Oxford Economic Papers 36 (1984), 131–145

IDENTIFICATION OF THE INVERSE RELATIONSHIP BETWEEN FARM SIZE AND PRODUCTIVITY: AN EMPIRICAL ANALYSIS OF PEASANT AGRICULTURAL PRODUCTION

By MICHAEL R. CARTER

Introduction

INDIAN agricultural data from the 1950s revealed different patterns of resource allocation and use on small and large farms, including an inverse relationship between farm size and annual production per hectare.[1] Because of their implications for agricultural development strategy, these findings provoked numerous studies designed to determine if production on small and large farms really differs and, if it does, why. These studies have often featured competing methodologies and conclusions. Using a pooled farm level dataset, this study tries to distinguish between alternative explanations of the inverse farm-size productivity relationship. Its basic conclusions are that the relationship is not a reflection of bias resulting from sample selection based on farmer literacy, nor is it a misidentification of village effects. The analysis favors what might be called a "mode of production" explanation of the inverse relationship.

The data

The data for this study are from farm management surveys taken in the Indian state of Haryana during the agricultural years 1969/70–1971/72.[2] For each of the three years, 162 holdings were selected through a multi-stage stratified random sampling procedure. However, no data were obtained from 22% of the selected holdings because of the unwillingness or inability of the farmers to keep the required records. Illiteracy was given as the major reason these observations were lost. The potentially pernicious effect of this non-random selection process, and a procedure which corrects for it, will be discussed below.

In total, data is available on 376 holdings. Each observation is identified only by its district and village location, and it cannot be determined if any individual holding was repeatedly sampled over time. But, there are at least two observations on all but fifteen of the 94 villages represented in the total sample. In anticipation of later analysis which requires multiple observations

[1] Sen. (1975) reviews much of the early Indian literature on the inverse relationship. Berry and Cline (1979) present evidence collected from studies of agriculture throughout the third world. Other recent studies of some notoriety are Yotopoulos and Lau (1971), Bardhan (1973) and Barnum and Squire (1978).

[2] The farm level data are included as appendicies in the annual issues of *Studies in the Economics of Farm Management.*

TABLE 1
Descriptive statistics

	Small farms	*Large farms*
Size (Net Hectares)	3.02	8.46
	(0.713)	(5.03)
Wage (Rs. per annum)	1189	1192
	(242)	(610)
Per-Hectare Figures:		
Production (Deflated Rs.)	2055	1474
	(801)	(711)
Labor (man years)	0.649	0.350
	(0.306)	(0.162)
Intermediate Inputs (Rs)	693	434
	(301)	(346)
Capital (Rs)	83.6	50.8
	(62.4)	(53.9)
Land values (Rs)	730	520
	(365)	(297)
Number of Observations	66	292

Figures in parentheses are estimated standard errors.

Farm size distribution

	State of Haryana: % of total area	*Selected data as a % of total random sample*
Below 2.02 has.	20.8	2.3
2.02 to 4.05 has.	23.1	12.0
4.05 to 6.07 has.	17.9	22.6
6.07 to 8.10 has.	12.6	15.9
8.10 to 10.12 has.	8.1	10.3
10.12 to 12.14 has.	5.7	4.3
Above 12.14 has.	11.8	9.1
	100%	77.5%

on villages, the fifteen unmatched observations, and three internally inconsistent ones, were dropped from the analysis, leaving a total of 358 observations.

Table 1 presents descriptive statistics computed from the data. The farm size variable measures the stock of land available to the producer, that is the net area owned and, or rented. This net farm size measure, rather than gross cultivated area which reflects multiple cropping decisions, is the variable of interest in studying whether economic behavior is different on different sized holdings (and whether production from the same land stock would vary as the size distribution of holdings varied). As the Table 1 figures report, small farms, defined as those of less than 10 acres (4.047 has.),[3] exhibit higher

[3] The 10 acre figure has been the traditional dividing line between large and small farms in studies using Indian data.

production per-(net) hectare than large farms, as well as substantially greater input intensities. Per-hectare land values are also on average higher for the smaller farms. The wage figures are based on payments to permanent workers. When a holding did not employ a permanent worker, its wage variable was valued at the average wage for its village. It can be seen that the measured wage on small holdings is on average the same as the wage paid on holdings greater than 10 acres in size.

Also included in Table 1 are figures on the actual distribution of agricultural holdings by size in Haryana. Under the initial random sampling procedure, the number of observations on each size strata should have been approximately proportional to each strata's area in the population.[4] The effect of the sample selection rule of farmer literacy is visible in the under-representation of the two smallest size strata relative to the proportion of land actually cultivated by small farms.

