E. PECICAN

Department of Economic Cybernetics
Academy of Economic Studies

Model of Prognosis Depending on Trend and Selfcorrelation

The analysis and prognosis model of the time series, proposed by the author, are based on the ground that, together with the factors imparting the evolutive trend of phenomena, there are also factors with discontinuous action. These latter factors bring about systematic deviations from trend. The selfcorrelation coefficients are considered suitable to retain in the drawn up model the systematic deviations from the general trend, with a view to attaining more accurate prognoses.

The evolution of phenomena in economy seldom presents a perfectly uniform trajectory in time. Deviations from the general evolutive trend occur frequently.

The model proposed by us was drawn up in view of obtaining prognoses, which include both the trend, and the more or less systematic errors, characteristic to the evolution described by the phenomena in the past. For exemplifying the calculations, we shall use the series of data concerning the mean production of maize by hectare in our country, in the time period 1960—1978.

We took into account the following hypotheses, when drawing up the model.

1. The evolution of phenomena in time is under the impulse of some systematical factors, with continuous action, imparting a general trend of development. Such factors are: progress in science, growth of the standard of living, growth of the population, economic development etc.

2. The deviations from trend, when they occur repeatedly, are due to some non-alleatory factors with discontinuous action (conjuncture, compensation pulsatory factors). Such factors are market saturation, lagging-behind recovery, climate, fashion, tradition etc.

The effect of systematic factors with discontinuous action, materialized by the evolutive trend, is considered to be well kept by the statistical analysis methods of chronological series. The analytical methods are adequate to retain the mean dynamics, presented by the followed up phenomena *during a time period. The problem arising in the stage of trend determining,* when a polynomial of the form:

$$y = \sum_{i=0}^{m} a_i t^i \tag{1}$$

is adopted (t representing the time), consists in selecting the degree of the polynomial; a too small degree has as a consequence the obtaining of a general trend much too approximate; a too high degree leads to obtaining values which include, beside the trend, also part of the effect of disturbing factors, a fact affecting the accuracy of prognoses [2] . The option between functions will be based upon the evolutive characteristic, stable during the whole period, and especially in the last time interval.

The trend was determined, in the case of the mean maize production by hectare (y), on the ground of data from the period 1960—1974, presented in Table 1.

Table 1

(in metric quintals)

Years	y	Years	y	Years	y
1960	15.5	1965	17.8	1970	21.2
1961	16.7	1966	24.4	1971	25.1
1962	15.9	1967	21.3	1972	30.7
1963	17.8	1968	21.3	1973	25.0
1964	20.2	1969	23.3	1974	25.1

Data source: Statistical Year Book of the S.R. of Romania, for the period 1961—1975.

The chronogram presented in Fig. 1, through a continuous line, was drawn up on the basis of the data in table 1. The stable element emphasized in the chronogram is the generally ascendent evolution of the obtained production. The polynomial of degree 1 was consequently adopted.

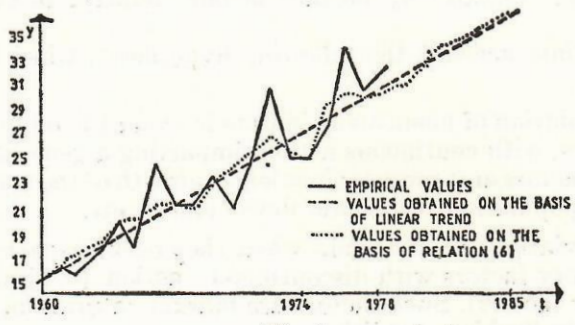

Fig. 1

The following regression equation was obtained after evaluating the parameters:

$$\tilde{} = 21 \cdot 42 + 0.83\, t \quad (t = -7, -6, \ldots, 0, \ldots, 6, 7).$$

The trend obtained by means of the linear function is represented in Fig. 1 by the dashed line.

The effect of the systematic factors with discontinuous action, materialized in deviations from trend, is more difficult to be determined, due to the fact that it appears in a similar manner as the effect of accidental factors. The deviations due to discontinuous factors are differing from the fortuituous deviations by the fact that the first ones appear repeatedly, they are, to a certain extent, systematic. It is necessary to state precisely to what extent and to what time interval the deviations appear systematically, so that this mode of manifestation may be held within the prognosis model. We shall refer to both aspects in the following paragraphs.

We begin by considering that a deviation in time period $t - 1$, generates a somewhat similar deviation in time period t.

$$e_t = \rho_1\, e_{t-1} + u_t \quad (t = 1, 2, \ldots, u) \tag{2}$$

where:

e_t represents the deviation from trend in the time period

$$t(e_t = y_t - \tilde{y}_t);$$

ρ_1 — a coefficient reproducing the extent to which e_t is similar to e_{t-1};

u_t — disturbance (a normally distributed variable, having the mean value $= 0$ and a dispersion σ_u^2, supposed to be independent from e_{t-1}).

