Vol XVII
No. 3

JULY-
SEPTEMBER
1962

ISSN 0019-5014

INDIAN JOURNAL OF AGRICULTURAL ECONOMICS

INDIAN SOCIETY OF
AGRICULTURAL ECONOMICS,
BOMBAY

A STUDY OF STATISTICAL METHODOLOGY IN FARM COST SURVEYS

S. K. Raheja*

INTRODUCTION

A number of farm cost and management surveys have been undertaken in India in recent years and these present several methodological problems. One of these is the choice between what are known as the Survey and the Cost Accounting (or the Farm Recording) methods of farm cost enquiry. The Directorate of Economics and Statistics of the Ministry of Food and Agriculture, Government of India in collaboration with the Research Programmes Committee of the Planning Commission carried out from 1954-55 to 1956-57 Farm Management Surveys of an exploratory nature in five regions, *viz.*, the States of U.P., Punjab, West Bengal, Bombay and Madras, employing both methods of collection of data. A study of the relative merits of the two methods was one of the principal objectives of the surveys. The data collected permitted a comparison of the two methods as well as a study of efficiency of the design of sampling adopted in the surveys and were also useful for a study of input-output relationship in respect of the crops covered. The present investigation utilizes the data obtained in the survey in the U.P. region in the crop year 1954-55 and aims at

(*i*) a comparison of the Survey and Cost Accounting methods of collecting primary data ;

(*ii*) a study of efficiency of the design adopted in the survey ;

(*iii*) calculation of sample size required for given precision ; and

(*iv*) a study of input-output relationship for certain crops.

Some data on sugarcane crop for the same region for the year 1955-56 collected by the Cost Accounting method were available from a cost of cultivation survey conducted by the Indian Central Sugarcane Committee. They were also utilized for a study of input-output relationship for sugarcane.

Details about the design of the surveys and the data collected are given in the next section and the items of investigation have been dealt with serially in the subsequent sections. It should be understood that the present investigation is primarily methodological and the economic conclusions, based as they are on only one year's data, would necessarily be tentative.

---

* The author is deeply indebted to Dr. V. G. Panse, Statistical Adviser, Indian Council of Agricultural Research, under whose guidance this paper has been prepared. But for his inspiration and encouragement it would not have been possible to present this paper.

He is also grateful to Dr. B. V. Sukhatme and Shri Daroga Singh of the Institute of Agricultural Research Statistics, who have been giving invaluable help and useful suggestions from time to time. He is further thankful to Shri S. D. Bokil of the same Institute for help in preparing the present paper for publication.

DESIGN OF SURVEY AND DATA STUDIED

The design of the Farm Management surveys was similar in all regions and is described below :—

Two contiguous districts were selected in each region and each district was divided into two strata nearly equal in size. In each stratum eight villages were selected with probability proportional to the cultivating population in a village. Of these 8, 3 each were assigned to Cost Accounting and Survey methods while both methods were tried in the remaining two villages. The holdings in each village were listed in the order of their size and then divided into 5 groups equal or nearly equal in size. Thus first one-fifth formed the first group, the second one-fifth the second group and so on. From each such group two holdings were selected randomly for the Cost Accounting and four for the Survey method, without replacement. Thus in each village, there were 10 holdings under the Cost Accounting method and 20 under the Survey method giving in all 200 holdings for the Cost Accounting and 400 for the Survey method in each region.

In U.P. the survey covered the districts of Meerut and Muzaffarnagar. Data were collected for a number of crops. Data for the major crops, wheat, sugarcane (plant) and gram, were utilized for the present investigation.

The survey of the Indian Central Sugarcane Committee was a very extensive survey covering several districts in U.P. For the purpose of the enquiry the State was divided into 3 regions, Western, Central and Eastern U. P., and 30, 15 and 30 villages were selected from the reserved sugar factory areas in the respective regions after suitable stratification. The selection of villages was done randomly with probability proportional to the sugarcane acreage in a village from lists of villages in the reserved areas available with the Cane Commissioner. In each selected village 12 sugarcane growing holdings were usually selected with equal probability from all sugarcane growing holdings in the village. Only data for sugarcane (plant) from 11 villages selected in Meerut and Muzaffarnagar districts were utilized for the study of input-output relationship as they were comparable with the data on sugarcane available from the other survey.