The inverse relationship and sample selection bias

The data, pooled across villages, was used to estimate the following linear regression function commonly found in the farm size-productivity literature:

$$y_i = \alpha + \beta h_i \tag{1}$$

where h_i is the log of farm size and y_i is the log of total annual farm output per-hectare. Output is measured as the current value of total marketed and non-marketed farm production, deflated by a Laspeyres price index specific to the year and region of the observation.[5] The regression relationship is assumed to be constant over time. The OLS estimate of β, reported in Table 2, shows a very strong inverse relationship between farm size and productivity, with per-hectare production declining nearly 40% as farm size doubles.

[4] The description of sampling procedure in *Studies in the Economics of Farm Management* is incomplete and requires interpretation. The interpretation presented in the text appears correct and is consistent with the figures reported in Table 1. In addition, an essay on agricultural survey methodology, prepared by the Indian National Sample Survey (1962) outlines a procedure which gives each physical unit of land an equal chance of selection, and would therefore be expected to yield observations on farm size strata in proportion to the area cultivated.

[5] The implicit assumption behind the aggregation of multiple crops into a one dimensional output measure is that crop mix is chosen to maximize revenues. Recent contributions to the economic theory of index numbers have been devoted to deriving index formulae which are consistent with (i.e. exact for) revenue maximizing substitution between elements of the aggregate production measure. The common pseudo-Laspeyres output index used here ("pseudo" because Laspeyres indicies do not fulfill Fisher's so called factor reversal test) does not possess virtuous aggregation/substitution properties. Given that aggregation may have induced some measurement error in the dependent variable, the important question is whether that error biases analysis of farm size issues. However, bias seems likely to result from inappropriate representation of substitution behavior than from the use of *market* prices as the weights for aggregating small farm, peasant production. As a low income, joint production/consumption unit, the small farm may internally assign a risk premium to food prices. Aggregation using market prices may thus understate small farm output. Output aggregation based on a more flexible representation of crop mix substitution would not remedy this possible bias. In any event, given the limitation of regional price data for index construction, it is not apparent that an alternative index would be more informative.

The size of this coefficient is surprising, particularly considering that the data postdate the green revolution. It is often argued that large farms have better access to the credit necessary to purchase yield-increasing inputs and that operators of larger holdings are generally less averse to the adoption of new techniques. G. S. Bhalla's survey of agricultural households in Haryana for the year 1970/71 found that the rate of adaptation of "progressive techniques" (as indicated by the use of improved seeds) was somewhat lower among the smaller farm size strata.[6] Such a pattern of differential technical change would be expected to diminish the inverse relationship relative to that measured in earlier data. However, the estimated elasticity of -0.39 is large even compared to results reported in pre-green revolution studies.[7]

This estimate must be viewed with particular skepticism because it is based on non-randomly selected data. The sample selection rule of farmer literacy, which censored 22% of the observations, may well be common to data used in other studies of the inverse relationship. The Table 1 figures on farm size distribution indicate that literacy is positively related to farm size. If literacy is also related to productivity, then inference from the systematically censored data would yield biased conclusions about the underlying population as a whole (literates and illiterates). Current research in fact suggests that education significantly affects agricultural productivity. In a summary of recent literature, the World Bank (1980) reports that four years of primary schooling may increase a farmer's productivity by 13%. The estimated inverse relationship may in part be a reflection of a sample selection process in which observations on small, low eductation, low productivity farms were censored. By explicitly modelling the sample selection process, it is possible to test for sample selection bias and estimate the underlying population relationship between farm productivity and size.

The probability of selection into the sample can be expressed as the probability that a latent literacy variable, l_i, exceeds an arbitrary value—assumed here to be zero. Let D_i be the binary variable which equals one if the operator of the ith holding is literate and included in the sample. Under this specification,

$$\text{Prob}\,(D_i = 1 \mid h_i) = \text{Prob}\,(l_i > 0 \mid h_i).$$

For notational simplification the constant term has been dropped. Assuming

[6] The ratio of progressive to non-progressive farmers was 0.44 in the smallest size strata (less than 5 acres), 0.61 for holdings of 5 to 10 acres, 0.53 for 10–20 acres, 1.0 for 20–30 acres, and 0.69 for holdings greater than 30 acres in size. Bhalla's definition does not, however, distinguish between degrees of adaptation to new techniques.

