ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT
STATE COMMITTEE OF THE RUSSIAN FEDERATION ON STATISTICS
GOVERNMENT OF ST. PETERSBURG AND LENINGRAD REGION
with the participation of FAO, UN/ECE, EUROSTAT

Sixth IWG.AGRI Seminar on Agricultural Statistics

Russian Federation, St.Petersburg, 29 June - 3 July 1998

STATISTICAL METHODS OF ANALYSING THE DEVELOPMENT OF AGRICULTURAL PRODUCTION

Prepared by A. Manallya, Head of Department, Centre for Economic Research attached to the Government of the Russian Federation

First day: Monday, 29 June 1998 *IWG.6 / Seminar paper 12*

STATISTICAL METHODS OF ANALYSING THE DEVELOPMENT OF AGRICULTURAL PRODUCTION

(Prepared by A. Manallya, Head of Department, Centre for Economic Research attached to the Government of the Russian Federation)

1. Agricultural production is a result of the complex interaction of biological, soil and climate and anthropogenic factors. This interaction is manifested in the nature of the development of production (static, increasing, decreasing) and its stability.

2. The results of productive activities in agriculture each year depend to a considerable extent on the prevailing agro-meteorological conditions during the sowing, growth and harvesting of crops. In extreme years, variations in individual crop yields relative to an evolving curve (smooth curve describing regular changes in the mean level) can be as much as 25-30% for the country as a whole. In livestock farming weather conditions are less influential on output than in crop farming. However, there is a seasonal component in both the output of livestock products and the productivity of farm animals, which means that deviations from the evolving curve may be just as large as in crop farming.

3. Let us look more closely at changes in agricultural production, and especially crop yields, in the light of the theory of dynamic series developed by the Soviet statisticians V.M. Obukhov, N.S. Chetverikov, Alb. L. Weinstein, S.P. Bobrov and B.S. Yastremsky. According to this theory, crop yields, as a specific statistical category, are the result of a non-stationary stochastic process which contains necessary and random elements. Necessity is expressed in the form of the trend of the series, and randomness in the form of the variations in the levels of the series relative to the curve expressing the trend. It is obviously not possible to identify all the causes of the trend, which can be viewed in theory as the cumulative impact on crop yields of all the factors which change over time and exert a positive or negative influence. In general terms, the trend expresses the influence on crop yields of changes in the standard of farming, science and technology and the organization of the production process, i.e. regular change in crop yields over time. In any inquiry the period for which crop yields are studied is limited, and hence representations of the real trend are also limited. The trend defined by one particular segment of time will differ, both in its parameters and in form, from the real trend. Thus, the trend is also subject to a certain uncontrolled spread. The complete separation of random and necessary elements exists only as a scientific abstraction. Variations in crop yields relative to the trend are due to random, elemental causes, principally agro-meteorological conditions, disasters, pests and diseases, which have an unpredictable influence. The stability of production depends primarily on this component of the time series and is measured relative to the mean level as it changes over time (evolving curve, trend).

4. The factors causing the trend and variability do not affect crop yields in isolation. For fertilizers to take a form that can be assimilated by plants, there needs to be precipitation. For the biological productivity potential of a new variety to be realized, the warmth and moisture conditions must be optimal. Thus, the influence of factors of intensification on crop yields is manifested not only directly, but also through a complex interaction with the agro-meteorological conditions prevailing each year. Separating the dynamics of crop yields into their component parts is a conventional descriptive approach that, as N.S. Chetverikov repeatedly emphasized, is not to be equated with any division of the phenomenon itself, which acts indivisibly in all its manifestations.

5. Studies by various authors nevertheless show that, despite the interdependence of the trend and variability and the large influence on crop yields of the interaction of anthropogenic factors and weather conditions, the decisive factor determining the trend is the standard of farming and the main cause of variability is change in weather conditions. According to the statistical theory of dynamic series, indicators of change may be divided into two groups: trend indicators and variability (or stability) indicators.

TREND INDICATORS

6. The trend of an economic phenomenon is characterized by indicators of the velocity, acceleration and rates of development. The mean annual chain and base indicators formally accepted in textbooks and manuals on statistics adequately reflect the development of the phenomenon over time when the dynamic series change gradually, i.e. experiencing no sharp fluctuations in individual periods. For series subject to considerable variation, these indicators may severely distort the actual trend, since their magnitude is determined by the values of the dynamic series at each end of the period studied.

