Vol XVI
No. 3

ISSN 0019-5014

JULY-
SEPTEMBER
1961

INDIAN JOURNAL OF AGRICULTURAL ECONOMICS

INDIAN SOCIETY OF
AGRICULTURAL ECONOMICS,
BOMBAY

INDEX NUMBERS OF AGRICULTURAL PRODUCTION —SOME METHODOLOGICAL CONSIDERATIONS

C. H. Shah

The importance of a study of trends in agricultural production needs hardly any emphasis. Two major attempts in this field may be noted. The Ministry of Food and Agriculture have constructed a series of index numbers with weights ıor the base year 1949-50. Their results are published from time to time in the monthly Journal of '*Agricultural Situation in India*'. In a recent article titled 'Index Numbers of Agricultural Production in India' published in the April-June, 1961, issue of the *Indian Journal of Agricultural Economics*, Dr. V. G. Panse and Mr. V. S. Menon have reported their results with weights for the base year 1949-50[1] but with different method for calculating the index.

Normally, construction of an index is not a difficult task, but in our country the task of construction of a series of index numbers of agricultural production is complicated by two factors. In the first place the area covered by official statistics increased during the period 1947 to 1957 and consequently inflated the crop production figures. Secondly, the method of estimation of crop production was changed. The scientific crop-cutting method replaced the traditional method ; the former is free from subjective bias, whereas the latter is not. The traditional method estimates production by the formula 'Area × Standard Yield × Condition Factor' ; and both the standard normal yield and the condition factor in the formula are influenced by subjective biases. The change in the method of estimation of production was gradual and was spread over the years 1947 to 1957.

Construction of a production index in India has to reckon with these two factors and make such adjustments as would make the index in its final form free from the vitiating influences of both the factors. Though this is a difficult task, it is made somewhat easy by the fact that the Ministry of Food and Agriculture provide for each year a set of two figures for farm production : one, which would compare both in statistical coverage and the proportions of area covered by two methods of production estimates, with the previous year, and two, which would incorporate in it the change in the coverage and the method of estimation. For the construction of the index free from the influences of the two disturbing factors, the Ministry of Food and Agriculture as well as Dr. Panse and Mr. Menon have utilized the duplicate series of the production data to give a series of comparable figures. Both of them use similar procedure, (but not the same) and essentially the same formula, with only one difference. Dr. Panse uses 1956-57 as the base year for linking, for he believes that most of the changes in coverage and the method of estimation were complete by that year and he does backward-linking between 1949-50 and 1956-57 for adjustment of the production data. The Ministry of Food and Agriculture works its adjustment forward as 1949-50 is the base year for their index. Both the indices are chain indices and give almost identical results. The difference in method of linking therefore has no perceptible influence.[2]

---

1. Please refer to the Appendix at the end.
2. Please refer to the Appendix at the end.

For the adjustment of the production data, Dr. Panse (and implicitly the Ministry of Food and Agriculture) uses the following formula for backward-linking :

$$Y^*_{n-1} = Y_{n-1} \times \frac{Y_n}{Y'_n}$$

where, Y_n = estimate of production for the year n

Y_{n-1} = estimate of production for the year n-1 (previous year).

Y'_n = estimate of production for the year n comparable to the year n-1 both in coverage and combination of methods of estimation.

$\frac{Y}{Y'_n}$ in this formula provides an inflator (or a deflator as the case may be) to raise Y_{n-1} by the difference in coverage and the difference in estimates by two different methods of estimation, *viz.*, the crop-cutting and the traditional and thus make the resulting Y^*_{n-1} comparable with Y_n. It will be claimed that the difference between Y^*_{n-1} and Y_n is then accounted for by a change in weather and material inputs and hence a comparison of Y^*_{n-1} with Y_n will give a pure trend between the two years, free from all biases. The logic is then extended to all adjustments. For instance for adjustment for the n-2 year we get:

$$Y^*_{n-2} = \frac{Y_{n-2} \times Y^*_{n-1}}{Y'_{n-1}}$$

We have to substitute in subsequent calculations like this, Y^*'s for Y's and then we get a series of adjusted data which can be compared over number of years.

The claim that the trend given by the above index is free from bias is however difficult to sustain. For the trend to be free from biases the formula would have to make the following assumptions :

(1) The trend for the reporting area as well as the non-reporting area is the same, or, if there is any difference it is consequential.

