

Reading Room Copy

Occasional Paper No. 31

**COBB-DOUGLAS AGRICULTURAL
PRODUCTION FUNCTIONS—
A SCEPTICAL NOTE**

N. KRISHNAJI

S/813
C-1



CENTRE FOR STUDIES IN SOCIAL SCIENCES, CALCUTTA

**PUBLICATIONS OF
CENTRE FOR STUDIES IN SOCIAL SCIENCES, CALCUTTA**

OCCASIONAL PAPER SERIES :

Mimeographed Occasional Papers for limited circulation for scholarly comments and critical evaluation of first drafts are meant for publication later in journals or books : reference to subsequent publication of each of the following Occasional Papers are given in brackets :

1. Iswarchandra Vidyasagar and his Elusive Milestones (Calcutta, Riddhi-India, 1977)
ASOK SEN
2. Budget Deficit, Money Supply and Inflation (*The Contents of Economic Growth and Other Essays*, Calcutta, Research India Publications, 1977)
BHABATOSH DATTA
3. Railway Network Growth in Eastern India, 1854-1910 (*International Geography*, Vol. VI, 1976)
SUNIL MUNSI
4. Sasipada Banerjee : A Study in the Nature of the First Contact of the Bengali Bhadrakol with the Working Classes of Bengal (*Indian Historical Review*, Vol. II No. 2, January, 1976)
DIPESH CHAKRABORTY
5. Reflections on Patterns of Regional Growth in India during the Period of British Rule (*Bengal Past and Present*, Vol. XCV, Part 1, No. 180, January-June, 1976)
AMIYA KUMAR BAGCHI
6. Social Groups and Social Relations in the Town of Murshidabad, 1765-1793 (*Indian Historical Review*, Vol. II, No. 2, January, 1976)
GAUTAM BHADRA
7. Contemporary Studies on the Indian Party System : An Evaluative Account (*Socialist Perspective*, Vol. VI, No. 3, December, 1978 and Vol. VI, No. 4, March 1979)
SOBHANLAL DATTA GUPTA
8. Studies in the Constitution and Government of India : A Methodological Survey (*Teaching Politics*, Vol. IV, No. 1-2, 1978)
SHIBANI KINKAR CHAUBE
9. Demand for Electricity
NIRMALA BANERJEE
10. Comintern and the Colonial Question : The Decolonisation Controversy (*Marxist Miscellany*, No. 8, 1977 and No. 11, 1978)
SOBHANLAL DATTA GUPTA
11. Communal Riots and Labour : Bengal's Jute Mill Hands in the 1890s (forthcoming in *Past and Present*, Oxford)
DIPESH CHAKRABORTY
12. An Enquiry into the Causes of the Sharp Increase in Agricultural Labourers in North Bengal (*Economic and Political Weekly*, Vol. XII, No. 53, December 31, 1977)
NRIPENDRANATH BANDYOPADHYAY
13. Research Notes and Documents Collected by the Late Prodyot Mukherjee
ARUN GHOSH, comp.
14. Choice of Techniques and Technological Development in Underdeveloped Countries : A Critique of the Non-Neoclassical Orthodoxy (*Cambridge Journal of Economics*, June, 1978)
AMIYA KUMAR BAGCHI

Occasional Paper No. 31

COBB - DOUGLAS AGRICULTURAL PRODUCTION FUNCTIONS
- A SCEPTICAL NOTE

N. KRISHNAJI

September, 1980

CENTRE FOR STUDIES IN SOCIAL SCIENCES, CALCUTTA
10, Lake Terrace, Calcutta-700 029.

COBB-DOUGLAS AGRICULTURAL PRODUCTION FUNCTIONS
- A SCEPTICAL NOTE

N. Krishnaji*

O. SUMMARY

The paper is an attempt to show that observations generated from a fixed-coefficients-technology can be approximated well by a Cobb-Douglas function with constant returns-to-scale. The approximation is shown to be robust with respect to both spatial inhomogeneity (in input-output ratios) and aggregation, under reasonable conditions. Further, an explanation is suggested for the wayward behaviour of some fitted Cobb-Douglas functions, reflected in negative, and hence nonsensical, exponents.

Thus two absolutely conflicting hypotheses, one which posits no substitution possibility and the other which allows for smooth substitution, and hence choice of input combinations with respect to varying prices, lead to the same empirical observational material. The empirical basis of the theories that go along with a Cobb-Douglas production function can be seen to be weak in this light.

*The paper is based on work initiated nearly a decade ago in collaboration with Nirmal Chandra. Discussions with him and Krishna Bharadwaj have been of considerable help.

1. INTRODUCTION

Let us begin with some common empirical observations :

- (A) The Cobb-Douglas function fits agricultural production data well;
- (B) the fitted functions usually exhibit constant returns-to-scale, i.e. the exponents add up roughly to unity;
- (C) a high degree of multi-collinearity between input variables is often noticed;
- (D) output per acre has a tendency to be uncorrelated with the size of the farm in respect of individual crops; and
- (E) sometimes negative (and statistically significant) exponents appear in the fitted functions but even in such cases observations (A) and (B) continue to hold.

Observations (A) and (B) (the good-fitting Cobb-Douglas function and constant returns-to-scale) seem to constitute an ubiquitous feature of wide varieties of empirical data : those based on cross-sections as well as time-series and at various levels of aggregation over crops and spatial units. Observation (D), i.e., the lack of correlation between yield per acre and the size of the farm in respect of individual crops, especially in spatially homogeneous cross-sections, is one of the end products of the long debate on the "inverse relationship" between productivity per acre and the size of farm. Empirical analysis of large bodies of data shows that the inverse relationship exists only in respect of total crop output per acre and farm-size and arises out of an inverse relationship between cropping intensity and size of holding as also differences in the cropping pattern; no relationship is found between yield per acre of individual crops and the size of the farm.¹ A related

1. For an excellent and up-to-date review see, Pranjoy Lal Roy, "The Relation between Farm Size and Productivity in the Context of Alternative Modes of Production in Indian Agriculture", Unpublished Ph.D. Thesis submitted to the University of Delhi, May 1979.

empirical observation suggests the possible independence of the labour input per acre and the size of the farm. An early attempt by Chandra points towards such a possibility. A more comprehensive analysis (by Chattopadhyay and Rudra) shows, however, that the two variables are generally inversely related.²

Before discussing the fixed-coefficients-technology let us briefly consider the Cobb-Douglas function. Let us suppose that the production function (interpreted as a blue-print of technical possibilities of transforming inputs into output) is given by

$$X_0 = \text{Constant } X_1^\alpha X_2^\beta, \dots (1)$$

where X_0 stands for the output and X_1 and X_2 for quantities of inputs — we shall, throughout this paper assume, for the sake of simplicity, that there are only two inputs.

