Testing Household-Specific Explanations for the Inverse Productivity Relationship

Juliano J. Assunção and Luis H. B. Braido

The inverse relationship between land productivity and farm size is an old and puzzling empirical regularity. Most explanations for this relationship rely on market imperfections that jointly determine the farm size and the household's shadow price of some productive inputs. We use plot-level data from the ICRISAT/VLS to assess whether these household-specific theories can explain the puzzle. The data exhibit plots of different sizes being simultaneously cropped by the same household. The inverse relationship is shown to hold true with the same magnitude across the plots of each household, thus cross-household heterogeneity does not suffice to explain the puzzle.

Key words: development, farm size, inverse, IP relationship, productivity.

The inverse relationship between land productivity and farm size has puzzled economists for a long time.[1] Chayanov (1926) first documented that small farms produced more output per unit of land in Russia. The same result was found in India by Sen (1962), Bardhan (1973), and Rosenzweig and Binswanger (1993); and in Brazil, Pakistan, and Malaysia by Berry and Cline (1979). This inverse relationship is intriguing as there is a large body of literature that estimates constant returns to scale for agricultural production in different countries (e.g., Hayami and Ruttan 1970; Bardhan 1973; Berry and Cline 1979; Fulginiti and Perrin 1993). Moreover, in the absence of market failures, farmers would voluntarily subdivide their lands in order to increase productivity thereby eliminating the inverse relationship.

Understanding this empirical regularity has important policy implications. Land redistribution would increase the agricultural productivity if small plots were intrinsically more productive than large pieces of land. However, this would not be effective if the puzzle was just a spurious statistical result; and alternative policies would be required if the inverse relation were caused by market failures in the labor and credit markets.

Feder (1985) noted that a single market failure is typically insufficient to generate the inverse relationship. Under constant returns to scale, the explanations for the puzzle are likely to depend on market failures that simultaneously prevent land subdivision and distort the shadow price of some productive factors. Chayanov (1926), Sen (1962), Carter (1984), and Carter and Wiebe (1990) argue that peasant households apply family labor more intensively because the opportunity cost of their time is low. If imperfections in the labor market cause the peasant's shadow price of time to differ from the market wages, and if failures in the land–rental market prevent them from managing lands owned by others, then the peasant mode of production would generate an inverse relationship.

In an alternative vein, Bardhan (1973), Feder (1985), Eswaran and Kotwal (1986), and Taslim (1989) theorize that labor is subject to increasing marginal cost of supervision, thus the optimal land-to-labor ratio is higher for large landowners. This argument generates the inverse relation when the land market is imperfect.

Moreover, as noted by Srinivasan (1972), Rosenzweig and Binswanger (1993), and Barrett (1996), risk concerns could also generate the inverse relationship. Consider, for instance, a scenario in which incomplete insurance markets hinder full hedging against agrarian risks and failures in the land market

---

Juliano J. Assunção is with the Department of Economics, Pontifical Catholic University of Rio de Janeiro – PUC-Rio and Luis H. B. Braido is with the Graduate School of Economics, Getulio Vargas Foundation.

The authors are especially indebted to the editor, Christopher Barrett, and three anonymous referees for insightful remarks that considerably improved the paper. They are also thankful for comments from Steven Helfand and Rodrigo Soares. Financial support from CNPq is gratefully acknowledged.

[1] As is usual in this literature, the term productivity refers to the value of the output per unit of land.

Amer. J. Agr. Econ. 89(4) (November 2007): 980–990

DOI: 10.1111/j.1467-8276.2007.01032.x

prevent small farmers from increasing the cropped area. In this case, small farmers experience food-security stress and then overapply productive inputs on their lands.

Assunção and Ghatak (2003) state that the heterogeneity of farmers skills, coupled with credit-market imperfections in an environment with constant returns to scale and no labor-market imperfection, is another explanation for the puzzle. In equilibrium, the occupational choice is such that high-skilled peasants end up cropping small farms because they have higher opportunity costs to become wage workers. In this context, there is a range in which small farms are profitable for skilled peasants and not profitable for unskilled peasants. Farmer self-selection would then generate the inverse relationship.

In this article, we empirically assess these theoretical explanations. Our main contribution is noticing that all of these theories depend on cross-household heterogeneity, and this should equally affect the lands cropped by the same household. We analyze a very special data set—from the International Crops Research Institute for Semi-Arid Tropics (ICRISAT)—which contains households cropping multiple plots in each season. This allows us to investigate the inverse relationship across different plots cropped simultaneously by the same household.

If the inverse relationship were due to either the peasant mode of production or increasing supervision costs, then the plot-level productivity should be related to the total area managed by the household in each period, rather than the area of each particular plot. Contrary to this prediction, we show that plot productivity is inversely related to plot area and unrelated to the total area managed by the household.

