Vol XIX
No. 2

APRIL-
JUNE
1964

ISSN 0019-5014

INDIAN JOURNAL OF AGRICULTURAL ECONOMICS

INDIAN SOCIETY OF
AGRICULTURAL ECONOMICS,
BOMBAY

to understand, for the method assumes no such thing. I would suggest to Dr. Shah to study the two series of all-India production estimates of various agricultural commodities, one being the series adjusted by the chain method (Table IV) and the other the original unadjusted series (Table V) in the paper on index numbers of agricultural production in India by myself and the late Shri V. S. Menon[6] to appreciate the nature of the correction made by this method. This comparison brings out incidentally the point that I have made earlier that the corrections are larger in the first few years and vanish completely towards the end of the series. Dr. Shah has, however, made a valid point that corrections to the traditional series of crop yields will not be available earlier than 1945, that is before the crop-cutting surveys were started; but he will find that towards the end of my note, I have suggested that interested workers might dig up old departmental crop-cutting data for this purpose.

Finally, I hope that the main point that I had made in my Note that crop-cutting surveys have brought out such great improvement in our yield and production statistics as to form a landmark in the improvement of agricultural statistics in India, will not be lost upon Dr. Shah and other readers of this correspondence.

V. G. PANSE*

COMPARISON OF YIELD ESTIMATES PREPARED ON THE BASIS OF CROP-CUTTING AND TRADITIONAL METHODS—FURTHER EMPIRICAL TESTING†

Since the discussion on the topic has stretched over more than a year, it would be helpful to recollect briefly the points at issue. It was maintained in the first Note[1] that the superiority of the crop-cutting method for estimating yields per acre is undisputable. However, a research worker would still be interested in knowing whether the traditional yield data and the new yield data can be used together in any profitable manner. To answer this question, in that Note, the variances and the means of the two series were computed and compared and their respective differences were tested by means of F and t tests. The conclusion arrived at was, since in most cases the variances and means are not statistically significantly different from each other, for research purposes the new series of yields can be extended backward with the help of the traditional yield series to observe at least long-term trends.

To this, Dr. V. G. Panse[2] took an exception. His major contention based on general consideration, was that the variance of the traditional series is likely to be low, village agency reporting yields being prone to under-estimate yields in good

6. This *Journal*, Vol. XVI, No. 2, April-June, 1961.

* Statistical Adviser, Institute of Agricultural Research Statistics, (ICAR), New Delhi.

† The author is indebted to Dr. V. R. Panchamukhi for the benefit of helpful discussions and valuable suggestions on statistical matters and Mr. R. H. Matta for statistical calculations. The author owns the responsibility for errors, if any.

1. C. H. Shah, "Comparison of Yield Estimates Prepared on the Basis of Traditional and Crop-Cutting Methods," this *Journal*, Vol. XVII, No. 4, October-December 1962, pp. 33-39.

2. "Why Crop-Cutting Surveys ? Comparison of Yield Estimates Prepared on the Basis of Traditional and Crop-Cutting Methods—A Rejoinder," this *Journal*, Vol. XVIII, No. 2, April-June 1963, pp. 33-36.

years and over-estimate them in bad years. Statistical testing had however negatived this contention so far as the particular data were concerned. The extended implication of Dr. Panse's contention might be that the long-term trends of the two series are likely to be different. Since our main objective was to investigate trends, in reply to Dr. Panse's objection[3] we fitted trends to the two series and tested statistical significance of the differences between the respective trend values. Of the limited number of comparisons none of the differences were found statistically significant. However, what is of importance is that the trend coefficients themselves were, in most cases, not significantly different from zero in both the series.

The major theoretically important issue raised by Dr. Panse in his Note above is that the two series might be highly correlated. If the two series are highly correlated the ordinary F and t tests would not be valid. And it may be added that even the comparison of trend coefficients cannot be considered valid.

To examine this contention of Dr. Panse, we carry out the empirical tests of correlations for our entire data. And where we find the correlation coefficients to be statistically significant we carry out the amended tests for comparing variances and means. These tests are endorsed by Dr. Panse in his Note appearing in this issue.

I

Before we proceed with the empirical testing we should state the logical implications of high correlations first. If the two series are highly correlated then it gives rise to the presumption that the two series can be used for extending one with the help of the other. To illustrate this let us take an extreme case. In an extreme situation of correlation coefficient being 1, all the paired observations would lie on a straight line. One can predict the value of an observation in one series from a given observation in the other series with complete certainty. If a research worker wants to combine these two series—say by splicing at a point of time, one has merely to apply a correction factor in case means of the two series are significantly different. Difference in means would imply a systematic bias—in this case it might be in traditional yields. If the variances are significantly different, the correction may have to be more elaborate (through regression of one series over the other) but we would still be right in using the two series in conjunction with each other in case the correlation is very high. Testing of differences of means and variances would then be for the purpose of evolving appropriate correction factor.[4]

II

We now proceed with our empirical testing. Table below gives the values of coefficients of correlation between two series of yields: (*i*) the traditional or the revenue estimates, and (*ii*) the crop-cutting survey estimates. Their significance is tested at 5 per cent and 1 per cent probability levels.

