
METHODS AND PROCEDURES FOR DATA ADJUSTMENT, BASED ON CHRONOLOGICAL SERIES USED IN THE ANALYSIS OF DEVELOPMENT TREND FOR VARIOUS SOCIAL – ECONOMIC DOMAINS

PhD Professor Constantin ANGHELACHE

„Artifex” University /Academy of Economic Studies - Bucharest

PhD Professor Gabriela-Victoria ANGHELACHE

Academy of Economic Studies - Bucharest

PhD Lecturer Florin Paul Costel LILEA

„Artifex” University - Bucharest

Ph.D. Student Andreea Gabriela BALTAC

Academy of Economic Studies - Bucharest

Abstract

By specific methods, statistics studies including the tendency of development, designed by the specific literature as the trend and, meantime, by trying to separate the influence of the major factors (with a systematic action), from the influence of the casual factors, which involve deviations between the empirical and theoretical terms. To the extent the casual factors influence is stronger, the degree of variation from one to another unit of time is higher and the line (curve) of tendency is more difficult to identify.

Practically, by the adjustment operation, calculated chronological series are achieved, by emphasizing the development trend, replacing the empirical series.

Key words: *dynamics, phenomena, term, data, dispersion*

As they are referring to social and economic phenomena, the **chronological series** may show very large variations from one to another period. The variations may indicate different tendencies which, in fact, are corresponding to the inner rules governing the development of the respective phenomena.

Typical for the social phenomena is the fact the systematic action of the essential is amended by the influence of the non-essential causes, not only from inside the same period of time but also by in their dynamics. The social and economic phenomena being influenced in their dynamics, not only by the action of the essential factors, are sometimes presenting larger or smaller

deviations from the theoretical general line which, statistically speaking, is expressing the development law specific to each stage.

When analyzing dynamic series, there is either the tendency of approaching to the same value of the absolute increases recorded from year to year, or the tendency of approaching to the increase rhythms with chain basis, which can result.

If the series data are represented graphically depending on the time, one can notice that some phenomena may develop in a rectilinear form and others may develop in a curvilinear form.

Should the series data refer to a longer period of development, it may happen that a tendency of rectilinear development turns into a tendency of exponential development and afterwards into a rectilinear tendency.

Other dynamic series are presenting certain repeating variations (of cyclical character) depending the seasons change, or depending on circumstantial production and market factors, while more rarely there are chronological series which terms seem to distribute themselves totally independently from one another. In the frame of dynamic series, the series terms can not absolutely independent between them, as it is the case of the distribution series where each term is independent as against the other one and is connected to a certain probability of occurrence, materialized by its occurrence frequency. In the frame of the dynamic series the interdependence between the series terms must be considered with the meaning that each term depends on the previous level and to a certain extent, all the terms of the series are determining the following one. For instance: the production recorded for February month is not entirely independent as against the conditions and, respectively, the level of the production of January month of the same year. Moreover, when analyzing the basic reasons and conditions of the phenomena occurrence, one can state out that they are changing systematically, from one to another unit, as a result of gradual quantitative accumulations which can be followed up within an entire stage of development. The objective interdependence between the terms of a dynamic series imparts a certain tendency to the evolution of the investigated phenomena which, due to the fact that the phenomena are influenced in their dynamics by casual factors as well, can not be known otherwise than by means of a theoretical and practical analysis of all the terms of the respective series.

By adjustment of the terms of a statistical data series, as the most general meaning, it is to be understood the operation of substituting the real terms with theoretical terms and it is expressing the specific law for the objective development of the phenomena the respective data are referring to[1]. As to the dynamic series, the development law is analyzed depending on

time. The time variable by itself should not be considered as a decisive factor but a mean of synthesizing successively the systematic influence of the factors acting in the frame of the same basic conditions but of different dimensions.

In the case of the adjustment of the dynamic series, the total dispersion which synthesizes the average size of the variation generated by the influence of all factors, is decomposed into the dispersion calculated on the basis of the real terms as against the adjusted values depending on time, plus the dispersion calculated on the basis of the variation of the adjusted values as from the mean of the dynamic series terms.

The total dispersion:

$$\sigma_y^2 = \frac{\sum (y_i - \bar{y})^2}{n}$$

The dispersion of the series terms from the adjusted values ($\sigma_{y/z}^2$) is synthesizing the influence of the residual factors – unrecorded factors - which, in the case of the dynamic series, are all the factors except the time factor:

$$\sigma_{y/z}^2 = \frac{\sum (y_i - Y_{t_i})^2}{n}$$

where:

Y_{t_i} = the theoretical value of the variable y depending on time.