Furthermore, according to all previous explanations, the inverse relationship is due to unobserved features in the household. We assess the importance of those explanations by using regression models with fixed effects to estimate the inverse relationship. We first use household fixed effects in order to account for household characteristics that are fixed over time. We then explore the fact that households harvest multiple plots in each season and introduce dummy variables for households in each period (season of the year), which accounts for unobserved household characteristics that are not fixed over time. The results show that the magnitude of the inverse relationship remains statistically unchanged. This evidence does not make the case for explanations based on cross-household heterogeneity.

Naturally, some of those explanations could be coupled with intrahousehold issues to generate the inverse relation. For instance, members with different characteristics could be allocated to supervise cropping activities in different plots of each household. We show, however, that the inverse relationship holds with the same magnitude when we restrict the analysis to plots cropped by households with one single adult member. Other intrahousehold issues—such as heterogeneous supervision costs due to geographical distance and differences in the cropping pattern across plots of each household—are also analyzed in a section of robustness checking. The results do not support those possibilities.

Our article is related to the work by Lamb (2003), which explores the ICRISAT/VLS sample at the aggregate farm level. In contrast, we explore the plot-level data to investigate the inverse relation across plots simultaneously cropped by the same household. This strategy leads us to obtain more conclusive results on the lack of importance of household-based explanations for the puzzle.

By rejecting household-based explanations for the inverse productivity relationship, our findings favor the literature that explains the puzzle by unobserved heterogeneity across plots and lands (e.g., Bhalla 1988; Bhalla and Roy 1988; Benjamin 1995; Chen, Huffman, and Rozelle 2003; Lamb 2003; and Kimhi 2006). In our view, future attempts to understand the economic content of the inverse relationship should focus on plot-specific unobservables as opposed to market failures affecting productivity at the household level. The policy implications of this research agenda depend crucially on understanding which are the specific unobservables associated with size at the plot level and which are the market forces behind this association.

Data

We use data from the longitudinal Village-Level Studies (VLS) conducted by the International Crops Research Institute for Semi-Arid Tropics (ICRISAT), in India, from 1975 to 1984. Six villages were initially selected from different agroclimatic zones, namely Aurapalle and Dokur (in the state of Andhra Pradesh); Kanzara, Kinkheda, Shirapur, and Kalman (in the state of Maharashtra). In 1980, the villages of Boriya Becharji and Rampura (in the state of Gujarat) were also included in the study. Farmers were randomly selected in each of these villages and resident investigators recorded information about all plots cultivated by them in each season of the year. Note that although the database is collected at the plot level, the household is the primary sampling unit. Farmers who moved out of the village during the period of data collection were randomly replaced. Further details about the data collection method can be found in Jodha, Asokan, and Ryan (1977) and Singh, Binswanger, and Jodha (1985).

The main data source is the ICRISAT's PS files, which contain plot-level information on cropping activities such as output value, cropped area, value of different nonlabor and labor inputs, estimated per acre value of the plot, irrigation, soil type, cropping pattern, village, year, and season. An auxiliary schedule, the C files, which contain information on household characteristic, is also used to measure the number of adult members in each household.

The ownership status is varied among the surveyed plots. We focus on plots cropped by their owners in order to avoid concerns about incentive problems sometimes associated with farms managed by tenants. The qualitative results, however, remain unchanged when we include these plots in the analysis.[2]

Farmers typically manage many different plots simultaneously. On average, each household harvests 5.6 plots per period. In order to study the importance of monitoring activities, we construct a variable describing the total area managed by the household in that period—i.e., for each plot, this variable sums the area of all plots cropped under the responsibility of the same household in that particular year and season. When constructing this variable, we include the plots rented by each household because, even if farmers faced incentive problems in rented farms, they would still expend part of their time with these plots. All results remain identical if we exclude the rented area from this variable.

Finally, some households have plots that produce no output in some seasons—the reported output is zero for about 6% of the plot observations. These are likely to be plots under rotation or temporarily abandoned after extreme shocks and are ignored in our analysis.[3] Table 1 describes the variables used throughout the article, and table 2 presents a few summary statistics.

The Inverse Relationship

Theoretical Framework

Consider a Cobb–Douglas production function. For each plot n, managed by household h, in period t (namely, the season of each year), one has

$$(1) \qquad Y_i = A_i T_i^{\alpha_t} K_i^{\alpha_k} L_i^{\alpha_l} \exp(\varepsilon_i)$$

where $i = (n, h, t)$ indexes the observations (plots, households, and periods); Y_i represents the total output; T_i is the cropped area; K_i and L_i represent the amount of nonlabor and labor input used; A_i is a technological factor that accounts for observable household and land characteristics as well as specific effects associated with different villages, years, seasons, and crops grown; and ε_i is an error term accounting for unobserved and idiosyncratic determinants of the output such as climatic shocks and infestations.

By multiplying Y_i, K_i, and L_i by their respective prices (namely, p, r, and w), we can represent the production function in monetary units, as follows:

$$(2) \qquad y_i = a_i T_i^{\alpha_t} k_i^{\alpha_k} l_i^{\alpha_l} \exp(\varepsilon_i)$$

where $y_i = pY_i$ represents the value of the output; $k_i = rK_i$ and $l_i = wL_i$ are the value of nonlabor and labor inputs (respectively); and $a_i = \frac{A_i p}{(r)^{\alpha_k}(w)^{\alpha_l}}$ is a price-adjusted technological term.

