smaller farms tend to overestimate their outputs while bigger farms underestimate theirs (Carletto et al., 2013). This has been proved by Desiere and Jolliffe (2017), who found that the inverse relationship only exists for self-reported data and not for crop-cut estimates.

Forth, empirical studies often ignore the use and adoption of technology, which is a crucial part of developing countries' agriculture sector (Dyer, 2013; Otsuka, 2013). They often assume identical use of technologies for farms of different sizes, which is highly inaccurate, as large farms tend to use machines more intensively while smaller farms with less fundings do not (Sen, 1964). In studies where modern technology transition is accounted for, the inverse relationship disappears (Chadha, 1978; Khan, 1979) or shrinks markedly (Liu et al., 2020), indicating that smaller farmers lose their advantage when technology is more intensively employed by large farms. The shift in machine-use behavior directly affects profitability, labor intensity, and the industrialization process, which are crucial in studying development programs. The assumption that all farms have the same technology is thus misleading as well as impractical.

Fifth, the inverse relationship between farm size and output, while important, offers little insight into policy recommendations. Sen (1962) found that the optimal farming area for productivity is 1ha, while that number is 3ha in Savastano and Scandizzo's analysis (2017). Such small sizes, however, are hardly sufficient for a standard four-people family whose only source of income is from farming. The impracticality is heightened when productivity is measured using total crop yield instead of gross profit. As Khusro (1964, p. 52) stated, "there is no reason why farmers should be interested in maximizing gross output per acre – a return which does no wholly accrue to them – rather than some other measure of returns such as farm business income per acre." Khusro's study shows that while crop yield exhibits an inverse trend to land size, gross profit per acre shows the opposite, implying the higher profitability of larger land sizes. Overall, productivity measurements concerning only crop yield fail to capture the actual effect of crop production on farmers' income, therefore limiting the scope of policy interpretations.

Given these problems, this paper attempts to examine the relationship between land size and productivity of rice farming by taking into account a wide range of land characteristics. Section 1 gives a brief overview of the data collection process and descriptive statistics. Section 2 investigates the relationship between *farm size* and *harvest*, followed by the relationship between *farm size* and *profit* in Section 3. Section 4 attempts to explain why the inverse relationship does not exist in the sample. Section 5 presents a brief discussion on Ha Tinh's land consolidation process. The conclusion, then, summarizes all preceding sections and proposes several policy recommendations to the rice farming sector in Ha Tinh.

1. Data overview

The data was collected from July to August 2021 in Ha Tinh Province, Viet Nam, using two rounds of survey. The first round interviewed 15 farm households and businesses, which served as a guide to modify survey questions to fit regional practices. Each household in this round reported in detail their farming process, which is then reflected in the survey questions to account for all possible costs. The second round consists of 504 households randomly selected from seven wards in four farming regions. The regions are close to each other, with highly similar weather conditions in 2021. A final set of n = 464 households remains for analysis.

In the second round, each survey is conducted by interviewing a household representative on their first rice farming season of 2020. Interviewers asked questions and wrote down the answers themselves to ensure accuracy. Each interview has 78 questions, including **household level** statistics (total cultivating area, total costs of the last farming season, etc.) and **plot level** statistics (plot quality, plot area, irrigation quality, machine rentals, etc.). At plot level, because most households own multiple plots, each only reported three: the biggest, the smallest, and another one of their choice (if owning more than two plots). Households also reported whether each plot went through the land consolidation process, with the reported *size* before and after the program was implemented. After eliminating outliers, n = 1,204 plots remain for analysis, with a total of 211.67 acres of land. Table 1 below shows the sample counts; Table 2 shows the descriptive statistics at household level. Survey questions are summarized in Appendix A.

Table 1. Household level and plot level samples count

Regions	No. of wards	No. of households interviewed	No. of plots		Area (acre)		No. of reported plots that are consolidated
			Total owned	Reported	Total owned	Reported	
#1	2	104	535	291	104.4	58.8	57
#2	1	83	281	205	40.2	29.0	0
#3	1	143	409	320	60.4	49.5	0
#4	3	134	851	388	154.1	74.3	268
Total	7	464	2,076	1,204	359.2	211.7	325
% of total			100%	58.00%	100%	58.94%	26.99%

Table 2. Descriptive statistics for household-level data

	Unit	Mean	S.D.	Median	Min	Max
Labor count						
Full-time	person	1.7	0.5	2	0	4
Part-time	person	0.2	0.5	0	0	2
Total cultivating area	acre	0.8	0.6	0.6	0.08	6.2
Number of plots owned						
Total (100%)	plot	4.5	2.7	4	1	19
Good quality (20%)	plot	0.9	1.2	0	0	7
Average quality (65%)	plot	2.9	2.2	3	0	15
Bad quality (15%)	plot	0.7	1.4	0	0	10
Total harvest	ton	1.7	1.4	1.4	0.2	12.5
Total sold	ton	0.6	0.9	0	0	8
Total costs	mil. VND*	4.8	3.1	4.4	0.6	25
Total other incomes	mil. VND*	46.0	37.6	35	0	270

* VND = Viet Nam Dong. Exchange rate as of October 30, 2021 is 1 USD = 22,928 VND.

From Table 2, we can see that on average, each household has 1.7 people working full-time and 0.2 person working part-time in rice production. This reflects the fact that normally, at least one parent within a two-parent household has to take on additional work besides rice farming. The mean cultivating area is 0.8 acre per household, distributed on 4.5 plots, indicating the very small size of rice plots in Ha Tinh. Among all plots, 20% are of good quality, 65% are of average, and 15% are of bad quality. Quality is self-assessed by household representatives; however, it is of considerable accuracy, as later regression analyses will show their significant correlation with harvest. Total harvest per household ranges from 0.2 to 12.5 tons, with a mean of 1.7 tons. Considering the rice price from 5 to 11 million VND (equivalent to US$220 – 480, depending on whether rice is sold on the market or sold to a third-party), a household's average revenue ranges from 8.4 to 18.5 million VND (US$370 – 810). Even for a small-sized family in suburban and rural areas, this is not a high number, and additional works are typically required to make ends meet. Total cost is also high compared to the revenue, averaging at 4.8 million VND per household.

While most parameters are self-explanatory, labor and output measurements need further clarification. Similar to the majority of agriculture studies, one of the most difficult parameters to obtain accurately is labor. This is because rice farmers generally involve themselves in a wide variety of activities, such as feeding cattle or planting other crops. Besides, rice production does not take the same amount of time every day; that is, there will be high times during the season where labor is needed all day, while the rest only requires minimal effort. Therefore, when asked to estimate their working hours during the four-month season, farmers showed little to no confidence in giving out an accurate answer. Furthermore, even when working hours can be measured, the qualities of those hours are far from similar due to the differences in workers' skills.

A substitution method for measuring labor in this paper is *full-time labor* and *part-time labor*, as well as *machinery substitutes*. Full-time labors are family members whose main job is planting rice, while part-time labors are family members who have other jobs but do supporting work on the farm. Machinery substitutes refer to the processes within the entire production that use machinery, either by renting or owning. In terms of renting, in Ha Tinh, machines are rented out with labor: for example, the businesses who rent out a seeding machine will also operate the machine. Households are not allowed to operate it themselves, due to their low level of familiarity with technology. Therefore, rented machines are perfect substitutes for labor in Ha Tinh. The formulas for labor and machine use are shown in Appendix B.

Measurements of output, on the other hand, use two indicators: *harvest* and *profit*. Raw *harvest* is measured by *dried tons of rice*. This takes care of the variations that many studies fail to account for, such as whether crops are of the same type, or whether they are weighed before or after being dried. *Profit* is another indicator of output. As mentioned in the literature review, it is the most important indication for poverty alleviation. Appendix B shows the calculation of profit.

