Oxford Economic Papers 40 (1988), 55–73

MIS-SPECIFICATION IN FARM PRODUCTIVITY ANALYSIS: THE ROLE OF LAND QUALITY

By SURJIT S. BHALLA *and* PRANNOY ROY*

A POPULAR "stylized fact" contained in studies on agriculture in developing countries is that there exists a strong inverse relationship between farm size and land productivity i.e. that small farmers are more productive per unit of land than large farmers. This fact has endured the test of time, and space. It has been observed for several countries, for different time periods, and for different technologies. In short, it has justifiably earned the status of "conventional wisdom".[1]

This "fact" has also endured in discussions because of its implications for theory and policy. On theoretical grounds, neo-classical assumptions would suggest that land productivity and farm size should be uncorrelated; a negative correlation , therefore, implies imperfections in factor markets.[2] Regarding policy, if small farmers are indeed more productive, then it follows that land redistribution (desirable from an equity standpoint) would also lead to a higher output (desirable from an efficiency standpoint).[3]

As far as factor markets are concerned, researchers have concentrated on the *labor* market for explanations of differing farm productivity. It is contended that labor dualism is a major contributor to the inverse relationship.[4] Small farmers are more productive because they face a lower cost of labor than large farms; hence, small farmers apply more labor on their land and a higher land productivity (but lower labor productivity) is obtained. Indeed, the inverse relationship has been used to test (albeit indirectly) the presence/absence of labor dualism.

Notwithstanding the near universal acceptance of the inverse relationship

* The Policy Group. This paper was written while the authors were at the World Bank and Delhi School of Economics, respectively. We would like to thank T. N. Srinivasan and Sanjiv Kumar for very useful suggestions on an earlier draft dated Nov. 1983. Our thanks to Julie Nelson and Kenneth Boese for patient and excellent research assistance. We would also like to thank two anonymous referees for their useful comments. The views expressed in this paper are the authors and not those of the institution to which they belong.

[1] The inverse relationship was first discovered by Chayanov (1966) in his analysis of Russian agriculture in the twenties. See Sen (1966), Rudra (1968) and Bardhan (1973) for observations on India; Cline (1970) and Berry (1973) for results on Latin America and Griffin (1974) for observations on Asia. Also, see Bhalla (1979) for results pertaining to the early seventies Green Revolution period. This is only a select listing; the literature on this subject is vast.

[2] Other possibilities include the presence of a non-homothetic production function. This is considered unlikely for Indian agriculture.

[3] In static terms, land reform can also improve the opportunity set of farmers (e.g. access to education, capital markets) and thus add to output gains. However, the total gain in output might be reduced in a dynamic context if saving rates increase with the level of permanent income. This is empirically supported for Indian farmers in Bhalla (1980).

[4] Other possible factor market imperfections, in capital and land, are presumed to help the large farmer and thereby make the relationship positive.

and the labor dualism explanation, there are two other possible explanations for the observed negative correlation. First, Srinivasan (1972) has theoretically shown that under considerations of uncertainty, (and no imperfections in input markets) it would be optimal for a small farmer to apply more labor and achieve a higher land productivity than a larger farmer. Second, the inverse relationship may simply be a reflection of differing land fertility between the small and large farms.

This paper will concentrate on empirically evaluating the certitude of the second explanation. Indeed, the land fertility argument is not a new one. Several authors have discussed it,[5] and Khusro (1964) in an important study, showed how adjustments for land quality diminish, and eliminate, the inverse relationship. Khusro's study, unfortunately, was based on highly aggregate land productivity and land tax revenue (i.e. land quality) data; further, adjustments to land taxes are conducted only once every twenty to thirty years. Consequently, it has been impossible to verify Khusro's important results with farm level data. Recent studies with village level data do tend to support the land fertility argument[6] though Carter's (1984) study on an Indian village reaffirms the existence of an inverse relationship.[7]

As its basis, this paper makes use of the unique Fertilizer Demand Survey (FDS) data set, with obserations on 21,500 farm households in different parts of India (See Appendix I for details). This data set contains detailed information on the soil characteristics of farms—an occurrence which is critical for assessing how much of the observed inverse relationship can be explained by purely *exogenous* land quality differences. By exogenous is meant factors that are outside the decision making process of individual farmers. In this category, the intrinsic qualities of land (e.g. soil color) are included, but excluded are several aspects of fertility (e.g. tubewell irrigation) that can be influenced by choice i.e. are endogenous.

If the inverse relation remains strong after controlling for these exogeneous differences, then our results, we feel, will provide the strongest case to date for arguing that higher land productivity on small farms is a permanent and universal feature of Indian agriculture and that its most likely cause is labor market imperfection. If on the other hand, land quality differences are a major (and perhaps sole) explanator of the inverse relation then a complete *re-evaluation* of the farm-size productivity debate is called for. Such results would undermine the labor dualism proponents[8] as well as

[5] Sen (1962) utilizes a Malthusian explanation, Bhagwati–Chakravarty (1969) refer to the phenomenon of "distress sales" and Bhalla (1986b) incorporates labor supervision costs in helping explain the presumed negative correlation between farm size and land quality.

[6] See Sen (1975), Bliss and Stern (1982).

[7] Carter (1984) finds that the inverse relationship weakens when intra-village soil quality differentials are included but inter-village differences do not alter it. Carter's findings, however, remain unconvincing mainly because he lacks direct data on soil quality. Moreover, since his study relates to only one region in India it is not clear whether his results can be generalized.