Consider now a competitive environment with no externality and constant return to scale, i.e., $\alpha_t = (1 - \alpha_k - \alpha_l)$. For any arbitrary plot size, farmers would maximize the expected profit, such that plot i's input choices would solve

$$(3) \qquad \max_{k_i, l_i} E\left(a_i T_i^{\alpha_t} k_i^{\alpha_k} l_i^{\alpha_l} \exp(\varepsilon_i) - k_i - l_i\right).$$

The optimal amount of nonlabor and labor inputs would be then given by

[2] Table A2 in the technical appendix (Assunção and Braido 2007) shows that our analysis is robust to the inclusion of plots managed under sharecropping and fixed rent.

[3] The ICRISAT/VLS documentation does not mention what could potentially explain the zero reported values. We elaborate on this topic in the technical appendix (Assunção and Braido 2007).

Table 1. Data Description

Variable	Description
Output	Nominal value of main output and by-products, measured in Indian rupees
Plot Cropped Area	Area of the plot actually cultivated (measured in acres)
Total Cropped Area	Area of all plots managed by the farmer in each season (include plots managed under ownership, fixed rent, and sharecropping)
Per Acre Land Value	Per acre value of the plot estimated by ICRISAT's investigators using information from village specialists about the potential sale value, topography, and location (nominal values expressed in 100 rupees per acre)
Irrigation Dummy	Dummy for irrigated plots
Soil Dummies	7.1% deep black; 33.9% medium black; 22.1% shallow black; 10.6% shallow red; 2.7% gravelly; 0.5% problem soil (saline, etc.); 10% sandy soil; 1.2% other soils; 11.9% undefined
Cropping Pattern	Qualitative variable (with 1,031 different codes) describing all products cropped in each plot
Main-Crop Dummies	Dummy variables constructed from the first letter of the cropping pattern code (which describes a general category for the dominant cropping product): 16.4% oilseeds; 52.4% cereals; 8.8% fiber crops; 0.5% garden crops; 15.1% pulses; 1% sugar cane; 4.4% vegetables and spices; 1.2% fodder crops; 0.2% missing information
Village Dummies	14% Aurepalle; 5.2% Dokur; 21.1% Shirapur; 15.9% Kalman; 14% Kanzara; 5.4% Kinkheda; 9.1% Boriya; 15.3% Rampura
Year Dummies	1975 (11%); 1976 (11.2%); 1977 (10.8%); 1978 (9.5%); 1979 (9.2%); 1980 (9.6%); 1981 (10.5%); 1982 (9.7%); 1983 (9.3%); 1984 (9.2%)
Season Dummies	35.19% planted from June to October; 59.22% from November to February; 5.34% from March to May; 0.21% perennial crops; 0.04% missing information
Adult Members	Number of members aged 18 years or more (in each particular year)

Note: Data from the ICRISAT/VLS. The primary sampling unit is the household, but the observations refer to plots managed by each household in each season of the year. Plots managed under fixed rent and sharecropping are not included in the analysis.

Table 2. Summary Statistics

Variable	Obs.	Mean	Std. Dev.	Min.	Max.
Per Acre Output	8,908	804.49	1,166.48	0.684	24,964
Plot Cropped Area	8,908	1.79	2.01	0	21
Total Cropped Area	8,908	13.08	14.25	0.08	83.87
Adult Members	7,319	3.12	1.41	1	8
Per Acre Land Value	8,908	34.38	24.92	0	160
Irrigation Dummy	8,908	0.34	0.47	0	1

Note: Data from the ICRISAT/VLS.

$$k_i^* = T_i\left(\alpha_k^{(1-\alpha_l)}\alpha_l^{\alpha_l} a_i E(\exp(\varepsilon_i))\right)^{\frac{1}{1-\alpha_k-\alpha_l}} \tag{4}$$

$$l_i^* = T_i\left(\alpha_l^{(1-\alpha_k)}\alpha_k^{\alpha_k} a_i E(\exp(\varepsilon_i))\right)^{\frac{1}{1-\alpha_k-\alpha_l}}. \tag{5}$$

Equation (2) can be written as

$$\frac{y_i}{T_i} = (\lambda a_i)^{\frac{1}{1-\alpha_k-\alpha_l}} \exp(\varepsilon_i) \tag{6}$$

where $\lambda = (\alpha_k)^{\alpha_k}(\alpha_l)^{\alpha_l}[E(\exp(\varepsilon_i))]^{(\alpha_k+\alpha_l)}$.

Equation (6) plays a central role in our empirical analysis. In principle, the technological term a_i and the production shocks ε_i should both be independent of the cropped area T_i. Under this assumption, the per acre value of the output should also be independent of the cropped area—that is, $\frac{y_i}{T_i} \perp T_i$.