2. The relationship between farm size and harvest

To investigate the relationship between farm size and productivity, we use the following model for household-level analysis:

$$\log(harvest) = \beta_0 + \beta_1 \log(area) + \beta_2 labor + \beta_3 log(costs) + \beta_4 X \quad (2.1)$$

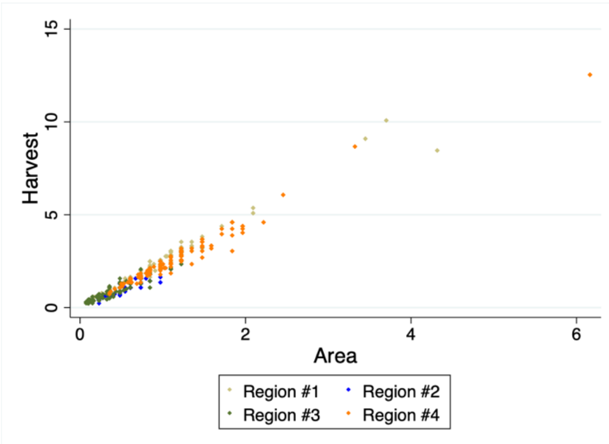
For plot-level analysis, the following model is used:

$$\log(harvest) = \beta_0 + \beta_1 \log(area) + \beta_2 log(labor) + \beta_3 log(machine\_costs) + \beta_4 log(other\_costs) + \beta_5 Y \quad (2.2)$$

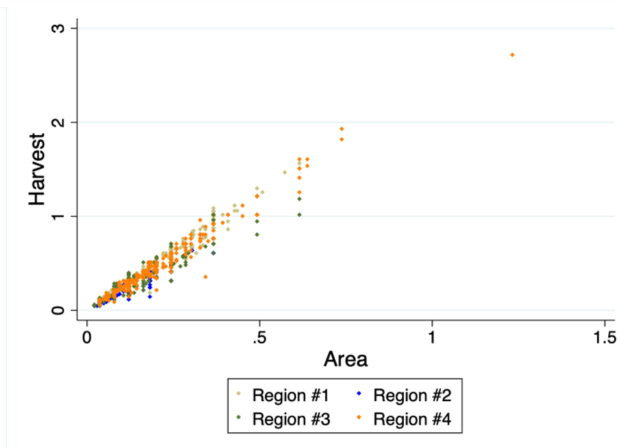
in which, *harvest* denotes tons of dried rice harvested after the first rice season, *area* is the cultivating area, *labor* is the number of people working full-time and part-time, *costs* account for all production costs and represent capital inputs besides land. Within *costs*, *machine_costs* are the costs used for renting or buying machines, and *other_costs* are capital costs besides those involving machine use. In (2.2), X is a matrix of household-related characteristics; in (2.2), Y is a matrix of plot-related characteristics. The reason why *labor* does not enter (2.1) in *log* value is because it is a discrete variable with a very small range (from 0 to 4). Using *log* in this case is thus inappropriate. Model (2.2), on the other hand, calculates *labor* by acre (Appendix B), which is not discrete and can be used in *log* value.

Model (2.1) and (2.2) are modified from the Cobb-Douglas model of production, in order to observe the behavior of β_1. The relationship between β_1 and 1 shows whether cultivating area exhibits increasing return to scale or not. If β_1 is significantly greater than 1, this means increasing area would increase output per acre. Graph 1 and Graph 2's upward concave trend suggests the validity of model (2.1). Regression results of (2.1) and (2.2) are presented in (6) under Table 3 (household level) and Table 4 & 5 (plot level). In addition, the linear-linear and linear-log relationships between *harvest per acre* and *area* are also examined, as suggested by Graph 3 and Graph 4. Regression results for these relationships are presented in (2) and (4) under Tables 3, 4, and 5.

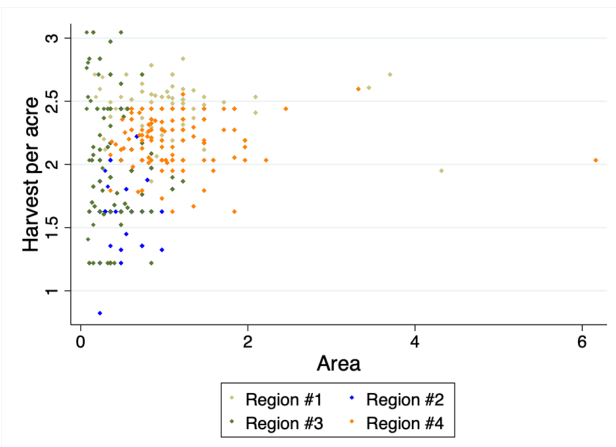
Graph 1. Harvest and area, household level



Graph 2. Harvest and area, plot level



Graph 3. Harvest/acre and area, household level



Graph 4. Harvest/acre and area, plot level

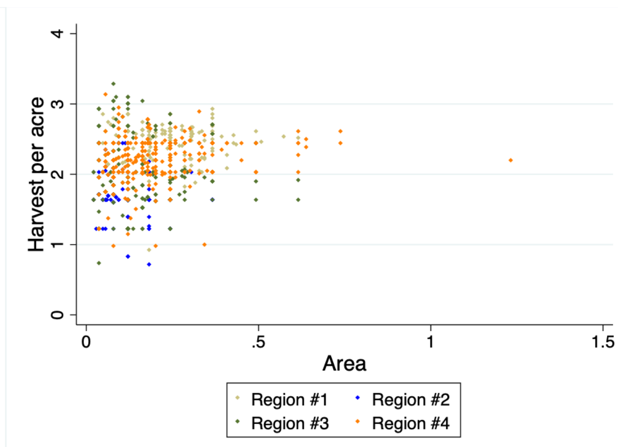


Table 3 indicates a positive relationship between *farm size* and *harvest* at household level. In particular, regression (3) and (4) show that as *area* increases, *harvest per acre* also increases. Regression (5) and (6), which link directly to our main model (2.1), have the same result: a one-sided hypothesis test shows that the coefficient β_1 of $\log(area)$ on $\log(harvest)$ is significantly larger than 1 at $p < 0.01$. Model (2.1) can be rewritten as $harvest = f(X, labor, costs) \times area^{1.129}$, in which *area* exhibits increasing return to scale on *harvest*. In general, at household level, increasing farm size benefits harvest per acre.

At plot level, when machine costs only include machine rental costs (Table 4), the same relationship holds. However, when household-owned machines are also included in the costs (regression (6) in Table 5), β_1 is only larger than 1 at $p < 0.1$. Overall, *area* exhibits increasing return to scale on *harvest* at plot level, but the trend is not as significant as at household level. It is possible that households maximize their overall harvest, but the maximization is less emphasized for individual plots within their ownership. Another plausible explanation is that, at household level, when asked to report costs, households do not take into account the costs of machines they own. This would explain why at household level (Table 3) and when owned machines are not accounted for (Table 4), the positive relationship between land size and productivity is more significant.

Another important observation is that in all models, when factors besides land area are taken into account, the positive effect of *area* on *harvest* is even larger. In particular, within Tables 3, 4, and 5, we can see that β_1 is larger in (4) than in (3), and larger in (6) than in (5). This emphasizes the importance of land size on productivity.

This relationship between land size and harvest is, however, not robust. When the models are regressed by region, the increasing return to scale of *area* on *harvest* is only significant for Region #1 and Region #3 when machine costs are not adjusted, and only for Region #3 when machine costs are adjusted (Table 6). For other regions, the relationship is either not significant, or showing a decreasing return to scale. Overall, the positive relationship between *area* and *harvest per acre* does not hold universally.

From Table 3, 4, and 5, we could also see some other relationships. *Land quality* affects *harvest* in all regressions, with bad quality significantly reducing harvest and productivity. The same result holds for *irrigation*, in which a lack of water supply lowers productivity and profitability. In terms of *labor*, its effect on *harvest* is not clear at plot level. This holds for different ways of calculating *labor* (Appendix B). *Machine costs* does not significantly affect *harvest* either. Another important, and perhaps surprising, result is that *consolidation* does not affect productivity. Possible explanations and further discussions for this relationship are presented in Section 4.