(2) The trend obtained from the comparison of the estimates given by the traditional method is unbiased. In other words, if there is a bias in the estimates for individual years, it remains constant over the entire period.

Violation of the second assumption will prove much more serious than that of the first one.[3] In fact the fulfilment of the second assumption is almost impos-

3. In the year 1949-50 details of land utilization were available for 615 million acres (though subsequently this figure was raised to 679 million acres when the Agricultural Statistics of India was published; we stick to the figure given by the First Five-Year Plan since this is a liberal assumption for illustration and in fact in the year 1949-50 the knowledge was confined to that figure) out of total geographical area of 810 million acres of the Indian Union. This means information was not available for 1/4 of the area. However it is estimated that owing to smaller proportion of cultivated area and lower per-acre-yields, the contribution of non-reporting area may amount to 10 per cent of the aggregate. If so, over the period of 1949-50 to 1958-59, the index which shows an increase of 30 points over the base year 1949-50 will be an over-estimate by 3 points if we assume an extreme situation of no increase in production in the non-reporting area. Thus the error caused by the difference in the trends in production in two areas is dwarfed as it is multiplied by a relatively much smaller figure, *viz.*, 0.1. On the other hand, error in trend regarding the traditional method which covered, say, 90 per cent of the area in the initial years will be reduced only to a much smaller extent as it is multiplied by 0.9.

sible, and in this paper we focus our attention on the violation of the second assumption. As we know and as Dr. Panse suggests (p. 18), the crop-cutting method made a small beginning in 1949-50. In 1949-50 obviously its coverage would be small. It gradually replaced the traditional method and only by 1956-57[4] covered substantial area. Thus during early years of the decade 1949-50 to 1958-59 the traditional method would predominate. It can be easily illustrated then that during these years when the traditional method predominated the aggregate trend reflected in a large measure the trend given by the traditional method.[5]

4. In fact in their article (*ibid.*) Dr. Panse and Mr. Menon show that only for major cereals and pulses (rice, wheat, barley, jowar, bajra and gram) crop-cutting method covered more than 90 per cent of the area in 1958-59. For all non-food crops except jute and small millets, and other pulses from among the group of foodgrains, this percentage was less than 50 and in castor seed it was zero.

5. Disaggregation of Dr. Panse's formula will promptly bring out the implied assumption two, *i.e.*, the trend given by traditional and scientific method should be identical and that the trend given by the traditional method should be free from bias.

$$Y^*_{n-1} = Y_{n-1} \times \frac{Y_n}{Y'_n} \text{ (Dr. Panse's formula)}$$

$$Y_{n-1} = Y^t_{n-1} + Y^c_{n-1}$$

$$Y'_n = Y'^t_n + Y'^c_n$$

$$Y_n = Y^t_n + Y^c_n + Y^s_n$$

Where Y^t_{n-1}'s, Y^t_n's are production estimates by traditional method for the years n-1, and n ; and Y^c_{n-1}'s, Y^c_n's are production estimates by crop-cutting method for the years n-1, and n ; and Y'^t_n, Y'^c_n are production estimates for year n by respective two methods for coverage comparably with the year n-1 ;

and Y^s_n is production estimate for year n for additional statistical coverage.

Now the trend (or strictly speaking the percentage increase as the trend in index numbers gives this) can be obtained by

$$\frac{Y^*_{n-1} - Y_n}{Y^*_{n-1}} = \frac{(Y^t_{n-1} + Y^c_{n-1})\ (Y^t_n + Y^c_n + Y^s_n) - Y^t_n + Y^c_n + Y^s_n}{Y'^t_n - Y'^c_n}$$

$$\frac{(Y^t_{n-1} + Y^c_{n-1})\ (Y^t_n + Y^c_n + Y^s_n)}{Y'^t_n + Y'^c_n}$$

$$= \left(\frac{(Y^t_{n-1} - Y'^t_n) + (Y^c_{n-1} - Y'^c_n)}{Y^t_{n-1} + Y^c_{n-1}} \right)$$

and if $\frac{Y^t_{n-1}}{Y^c_{n-1}} > 1$, the trend (in the index numbers) will be dominated by the trend given by $Y^t_{n-1} - Y'^t_n$, *i.e.*, the trend given by the traditional method, which was the case during the few years subsequent to 1949-50. If the production for 1949-50 (Y^t_{n-1}) was low compared to any subsequent year (corrected for coverage Y'^t_n) the comparison will show unduly large increase in production.