[8] It should be emphasized that the existence/absence of an inverse relationship is only an *indirect* test of the hypothesis of labor dualism. Lack of labor use data prevents direct tests. However, this drawback (indirect testing) is common to most studies which explain dualism via the inverse relationship.

the efficiency argument for land reform. Of course, the equity argument for land reform remains completely unaffected by any such explanation.

2. Land productivity differences—specification

Conventionally, the inverse relationship is documented via estimation of an equation like (1):

$$\ln Y' = \ln \frac{Y}{H} = a + b \ln H + e \tag{1}$$

where Y is value of output and H represents farm size.[9] The coefficient b, is presumed to equal zero; a negative b signifies the existence of the inverse relationship.

It is important to differentiate (1) from a production function. The latter would include all variable and fixed inputs.[10] Equation (1), on the other hand, is a reduced form equation which assumes that farmers have adjusted to their environment by making the relevant choices and that the exogenous non-choice determinants of farm productivity (e.g. weather, soil, ability etc.) are uncorrelated with farm size. Further, the technology and factor prices faced by the farmers are identical. Hence, an estimate of b is unbiased and should equal zero, since factor productivities are not assumed to be a function of the use of any factor.

A few comments on the included variables Y and H are in order. Differences in output can result from three factors (i) differences in cropping pattern (ii) differences in cropping intensity and (iii) differences in yields of various crops. The concern of this paper is with the estimation of a reduced form equation like (1) and, hence, the composition of Y, and differences in H (due to cropping intensity) are ignored. Only an answer to the following simple question is attempted: Is land productivity (Y/H) related to cultivated land (H)?

An assumption implied by equation (1) is that land heterogeneity, if present, is uncorrelated with farm size H. The consequences of a non-zero correlation can easily be interpreted; in particular, a negative correlation would bias the estimate of b downwards. In other words, *land quality differences, if negatively correlated with H, may have yielded the observed inverse relationship for Indian agriculture.* The issue can only be settled empirically. If data on land quality, Q, were available, then the following equation could be estimated:

$$\ln \frac{Y}{H} = a' + b' \ln H + c \ln Q + e' \tag{2}$$

[9] H represents the area available for cultivation i.e. it is equal to land owned plus land leased in minus land leased out. In this paper, the terms farm size, land size and net cultivated area are used interchangeably.

[10] The FDS data set does not allow the estimation of a production function since data on labor inputs are not available.

Estimate of b', rather than b, establishes the nature of the inverse relationship after controlling for exogenous factors and, by implication, tests for the importance of the effects of labor dualism.

Lack of data has prevented estimation of equation (2) for any varied sample of farm households; though as mentioned earlier, village based studies have estimated such an equation. This paper proposes to estimate equation (2) for the different parts of India.

What variables should be contained in land quality Q? A primary consideration is that the Q variables should be exogenous. Small irrigation facilities such as tubewells and pumpsets come within the purview of the investment decisions of individual farmers and are clearly endogenous. On the other hand, larger irrigation works, such as canals, tanks and village wells, are typically not the investment decision of the individual farmers but arise from investment by the government or the village group. Thus, such irrigation can be considered exogenous. This assumption is not valid if farms are frequently purchased and sold. Given that a vast majority of farms in India have not been acquired in the market, the assumption of exogeneity of canal and tank irrigation may be warranted. In any case, equations are estimated for both including, and excluding, the presence of exogenous irrigation.

An example of how it can be misleading to aggregate all sources of irrigation is revealed by the following exercise. Table 1 reports for each state in India the fraction of land irrigated by exogenous/endogenous irrigation, and the sample correlations of irrigation with farm size. For ten such states, there is a significant negative correlation between exogenous irrigation and farm size; only for one state (Assam) is the correlation significantly positive.[11] Endogenous irrigation, on the other hand, shows a significant negative correlation for only one state, and a positive correlation for two states. Thus, with exogenous irrigation, a negative correlation exists, but no relationship exists between farm size and endogenous irrigation.

In addition to the source of irrigation, it is fortunate that the Fertilizer Demand Survey collected detailed information on other aspects of soil quality. For each farm, the following information is available: (i) soil texture, SOIL1—coded as sand, loam, light clay and heavy clay (ii) soil color, SOIL2—coded as red, black, grey and yellow (iii) depth of soil, SOIL3—less than 1 foot; between 1 and 3 feet, and more than 3 feet. These characteristics are clearly exogenous, and fixed. In addition, the survey collected information on three other soil variables—soil salinity, surface drainage and rate of percolation. However, these latter variables can be affected by farmer decisions, and cannot be considered exogenous.

In addition, the number of fragments per farm F was also added as an

[11] Unless otherwise indicated, all tests in this paper are conducted at the 5% level of significance.

TABLE 1
Irrigation and Farm Size

State	Farm size acres	Exogenous Irrigation % of land	Endogenous Irrigation % of land	Correlations with farm size: Percent Exogenous Irrigation	Percent Endogenous Irrigation
1. Andhra	7.84	23.3	12.8	−0.226*	0.067*
2. Assam	3.77	1.9	0.3	0.065*	0.011
3. Bihar	7.08	8.8	18.5	−0.014	0.112*
4. Gujarat	9.99	5.3	23.7	−0.089*	−0.153*
5. Haryana	10.52	33.1	29.6	−0.041	−0.026
8. Karnataka	6.15	18.5	7.5	−0.191*	−0.004
9. Kerala	2.21	41.2	6.8	0.003	−0.067
10. M.P.	9.56	9.3	4.1	−0.141*	0.075
11. Maharashtra	7.82	7.3	6.0	−0.095*	−0.006
12. Orissa	7.16	9.2	0.4	0.024	−0.007
13. Punjab	8.81	20.8	44.2	0.068	−0.027
14. Rajasthan	11.95	16.6	6.9	−0.146*	0.006
15. Tamil Nadu	5.49	15.7	32.6	−0.240*	0.025
16. U.P.	5.42	23.1	44.3	−0.108*	−0.008
17. W. Bengal	5.10	24.6	10.4	−0.056*	0.044
India	7.14	15.6	16.9	−0.103*	−0.004

