Vol XLIII
No. 4

ISSN 0019-5014

OCTOBER-
DECEMBER
1988

INDIAN JOURNAL OF AGRICULTURAL ECONOMICS

INDIAN SOCIETY OF
AGRICULTURAL ECONOMICS,
BOMBAY

Ind. Jn. of Agri. Econ.
Vol. 43, No. 4, Oct.-Dec. 1988

RESEARCH NOTES

EXPLAINING VARIATIONS IN COTTON YIELDS : A FACTOR ANALYSIS APPROACH

Production relationships in the cultivation of cotton are more complex than portrayed in simple production function or farm size analysis. Information collected on many other aspects of cotton cultivation as part of comprehensive farm level surveys remains unutilised in applied work due to large non-response or doubt on its quality. Consequently, empirical assessment of the influence of labour-intensive cultural practices, such as planking, ploughing, levelling and weeding operations, with yields has remained scanty in comparison to studies based on a few key inputs.[1] The purpose of this paper is to rectify, to the extent possible, this neglect by analysing within a more integrated framework, the relationship of numerous types of inputs with cotton yields.

The survey of cotton growers conducted during the three-year period 1983-85 in the Tharparkar district of Sind province provides the data base for the present analysis. The data are factor analysed to *(a)* examine whether the available information on cross-section of cotton farms has some meaningful interpretive dimensions, *i.e.,* can a few identifiable factors account for a greater part of variation in the entire data and specifically yields ?, *(b)* sift out a sub-set of variables that distinguish themselves more than others in the explanation of differences in cotton yields; and *(c)* within the sub-set identify some policy relevant instruments for the benefit of agriculture and extension agencies. After a brief description of the data and their limitations, we directly move on to the discussion of results and a simple sensitivity exercise. The findings are summarised in the last section of the paper.

DATA AND VARIABLES

Since 1980-81, the Applied Economics Research Centre of University of Karachi has been conducting a yearly survey of cotton growers in the Tharparkar district as part of a study on the Cotton Maximisation Project sponsored by the Pakistan Central Cotton Committee. Each season three rounds of field investigation at different stages, *i.e.,* post-sowing, flowering and post-harvest, are undertaken to collect data on many aspects of cotton cultivation. A two-stage stratified random sampling procedure is used to draw out a new sample every year from a list of farmers in the project area. For comparison purposes, the same questionnaire is also administered to farms situated contiguously but outside the project area.[2] In our sample of 301 farmers, 36 per cent are from the non-project area. Of the total sample, 36 per cent are small farmers (0-12.5 acres), 24 per cent own medium size holdings (12.5-25.0 acres) and 40 per cent have large farms (above 25 acres). The

entire sample area is irrigated and located at the tail-end of the canal system in Sind province.

At the time of surveys, only a single variety of cotton seed popularly known as NT was sown by the majority of the farmers. Given the specific purpose of this study, the information extracted from the surveys for use in the analysis can broadly be categorised as : *(i)* cultural practices including timing of sowing and sprays, *(ii)* input use—chemical, mechanical, human and animal as also the farmer's estimate of pest and rain damage as proxy for negative inputs, *(iii)* an indicator of extension services, *(iv)* tenancy and land use characteristics and *(v)* the farmer's education. *A priori*, the hypothesis is that the above variety of inputs interact directly or indirectly in influencing the final yields at the farm level.[3]

Initially, 57 variables were employed to search for a meaningful factor solution.[4] However, the determinant of correlation matrix was nearly singular.[5] A second look at the data and their correlation matrix revealed that the correlation of several variables with yield was near to zero, and the variability in a few others was negligible. Subsequently, these were excluded to obtain a solution which would not only facilitate interpretation but account for variation in the data set with a minimum number of factors. A description of both the 40 variables included in the final solution and those dropped is given in the Appendix. The discussion of results in the next section is based on 221 farms sampled during 1983 and 1984 in the five sub-districts (*i.e.*, talukas) of Tharparkar district. In 1985, the project area was shifted to another location within the same district, while retaining one sub-district from the previous project area.[6] This provided an opportunity to test the robustness of results on the same set of information on a larger area and a sample of 301 farms over a period of three years.