The generalized formula will be

$$Y^*_{N-n} = Y_{N-n} \times \frac{Y^*_{N-n+1}}{Y'_{N-n+1}}$$

Where N = for the base year
n = particular year and
-n = so many years back from the base year.

$$Y_{N-n} = Y^t_{N-n} + Y^c_{N-n}$$

$$Y'_{N-n-1} = Y'^t_{N-n+1} + Y^c_{N-n+1}$$

$$Y^*_{N-n-1} = Y^{*t}_{N-n+1} + Y^{*c}_{N-n+1} + Y^{*s}_{N-n+1}$$

Now $$\frac{Y^*_{N-n} - Y_{n-n+1}}{Y^*_{n-n}} = \left(\frac{(Y^t_{N-n} - Y'^t_{N-n+1}) + (Y^c_{N-n} - Y'^c_{N-n+1})}{Y^t_{N-n} + Y^c_{N-n}} \right)$$

This means that if $\frac{Y^t_{N-n}}{Y^c_{N-n}} > 1$ the traditional method predominates.

The trend in the estimates of production given by traditional method will be obtained mainly from difference given by $Y^t_{N-n} - Y'^t_{N-n} + 1$ and since this is true for all the pairs of years compared, the assumption regarding constancy of the bias in the traditional estimates is applicable to all the years.

We can illustrate this with the help of a simple example. Let us assume during the two years 1949-50 and 1950-51 the traditional method covered 90 per cent of production, the remaining 10 per cent being covered by the crop-cutting method. If between these two years the traditional method estimated an increase of 10 per cent but the latter only 5 per cent (for area comparable for two years) then the overall index would show an increase of $(.1 \times .9) + (.05 \times .1) = (.09 + .005)$, *i.e.*, .095 or 9.5 per cent. Gradually the weight of the traditional method will decline as it is replaced by the crop-cutting method. We may have an index reflecting in large measure the trend given by the scientific crop-cutting method only from 1956-57 onward, since only from this year the improved method covered substantial area.

The table below illustrates the entire procedure. Adjusted figures are worked out on the basis of Dr. Panse's formula. The entire data are hypothetical. We have assumed the two methods cover .9 and .1 of the total output in the first year, the proportion of traditional method declines thereafter. We consider here only one homogeneous commodity to abstract from the problem of weights.

SIMULATED PRODUCTION INDEX

	t	c	s	Total	Adjusted	Index
Yn-2	90	10	0	100	113.4	100
Y'n-1	99	10.5	0	109.5	—	—
Yn-1	40	58.5	20	118.5	124.19	109.5
Y'n	44	61.4	22	127.4	—	—
Y'n (rearranged)	66	61.4	0	127.4	—	—
Yn	10	118.6	5	133.6	133.6	107.5

t = estimate of farm production by traditional method
c = estimate of farm production by crop-cutting method
s = estimate for increased coverage, the method used is assumed to be traditional

Y'_n (rearranged) for convenience of comparison only.

In the above table, we essentially kept the percentage increase in t and c constant for two comparisons at 10 and 5 respectively. In spite of this, it will be noticed that between the years n-2 and n-1 index increased by 9.5 per cent and between n-1 and n years by 7.5 per cent. The table incidentally illustrates that though in the year n crop-cutting method covered little less than 90 per cent of production, the comparison of production between years n-1 and n is dominated by the proportions of two methods obtaining in the year n-1. From this point of view 1955-56 may not be a happy choice since even in 1958-59 sizable

areas reported production through traditional method. On the whole, our illustration shows that the bias in the trend of the traditional method will affect the trend given by the index number based on adjusted data.

It is admitted that the bias in the traditional method is purely subjective and hence likely to be influenced by all subjective considerations of those in charge of preparing the production estimates. It fact in some other context, Dr. Panse suggests prevalence of a tendency of over-estimation during the period of low production and of under-estimation during the period of high production.[6] This is an admission of non-constancy of the bias. But this hypothesis of Dr. Panse which may be termed as 'damped fluctuations' hypothesis will have an effect of deflating a trend in the index numbers when the traditional method predominates and is being replaced by a scientific method and also simultaneously efforts are made in the economy for raising the level of production. Later when the traditional method is fully replaced the unbiased trend may be fully reflected. The overall trend then will be initially low and then rising.