COMPARISON OF THE SURVEY AND COST ACCOUNTING METHODS OF COLLECTION OF PRIMARY DATA

As stated earlier the comparison of the two methods of collecting primary data was one of the principal objectives of Farm Management Surveys undertaken by the Directorate of Economics and Statistics. In the Cost Accounting method detailed records of day to day operations on the farm are maintained by whole time fieldmen residing in or near the selected villages, on the basis of regular observation of the agricultural operations supplemented by enquiry regarding wages, prices, etc., to the extent necessary. In the Survey method the field investigator visits the sample villages periodically (usually 3 or 4 times during a year) and collects information pertaining to various agricultural operations carried out during the interval between two visits by enquiry only. If the method could yield reliable data it will be the one to be preferred on account of its cheapness.

The purpose of this section of the paper is to make a critical comparison of the two methods on the basis of data at hand.

The design of the survey provides two sources of data for the comparison of methods: (*i*) on the basis of independent villages, and (*ii*) on the basis of common villages. The comparison based on independent villages is made by comparing the means of different factors of cost for the two methods by application of Student's t-test on the basis of standard errors of the means worked out for the two sets of independent villages in a stratum. The overall estimates of the differences between means for the two methods are obtained by taking the differences stratum-wise and calculating their weighted average with weights proportional to strata sizes. The factor-wise means, their differences, corresponding standard errors and the t-values for the three crops are given in Table I. It is seen from Table I that for costs per acre, estimates obtained by the Survey method are significantly larger than those obtained by the other method for 'bullock labour' in wheat and sugarcane (plant) and 'total input' in wheat whereas for cost of 'fertilizers and manures' in sugarcane the Survey method estimate is significantly smaller than the other.

With regard to comparison based on common villages it is observed that the variation in the estimated value of any character within a village might be assigned to (*i*) methods, (*ii*) size-groups of holdings, (*iii*) interaction between methods and groups, and (*iv*) the residual variation. Following the usual approach of two-way classifications with unequal cell frequencies, the analysis of variance takes the form :

Source	Degrees of freedom
Methods	1
Groups	q-1
Interaction	q-1
Error (residual variation)	n-2q
Total	n-1

where q ($\leqslant$ 5) is the number of size-groups and n, the total number of observations in all the cells.

To test the significance of difference due to methods the interaction mean square is first tested against the error mean square. If it is not significant it is pooled with error mean square to get a better estimate of the latter and the mean square due to methods is tested against this pooled estimate. If on the other hand the interaction mean square turns out to be significant the mean square for methods needs to be recalculated on the basis appropriate for the presence of interaction and then tested against corresponding error mean square. Since the comparison of methods (*i.e.*, methods mean square) involves a single degree of

TABLE I—COMPARISON OF COST ACCOUNTING AND SURVEY METHODS OF COLLECTING PRIMARY DATA (SEPARATELY FOR INDEPENDENT AND COMMON VILLAGES)

	Independent Villages						Common Villages					
	S@	C@	S—C	S.E.	d.f.	t	S	C	S—C	S.E.	d.f.	t
Wheat												
Human Labour	43.8	41.4	2.4	3.4	16	0.7	50.6	41.0	9.6	2.0	123	4.5**
Bull ck Labour	172.5	101.8	70.7	13.2	16	5.4**	195.0	117.5	77.5	9.4	119	8.3**
Se d	11.2	11.9	—0.7	0.8	16	0.9	13.6	11.9	1.7	0.4	123	4.2**
Total Input	257.4	180.0	77.4	16.3	16	4.7**	297.3	198.0	99.3	11.1	115	8.9**
Total Output	218.0	232.5	—14.5	16.8	16	0.9	236.0	213.2	22.8	9.9	123	2.3**
Sugarcane (Plant)												
Human Labour	108.5	109.0	—0.5	11.1	16	0.04	123.9	96.5	27.4	7.2	112	3.8**
Bullock Labour	144.8	95.9	48.9	16.3	16	3.0	150.7	98.9	51.8	11.0	112	4.7**
Seed	67.8	67.4	0.4	5.7	16	0.1	70.7	75.8	—5.1	3.1	113	1.7
Fertilizers and Manures	22.8	43.4	—20 6	7.0	16	2.9**	37.8	44.6	—6.8	4.6	86	1.5
Total Input	400.3	372.8	27.5	25.6	16	1.0	435.8	357.0	78.8	21.2	112	3.7**
Total Output	593.9	624.3	—30.4	62.7	16	0.5	606.2	474.0	132.2	39.8	116	3.3**
Gram												
Human Labour	15.8	18.9	—3.1	1.9	16	1.6	20.4	20.8	—0.4	1.9	90	0.2
Bullock Labour	45.5	37.0	8.5	4.9	16	1.9	57.7	39.9	17.8	4.6	94	3.8**
Seed	5.1	5.9	—0.8	0.5	16	1.6	6.0	5.9	0.1	0.4	95	0.2
Total Input	72.3	67.3	5.0	7.2	16	0.7	90.0	74.4	15.6	6.9	95	2.2*
Total Output	86.0	91.1	—5.1	16.8	16	0.3	81.9	78.1	3.8	7.5	94	0.5

@ S and C represent average input/output in Rupees per acre estimated by the Survey and Cost Accounting methods respectively.

