# TEXTO PARA DISCUSSÃO

No. 500

Testing competing explanations for the
inverse productivity puzzle

Juliano J. Assunção
Luis H. B. Braido

**PUC**
RIO

## DEPARTAMENTO DE ECONOMIA
www.econ.puc-rio.br

# TESTING COMPETING EXPLANATIONS FOR THE INVERSE PRODUCTIVITY PUZZLE

JULIANO J. ASSUNÇÃO
LUIS H. B. BRAIDO

# Testing Competing Explanations for the Inverse Productivity Puzzle

Juliano J. Assunção[1]        Luis H. B. Braido[2]

February 9, 2005

[1]Department of Economics, Pontifical Catholic University of Rio de Janeiro; email: juliano@econ.puc-rio.br.

[2]Graduate School of Economics, Getulio Vargas Foundation; email: lbraido@fgv.br.

**Abstract**

We use plot-level data from ICRISAT to assess competing explanations for an old empirical regularity—the inverse relationship between land productivity and farm size. The presence of farmers who simultaneously crop multiple plots with different sizes is used to test (and reject) explanations based on household heterogeneity. The panel nature of the data is explored to test (and refuse) explanations based on plot fixed characteristics. We are then left with explanations based on time-varying plot features or measurement errors in the plot size. Theoretically, the input choices should reflect both plot-specific features and the true plot size. Empirically, the inverse relationship vanishes when we control for input use.

Keywords: Development, farm size, productivity. JEL Classification: C52, D82, O12, Q15.

# 1    Introduction

The inverse relationship between farm size and productivity has been puzzling economists for a long time.[1] Chayanov (1926) first documented that small farms produce more output per cropped area in the Russian agriculture. For India, the same evidence is found by Sen (1962), Bardhan (1973), and Rosenzweig and Binswanger (1993).[2] This inverse relation is considered a puzzle because there is a large body of the literature that estimates constant returns to scale for agricultural production in different countries (see Hayami and Ruttan, 1970; Bardhan, 1973; Berry and Cline, 1979; and Fulginiti and Perrin, 1993). The contents of this empirical regularity have important policy implications, especially for developing countries willing to improve agricultural efficiency through land reforms that reduce land concentration.

We examine five competing explanations for this puzzle. The first theory is based on imperfect labor supervision. Eswaran and Kotwal (1986) model an environment in which labor is subject to supervision problems and land provides a better access to credit. Because of increasing marginal cost of supervision, the optimal land-to-labor ratio is higher for large landowners, which generates the inverse relation.[3] The second theory relies on imperfections in the insurance market. Poor farmers experiencing risk as food-security stress would be induced to apply more labor input per acre than large landholders (see Srinivasan, 1972; and Barret, 1996). A third possible explanation is based on missing variables regarding farmer skills. Assunção and Ghatak (2003) use a simple self-selection argument to show that heterogeneity in skills coupled with credit-market imperfections would generate the inverse relationship in an environment with constant returns to scale and no labor-market imperfection.

Furthermore, the oft-observed inverse relationship could be simply a statistical artifact generated by the existence of plot-specific features which are privately observed by the farmer and affect the cropping-area decision (see Bhalla, 1988; Bhalla and Roy, 1988; Benjamim, 1995; and Lamb, 2003). For instance, land quality might be imperfectly described by the variables usually available for the econometrician, and farmers could have private information about the weather conditions and other shocks. In both cases, a spurious inverse relationship would appear if poor lands or bad shocks were optimally associated with the choice of extensive cultures—which are typically less productive. Finally, a fifth possible explanation found in the lit-

---

[1] As usual in this literature, the term productivity refers to output value per cropped area.

[2] Berry and Cline (1979) compute the ratio of productivity of small farms to the largest farms for many countries. The index is 5.63 in Northeast Brazil, 2.74 in Punjab, Pakistan, and 1.48 in Muda, Malaysia.

[3] See Feder (1985) for a related model.