If we consider a cross-section of profit-maximising farm-firms, all facing the same set of prices, and assume further that $\alpha + \beta = 1$ (i.e., the returns-to-scale are constant), it follows from standard theory that both the input ratio (X_1/X_2) and the output-input ratios (X_0/X_1 , X_0/X_2) will be constant under optimal conditions. We thus see that the empirical observations (A), (B), (C) and (D) are all mutually consistent under the Cobb-Douglas technology (note that the observation on multi collinearity is a consequence of the constancy of the input ratio). Observation (E), relating to the appearance, sometimes, of negative exponents needs to be explained, but, more importantly, the empirical validity of the Cobb-Douglas approximation, under both aggregation and mixing of inhomogeneous data, needs to be theoretically established.

Even if we assume that this can be done, the interpretation of fitted Cobb-Douglas production functions is beset with numerous difficulties.

2. Manabendu Chattopadhyay and Ashok Rudra, Size Productivity Revisited, EPW, Review of Agriculture, Vol. XI, September 1976. For earlier work see, N.K. Chandra, "Farm Efficiency Under Semi-Feudalism : A Critique of Marginalist Theories and Some Marxist Formulations", Economic and Political Weekly, Vol. 9, August 1974.

These difficulties have been listed by Bharadwaj and there is no need repeat them here.³ We need, however, to remember a well known fact concerning the non-identifiability of the parameters of a Cobb-Douglas function in statistical estimation. This arises from the mathematical form of the production function as a result of combining it with the marginal productivity conditions (for profit-maximization): for given prices the optimally chosen input ratios remain constant across all farms. This may be regarded as the artifact underlying multicollinearity but it makes the estimated parameters meaningless: they can no longer be interpreted as the elasticities of the output with respect to the corresponding inputs. Put in a slightly different way, the range of variation in the input ratio becomes so narrow that on each production isoquant all the observations reduce practically to a single point and they would thus contain practically no "information" on technical substitution possibilities. It follows that if observations do summarise a true Cobb-Douglas technology and are consistent with profit-maximization, such a consistency itself goes against generating empirical knowledge of the underlying production function.⁴ From an economic as well as a statistical theoretical point of view, there is thus a need to look for an alternative explanation for the empirical observations and their mutual consistency.

-
3. Krishna Bharadwaj, "Technical Relations in Agriculture", in C.H. Shah (Ed.), Agricultural Development of India, Policy and Problems, Orient Longmans, 1979. An important observation that she makes is: "The different cultivating households with their diverse resource positions and their varying economic status arising therefrom face very different kind and range of choices on the market. Moreover, the markets in our agrarian economy can be, by no means, characterized as affording equal and uniform opportunity to all. Neither are 'prices' the only basis of transactions nor the participants equally free to make choices. Elsewhere we have noted how the resource position of sections of peasantry affects their market involvement and how, in turn, the peculiar characteristics of market forms and market involvements might constrain production and investment choices. Under such conditions factor and product prices play a varying role for different sections of peasants and are only one element in the process of decision-making".
 4. Marc Nerlove, Estimation and Identification of Cobb-Douglas Production Functions, Chicago, Rand McNally, 1965.

The case for considering the fixed-coefficient-technology in this context, with which we are concerned in this paper, is easily established in a heuristic manner. Let the production function for a given crop be

$$X_0 = \text{Min}(aX_1, bX_2) \quad \dots (2)$$

where we shall, for fixing ideas, call X_1 as land and X_2 as the labour inputs. This function does not allow for substitution of inputs; the isoquants will be L-shaped. Maximising the level of output would require the labour intensity per acre (the input ratio X_2/X_1) to be (a/b) . The input ratio would, under these conditions, be price-invariant and may be regarded as "traditional-knowledge-based". Actual observations can be expected to cluster around the optimum (with more or less variability - a point we discuss in detail later. A more precise formulation merely requires the independence the pairs $(X_0/X_1, X_1)$ and $(X_2/X_1, X_1)$ and does not assume that all farmers are output-maximisers - see section 2) and satisfy the approximate equality

$$X_0 \approx a X_1 \approx b X_2 \quad \dots (3).$$

This would lead to the empirical observations (C) and (D), i.e., those concerning multicollinearity and the lack of correlation between yield per acre (X_0/X_1) and farm-size (X_1). It also leads to the rough constancy of the input-ratio (i.e., the labour input per acre) across farms (Chandra, op. cit.).

From (3) we can also see that, for any λ

$$\begin{aligned} X_0 &\approx (a X_1)^\lambda (b X_2)^{1-\lambda} \\ &\approx \text{constant } X_1^\lambda X_2^{1-\lambda} \quad \dots (4) \end{aligned}$$

which is a Cobb-Douglas function with constant returns-to-scale. This leads to empirical observations (A) and (B).

Since the approximate equality (4) holds for all λ the question that arises is : what can be said, about the least squares estimate of λ (or the parameters of a Cobb-Douglas function, more generally) for a given set of data? We need to note that no meaning can be attached to λ in terms of the function (2). The value realised through estimation would obviously depend on the exact nature of (3), i.e., the range of variation in the input ratio. The answer requires a more precise formulation of (2), which we attempt in the next section.

All the foregoing is based on heuristic reasoning; precise formulations follow. However, it is not being suggested here that the production and decision-making processes for Indian agriculture can be summarised wholly by a fixed-coefficients-technology. The aim of the paper is to demonstrate, in theoretical terms, that a no-substitution technology can lead to empirical observations commonly taken as strong support for the Cobb-Douglas function combined with marginal productivity theory; for understanding production and decision processes much more than this negative demonstration is required.