[7] Saini (1971), Rani (1971) and Bhattacharya and Saini (1972) present numerous productivity-farm size regressions using Indian farm management data from the 1950s and 1960s. Several regressions using data from the 1950s show an inverse relationship comparable in size to that estimated here. Most estimates, however are somewhat smaller (in the -0.10 to -0.20 range). Their estimates using data from the first few years of the green revolution (the late 1960s) show no change from the earlier periods. Bliss and Stern (1982) do not find an inverse relationship in their intravillage data. Their recent work provides evidence on other issues studied here. Unfortunately this work was completed before the author encountered Bliss and Stern's interesting study.

that l_i is a linear function of the log of farm size, the complete model for the process which generated the data is

$$l_i = \rho h_i + \varepsilon_{i1}$$
$$y_i = \beta h_i + \varepsilon_{i2}$$

While under the assumptions of the model, it is true that $E(y_i \mid h_i) = \beta h_i$, y_i is only observed when

$$l_i = \rho h_i + \varepsilon_{i1} > 0$$
$$\varepsilon_{i1} > -\rho h_i.$$

For the censored data, the relevant regression function is conditional on the selection criteria

$$E(y_i \mid h_i, \varepsilon_{i1} > -\rho h_i)$$

which equals

$$\beta h_i + E(\varepsilon_{i2} \mid \varepsilon_{i1} > -\rho h_i). \tag{2}$$

If in fact literacy, and attributes correlated with it, positively affect farm productivity (i.e. $E(\varepsilon_{i2} \mid \varepsilon_{i1}) > 0$), then the unconditional estimate of β from equation (1) will be a biased estimator of the population relationship.

In order to estimate the conditional regression function (2), some further assumption about the joint distribution of ε_i is necessary. For simplicity's sake, it is assumed here that ε_i has a multivariate normal distribution with variance-covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}.$$

While this assumption is arbitrary, the resulting conditional distribution is simple

$$\varepsilon_{i2} \mid \varepsilon_{i1} \sim N(\tau\varepsilon_{i1}, \sigma^2),$$

where $\tau = \sigma_{12}/\sigma_{11}$ and $\sigma^2 = \sigma_{22} - \sigma_{12}^2/\sigma_{11}$. The conditional regression function becomes

$$\beta h_i + \tau E(\varepsilon_{i1} \mid \varepsilon_{i1} > -\rho h_i) = \beta h_i + \delta\lambda(-\gamma h_i), \tag{3}$$

where $\delta = \tau\sqrt{\sigma_{11}}$, $\gamma = \rho/\sqrt{\sigma_{11}}$ and $\lambda(\cdot) = \phi(\cdot)/1 - \Phi(\cdot)$, where $\phi(\cdot)$ and $\Phi(\cdot)$ are the standard normal density and cumulative density functions respectively. If $\rho > 0$ (i.e. if literacy increases with farm size), then $\mathrm{d}\lambda(-\gamma h_i)/\mathrm{d}h_i < 0$. The standard omitted variable estimator expression then shows that the omission of $\lambda(\cdot)$ in equation (1) negatively biases the estimate effect of farm size on productivity if σ_{12} (and therefore δ) is greater than zero, as expected on the basis of education-productivity studies.

Equation (3), and an estimate of β unbiased by sample selection, can be approached either directly through maximum likelihood, or through the two-stage, probit/OLS procedure proposed by Heckman (1979).[8] The Heckman procedure can only be implemented if h_i is observed for those farms excluded from the sample. Otherwise the probit estimate of the probability of selection into the sample given farm size degenerates for lack of variance in the binary dependent variable. While the published data do not report the size of the excluded holdings, information on them can be inferred from the population distribution of farms by size strata. Using the assumption that the original random sample contained observations on farm size strata in proportion to the area of each strata in the population, the 108 censored observations were divided between the seven farm size categories, and assigned the value of their respective strata mean.

The estimates of the conditional regression function reported in Table 2 were computed using the Heckman two stage procedure. The conditional estimate of β is virtually identical to the unconditional estimate from equation (1). The estimated γ has the expected sign, indicating that $\rho > 0$—i.e., that literacy is positively related to farm size. However, the estimate of δ indicates that σ_{12} (the relation between literacy and productivity) is negative—although $\hat{\delta}$ itself is only insignificantly different from zero. The reported standard errors for $\hat{\beta}$ and $\hat{\delta}$ are the usual OLS estimates, and in this case they are biased downwards. Heckman (1979) presents the appropriate correction, which is somewhat complicated. Alternatively, maximum likelihood would estimate the true asymptotic standard errors. But given the estimated economic insignificance of sample selection bias, the additional expense of calculating correct standard errors was foregone.

After taking into consideration the sample selection procedure which overwhelmingly eliminated small holdings, the data still exhibit a substantial inverse relationship between farm size and productivity. Subject to the accuracy of the multinormal distributional assumption used to implement the sample selection model, a 1% increase in farm size unconditionally diminishes farm productivity by 0.4%. The next step in the analysis is to identify the factors which explain this relationship.