This unconditional independence should be verified at the plot level as well as at the household aggregated level. To emphasize this point, assume that the technological factor a_i and shocks ε_i are common across all plots (n) cropped by each household (h) in a certain period (t); that is, $a_i = a_{h,t}$ and $\varepsilon_i = \varepsilon_{h,t}$, $\forall\, i = (n, h, t)$. This assumption is natural, for instance, if all those plots belong to a contiguous and homogenous farm. In this scenario, equation (6) could be aggregated as follows:

$$\frac{y_{h,t}}{T_{h,t}} = (\lambda a_{h,t})^{\frac{1}{1-\alpha_k-\alpha_l}} \exp(\varepsilon_{h,t}) \tag{7}$$

where $y_{h,t} \equiv \sum_{i \in \mathbb{I}_{h,t}} y_i$; $T_{h,t} \equiv \sum_{i \in \mathbb{I}_{h,t}} T_i$; and $\mathbb{I}_{h,t}$ represents the set of plots cropped by household h in period t.

The farm-level model (7) has been predominantly used in the literature because aggregate data are more frequently available. Again, the per acre productivity $\frac{y_{h,t}}{T_{h,t}}$ should (in principle) be unrelated to the cropped area $T_{h,t}$. We show, however, that, for the ICRISAT/VLS data, the inverse relationship is present in both econometric specifications. The disaggregated plot-level specification will be used in this article since it allows us to test for household-specific explanations.

Empirical Characterization

We start the empirical analysis by showing the nonparametric relationship between the logarithm of the per acre output and the logarithm of the cropped area. Similar to Barrett (1996), we show in figure 1 the curves obtained by the Nadaraya–Watson estimator with an Epanechnikov kernel of bandwidth 1.25. They show the existence of an inverse relationship between per acre output and cropped area, both at the plot level and at the aggregated household level. In both cases, the relationship is approximately log-linear.

The existence of the inverse relationship as depicted in figure 1 can potentially be explained by a negative correlation between the technological factor (A_i) and the cropped area. In this case, regressions controlling for observed regressors (such as land value, soil type, irrigation, village, year, season, and crop grown) are needed. Furthermore, as noted before, the cropped area could be negatively correlated to household-specific features that affect land productivity (such as farming skills, monitoring capability, stress-induced effort, etc.). In the remainder of the article, we examine these different possibilities in detail.

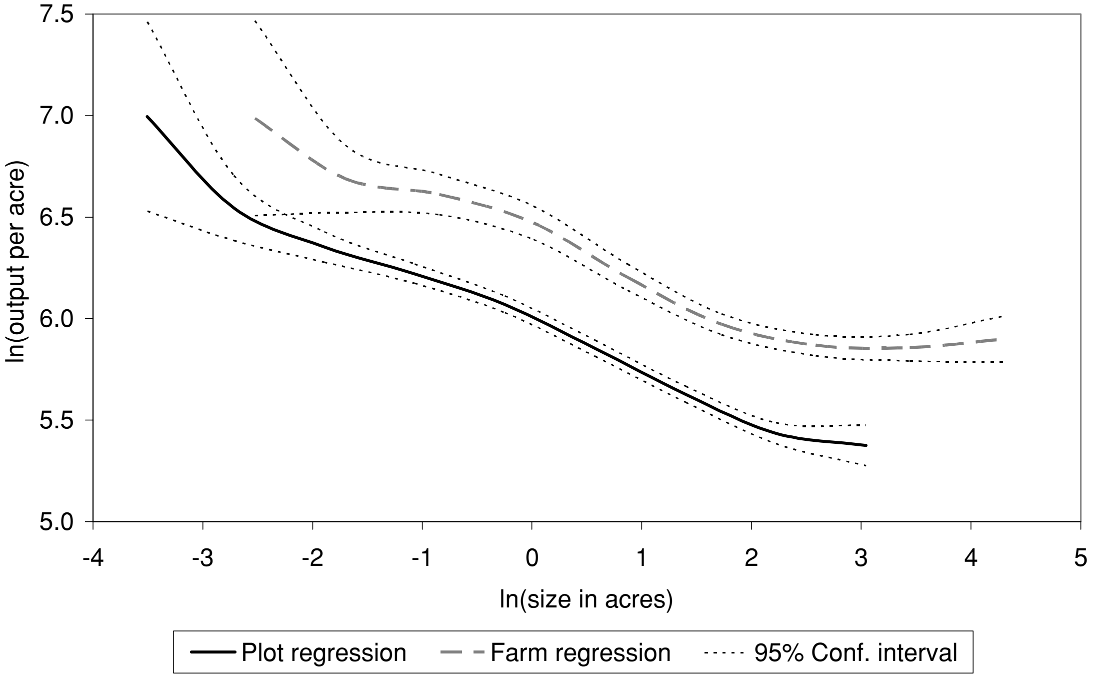
Testing for Household-Specific Explanations

Throughout the article, we consider the plot-level model of production and use the log-linear version of (6), namely,

$$\ln\left(\frac{y_i}{T_i}\right) = \beta_0 + \beta_1 \ln(a_i) + \varepsilon_i \tag{8}$$

where $\beta_0 = \frac{\ln(\lambda)}{1-\alpha_k-\alpha_l}$ and $\beta_1 = \frac{1}{1-\alpha_k-\alpha_l}$.