2 HOMOGENEOUS CROSS-SECTIONS

(2.1) The Model

Let us suppose that we have a random sample of farms from a population in which the distribution of land (X_1) is given. We shall formulate the model in terms of the logarithms of the variables : $x_0 = \log X_0$, $x_1 = \log X_1$ and $x_2 = \log X_2$. All the variables refer to a single crop. We assume that the observations are generated by the following model :

(i) x_1 is a given random variable ; (5.1)

(ii) $x_2 = x_1 + \alpha - \beta + v$, where α and β are constants and v is a random variable independent of x_1 ; and (5.2)

(iii) $x_0 = \text{Min} (\alpha + x_1, \beta + x_2)$ (5.3)

The interpretation is straightforward. Since $(x_2 - x_1)$ is the logarithm of the labour input per acre, $(\alpha - \beta)$ measures the logarithm of the optimal input ratio for from (5.3), at the optimal combination, $x_0 = \alpha + x_1 = \beta + x_2$. Hence v measures the deviation of the logarithm of actual input ratio from the logarithm of its optimal value.

NOTE

From now on we shall drop the prefix "logarithm of " and speak of $(x_2 - x_1)$ as the input ratio, $(x_0 - x_1)$ as the output per acre etc.

Remark 1. (5.1) to (5.3) constitute a simple stochastic version of the fixed-coefficient-technology. Given land, the labour input is determined by (5.2) and the output by (5.3). The model allows for variation in the input ratio from farm to farm. Instead of (5.2) one can write $X_2 = X_1 P$ where P and X_1 are independent random variables (i.e., the input ratio and farm-size are independent). But this can be reduced to the form (5.2). The assumption in (5.2), viz., the independence of v (i.e., "errors" in labour intensity per acre) and x_1 (size of farm) is crucial to the model, as we shall see. We may regard this along with constancy of α and β , as the defining criterion for the homogeneity of the data. The assumption is crucial in another sense. The errors will be small and concentrated around the optimum only if all the farmers are assumed to maximise the output. The model does not require this assumption. For this reason, the deviations from the optimum can have a large variance. What we do stipulate is that these deviations should not be systematically related to the size of the farm. The model is thus free of uniform behaviouristic assumptions for all households.

Remark 2. (5.3) can be rewritten as

$$x_0 = \alpha + x_1 + u, \text{ where } u = \text{Min}(0, v). \quad \dots \quad (6)$$

Since x_1 and v are independent, it follows that x_1 and u are also independent. Since α measures (see 5.3) the optimal output per acre, u is a random variable which measures the deviation of the actual yield per acre from its optimal value.

Remarks 3. Since ν measures the deviation in the labour intensity per acre from its optimal value $E(\nu) \geq 0$ are all possible: it is quite possible that all the farms may apply sub-optimal labour intensities. On the other hand $E(u) \leq 0$ always, since by definition $u = 0$ if $\nu > 0$ (i.e., output per acre does not change from its optimal level with an increase in the labour input per acre once the latter crosses the optimal input ratio) and $u = \nu$ if $\nu < 0$.

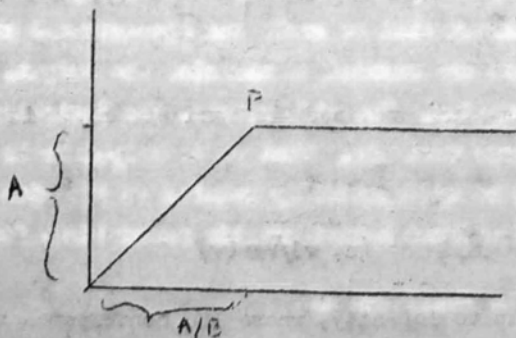
The production function (5.3) can be rewritten in terms of the basic variables as

$$X_0 = \text{Min} (A X_1, B X_2)$$

where $A = \exp(\alpha)$ and $B = \exp(\beta)$. From this it is easy to sketch the graph of output per acre (X_0/X_1) as a function of labour input per acre (see Figure (1)).

Figure (1)

Output per acre as a function of labour input per acre



The figure represents the function $X_0/X_1 = \text{Min} (A, B X_2/X_1)$

$$\text{i.e., } (X_0/X_1) = A, \text{ if } (X_2/X_1) \geq (A/B)$$

$$= B (X_2/X_1), \text{ if } (X_2/X_1) < (A/B).$$

We note that in the range of sub-optimal labour use, the relationship is linear and further, after logarithmic transformation, $(x_0 - x_1) = \beta + (x_2 - x_1)$ for $(x_2 - x_1) < (\alpha - \beta)$; the relationship thus remains linear. This has implications for the results of Cobb-Douglas approximation.

(2.2) Fitting a Cobb-Douglas Function

We may now raise the question : what happens when a Cobb-Douglas function is fitted to the data generated by this model? The basic equations are

$$x_2 = x_1 + \alpha - \beta + v, \quad (x_1, v) \text{ independent} \quad \dots (7.1)$$

$$x_0 = x_1 + \alpha + u \quad (x_1, u) \text{ independent} \quad \dots (7.2)$$

$$u = \text{Min}(0, v) \quad \dots (7.3)$$

Suppose we estimate, by the method of least-squares, the equation

$$x_0 = \text{constant} + b_1 x_1 + b_2 x_2 \quad \dots (8)$$

which is a Cobb-Douglas function with no restriction on the parameters (constant returns-to-scale would require $b_1 + b_2 = 1$). Let the estimated parameters be \hat{b}_1 and \hat{b}_2

We shall show that, provided the variance of v is finite.

$$\hat{b}_1 + \hat{b}_2 \xrightarrow{P} 1 \quad \dots (9)$$

$$\text{and } \hat{b}_2 \xrightarrow{P} \text{cov}(u, v)/\text{Var}(v) \quad \dots (10)$$

as the sample size tends to infinity, where the convergence refers to convergence in probability. (Throughout this paper convergence means only this type of convergence; we shall accordingly delete the symbol 'p' over the arrow, as well as the expression $n \rightarrow \infty$).