* Significant at 5% level.

** Significant at 1% level.

freedom, the Student's t-test is also appropriate for testing significance in this case and it takes the following form in the absence and presence of interaction :

With interaction absent[1]

$$t = \bar{d} / s\sqrt{P},$$

and with interaction present[2]

$$t = \bar{d} / s\sqrt{\frac{1}{N} + \frac{1}{M}}$$

where $\bar{d}$ = the difference of means for the two methods.

s = square root of error mean square.

$$P = \sum_{i}^{L} \left(\frac{1}{n_i} + \frac{1}{m_i} \right)$$

$$N = L^2 \Big/ \sum_{i}^{L} \frac{1}{n_i}$$

and $$M = L^2 \Big/ \sum_{i}^{L} \frac{1}{m_i}$$

n_i and m_i being the number of observations under the two methods in the i-th group and L the number of groups (5 in the present case).

The differences and standard errors are initially worked out village-wise by the procedure indicated earlier. The stratum-wise mean differences are calculated by taking simple average of village-wise differences (since villages were selected with probability proportional to size). The overall differences for the whole region are obtained by taking weighted averages of stratum-wise differences weighting by strata sizes. The overall differences are tested against their standard errors calculated from the village-wise errors by a corresponding procedure. The results so obtained are included in Table I alongwith those obtained for the first type of comparison.

As is to be expected, comparisons based on common villages have a higher precision and have consequently revealed significant differences in several factors in addition to those shown by the comparison of independent villages. Otherwise the trend of results is similar in both types of comparisons.

To obtain an overall picture of the differences due to the two methods for the various cost factors in the three crops, the results for the two types of comparison were combined by calculating weighted averages of differences, weighting the

1. M. N. Das, "Analysis of Covariance in Two-way Classification with Disproportionate Cell Frequencies," *Journal of Indian Society of Agricultural Statistics*, Vol. V, No. 2, 1953.

2. M.G. Kendall : The Advanced Theory of Statistics, Charles Griffin & Co., London, 1951.

respective method differences in proportion to the inverse of corresponding variances and these average differences were tested against their standard errors. The results so obtained are given in Table II.

TABLE II—COMPARISON OF COST ACCOUNTING AND SURVEY METHODS (COMBINED OVER INDEPENDENT AND COMMON VILLAGES)

	S	C	S—C	S.E.	d.f.	t
Wheat						
Human Labour	48.8	41.5	7.3	1.7	139	4.2†
Bullock Labour	187.5	112.2	75.3	7.6	135	9.8†
Seed	13.1	11.9	1.2	0.4	139	3.0†
Total Input	284.6	192.3	92.3	9.2	131	10.0†
Total Output	231.4	218.1	13.3	8.5	139	1.6
Sugarcane (Plant)						
Human Labour	119.4	100.1	19.3	6.0	128	3.2†
Bullock Labour	148.9	98.0	50.9	9.1	128	5.6†
Seed	70.0	73.9	−3.9	2.7	129	1.4
Fertilizers and Manures	33.3	44.3	−11.0	3.8	102	2.8†
Total Input	416.2	363.8	52.4	15.0	128	3.5†
Total Output	602.7	517.2	85.5	33.6	132	2.5†
Gram						
Human Labour	18.0	19.8	−1.8	1.3	106	1.4
Bullock Labour	51.9	38.5	13.4	3.4	110	4.0†
Seed	5.8	5.9	−0.1	0.3	111	0.3
Total Input	81.5	71.0	10.5	5.0	111	2.1*
Total Output	82.6	80.2	2.4	6.8	110	0.3

† Significant at 1% level.
* Significant at 5% level.