7. Let us consider the widely applied indicator of mean annual increment (h) =

$$h = \frac{Y(n) - Y(1)}{n - 1},$$

where Y(n) and Y(1) are the first and last levels in the time segment in question, and n is the number of levels in the series.
It is clear from the formula that the mean annual increment does not depend on the intermediate levels. As an example, let us look at the mean annual increment in grain crop yields in the Russian Federation for 1988-1997 (table 1).

Table 1

CHANGES IN GRAIN CROP YIELDS IN THE RUSSIAN FEDERATION
on farms of all categories

Years	Yield (weight after processing), quintals/hectare		Deviations from trend	
	Actual	Calculated from equation F(t)=16. 42-0.22t	Absolute	Relative %
1988	14.2	16.2	-2.0	12.3
1989	16.1	16.0	0.1	0.6
1990	18.5	15.8	2.7	17.1
1991	14.4	15.5	-1.1	7.1
1992	17.2	15.3	1.9	12.4
1993	16.3	15.1	1.2	7.9
1994	14.4	14.9	-0.5	3.4
1995	11.6	14.7	-3.1	21.1
1996	12.9	14.4	-1.5	10.4
1997	16.5	14.2	2.3	16.2

$$h = \frac{16{,}5 - 14{,}2}{10 - 1} = 0{,}26 \text{ (q/ha)}$$

From this figure it might be concluded that crop yields over the period under consideration grew at a mean velocity of 0.26 quintals/hectare per year. In actual fact, however, it is obvious even from the dynamic series itself that there was an overall downward trend in crop yields, and the result obtained is merely the consequence of the relatively high yields in 1997, which were due to favourable agro-meteorological conditions. If the last year of the period under review had been 1995 or 1996, then the mean annual increment would have been negative (-0.37 or -0.16 quintals/hectare, respectively).

8. A more stable indicator can be obtained from analytical smoothing of the time series. Let us call the mean annual increment obtained in this way the "smoothed increment". To define it, we must resolve the question of the form of the curve describing the trend. The mean annual increment, as a measure of the mean velocity of development of the phenomenon, can be appropriately determined only by the linear function:

$$f(t) = a(0) + a(1)t$$

The mean annual increment is equal to the first time derivative of this function, i.e. the coefficient of regression "a(1)". The imputed level at the end of the series is obtained by adding the coefficient "a(1)" (n-1) times to the imputed level at the beginning of the series.

9. Thus, however complex the trend of the series and however poor the approximation given by the straight line, the best way of expressing the mean annual increment (mean velocity) using the least squares method can only be a linear equation.

10. The smoothed mean annual increment in yields is -0.22. Using the absolute mean annual increment (0.26 q/ha) to analyse the changes in crop yields may lead to the erroneous conclusion that there was an increase, whereas grain crop yields for the period 1988-1997 actually fell at a mean velocity of -0.22 q/ha.

11. There is a similar drawback with the mean annual growth coefficient calculated by the geometric mean formula:

$$K(g) = n-1 \quad \frac{y(n)}{y(1)}$$

Where there is a large variability in the levels of the series, "K(g)" becomes unstable and dependent on random variations in the levels of the series at the ends, which often leads to a distorted representation of the development of the phenomenon in question. To avoid this, it is advisable for the growth coefficient to be calculated by analytical smoothing. By analogy with the smoothed increment, let us call this coefficient the "smoothed growth coefficient". As in the case of the smoothed increment, it would be wrong to make an arbitrary choice of analytical function.
The mean annual growth coefficient shows by how many times on average each successive level is greater (or smaller) than the preceding one. This is found from the exponential curve:

$$f(t) = ak^t$$

In that equation "k" is the constant expressing the relations of the adjacent levels of the series. The coefficient "k" will be the mean annual smoothed growth coefficient, and T = 100*k will be the analogous rate of growth.
For the above series of changes in grain crop yields for 1988-1997, the mean annual growth coefficient calculated by the geometric mean formula will be:

$$K(g) = 10-1 \quad \frac{y(10)}{y(1)} = 9 \quad \frac{16{,}5}{14{,}2} = 1{,}017$$

The coefficient shows the growth of crop yields, but takes no account of the levels of the series between the first and the tenth. The smoothed growth coefficient does not have this drawback. The equation of the exponential curve obtained by the least squares method has the form: $f(t) = 16.434 * 0.986^t$. The growth coefficient of 0.986 shows that what actually occurred was not an increase but a decrease in crop yields.