Relation (2) represents a selfregressive scheme of order 1. We take into account the evolution in time of the "couple" of values e_t, e_{t-1}; then cov (e_t, e_{t-1}) expresses the common part during the variation of the deviations being at a distance of one time interval. The covariance can be expressed under the form of a quota from dispersion [1], for the analysis of the selfregressive evolution:

$$cov\,(e_t,\ e_{t-1}) = \rho_1 \tau_e^2. \tag{3}$$

It results from expression (3):

$$\rho_1 = \frac{cov\,(e_t,\ e_{t-1}).}{\sigma_e^2}. \tag{4}$$

Relation (4) is used to define the selfcorrelation coefficient. This coefficient measures the dependency between the values $e_t (t = 1, 2, \ldots, n)$, being at an interval of a period (lag 1). The interval to which a deviation becomes systematic can be nevertheless also of two periods (lag 2), of three periods (lag 3) etc. We can proceed accordingly to the calculation of coefficients lag 2 (ρ_2), lag 3(ρ_3) etc. We consider, if we take into account both the significance of the selfcorrelation coefficient and its range of variation ($-1 \leqslant \rho \leqslant 1$),

that the most adequate method is to measure the degree within which a deviation comes back during the time period for which data are known.

It is necessary to proceed to the analysis of the selfcorrelation coefficients of lag k ($k = 1, 2, \ldots, j$), as concerns the second aspect, regarding the determination of the interval to which repeatability occurs. The analysis will take into account the fact that the values of the coefficients are estimations, obtained for a relatively restricted number of observations, reproduced under the form of chronological series. At least two problems derive therefrom:

— the deviations, being at a distance of k intervals, may present analogies in variation, due to hazard. It is necessary, consequently, to test the significance of the selfcorrelation coefficients, with a view to holding back within the model only the magnitudes ρ_k, significantly differing from zero;

— the dependency of deviation e_t with respect to deviation e_{t-k} implies the dependency between e_t and $e_{t-mk}(m = 2, 3, \ldots)$. For instance, the connection between the deviations being at a distance of one interval, assumes that e_1 influences e_2, e_2 influences e_3 etc. We deduce herefrom that e_1 influences deviation e_3 too. Similarly, for lag = 2, the dependency between e_t and e_{t-2} implies the dependency between e_t and e_{t-4}, e_{t-6}. It results that, so as to obtain coefficients ρ_k, able to express the exclusive dependency of the deviations being at a distance of k intervals, a previous calculation is necessary, in order to eliminate the common elements propagating from intervals equal to the divisors of k.

Taking into account relation (2), the variable which excludes the delayed effect of the deviation occurring at the distance of k intervals, is obtained as follows:

$$e'_t = e_t - \rho_k \; e_{t-k}. \tag{5}$$

The determination of coefficients $\rho_{mk}(m = 2, 3, \ldots)$ will be done further on, with the help of variable e'_t. The successsion of the self correlation, obtained as showed above, closes when the values of the coefficients become statistically zero. The left deviation, which does no longer present selfcorrelation, is in fact, disturbance (u_t) — result of the action of random factors*.

Starting from the presented premises, it results that, in order to obtain level y_t which may include the effect of the systematic factors, with continuous and discontinuous action, it is sufficient to add to the trend from period $t(\tilde{y}_t)$, the weighted level of the deviations recorded k intervals before. The weight is represented by the magnitude of the selfcorrelation coefficient of lag k. The model can be represented as follows, in a general formulation:

$$y_t = \tilde{y}_t + \sum_{k=1}^{m} \rho_k e'_{t-k} + u_t \tag{6}$$

where $\rho_k = \begin{cases} = 0, \text{ in case of non-significance;} \\ \neq 0, \text{ in case it presents significant values;} \end{cases}$

e'_{t-k} represents the deviations which were used for obtaining ρ_k;

m — the magnitude of the lag up to which significant coefficients were obtained.

---

* The check of the non-existence of the disturbance selfcorrelation phenomena can be made on the basis of Durbin-Watson's test [1].

For the considered example, the selfcorrelation coefficient, calculated on the basis of the deviations from the linear trend, presents the following values: $\rho_1 = -0.232$, $\rho_2 = -0.42$, $\rho_3 = 0.0269$. $\rho_4 = 0.04$. The first two coefficients were kept within the model. Coefficient ρ_1 though feeble from the viewpoint of significance, was maintained, because its introduction into the computation lead to an improvement of model performances.