RESULTS AND INTERPRETATION

The final factor solution yielded 16 factors. The matrix of factor loadings, communalities and eigen values shown in Table I relate to four factors and 14 variables picked out from the final solution. The selection of factors and variables presented in the above sub-matrix is based on the following considerations : *(i)* these four factors account for 45 per cent of the aggregate variation in the data, *(ii)* they also account for 77 per cent ($h^2 = 0.54$) of the total variation ($h_T^2 = 0.70$) in cotton yields explained by 16 factors, and *(iii)* given the low correlation among variables, only those variables (except cotton yields) are assigned to common factors, whose 'factor loading' or a correlation with a factor is equal to or greater than 0.35. The left-out variables either correlate poorly with these four factors or strongly with the remaining unimportant 12 factors. Moreover, an interesting feature of the partial factor solution in Table I is that no single factor accounts for an overwhelming majority of the explained variation in the data or cotton yields. Secondly, ten out of 14 variables are 'bunched' together on two factors. Once the variables have been assigned to various factors, the

TABLE I. MATRIX OF FACTOR LOADINGS, COMMUNALITIES AND EIGEN VALUES

Variables	F1	F2	F3	F5	h^2	h^2_T
YIELD	0.197	0.350	− 0.450	0.422	0.542	0.704
PTUREAB 2	0.816	− 0.067	0.016	− 0.190	0.707	0.762
PTACHWED	0.774	− 0.064	0.193	− 0.208	0.684	0.768
PTACPWED	0.858	− 0.019	0.139	− 0.153	0.779	0.818
SPRAY	0.467	0.205	0.013	0.335	0.373	0.554
ACRR	− 0.011	− 0.446	0.071	0.014	0.207	0.452
PTSHAC	− 0.074	0.638	0.384	0.053	0.563	0.738
OWNPT	− 0.081	− 0.352	− 0.160	− 0.160	0.182	0.506
EDUCA	− 0.095	0.485	0.357	− 0.086	0.379	0.495
TDTPLOW	0.134	0.357	0.119	0.044	0.162	0.401
PSTDAM	0.088	− 0.451	0.428	0.181	0.427	0.668
TCOTAC	− 0.199	0.272	0.379	0.024	0.257	0.390
NIFCLD	− 0.068	− 0.396	0.489	0.159	0.430	0.478
RAINDM	− 0.012	0.163	− 0.188	− 0.371	0.199	0.441
AUG 21	0.252	0.067	0.076	0.368	0.209	0.996
Eigen values	2.851	2.548	1.716	1.373		
Variation explained (per cent)	15.3	13.6	9.2	7.4		

next step is to give these factors some meaningful interpretation in terms of underlying forces or dimensions they represent towards explaining the variation in cotton yields.

Factor 1

Per acre use of urea at flowering stage (PTUREAB 2), the proportion of area hand and plough weeded (PTACHWED, PTACPWED) and the number of sprays (SPRAY) load positively and strongly on this factor. Given that all these inputs are important at the maturity phase of the cotton crop, Factor 1 expresses a strong interaction between cotton yields and its nurturing at the maturity stage.

Factor 2

Three out of six variables, *i.e.*, ownership pattern (OWNPT), the proportion of cotton area share-cropped (PTSHAC) and the farmer's education (EDUCA), associated with this factor are indicators of tenancy and the farmer's characteristics. The member of tractor ploughing on dryland (TDTPLOW)[7] interacting positively with yields suggests a useful role of technology in response to physical and soil characteristics of land under preparation. The opposite correlation signs of share-cropped area (PTSHAC) and ownership pattern (OWNPT) on Factor 2 reinforce each other.[8] Plausibly, assured labour supply at important stages of crop production in tenant-operated farms helps to improve cotton yields. Factor 2 is thus more indicative of the role of land allocation, tenancy, farm and the farmer's characteristics in influencing cotton yields.

Remaining Factors

The loading of two variables on each of the remaining two factors, *i.e.*, F3 and F5 renders it difficult to give each a meaningful interpretation. However, some variables, if considered in isolation, present interesting insights into the inputs-yield relationship. *A priori*, a positive association between the number of farmer's visits to the extension staff (NTFCLD) and yield is expected.

However, the negative association of NTFCLD with cotton yields appearing in Factor 3 suggests a different explanation. It seems that those farmers who expect poor yields during the current harvest, either due to the past year's experience or the constraints developed in the year, pay a greater number of visits to the extension staff during the year than the other farmers. The negative correlation of the proportion of acreage in cotton to total acreage (ACRR) and total cotton acreage (TCOTAC) with yields in F2 and F3 respectively indicates that cotton yields are inversely related to cotton acreage both in absolute and relative terms. These inverse relationships do not unambiguously favour the efficiency argument of small farms. Implicit in these relationships is probably the fact that the expansion in acreage under a crop often involves bringing marginal lands under cultivation.