Notes

1. For definitions of exogenous and endogenous sources of irrigation, see text.
2. Starred values represent significance at the 5% level.

additional indicator of exogenous land quality. Bhagwati–Chakravarty (1969) stress the importance of this variable as a quality variable—specifically, it is hyypothesized that land fragmentation occurs due to "distress sales" of land by small farmers and it is likely that the relatively bad land gets transferred in this process. Hence, the greater the number of fragments, the more likely it is that the land is of lower average quality. Bardhan (1973) uses number of fragments as a state or fixed variable in his estimation of production functions in Indian agriculture. The decision to include F was therefore made on the grounds that it might include information not contained in other variables. (The efficiency loss in case F was an extraneous variable is negligible for such a large data set).

The exogenous irrigation and soil quality variables together form an important addition to the estimation of farm productivity equations. Econometrically, these variables substitute for the land quality variable Q in equation (2). As mentioned earlier, according to theory, no specific functional relationship is postulated between output per acre and farm size because the two are not presumed to be related in the first place. Equation (2) postulates a constant elasticity relationship between Y/H and H.

Alternative hypotheses about the causes of the relationship between land and farm productivity will dictate different estimating forms. Since the dual labor market hypothesis is the most important (assumed) explanator of the inverse relationship, a functional form dictated by this hypothesis may be appropriate. If farms are classified as family (all family labor), intermediate (some hired labor) and capitalist (all hired labor), then marginal product of labor is highest (and equal) on large farms, declines as share of family labor increases, and flattens out for family farms, i.e. suggesting a semi-logarithmic relationship between Y/H and H. The contrast between the log-log and semi-log models is illustrated by the different land elasticities yielded by the two models. In the semi-log form, $(Y/H = A + B \log H)$ the elasticity of output per acre with respect to farm size is not constant (as in the log-log form) but is dependent on output/acre i.e. it is equal to $B(Y/H)$. Thus, at lower levels of yield/acre (larger farms) the decline is more rapid than at higher levels (smaller farms) for the semi-log form.[12]

3. Emprical results

The basic equation estimated in this paper is the following, Model A;[13]

$$Y'_{ij} = \frac{Y}{H} = \alpha_i + \beta_i \ln H_{ij} \tag{3A}$$

where i represents a homogenous region and j an individual farm. In addition, soil quality variables are incorporated in Model B:

$$Y'_{ij} = \alpha_i + \beta_i \ln H_{ij} + \delta_{1_i} S1_{ij} + \delta_{2_i} S2_{ij} + \delta_{3_i} S3_{ij} + \gamma_i F_{ij} \tag{3B}$$

whree $S1$ to $S3$ are vectors representing soil type, soil color, and soil depth, respectively, and $\delta_1 - \delta_3$ the corresponding vector of coefficients, and F represents fragmentation (number of fragments per farm).

Finally, if the assumption of the exogeneity of certain types of irrigation is accepted, then Model C results:

$$Y'_{ij} = \alpha_i + \beta_i \ln H_{ij} + \delta_{1_i} S1_{ij} + \delta_{2_i} S2_{ij} + \delta_{3_i} S3_{ij} + \gamma_i F_{ij} + \lambda_i I_{ij} \tag{3C}$$

where I represents the fraction of cultivated area that is irrigated by canals, tanks, community wells and ponds.

Though comparatively exhaustive, equation (3C) is still misspecified. It ignores interaction amongst the soil variables and between the soil variables and irrigation. In other words, the assumption is made that farms of a given soil type have the same impact on yield irrespective of soil color or soil depth. Further, the impact of irrigation is not conditional on the nature of the soil. The empirical importance of these interaction affects can easily be

[12] Though results are generally reported for the semi-log form, tests indicated that the choice of functional form does not affect the results.

[13] For ease of comparison with other studies, results for the log-log formulation are also reported (Table 5). No difference in the results are observed.

incorporated into equation (3C). Before estimating such an equation, however, it should be noted that even such an extended equation may be mis-specified. In particular, such an equation ignores differences in the soil moisture content of the farms. In a country like India, the weather is a major variable affecting yields. However, soil moisture content is an inherently intractable variable. This drawback is handled in this paper at two levels—firstly, equation (3C) is estimated for households organized by districts—a classification that assures homogeneity of agro-climatic conditions. However, as discussed in the next section, this classification reduces degrees of freedom. A second, and more 'efficient' method of estimation is to classify households into agronomic zones. As outlined in Prasad (1983), India can be classified into 78 agronomic zones on the basis of three criteria: (i) moisture index, (ii) thermal index, and, (iii) soil class.[14] Though this zonal classification captures some of the same factors as SOIL1, SOIL2 and SOIL3, its addition to the specification of equation (3C) should, *a priori*, be quite important.