Table 3 presents OLS regressions for (8), where different variables are sequentially introduced to control for $a_i = \frac{A_i p}{r^{\alpha_k} w^{\alpha_l}}$. All



Note: Curves from Nadaraya–Watson regressions using the Epanechnikov kernel with bandwidth 1.25. Confidence intervals obtained from 200 bootstrap samples.

Figure 1. Inverse relationship

Table 3. Household-Based Explanations

OLS Dependent Variable: Log per Acre Output	Without Soil Quality (1)	With Soil Quality (2)	Total Area (3)	Fixed Effects I (Household) (4)	Fixed Effects II (Household & Period) (5)
Log plot cropped area	−0.305***	−0.160***	−0.180***	−0.167***	−0.160***
	(0.030)	(0.023)	(0.024)	(0.025)	(0.026)
Log total cropped area			0.053***	0.020	
			(0.017)	(0.018)	
Log per acre land value		0.386***	0.368***	0.340***	0.371***
		(0.048)	(0.048)	(0.052)	(0.069)
Dummies for irrigation and soil type	No	Yes	Yes	Yes	Yes
Constant and dummies for the main-crop, village, year, and season	Yes	Yes	Yes	Village dropped	Village, year, and season dropped
Number of observations	8,908	8,906	8,906	8,906	8,906
Number of groups				268	2,633
R^2	0.36	0.52	0.52	0.56	0.71

Note: Robust standard deviation (in parenthesis) account for the fact that farmers, rather than plots, are the primary sampling unit (* significant at 10%; ** significant at 5%; *** significant at 1%). Fixed effects I refer to 268 household dummies; while fixed effects II refer to 2,633 dummy variables generated through the iteration of the household and period codes (household-village, year, and season).

regressions include a constant term and dummies for the main crop, village, year, and season. These variables account for differences in A_i as well as for differences in prices across villages and periods (i.e., seasons of each year).[4] The first regression does not control for plot attributes, whereas the second one controls for land value, irrigation, and soil type. In both cases, the per acre output is negatively correlated with the plot cropped area. One must notice that the inverse productivity relation is considerably smoothed when one controls for observed plot attributes—the point estimates drop from −30% to −16%. This suggests that larger plots have worse productive attributes.

Next, we note that the explanations listed in the Introduction are all based on household-specific features (namely, the peasant mode of production, imperfect labor supervision, food-security stress, or unobserved farming skills). The first two theories are directly linked to the total area managed by the household (instead of the area of each plot). The other theories are also indirectly linked to the total area cropped by the household. We then introduce this variable into the regression in column (3) and notice that the coefficient associated with the total cropped area is positive (rather than negative), while the coefficient associated with the plot area remains unchanged (i.e., negative and with the same statistical magnitude).

In the fourth regression of table 3, we estimate equation (8) with household fixed effects (based on the household codes), exploring the variation in the area cropped across all plots cultivated by each household in all periods.[5] Observed and unobserved effects that are specific to the household and constant over time are considered in this regression. The coefficient of the total cropped area turns out to be nonsignificant, but there is virtually no change in the inverse relationship with respect to the plot size.

However, it could still be possible that the inverse relationship is driven by the unobserved household characteristics that are not fixed over time. The ICRISAT/VLS data provide a striking means of control for this possibility, since there are many farmers cultivating multiple plots with different sizes in each period (year and season). The estimates of the regression with household-period fixed effects (based on codes for the household, year, and season) are reported in the last column and reinforce the results of the previous regres-

[4] Since we consider nominal values throughout the article, the year dummies also account for inflation.

[5] Household fixed effects account for 23% of the variation of the logarithm of the per acre output. See the technical appendix (Assunção and Braido 2007) for detailed variance decomposition of the main variables.

sion.[6] The effect of the total cropped area on $\frac{y_i}{T_i}$ is not significant, while the inverse productivity relation still holds steady with the same magnitude.[7] These results show that theoretical explanations focusing exclusively on household-specific features do not account for the inverse productivity puzzle.

Intrahousehold Issues

Table 3 suggests that the inverse relationship between productivity and farm size is not explained by household-specific theories such as those based on the peasant mode of production, imperfect labor supervision, food-security stress, and skill bias. Conditional on land value and other plot attributes (namely, soil type and presence of irrigation), household-specific features have virtually no impact on the estimated magnitude of the inverse relationship. However, our empirical strategy depends on some identification hypotheses about how resources are allocated inside each household. Intrahousehold issues are now submitted to falsification tests.

Intrahousehold Allocation of Managerial Resources

We have implicitly assumed that a single manager controls all plots of each household. If so, household-period fixed effects could account for all unobservable characteristics of the manager. However, different members could be systematically assigned to plots of different sizes, based on a combination of managerial skills and soil characteristics or as a result of intrahousehold bargaining. This possibility is studied, for instance, by Carter (1984) and Udry (1996). If this is valid here, our interpretation for the results of table 3 needs to be revised.