SENSITIVITY ANALYSIS

As mentioned earlier, the shifting of the Cotton Maximisation Project in 1985 to another location within the same district provides one more variant of the solution on the same set of information, but on a larger sample. Moreover, the addition of another year to the same sample further smooths out any effects of weather fluctuation on cotton yields. The final factor solution based on the original 40 variables contained 16 factors. Table II presents a sub-matrix of eight factors and 18 variables selected on the same criteria as the previous factor solution.

Though the explained variation in cotton yields ($h^2 = 0.558$) is 90 per cent of the total variation ($h^2 = 0.622$), the proportion of overall variation accounted for by these eight factors is 73 per cent. Comparing with results in Table I, we observe that *(i)* ten of the 15 inputs identified in the earlier solution retain their importance in the enlarged sample. *(ii)* The structure and contribution of the first two factors are fairly similar to the corresponding factors in Table I, *i.e.*, the underlying forces represented by factor 1 and 2 are invariant to change in location (within the district) and time. *(iii)* Except ownership pattern (OWNPT), the correlation or loading of other inputs, *i.e.*, the farmer's estimate of rain and pest damage (RAINDM, PSTDAM) and the number of farmer's visits to the extension staff (NTFCLD), on various factors reinforce the results obtained in the first solution. Positive, in contrast to a negative, loading of OWNPT in Table I supports a tentative conclusion that interaction of tenurial arrangements on yields varies with location. *(iv)* The favourable influence of mechanical technology in cultural practices with better yields is supported by the positive correlation of tractor use (TWTPLOW) as opposed to the negative relationship of bullock use (TWBPLOW) in the preparation of wet land.[9] Of course, causality can flow in both directions. Higher income due to improved yields can also encourage the use of tractors or vice versa. *(v)* The first factor solution includes only a single indicator of spray timing, *i.e.*, AUG 21, while in Table II their number rises to six with only AUG 21 common in both the solutions.[10] Thus it is difficult to identify a robust pattern of spray timing that may be ideal for cotton crop across years and locations even within the confines of a single district. The above pattern is also inconsistent with the findings based on a sub-set of this data (Akhtar, 1985).[11] From the results in Table II a combination of first spray during the third or fourth week of July, followed by a second spray in the first week of August, appears to correlate positively with cotton yields. Formulating a feasible combination of spray timing every year and for every location may remain a serious challenge to administrative and technical capabilities of extension agencies.

SUMMARY AND CONCLUSION

Information contained in farm level surveys conducted during 1983-85, and covering 301 cotton growers in Tharparkar district of Sind province formed the data base for this analysis. Forty variables representing various facets of farm level production practices, input use and tenancy

TABLE II. MATRIX OF FACTOR LOADINGS : ENLARGED SAMPLE

Variables	F1	F2	F3	F4	F5	F6	F7	F8	h^2	h^2_T
YIELD	0.169	0.319	− 0.338	0.116	0.421	0.112	0.167	0.287	0.558	0.622
PTUREAB 2	0.811	− 0.023	− 0.041	− 0.150	− 0.068	− 0.057	− 0.087	− 0.089	0.706	0.725
PTACHWED	0.718	0.023	0.361	− 0.142	− 0.054	− 0.179	0.245	− 0.020	0.762	0.784
PTACPWED	0.809	0.053	0.260	− 0.167	− 0.009	− 0.113	0.116	− 0.029	0.780	0.788
ACRR	0.001	− 0.378	0.106	− 0.147	0.007	0.060	0.096	0.163	0.215	0.412
PTSHAC	− 0.177	0.614	0.289	0.145	0.029	− 0.225	0.016	− 0.199	0.604	0.681
EDUCA	− 0.182	0.457	0.260	− 0.023	− 0.080	− 0.139	− 0.043	0.062	0.341	0.448
PSTDAM	0.049	− 0.395	0.430	− 0.021	0.193	0.022	− 0.246	− 0.209	0.486	0.606
NTFCLD	0.109	− 0.317	0.454	0.078	0.066	0.037	− 0.064	− 0.007	0.334	0.370
JULY 12	0.085	− 0.058	0.054	0.353	− 0.027	0.218	0.162	− 0.090	0.221	0.418
AUG 21	0.246	0.075	0.183	0.777	− 0.080	0.405	− 0.033	− 0.084	0.882	0.992
RAINDM	0.114	0.145	− 0.387	− 0.033	− 0.474	0.038	− 0.106	− 0.108	0.434	0.535
JULY 13	0.146	0.130	− 0.094	− 0.175	0.395	0.099	− 0.014	− 0.251	0.307	0.378
TWBPLOW	− 0.020	− 0.312	− 0.038	0.384	0.126	− 0.455	0.190	− 0.128	0.522	0.586
TWTPLOW	0.021	0.260	0.223	− 0.346	− 0.170	0.405	− 0.204	− 0.030	0.473	0.523
JULY 11	0.365	0.088	− 0.348	0.180	− 0.081	− 0.208	− 0.469	0.221	0.613	0.709
JULY 14	− 0.036	0.009	− 0.299	− 0.092	− 0.194	− 0.062	0.527	− 0.019	0.419	0.677
OWNPT	− 0.140	− 0.328	0.154	− 0.103	− 0.080	0.120	0.149	0.425	0.385	0.577
JULY 23	0.085	− 0.191	− 0.048	− 0.019	0.143	− 0.135	− 0.200	0.364	0.258	0.396
Eigen values	2.64	2.15	1.78	1.41	1.18	1.13	1.05	0.98		
Variation explained (per cent)	15.30	12.50	10.30	8.20	6.80	6.50	6.10	5.70		