Table 3. Factors affecting harvest, household level

Parameters	Harvest		Harvest per acre		Log(Harvest)	
	(1)	(2)	(3)	(4)	(5)	(6)
Cultivating area	2.30***	1.84***	0.17***	0.20*	1.10***	1.13***
(log for (3)-(6))	(0.078)	(0.11)	(0.030)	(0.10)	(0.015)	(0.054)
Labor count						
Full-time		0.056*		0.088***		0.043**
		(0.029)		(0.034)		(0.018)
Part-time		-0.0060		0.0047		0.0021
		(0.025)		(0.033)		(0.017)
Total costs		0.079***		-0.28***		-0.16***
(log for (3)-(6))		(0.021)		(0.098)		(0.052)
Other household incomes		0.00071**		0.055***		0.029***
(log for (3)-(6))		(0.00030)		(0.019)		(0.0095)
Plot count						
Good quality		0.033*		0.078***		0.037***
		(0.019)		(0.017)		(0.0081)
Average quality		-0.013		0.0010		0.00056
		(0.012)		(0.0090)		(0.0044)
Bad quality		-0.057***		-0.060***		-0.032***
		(0.016)		(0.013)		(0.0068)
Region (base = #1)						
#2		-0.39***		-0.82***		-0.41***
		(0.043)		(0.049)		(0.026)
#3		-0.23***		-0.37***		-0.18***
		(0.034)		(0.041)		(0.02)
#4		-0.17***		-0.29***		-0.13***
		(0.053)		(0.034)		(0.016)
Constant	-0.10*	-0.016	2.17***	2.54***	0.76***	0.98***
	(0.054)	(0.056)	(0.017)	(0.21)	(0.0084)	(0.11)
Observations	464	426	464	425	464	425
R-squared	0.96	0.98	0.095	0.50	0.94	0.97

Robust standard error in parentheses, with significant levels *** p<0.01, ** p<0.05, * p<0.1. Each column is a robust OLS regression. Costs are not adjusted for machine costs. For regression (3), (4), (5), and (6), *Cultivating area*, *Total costs*, and *Other household incomes* are in *log* value. Regression (6) shows results for model (2.1).

Table 4. Factors affecting harvest, plot level (machine costs not adjusted)

Parameters	Harvest		Harvest per acre		Log(Harvest)	
	(1)	(2)	(3)	(4)	(5)	(6)
Cultivating area	2.35***	2.24***	0.14***	0.27***	1.071***	1.15***
(log for (3)-(6))	(0.032)	(0.051)	(0.019)	(0.064)	(0.010)	(0.034)
Labor		-0.019**		0.025		0.0086
(log for (3)-(6))		(0.0094)		(0.024)		(0.013)
Costs						
Machines		-0.013		-0.30***		-0.16***
(log for (3)-(6))		(0.021)		(0.052)		(0.027)
Others		0.017		0.014		-0.0012
(log for (3)-(6))		(0.015)		(0.043)		(0.024)
Land quality (base = good)						
Average		-0.027***		-0.19***		-0.089***
		(0.0052)		(0.028)		(0.013)
Bad		-0.059***		-0.41***		-0.21***
		(0.0075)		(0.038)		(0.021)
Didived into sub plots		0.0090		-0.023		-0.00036
		(0.015)		(0.082)		(0.042)
Consolidated		0.0066		0.023		0.014
		(0.0056)		(0.025)		(0.012)
Irrigation (base = surplus supply)						
Enough		0.00043		-0.025		-0.021
		(0.0076)		(0.035)		(0.019)
Insuficient		-0.0092		-0.14***		-0.080***
		(0.0094)		(0.046)		(0.026)
Region (base = #1)						
#2		-0.11***		-0.70***		-0.35***
		(0.0063)		(0.033)		(0.019)
#3		-0.078***		-0.47***		-0.23***
		(0.0066)		(0.035)		(0.018)
#4		-0.048***		-0.26***		-0.12***
		(0.0085)		(0.037)		(0.019)
Constant	-0.034***	0.068***	2.36***	2.87***	0.86***	1.123***
	(0.0049)	(0.011)	(0.037)	(0.083)	(0.019)	(0.044)
Observations	1,204	1,204	1,204	1,190	1,204	1,190
R-squared	0.92	0.95	0.038	0.44	0.90	0.94

Robust standard error in parentheses, with significant levels *** p<0.01, ** p<0.05, * p<0.1. Each column is a robust OLS regression. Machine costs only include rental costs and treat household-owned machines as zero cost. Part-time labor is treated as 0.5 full-time labor (interpretations of all regression results are the same when part-time labor is treated as 0.25 and 0.75 full-time labor). For regression (3), (4), (5), and (6), *Cultivating area*, *Labor*, *Machine costs*, and *Other costs* are in *log* value. Regression (6) shows results for model (2.2).

Table 5. Factors affecting harvest, plot level (machine costs adjusted)

Parameters	Harvest		Harvest per acre		Log(Harvest)	
	(1)	(2)	(3)	(4)	(5)	(6)
Cultivating area	2.35***	1.89***	0.14***	0.18	1.07***	1.11***
(log for (3)-(6))	(0.032)	(0.17)	(0.019)	(0.15)	(0.010)	(0.084)
Labor		-0.023**		0.014		0.0031
(log for (3)-(6))		(0.0096)		(0.024)		(0.013)
Costs						
Machine		0.13**		-0.19		-0.11
(log for (3)-(6))		(0.060)		(0.15)		(0.082)
Others		0.0058		0.012		-0.0017
(log for (3)-(6))		(0.015)		(0.042)		(0.023)
Land quality (base = good)						
Average		-0.029***		-0.19***		-0.089***
		(0.0053)		(0.028)		(0.013)
Bad		-0.061***		-0.40***		-0.21***
		(0.0075)		(0.038)		(0.021)
Divided into subplots		0.0089		-0.050		-0.015
		(0.011)		(0.086)		(0.044)
Consolidated		0.0052		0.014		0.0092
		(0.0055)		(0.025)		(0.013)
Irrigation (base = surplus supply)						
Sufficient		-0.00063		-0.0067		-0.011
		(0.0077)		(0.035)		(0.019)
Insufficient		-0.0078		-0.11**		-0.063**
		(0.0093)		(0.045)		(0.025)
Region						
#2		-0.093***		-0.75***		-0.38***
		(0.0080)		(0.049)		(0.027)
#3		-0.067***		-0.44***		-0.21***
		(0.0072)		(0.034)		(0.018)
#4		-0.031***		-0.28***		-0.14***
		(0.0092)		(0.049)		(0.026)
Constant	-0.034***	0.059***	2.36***	2.82***	0.86***	1.11***
	(0.0049)	(0.011)	(0.037)	(0.19)	(0.019)	(0.10)
Observations	1,202	1,202	1,202	1,199	1,202	1,199
R-squared	0.92	0.95	0.038	0.43	0.90	0.94

Robust standard error in parentheses, with significant levels *** p<0.01, ** p<0.05, * p<0.1. Each column is a robust OLS regression. Machine costs treat household-owned machines as rentals, and cost per acre is computed as average renting costs. Compared to Table 4, two observations were omitted in Table 5 because there is no reference cost for owned weeding machines. Part-time labor is treated as 0.5 full-time labor (interpretations of all regression results are the same when part-time labor is treated as 0.25 and 0.75 full-time labor). For regression (3), (4), (5), and (6), *Cultivating area*, *Labor*, *Machine costs*, and *Other costs* are in *log* value. Regression (6) shows results for model (2.2).