The ICRISAT/VLS survey has no information about the actual manager of each plot. However, we can check the robustness of our results by restricting the previous analysis to households with the same numbers of adult members. For example, intrahousehold bargaining or managerial heterogeneity should not arise in households with a single adult member.

Table 4 presents the regressions for subsamples of households with 1, 2, 3, and 4 adult members. The econometric specification is the same that is used in the last regression of table 3—that is, it controls for soil attributes and household-period fixed effects. The inverse relationship is present in all four subsamples. Moreover, its magnitude does not change (when compared to the last regression of table 3). The p-values for the hypothesis test that the estimated coefficient for the log area cropped is statistically equal to -0.16 (our best estimate from table 3) are reported in table 4. In all subsamples, these estimated coefficients are either statistically equal to this value or significantly more negative. This does not support the hypothesis that the inverse relationship estimated in table 3 is due to differentiated intrahousehold allocation of managerial resources.[8]

Supervision Costs across Plots of Each Household

A key aspect of our empirical analysis is the fact that we observe farmers cultivating multiple plots in each given period (season of the year). This provides a powerful means of controlling for unobserved household characteristics. However, the possibility of having heterogeneity among the distances to each plot is another potential source of bias if these differences are systematically related with size.

The ICRISAT/VLS database does not contain information about the geographical dispersion of the plots cultivated by each household. Plots might be contiguous or not. If the largest plots available to each household are isolated and harder to be monitored, the results presented in table 3 would not necessarily be rejecting the labor supervision explanation. In this case, the plot size could be correlated with unobserved monitoring capability through the geographical dispersion of the plots.

We are not able to directly address this issue. We are restricted to an indirect assessment

[6] As is shown in the technical appendix (Assunção and Braido 2007), household-period fixed effects explain 57% of the variation in the logarithm of the per acre output.

[7] Our estimates for the inverse relationship are robust to the specifications with random effects. The Hausman's specification tests favor the models with fixed effects, thus we left these estimates in the technical appendix (Assunção and Braido 2007).

[8] When comparing the results across households with different numbers of adults, it is important to note that the gender composition is systematically related to the number of adult members. For instance, 61% of the single-adult households are headed by females, while 99% of the two-adult households are couples. For further details on gender composition, see the technical appendix (Assunção and Braido 2007).

Table 4. Intrahousehold Managerial Resources

OLS Dependent Variable: Log per Acre Output	1 Adult	2 Adults	3 Adults	4 Adults
Log plot cropped area	−0.191**	−0.260***	−0.126**	−0.169***
	(0.077)	(0.042)	(0.059)	(0.059)
Log per acre land value	0.381	0.570***	0.599***	0.401**
	(0.360)	(0.122)	(0.152)	(0.199)
Dummies for irrigation, soil type, and main crop	Yes	Yes	Yes	Yes
Constant and fixed effects (household & period)	Yes	Yes	Yes	Yes
Hypothesis test: β = −0.160 (*p*-value)	(0.689)	(0.020)**	(0.560)	(0.882)
Number of observations	399	2,689	1,854	1,173
Number of groups	136	879	571	295
R^2	0.75	0.74	0.69	0.69

Note: Robust standard deviation (in parenthesis) account for the fact that farmers, rather than plots, are the primary sampling unit (* significant at 10%; ** significant at 5%; *** significant at 1%). All regressions include a constant term, and fixed effects generated through the iteration of the household and period codes (household-village, year, and season).

Table 5. Number of Plots and Crop Mix

OLS Dependent Variable: Log per Acre Output	Number of Main Plots		Crop Mix (Cereals)	
	2 plots	3 plots	Jowar Sorghum	Paddy
Log plot cropped area	−0.171***	−0.124**	−0.261***	−0.148**
	(0.039)	(0.065)	(0.068)	(0.074)
Log per acre land value	0.361***	0.454***	0.140	0.310***
	(0.111)	(0.132)	(0.238)	(0.084)
Dummies for irrigation, soil type, and main crop	Yes	Yes	Yes	Yes
Constant and fixed effects (household & period)	Yes	Yes	Yes	Yes
Hypothesis test: β = −0.16 (*p*-value)	(0.732)	(0.608)	(0.135)	(0.902)
Number of observations	1,829	1,325	1,079	680
Number of groups	632	344	476	437
R^2	0.72	0.69	0.70	0.92