The implications are that in large samples the returns-to-scale parameter of the estimated Cobb-Douglas function approximates unity and that the estimated exponent of labour approximates the same value which will be

obtained by regressing output per acre on labour-input per acre in a constrained Cobb-Douglas form; i.e.

$$X_0 = \text{constant } X_1^{1-b_2} X_2^{b_2}$$

or $(X_0/X_1) = \text{constant} \left(\frac{X_2}{X_1} \right)^{b_2} \dots \dots (11)$

To demonstrate these results we note that

$$\hat{b}_1 = \Delta^{-1} (S_{22} S_{01} - S_{12} S_{02}) \dots \dots (12)$$

and $\hat{b}_2 = \Delta^{-1} (S_{11} S_{02} - S_{12} S_{01})$

where $\Delta = S_{11} S_{22} - S_{12}^2$ and S_{ij} denotes the corrected sum of products between x_i and x_j ($i, j = 0, 1, 2$).

But we have from (7.1) and (7.2)

$$\left. \begin{aligned} S_{12} &= S_{11} + S_{1v}, \quad S_{22} = S_{11} + S_{vv} + 2S_{1v}, \quad S_{01} = S_{11} + S_{1u}, \\ \text{and } S_{02} &= S_{11} + S_{1u} + S_{1v} + S_{uv} \end{aligned} \right\} \dots \dots (13)$$

where S_{1u} is the corrected sum products between x_1 and u (S_{1v} and S_{uv} being defined in the same way).

It is easy to verify that substitution of (13) in (12) yields

$$\hat{b}_2 = (S_{11} S_{uv} - S_{1u} S_{1v}) / (S_{11} S_{vv} - S_{1v}^2) \dots \dots (14.1)$$

$$\text{and } \hat{b}_1 + \hat{b}_2 - 1 = (S_{1u} S_{vv} - S_{1v} S_{uv}) / (S_{11} S_{vv} - S_{1v}^2) \dots \dots (14.2)$$

Since (x_1, u) and (x_1, v) are independent and the variance of v is assumed to be finite, we have

$$\left. \begin{aligned} S_{1u}/n &\rightarrow 0, & S_{1v}/n &\rightarrow 0 \\ S_{11}/n &\rightarrow \text{Var}(x_1), & S_{vv}/n &\rightarrow \text{Var}(v) \\ \text{and } S_{uv}/n &\rightarrow \text{Cov}(u, v) \end{aligned} \right\} \dots \dots (15)$$

The results displayed at (9) and (10) now follow from (14.1), (14.2) and (15).

(2.3) The Goodness of Approximation

Since $x_0 = x_1 + \alpha + u$ and x_1 and u are independent it follows that

$$S_{00}/n \rightarrow \text{Var}(x_1) + \text{Var}(u)$$

$$\text{and } S_{01}/n \rightarrow \text{Var}(x_1)$$

Similarly, from (7.1) and (13), it follows that

$$S_{02}/n \rightarrow \text{Var}(x_1) + \text{Cov}(u, v).$$

Since $R^2 = \frac{(\hat{b}_1 S_{01} + \hat{b}_2 S_{02})/n}{S_{00}/n}$, it is easy to see from (9) and (10)

$$\text{that } R^2 \rightarrow \frac{(1-b_2^*) \text{Var}(x_1) + b_2^* [\text{Var}(x_1) + \text{Cov}(u, v)]}{\text{Var}(x_1) + \text{Var}(u)} \dots \dots (16)$$

where $b_2^* = \text{Cov}(u, v)/\text{Var}(v)$, the limit of \hat{b}_2

A simplification of (16) now yields

$$R^2 \rightarrow \frac{[\text{Var}(x_1)/\text{Var}(u)] + r^2(u, v)}{[\text{Var}(x_1)/\text{Var}(u)] + 1} \text{ as } n \rightarrow \infty \dots \dots (17)$$

where $r(u, v)$ is the correlation coefficient between u and v . This shows that the Cobb-Douglas approximation will be good provided the productivity per acre and labour input per acre (both measured in logarithms) are highly correlated. If the data are generated by a Cobb-Douglas function this is what is expected but as we can readily see from Fig. (i) even data generated by the fixed-coefficient technology can yield high correlations between (X_0/X_1) and (X_2/X_1) : this would happen especially if there are not too many observations in the sample with values of labour input per acre far higher than the optimal input ratio (for such observations would fall along the horizontal section of

the curve in Fig (i) and may weaken the correlation if they are far removed from the break in the curve : the point P).

But what is more important from our point of view here is that the expression in (17) shows that even when $r(u, v)$ is not very high, R^2 can be high provided $[\text{Var}(x_1)/\text{Var}(u)]$ is large, i.e., when the variation in the farm-size is large relative to the variation in productivity per acre. Since observations drawn from a fixed-coefficients - technology can be expected to yield low variations both in the input-ratio and the output-input ratios, the largeness in $\text{Var}(x_1)/\text{Var}(u)$ will be guaranteed so long as the sample includes both small and big farms. This is a curious aspect of the confounding between the two models : the stronger the fixed-coefficient model holds, the better the Cobb-Douglas function fits.

(2.4) The Meaning of the Estimated Parameters

The estimated parameters \hat{b}_1 and \hat{b}_2 of the Cobb-Douglas function have no meaning in terms of the model assumed to generate the data, i.e., (5.1) to (5.3). All the same since they are interpreted by those who fit the Cobb-Douglas function as the elasticities of the output (with respect to the corresponding inputs) it is necessary to ask under what conditions they appear "sensible" from this point of view. We know that $\hat{b}_1 + \hat{b}_2 \rightarrow 1$ and $\hat{b}_2 \rightarrow \text{Cov}(u, v)/\text{Var}(v)$.

We shall show that the model (5.1) to (5.3) guarantees

$$0 \leq b_2^* = \frac{\text{Cov}(u, v)}{\text{Var}(v)} \leq 1 \quad \dots (18)$$

so that the parameters always appear to be sensible in the light of a Cobb-Douglas interpretation.