Table 6. Factors affecting harvest, plot level, by Region (model 2.2)

Region	#1	#2	#3	#4	#1	#2	#3	#4
	(machine costs not adjusted)				(machine costs adjusted)			
Log(Cultivating area)	1.12***	0.77***	1.31***	0.96***	0.73***	0.81***	1.81***	0.95***
	(0.082)	(0.15)	(0.042)	(0.069)	(0.13)	(0.15)	(0.22)	(0.13)
Log(Labor)	0.013	-0.016	0.058**	-0.0088	0.0029	-0.014	0.053**	-0.012
	(0.016)	(0.024)	(0.023)	(0.020)	(0.015)	(0.024)	(0.023)	(0.020)
Log(Costs)								
Machine	0.0049	0.16***	-0.30***	0.18***	0.33**	0.088	-0.67***	-0.066
	(0.065)	(0.029)	(0.035)	(0.057)	(0.15)	(0.13)	(0.23)	(0.13)
Others	-0.095	0.12	-0.18***	-0.10**	-0.025	0.16***	-0.28***	0.17***
	(0.077)	(0.13)	(0.035)	(0.045)	(0.068)	(0.029)	(0.040)	(0.053)
Land quality (base = good)								
Average	0.022	-0.14***	-0.20***	-0.019	-0.00072	-0.15***	-0.20***	-0.015
	(0.031)	(0.032)	(0.026)	(0.015)	(0.029)	(0.031)	(0.026)	(0.015)
Bad	-0.0098	-0.27**	-0.41***	-0.091***	-0.0044	-0.29**	-0.39***	-0.091***
	(0.033)	(0.12)	(0.035)	(0.025)	(0.034)	(0.12)	(0.036)	(0.024)
Divided into sub plots	-0.058**	(omitted)	(omitted)	-0.015	-0.023	(omitted)	(omitted)	-0.023
	(0.026)			(0.052)	(0.021)			(0.052)
Consolidated	-0.010	(omitted)	(omitted)	0.035**	-0.0011	(omitted)	(omitted)	0.030*
	(0.015)			(0.017)	(0.014)			(0.017)
Irrigation (base = surplus)								
Sufficient	0.019	-0.047	-0.30***	-0.023	-0.026	-0.039	-0.26***	-0.017
	(0.037)	(0.032)	(0.034)	(0.024)	(0.038)	(0.031)	(0.038)	(0.023)
Insufficient	-0.020	-0.0047	-0.29***	(omitted)	-0.048	-0.0036	-0.25***	(omitted)
	(0.045)	(0.035)	(0.046)		(0.040)	(0.034)	(0.050)	
Constant	0.99***	0.44**	1.51***	0.74***	0.60***	0.49***	2.10***	0.73***
	(0.095)	(0.17)	(0.059)	(0.078)	(0.14)	(0.17)	(0.23)	(0.13)
Observations	288	200	320	382	288	205	319	387
R-squared	0.97	0.90	0.92	0.95	0.97	0.91	0.92	0.95

Robust standard error in parentheses, with significant levels *** p<0.01, ** p<0.05, * p<0.1. Each column is a robust OLS regression using model (2.2). Parameters denoted (omitted) are omitted due to insufficient numbers of observations. Harvest, Cultivating area, and Costs are in log values.

3. The relationship between farm size and profit

The models used here are similar to (2.1) and (2.2) due to the Cobb-Douglas model and Figure 5 and 6, except that *harvest* is replaced by *profit*. We have at household level:

$$\log(profit) = \beta_0 + \beta_1 \log(area) + \beta_2 labor + \beta_3 log(costs) + \beta_4 X \tag{3.1}$$

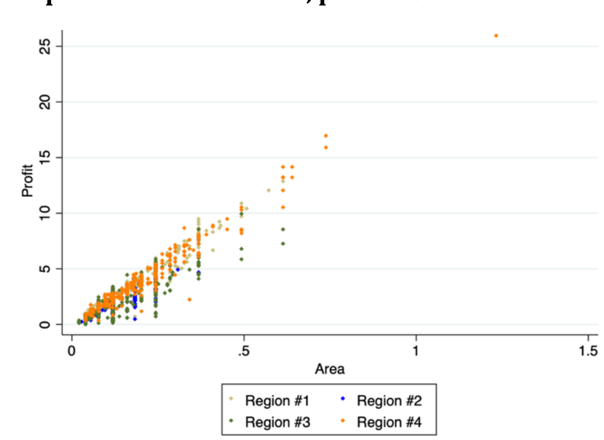
And at plot level:

$$\log(profit) = \beta_0 + \beta_1 \log(area) + \beta_2 log(labor) + \beta_3 log(machine\_costs) + \beta_4 log(other\_costs) + \beta_5 Y \tag{3.2}$$

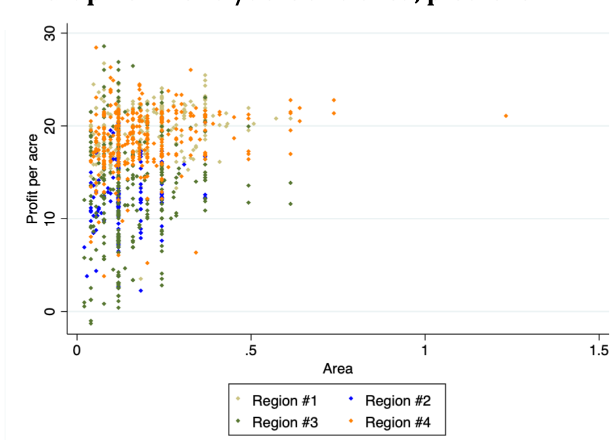
Model (3.1) is shown in Table 7 (6); model (3.2) is shown in Table 8 (6) (machine costs not adjusted) and Table 9 (6) (machine costs adjusted). In all models, we can see that *area* exhibits increasing return to scale on *profit*, and the magnitude is much bigger than what we saw earlier on *harvest*. In particular, $\beta_1 > 2$ in all cases, and is significantly larger than 1 for $p < 0.01$. This implies that variable costs are generally low in Ha Tinh's agriculture sector, and thus increasing farm size significantly improves profit.

From Tables 7, 8, and 9, we can also see that the coefficients of *costs* and *machine costs* are significantly smaller than zero. This indicates that while machine use makes up for the lack of labor in larger farms, using more machines lowers *profit*. Similar results are found for other capital inputs besides machinery and land. Bad *land quality* and a lack of *irrigation* also significantly reduce *profit*. Last but not least, *profit per acre* of consolidated plots shows no significant differences from the rest.

Graph 5. Profit* and area, plot level



Graph 6. Profit*/acre and area, plot level



*Graph 5 and 6 use rice price of 11 million VND per ton, which is the median price of rice sold directly on the market. Costs are adjusted.

Table 7. Factors affecting profit, household level

Parameters	Profit		Profit per acre		Log(Profit)	
	(1)	(2)	(3)	(4)	(5)	(6)
Cultivating area	15.76***	16.58***	2.54***	8.45***	1.30***	2.16***
(log for (3)-(6))	(0.43)	(0.97)	(0.30)	(1.05)	(0.040)	(0.21)
Labor count						
Full-time		0.50*		0.75**		0.032
		(0.27)		(0.31)		(0.045)
Part-time		-0.054		0.036		-0.00072
		(0.23)		(0.30)		(0.039)
Total costs		-0.29		-9.11***		-1.22***
(log for (3)-(6))		(0.19)		(1.01)		(0.22)
Other household incomes		0.0064**		0.46***		0.057**
(log for (3)-(6))		(0.0027)		(0.18)		(0.024)
Plot count						
Good quality		0.30*		0.70***		0.072***
		(0.18)		(0.16)		(0.018)
Average quality		-0.12		0.012		0.011
		(0.11)		(0.083)		(0.011)
Bad quality		-0.51***		-0.57***		-0.069***
		(0.14)		(0.12)		(0.019)
Region (base = #1)						
#2		-3.53***		-7.53***		-0.70***
		(0.39)		(0.45)		(0.063)
#3		-2.09***		-3.55***		-0.37***
		(0.31)		(0.38)		(0.050)
#4		-1.48***		-2.74***		-0.26***
		(0.47)		(0.31)		(0.043)
Constant	-1.87***	-0.15	13.43***	28.90***	2.56***	4.66***
	(0.30)	(0.50)	(0.15)	(2.09)	(0.016)	(0.41)
Observations	463	426	463	425	460	422
R-squared	0.94	0.96	0.20	0.60	0.78	0.86

Robust standard error in parentheses, with significant levels *** p<0.01, ** p<0.05, * p<0.1. Each column is a robust OLS regression. Costs are not adjusted for machine costs (interpretations of all regression results are the same when part-time labor is treated as 0.25 and 0.75 full-time labor). The price of rice here is 9 million VND, or US$395 (interpretations of all regression results are the same when price is calculated at 5, 7, 11, and 13 million VND). For regression (3), (4), (5), and (6), *Cultivating area*, *Total costs*, and *Other household incomes* are in *log* value.