More serious results will follow if there is under-estimation or low reporting of production during the initial years for the period beginning with 1949-50. According to one of the recent studies reported by Prof. Dantwala[7], this probably seems to be the case. He feels that during the years around 1949-50, yields-per-acre were abnormally depressed. This he shows with the help of long term (1932-33 to 1958-59) series of yields-per-acre for nearly 16 principal crops both in the form of a table and charts.[8] On evidence of trade data regarding cotton, this phenomenon of 'dip' is attributed to a tendency of under-estimation of production—a result of the play of subjective biases. That this tendency of under-estimation has affected the crop production data and especially foodgrains is also suspected on the ground that during the years 1949-50 to 1952-53 there was Govern-

---

6. 'Areas and Yields of Principal Crops in India', V. G. Panse, *Agricultural Situation in India*, June 1952, p. 144.

7. "Trends in Yields Per Acre," M. L. Dantwala, published in 'Changing India,' Ed. V. M. Dandekar, and N. V. Sovani, Asia Publishing House, Bombay, 1961. Pp. 356. Rs. 18.

8. The data indicate the following position.

	Percentage increase (+) or decrease (—) in yields per acre during 1948-51 over 1936-39
Rice	—15.3
Wheat	— 6.0
Jowar	—26.8
Bajra	—24.8
Maize	—19.0
Barley	—13.0
Gram	—16.5
Sugarcane	— 1.6
Cotton	— 2.6
Jute	+ 7.2
Linseed	— 4.7
Rape and Mustard	— 8.6
Sesamum	—17.2
Castor seed	—12.7
Groundnut	—17.7
Tobacco	—19.0

ment control on production, prices and distribution of selected commodities.[9] The Government controls would encourage under-reporting because that would mean less responsibility for compulsory procurement and more ease of obtaining supplies from the central pool.

The effect of under-estimation of output, when the Government controls are lifted will be to exaggerate the trend in farm production. This result follows from the predominance of the traditional method of estimating production during the years (when the tendency to under-estimate prevailed) as already illustrated earlier. The formula adopted by Dr. Panse and Mr. Menon (as well as the Ministry of Food and Agriculture) does not eliminate this bias in the trend. Their method in fact removes only the influence of the increased coverage and the bias in traditional method only if it is constant over the years considered. Thus, despite the efforts of Dr. Panse and the Ministry of Food and Agriculture to provide a pure trend through an unbiased index, we do not have even now a truly unbiased index of agricultural production.[10]

---

9. Even if one argues that the low level of production in 1949-50 was genuinely so, it does not affect the main theme of our argument since a year which temporarily recorded abnormally low production is a bad choice as a base year. This applies to Dr. Panse's index too. Even though the base for linking is shifted to 1956-57, the comparison will be made regarding the change between the years 1949-50 and 1958-59 to get an idea of production increase.

In fact, the Ministry of Food and Agriculture argues that in the year 1949-50, the traditional method did not under-estimate production. "Official and crop-cutting estimates of production in 1949-50 were *very nearly* (italics ours) equal and as such the official estimates of production are comparatively free from bias of under-estimation." (*Agricultural Situation in India*, July 1954). Hence the selection of the year 1949-50 as the base period. Nevertheless, we have no evidence to show that the same relationship obtained during all the years. The method employed for constructing index assumes bias in the traditional method to remain constant over the entire period. In any case, if the production for 1949-50 is comparable with the subsequent period, it is also comparable with previous years if we grant the comparability on the ground of no significant difference between official and crop-cutting estimates and if so the low level of yields in 1949-50 becomes a truly pathological case.

10. If one imposes a restriction of 'no increase in area under production' and substitute average of yields per acre for 1936-37 to 1938-39 for those obtaining in 1949-50 and calculate the index and compare it with Dr. Panse's index with 1949-50 = 100 one would get the following results:

INDEX NUMBERS OF AGRICULTURAL PRODUCTION

	Estimate on the basis of yields per acre[1]	Dr. Panse's estimate[2]
1949-50	100.0*	100.0
1953-54	101.7	114.3
1955-56	99.1	116.8
1958-59	112.5	132.3

* Average for 1936-37 to 1938-39 instead of 1949-50.

1. The items covered are rice, wheat, bajra, jowar, maize, barley, gram, oilseeds, tobacco, cotton, jute, sugarcane.

2. The items covered are rice, wheat, jowar, bajra, maize, ragi, small millets, barley, gram, tur, other pulses, groundnut, sesamum, rapeseed and mustard, linseed, castorseed, cotton, jute, mesta, tea, coffee, rubber, sugarcane, tobacco, potato, pepper, chillies and ginger.