From this table we observe that the factors 'bullock labour' and 'total input' show significant difference in all the three crops. The magnitude of the difference expressed as percentage excess of the Survey over the Cost Accounting value for 'bullock labour' is about 67 per cent for wheat, 52 per cent for sugarcane and 35 per cent for gram. The corresponding figures of excess for 'total input' are 48 per cent, 15 per cent and 15 per cent respectively. The factor 'human labour' shows a significant excess for the Survey method amounting to 18 per cent for wheat and 20 per cent for sugarcane. Similarly there is a significant excess of about

10 per cent for 'seed' in wheat. Only the factor 'fertilizers and manures' shows a significant difference in the opposite direction in sugarcane (plant), its magnitude being of the order of 25 per cent. Thus it appears that, in general, the Survey method leads to an overestimation of inputs in comparison to the Cost Accounting method. All the same it seems hazardous to try to adjust these estimates by any statistical process because of the unpredictable nature of the magnitude and direction of the bias.

EFFICIENCY OF STRATIFICATION AT SECOND STAGE SELECTION

The system of stratification adopted at second stage selection has been described earlier. In the present section are presented the results of study to determine whether this stratification has helped in reducing the sampling variation of the overall estimate in comparison to simple random selection of holdings within villages.

The gain in precision is given by the difference in the variance of the estimate of the population mean when the stratification of holdings within a village is taken into account and when it is not. Following the notation used by Sukhatme[3] and further denoting by z_{tiuj}, the value of a character for the j-th holding of the u-th group in the (t, i)th village, the variance of the sample mean when the stratification is not taken into account is given by

$$(1) \quad V(\bar{z})_{us} = \sum_{t=1}^{k} \frac{\lambda_t^2}{n_t} \left[\sigma_b^2 + \sum_{i=1}^{Nt} P_{ti} \left(\frac{1}{m_{ti}} - \frac{1}{M_{ti}} \right) S_{ti}^2 \right]$$

the corresponding expression when stratification is taken into account being

$$(2) \quad V(\bar{z})_s = \sum_{t=1}^{k} \frac{\lambda_t^2}{n_t} \left[\sigma_b^2 + \sum_{i=1}^{Nt} P_{ti} \sum_{u=1}^{L_{ti}} \left(\frac{1}{m_{tiu}} - \frac{1}{M_{tiu}} \right) p_{tiu}^2 S_{tiu}^2 \right]$$

where M_{tiu} is the total number of holdings in the u-th group belonging to the (t, i)th village and

$$p_{tiu} = \frac{M_{tiu}}{M_{ti}}, \quad S_{tiu}^2 = \frac{1}{M_{tiu} - 1} \sum_{j=1}^{M_{tiu}} (z_{tiuj} - \bar{z}_{tiu.})^2.$$

Expressing S^2_{ti} in (1) in terms of between groups and within groups mean squares and subtracting (2) from (1), it can be shown that the gain in precision is approximately given by

$$(3) \quad V(\bar{z})_{us} - V(\bar{z})_s = \sum_{t=1}^{k} \frac{\lambda_t^2}{n_t} \sum_{i=1}^{Nt} P_{ti} \left(\frac{1}{m_{ti}} - \frac{1}{M_{ti}} \right) \sum_{u=1}^{L_{ti}} p_{tiu} (\bar{z}_{tiu.} - \bar{z}_{ti..})^2.$$

3. P. V. Sukhatme : Sampling Theory of Surveys with Applications, Indian Society of Agricultural Statistics, New Delhi, 1953.

To estimate the gain, we have to estimate $\sum_{u=1}^{L_{ti}} p_{tiu} (\bar{Z}_{tiu.} - \bar{Z}_{ti..})^2$

where L_{ti} is the number of groups in the (t, i)th village.

(4) Let $\bar{Z}_{tiu}(m_{tiu}) = \bar{Z}_{tiu.} + \epsilon_u$

where $\bar{Z}_{tiu.}$ is the population mean of the (t, i, u)th group and $\bar{Z}_{tiu}(m_{tiu})$ is its sample estimate, and

$$E(\epsilon_u) = o, \; E(\epsilon_u)^2 = \left(\frac{1}{m_{tiu}} - \frac{1}{M_{tiu}}\right) S^2_{tiu}$$

Squaring both sides of (4) and taking expectations, we have after multiplying by p_{tiu} and taking summation over the groups,

$$E\left[\sum_{u=1}^{L_{ti}} p_{tiu} \bar{Z}^2_{tiu}(m_{tiu})\right] = \sum_{u=1}^{L_{ti}} p_{tiu} \bar{Z}^2_{tiu.} + \sum_{u=1}^{L_{ti}} p_{tiu} \left(\frac{1}{m_{tiu}} - \frac{1}{M_{tiu}}\right) S^2_{tiu.}$$