The above analysis results (on page 15) remain valid when the price of rice is varied to be at 5, 7, 9, 11, and 13 million VND per ton, and when part-time workers are counted as 0.25, 0.5, and 0.75 a full-time worker. All results hold when each region is analyzed separately.

Table 8. Factors affecting profit, plot level (machine costs not adjusted)

Parameters	Profit		Profit per acre		Log(Profit)	
	(1)	(2)	(3)	(4)	(5)	(6)
Cultivating area	16.18***	20.18***	2.19***	9.42***	1.20***	2.23***
(log for (3)-(6))	(0.42)	(0.46)	(0.21)	(0.62)	(0.025)	(0.11)
Labor		-0.17**		0.18		-0.020
(log for (3)-(6))		(0.084)		(0.23)		(0.029)
Costs						
Machine		-1.12***		-4.83***		-0.53***
(log for (3)-(6))		(0.19)		(0.49)		(0.066)
Others		-0.85***		-4.68***		-0.72***
(log for (3)-(6))		(0.14)		(0.41)		(0.089)
Land quality (base = good)						
Average		-0.24***		-1.68***		-0.16***
		(0.047)		(0.26)		(0.026)
Bad		-0.53***		-3.76***		-0.48***
		(0.068)		(0.35)		(0.054)
Divided into sub plots		0.081		-0.33		-0.000078
		(0.13)		(0.77)		(0.089)
Consolidated		0.059		0.19		0.025
		(0.051)		(0.23)		(0.024)
Irrigation (base = surplus supply)						
Sufficient		0.0038		-0.22		-0.047
		(0.068)		(0.33)		(0.036)
Insufficient		-0.083		-1.40***		-0.21***
		(0.085)		(0.44)		(0.074)
Region (base = #1)						
#2		-0.97***		-6.58***		-0.59***
		(0.057)		(0.31)		(0.046)
#3		-0.70***		-4.19***		-0.31***
		(0.059)		(0.31)		(0.040)
#4		-0.43***		-2.51***		-0.25***
		(0.077)		(0.35)		(0.038)
Constant	-0.53***	0.61***	16.60***	27.88***	2.82***	4.27***
	(0.067)	(0.10)	(0.38)	(0.79)	(0.044)	(0.14)
Observations	1,204	1,20	1,204	1,19	1,193	1,179
R-squared	0.85	0.92	0.086	0.58	0.65	0.81

Robust standard error in parentheses, with significant levels *** p<0.01, ** p<0.05, * p<0.1. Each column is a robust OLS regression. Machine costs only include rental costs and treat household-owned machines as zero cost. Part-time labor is treated as 0.5 full-time labor (interpretations of all regression results are the same when part-time labor is treated as 0.25 and 0.75 full-time labor). The price of rice here is 9 million VND, or US$395 (interpretations of all regression results are the same when price is calculated at 5, 7, 11, and 13 million VND). For regression (3), (4), (5), and (6), *Cultivating area*, *Labor*, *Machine costs*, and *Other costs* are in *log* value.

Table 9. Factors affecting profit, plot level (machine costs adjusted)

Parameters	Profit		Profit per acre		Log(Profit)	
	(1)	(2)	(3)	(4)	(5)	(6)
Cultivating area	15.38***	16.97***	2.14***	9.61***	1.21***	2.68***
(log for (3)-(6))	(0.32)	(1.53)	(0.21)	(1.51)	(0.025)	(0.25)
Labor		-0.21**		0.094		-0.066*
(log for (3)-(6))		(0.086)		(0.23)		(0.039)
Costs						
Machine		0.15		-5.00***		-0.86***
(log for (3)-(6))		(0.54)		(1.45)		(0.21)
Others		-0.95***		-4.58***		-0.84***
(log for (3)-(6))		(0.13)		(0.42)		(0.10)
Land quality (base = good)						
Average		-0.26***		-1.67***		-0.16***
		(0.047)		(0.25)		(0.030)
Bad		-0.55***		-3.73***		-0.66***
		(0.067)		(0.35)		(0.091)
Divided into sub plots		0.080		-0.58		-0.034
		(0.099)		(0.82)		(0.14)
Consolidated		0.047		0.15		0.034
		(0.050)		(0.23)		(0.028)
Irrigation (base = surplus supply)						
Sufficient		-0.0057		-0.022		-0.0078
		(0.070)		(0.33)		(0.041)
Insufficient		-0.070		-1.10**		-0.21**
		(0.083)		(0.43)		(0.11)
Region (base = #1)						
#2		-0.84***		-7.01***		-0.71***
		(0.072)		(0.46)		(0.067)
#3		-0.61***		-4.00***		-0.32***
		(0.065)		(0.32)		(0.046)
#4		-0.28***		-2.78***		-0.36***
		(0.083)		(0.47)		(0.063)
Constant	-0.47***	0.53***	16.03***	28.41***	2.77***	4.84***
	(0.050)	(0.096)	(0.38)	(1.88)	(0.044)	(0.30)
Observations	1,202	1,202	1,202	1,199	1,190	1,19
R-squared	0.85	0.91	0.083	0.56	0.53	0.72

Robust standard error in parentheses, with significant levels *** p<0.01, ** p<0.05, * p<0.1. Each column is a robust OLS regression. Machine costs treat household-owned machines as rentals, and cost per acre is computed as average renting costs. Compared to Table 8, two observations were omitted in Table 5 because there is no reference cost for owned weeding machines. Part-time labor is treated as 0.5 full-time labor (interpretations of all regression results are the same when part-time labor is treated as 0.25 and 0.75 full-time labor). The price of rice here is 9 million VND, or US$395 (interpretations of all regression results are the same when price is calculated at 5, 7, 11, and 13 million VND). For regression (3), (4), (5), and (6), *Cultivating area*, *Labor*, *Machine costs*, and *Other costs* are in *log* value.

4. Explaining the absence of the inverse relationship

The widely-acclaimed inverse relationship between size and harvest (raw output) does not hold in our sample. It is possible that by including a wide array of variables, most biases are accounted for and thus the true relationship is shown. This paper proposes another possible explanation, which can be directly linked to Sen's model (1962). As discussed in the literature review, the most widely claimed reason behind the inverse relationship is small plots' higher labor intensity per unit of land. However, the following model proposes that, with machine substitutes and alternative job opportunities available, small plots no longer accumulate higher labor intensity, and thus the inverse relationship disappears.

Assume that a household owns a single plot and wishes to maximize their profit by adjusting labor use and machine use. Any labor use saved by machines is then applied in alternative jobs, such as raising cattle or planting different crops. *Harvest* is determined by the Cobb-Douglas function. Assuming a fixed rice price p, we have harvest and revenue from rice production is:

$$harvest(l, l_m, m) = AK^{\alpha}(l + l_m + m)^{\beta}$$

$$revenue(l, l_m, m) = AK^{\alpha}(l + l_m + m)^{\beta}p$$

in which, total factor of productivity A, capital inputs K, capital's return on output α, and labor's return on output β are fixed. Labor used in rice production has three components: m is the machine used for a part of the production process, l_m is the labor work that can be substituted by that machine, and l is the labor used in the rest of the production process. We have $l_m + m = M$ is fixed. To understand this better, consider when a tractor (used for plowing) is the only machine available for rent. Labor l is the labor hours used for any production process besides plowing. M is equivalent to the total hours of manual plowing work. A household might rent the tractor to plow half of the land and plow the other half manually; in which $l_m = m = \frac{1}{2}M$. If they decide to use the tractor entirely, we have $m = M$ and $l_m = 0$. On the other hand, if they plow the whole plot manually, $m = 0$ and $l_m = M$. Notice that m is included in *labor* because it is a perfect substitute for labor work, as explained in Section 1.