12. Thus, as with the mean annual increment, we may conclude that whatever the form of the curve describing the trend of the dynamic series and however good the approximation it provides, the only true way of expressing the mean annual coefficient or rate of growth involves exponential smoothing.

If there is an acceleration or deceleration of the economic phenomenon, then together with the mean velocity it is also useful to calculate the mean annual acceleration, which is equal to the second time derivative of the function

$$f(t) = a(0) + a(1)t + a(2)t^2$$

or 2a(2). The parabolic trend is characterized by the constancy of the second differences of the levels of the dynamic series, which also express the acceleration (increase in velocity). Regardless of the form of expression of the trend, therefore, when calculating the mean annual acceleration, use is made of smoothing of the dynamic series with the second-order parabola.

13. For the series showing changes in grain crop yields, the equation of the parabola describing the trend has the form:

$$f(t) = 15.755 + 0.121t - 0.0314t^2$$

The doubled coefficient of regression with t^2 equal to 0.0628 shows that the crop yields decreased with a slight negative acceleration.

VARIABILITY INDICATORS

14. In a dynamic series where there is no trend, the indicators of variability are calculated with respect to the unchanging average level. In most dynamic series characterizing agricultural production, the average level changes according to a specific law, which is expressed in the form of a trend. It was argued above that the trend, for example, of crop yields, is linked to the action of primarily random causes. For this reason, the variations in the levels of the series resulting largely from the influence of random causes should be viewed in relation to the trend and not to the unchanging average level. Otherwise the variations would be exaggerated artificially on account of the regular changes of the levels of the dynamic series.

15. The question of the analytical expression of the trend is not to be confused with the question of calculating trend indicators by means of analytical smoothing. The form of the trend depends on the characteristic properties of the dynamic series, whereas the trend indicators are strictly related to the particular properties of the analytical functions. Let us suppose that an indicator grows at an accelerating rate. In this case the trend is non-linear. However, if the question concerns the mean velocity of change of the indicator over a given period of time, it is conventionally assumed that the indicator grows in a linear fashion and the mean velocity is characterized by the regression coefficient of the linear equation. The variability of the indicator with respect to the trend depends not only on the choice of the form of analytical expression of the trend, but also on what years (anomalous or average) come at the beginning and end of the period studied. This is especially important for relatively short series. Let us look at the dynamic series of gross milk yields, which has a marked seasonal character: the gross yields rise from the beginning of the year to the middle and then fall back towards the end of the year (table 2).

Table 2

CHANGES IN GROSS MILK YIELDS IN THE RUSSIAN FEDERATION
on farms of all categories, thousands of tonnes

	1994	1995	1996	1997
January	2 354	2 037	1 894	1 761
February	2 470	2 162	2 039	1 897
March	3 272	2 807	2 613	2 454
April	3 856	3 677	3 387	3 268
May	4 757	4 476	3 915	3 793
June	5 477	5 175	4 580	4 385
July	5 121	4 865	4 340	4 168
August	4 426	4 186	3 973	3 818
September	3 708	3 523	3 306	3 124
October	2 658	2 525	2 338	2 195
November	2 016	1 915	1 714	1 604
December	2 061	1 893	1 720	1 599

16. A yearly cycle can be observed for all years and this is connected with the movement of the cattle from stabling to pasturing and back again. However, we can also see that the yields for each month are falling, and this reflects the worsening overall conditions of livestock management. Let us assume that the trend has a linear form. If one end of the series coincides with the initial phase of the yearly seasonal cycle (January 1994) and the other with the point of greatest increase (June 1997), then the trend equation will have the form:

$$f(t) = 3615.2 - 18.88t$$

Let us now look at the opposite case, where the beginning of the period coincides with the extreme point (June 1994) and the end with the low point of the seasonal cycle (December 1997). In this case we obtain the following equation:

$$f(t) = 3814.3 - 25.34t$$

In the second case, the velocity of the decrease in milk yields rose sharply. The trend may rise above its "true" position or, on the contrary, drop relative to this position if the beginning and end of the series coincide with the maximum or minimum phases of the yearly cycle. To bring the trend closer to the "true" position it is necessary either to have a very long series or to trace an empirical broken line on the graph "by eye", intersecting it roughly in the middle of the smooth curve and choosing the actual values of the indicator at the beginning and end of the series which lie closest to this curve, for example, March 1994 and September 1997.
The equation will then have the form:

$$f(t) = 3729.7 - 18.7t.$$

Indicators of variability also need to be calculated here. Since these are to be defined in terms of deviations from the trend, it would be logical, following the proposal of A.M. Gataulin, to add the word "residual" to the names of the indicators - for example, residual mean square deviation, residual coefficient of variation, etc. However, for the sake of brevity, and also to avoid any confusion regarding the terminology used in this paper, let us keep to the generally accepted terms, although on the understanding that these indicators of variability will be determined in relation to the trends.