---

3. C. H. Shah, this *Journal*, April-June 1963, pp. 37-42.

4. Normal assumption is of course implied that the relation between the two series that obtains during the period under study, also obtained during the period over which one series is extended with the help of the other.

CORRELATION COEFFICIENTS

	Rice	Wheat	*Kharif*-Jowar	*Rabi*-Jowar	Bajri
Bombay State	0.9595**	0.9587**	0.8412**	0.7345**	0.6602
Districts :					
Ahmedabad	0.8134*	-0.0957	0.9023*	0.8832**	0.2029
Kanara	0.0468	—	—	—	—
East Khandesh	0.5906	0.9401**	0.9347**	0.7734**	0.6338**
Nasik	0.7657*	0.9324**	0.0704	-0.5387	0.8002*
Ahmednagar	0.7730	0.6133*	0.0143	0.4513	0.5927

* Significant at 5 per cent probability level.
** Significant at 1 per cent probability level.

Out of 26 coefficients of correlations worked out above, about 15 are significantly different from zero, 10 at 1 per cent level and 3 at 5 per cent level. The contention that the two series of yields would be highly correlated will have a little more than 50 per cent support from the above data. However, for the State as a whole, we find in 4 out of five cases the two series are highly correlated.

We have applied the amended test[5] for comparing variances in 15 cases in which correlation coefficient was high and significant. The results are as under:-

COMPARISON OF VARIANCES

	Rice	Wheat	*Kharif*-Jowar	*Rabi*-Jowar	Bajri
Bombay State	**	NS	*	*	—
Districts :					
Ahmedabad	*	—	**	NS	—
Kanara	—	—	—	—	—
East Khandesh	—	NS	*	NS	NS
Nasik	*	**	—	—	NS
Ahmednagar	—	**	—	—	—

NS: Not significant.
* Significant at 5 per cent level.
** Significant at 1 per cent level.

5. $\sigma_1^2 = \sigma_2^2 \mid Z_r - Z_p \mid \sqrt{n-3}$ taken as the normal variate. For details see *Biometrika*, 30, 1938, pp. 188-190.

Out of 15 comparisons, differences in variances in 6 cases are found statistically not significant. Only in four cases differences in variances are found significant at 1 per cent level. For Bombay State as a whole, in three cases the variances of the two series are found significantly different.

Significantly different variances would hinder comparison of the means of the two series since 't' test applied for this purpose presumes equality of variances. Behrens-Fisher test can be applied where variances are unequal as shown below. Leaving out these cases, we apply generalized t-test[6] in case of 6 comparisons where correlation is found high but variances are not found significantly different. Of them, only in *one* case we find the means to be statistically significantly different for two series.

Behrens-Fisher Means Test[7]

In 13 cases of variances of the two series being significantly different, we could not apply t-test for comparing means of two series. A test evolved by Behrens and Fisher is relevant in these cases. Applying the test, we get the following results.

RESULTS OF BEHRENS-FISHER TEST

	Rice	Wheat	Jowar		Bajri
			Kharif	*Rabi*	
Bombay State	0.2607	..	3.228**	0.79485	—
Districts :					
Ahmedabad	7.0923**	..	6.6973**	..	..
Kanara	6.5587**	..	..	..	..
East Khandesh	..	..	4.0384**	..	..
Nasik	1.8077	7.727**	1.9166	..	..
Ahmednagar	..	4.209**	2.1574*	2.138*	..

* Significant at 5 per cent level.
** Significant at 1 per cent level.

Of 13 comparisons given above, in 9 cases means are significantly different, of them in 2 cases at 5 per cent level.

6. This takes account of co-variances. See Mills : Statistical Methods, pp. 213-34.
7. T. W. Anderson : Multivariate Analysis, p. 118.

Combined Results

We may now combine the results of the ordinary F and t tests where correlation is not high, with amended variance and mean tests where correlation is high to get an overall picture.