1

erature relies on measurement errors in the cropped area (see Lamb, 2003). This problem would arise if the field researchers tended to systematically overestimate the size of larger plots relatively to the smaller ones.

In this paper we access a well-known agricultural database (the ICRISAT Village Level Studies, from India) in order to test among these five possible explanations— namely, (i) Labor-Supervision Capability; (ii) Food-Security Stress; (iii) Farmer-Skill Bias; (iv) Plot-Specific Features; and (v) Measurement Errors. Our contributions are based on two key strategies that were not pursued by previous studies. First, we notice that explanations (i)-(iii) are all based on farmer-specific unobservabilities. Supervision capability depends on the total area managed by the farmer in each season and on one's ability to perform supervision tasks—both features are constant for plots cropped in the same period. Similarly, labor-inducing stress and productive skills also depend on variables that are specific to the farmer in each period. We then explore the presence of farmers cultivating multiple plots in the same period to control for those farmer-specific features. This peculiarity of the data allows us to go beyond the usual fixed-effect specifications and control for household characteristics that evolve over time. Second, we use production theory to show that plot-specific features (that were privately observed by the farmer) as well as the true plot size should be embedded in the input choice. Therefore, the amount of labor and nonlabor inputs per acre contain useful missing information regarding land quality, plot-specific shocks, and measurement errors in the cropped area. The guidelines of the empirical analysis and our main results are as follows.

*Labor-Supervision Capability*

The ICRISAT data display disaggregated information on each plot cropped by a farmer. Under the labor-supervision explanation, the productivity of each plot should be related to the total area managed by the farmer in each period, rather than the area of each particular plot. Contrary to this prediction, we show that the plot productivity is inversely related to the plot area and positively related to the additional area managed by the farmer.

Furthermore, we use the presence of farmers cultivating multiple plots within each period and over periods to control for differences in the supervision ability across farmers. The regressions with farmer fixed effects and fixed effects per farmer in each period display a significant inverse relation between productivity and the plot area and a positive and nonsignificant relationship between productivity and the additional area cropped by the farmer.

*Food-Security Stress and Farmer-Skill Bias*

According to these explanations, the inverse relation is caused by unobserved

characteristics of the farmer (respectively, stress and skill). We use farmer fixed effects to account for household characteristics that are fixed over periods. We also introduce dummy variables for each farmer and period. Since many farmers in the sample harvest multiple plots in each season, this latter exercise controls for farmer unobserved characteristics even if they were not fixed over periods. The results show that the plot productivity remains inversely related to the plot area—rejecting these two theories as possible explanations for the puzzle.[4]

*Plot Heterogeneity and Measurement Errors*

We also investigate the possibility of the puzzle being generated by plot characteristics which are privately observed by the farmers. The data display plot-level information in different periods and the cropped area of each plot is not constant over time. We show that the inverse productivity puzzle is still significant after introducing plot fixed effects. This shows that the inverse relationship is not due to unobserved land quality unless it evolved over time. Unfortunately, we cannot directly account for time-varying land unobservabilities since, unlike farmers, there is only one plot observation per period.

We remain then with two possible explanations for the puzzle: (i) a potential interaction between privately observed land characteristics and shocks (such as rainfalls or monsoon arrivals) that are not constant over time; and (ii) measurement errors in the plot size. Being unable to disentangle these potential explanations, we attempt to test whether either of them could explain the puzzle. Privately observed land quality and shocks affect the optimal choice of nonlabor and labor inputs. Hence, input choices would partially reveal plot-level heterogeneity that was observed by the farmer but not by the econometrician. Moreover, inputs must also be correlated to the true plot size. Therefore, introducing inputs into the regressions would reduce the effect of both plot-specific features and measurement errors in the plot size. In fact, the inverse-relation puzzle vanishes when we introduce the per acre value of nonlabor and labor inputs as control variables. The negative correlation between output per acre and farm size is due to a negative correlation between farm size and the intensity of input use.[5]

Our findings are very conclusive in rejecting all explanations based on farmers' unobservables and point out to the same direction of some recent papers relating the inverse relationship to unobserved plot-specific features and potential measurement errors in the cropped area (see Chen, Huffman, and Rozelle, 2003; Kimhi, 2003; and

---

[4] Rosenzweig and Binswanger (1993) also reject the theory based on food-security stress. They show that the productive advantage of small farmers is decreasing with risk in the ICRISAT farms.