(5) *i.e.*, $$\text{Est.} \sum_{u=1}^{L_{ti}} p_{tiu} \bar{Z}^2_{tiu.} = \sum_{u=1}^{L_{ti}} p_{tiu} \bar{Z}^2_{tiu}(m_{tiu}) - \sum_{u=1}^{L_{ti}} p_{tiu} \left(\frac{1}{m_{tiu}} - \frac{1}{M_{tiu}}\right) s^2_{tiu}$$

since Est. $S^2_{tiu} = s^2_{tiu}$

Proceeding similarly, we have

(6) $$\text{Est.} \bar{Z}^2_{ti..} = \bar{Z}^2_{ti}(m_{ti}) - \sum_{u=1}^{L_{ti}} p^2_{tiu} \left(\frac{1}{m_{tiu}} - \frac{1}{M_{tiu.}}\right) s^2_{tiu.}$$

Subtracting (6) from (5)

(7) $$\text{Est.} \sum_{u=1}^{L_{ti}} p_{tiu} (\bar{Z}_{tiu.} - \bar{Z}_{ti..})^2 = \sum_{u=1}^{L_{ti}} p_{tiu} \left\{ \bar{Z}_{tiu}(m_{tiu}) - \bar{Z}_{ti}(m_{ti}) \right\}^2 - \sum_{u=1}^{L_{ti}} p_{tiu} (1 - p_{tiu}) \left(\frac{1}{m_{tiu}} - \frac{1}{M_{tiu}}\right) s^2_{tiu}.$$

Whence, finally from (3) and (7), the estimate of the gain in precision is given by

$$\sum_{t=1}^{k} \frac{\lambda^2_t}{n_t} \sum_{i}^{n_t} \frac{1}{m_{ti}} \left[\sum_{u=1}^{L_{ti}} p_{tiu} \left\{ \bar{Z}_{tiu}(m_{tiu}) - \bar{Z}_{ti}(m_{ti}) \right\}^2 - \sum_{u=1}^{L_{ti}} p_{tiu} (1 - p_{tiu}) \frac{s^2_{tiu}}{m_{tiu}} \right]$$

assuming M_{tiu} and M_{ti} large.

Percentage gain in efficiency by stratification at second stage calculated for various factors for the three crops by the above formula is shown in Table III.

TABLE III—PERCENTAGE GAIN IN EFFICIENCY BY SECOND STAGE STRATIFICATION

Crop	Wheat	Sugarcane (Plant)	Gram
Factors			
Human Labour	1.4	2.1	0.0
Bullock Labour	6.7	7.1	0.0
Total Input	5.4	1.2	0.0
Total Output	0.8	2.1	27.9

Results for gram are a little curious. The possible explanation for this result is that the output data were not quite reliable. The results for the other two crops, however, indicate that there is practically no gain due to stratification of holdings within a village. Further with this type of stratification it is not possible to estimate any averages for specific size-classes of holdings as the strata limits vary from village to village. Stratification of holdings within villages according to uniform size-classes such as < 2.5 acres, 2.5—5.0 acres, etc., would, on the other hand, make estimation of averages for such classes possible. Whether such stratification will lead to any gain in precision needs to be explored from future surveys appropriately designed for the purpose.

ESTIMATION OF SAMPLE SIZE FOR GIVEN PRECISION

In view of the negligible gain in precision due to second stage stratification, the sample size required for given precision might be determined on the basis of equation (1) (given in the last section) ignoring second stage stratification. Denoting by B_t the estimate of σ^2_{tb} and by W_t, the estimate of the mean square between holdings within villages, the equation takes the form

$$(8) \quad V(\bar{z}) = \sum_{u=1}^{k} \frac{\lambda^2_t}{n_t} \left(B_t + \frac{W_t}{m_t} \right)$$

where m_{ti} is given the constant value m_t in the selected villages within the t-th stratum.

As $\frac{1}{n_t} \left(B_t + \frac{W_t}{m_t} \right)$ is the estimate of the variance of the mean for the t-th stratum, values of n_t can be calculated corresponding to convenient values of m_t, assumed constant from stratum to stratum, so that the precision of this estimate could be of a desired degree. The overall sample size n is then obtained by adding the n_t's

If B_t and W_t are further assumed to be uniform from stratum to stratum, say, B and W, and a constant number of holdings m, are selected in each village it can be shown that the required sample size, n, can be determined from the equation

$$(9) \quad Vo = \frac{B}{n} + \frac{W}{mn}$$

where Vo is the desired sampling error. Values of B and W were calculated for a number of input factors and output for the three crops and values of n were calculated corresponding to values of m ranging from 1 to 12 for 5 per cent precision. These are shown in Table IV.