Thus, equation (3C) is estimated at four levels—district, sub-zonal, zonal and state.[15] Obviously, the district classification is the most preferable, but it has the drawback of fewer observations.[16] As one pools (or aggregates), the land quality coefficients become less precise, unless one allows them to vary with each level of aggregation.[17] The importance of whether, for example, black soils have the same effect on yields irrespective of their presence in a dry or wet zone, can be tested by comparing the coefficients at the different aggregation levels. However, other "cross-effects" are difficult to estimate since the possible configurations are numerous (e.g. blacksoil with loamy texture and depth 1–3ft, compared with depth $<$ 1ft etc.). The estimation strategy employed is dictated by interest in the coefficient of land, β_i. The district, sub-zonal and zonal regressions allow for sufficient 'variability' in the slope coefficients, and the value of β under these diverse circumstances may be sufficient to establish the nature of the inverse relationship.

In essence, therefore, the question being asked is whether observations can be pooled across districts and/or agronomic zones. F tests reject the assumption of homogeneity of coefficients at the zonal and district level for

[14] The moisture index is based on the formula $(P - P.E.) \times 100/P.E.$ where P is the mean annual precipitation received and $P.E.$ is the mean annual potential evapotranspiration computed with Thornwaite's formula. The index represents 8 zones: <-80, extremely dry, to >100, extremely wet. The thermal index is based on mean annual temperature and soil class represents soil texture e.g. arid soils, alluvial soils, vertisoils etc.

[15] The sub-zonal classification deserves an explanation. If soil type is the most important soil variable, then farms can be aggregated according to the dominant soil type *within* a zone, i.e. all farms in zone 1 with sandy soil are in one group, and all farms in zone 1 with loamy soil are in another. This kind of classification is a "half way" measure between the district and zonal classifications.

[16] Models A, B and C were estimated for only those classifications that had a minimum of forty observations.

[17] Throughout this paper, the term aggregation is means to refer to pooling of observations into a broader regional classification with the restriction that coefficients of the included variables are constant across the classifications which range from district, to sub-zone to zone.

TABLE 2
Farm Size and Productivity—State Equations

State	*Mean yield Rs/acre*	*Coefficient of log land Model A*	*Land Elasticity Model A*	*Coefficient of log land Model C*
1. Andhra	1203	−267	−0.22	−234
		(−7.9)		(−6.6)
2. Assam	646	−63	−0.10	−64
		(−3.7)		(−3.6)
3. Bihar	636	−57	−0.09	−59
		(−4.7)		(−4.5)
4. Gujarat	417	−109	−0.26	−100
		(−7.5)		(−6.2)
5. Haryana	211	−111	−0.09	−87
		(3.7)		(−3.0)
6. Himachal	929	−164	−0.18	−167
		(−4.7)		(−4.9)
7. J & K	468	−159	−0.34	−171
		(−7.2)		(−6.8)
8. Karnataka	1255	−379	−0.30	−256
		(−6.1)		(−3.9)
9. Kerala	2425	−1032	−0.42	−1063
		(−3.8)		(−3.5)
10. M.P.	256	−46	−0.18	−36
		(−7.8)		(−5.5)
11. Maharashtra	557	−193	−0.35	−212
		(−9.4)		(−9.9)
12. Orissa	351	−76	−0.22	−89
		(−7.8)		(−9.1)
13. Punjab	1124	−39	−0.03	−33
		(−1.6)		(−1.2)
14. Rajasthan	453	−107	−0.24	−54
		(8.8)		(−4.0)
15. Tamil Nadu	1297	−294	−0.23	−270
		(−5.6)		(−4.5)
16. U.P.	942	−101	−0.11	−101
		(−7.6)		(−7.0)
17. W. Bengal	989	−132	−0.13	−157
		(−5.1)		(−53)

Notes

1. Land elasticity is computed at the mean and is for Model A.

2. $\bar{R}^2$ are not reported. In general they were quite small i.e. in the range 0.05 to 0.10.

3. Figures in parentheses represent *t*-statistics.

4. Model A and Model C refer to equations 3A and 3C in the text.

5. The land output elasticity according to the conventional log-log models reported in Table 5.

the intercept term, α, for all the states in India except Kerala; for the coefficients of SOIL1, F-tests reject homogeneity at the zonal level for 8 out of 14 states.[18] In general, the assumption of homogeneity of intercept and slope coefficients is not warranted. Thus, the models were estimated at four different levels of pooling (aggregation).[19]

First, the results at the broadest aggregation level—the state—are reported. Table 2 contains these results for seventeen states of India. Incorporation of soil quality variables (Model C) diminishes in some cases the land elasticity but in general these elasticities remain strongly negative. Thus, the inverse relationship appears confirmed at the state level; soil quality variables affect the estimated elasticities, but do not eliminate their significance.

A reclassification of observations according to their residence in an agronomic zone rather than a state (a more heterogeneous and political classification) has a substantial impact on the universality of the inverse relation. The traditional specification (model A) has a significant negative coefficient for LAND in 54 of the 78 agronomic zones, much fewer than the 95 per cent (16 out of 17 states) reported at the state level (Table 3). When soil quality and exogenous irrigation are incorporated in the agronomic zone estimation (model C), the cases for which an inverse relation exists drops further to only 44 per cent.

If soil type is the most important soil variable, then farms can be aggregated according to the dominant soil type within a zone, i.e. all farms in zone 1 with loamy soil are in a particular sub-zone. This classification ensures greater homogeneity at the sub-zone level since it controls both for soil characteristics as well as climatic factors. Of the 142 subzones, the traditional specification reports an inverse relation in only 73 and this drops to 53 (37 percent) when other exogenous factors (model C) are incorporated.

Finally, the models were estimated at the lowest level of aggregation (or pooling)—a district. According to Model A, (estimated for 176 districts across India) less than half of the districts report a significant negative relationship. Moreover, when soil quality variables are controlled for (Model C), a significant inverse relation exists for only 29% of the districts.