[5] It is worth stressing that this exercise identifies correlations between economic variables without establishing causal relations.

Lamb, 2003).

The remainder of this paper is organized as follows. Section 2 describes the data; Section 3 shows that the puzzle is not explained by household unobservabilities and plot-fixed effects. Section 4 uses standard production theory to link plot unobservabilities to optimal input choices and shows that the puzzle becomes statistically nonsignificant when one controls for input use. Concluding remarks are presented in Section 5.

## 2 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 in different agroclimatic zones, namely: Aurapalle and Dokur (in the state of Andhra); Kanzara, Kinkheda, Shirapur, and Kalman (in the state of Maharashtra). In 1980, the villages of Boriya 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. Thus, although the database is collected at the plot level, the household is the primary sampling unit here. Farmers who moved out of the village during the period of data collection were randomly replaced. Details about the data collection method can be found in Jodha, Asokan, and Ryan (1977) and Singh, Binswanger, and Jodha (1985).

We use a schedule (the PS files) containing plot-level information on cropping activities—e.g., output value, cropped area, estimated per acre value of the plot, irrigation, value of labor and nonlabor inputs, village, season, year, and cropping pattern. There are different ownership statuses among the surveyed plots. We focus on plots cropped by their owners in order to avoid concerns about incentive problems sometimes associated with farms cropped by tenants. Farmers typically manage many different plots simultaneously.[6] In order to study the importance of monitoring activities, we construct a variable describing the additional area managed by the household in that period—i.e., for each plot, this variable sums the area of all other plots cropped under its farmer responsibility in that particular season and

---

[6]On average, each farmer harvests 3.4 plots per period. The maximum number of plots cultivated in the same season in our sample is 18.

year.[7] Moreover, some households have plots delivering no output in some seasons.[8] These are likely to be plots under rotation or temporarily abandoned after extreme shocks (such as infestations). They are ignored in our analysis—mainly because of the log-linear specification used throughout the paper. Table 1 describes all variables used throughout the paper, and Table 2 displays a few summary statistics.

[**Tables 1-2**]

## 3  Household Unobservabilities and Plot Fixed Effects

We model the output of a plot by a Cobb-Douglas production function, as follows:

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

where $i$ indexes the observations (plots in different 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 $\varepsilon_i$ is an error term accounting for unobserved characteristics of farms and farmers, climatic shocks, and infestations.

Multiplying quantities by their respective prices ($p_y$, $p_k$, and $p_l$) and dividing both sides by $T_i$, the production function can be represented in monetary units per acre as follows:

$$y_i = a_i T_i^{\beta} k_i^{\alpha_k} l_i^{\alpha_l} \exp(\varepsilon_i), \tag{2}$$

where $y_i = p_y \frac{Y_i}{T_i}$ represents the value of plot $i$'s per acre output; $k_i = p_k \frac{K_i}{T_i}$ and $l_i = p_l \frac{L_i}{T_i}$ are the per acre value of nonlabor and labor inputs (respectively); $a_i = \frac{A_i p_y}{p_k^{\alpha_k} p_l^{\alpha_l}}$ is a price-adjusted technological term; and $\beta = \alpha_t + \alpha_k + \alpha_l - 1$. Notice that $\beta = 0$ when technology displays constant return to scale.