If the total labor available is L and any remaining labor $(L - l - l_m)$ is used in other works (such as industrial jobs, raising cattle, planting different crops, etc.), we have additional revenue equals:

$$additional\ revenue(l, l_m) = i(L - l - l_m)$$

with i is the hourly rate of alternative income streams. If we fix machine rental cost r, value of labor w, and capital costs $f(K)$, then the total costs for a household is:

$$cost(m) = f(K) + wL + rm$$

The maximization of a household's profit is then:

$$\begin{aligned} profit(l, l_m, m) &= revenue(l, l_m, m) + additional\ revenue(l, l_m) - cost(m) \\ &= AK^{\alpha}(l + l_m + m)^{\beta}p + i(L - l - l_m) - [f(K) + wL + rm] \end{aligned} \tag{4.1}$$

We can rewrite this profit function as:

$$profit(l, l_m, m) = \left[AK^{\alpha}(l + M)^{\beta}p - il\right] - [il_m + rm] + [iL - f(K) - wL] \tag{4.2}$$

Here we have decomposed *profit* into three components. The first component is $\left[AK^{\alpha}(l + M)^{\beta}p - il\right]$, which is the regular maximization of output with respect to labor l, and it does not depend on machine use. The second component is $[il_m + rm]$, which a household can adjust by deciding how much labor l_m and machine m should be used. The last component is $[iL - f(K) - wL]$, which is fixed based on our assumptions.

To maximize (4.2), households will then need to minimize $[il_m + rm]$, subjecting to $l_m + m = M$ and $l_m, m \geq 0$. When $r > i$, we have

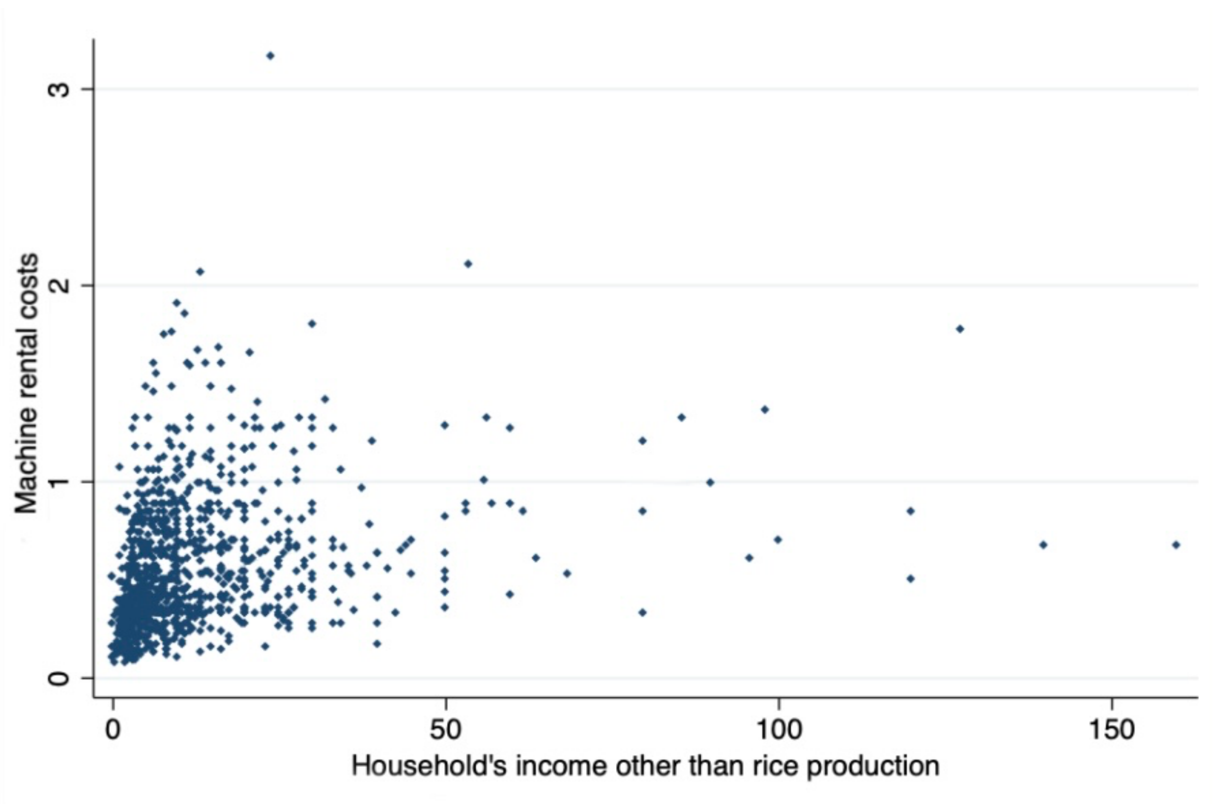
$$il_m + rm = (il_m + im) + (-im + rm) = i(l_m + m) + (r - i)m \geq i(l_m + m) = iM$$

We obtain the minimum when $m = 0$. The other case when $i > m$ is similar, with the minimum obtained when $l_m = 0$. When $i = m$, it does not matter how much l_m and m are chosen.

Intuitively, when the machine rent price is cheap, and households find that spending time on other jobs can make more money $(i > r)$, they will rent the machine whenever possible $(m = M, l_m = 0)$. On the other hand, when machine rent price is expensive relative to incomes from other jobs $(r > i)$, households will not use any machine $(m = 0, l_m = M)$.

When machines are extended to account for n machines $m_1, m_2, m_3, \dots, m_n$ with respective rental price $r_1, r_2, r_3, \dots, r_n$, we will need to minimize $\sum_{k=1}^{n}(il_{m_k} + rm_k)$. Similarly, a household will use whichever machine has the cheapest relative price compared to other machines and compared to the potential alternative incomes. Graph 7 verifies this behavior: there is a positive relationship between a household's alternative incomes and machine costs in the last farming season. The more profitable alternative income opportunities are, the more labor will be employed in such opportunities, and more machines are used in rice production. This also implies a necessary step in poverty alleviation: while expanding plot size can boost industrialization, providing farmers with other job opportunities would also help speed up the process.

Graph 7. Machine used in rice production and household's other incomes



How does model (4.1) link to the inverse relationship? In Sen's model (1962), one of the assumptions is that there is no alternative income available, or $i = 0$. In this case, the only way to maximize (4.1) is to set $l + l_m = L$ for rice production. In other words, if there is no alternative job available, a household will spend all of their labor on rice production. As a result, in smaller plots, labor use L will be distributed more densely per acre, leading to higher productivity and subsequently the widely-acclaimed inverse relationship between farm size and output.

However, when alternative incomes and machine substitutes are available, this change in labor intensity does not hold. This is observed by looking at the first component in (4.2), which is $\left[AK^{\alpha}(l + M)^{\beta}p - il\right]$. The optimization of this component would be achieved at a level of labor use dependant on land area. Call this optimal level $l_0(area)$. For small farms, l_0 is small, and any remaining labor $L - l_0$ will be transferred to other jobs besides rice production. On the other hand, for large farms that requires more labor than available ($l_0 > L$), households only need to increase machine use m to make up for the insufficient labor supply. Therefore, the change in labor intensity as farm size varies, as observed in Sen's model, does not hold, and the inverse relationship no longer exists.

In our model, we have not accounted for rented labor. Rented labor is a perfect substitute for household labor, and thus in our case, the same as machine substitute. Similarly, to maximize profit, households will use hired labor whenever their alternative income is more per hour than the costs of hourly labor rent. If the two rates are the same, the model will be identical to (4.1), except that there is no upper constraint for labor. In the case of Sen, which assumes there is no alternative income stream, rented labor would not be available either because it is a form of alternative job. This provides possible directions for future study in testing the relationship between land size, hired labor, and machine use. The dataset in this paper, however, does not have sufficient information to analyze this relationship.