Optimum allocation of the sample between two stages of sampling could be considered as usual on the principle of minimum cost for given precision. Thus if C_1 is the cost of field staff and other costs incidental to each selected village and C_2 the cost of canvassing each selected holding, the total cost of the survey within the t-th stratum will be given by

$$C_1 n_t + C_2 n_t m_t$$

and for least cost

$$\frac{1}{n_t}\left(B_t + \frac{W_t}{m_t}\right) + \mu \left(C_1 n_t + C_2 n_t m_t \right)$$

should be a minimum. This gives us n_t and m_t for each stratum in terms of B_t, W_t and ratio $C = C_1/C_2$. The optimum sample number of villages is then obtained by adding n_t's. As regards the number of holdings to be selected from each village, for convenience, the average of m_t's, say m, may be taken as the uniform sample size of holdings. However, if B_t and W_t can be assumed uniform from stratum to stratum say, B and W, we can start with equation (9) and assuming the cost function to be of the form $(C_1 n + C_2 mn)$, obtain optimum values of n and m by minimising the expression

$$\left(\frac{B}{n} + \frac{W}{mn}\right) + \mu \left(C_1 n + C_2\, mn \right)$$

This gives us

$$n = \frac{1}{Vo}\left(B + \sqrt{\frac{WB}{C}}\right)$$

$$\text{and } m = \sqrt{\frac{WC}{B}}$$

TABLE IV—NUMBERS OF SAMPLE VILLAGES REQUIRED FOR 5 PER CENT PRECISION CORRESPONDING TO VARIOUS VALUES OF NUMBER OF HOLDINGS SELECTED PER VILLAGE (m)

No. of holdings (m)	Wheat				Sugarcane (Plant)						Gram			
	Human labour	Bullock labour	Total input	Total output	Human labour	Bullock labour	Seed	Manures etc.	Total input	Total output	Human labour	Bullock labour	Total input	Total output
1	67	94	87	38	97	179	40	264	55	93	155	151	77	156
2	44	54	51	26	63	97	28	151	36	66	89	80	42	100
3	36	40	39	22	51	70	24	113	30	57	62	56	30	82
4	33	33	33	20	46	57	22	94	27	52	51	45	24	72
5	30	29	30	19	42	49	21	83	25	49	44	37	20	67
6	29	27	27	18	40	43	20	75	24	47	39	33	18	63
7	28	25	26	18	38	39	19	70	23	46	36	29	16	60
8	27	23	24	17	37	36	19	66	22	45	33	27	15	58
9	26	22	23	17	36	34	18	63	22	44	31	25	14	57
10	26	21	23	17	35	32	18	60	21	44	30	23	13	56
11	25	21	22	17	35	31	18	58	21	43	28	22	13	55
12	25	20	21	17	34	30	18	57	21	43	27	21	12	54

Optimum values of n and m for 5 per cent precision were worked out on the basis of above formulae for a likely range of values of the ratio C. These are given in Table V. The actual problem of Cost Accounting is, however, more complex

TABLE V—OPTIMUM VALUES OF m AND n FOR CERTAIN VALUES OF COST RATIO C ($= C_1/C_2$) FOR 5 PER CENT PRECISION

Wheat

C	Human labour m	Human labour n	Bullock labour m	Bullock labour n	Total input m	Total input n	Total output m	Total output n
2.0	2	43	4	36	3	39	2	27
2.5	3	41	4	34	4	36	2	26
3.0	3	39	5	32	4	34	3	25
3.5	3	39	5	31	4	33	3	24
4.0	3	37	5	29	5	32	3	24

Sugarcane (Plant)

C	Human labour m	Human labour n	Bullock labour m	Bullock labour n	Seed m	Seed n	Fertilizer and manures m	Fertilizer and manures n	Total input m	Total input n	Total output m	Total output n
2.0	3	60	5	52	2	30	4	103	3	36	2	71
2.5	3	56	6	48	2	28	4	94	3	34	2	67
3.0	3	54	6	45	3	27	5	91	3	32	2	65
3.5	3	52	6	43	3	27	5	87	3	31	3	63
4.0	4	51	7	41	3	25	5	84	3	30	3	61