Village effects and the inverse relationship

Exploiting the structure of the data as pooled village cross-sectional surveys, it is possible to discriminate between two classes of explanations of the inverse relationship. The first class hypothesizes that the observed relationship is a misidentification of village factors which are correlated with

[8] The likelihood function for the selectivity model is

$$\prod_{i\in M_1}[1-\Phi(\gamma x_i)]\prod_{i\in M_2}\left[\sigma\int_{-\gamma x_i}^{\infty}\phi(y_i-\beta x_i-\tau u)/\sigma\phi(u/\sigma)\,du\right],$$

where $M_1=\{i\mid D_i=0\}$ and $M_2=\{i\mid D_i=1\}$. The Heckman procedure uses a first stage probit to estimate γ and the probability of selection into the sample given farm size. The estimate of γ is then used to construct a $\lambda(-\gamma h_i)$ variable which is used as the second regressor for OLS estimation of the conditional regression function, equation (3).

TABLE 2
Farm size productivity relationship

		Estimate of	R^2
Equation (1)			
	β	−0.393	0.182
		(0.044)	
Selectivity model			
	β	−0.416	0.182
		(0.124)	
	δ	−0.067	
		(0.333)*	
	γ	1.19	
		(0.098)	
Village effects model			
	β	−0.337	0.785
		(0.041)	
	ξ	−0.414	
		(0.105)	
	$\bar{\theta}$	−0.077	
		(0.146)	

Figures in parentheses are estimated standard errors.
* Downward biased estimate.
† Estimated asymptotic standard error.

farm size. For example, in areas of greater soil quality, population may have grown relatively rapidly, leading to a subdivision of land into small holdings. The inverse relationship may then simply be a misidentification of the effect of soil quality on production. Similarly, villages with small holdings may have cheap and abundant labor which would allow both small and large units to employ more labor per-hectare and exhibit higher farm productivity. Either way, after controlling for village factors, the inverse relationship would disappear, or at least be diminished, according to these hypotheses. Sen (1975) reports that in fact estimates using data from within a single village often show a lesser effect of farm size on productivity.

Contrary to these misidentification hypotheses is the view that the inverse relationship is generated by characteristics of small farms themselves. Small farms may uniquely enjoy higher quality soil within the village, creating the observed size-productivity relationship. The inverse relationship could also result if farm size is a proxy for mode of production. Small, peasant mode of production farms arguably utilize their fixed resources more intensively (e.g., through heavy multiple cropping) and generate higher production per hectare than do large market-oriented farms (Sen [1966]).

Discriminating between the two classes of explanations of the inverse relationship requires separate identification of village and farm specific effects on productivity. Within villages, average soil quality and wage levels should be relatively constant over the sample period. Their effect on observed farm productivity can be separately identified from effects of farm-specific factors by using the group structure of the data. Per-hectare production for the jth farm in the ith village is specified as:

$$y_{ij} = \alpha_i + \beta h_{ij} + \varepsilon_{ij}, \tag{4}$$

where the latent variable α_i measures the effect of factors specific to village i on farm productivity. The effect of omitting α_i on the estimated inverse relationship depends on the auxillary linear regression function relating the latent village effects variable to observations on farm size in the village:

$$E^*(\alpha_i \mid \underline{h}_i) = \underline{\theta}' \underline{h}_i.$$

But, because the ordering of observations on each village is arbitrary, this unrestricted form of the regression function makes little sense. There is no apparent reason for arbitrarily denoted second observations on villages to have different effects (θ_2) from first observations (θ_1). Imposing the restriction that all elements of $\underline{\theta}$ are equal to some $\bar{\theta}$ seems appropriate and simplifies the auxiliary regression function considerably:

$$E^*(\alpha_i \mid \underline{h}_i) = \bar{\theta}\bar{h}_i,$$

where $\bar{h}_i = \Sigma_j h_{ij}$. In terms of this model, the omission of α_i in equation (1) leads to a biased estimate of β, which misidentifies village effects as an effect of farm size per se, if $\bar{\theta} \neq 0$. Substituting for α_i in equation (4) and rearranging terms gives:

$$y_{ij} = \beta(h_{ij} - \bar{h}_i) + \xi\bar{h}_i + (\omega_i + \varepsilon_{ij}) \tag{5}$$

where $\xi = \bar{\theta} + \beta$, and $\omega_i = \alpha_i - E^*(\alpha_i)$. Mundlak (1978) demonstrates that the GLS estimators of β and ξ are independent and are equivalent to the within and between estimators of the analysis of covariance, respectively. Using these estimators it is possible to separately identify village and farm specific effects.

The estimated parameters for equation (5), reported in Table 2, show that the within estimator of β is less than the between estimator, ξ.[9] The estimated value of $\bar{\theta}$ is -0.077, indicating that productivity enhancing village effects are inversely related to farm size. The omission of α_i in equation (1) does apparently cause an overstatement of the inverse relationship. However, the hypothesis that $\bar{\theta} = 0$ (i.e. that the bias is insignificant) is not rejected, as the relevant computed t-statistic is 0.69. In addition, the difference between the within estimator of β and the omitted variable estimator from equation (1) is of minor economic significance.