arrangements were factor analysed to identify the underlying forces related to improved cotton yields. The main findings of this paper are:

(a) Cultural practices pursued during the maturity or flowering stage of cotton crop, *i.e.*, plough and hand weeding practices, and use of urea notably influenced cotton yields as compared to various other practices. Similarly, the use of tractors instead of bullock in seed-bed preparation after initial watering of land is positively related to yields.

(b) Farm characteristics including the farmer's education is an important dimension affecting the difference between below and above average yields.

(c) A positive link between the farmer's visits to extension agency for advice and his end-season cotton output is not as apparent as commonly hypothesised.

(d) Attempts to identify a common and robust pattern of spray timings applicable for the entire area and over time were unsuccessful. A tentative finding is that a combination of first spray during the third or fourth week of July, followed by second spray in the first week of August, is suitable for cotton growing areas of Tharparkar.

It is worth emphasising that *(i)* the results of this analysis are specific to agro-climatic conditions of a semi-desert region, *i.e.*, Tharparkar, and need to be verified across a number of years to serve as useful policy prescriptions for the benefit of cotton growers situated in the district. *(ii)* The technique employed and tentative interpretations of various dimensions should be regarded as a first step in formulating a more integrated testable hypothesis for further research into the determinants of cotton yields.

Sajjad Akhtar and Mohammed Nishat*

APPENDIX

VARIABLES INCLUDED

YIELD	:	Average yield in maunds per acre.
PTUREAB 2	:	Total urea bags used per acre at flowering stage.
PTACHWED	:	Percentage of cotton acreage hand weeded.
PTACWED	:	Percentage of cotton acreage plough weeded.
SPRAY	:	Total number of sprays.
ACRR	:	Ratio of cotton acreage to total acreage.
PTSHAC	:	Percentage of cotton under share-cropping

* Research Economists, Applied Economics Research Centre, University of Karachi, Karachi, Pakistan.

The authors are grateful to Dr. S.A. Sarwar Rizvi for releasing the data. Comments by the referees of this Journal led to improvement in the paper. Noor-ul-Hasan provided programming assistance. N.A. Khokar at the World Bank Resident Mission in Islamabad gave secretarial support. The errors are the authors' responsibility.

EDUC	:	Level of education : 0 = no education, 1 = maderssah, 2 = primary, 3 = middle, 4 = matric, 5 = intermediate, 6 = graduate, 7 = diploma.
PSTDAM	:	Percentage of damage due to pest attack.
NTFCLD	:	Number of times the farmers called on the CMP (Cotton Maximisation Project) extension staff during one crop season.
JULY 12	:	First spray done in the second week of July.
AUG 21	:	Second spray done in the first week of August.
RAINDAM	:	Percentage of damage due to rain.
JULY 13	:	First spray done in the third week of July.
TWBPLOW	:	Number of wet bullock ploughing.
TWTPLOW	:	Number of wet tractor ploughing.
JULY 11	:	First spray done in the first week of July.
JULY 14	:	First spray done in the fourth week of July.
OWNPT	:	Ownership pattern : 1 = tenant-operated, 2 = owner-operated, 3 = owner-cum-tenant-operated.
JULY 23	:	Second spray done in the third week of July.
TCOTAC	:	Total cotton acreage.
JULY 24	:	Second spray done in the fourth week of July.
PTDAPBS	:	Number of DAP bags used per acre at sowing time.
TWBPLOW	:	Number of dry bullock ploughing.
TWTPLOW	:	Number of dry tractor ploughing.
STYP	:	Soil type : 0 = light, 1 = medium, 2 = heavy, 3 = combination.
TDTPLOW	:	Number of dry tractor ploughing
TDBPLOW	:	Number of dry bullock ploughing.
TSOKDOS	:	Number of soaking dose.
PSOWMAY	:	Percentage of cotton acreage sown in May.
PTNPB 1	:	Number of NP (nitrogen-phosphate) bags used per acre at Banjoo (first irrigation).
AUG 11	:	First spray done in the first week of August.
AUG 12	:	First spray done in the second week of August.
JULY 23	:	Second spray done in the third week of July.
JULY 24	:	Second spray done in the fourth week of July.
AUG 22	:	Second spray done in the second week of August.
AUG 23	:	Second spray done in the third week of August.
AUG 34	:	Third spray done in the fourth week of August.
SEP 41	:	Fourth spray done in the first week of September.
SEP 42	:	Fourth spray done in the second week of September.