Table 4 summarises the information on the existence, or lack of it, of the inverse relation for various specifications of the land productivity model (Models A, B, and C) and for different levels of pooling (state, zone, sub-zones and district).

A clear pattern emerges from Table 4. The smaller the unit of estimation (i.e. the more valid the assumption of homogeneous environment) the lower

[18] All F-tests are at the 5% level of significance and are available from the authors.

[19] The estimated results for the various models, and the different classifications, are not reported but are available from the authors. Only the more important findings are reported below. Table 3, nevertheless, summarizes the various results.

TABLE 3

Summary Results of Farm Productivity Equations—By District, Zone, and Subzone (Number of classifications with a negative coefficient on log land)

State	Zones				Subzones				Districts			
	Number of zones	*Model A*	*Model B*	*Model C*	*Number of subzones*	*Model A*	*Model B*	*Model C*	*Number of districts*	*Model A*	*Model B*	*Model C*
Andhra Pradesh	7	4	5	4	13	6	7	5	14	6	7	7
Assam	5	2	1	1	7	2	1	1	9	3	1	1
Bihar	4	3	3	3	12	6	4	4	14	10	4	2
Gujarat	5	4	3	3	12	7	5	6	10	5	4	4
Haryana	3	2	1	1	4	3	0	1	5	2	1	1
Himachal	2	2	2	2	4	1	1	1	3	2	2	2
J and K	1	1	1	1	4	4	4	4	4	2	3	2
Karnataka	5	4	3	3	6	4	3	3	15	4	4	4
Kerala	1	1	1	1	2	2	1	2	0	—	—	—
M.P.	9	5	4	4	15	7	7	6	19	8	6	5
Maharashtra	5	5	5	5	8	5	4	4	6	3	2	2
Orissa	4	3	3	3	7	2	3	3	9	3	3	2
Punjab	3	1	0	0	6	1	0	0	9	2	0	0
Rajasthan	7	5	2	2	10	6	2	2	12	6	4	3
Tamil Nadu	6	4	4	4	12	7	8	7	11	7	6	6
U.P.	5	5	5	5	12	7	7	7	22	10	5	6
W Bengal	6	3	3	2	8	3	1	1	14	4	4	4
India	78	54	46	44	142	73	58	53	176	83	56	51

Notes

1. The order in which the results are reported (zones, sub-zones and districts) reflects decreasing levels of pooling/aggregation. The results for the most heterogeneous classification, the state, are reported in Table 2.
2. Models A, B and C reflect equations 3A, 3B and 3C respectively.
3. All significant tests are at the 5% level of confidence.

TABLE 4
Percent of Regions Reporting a Negative Relationship Between Farm Size and Farm Productivity

Level of Pooling (Aggregation)	Traditional Model A % regions	Extended Model C % regions
State	94	94
Agronomic zones	69	56
Sub-zones	51	37
Districts	47	29

Notes

1. The 5% level of significance was used as a basis for classifying whether the coefficient of log land was significant in the different equations.
2. Detailed state wide results are given in Table 2.

the number of samples reporting a negative relationship for the traditional specification—Model A. Further, the addition of soil quality variables always decreases the number of aggregation units reporting an inverse relationship. Thus, these results support the contention that soil quality is an important variable in explaining farm productivity differences between small and large farms.

At a micro level, a closer look at one of the states, Karnataka, shows why the results differ at the district and state level. Figure 1 plots the average farm size and average output per acre of each district (marked by crosses); the line through these averages represents the slope derived from the coefficient of log land in the traditional specification (Model A) estimated for each district.

Figure 1 shows that even though only four (districts 2, 4, 8, and 13) of the 15 districts report an inverse relation, there is a negative relation between the mean values of output per acre and farm size of each district. Consequently, when pooling across districts this systematic bias in inter-district variations results in a 'spurious' and strong negative relation being observed at the state level.

The pattern of relationship observed in Figure 1 has not gone unobserved. It has been noted, and documented, that areas with high rainfall and inherently high land productivity are densely populated and characterized by small holdings (e.g. the coastal belt and the 'wet' states of West Bengal and Kerala). Conversely, dry areas with low levels of yield usually have large farms and are sparsely populated (e.g. Rajasthan, Madhya Pradesh). Thus, at the all-India level there is a strong inverse relationship which has long been noticed but has evoked little interest. This lack of interest stems mainly from the conclusion that the *cause* of the inverse relation is the heterogeneity of natural (or exogenous) factors and hence of little policy relevance. But, at the less aggregate (e.g. state) level, the assumption