It should not be concluded that village effects are unimportant. They do explain a substantial amount of the total variation in productivity. The restriction that all the α_i are equal (implicitly imposed in equation (1)) is

[9] Even if the restrictions imposed on $E^*(\alpha_i \mid h_i)$ are inappropriate, the within estimator reported in Table 2 is an unbiased estimator of β in equation (4), although it would not generally be efficient.

statistically rejected with a calculated F-statistic of 10.18.[10] Including village dummies in equation (1) would have increased R^2 from 0.168 to 0.785. However, while village factors explain much of the variance in productivity—over two thirds of the total variation in per hectare farm productivity is between village means—they do not substantially diminish the measured impact of farm size on productivity. Contrary to the findings reported by Sen (1975), the within village inverse relationship is not significantly different from the relationship between villages.

Farm specific factors

The evidence so far considered militates against an explanation of the inverse relationship as a misidentification of the effects of village soil quality or labor costs. This section will examine patterns of resource use and allocation in order to locate farm specific factors which could explain the relationship. At the farm level, there are several possible explanations, each of which will be examined in turn. First, it could be that small farms are technically more efficient in the sense that they produce more output from given inputs. Alternatively, agricultural production could exhibit decreasing returns to scale. A third possible explanation is that small farms use greater quantities of variable inputs per hectare, either because they face different input prices than do large farms, or because they operate under a peasant mode of production. Finally, the inverse relationship could be a reflection of higher quality land on small farms.

These explanations have been examined in earlier studies. S. Bhalla (1979) finds that an inverse size productivity relationship remains even after controlling for land quality and irrigation. Bardhan (1973) locates some evidence of decreasing returns to scale. Using a production function which controls for irrigation, Bardhan also finds that small farms may be technically more efficient than large farms. Barnum and Squire (1978) reject the hypothesis of decreasing returns to scale, and based on production function estimates conclude that the technical and allocative efficiency of small and large farms is identical. (Unfortunately, it is not clear if their sample even exhibits an inverse relationship which requires explanation.) In contrast Yotopoulos and Lau (1973) find, using profit function analysis, that small farms are "economically" more efficient than large farms. They do not find any differences in allocative efficiency between small and large farms and conclude that small farms must be technically more efficient.

The mixed conclusions of these earlier studies may simply reflect contradictory data, or they may result from differences in the statistical

[10] The F-statistic was calculated as

$$\frac{(SSR_R - SSR_U)/78}{SSR_U/283},$$

where SSR_U is the sum of squared residuals from the "within" regression and SSR_R is from the "total" regression, equation (1). The 5% critical value of F(78,283) is approximately 1.47.

methodologies used. Direct estimation of production function parameters (as in the Bardhan and Barnum and Squire studies) is inconsistent if variable inputs are correlated with omitted variables (such as weather) and stochastic disturbances. While Yotopoulos' profit function analysis resolves this inconsistency by incorporating the endogeneity of variable inputs, its statistical accuracy and power is limited by the questionable quality of price data used as instrumental variables. Motivated largely by data deficiencies, this study will rely on direct methods to estimate production and efficiency parameters.

The choice of functional form to represent production possibilities can be analytically significant. The commonly used Cobb-Douglas form imposes the assumption that all partial elasticities of substitution equal one at every point in the input space. However, if, for example, labor can better substitute for other factors as irrigation facilities improve, a Cobb–Douglas analysis might misidentify rational substitution into labor as allocative inefficiency on well irrigated holdings. To lessen the possibility of such problems, a translog approximation to an arbitrary production function is used here to specify production:[11]

$$y = e + \Sigma_i \beta_i x_i + \tfrac{1}{2}\Sigma_1 \Sigma_j \gamma_{ij} x_i x_j + u, \qquad i, j = 1, 3, \tag{6}$$

where the input vector $\underset{\sim}{x} = (\ln L \ \ln I \ \ln R)'$, and all variables have been divided by their sample mean. The input L is the value of payments to casual and permament workers plus the imputed value of family labor, all deflated by the permanent worker wage. The variable I is the actual and/or imputed value of intermediate inputs and R is the actual and/or imputed rental value of fixed farm assets, including land. Ideally, these latter two variables should be deflated into quantity measures, but the necessary price information is not available. The asset variable should account for land quality and irrigation facilities. Finally, the term u represents stochastic variation in production, and e is a latent technical efficiency variable to be discussed in more detail after a few comments on the concept of technical efficiency.