As already noted $\mu = E(v) \geq 0$ are all possible. Let us write $f(\cdot)$ for the density of v and

$$\text{Prob}(v \leq 0) = p; \quad q = 1-p; \quad E(\sqrt{v} \leq 0) = \mu_-$$

$$E(\sqrt{v} > 0) = \mu_+; \quad \text{Var}(\sqrt{v} \leq 0) = \sigma_-^2 \quad \text{and} \quad \text{Var}(\sqrt{v} > 0) = \sigma_+^2$$

Since $u = 0$ if $v \geq 0$ and $u = v$ if $v < 0$, we have

$$(a) E(u) = \int_{-\infty}^0 v f(v) dv = p \mu_- \leq 0 \text{ always}$$

$$(b) E(u^2) = \int_{-\infty}^0 v^2 f(v) dv = p (\sigma_-^2 + \mu_-^2) \text{ and}$$

$$\text{Var}(u) = p (\sigma_-^2 + \mu_-^2) - p^2 \mu_-^2 = p \sigma_-^2 + pq \mu_-^2$$

$$(c) \text{Cov}(u, v) = \int_{-\infty}^0 v^2 f(v) dv - p \mu_- \mu_- = p \sigma_-^2 + p \mu_- (\mu_- - \mu)$$

But since $\mu = p \mu_- + q \mu_+$ the last expression simplifies to

$$p \sigma_-^2 + pq (\mu_- - \mu_+) \mu_- \geq 0 \dots (19)$$

since $\mu_- \leq 0$ and $\mu_+ \geq 0$

Therefore, $\text{Cov}(u, v) \geq 0$ always.

(d) If $\text{Prob}(v < 0) = 1$, then $u = v$ with probability unity and then $\text{Cov}(u, v) = \text{Var}(v)$ implying $b_2^* = 1$. Let us rule out this case.

Since $v - u = v$ if $v > 0$ and $v - u = 0$ if $v \leq 0$, we have

$$\text{Var}(v) - \text{Cov}(v, u) = \text{Cov}(v, v - u)$$

$$= \int_0^{\infty} v^2 f(v) dv - E(v) E(v - u)$$

$$= q (\sigma_+^2 + \mu_+^2) - \mu (\mu_- - p \mu_+)$$

$$= q \sigma_+^2 + pq \mu_+^2 - pq \mu_+ \mu_- > 0 \dots (20)$$

since $\mu_- < 0$ (the last equation is derived by substituting $\mu = p \mu_- + q \mu_+$)

From (19) and (20) it follows that

$$0 \leq b_2^* = \text{Cov}(u, v) / \text{Var}(v) \leq 1$$

As already noted $b_2^* = 1$ if $\text{Prob}(v \leq 0) = 1$, i.e., when all the farms use labour intensity per acre lower than the optimal level. That in this case \hat{b}_2 will be close to unity is otherwise obvious from Fig (i), since we have in the range $x_2 - x_1 < \alpha - \beta$,

$$x_0 - x_1 = \beta + (x_2 - x_1)$$

and the slope of $(x_0 - x_1)$ on $(x_2 - x_1)$ is unity.

The other extreme case occurs when $q = 1$, i.e., when all farms have higher than optimal input ratios. From (19), we see that in this case $\text{Cov}(u, v) = 0$, and hence $b_2^* = 0$, which is also obvious for in Fig (1) this corresponds to the case of all observations falling along the horizontal section of the curve.

Barring these two extreme cases we get exponents lying between 0 and 1, provided the sample size is large.

3. MIXED DATA

The independence of the input-ratio and the size of the farm, and the constancy of the optimal ratios (α and β), are the defining criteria for the homogeneity of data. While independence may hold in subsets of data (say, within villages) the optimal ratios may differ between subsets, destroying the overall homogeneity. We shall briefly investigate the consequences of fitting Cobb-Douglas functions to such data.

(3.1) The Case when Mean Farm-size is Homogeneous

Let us suppose that the population from which the sample is drawn is a mixed one of, say k , villages, with π_j and p_j being respectively the population and sample proportions in the mixture.

We may rewrite the model simply as

$$x_2 = x_1 + \theta \quad \dots \dots (21.1)$$

$$x_0 = x_1 + \delta \quad \dots \dots (21.2)$$

where θ and δ can no longer be assumed to be independent of x_1 . For

village but varies from village to village and v , a random deviation from the optimal value which is assumed to be independent of x_1 within each village. Similarly δ is the sum of a constant α which varies from village to village and a random deviation u (independent of x_1 within villages).

The correlation between the pairs (x_1, θ) and (x_1, δ) arises out of inter-village heterogeneity.

From (21.1), if the overall sum of products between x_1 and θ is S and the sum within village i is S_i , it easily follows from the decomposition of S into its "within" and "between" components that

$$S = \sum_{i=1}^k S_i + (\sum n_i \bar{x}_{1i} \bar{\theta}_i - n \bar{x}_1 \bar{\theta}) \dots (22)$$

where n_i is the sample size for the i th village, \bar{x}_{1i} and $\bar{\theta}_i$ are the corresponding village means and \bar{x}_1 and $\bar{\theta}$ the overall means (i.e., $\bar{x}_1 = \sum p_i \bar{x}_{1i}$ etc).

Dividing (22) by the sample size we get

$$S/n = \sum p_i (S_i/n_i) + (\sum p_i \bar{x}_{1i} \bar{\theta}_i - \bar{x}_1 \bar{\theta}) \dots (23)$$

If we assume intra-village independence between x_1 and v , it follows that $S_i/n_i \rightarrow 0$ but it will no longer be true in general that $S/n \rightarrow 0$ in view of the "between-village" contribution in (23). However, it is easy to see that if all the villages have the same mean farm-size (say μ') then $S/n \rightarrow 0$. For, in this case

$$p_i \rightarrow \pi_i, \bar{x}_{1i} \rightarrow \mu', \bar{\theta}_i \rightarrow \mu' \dots (24)$$

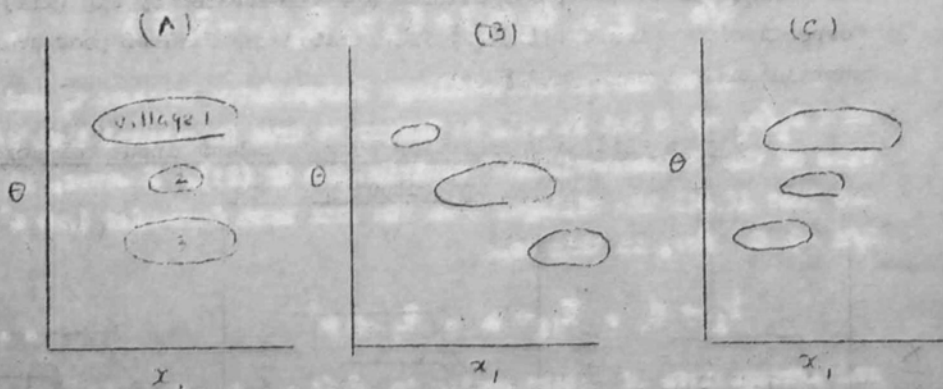
and irrespective of heterogeneity in θ_i (i.e., in α_i, β_i and the possibly in the distribution of v_i)

$$(\sum p_i \bar{x}_{1i} \bar{\theta}_i - \bar{x}_1 \bar{\theta}) \rightarrow 0. \text{ (this follows from (24) and the fact that } \bar{\theta} = \sum p_i \bar{\theta}_i \text{).}$$