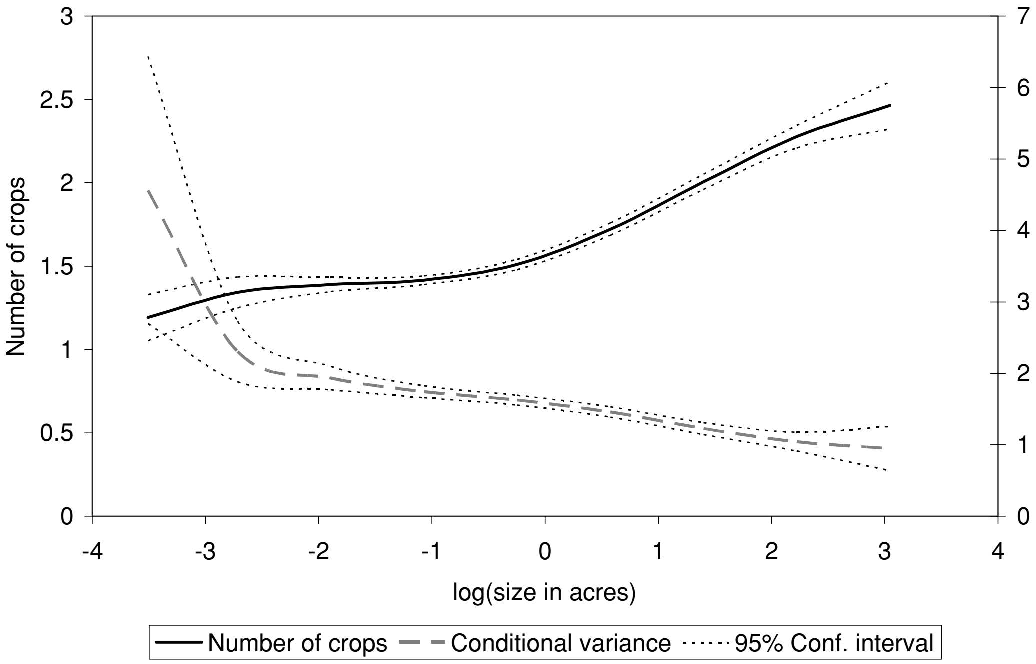
Note: Robust standard deviation (in parenthesis) account for the fact that farmers, rather than plots, are the primary sampling unit (* significant at 10%; ** significant at 5%; *** significant at 1%). All regressions include a constant term, and fixed effects generated through the iteration of the household and period codes (household-village, year, and season).

of this possibility. Our test is based on the assumption that the possibility of having significant differences in the supervision costs due to geographical dispersion of the plots depends on the number of main plots cultivated by the household (where the main plot is characterized by the first letter of the plot identification code). The underlying hypothesis is that the transportation costs of supervision are similar to farmers harvesting the same number of plots.

The first two columns of table 5 show the estimates of the log-linear model with household-period fixed effects, restricted to the subsamples of farmers cultivating two and three plots, respectively. We also test whether the coefficient of the logarithm of the cropped area is statistically equal to −0.16 (our best estimate from table 3). The *p*-values for these tests are displayed in table 5. The inverse relationship and its magnitude are robust to the possibility of heterogeneity in supervision costs due to the geographical dispersion of plots.

Intrahousehold Crop Mix

We also consider the possibility of aggregation bias caused by heterogeneity in the crop mix across plots of each household.



Note: Curves from Nadaraya–Watson regressions, using the Epanechnikov kernel with bandwidth 1.25. Confidence intervals obtained from 200 bootstrap samples.

Figure 2. Plot size, crop mix, and output risk

Bharadwaj (1974) first documented that plots with different sizes are typically used to produce different crops. The differences in the cropping patterns could also be related to insurance motives, as noted by Barrett (1996). Facing a risky environment, small and net buyer farmers might change the crop mix in order to avoid food-security stress. In this case, the inverse relationship could be generated by an aggregation bias determined by a systematic association between crop mix and size.

Figure 2 shows that the number of products cropped in each plot increases with the cropped area, while the plot risk decreases with the area cropped. When constructing figure 2, we used a qualitative variable (with 1,031 different codes) describing all products cropped in each plot; and our risk measure was constructed by computing the squared error of the regression of the logarithm of the per acre output against village dummies. The expected value of this variable gives the plot variance of the log per acre output, conditional on the village were the plot is located.

Figure 2 suggests that the largest plots might have been used for risk diversification. Hence, if there is a trade-off between crop risk and return, it is possible to observe small plots with higher (although riskier) returns than large plots. We then analyze subsamples of plots cropping the same product. We benefit from the existence of the large number of plots growing two types of cereals—namely, jowar/sorghum (1,079 observations) and paddy (680 observations). Columns (3) and (4) in table 5 show that the inverse relationship is statistically the same in the subsample of plots cropping only paddy or only jowar/sorghum. In both cases, the estimated coefficients are not statistically different from −0.16 (our best estimate from table 3) at the conventional levels of significance.

Conclusion

This article tests household-based theories for the inverse productivity puzzle using the ICRISAT/VLS data. Considering multiple plots cultivated by a single household in a given season, our evidence does not support explanations that hinge on household level characteristics such as the peasant mode of production or increasing supervision costs.

In our first estimation, a doubling of the plot area is associated with a 30% decrease in the output per acre. When we control for observed plot attributes, this coefficient is reduced to 16%. As suggested in the literature, observed plot attributes play an important role in explaining the inverse relationship, although it does not account for the entire effect.