COMPARISON OF VARIANCES AND MEANS

	Rice	Wheat	*Kharif*-Jowar	*Rabi*-Jowar	Bajri
			Variances		
Bombay State	**@	NS@	*@	*@	NS
Districts :					
Ahmedabad	*@	NS	**@	NS@	NS
Kanara	*	--	--	--	--
East Khandesh	NS	NS@	*@	NS@	NS@
Nasik	*@	**@	*	NS	NS@
Ahmednagar	NS	**@	*	*	NS
			Means		
Bombay State	NSb	NS@	**b	NSb	NS
Districts:					
Ahmedabad	**b	NS	**b	*@	**
Kanara	**b	--	--	--	--
East Khandesh	NS	NS@	**b	NS@	NS@
Nasik	NSb	**b	NSb	NS	NS@
Ahmednagar	NS	**b	*b	*b	NS

@ relate to amended tests for comparison of variances and means.
b relate to Behrens-Fisher Test for comparing means.

Taking F test results and the results of the amended test of variance comparisons and Behrens-Fisher Test for comparing means, we find that out of 26 comparisons, in 13 cases variances of the two series are not accepted to be significantly different. In 9 more cases, they were significantly different at 5 per cent level. Only in remaining 4 cases the difference between the variances of the two series was highly significant—*i.e.*, at 1 per cent level. Comparison of means shows that only in 10 cases means are significantly different in two series, in 3 cases, they were significantly different at 5 per cent and in 8 remaining cases, they were significantly different at 1 per cent level or less.

There is again no consistent tendency regarding estimates of crop-cutting methods being higher or lower in cases where variances or means are significantly different. Out of 13 cases of variances being significantly different, in 2 cases, the variances of revenue estimates were larger. Of two cases of means being significantly different, in one case the mean of traditional estimate was larger ; in the other, the mean of the crop-cutting estimate was larger.

And if we ignore the results of the tests, we find that out of 26 comparisons of variances in 10 cases the variances of traditional estimates were larger. In case of means, out of 26, in 9 cases means of yield estimates of crop-cutting method were larger than those of revenue or traditional estimates. In 17 cases the means of traditional estimates were larger.

III

Thus we find from our study that none of the three contentions, *viz.*, (*i*) revenue estimates and survey estimates, are highly related, (*ii*) the variance of revenue estimate is likely to be low and hence different from that of the survey estimate, and (*iii*) yields according to the revenue agencies are systematically under-estimated, are wholly accepted. There is slightly greater evidence in favour of contention (*i*), less in favour of contention (*ii*), and still less in favour of (*iii*).

What is also necessary to note is that the conclusion cannot be drawn confidently against the above three contentions as well since the evidence is divided. From what we know about the crop-cutting surveys, we are given to understand that the sample is designed to give a confident estimate for the State rather than for the individual districts. For this reason in I.A.D.P. districts sample size has been enlarged to obtain confident estimates of yields. If this is correct, it would follow that there is a weak basis for comparison of yields at district level. At the State level, we find means to be significantly different in one case and variances to be significantly different in three cases out of five though correlations were high and significant in four cases. Hence, our conclusions already stated apply *a fortiori* to the State data.

At the cost of even repetition, we would add that our observations have no implication for general superiority of a scientific crop-cutting method of estimating yields since our inquiry is not intended for it. The data given by this method would be used where they are available. Secondly, we have restricted our study to one State and few selected districts since to us the data for them were easily available. The districts were selected on the basis of the fact that during the period under study (1945-46 to 1955-56), only for these districts their physical boundaries were more or less in tact. For all other districts in the erstwhile Bombay State, the boundaries altered making comparisons over time difficult. Those who have got easy access to much larger data would make a substantial contribution to this important but vexed problem of comparability of the two series of yield estimates, if they extend the empirical testing to much larger area. Till then we have to content with the results obtained here.

C. H. SHAH*

* Reader in Agricultural Economics, Department of Economics, University of Bombay, Bombay.

APPENDIX

VALUES OF VARIANCE TEST (IN CASES WHERE CORRELATION WAS FOUND HIGH AND SIGNIFICANT)

	Rice	Wheat	*Kharif*-Jowar	*Rabi*-Jowar	Bajri
Bombay State	3.400**	1.217	Sign. *	Sign. *	—
Districts :					
Ahmedabad	2.354	—	3.6035**	0.575	—
Kanara	—	—	—	—	—
East Khandesh	- -	1.626	Sign. *	0.309	0.005
Nasik	Sign. *	4.41**	—	—	1.892
Ahmednagar	—	5.103**	—	—	—

VALUES OF MEAN TEST (IN CASES WHERE CORRELATION WAS FOUND HIGH AND SIGNIFICANT)

	Rice	Wheat	*Kharif*-Jowar	*Rabi*-Jowar	Bajri
Bombay State	—	0.7968	—	—	0.0451
Districts :					
Ahmedabad	—	0.2983	—	2.0032*	2.7258*
Kanara	—	—	—	—	
East Khandesh	1.2335	0.3567	—	0.4057	1.1775
Nasik	—	—	—	0.3779	0.4248
Ahmednagar	1.2054	—	—	—	0.2713