Since the same argument applies to the correlation between x_1 and θ , it follows that, so long as the land distributions in the different villages have the same mean, mixing of data based on heterogeneous optimal input and input-output ratios does not alter the conclusions of the last section: with $S_{1\theta}/n \rightarrow 0$ and $S_{1\delta}/n \rightarrow 0$, it will be still be true that $\hat{b}_1 + \hat{b}_2 \rightarrow 1$ and $\hat{b}_2 \rightarrow \text{Cov}(u, v)/\text{Var}(v)$ (but the covariance and variance in the latter would now contain "between-village" components introduced by inter-village differences). These follow from the validity of (15) with θ and δ replacing v and u respectively.

The robustness of Cobb-Douglas fits to mixed data is thus preserved when mean farm size remains constant in the different parts of the mixture. When the mean farm size also varies then the "between-village" component can introduce a correlation which can be either positive or negative. This is illustrated in Figure (ii).

Figure (ii)
Correlation between x_1 and θ for mixed data



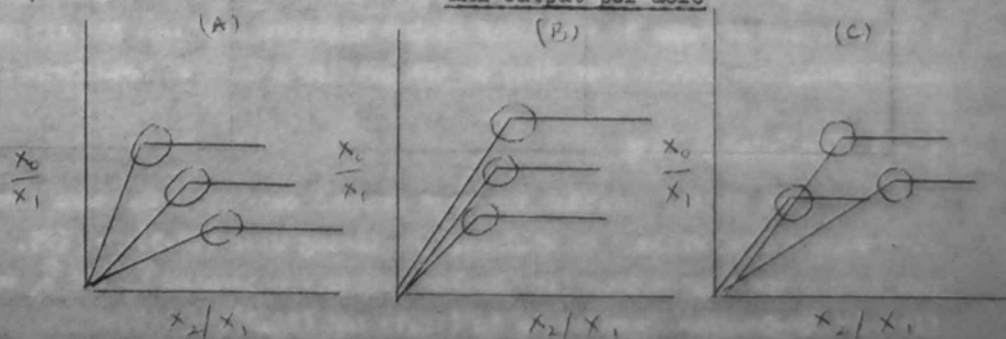
With observations for each village falling within the corresponding elliptical region, Fig (ii) shows the results of independence of x_1 and θ within each village combined with heterogeneity in the input ratio. (A) illustrates the case we have already discussed: i.e., mean farm size is the

same in all villages. In this case (x_1, θ) will continue to be uncorrelated. (B) and (C) show the possibility of inverse and positive correlations respectively. The inverse relationship between labour intensity per acre and size of farm in mixed data arises, inspite of their independence within each villages, out of high optimal input ratios being a feature (for whatever reason) of villages with a small average farm-size and vice-versa. Note that these phenomena do not depend, for their validity, on the differences in the variance of the distribution of land (i.e., these results are independent of the spread of the x_1 values within the elliptical regions shown in Fig (ii)).

Negative Exponents

However it will no longer be true in this case that $0 \leq b_2^* \leq 1$, where $b_2^* = \text{Cov}(\theta, \delta) / \text{Var}(\theta)$, the limit of the exponent of labour. For the heterogeneity in α and β can give rise to inverse correlation between θ = labour input per acre and δ = output per acre; the exponent can also be higher than unity (with $\hat{b}_1 + \hat{b}_2 \rightarrow 1$, this will imply a negative exponent for land). These two possibilities are illustrated by Fig (iii) (A) and (B) respectively. Figure (iii) (C) illustrates the case of poor over all correlation between θ and δ .

Figure (iii) : Correlation between Labour input per acre and output per acre



For each village the break in the curve represents the point, the x co-ordinate of which is the optimal input ratio and the y co-ordinate is the optimal output per acre. The actual observations are expected to cluster around the optimal points within the encircled regions. Figure (iii) illustrates the possible effect of heterogeneity in the input and input-output ratios.

In both cases (A) and (B), $r(\theta, \delta)$ can be expected to be high and hence the overall R^2 obtained by fitting a Cobb-Douglas function will be high (the limit to R^2 , shown in (17) being valid with θ and δ replacing v and u respectively). (A) would yield a negative exponent for labour while the case (B) would correspond to an estimated negative exponent for land (the exponent for labour being more than unity).

Case (C) yielding poor overall correlation between θ and δ , resulting from heterogeneity requires a comment. As we have already remarked on the basis of (17), the overall R^2 can be high even when $r(\theta, \delta)$ is low; this will happen provided $\text{Var}(x_1)/\text{Var}(\delta)$ is high, i.e., the variance of farm size is high in relation to the variance of the productivity per acre. Since the latter is expected to be high in mixed data, a pattern of empirical data such as that in Fig (iii) C cannot usually be well approximated by a Cobb-Douglas function.

However, the cases (A) and (B) show that a Cobb-Douglas function with constant returns to scale and with a negative exponent can fit some mixed data, the result depending on the precise nature of heterogeneity.

(3.2) The General Case

Let us now consider the more general case when the mean farm-size also varies from village to village. Beginning with (21.1) and (21.2) we can derive the expressions for the last squares estimates \hat{b}_1 and \hat{b}_2 of a fitted Cobb-Douglas function as in (14.1) and (14.2).

$$\hat{b}_2 = (s_{11} s_{\theta\delta} - s_{1\theta} s_{1\delta}) / \Delta \quad (25.1)$$

$$\text{and } \hat{b}_1 + \hat{b}_2 - 1 = (s_{1\theta} s_{\theta\theta} - s_{1\theta} s_{\theta\delta}) / \Delta \quad (25.2)$$

$$\text{where } \Delta = s_{11} s_{\theta\theta} - s_{1\theta}^2 \quad (25.3)$$

The expression in (23) shows that because of inter-village variation in mean land size, it will no longer be true that $s_{1\theta}/n \rightarrow 0$.