17. To calculate the individual indicators of variability use has been made of the data on changes in grain crop yields (table 1).

(1) Range of variation (R)

R = E(max) - E(min),

where E(max), E(min) are the maximum positive and negative deviations in crop yields from the trend.

E(max) = 2.7; E(min) = -3.1; R = 2.7 - (-3.1) = 5.8 q/ha

(2) Mean absolute deviation (d)

$$d = \frac{E(t)}{n} = 1{,}6,$$

where E(t) = y(t) - f(t) gives the deviation from the trend in the (t)th year

(3) Mean square deviation (MS)

$$MS = \frac{E(t)}{n} = 1{,}9 \quad q/ha$$

(4) Coefficient of variation (V)

$$V = \frac{\quad}{y(m)} 100 = 12{,}5\ \%$$

where y(m) = mean value, y(m) =

$$y(m) = \frac{y(t)}{n} = 15{,}2$$

18. As follows from the rules of majorization, the mean square deviation is greater than the mean linear deviation, by a ratio of about 5/4. If the absolute measure of variability is used for a comparison of the stability of the dynamic series, then it is preferable to use the mean square deviation, which is more sensitive to sharp deviations from the trend. The coefficient of variation characterizes the intensity of the variation process. This property allows us to compare the variability of one and the same indicator under different conditions - for example, crop yields in different climates. In addition, the so-called coefficient of alienation indicates the stability of the dynamic series A = 1-V. The greater the value of A, the lower is the variability

of the series relative to the trend and the higher its stability. An increase in the absolute variability of the crop yields should not always be seen as a negative phenomenon. If this increase is accompanied by a rise in the level of the crop yields (i.e. there is an upward trend), then the stability of production may also be higher. This phenomenon will be expressed in terms of a lower relative variability and higher absolute values for the deviations of the crop yields from the trend in poor crop years (rise of so-called "guaranteed minimum" yields). There is thus another indicator of stability:

$$E = E(min,i) - E(min,j),$$

where E(min,i), E(min,j) are the greatest negative anomalies of crop yield relative to the trend in the (i)th and (j)th poor crop years.

19. The intensification and scientific and technical progress of agriculture do not in all respects contribute to increasing the stability of production. Some of the unpredictable effects are waterlogging of soils or build-up of toxic substances in the air and on the land. The result is smaller harvests and greater mutagenicity of plants, leading to more variability. The positive results of the intensification and hence sharp increase in productivity in the early stages also lead to lower stability of production. However, if the greater variability is offset by an increase in mean annual production, the stability of production may be said to be increasing.

References

1. Bobrov, S.P. Ekonomicheskaya statistika. M-L., Gosizdat, 1930.

2. Gataulin, A.M. Osnovnye tendentsii dinamiki urozhainosti zernovykh kultur v Tsentralnom raione Nechernozemnoi zony. // Statistichesky analiz selskokhozyaistvennogo proizvodstva. M., Nauka, 1984.

3. Manellya, A.I. i dr. Dinamika urozhainosti selskokhozyaistvennykh kultur v RSFSR. M., Statistika, 1972.

4. Obukhov, V.M. Urozhainost i meteorologicheskie faktory. M., Gosplanizdat, 1949.

5. Paskhaver, I.S. O priemakh sravnitelnoi otsenki vysoty i ustoichivosti urozhainosti selskokhozyaistvennykh kultur. M., Vestnik statistiki, 1954, No.6.

6. Chetverikov, N.S. Statisticheskie issledovaniya. M., Nauka, 1975.

7. Yuzbashev, M.M. Metody izucheniya dinamiki raspredelenii i zavisimostei. M., Statistika, 1974.

8. Yastremsky, B.S. Nekotorye voprosy matematicheskoi statistiki. M., Gosstatizdat, 1961.