According to the Government data less than 2 per cent has been added to the net sown area through reclamation. If we assume 5 per cent addition to gross cropped area (to provide for increased double cropping and reclamation through private efforts) the index of 112.5 will be raised to 117.06 $\left(1.00 + .12 + .05 + (.12 \times .05 = .006)\right)$ *i.e.*, 1.1706 or 117.06 in terms of Index. The difference between 117.06 and 132.3 can be one (though crude) measure of exaggeration of the increase due to under-estimation during the early years.

APPENDIX

Are the Index Numbers of Agricultural Production prepared by the Ministry of Food and Agriculture and by Dr. Panse different?

First attempt of Ministry of Food and Agriculture to prepare the index numbers of agricultural production was reported in 1954, in an article "Index Numbers of Agricultural Production" published in *Agricultural Situation in India*, July 1954. The method used for preparing the index number is described in this article as under. ". . . . the Index Numbers have been constructed by the chain base method. For each crop all-India production during a year has been expressed as a production relative with the *corresponding* (italics ours) production in the preceding year as base. These production relatives for each crop have been linked to the production in the base period through the intervening chain relatives to give the production index for that year. . . . " (p. 206). In other words, using the notations given in the preceding text of our paper, the formula used is,

$$YN + 4 = \frac{Y_{n+3}}{Y'_{n+4}} \times Y^*_{n+3}$$

This is what is described as forward linking. The article states clearly that "the year 1949-50 has been chosen as the base period" (p. 205) and "weights have been assigned to the different commodities in proportion to the total value of production of each during the base period. The production has been evaluated at the annual harvest prices prevailing during the year." (p. 206). This leaves no doubt regarding the base period being 1949-50.

Subsequently, in *Agricultural Situation in India*, August 1957, we find the following statement: "Index Numbers given in this article are so constructed as to link up the production of each year with that of 1949-50 through pairs of comparable figures for each two successive years which are available under our system of crop estimation. Therefore, in any analysis of trends in the production of any commodity or in agricultural production as a whole, we must rely on these indices rather than on absolute figures which are not strictly comparable." (p.450). The above statement is repeated *verbatim* in *Agricultural Situation in India*, August 1958 (p.421) and in August 1960 issue (p.555).

We have already discussed the method of backward linking from 1956-57 to 1949-50 used by Dr. Panse. Regarding base period, he observes thus: "The year 1949-50 has been taken by the Ministry of Food and Agriculture as the base period for the index number of agricultural production; but since for calculating adjusted production relatives for individual commodities the base taken is 1956-57, it is logical to shift the base for the index number of agricultural production also to that year. The index numbers calculated for the period 1949-50 to 1958-59 with 1949-50 as the base and also with 1956-57 as the base are shown in Table VII. The prices for 1949-50 used as weights in the calculation are shown in Table VIII." (*Ibid*, p. 25). Dr. Panse has not given weights or prices for 1956-57. From the above statement, it is not clear whether for the base year 1956-57 different weights are used. We have to infer that the same weights have been used for both the base periods 1949-50 and 1956-57. If this is so, then the differences in two in-

dices, if any, would be accounted for by the difference in the method used for linking the figures for a given year with the base period. For comparison, figures for 1956-57 as base are recalculated with 1949-50=100. Two series compare as under.

Year	1949-50 = 100 Index Numbers of Agricultural Production (28 commodities) calculated by :	
	Ministry of Food and Agriculture*	Dr. Panse and Mr. Menon†
1949-50	100.0	100
1950-51	95.6	95.657
1951-52	97.5	97.5186
1952-53	102.0	102.1092
1953-54	114.3	114.392
1954-55	117.0	117.122
1955-56	116.8	116.8734
1957-58	114.6	124.069
1958-59	132.3	132.382
1959-60	127.2	

* *Agricultural Situation in India*, August, 1960, p. 560.
† *Indian Journal of Agricultural Economics*, Vol. XVI, No. 2, April-June, 1961, p. 35. (Recalculated.)

As can be seen from the table, the two series give figures which compare with each other except in the decimal point. (This difference may also be due to the fact that we do not have original detailed data with 3 or 4 decimal points for recalculation). We can conclude, therefore, there is no perceptible difference between the two indices, and hence between the two methods.