The log-linear version of the production function is:

$$\ln(y_i) = \ln(a_i) + \beta \ln(T_i) + \alpha_k \ln(k_i) + \alpha_l \ln(l_i) + \varepsilon_i. \tag{3}$$

Under a Cobb-Douglas technology, the profit-maximizing amount of each input is an exponential function of $a_i$, so that $\ln(k_i)$ and $\ln(l_i)$ are colinear with $\ln(a_i)$—see details in Section 4. In this case, and if all elements in $a_i$ were observed, one could neglect $k_i$ and $l_i$ when testing the sign of the parameter $\beta$. This is the type of

---

[7]When constructing this variable, we include the plots rented by each household. We argue that even if farmers faced incentive problems in rented farms, they would still expend part of their time in these plots. The results are unchanged if we exclude the rented area from this variable.

[8]There are 813 plots producing no output out of 11,517 plots sampled.

exercise usually performed in the literature, which typically finds a negative value for $\beta$ (the inverse farm size-productivity relationship). We follow this methodology in this section.

We include many variables to control for $a_i = \frac{A_i p_y}{p_k^{\alpha_k} p_l^{\alpha_l}}$. Dummy variables for the main cropping pattern, village, year, and season are introduced in the regressions. They account for differences in the technological factor $(A_i)$ as well as differences in prices across villages and periods. The plot value (per acre), the irrigated fraction of the plot, and dummies for the type of soil account for the observed land characteristics. Besides plot area $(T_i)$, we also introduce the additional area cultivated by the farmer in that particular period. This variable is used to test whether the puzzle is due to the labor-supervision explanation.

The results are displayed in Table 3. One must notice from the first regression that productivity is negatively affected by the plot area and positively related to the additional area managed by the farmer. This does not support imperfect labor supervision as an explanation for the puzzle. According to this theory, landowners have limited monitoring capabilities which cannot be replicated as the farm size increases (see Eswaran and Kotwal, 1986; and Feder, 1995). If the effective labor is affected by supervision, we should observe a negative correlation between output per acre and the additional cropped area, which is not the case in our sample.

From the second regression, one must notice that the inverse productivity relation is considerably reduced (from $-33\%$ to $-18\%$) when one controls for observed land quality. The coefficient for the additional cropped area is still positive and significant. This exercise suggests that larger plots display lower observed quality.

The rejection of the supervision-based argument is further stressed in the next two regressions, where we introduce household fixed effects and fixed effects per household and period. In the third regression of Table 3, we estimate equation (3) with household fixed effects, exploring the variation among all plots cultivated by each farmer in all periods. Any aspect of the household (observed or not) which is constant over time is considered in this regression. Although the coefficient of the additional cropped area turns out to be nonsignificant, there is virtually no change in the inverse relationship. Moreover, one must also worry about unobserved household characteristics that are not fixed over time. The ICRISAT data provide a striking means to control for this, since there are many farmers cultivating two or more plots with different sizes within a season. The estimates of the regression with farmer-period fixed effects are reported in the fourth column and reinforce the results of the previous regression—i.e., the effect of the additional cropped area on $y_i$ is not significant while the inverse productivity relation still holds with the same magnitude. These regressions reject the explanations based not only on imperfect

6

labor supervision but also on any other household-specific feature, such as food-security stress and farmer-skill bias.

Finally, the last regression in Table 3 shows that plot fixed effects are also ineffective to explain the inverse relationship. Since the household is the primary sampling unit, the plots in the sample are linked to the same farmer over years. Therefore, this exercise shows that the inverse productivity relation is unrelated to plot fixed characteristics, household fixed unobservabilities, or any combination of these explanations. Unfortunately, unlike farmers, there is only one plot observation per period and, therefore, we cannot directly evaluate the importance of plot-specific features that evolve over time.

[**Table 3**]

# 4  Time-Varying Plot Features and Measurement Errors

From Section (3), we are left with two possible explanations for the puzzle, which are based on: (i) privately observed land quality and shocks that are not constant over time; and (ii) measurement errors in the plot size. We perform here an exercise to check whether either of these two features could in fact make the inverse relationship compatible with constant return to scale.