In general, not all production inputs are observed, and those that are observed are not measured in productivity units (e.g., labor is measured in time rather than effort units). Variation among producers in the employment of unobserved inputs, and in the intensity and effort with which observed inputs are utilized, generates variation in the apparent technical efficiency of the (observable) production function. Regardless of whether such variation in resource utilization is appropriately labelled as "technical inefficiency," so called technical efficiency measures are useful characterizations of how firms

[11] Berndt and Christensen (1973a and 1973b) discuss the properties of the translog approximation in detail. The translog imposes no restrictions on elasticities of substitution. Assuming symmetry, i.e. $\gamma_{ij} = \gamma_{ji}$, the translog function is homogenous if $\Sigma_i \gamma_{ij} = 0$ for all i. Further, if $\Sigma_i \beta_i = 1$ the function exhibits constant returns to scale. Given homogeneity, if $\gamma_{jk} = \gamma_{ik} = 0$ then factors i and j are separable from k and $\sigma_{jk} = \sigma_{ik} = 1$, where σ represents the Allen partical elasticity of substitution. The translog collapses to a Cobb-Douglas if all $\gamma_{ij} = 0$.

use their observed inputs.[12] Estimation of technical efficiency measures is here approached as a problem of recovering information on the latent variable "e". As specified in (6), farms with larger values of e produce more output from the same inputs and in that sense can be labelled as technically more efficient.

In equation (6), the latent technical efficiency variable has been specified to be strongly separable from the other factors of production. It can be written as the sum of two orthogonal components: $e=(\alpha_0+\alpha_s d)+\omega$, $E^*(\omega \mid 1, d)=0$, where E^* is the linear expectation operator and d is a dummy variable which equals one if the holding is less than ten acres in size and equals zero otherwise. The parameter α_s measures the difference between the unconditional mean of e for small and large farms. A non-zero value of α_s would indicate systematic differences in the technical efficiency of small and large farms. Direct OLS estimation of equation (6) given $\underline{x}$ and d will yield coefficients subject to two sources of bias. First, if $E^*(\omega \mid \underline{x}) \neq 0$, the estimates will be affected by "residual omitted variable bias" which would likely cause an over-statement of returns to scale. Second, as mentioned earlier, simultaneous bias resulting from correlation between variable inputs and other error components (e.g. weather) might also bias direct OLS estimates, although the direction of bias is uncertain.

Table 3 presents direct estimates of the general translog function and of several restricted versions. Statistically it is impossible to reject the restriction that the translog is homogeneous of degree one (i.e. exhibits constant returns to scale). Complete global separability (which reduces the translog to a Cobb–Douglas function) is rejected, but the hypothesis that labor and intermediate inputs are separable from farm assets ($\gamma_{LR}=\gamma_{LI}=0$) is accepted. This implies that Allen partial elasticities of substitution between assets and labor, and between labor and intermediate inputs, equal one everywhere in the input space (see footnote 11 above). Subject to that restriction, the elasticity of substitution between assets and intermediate inputs is calculated to average 8.6 in the sample.

The estimates of α_s reported in Table 3 indicate that small farms would produce 15% *less* output than large farms given the same inputs.[13] This result may reflect better access of large farms to yield increasing green revolution technology. Certainly no explanation of the inverse relationship is to be found in terms of technical efficiency. The inverse farm size productivity relationship, which exists despite the estimated technical inefficiency of

[12] Carter (1982) discusses in more detail the conceptual issue of what is or is not defined as technical efficiency and how econometric procedures designed to generate technical efficiency measures, such as frontier estimation, might be appropriately modified.

[13] Stochastic frontier techniques were also used to obtain information on technical efficiency. Frontier estimates of the production function gave the puzzling result of practically no variation in technical efficiency across the sample. This finding contradicts the dummy variable estimates and is interpreted more as a reflection of the unreliability of frontier estimation (given its dependence on arbitrary distributional assumptions), rather than as a characteristic of production in the sample.

TABLE 3
Resource allocation and use

Estimate of	*Translog*	*CRTS Translog*	*CRTS, $\gamma_{LR} = \gamma_{LI} = 0$ Translog*	*Cobb–Douglas*
α_s	−0.143	−0.146	−0.147	−0.163
	(0.035)	(0.033)	(0.033)	(0.039)
β_L	0.254	0.244	0.264	0.283
	(0.042)	(0.030)	(0.026)	(0.040)
β_I	0.348	0.345	0.352	0.302
	(0.022)	(0.021)	(0.021)	(0.023)
β_R	0.420	0.411	0.384	0.398
	(0.032)	(0.029)	(0.024)	(0.027)
γ_{LL}	−0.074	0.048	0.0	0.0
	(0.154)	(0.079)	—	—
γ_{LI}	0.059	0.034	0.0	0.0
	(0.058)	(0.048)	—	—
γ_{LR}	−0.033	−0.082	0.0	0.0
	(0.072)	(0.053)	—	—
γ_{II}	0.164	0.159	0.175	0.0
	(0.024)	(0.023)	(0.019)	—
γ_{IR}	−0.180	−0.192	−0.175	0.0
	(0.041)	(0.040)	(0.019)	—
γ_{RR}	0.269	0.274	0.175	0.0
	(0.058)	(0.056)	(0.019)	—
SSR	19.77	19.94	20.14	24.77
R^2	0.830	0.828	0.826	0.787
Mean Labor Misallocation: $E(\varepsilon_1)$				
Small Farms			0.111	0.092
			(0.017)	(0.017)
Large Farms			0.019	0.0002
			(0.008)	(0.008)
Mean Intermediate Input Misallocation: $E(\varepsilon_2)$				
Small Farms			−0.038	0.028
			(0.010)	(0.015)
Large Farms			−0.057	−0.028
			(0.005)	(0.007)