The relation, with the help of which adjusted values ($\tilde{\tilde{y}}_t$) were obtained, was:

$$\tilde{\tilde{y}}_t = \tilde{y}_t - 0.232 e_{t-1} - 0.42 e'_{t-2}.$$

The graphical representation (Fig. 1) gives, through the dotted line, the evolution of values $\tilde{\tilde{y}}_t$. We notice that the dotted line follows, in most cases, the continuous line, reproducing, to a certain extent, the changes of sense. The adjustments made on the basis of the proposed model are, from the viewpoint of accuracy, superior to the adjustments carried out on the basis of the linear model. The level of the mean square deviation of the adjusted values ($\tilde{y}_t$), as against the empirical ones (y_t) is $\sigma = 1.786$, lesser than $\sigma = 2.06$, obtained when using the linear function.

It is necessary, so as to obtain prognoses on the basis of relation (6), to obtain previously extrapolated values of the trend. The deviations from trend, for moments $n + 1, n + 2, \ldots$ are determined with the help of relation:

$$e_t = y_t^{**} - y_t^*$$

where: y_t^{**} represents the prognosis obtained on the ground of relation (6);
y_t^* — the level obtained through extrapolation of the linear trend.

The modality of obtaining the deviations make it necessary to determine the forecast levels in successive steps. For example, the maize production prognosis in the second prognosis year (1976) can be obtained only after having previously determined the level forecast for 1975. Concretely, the calculations develop as follows:

— for year 1975:

$$y_{16}^{**} y^* - 0.232\, e_{15} - 0.42\, e'_{14} = 28.06 - 0.232(-2.13) - 0.42(-0.21) = 28.64$$

$$e_{16} = y_{16}^{**} - y_{16}^* = 28.54 - 28.06 = 0.58$$

— for year 1976

$$y_{17}^{**} = y_{17}^* - 0.232\, e_{16} - 0.42\, e'_{15} = 28.89 - 0.232.$$

$$0.58 - 0.42.\ (-2.455) = 29.7866.$$

Table 2 presents, in the prognoses obtained on the ground of model (6) (y_t^{**}), the extrapolations of trend (y_t^*) as well as the real levels (y) from the time period 1975—1978.

The levels forecast by using relation (6) present a higher accuracy with respect to the extrapolations based upon the linear function. The allegation takes into account both the measure within which prognoses get near the actual values (and, in this sense, the level of the mean square deviation, regis-

Table 2
in metric quintals

Years	1975	1976	1977	1978	σ
y	27.8	34.1	30.4	32.1	
y^{**}	28.64	29.78	29.40	30.45	2.40
y^{*}	28.06	28.89	29.72	30.55	2.75

tered in the last column of the table, is conclusive), and the values y_t^{**} (see Fig. 1 — dotted line).

The forecastings obtained through both methods, for the time period 1979—1985, are presented in Table 3.

Table 3
metric quintals

Years	1979	1980	1981	1982	1983	1984	1985
y^{**}	30.22	31.76	33.48	34.07	34.51	35.44	36.44
y^{*}	31.38	32.21	33.04	33.87	34.70	35.53	36.35

The forecasts y_t^{**} in Table 3, as can be seen, tend to extrapolated levels according as the forecasting horizon moves off. This makes it necessary to bring up to date continuously the prognoses, as far as new actual data appear, so as to maintain the increased accuracy, brought about by the model based upon trend and selfcorrelation.

REFERENCES

1. KMENTA T.: *Elements of Econometrics.* The Mc. Millan Co., New York, 1971.
2. RUMSISKI L. Z.: *Prelucrarea matematică a datelor experimentale* (Mathematical Processing of Experimental Data). Editura tehnică, București, 1974.

MODELE DE PREVISIBILITÉ UTILISANT LE TREND ET L'AUTOCORRELATION

Le modèle d'analyse et de prévisibilité des séries de temps proposé par l'auteur a pour base l'idée qu'il existe des facteurs à action discontinue à coté des facteurs qui impriment la tendance évolutive des phénomènes. Ces facteurs provoquent des écarts systématiques au trend Les coefficients d'autocorrélation sont considérés indiqués pour retenir, dans le modèle élaboré, les écarts systématiques de la tendance générale ayant pour but d'aboutir à des prévisibilités plus précises.

МОДЕЛЬ ПРОГНОЗИРОВАНИЯ, ОСНОВАННАЯ НА ТЕНДЕНЦИИ И АВТОКОРРЕЛЯЦИИ

Модель анализа и прогнозирования временных рядов, предложенная в работе, основывается на том предположении, что вместе с факторами, определяющими эволюционную тенденцию явлений, существуют и факторы с непостоянным действием. Эти факторы определяют систематичные отклонения от тенденции. Автокорреляционные коэффициенты считают подходящими для отражения в разработанной модели систематических отклонений от общей тенденции, с целью получения более точных прогнозов.

5.04.1980.