We consider an environment with constant return to scale ($\beta = 0$), a complete financial market, and no externality. Under these assumptions, the plot size does not affect the expected profit. Moreover, given an arbitrary plot size, farmers choose nonlabor and labor inputs in order to maximize expected profit in each plot. Considering the production function in (1), the optimal input level must solve:

$$\max_{k_i, l_i} E\left(a_i k_i^{\alpha_k} l_i^{\alpha_l} \exp\left(\varepsilon_i\right) - k_i - l_i \mid \mathcal{I}_i\right), \tag{4}$$

where $\mathcal{I}_i$ is the information set available for plot $i$'s farmer at the time of choosing the inputs. (Notice output and inputs are expressed in monetary units per acre.)

Hence, the optimal amount of nonlabor and labor inputs must satisfy:

$$k_i^* = \left(\left(\alpha_k\right)^{1-\alpha_l}\left(\alpha_l\right)^{\alpha_l} a_i \theta_i\right)^{\frac{1}{1-\alpha_k-\alpha_l}}, \tag{5}$$

and

$$l_i^* = \left(\left(\alpha_l\right)^{1-\alpha_k}\left(\alpha_k\right)^{\alpha_k} a_i \theta_i\right)^{\frac{1}{1-\alpha_k-\alpha_l}}, \tag{6}$$

where $\theta_i = E\left(\exp\left(\varepsilon_i\right) \mid \mathcal{I}_i\right)$.

The term $\theta_i$ captures private information regarding land characteristics and shocks. It will be constant whenever farmers have no private information about

7

$\varepsilon_i$ (that is $\mathcal{I}_i = \mathcal{I}$ for all $i$). In this case, $\ln\left(k_i^*\right)$ and $\ln\left(l_i^*\right)$ could be expressed as a linear combination of $\ln\left(a_i\right)$ and a constant. However, if farmers accessed private information regarding land quality or shocks, the input choices would partially reveal them. Furthermore, for each plot $i$, the optimal input choices ($L_i^*$ and $K_i^*$) should also be colinear with the correctly measured cropped area.

Given these considerations, we introduce the per acre value of labor and nonlabor inputs into the regressions. Since inputs are endogenous, this exercise does not identify the coefficients of the production function. It does however take privately observed plot features into account when computing the conditional correlation between farm size and output per acre.

The result is presented in Table 4, where the last column of Table 3 is reproduced in order to facilitate comparisons. Notice that the estimated coefficient for cropped area drops from $-24\%$ (significant at a $1\%$ level) to a nonsignificant value of $+1\%$. The information embedded in the inputs is able to account for the puzzle. Large farms are less productive because they use inputs less intensively, and the input choices are likely to reflect plot features (land quality, shocks, or size) that are not perfectly observed by the econometrician.

[**Table 4**]

# 5    Conclusion

This paper tests competing explanations for the inverse productivity puzzle using the ICRISAT village level studies. In our first estimation, an increase of $1\%$ in the cropped area is associated with a $33\%$ decrease in the output per acre. When we control for observed land quality, this coefficient is reduced to $18\%$. As suggested by the literature, observed land quality plays an important role in explaining the inverse relationship, although it does not account for the entire effect.

A second set of regressions tests theories based on household-specific effects such as labor-supervision capability, food-security stress, and farmer skill. We strongly reject all those theories. The inverse relationship remains virtually unchanged when we introduce farmer fixed effects and farmer-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.

We then consider the unobserved characteristics of each plot. We show that plot fixed effects are also unable to solve the puzzle—the point estimate for the inverse relationship is even higher ($24\%$). Therefore, plot-specific features that evolve over time (such as privately observed land quality and weather conditions) as well as

8

measurement errors in the plot size remained as the possible explanations for the puzzle.

Our last estimation is based on the fact that input choices should be related to both privately observed features of the plot and the true plot size. We show that the coefficient for the inverse relation vanishes when we introduce inputs as control variables—suggesting that either of these facts could in fact make the estimated inverse relationship compatible with constant return to scale.