VARIABLES EXCLUDED

TDBPLOW	:	Number of dry bullock ploughing.
TASPB 1	:	Number of ASP bags used per acre at Banjoo (first irrigation).
TNOTHWED	:	Number of times hand weeded.
TNOTPWED	:	Number of times plough weeded.
AUG 32	:	Third spray done in the second week of August.
AUG 33	:	Third spray done in the third week of August.
TNPBS	:	Number of NP (nitrogen-phosphate) bags used per acre at sowing.
TSSPB 2	:	Number of SSP bags used per acre at flowering stage.
TWBLEVL	:	Number of wet bullock levelling.
TWTLEVL	:	Number of wet tractor levelling.
PTLINSOW	:	Percentage of cotton acreage under line sowing.
TSEDR	:	Seed rate (kg./acre)
PSOWAPL	:	Percentage of cotton sown in April.
TGERM	:	Germination rate (percentage).
PTUREAB 1	:	Number of urea bags used per acre at Banjoo (first irrigation)
AUG 23	:	Second spray done in the third week of August.
AUG 24	:	Second spray done in the fourth week of August.

NOTES

1. According to the cost of production estimates prepared by the Agricultural Price Commission (1986), cultural practices in cotton constitute 21 per cent of the total cost per acre.

2. In this sampling procedure the probability of the same farms being interviewed consecutively for two or three years is not zero.

3. Pricing information on inputs and output remained largely unutilised.

4. Eighty-one variables or indicators were listed for the purpose of inclusion in the factor analysis. After detailed examination, many were found with insufficient values or inconsistent with normal cultivation practices. For example, except at the flowering stage, urea is not commonly applied. Thus questions related to its application at other stages of cultivation contained zeros. Similarly, information collected on the use of seven or eight types of fertiliser proved redundant as only three or four types were commonly used.

5. Its value was 0.9728D−10. In the Principal Factoring with Iteration Technique, if the determinant of correlation matrix is 10−8, the absolute value of the largest element in each column is replaced in the diagonal of the matrix, instead of the square of multiple correlation.

6. The share of farms from a retained sub-district, *i.e.*, Mirpurkhas taluka in the total sample fell to 20 per cent in 1985 against 35 per cent in the sample of the previous two years.

7. Preparation of dryland before sowing.

8. OWNPT is a category variable with 1 = tenant-operated, 2 = owner-operated and 3 = tenant-cum-owner-operated farm.

9. It is to be noted that these two inputs closely missed our inclusion criteria in the sub-matrix in Table I. On factor 7, which explained 0.03 ($h^2 = 0.0575$) per cent of the variation in yields. TWBPLOW and TWTPLOW loadings were 0.54 and 0.38 respectively.

10. The first digit with the month indicates the number of sprays, *i.e.*, first or second and the second digit refers to the week of the month.

11. Regression analysis on 1983 sample comprising 114 farmers shows that a combination of three sprays and spraying during the last fortnight of August and the first fortnight of September result in significantly higher yields.

REFERENCES

1. Agriculture Price Commission : Cost of Production of Field Crops : Methodology and Empirical Results (Rice, Cotton and Wheat), APCOM Series No. 52, Government of Pakistan, Islamabad, December 1986.

2. S. Akhtar, "Effects of the Timing and the Number of Sprays on Cotton Yields in Sind: An Exploratory Analysis", *Pakistan Development Review*, Vol. XXIV, No. 2, Summer 1985.