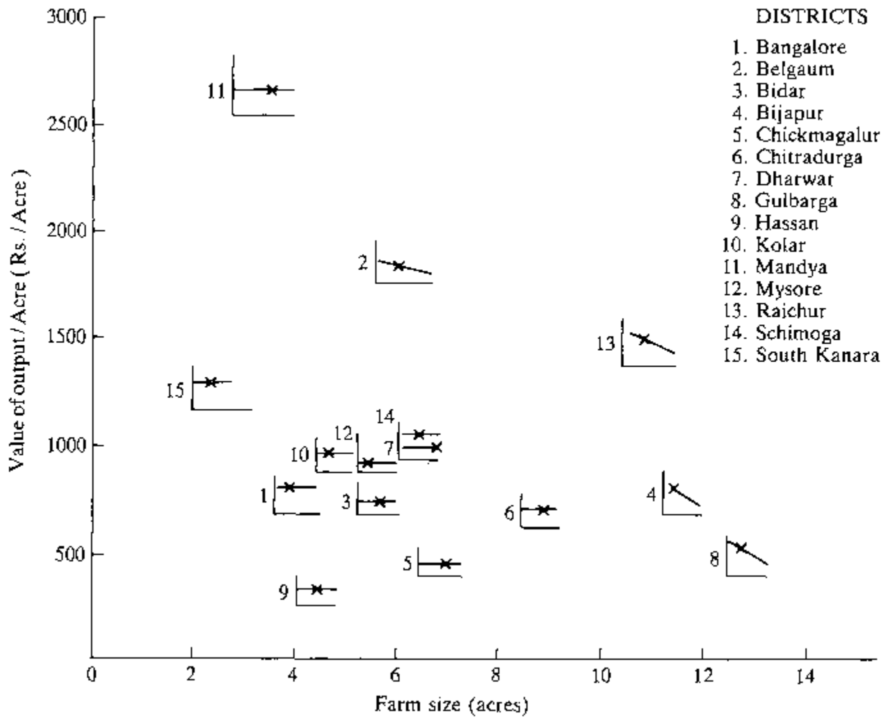

FIG. 1. Karnataka–District level regression.

apparently changes to one of *homogeneous* land, and the explanation of the inverse relationship is attributed to factors endogenous to farm size differentials i.e. the explanation runs from farm size to productivity rather than the reverse.

Our results indicate that their is little justification for the above reversal in causation i.e. it is inappropriate to consider land quality to be homogeneous within states. That there is a wide variation in land quality within a state is documented elsewhere (see Roy (1981), Bhalla (1986a)). And it may be that these land quality differences are the major factor explaining land productivity differentials. As a corollary, labor market imperfections may have little, if any, effect on land productivity.

Two questions remain unanswered: (1) what causes a negative relationship to exist so persistently between farm size and farm quality, and, (2) how can these results be used to interpret the debate on labor market imperfections?

As noted earlier, several writers have tried to offer explanations for the farm size-land quality relationship. They all have merit, yet leave certain questions unanswered. In another paper, Bhalla (1986b) sets up a farmer profit maximization model in which no market imperfections are assumed except to note that supervision of labor costs are present and that a perfect market for managers does not exist. Using this model, it is shown that under plausible conditions, a negative relationship between size and quality will be observed.

Regarding the second question, labor dualism, the results of the present study *cannot* be used to assert that labor market imperfections are absent. However, it is fair to state that the existence of an inverse relationship (and its universality) was used by many authors as a prime piece of evidence to support the hypothesis about labor dualism (e.g. Sen (1966)). If, as it now appears, the inverse relationship is not regularly observed, then other evidence will have to be offered to support dualism. Further, direct evidence on rural India, and one which rejects dualism, is offered by Rosenzweig (1980).

One further objection can be raised to the results contained in this paper. The analysis pertains to data for a survey conducted more than a decade after the advent of the green revolution. Since several authors have noted (and offered explanations for) a weakening of the historical negative relationship with the advent of the new technology,[20] could it not be that our results merely reflect this phenomenon? In other words, an inverse relationship and the conclusion about labor market dualism may still be valid for results obtained from pre-green revolution days.

Our conclusion is that the present results *can* be used to "correctly" interpret prior findings. The FDS was a national survey. The different states of India represent farmers at different stages of the new technology. In Madhya Pradesh, only 18% of the farmers use fertilizer and only 15% use high yielding variety seeds. In contrast, the figures in progressive states like Punjab and Haryana are virtually 100% adoption. This cross-section of experience is analogous to time-series data, and, as is made clear from Tables 2 and 3, there is *no* systematic relationship between the progressivity of a state or district and the nature of the inverse relationship. Out of 8 subzonal regressions in Maharashtra, 4 show an inverse relation; in Tamil Nadu (a progressive state) the corresponding figures are 12 and 7, i.e. an almost equal proportion. The same kind of conclusion can be derived on the basis of district data e.g. in Madhya Pradesh (a backward state) 5 out of 19 districts show a negative relationship; in West Bengal (a relatively progressive state) an almost equal proportion show a similar pattern.

Another indication that land quality is the likely major determinant of the inverse relationship and *not* technology and/or the date of survey, is provided by the gross elasticities of output acre with respect to farm size[21]

[20] For example, see Bhalla (1979).

[21] The reason a log-log formulation is estimated here rather than a semi-log one is because the former is scale free and can therefore be directly used for comparing 1970 and 1976 data—a procedure not possible with the semi-log formulation if inflation in price of output is present.

TABLE 5

Farm Size and Farm Productivity—1970–71 and 1976–77

	1970–71 ARIS Survey				*1976–77 FDS Survey*			
State	*Number of observations*	*Mean of farm size acres*	*Elasticity farm size*	$\bar{R}^2$	*Number of observations*	*Mean of farm size acres*	*Elasticity farm size*	$\bar{R}^2$
Andhra pradesh	155	6.6	−0.24 (−3.5)	0.069	1669	5.2	−0.27 (−11.5)	0.073
Assam	140	3.9	−0.14 (−1.7)	0.013	1025	2.9	−0.09 (−3.7)	0.012
Bihar	88	4.7	−0.33 (−5.0)	0.214	1849	4.5	−0.7 (−3.9)	0.008
Gujarat	193	12.6	0.04 (0.7)	−0.003	1134	6.9	−0.26 (−7.8)	0.050
Haryana	97	14.2	−0.25 (−3.4)	0.101	703	7.6	−0.10 (−3.3)	0.014
Himachal	33	1.9	−0.56 (−3.7)	0.278	480	1.6	−0.21 (−5.7)	0.063
J & K	41	3.0	−0.26 (−1.9)	0.065	529	3.3	−0.22 (−5.2)	0.048
Karnataka	212	7.7	−0.18 (3.5)	0.050	1216	4.1	−0.34 (−10.8)	0.087