However, the exact expressions for R^2 is easily derived from the structure (21.1) and (21.2), the identities in (13) (with θ and δ replacing v and u respectively) and the solution to the normal equations given by (25.1), (25.2) and (25.3).

Starting from $R^2 = (\hat{b}_1 s_{01} + \hat{b}_2 s_{02}) / s_{00}$ and making the relevant substitutions, it is easy to show after some algebraic manipulation - omitted here) that

$$1 - R^2 = \frac{s_{\delta\delta}(1 - r_{\theta\delta}^2) - s_{11}(1 - r_{1\theta}^2)(1 - \hat{b}_1 - \hat{b}_2)^2}{s_{11} + 2s_{1\delta} + s_{\delta\delta}} \quad (26)$$

That a high correlation between θ and δ is a sufficient condition for a good approximation by a Cobb-Douglas function with constant returns-to-scale can be easily seen from (26). First note that $1 - R^2 > 0$ always. Now the second expression in the numerator of the RHS in (26) appears with a negative sign and all its factors are positive. Hence if $r(\theta, \delta)$ is high and, as a consequence, the first expression in the numerator low, the second expression has also be to be low (since $1 - R^2$ is positive). This shows that

$$r^2(\theta, \delta) \approx 1 \Rightarrow (1 - r_{1\theta}^2)(1 - \hat{b}_1 - \hat{b}_2)^2 \approx 0 \quad \dots (27)$$

$$\Rightarrow 1 - R^2 \approx 0 \quad \dots (28)$$

From (27) it follows that in this case

$$r_{1\theta}^2 \approx 1 \text{ and/or } \hat{b}_1 + \hat{b}_2 \approx 1 \quad \dots (29)$$

We thus see that $r(\theta, \delta)$ being high is a sufficient condition for obtaining a good fitting Cobb-Douglas function with exponents roughly adding up to unity even in the case of mixed data with heterogeneity in farm size distributions as well as in the input and input-output ratios. In Figure (iii) cases (A) and (B) illustrate the possibility of $r(\theta, \delta)$ being high in such data. We can have, at the same time, a high correlation (both positive and negative correlations being possible) between farm size (x_1) and productivity (δ).

Remark $r(\theta, \delta)$ is expected to be high, if the observations are generated by a Cobb-Douglas technology. The above arguments show, however, that it can be high and lead to good Cobb-Douglas fits, even when the data are generated by a no-substitution technology.

The Cobb-Douglas approximation thus remains robust for certain kinds of mixed data and all empirical observations, including the appearance, on occasion, of negative exponents remain consistent with a fixed-coefficient technology.

4 AGGREGATION

(4.1) Aggregation over Crops

7some-
times Thus far we have considered a single crop. While output per acre (and labour input per acre) are empirically found to be uncorrelated with farm-size in respect of individual crops, inverse correlations between both these pairs are found to emerge as a result of aggregation over crops. This is because of higher cropping intensities being associated with small holdings.⁵

However we can still write $x_2 = x_1 + \theta$ and $x_0 = x_1 + \delta$, where the variables now refer to values aggregated over all crops, with (x_1, θ) and (x_1, δ) being inversely correlated (an assumption of positive correlations does not alter the conclusions). Formally, this case has already been

5. Prannoy Lat Roy, op.cit.

5/8/13

covered in the last section and the conclusions arrived at there apply to the case of aggregation over crops.

(4.2) Aggregation over spatial units

Let us rewrite the model as

$$X_2 = PX_1 \dots (30.1)$$

$$\text{and } X_0 = QX_1 \dots (30.2)$$

where random variables (X_1, P) and (X_1, Q) are independent, X_1 denoting farm-size, X_2 labour input and X_0 output as before.

Consider (30.1) first. If we have data on k units we can write

$$\begin{aligned} \sum_{i=1}^{k_1} X_{2i} &= \sum_{i=1}^k P_i X_{1i} \\ &= X_{1.} \sum (P_i X_{1i} / X_{1.}) \dots (31) \end{aligned}$$

where $X_{1.} = \sum X_{1i}$ and X_{1i} is independent of X_{1j} ($j \neq i$) and X_{1i} is independent of all P_j .

Now, it is well-known in probability theory that if Z_1, Z_2, \dots, Z_n are independent then

$Z_i / (Z_1 + Z_2 + \dots + Z_n)$ is independent of $(Z_1 + Z_2 + \dots + Z_n)$ for each i if and only if each Z_i has a gamma distribution with the same scale parameter.⁶

By appealing to this result, we see that if land is assumed to be distributed as a gamma variable, it follows from (31) that

6. The case $n = 2$ was first proved in, Eugene Lukacs (1955), A characterization of the gamma distribution, Annals of Mathematical Statistics, 26, pp. 319-324.

$P_i X_{1i}/X_1$ is independent of X_1 , for each i and hence (31) can be rewritten as

$$X_2 = P \cdot X_1 \dots (32)$$

where P and X_1 are independent and the dots in the suffix indicate aggregation (X_1 and X_2 represent total land and total labour while P would give a weighted average of the labour input acre, the weights being the values of the farm size).

The same argument applies to the aggregation of (30.2) and we thus see that the basic model represented for homogeneous cross-sections (displayed by (7.1) and (7.2) remains valid under aggregation. This is especially true of aggregation over villages or even larger units leading to larger cross-sections : it is fairly common to use, for example, inter-district or inter-state cross-sectional averages of land-size etc. for carrying out Cobb-Douglas analysis.

Aggregation over spatial leading to time-series data (for example data for a state such as West Bengal) cannot, however, be covered by (32) since the input-ratio can change systematically over time. This only means that we now have to discard the assumption of independence between P and X_1 , but the assumption of correlation, as we have seen in the last section, does not invalidate the consistency between the empirical observations and the hypothesis of a fixed-coefficients-technology.