Gram

C	Human labour m	Human labour n	Bullock labour m	Bullock labour n	Total input m	Total input n	Total output m	Total output n
2.0	5	49	6	34	5	21	3	94
2.5	5	45	7	31	6	19	3	89
3.0	6	43	7	29	6	18	3	85
3.5	6	41	8	28	7	17	3	82
4.0	6	39	9	26	7	16	4	79

because of the necessity of stationing a whole-time field investigator in each village and giving him a more or less uniform work-load. The foregoing analysis is more appropriate to the problem of design with the Survey method of investigation. For the Cost Accounting method, results given in Table IV provide the required guidance. Taking 6-8 holdings as the manageable work-load for a whole-time resident investigator, it is observed from that table that the character 'total input' per acre which is basic to the calculation of cost of cultivation of a crop can be estimated with precision of the order of 5 per cent sampling error on the basis of a sample of 20-30 villages for various crops where the investigation is confined to a compact region such as in the present case. These findings are of value in showing that estimates with useful precision can be obtained by the Cost Accounting method with a sample of not an unreasonably large size.

INPUT-OUTPUT RELATIONSHIPS

The investigation also included the study of input-output relationship on the basis of what appeared to be the most appropriate production function. Of the various production functions which might be fitted to input-output data, such as the Spillman's function,[4] quadratic function, etc., the Cobb-Douglas type of production function is probably the one most commonly fitted. With this type of function effect of four or five variable input factors on the output can be studied simultaneously without any appreciable difficulty as the function reduces to a simple linear form on the logarithmic scale. Further, in the present case, a comparison of the corresponding scatter diagrams between output per acre and various factors of input per acre drawn on the original scale as well as the logarithmic scale showed that the variability reduced to some extent after transformation. This type of function was therefore chosen for fitting to the data.

The method of fitting was that of the least squares, *i.e.*, the various constants were obtained by minimising the residual sum of squares on the transformed scale. Thus a function such as

$$y = a\ x_1^{b_1}\ x_2^{b_2}\ \ldots\ldots$$

takes the form

$$\log y = \log a + b_1 \log x_1 + b_2 \log x_2 + \ldots\ldots$$

on the transformed scale ; or indicating the variates on this scale by capital letters it might be expressed as

$$Y = A + b_1 X_1 + b_2 X_2 + \ldots\ldots$$

The coefficients A, b_1, b_2, etc., were obtained by minimising

$$\Sigma\ [Y - (A + b_1 X_1 + b_2 X_2 + \ldots\ldots)]^2$$

the summation being taken over sample holdings constituting the basic data. This involves solving normal equations of the type

$$\Sigma X_1 Y = b_1 \Sigma X_1^2 + b_2 \Sigma X_1 X_2 + b_3 \Sigma X_1 X_3 + \ldots\ldots$$

$$\Sigma X_2 Y = b_1 \Sigma X_1 X_2 + b_2 \Sigma X_2^2 + b_3 \Sigma X_2 X_3 + \ldots\ldots \text{etc.}$$

Taking X's as normally distributed, the partial regression coefficients, b's are tested with appropriate standard errors by the usual t-test. Table VI gives the coefficients of various factors entering the production function fitted to the different

4. This function is of the type $y = \prod_{i=1}^{n} (m_i - a_i r_i^{x_i})$ where y is the output and x_i is the i-th factor of input ; m_i and a_i are respectively the maximum attainable and the maximum additional outputs due to indefinite increments in the input factor x_i and r_i is the ratio of the increment in the output for a unit increment in the input x_i to the preceding increment in the output due to the same increment in the input. Thus in this function, y never decreases so that negative returns are ruled out since, for diminishing returns, r_i is less than unity. However it becomes very cumbersome to fit this function to more than one variable input and therefore effect of two or more variable inputs on the output cannot be studied simultaneously. *Vide* E. O. Heady : Economics of Agricultural Production and Resource Use, Prentice Hall, Inc., New York, 1952, pp. 58-59.

crops and also indicates their significance. In case of sugarcane (plant) the function was also fitted to data obtained from the Sugarcane Committee as explained earlier and those results are also given in the table. The coefficient of determination, R^2, which indicates the fraction of variability accounted for by the relationship is also given in the table.