6. Consolidation program's results

In all regressions we have examined, *consolidation* shows no significant effect in both *harvest* and *profit*. This is different from all government reports in Ha Tinh, which have shown only positive results from the program. One reason for such differences can be a regional bias. In particular, in all regressions, *regions* significantly affect *harvest* and *profit*, with marginal productivity highest in region #1, followed by #4, then #3, and lowest in #2. From Table 1, we also see that the most consolidated land happened in region #1 then #4, followed by #3 and #2. This might have created a bias in government reports: consolidated plots are from regions with more favorable characteristics for rice farming, and their higher productivity (compared to provincial average data) might have come solely from regional characteristics instead of consolidation.

Because there is no longitudinal data to evaluate the program's effects, a series of t-tests and regressions were conducted instead. Table 11 reports the average plot size and harvest per acre of the four regions; Table 10 reports t-values of plot size comparisons between regions (each entry represents the t-value of a t-test between two regions). A positive value means the region on the first column has a higher mean value compared to the region of the first row, and vice versa. Plot sizes in these tables are the sizes before consolidation.

Table 10. T values for Area (t-test)

Region	#2	#3	#4
#1	7.12***	5.77***	1.12
#2		-1.76**	-5.20***
#3			-4.26***

Significant level: *** p<0.01, ** p<0.05, * p<0.1. Each entry represents the t-value of a t-test between two regions. A negative value means the region on the first column has a higher mean value than that of the first row.

Table 11. Regional plot area & harvest

Region	Average plot size	Average harvest per acre
#1	0.20	2.41
#2	0.14	1.72
#3	0.15	1.96
#4	0.19	2.19

Table 11 shows that the average farm size before consolidation is largest in region #1, followed by #4, #3, then #2, and the difference is mostly significant (Table 10). This confirms the hypothesis that the program was not implemented randomly. Regressions using model (2.2) were conducted in Region #1 and #4, the only two regions that went through consolidation. Table 12 shows that consolidation benefits Region #4 significantly for both *harvest* and *profit*, while having no effect in Region #1. More notably, even though not significant, the coefficient for *harvest* in Region #1 is negative. The linear-linear and linear-log regressions for these two regions yield similar results.

Table 12. Effects of consolidation on harvest and profit (plot level)

Parameters	Region #1		Region #4	
	Log(Harvest)	Log(Profit)	Log(Harvest)	Log(Profit)
Log(Cultivating area)	0.73***	1.13***	0.95***	1.33***
	(0.13)	(0.21)	(0.13)	(0.19)
Log(Labor)	0.0029	0.011	-0.012	-0.0068
	(0.015)	(0.024)	(0.020)	(0.032)
Log(Costs)				
Machine	0.33**	0.20	-0.066	-0.28
	(0.15)	(0.27)	(0.13)	(0.19)
Others	-0.025	-0.28**	0.17***	0.014
	(0.068)	(0.12)	(0.053)	(0.086)
Land quality (base = good)				
Average	-0.00072	0.0021	-0.015	-0.022
	(0.029)	(0.044)	(0.015)	(0.022)
Bad	-0.0044	-0.0046	-0.091***	-0.15***
	(0.034)	(0.048)	(0.024)	(0.041)
Divided into sub plots	-0.023	-0.042	-0.023	-0.010
	(0.021)	(0.031)	(0.052)	(0.078)
Consolidated	**-0.0011**	**0.00069**	**0.030***	**0.049***
	(0.014)	**(0.022)**	**(0.017)**	**(0.029)**
Irrigation (base = surplus supply)				
Sufficient	-0.026	-0.028	-0.017	-0.029
	(0.038)	(0.057)	(0.023)	(0.037)
Insufficient	-0.048	-0.063	(omitted)	(omitted)
	(0.040)	(0.059)		
Constant	0.60***	2.92***	0.73***	2.988***
	(0.14)	(0.22)	(0.13)	(0.19)
Observations	288	288	387	387
R-squared	0.97	0.92	0.95	0.89

Robust standard error in parentheses, with significant levels *** p<0.01, ** p<0.05, * p<0.1. Each column is a robust OLS regression using model (2.2) for *harvest* and (3.2) for *profit*. All regressions use adjusted machine costs. *Harvest*, *Profit*, *Cultivating area*, *Labor*, and *Costs* are in log values.

We have seen the positive relationship between land size and productivity, yet why has the land consolidation process shown no significant effect in Region #1? It turns out that, as reported in the interviews, while the land borders were erased and small plots were merged, there is no change in their actual land size. Different households own different small plots within a merged one, and they refused to trade, sell, or work together after the consolidation process. In other words, despite the land being merged, households continue to work individually within a larger plot. Several households even reported that their productivity has been lower than before consolidation. The detailed reason is that their plot of land is higher than their neighbors', and thus when borders were erased, their plots suffered from a loss of water supply.

This is not to say that the consolidation program should not be implemented, but that a more thorough approach is needed. Several distinct cultural and historical factors potentially hindered the success of this program. In Ha Tinh, family lands are normally owned for generations and have great sentimental values. In addition, there is a widely-believed superstition that there is a land god that attaches with the family land, which prevents them from trading or selling their rice plot. Many fear that once they trade to a different plot, there would be some types of incompatibility between themselves and the new land's god.

Another reason that hinders Ha Tinh's land consolidation process is its legal system, which is still in its infancy, lacks clarity, and is oftentimes infiltrated with corruption. The legal process gets even more complicated when each region consists of multiple plots of very small sizes. Because there are two consecutive rice farming seasons in Ha Tinh, and since the water supply ties closely to the exact timing of the year, any delay in the land selling, buying, or trading process can significantly affect productivity. Furthermore, there is no official assessment of land quality, which makes families hesitate to trade their plots. Therefore, in order to achieve the desired benefits of consolidation, these problems need to be properly addressed.

7. Conclusions

This paper uses a sample of 464 households and 1204 rice plots in Ha Tinh, Viet Nam, to study the relationship between land size and productivity. Productivity is measured using *harvest* and *profit* per acre of land. Overall, at household level, cultivating *area* exhibits increasing return to scale on both *harvest* and *profit*. At plot level, the relationship between *area* and *harvest* is mixed. *Area* shows increasing return on *harvest* when all plots are regressed, but this relationship disappears when each region is analyzed. However, *area* exhibits increasing return to scale on *profit* in all regions. This implies that increasing land size helps increase profit per acre of land, which is crucial for poverty alleviation.

To explain the positive relationship between *profit per acre* and *area*, a model was constructed using the Cobb-Douglas production function. The model was tailored to fit the regional characteristics of Ha Tinh where machines are mostly rented and is a perfect substitute for labor. To maximize profit, a household use machine(s) whenever the labor saved from such machine use can be utilized elsewhere to make more money compared to machine rental costs. This model also explains why in Sen's model, there is an inverse relationship between land size and productivity, and why it disappears in our sample.

This paper also studies the effects of the land consolidation program in Ha Tinh. The sample above was taken from four regions, and two of which went through consolidation partially. The program, however, does not show any definitive result for *harvest* and *profit*. In one region, consolidation has a significantly positive effect on *harvest per acre* and *profit per acre*, while it is unclear in another region. This result contradicts all government reports. There might have been a potential bias in those reports, which is caused by a non-random selection of plots to be consolidated.

There are several limitations to this study. First, because each region is surveyed by a different interviewer, there could be systematic bias within the dataset. Second, there is no longitudinal data, and thus the land consolidation program cannot be assessed accurately. And third, while the model developed in Section 4 is more inclusive than Sen's original model, it is still subjected to many underlying assumptions that might not hold in regions outside of Ha Tinh, Viet Nam.

References

Barrett, C. B. (1996). On price risk and the inverse farm size-productivity relationship. *Journal of Development Economics*, *51*(2), 193–215. https://doi.org/10.1016/S0304-3878(96)00412-9

Berry, R. A., & Cline, W. R. (1979). *Agrarian Structure and Productivity in Developing Countries: A Study Prepared for the International Labour Office Within the Framework of the World Employment Programme*. Johns Hopkins University Press.