Remark Some empirical analysis suggests that the land distributions (based on the National Sample Surveys) can be graduated well by gamma distributions and that, in this respect, the gamma distribution performs uniformly better than the log-normal distribution which is commonly used for the purpose.⁷

The empirical fits of the gamma law are sufficient for our purpose to demonstrate the robustness of the Cobb-Douglas approximation. Whether they

7. N. Krishnaji, "Some Analytical Characteristics of the Distribution of land: An Exercise in Graduation by the Lognormal, Gamma and Loggamma Laws" Working Paper No.18, Centre for Development Studies, Trivandrum, 1974.

show something more and whether the theories underlying the gamma distribution can be regarded as superior to those underlying the log-normal law, is a different matter.⁸

5. CONCLUDING REMARKS

We can give a purely statistical description of our main results : when the range of variation in the input and input-output ratios is narrow a Cobb-Douglas function with constant returns-to-scale fits the data well, but the estimated parameters convey no information (the input-ratio variation being small) about substitution possibilities. There is no way of dismissing the possibility of the data being generated by a process characterised by fixed coefficients. These conclusions hold for wide varieties of data regardless of heterogeneity (in the ratios as well as the distribution of land) and aggregation (over crops as well as spatial units).

The question of substitution possibilities is important, but to demonstrate the latter we need data which exhibit a wide range of variation in the input-ratio corresponding to the same level of output. But marginal-productivity theory, with which the constant returns-to-scale Cobb-Douglas function is combined to demonstrate the consistency between theory and empirical observations, itself goes against the generation of such data in homogeneous cross-sections.

On the other hand, the fixity of coefficients alone suffices to explain the empirical observations without the intervention of any theory of uniform behaviour of profit maximization on the part of all cultivating households. The extreme counter-example of no-substitution that we have discussed in this paper leads to the conjecture that empirical observations will remain consistent under intermediate hypotheses which allow for limited

8. In fact, both sets of theories depend on purely chance mechanisms to explain the distribution; since they ignore historical sources of inequality in land distribution, there is here, as in the Cobb-Douglas case, more to it than meets the eye in the good fits.

substitution possibilities. But it is doubtful if knowledge of such possibilities, or of the production process in general, can be generated by fitting regression equations. Likewise, in the light of questions raised by Bharadwaj concerning the diversity of choices open to different households, it is doubtful whether a production function approach, combined with uniform behavioural assumptions, is at all useful.

15. On the Scientific Study of Politics : A Review of the Positivist Method (*The State of Political Theory : Some Marxist Essays*. Calcutta, Research India Publications, 1978)
PARTHA CHATTERJEE
16. Trade and Empire in Awadh, 1756-1804
RUDRANGSHU MUKHERJEE
17. The Ethnic and Social Bases of Indian Federalism
SHIBANI KINKAR CHAUBE
18. বাংলা সংবাদ-সাময়িকপত্রে যতিচিহ্নের ব্যবহার, ১৮১৮-১৮৫৮
(Use of Punctuation Marks in the Bengali Journalistic Prose, 1818-1858)
DEBES ROY
19. Medieval Northeast India : Polity Society and Economy, 1200-1750 A. D. (forthcoming in *The Cambridge Economic History of India*, Vol. I)
AMALENDU GUHA
20. The Colonialist Premise in the British Occupation of Bengal : Contributions by Clive and Pitt, the Elder, during 1757-59 (*Proceedings of the Indian History Congress*, Bhubaneswar, 1977)
BARUN DE
21. Thinking About Ideology : In Search of an Analytical Framework
PARTHA CHATTERJEE
22. Material Conditions and Behavioural Aspects of Calcutta Working Class 1875-1899
RANAJIT DAS GUPTA
23. An Essay on John Rawls' Theory of Distributive Justice and its Relevance to the Third World
A. P. RAO
24. Impact of Plantations on the Agrarian Structure of the Brahmaputra Valley
KEYA DEB
25. Assamese Peasant Society in the late Nineteenth Century : Structure and Trend (*The Indian Economic and Social History Review*, Vol. XVII, No. 1)
AMALENDU GUHA
26. Of Trade and Traders in Seventeenth Century India : An Unpublished French Memoir by George Roques
INDRANI RAY
27. Pattern of Organisation in Hardloom Industry of West Bengal (forthcoming in *Social Scientist*, Trivandrum)
ABANTI KUNDU
28. Foreign Technical Collaboration in Indian Business Houses, 1957-76 : A Quantitative Analysis (*Business Standard*, Annual Number, 1980)
SUBHENDU DASGUPTA
29. The Multiple Faces of the Early Indian Merchants
INDRANI RAY
30. Agrarian Relations and Politics in Bengal : Some Considerations on the Making of the Tenancy Act Amendment 1928
PARTHA CHATTERJEE

PROCEEDINGS OF CONFERENCES & SEMINARS :

1. Problems of the Economy and Planning in West Bengal (CSSSC, 1974)
2. Change and Choice in Indian Industry (in press)

PERSPECTIVES IN SOCIAL SCIENCES ;

Volumes of essays on a common theme by scholars in the Centre to be periodically published :

Historical Dimensions (Calcutta, Oxford University Press, 1977)

MONOGRAPHS

Results of research work individually undertaken by the Centre's staff :

1. SUNIL MUNSI : Geography of Transportation in Eastern India under the British Raj. Calcutta, K. P. Bagchi & Co. 1980
2. NIRMALA BANERJEE : Demand for Electricity. Calcutta, K. P. Bagchi & Co., 1979
3. SOBHANLAL DATTA GUPTA : Comintern, India and the Colonial Question : 1920-1937. Calcutta, K. P. Bagchi & Co , 1980

PUBLIC LECTURES ;

1. ASHOK MITRA : Terms of Exchange and Accumulation : The Soviet Debate (R. C. Dutt Lectures on Political Economy, 1975). Calcutta, Orient Longman, 1977
2. KRISHNA BHARADWAJ : Classical Political Economy and Rise to Dominance of Supply and Demand Theories (R. C. Dutt Lectures on Political Economy, 1976). Calcutta, Orient Longman, 1978
3. B. N. GANGULI : Some Aspects of Classical Political Economy in the Nineteenth Century Indian Perspective (R. C. Dutt Lectures on Political Economy, 1977) Calcutta, Orient Longman, 1979
4. I. S. GULATI : International Monetary Development and the Third World : A Proposal to Redress the Imbalance (R. C. Dutt Lectures on Political Economy, 1978) in press
5. V. M. DANDEKAR : Peasant-Worker Alliance : Its Basis in the Indian Economy (R. C. Dutt Lectures on Political Economy, 1979) in press