A second set of regressions assesses explanations based on household-specific effects. Our results show that household-specific theories do not explain the puzzle. The inverse relationship remains virtually unchanged when we introduce household fixed effects and household-period fixed effects into the model. This latter result explores the presence of farmers cultivating multiple plots in the same year and season, which allows us to account for time-varying unobserved characteristics of the households, going beyond the traditional fixed-effect estimates.

In a robustness exercise, the inverse relationship is shown to hold with the same magnitude in subsamples containing: (*a*) plots cropped by households with one single adult member; (*b*) plots cropped by households cropping only two or three main plots; and (*c*) plots cropping the same main product.

A consequence of these results is that the content of the inverse relationship is related to unobserved characteristics of the plot rather than the household. Further analyses should then focus on the economic forces that associate the area cropped with plot-specific productive features.

[Received November 2005; accepted February 2007.]

References

Assunção, J.J., and L.H.B. Braido. 2007. "*AJAE* Appendix: Testing Household-Specific Explanations for the Inverse Productivity Relationship." Unpublished manuscript. Available at http://agecon.lib.umn.edu/.

Assunção, J.J., and M. Ghatak. 2003. "Can Unobserved Heterogeneity in Farmer Ability Explain the Inverse Relationship between Farm Size and Productivity?" *Economics Letters* 80:189–94.

Bardhan, P.K. 1973. "Size, Productivity and Returns to Scale: An Analysis of Farm-Level Data in Indian Agriculture." *Journal of Political Economy* 81:1370–86.

Barrett, C.B. 1996. "On Price Risk and the Inverse Farm Size-Productivity Relationship." *Journal of Development Economics* 51:193–215.

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

Berry, R.A., and W.R. Cline. 1979. *Agrarian Structure and Productivity in Developing Countries*. Baltimore: Johns Hopkins University Press.

Bhalla, S.S. 1988. "Does Land Quality Matter? Theory and Measurement." *Journal of Development Economics* 29:45–62.

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

Bharadwaj, K. 1974. *Production Conditions in Indian Agriculture*. Cambridge, UK: Cambridge University Press.

Carter, M. 1984. "Identification of the Inverse Relationship between Farm Size and Productivity: An Empirical Analysis of Peasant Agricultural Production." *Oxford Economic Papers* 36:131–45.

Carter, M., and K. Wiebe. 1990. "Access to Capital and Its Impact on Agrarian Structure and Productivity in Kenya." *American Journal of Agricultural Economics* 72:1146–50.

Chayanov, A.V. 1926. *The Theory of Peasant Economy*, In D. Thorner, B. Kerblay, and R.E.F. Smith, eds. Irwin: Homewood.

Chen, Z., W.E. Huffman, and S. Rozelle. 2003. "The Relationship between Farm Size and Productivity in Chinese Agriculture." Working paper, Dept. of Agr. Econ., Iowa State University.

Eswaran, M., and A. Kotwal. 1986. "Access to Capital and Agrarian Production Organization." *Economic Journal* 96:482–98.

Feder, G. 1985. "The Relation between Farm Size and Farm Productivity: The Role of Family Labor, Supervision and Credit Constraints." *Journal of Development Economics* 18:297–313.

Fulginiti, L.E., and R.K. Perrin. 1993. "Prices and Productivity in Agriculture." *Review of Economics and Statistics* 75:471–82.

Hayami, Y., and V.W. Ruttan. 1970. "Agricultural Productivity Differences among Countries." *American Economic Review* 60:895–911.

Jodha, N.S., M. Asokan, and J.C. Ryan. 1977. "Village Study Methodology and Resource Endowment of the Selected Villages in ICRISAT's Village Level Studies." *Economics Program Occasional Paper no. 16*. Hyderabad: International Crops Research Institute for Semi-Arid Tropics.

Kimhi, A. 2006. "Plot Size and Maize Productivity in Zambia: Is There an Inverse Relationship?" *Agricultural Economics* 35:1–10.

Lamb, R.L. 2003. "Inverse Productivity: Land Quality, Labor Markets, and Measurement Error."

Journal of Development Economics 71:71–95.

Rosenzweig, M.R., and H.P. Binswanger 1993. “Wealth, Weather Risk and the Composition and Profitability of Agricultural Investments.” *Economic Journal* 103:56–78.

Sen, A.K. 1962. “An Aspect of Indian Agriculture.” *Economics Weekly* Annual Number: 243–66.

Singh, R.P., H.P. Binswanger, and N.S. Jodha. 1985. “Manual of Instructions for Economic Investigators in ICRISAT’s Village Level Studies.” Technical Report, International Crops Research Institute for Semi-Arid Tropics.

Srinivasan, T.N. 1972. “Farm Size and Productivity: Implications of Choice under Uncertainty.” *Sankhya—The Indian Journal of Statistics* 34:409–20.

Taslim, M. 1989. “Supervision Problems and the Size-Productivity Relation in Bangladesh Agriculture.” *Oxford Bulletin of Economics and Statistics* 51:55–71.

Udry, C. 1996. “Gender Agricultural Production, and the Theory of the Household.” *Journal of Political Economy* 104: 1010–46.