Figures in parentheses are estimated standard errors.

small farms, must therefore be the result of greater input intensities per hectare on small farms. Still to be determined is whether these input intensities, and the inverse relationship, can be explained in terms of profit maximizing response to variation in fixed assets and market prices, or whether small farms exhibit characteristics of a peasant mode of production.

Given a production function, efficient allocation of variable inputs at market prices is equivalent to profit maximization at those same prices. Treating labor and intermediate inputs as freely variable in the short run period of the crop season, stochastic first order conditions for profit maximization given a translog production function are:

$$\begin{aligned} S_{\mathrm{L}}-(\beta_{\mathrm{L}}+\gamma_{\mathrm{LL}} \ln L+\gamma_{\mathrm{LI}} \ln L+\gamma_{\mathrm{LR}} \ln R)=\varepsilon_1 \\ S_{\mathrm{I}}-(\beta_{\mathrm{I}}+\gamma_{\mathrm{LI}} \ln L+\gamma_{\mathrm{II}} \ln I+\gamma_{\mathrm{IR}} \ln R)=\varepsilon_2, \end{aligned} \tag{7}$$

where S_{L} and S_{I} are the ratios of labor and intermediate input costs (at market prices) to the value of output, respectively. Some deviation between actual and optimal factor shares (captured by ε_1 and ε_2) would be expected even for the allocatively efficient firm. However, a non-zero mean for either ε_1 and ε_2 would indicate systematic resource misallocation at the observed market prices.

Table 3 presents separate estimates by farm size of mean resource misallocation. The estimates were calculated by substituting the directly estimated production function parameters into (7). When labor is valued at the market wage, small farms are estimated to systematically overemploy labor such that the actual factor share averages 11 percentage points above the optimal share. At the average optimal share of about 0.26, this implies that labor is employed about 36% above its optimal quantity. Large farms, by comparison, are estimated to overutilize labor by a relatively small amount. Based on the translog estimates, small and large farms do not behave very differently in their choice of intermediate inputs. (The restricted Cobb-Douglas estimates do indicate a more significant behavioral difference in the allocation of intermediate inputs.)

Based on these estimates, the relatively massive inputs of labor per-hectare on small farms shown in Table 1 result in part from apparently non-profit maximizing allocative behavior on small farms. Another way of stating this result is to say that operators of small holdings act as if they were maximizing profits at a wage some 36% below market level. The large per-hectare use of intermediate inputs on small farms shown in Table 1 does not result from any behavioral difference, in the face of market prices, between farm size strata, and on average it does not exceed profit maximizing levels given the level of other inputs and substitution possibilities.

This intensive use of variable inputs contributes to the farm size productivity relationship. The third production function input, farm assets (including soil quality), has been treated as a fixed factor. Inputs of this factor per-hectare are also higher on small farms (see Table 1), and residually explain that portion of the inverse relationship which is not accounted for by the intensity of variable input use. As an indicator of the impact of small farm assets on production per-hectare, the inverse relationship was reestimated with assets per-hectare as an additional explanatory variable. The inclusion of this variable reduced the estimated inverse relationship (β) to just below -0.20 for both the simple model of equation (1) and the village effects model of equation (4).

Conclusion

The inverse relationship between farm size and productivity has been found to be a reflection of substantive differences between small and large farms, rather than a product of sample selection bias or village factors correlated with farm size. Intravillage soil quality differences and other farm assets explain part of the size-productivity relationship, although per-hectare production is still estimated to decline 20% as farm size doubles controlling for these factors. Small farms are estimated to be technically inefficient in the sense that they would produce approximately 15% less output given the same inputs. In fact, small farms use far more inputs per-hectare than do large farms, producing the observed inverse relationship given the estimate of constant returns to scale. Labor on small farms is employed 36% beyond the optimal level defined by profit maximization at market prices, dissipating returns on fixed factors of production. This finding of apparently non-profit maximizing behavior by small farmers is consistent with the notion of a peasant mode of production advanced by Vergopoulos (1978) among others. Yet, it may also be that market prices overstate the actual opportunity costs of the peasant's factors of production, and that the observed input levels are profit maximizing at the effective price level. This latter interpretation is consistent with de Janvry's (1981) analysis of peasant agriculture within a capitalist mode of production. de Janvry rejects the notion that peasants are constitutionally disinterested in profits and accumulation, and attributes their rent-dissipating "self-exploitation" to their weak terms of trade and market position.