Kerala	180	1.8	−0.14 (−6.2)	0.173	282	1.6	−0.17 (−3.1)	0.030
M.P.	302	9.3	−0.13 (−4.2)	0.052	2068	6.6	−0.16 (−8.6)	0.034
Maharashtra	179	8.4	−0.15 (−3.1)	0.048	1006	5.0	−0.34 (−11.5)	0.115
Orissa	147	5.9	−0.19 (−4.9)	0.135	946	4.7	−0.22 (−7.6)	0.058
Punjab	98	10.2	−0.24 (−5.0)	0.197	805	5.8	−0.07 (−2.6)	0.007
Rajasthan	289	8.2	−0.25 (−5.1)	0.081	1191	6.8	−0.34 (−13.1)	0.125
Tamil Nadu	178	3.2	−0.07 (−1.7)	0.010	1732	4.1	−0.20 (−8.3)	0.038
U.P.	377	5.3	−0.06 (−2.7)	0.019	2608	3.7	−0.11 (−7.3)	0.020
W. Bengal	133	4.5	−0.08 (−1.8)	0.018	1236	3.1	−0.05 (−2.8)	0.005

Notes

1. Both surveys were conducted by the same organization, National Council of Applied Economic Research, NCAER. The FDS data are described in Appendix I, and the ARIS data are briefly described in the text.

2. The reported farm size means are *geometric* (i.e. exp (Lland)) where Lland refers to the average of values for log Land. These differ from the arithmetic means reported in Table 1.

3. Elasticity of farm size refers to the coefficient β in the regression $\log Y/H = \alpha + \beta \log H$ where Y is the output per acre and H is farm size.

4. Figures in parentheses represents t-statistics.

reported in Table 5 for two separate survey dates—1970[22] and 1976. These results show that there has been little change in the apparent negative relationship between farm size and productivity for the different states, even though technology had changed considerably over the ensuing years. Thus, this aggregate, and indirect evidence does not support the contention that technical change was a cause of the apparently weak negative relationship obtained at a disaggregated level in India in 1976–77.

If the coefficient of log land (equation 3C) is taken as the index of the inverse relationship, then it can be regressed on indices of technical progress to yield a direct test of the relationship between the two.[23] Such a procedure was tried,[24] but yielded inconclusive results—the variables had the correct signs, but the equation had low explanatory power. This result suggests that the inverse relationship observed at the district level (29% of the cases) is more likely to be a function of random and/or unobserved factors rather than due to labor dualism or lack of technical change.

6. Conclusions

This paper makes a modest attempt at analyzing the farm size–farm productivity relationship in the context of heterogeneous land. Lack of data has prevented the resolution of the important question of whether the observed inverse relationship between farm size and farm productivity is due to labor dualism or land fertility. Availability of a large national survey of farm households in India allows for a test which *does* distinguish between the above two hypotheses. In particular, the data permit a *direct* test of the role of land quality in explaining land productivity.

Exogenous land quality is difficult to define and measure. Further, it is difficult to estimate the myriad ways in which different components of land quality (e.g. soil type, soil color and soil depth) interact. An assessment of the errors that can result from an inappropriate specification of the land quality variables is conducted in two different ways. First, soil quality variables are explicitly introduced into equations explaining farm productivity. Second, farm productivity models are estimated at different levels of pooling—households are classified according to state, zone, sub-zones and district. The district is the lowest level of pooling, and it is hypothesized that at this level, the mis-specification due to heterogeneity of regional variables (e.g. soil moisture) is the least.

Three results emerge from this research: (1) agro-climatic and soil factors

[22] The 1970 data are from another survey conducted by the same organization, NCAER, in 1970–71 as part of a three year panel survey. The sample size was approximately 3000 cultivating households. The results based on these data are contained in Bhalla (1979) and Bhalla (1980).

[23] In such an instance, weighting the index (which is a parameter in a regression) by its accompanying standard error will yield consistent estimates. (See Saxonhouse (1976)).

[24] The variables chosen were fertilizer use and irrigation (to reflect technical change) and number of adults per farm (to reflect labor availability).

are important determinants of farm productivity; (2) once proper account is made of exogeneous land quality variables, the inverse relationship is observed to weaken, and in many cases, to disappear. It is not the case, however, that no relationship remains between size and producticity—but the universality of the "stylized fact" is not 100%, but only 30% of the districts in India. (3) Evidence is also presented to reject the view that this weakened inverse relationship is due to the advent of the green revolution.

Our results are the first of their kind because lack of data has prevented other researchers to conduct the kind of analysis reported here. In an important way, however, our results suggest that past research may have suffered from a specification error problem i.e., exclusion of land quality, a variable negatively correlated with farm size (see Bhalla (1986a)) results in the coefficient of land being biased downward. Thus, the stylized fact of a negative relationship between farm size and farm productivity may in large part have been due to the *omission* of soil quality variables from the estimated equations. Further, though our results cannot reject the hypothesis of labor market dualism, they do succeed, in our opinion, in questioning the presence of dualism by providing a consistent alternative explanation for the existence of the inverse relationship in a large proportion of the regions in India.