# References

[1] Assunção, J. J. and Ghatak, M. (2003) "Can unobserved heterogeneity in farmer ability explain the inverse relationship between farm size and productivity?," *Economics Letters*, 80, 189-194.

[2] 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.

[3] Barret, C. B. (1996) "On price risk and the inverse farm size-productivity relationship," *Journal of Development Economics*, 51, 193-215.

[4] Benjamim, D. (1995) "Can unobserved land quality explain the inverse productivity relationship?," *Journal of Development Economics*, 46, 51-84.

[5] Berry, R. A. and Cline, W. R. (1979) *Agrarian structure and productivity in developing countries*, Johns Hopkins University Press, Baltimore, MD.

[6] Bhalla, S. S. (1988) "Does land quality matter? Theory and measurement," *Journal of Development Economics*, 29, 45-62.

[7] Bhalla, S. S. and Roy, P. (1988) "Misspecification in farm productivity analysis: the role of land quality," *Oxford Economic Papers*, 40, 55-73.

[8] Chayanov, A. V. (1926) *The theory of peasant economy*, In D. Thorner et al. (eds.), Irwin, Homewood, IL.

[9] Chen, A. Z., Huffman, W. E. and Rozelle (2003) "The relationship between farm size and productivity in Chinese agriculture", Iowa State University, mimeo.

[10] Eswaran, M. and A. Kotwal (1986) "Access to capital and agrarian production organization," *Economic Journal*, 96, 482-98.

[11] 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.

[12] Fulginiti, L. E. and Perrin, R. K. (1993) "Prices and productivity in agriculture," *Review of Economics and Statistics*, 75, 471-82.

[13] Hayami, Y. and Ruttan, V. W. (1970) "Agricultural productivity differences among countries," *American Economic Review*, 60, 895-911.

[14] Jodha, N.S., Asokan, M. and Ryan, J.C. (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.

[15] Kimhi, A. (2003) "Plot size and maize productivity in Zambia: the inverse relationship re-examined," Hebrew University of Jerusalem, mimeo.

[16] Lamb, R. L. (2003) "Inverse productivity: land quality, labor markets, and measurement error," *Journal of Development Economics*, 71, 71-95.

[17] Rosenzweig, M. R. and H. P. Binswanger (1993) "Wealth, weather risk and the composition and profitability of agricultural investments," *Economic Journal*, 103, 56-78.

[18] Sen, A. K. (1962) "An aspect of Indian agriculture," *Economics Weekly*, Annual Number, 243-66.

[19] Singh, R. P., Binswanger, H. P. and Jodha, N.S. (1985) "Manual of instructions for economic investigators in ICRISAT's village level studies," Technical Report, International Crops Research Institute for Semi-Arid Tropics.

[20] Srinivasan, T. N. (1972) "Farm size and productivity: Implications of choice under uncertainty," *The Indian Journal of Statistics*, 34, 409-420.

## Table 1. Data Description

| Variable | Description |
|---|---|
| Output (per acre) | Value of main output and byproducts per area cropped (in Rupees per acre, R$/acre) |
| Plot Cropped Area | Area of the plot actually cultivated (in acres) |
| Additional Cropped Area | Area of all other plots (owned and rented) which, in that season, were under the responsibility of the household that cultivates the present plot |
| Labor Input (per acre) | Per acre value of family and hired labor, i.e., number of hours worked times village wages for males, females, and children (in R$/acre) |
| Nonlabor Input (per acre) | Per acre value of seeds, fertilizers, pesticides, organic and inorganic manures, plus rental value of owned and hired bullock, and other machineries (in R$/acre) |
| Plot Value (per acre) | Per acre value of the plot (in 100R$) estimated by ICRISAT's investigator using information about topography, location, etc., obtained from a village specialist |
| Irrigation | Total irrigated area in the plot divided by the cropped area |
| Soil Dummies | 7.1% deep black; 33.9% medium black; 22.1% shallow black; 10.6% shallow red; 2.7% gravelly; .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; .5% garden crops; 15.1% pulses; 1% sugar cane; 4.4% vegetables and spices; 1.2% fodder crops; .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; .21% perennial crops; .04% missing information |

*Note:* Data from ICRISAT. The unit of analysis is a plot in a certain season and year. Only owned plots were included.