Empirically it is difficult to sharply distinguish between these two competing theories of the mode of peasant agricultural production. The finding here that intermediate inputs are used by small farmers in approximately profit maximizing proportions weakly favors de Janvry's capitalist mode interpretation of peasant agriculture. However, whichever interpretation is correct, the overall statistical analysis strongly supports an explanation of the inverse relationship in terms of the specific characteristics (either behavioral rules, or market position) of small farm production.

The strength of the inverse relationship relative to that in earlier studies is intriguing given that the data used here were collected several years into India's green revolution. Its increase compared to pre-green revolution estimates is perhaps a reflection of further adaptation of peasant agriculture to increased pressure on the land (its involution to use Geertz's term). The finding on technical efficiency seems to ratify G. S. Bhalla's evidence that the rate of adaptation to technical change has been lower on small holdings. But if, as Bhalla and others have suggested, these techniques are scale neutral, service organizations designed to enhance the access of small farms to new technology might overcome the discrepancy in technical efficiency. Together these results suggest that small scale agriculture warrants attention as a base for agriculture development in a land scarce economy.

Georgetown University, Washington

BIBLIOGRAPHY

BARDHAN, P. K. (1973). "Size, Productivity and Returns to Scale: An Analysis of Farm Level Data from Indian Agriculture," *Journal of Political Economy* 81 (November-December): 1370–1386.

BARNUM, R. and SQUIRE, L. (1978). "Technical and Relative Economic Efficiency," *Oxford Economic Papers* 30 (July): 181–198.

BHALLA, G. S. (1974). *Changing Agrarian Structure in India: A Study of the Impact of Green Revolution in Haryana.* Delhi: Meenaskshi Prakashan.

BHALLA, S. S. (1979). "Farm Size, Productivity and Technical Change in Indian Agriculture," appendix in Berry and Cline, p. 141–193.

BHATTACHARYA, N. and SAINI, G. R. (1972). "Farm Size and Productivity: A Fresh Look," *Economic and Political Weekly* 24 (June).

BERNDT, E. and CHRISTENSEN, L. (1973a). "The Internal Structure of Functional Relationships: Separability, Substitution and Aggregation," *Review of Economic Studies* 40 (July): 403–11.

BERNDT, E. and CHRISTENSEN, L. (1973b). "The Translog Function and the Substitution of Equipment, Structure and Labor in U.S. Manufacturing, 1929–68," *Journal of Econometrics* 1 (May): 81–114.

BERRY, R. and CLINE, W. (1979). *Agrarian Structure and Production in Developing Countries.* Baltimore: Johns Hopkins University Press.

BLISS, C. and STERN, N. (1982). Palanpur: *The Economy of an Indian Village.* Oxford: Clarendon Press.

CARTER, M. R. (1982). "On the Estimation and Definition of Technical Efficiency Using Frontier and Other Omitted Variable Estimators," Unpublished.

DE JANVRY, ALAIN (1981). *The Agrarian Question and Reformism in Latin America.* Baltimore: Johns Hopkins University Press.

Economic and Statistical Organization of the Government of Haryana (1974–1976). *Studies in the Economics of Farm Management,* 1969/70–1971/72. Chandijarh: Government Press.

HECKMAN, J. (1979). "Sample Selection Bias as a Specification Error," *Econometrica* 47 (January): 153–161.

MUNDLAK, Y. (1961). "Empirical Production Function Free of Management Bias," *Journal of Farm Economics* 43 (March): 45–56.

MUNDLAK, Y. (1978). "On the Pooling of Time and Cross Section Data," *Econometrica* 46 (January): 69–85.

National Sample Survey (1962). *Technical Report on Sample Design.* Calcutta: Eka Press.

SAINI, G. R. (1971). "Holding Size, Productivity and Some Related Aspects of Indian Agriculture," *Economic and Political Weekly* 26 (June).

SEN, A. K. (1966). "Peasants and Dualism With or Without Surplus Labor," *Journal of Political Economy* 74 (October): 425–450.

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

RANI, U. (1971). "Size of Farm and Productivity," *Economic and Political* Weekly 26 (June).

VERGOPOULOS, K. (1978). "Capitalism and Peasant Productivity," *Journal of Peasant Studies* 5 (July): 446–65.

World Bank (1980), *World Development Report.* New York: Oxford University Press.

YOTOPOULOS, P. and LAU, L. (1971). "A Test for Relative Efficiency and Application of Indian Agriculture," *American Economic Review* 61 (March): 94–109.

YOTOPOULOS, P. (1973) "A Test for Relative Efficiency: Some Further Results," *American Economic Review* 63 (March): 214–23.