The World Bank

APPENDIX I. DATA AND DEFINITIONS

Data

The National Council for Applied Economic Research (NCAER) undertook a survey (known as the Fertilizer Demand Survey (FDS)) of some 22,791 households for the agricultural years 1975–76 and 1976–77. A principal objective of the study was to estimate "fertilizer comsumption at farm level for major crops by varieties, source of irrigation and size of farm" NCAER (1978, p. 18). Consequently, a three-stage stratified sample design with community development (CD) blocks consisting of a cluster of villages as the first, villages with CD blocks as the second and farm households within villages as the final stage was used for the survey. Re-sampling of the households in the second year led to a drop out of 1293 households—a small proportion. Consequently, the sample used in this study consists of 21,499 households.

The NCAER collected detailed information on region of location, land characteristics, irrigation by nine different sources, and ownership of physical assets during the first year of the survey. In addition, detailed cropping pattern data was collected. This included value of output, quantity of output, source of irrigation, fertilizer use, etc. No information was gathered on the use of labor input on the farm on the grounds that such data are difficult to gather and inherently unreliable. The survey was repeated in 1976–77, but questions on soil and assets were not repeated.

Definitions

Definitions and measurements affect estimates, and it is important that the basis for the definitions of the variables used in this study be clearly outlined.

(i) Land—detailed information on ownership and disposition (operated, leased out etc.) of land was collected.

(ii) *Irrigation*—Detailed information on sources, and uses, of irrigation was collected. Nine irrigation sources were identified—canal, canal and tubewell/pumpset, tubewell, pumpset from well, pumpset from other sources, hired tubewell/pumpset, well, tanks/ponds and other.

(iii) *Quality of Soil*—Each farmer was asked detailed questions on the characteristics of his land holdings. These included the following: (a) Soil Texture, SOIL1—sand, loam, light clay and heavy clay; (b) Soil Color—SOIL2—red, black, grey and yellow; (c) Depth of Soil—SOIL3—less than 1 foot; between 1 and three feet, and more than 3 feet; (d) Soil salinity—nil, moderate and high; (e) Surface Drainage—good, not so good and poor; (f) Rate of Percolation—fast, medium and slow.

Sample size

Though observations on 21,499 are available data, only 17,147 were used. Observations were excluded if they had information missing on *any* of the soil variables included in this study. In addition, some observations had to be omitted because an agronomic zone classification was not available. However, there appears to be little censorship bias for the major relationship of interest in this paper i.e. the inverse relationship. When such a relationship is estimated at the state level, very little different, if any, is observed between elasticities obtained via the complete, and reduced sample set.

REFERENCES

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

Berry, R. A. "Land Distribution, Income Distribution, and the Productive Efficiency of Colombian Agriculture," *Food Research Institute Studies* 12 (3), 1973.

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

Bhagwati, J. N. and Chakravarty, S. "Contributions to Indian Economic Analysis: A Survey," *American Economic Review* (supplement) September 1969.

Bhalla, Surjit S., "Farm Size, Productivity, and Technical Change in Indian Agriculture", in Berry, A. and Cline, W., *Agrarian Structure and Productivity in Developing Countries*, Johns Hopkins University Press, Baltimore, 1979.

——, The Measurement of Permanent Income and its Application to Savings Behavior", *Journal of Political Economy*, Vol. 88, #4, Aug. 1980.

——, "Size Determination of Farms—Theory and Evidence", World Bank, March 1983a; revised Aug. 1986b.

——, "Does Land Quality Matter—Theory and Measurement", World Bank, mimeo, December 1986a; forthcoming, *Journal of Development Economics*.

Bharadwaj, K., *Production Conditions in Indian Agriculture: A Study Based on the Farm Management Surveys*, Cambridge: Cambridge University Press, 1974.

Bliss, C. and Stern, N., *Palanpur: The Economy of an Indian Village*, Oxford, Clarendon Press, 1982.

Chayanov, A. V., *The Theory of Peasant Economy*, ed. Daniel Thorner, B. Kerblay, R. E. F. Smith (Homewood: R. D. Irwin) 1966.

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

Cline, W. R. *Economic Consequences of a Land Reform in Brazil*, North Holland, 1970.

Griffin, K. *The Political Economy of Agrarian Change: An Essay on the Green Revolution*, Cambridge, Mass.: Harvard University Press, 1974.

Khusro, A. M., "Returns to Scale in Indian Agriculture", *Indian Journal of Agricultural Economics*, July–December 1964.

PRASAD, R. "Role of Soil, Irrigation and Climatic Factors in the Adoption of HYV's of Wheat and Rice in India", mimeo, World Bank, 1983.

ROSENZWEIG, MARK R. "Neoclassical Theory and the Optimizing Peasant: An Econometric Analysis of Market Family Labor Supply in a Developing Country", *Quaterly Journal of Economics*, February 1980.

ROY, PRANNOY, "Transition in Agriculture: Empirical Indicators and Results (Evidence from Punjab, India)", *Journal of Peasant Studies*, January, 1981.

RUDRA, ASHOK, "More on Returns to Scale in Indian Agriculture", *Economic and Political Weekly*, Oct. 26, 1968.

SAXONHOUSE, G., "Estimated Parameters as Dependent Variables", *American Economic Review*, 66, March 1976.

SEN, A. K., *Employment, Technology and Development* London, Oxford University Press, 1975.

——, "Peasants and Dualism with or without Surplus Labor," *Journal of Political Economy*, October, 1966.

——, "An Aspect of Indian Agriculture," *Economic Weekly* Annual Number 1962.

SRINIVASAN, T. N., "Farm Size and Productivity—Implications of Choice under Uncertainty", *Indian Journal of Statistics*, Series B, Vol. 34, part 4, 1972.