TABLE VI—COEFFICIENTS OF COBB-DOUGLAS FUNCTIONS FITTED TO WHEAT, SUGARCANE AND GRAM

Factor	A	X_h	X_b	X_s	X_f	X_i	X_l	R^2	E_p
Wheat‡ (Irrigated) ..	1.856	0.056	0.041	0.214	—	0.063	—	0.063	0.373
Sugarcane‡ (Plant) ..	2.212	−0.027	0.024	−0.037	0.153	0.209	—	0.032	0.322
Gram‡ ..	1.016	0.165*	0.210	0.407*	—	—	0.039	0.129*	0.821
Sugarcane† (Plant) ..	0.871	0.317**	−0.079	0.623**	0.055	0.022	—	0.287**	0.938

Where

Y	stands	for	Output per acre	(money value in rupees)
X_h	,,	,,	Human labour/acre	(-do-)
X_b	,,	,,	Bullock Labour/acre	(-do-)
X_s	,,	,,	Seed/acre	(-do-)
X_f	,,	,,	Fertilizer and Manures/acre	(-do-)
X_i	,,	,,	Irrigation charges/acre	(-do-)
X_l	,,	,,	Land Revenue/acre	(-do-)
A	,,	,,	Output when all the factors in the production function are put equal to unity.	

‡Collected under the Farm Management Enquiry by the Directorate of Economics and Statistics.
†Collected under the Cost of Cultivation Scheme by the Indian Central Sugarcane Committee.
*Indicates significance at 5%.
**Indicates significance at 1%.

Elasticity of production is defined as the ratio of percentage increase in the output to percentage increase in the input with reference to any factor. In case of Cobb-Douglas type of function it equals the coefficient 'b' corresponding to the factor (Heady)[5] and the overall elasticity of production is given by $\sum_{i=1}^{n} b_i$, n being the number of factors entering the function. The production function represents increasing, constant or decreasing marginal returns according as the elasticity is greater than, equal to or less than unity. The overall elasticity of production is shown in the last column of Table VI (denoted by E_p).

5. E. O. Heady: *Op. cit.*

It is observed from Table VI that for the three functions, *viz.*, for wheat, sugarcane and gram, based on Farm Management Enquiry data, the individual elasticities as well as overall elasticities of production are each appreciably less than unity indicating decreasing marginal returns to all input factors individually as well as jointly for the three crops. It is also noted that only the elasticities (coefficients) for 'seed' in wheat and 'human labour', 'bullock labour' and 'seed' in gram are significant, the rest being non-significant. The coefficient of determination, R^2, is highly significant for gram although it accounts for only 13 per cent of the variation in output.

From the function fitted to data on sugarcane obtained from the Indian Central Sugarcane Committee, it is noted that the individual elasticities as well as the overall elasticity are appreciably larger than corresponding values for the function fitted to data from the former source. The elasticities for the factors 'human labour' and 'seed' are highly significant and the overall elasticity approaches unity. The coefficient of determination, R^2, is significant at 5 per cent and accounts for about 29 per cent of the variation in output. These data thus reveal a stronger relationship and the function obtained from these would be more useful for prediction purposes. The difference in the behaviour of the two functions fitted to the same crop in the same region could possibly be due to the fact that, as mentioned in the report of the enquiry, the data which were obtained in the first year of the enquiry might have suffered in quality for various reasons.

An important conclusion that emerges from the above results is that the input factors included in the present study actually contribute to production only to a limited extent and that it might be worthwhile in any future study to include other factors such as, for example, observations on plant nutrients, soil fertility, etc., in selected holdings.

Summary

Some of the major problems in applying the statistical methodology to farm cost surveys are discussed in this paper. The statistical methodology relevant to the following four problems has been developed and illustrated with the help of data collected from various sources.

(i) *The comparison between the Cost Accounting and Survey methods of collecting the data :* The comparison was made on the basis of both independent villages and common villages. Of the two types of comparisons, the latter being a more precise comparison has yielded more significant results. The methods were found to differ in almost all the input factors. However, the difference was particularly marked in the case of bullock labour/acre and total input/acre, the Survey method giving estimates appreciably in excess of those obtained by the other method.

(ii) *The efficiency of the design adopted in the survey :* It was found that, on an average, there was no gain in precision due to stratification at the second stage. The efficiency of stratification of holdings according to fixed size-classes needs to be explored.

(iii) *The sample size required for a given precision :* It was observed that the factor total input/acre could be estimated with 5 per cent standard error with a sample of 23, 21 and 13 villages for the three crops wheat, sugarcane, and gram assuming that 10 holdings were selected in each village. The corresponding sample size for total output/acre was 17, 44 and 56 villages respectively. Sample size for other input factors was also calculated.

(iv) *Input-Output relationship :* In an effort to find out the relationship existing between the factors of input and output per acre, Cobb-Douglas function was fitted to all the crops. No close relationship could be established between the input factors and output.