Carletto, C., Savastano, S., & Zezza, A. (2013). Fact or artifact: The impact of measurement errors on the farm size–productivity relationship. *Journal of Development Economics*, *103*, 254–261. https://doi.org/10.1016/j.jdeveco.2013.03.004

Chadha, G. K. (1978). Farm Size and Productivity Revisited: Some Notes from Recent Experience of Punjab. *Economic and Political Weekly*, *13*(39), A87–A96.

Chen, Z., Huffman, W. E., & Rozelle, S. (2011). Inverse relationship between productivity and farm size: The case of China. *Contemporary Economic Policy*, *29*(4), 580–592. https://doi.org/10.1111/j.1465-7287.2010.00236.x

Desiere, S., & Jolliffe, D. (2017). *Land Productivity and Plot Size: Is Measurement Error Driving the Inverse Relationship?* (Vol. 8134). World Bank, Washington, DC. http://hdl.handle.net/10986/27638

Dyer, G. (1991). Farm size - farm productivity re-examined: Evidence from rural Egypt. *The Journal of Peasant Studies*, *19*(1), 59–92. https://doi.org/10.1080/03066159108438471

Dyer, G. (2013). *Class, state, and agricultural productivity in Egypt: A study of the inverse relationship between farm size and land productivity*. Routledge.

Eastwood, R., Lipton, M., & Newell, A. (2010). Chapter 65 Farm Size. In *Handbook of Agricultural Economics* (Vol. 4, pp. 3323–3397). Elsevier. https://doi.org/10.1016/S1574-0072(09)04065-1

Foster, A., & Rosenzweig, M. (2011). *Are Indian Farms Too Small? Mechanization, Agency Costs, and Farm Efficiency*. Brown University, Yale University.

Gollin, D. (2019). IFAD Research Series No. 34 - Farm size and productivity: Lessons from recent literature. *IFAD Research Series*.

Gollin, D., & Udry, C. (2017). Heterogeneity, Measurement Error, and Misallocation: Evidence from African Agriculture. *Journal of Political Economy*, *129*(1), 1–80. https://doi.org/10.1086/711369

Khan, M. H. (1979). Farm size and land productivity relationships in Pakistan.(Report). *Pakistan Development Review*, *18*(1), 69–77. https://doi.org/10.30541/v18i1pp.69-77

Khusro, A. M. (1964). Returns to Scale in Indian Agriculture. *Indian Journal of Agricultural Economics*, *19*(3–4), 1–33.

Kilic, T., Zezza, A., Carletto, C., & Savastano, S. (2013). *Missing(ness) in Action: Selectivity Bias in GPS-Based Land Area Measurements*. The World Bank. https://doi.org/10.1596/1813-9450-6490

Liu, Y., Barrett, C. B., Pham, T., & Violette, W. (2020). The intertemporal evolution of agriculture and labor over a rapid structural transformation: Lessons from Vietnam. *Food Policy*, *94*, Food policy. Volume 94 (2020). https://doi.org/10.1016/j.foodpol.2020.101913

Mahmood, H., Qasim, M., Khan, M., & Husnain, M. (2014). Re-examining the inverse relationship between farm size and productivity in Pakistan. *Journal of Animal and Plant Sciences*, *24*, 1537–1546.

Otsuka, K. (2013). Food insecurity, income inequality, and the changing comparative advantage in world agriculture. *Agricultural Economics*, *44*(s1), 7–18. https://doi.org/10.1111/agec.12046

Rada, N., Wang, C., & Qin, L. (2015). Subsidy or market reform? Rethinking China's farm consolidation strategy. *Food Policy*, *57*, 93–103. https://doi.org/10.1016/j.foodpol.2015.10.002

Rudra, A. (1968a). Farm Size and Yield per Acre. *Economic and Political Weekly*, *3*(26/28), 1041–1044.

Rudra, A. (1968b). More on Returns to Scale in Indian Agriculture. *Economic and Political Weekly*, *3*(43), A33–A38.

Rudra, A., & Sen, A. (1980). Farm Size and Labour Use: Analysis and Policy. *Economic and Political Weekly*, *15*(5/7), 391–394.

Savastano, S., & Scandizzo, P. L. (2017). *Farm Size and Productivity: A "Direct-Inverse-Direct" Relationship*. *8127*. http://hdl.handle.net/10986/27630

Scandizzo, P. L., & Savastano, S. (2017). Revisiting the farm size–productivity relationship: New evidence from sub-Saharan African countries. In *Agriculture and Rural Development in a Globalizing World*. Routledge.

Sen, A. (1962). An Aspect of Indian Agriculture. *Economic and Political Weekly*, *14*(4-5–6), 7–8.

Sen, A. (1964). Size of Holdings and Productivity. *The Economic Weekly, Indian Institute of Management*, 4.

Sheng, Y., Ding, J., & Huang, J. (2019). The Relationship between Farm Size and Productivity in Agriculture: Evidence from Maize Production in Northern China. *American Journal of Agricultural Economics*, *101*(3), 790–806. https://doi.org/10.1093/ajae/aay104

Srinivasan, T. N. (1972). Farm Size and Productivity Implications of Choice under Uncertainty. *Sankhyā: The Indian Journal of Statistics, Series B (1960-2002)*, *34*(4), 409–420.

Wang, J., Chen, K., & Huang, Z. (2015). Is small still beautiful? A comparative study of rice farm size and productivity in China and India. *China Agricultural Economic Review*, *7*(3), 484–509. https://doi.org/10.1108/CAER-01-2015-0005

Appendix A: Survey questions

General household-level questions	Household ID Name; Year born; Village; Region Full-time labor count; Part-time labor count Total plots owned; Total area Number of good/average/bad quality plots Total harvest (dried/before dried); Total sold Total costs Total household's alternative incomes
Largest-plot questions	Plot quality: good/average/bad Area: before and after consolidation Is the plot divided into subplots? Is the plot consolidated? If yes, does the household work with other families or separately? Irrigation: Surplus/Sufficient/Insufficient/Very insufficient Seed type Costs of seeds by kilogram; Total kilogram of seeds used Costs of fertilizer and other chemical treatment for land Weeding: Manual/Owned machine/Rented machine: unit cost Herbicide Spray: Manual/Owned machine/Rented machine: unit cost Pesticide Spray: Manual/Owned machine/Rented machine: unit cost Plowing: Manual/Owned machine/Rented machine: unit cost Seeding: Manual/Owned machine/Rented machine: unit cost Harvesting: Manual/Owned machine/Rented machine: unit cost Threshing: Manual/Owned machine/Rented machine: unit cost Other process(es): Manual/Owned machine/Rented machine: unit cost Other costs Total harvest (dried/before dried) Harvest compared with the previous year: More/The same/Less Reason for change in harvest
Smallest-plot questions	(Same as in Largest plot questions)
Other plot questions	(Same as in Largest plot questions)

Appendix B: Some measurements and calculations

In model (2.2) and (3.2):

$$labor = [(total\ full\ time\ labor) + \ \partial \times (total\ part\ time\ labor)] \div total\ area \times plot\ area$$

In which, ∂ is varied as 0.25, 0.5, and 0.75. “Total” here means the total number owned by the household. For example, $total\ full\ time\ labor$ means the total number of full-time labors within the household.

In model (3.1) and (3.2):

$$profit = harvest \times price\ of\ rice$$

In which, $price\ of\ rice$ is in millions Viet Nam Dong (VND), varying from 5, 7, 9, to 11 million VND per ton. This is because the price of rice fluctuates based on whether rice is sold directly on the market or to a third party.

In model (3.2):

$$machine\ costs = \sum_{i=1}^{8}(unit\ rent\ price\ of\ machine\ i) \times (area)$$

For each machine, there are three scenarios: either rented, owned, or no machine was used. In the case of “machine costs not adjusted,” $unit\ rent\ price\ of\ machine\ i = 0$ if households own the machine or no machine was rented. In the case of “machine costs adjusted,” $unit\ rent\ price\ of\ machine\ i = 0$ if no machine was rented. If market unit rent price varies, then for household-owned machines, unit price is the mean of market price within the local area.