**Table 2. Summary Statistics**

| Variable | Obs. | Mean | Std. Dev. | Min. | Max. |
|---|---|---|---|---|---|
| Output (per acre) | 8908 | 804.49 | 1166.48 | 0.684 | 24964 |
| Plot Cropped Area | 8916 | 1.79 | 2.01 | 0 | 21 |
| Additional Cropped Area | 8916 | 13.08 | 14.25 | 0.08 | 83.87 |
| Labor Input (per acre) | 8908 | 158.63 | 187.64 | 1 | 3064 |
| Nonlabor Input (per acre) | 8908 | 332.72 | 524.44 | 0 | 16478.8 |
| Plot Value (per acre) | 8916 | 34.36 | 24.92 | 0 | 160 |
| Irrigation | 8907 | 0.34 | 0.47 | 0 | 3 |

*Note:* Data from ICRISAT.

**Table 3. Household Unobservabilities and Plot Fixed Effects**

Dependent Variable: Log Output (per acre)

| | OLS without Soil Quality | OLS with Soil Quality | Farmer Fixed Effects | Farmer-Season Fixed Effects | Plot Fixed Effects |
|---|---|---|---|---|---|
| Log Plot Cropped Area | -0.331*** (0.033) | -0.180*** (0.024) | -0.167*** (0.025) | -0.158*** (0.031) | -0.242*** (0.052) |
| Log Additional Cropped Area | 0.071*** (0.026) | 0.051*** (0.017) | 0.019 (0.018) | 0.000 (0.000) | -0.021 (0.028) |
| Log Plot Value (per acre) | | 0.377*** (0.048) | 0.347*** (0.053) | 0.379*** (0.082) | 0.131 (0.090) |
| Irrigation | | 1.086*** (0.053) | 1.031*** (0.055) | 1.081*** (0.072) | 0.815*** (0.076) |
| Soil Dummies | No | Yes | Yes | Yes | No |
| Constant and Dummies for the Main-Crop, Village, Year, and Season | Yes | Yes | Yes | Yes | Yes |
| **Number of Observations** | 8,908 | 8,905 | 8,905 | 8,905 | 8,905 |
| **Number of Groups** | | | 268 | 2,633 | 3,956 |
| **R²** | 0.36 | 0.51 | 0.56 | 0.71 | 0.80 |

*Note*: Robust standard deviation in parenthesis (* significant at 10%; ** significant at 5%; *** significant at 1%); 268 household clusters used.

## Table 4. The Role of Labor and Nonlabor Inputs

Dependent Variable: Log Output (per acre)

| | Plot Fixed Effects without Inputs | Plot Fixed Effects with Inputs |
|---|---|---|
| Log Plot Cropped Area | -0.242***<br>(0.052) | 0.014<br>(0.037) |
| Log Additional Cropped Area | -0.021<br>(0.028) | 0.015<br>(0.023) |
| Log Labor Input (per acre) | | 0.879***<br>(0.044) |
| Log Nonlabor Input (per acre) | | 0.055<br>(0.035) |
| Log Plot Value (per acre) | 0.131<br>(0.090) | 0.081<br>(0.068) |
| Irrigation | 0.815***<br>(0.076) | 0.117*<br>(0.061) |
| Constant and Dummies for the Main-Crop, Village, Year, and Season | Yes | Yes |
| **Number of Observations** | 8,905 | 8,892 |
| **Number of Groups** | 3,956 | 3,955 |
| **R$^2$** | 0.80 | 0.86 |

*Note*: Robust standard deviation in parenthesis (* significant at 10%; ** significant at 5%; *** significant at 1%); 268 household clusters used.