The dispersion of the adjusted values from the average value ($\sigma_{y/t}^2$), which is synthesizing the variation generated by the modification of the time factor only, considering all the other factors as non-essential and with a steady action for all the cases:

$$\sigma_{y/t}^2 = \frac{\sum (Y_{t_i} - \bar{y})^2}{n}$$

A fact to be underlined is that the theoretical (adjusted) values depending on time can be set up by using several calculation proceedings.

The essential condition of a correct application of one or another of these procedures is that the number of the series terms is big enough in order to enter thus the field of action of the **big number law**, insuring thus a real compensation for the casual deviations[1]. It is also known the fact that the law governing the development of a phenomenon cannot be followed up otherwise than within a full stage, in which frame the basic conditions of the phenomena occurrence are changing quantity wise only. Only when this condition – a enough large number of terms – is met, the tendency (trend) of development of the respective phenomenon can veraciously result.

The statistical theory and practice show that the adjustment methods and proceedings are more frequently used for adjustment: adjustment on the basis of mobile means; adjustment through the graphic method; adjustment on the basis of the average growth; adjustment on the basis of the growth average index; adjustment through the equations of the analytical functions computed by the smallest squares proceeding.

• *Adjustment on the basis of mobile means*

Is applied mainly when the variation of the terms of a dynamic series is showing a cyclical regularity aspect. By calculating the mobile means, the cyclical variation is removed and so the data series is presented with a smooth, continuous variation.

The mobile means are partial means, calculated out of a preset number of terms, in which the first term is gradually replaced by the term following within the series to be adjusted.

The mobile means are known also as sliding or slipping means.

Consider, for instance, a series formed by 8 terms (y_i), which has to be adjusted through the mobile means proceeding, y_i , computed out of three terms each.

The mobile means:

$$\begin{aligned} \bar{y}_1 &= \frac{y_1 + y_2 + y_3}{3}; \bar{y}_4 = \frac{y_4 + y_5 + y_6}{3}; \\ \bar{y}_2 &= \frac{y_2 + y_3 + y_4}{3}; \bar{y}_5 = \frac{y_5 + y_6 + y_7}{3}; \\ \bar{y}_3 &= \frac{y_3 + y_4 + y_5}{3}; \bar{y}_6 = \frac{y_6 + y_7 + y_8}{3}. \end{aligned}$$

The adjustment will be made according to the following pattern:

Valori empirice (y_i)	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8
Valori ajustate (\bar{y}_i)	–	\bar{y}_1	\bar{y}_2	\bar{y}_3	\bar{y}_4	\bar{y}_5	\bar{y}_6	

It results that there is a smaller number of mobile means being obtained as comparatively with the number of the empirical series terms, out of 8 terms there are only 6 mobile means obtained. Generally speaking, this means that there are so many mobile means obtained as how many terms the series has, less the number of terms out of which the means have been

calculated, diminished by one unit. If n represents the number of the series terms and the number of terms for each mobile mean (n'), then the number of mobile means will be: $n - (n'-1)$.

In this case the adjusting operation was relatively simple since, by calculating the means out of an uneven number of terms, each mean placed itself right in front of a real term of the series, *corresponding with the position of the central term*. If the mobile means are calculated out of an even number of terms, each mobile mean will be placed in the middle of the terms (between the two central terms). In order to manage adjusting the terms under such circumstances the mobile means are calculated out of two terms of the adjusted series. The initial mobile means will be called provisional mobile means (noted by \bar{y}_i), while the final or adjusted mobile means (noted by $\bar{\bar{y}}_i$) will be obtained at a second stage:

$$\bar{\bar{y}}_1 = \frac{\bar{y}_1 + \bar{y}_2}{2}; \quad \bar{\bar{y}}_2 = \frac{\bar{y}_2 + \bar{y}_3}{2} \text{ etc.}$$

The adjustment of the terms is made according to the following pattern (series of 8 terms, mobile means of four terms each):

Empirical values (y_i)	y_1	y_2		y_3		y_4		y_5		y_6		y_7	y_8
Provisional mobile means				↓	↓	↓	↓	↓	↓	↓			
\bar{y}_i	-	-	\bar{y}_1	↓	\bar{y}_2	↓	\bar{y}_3	↓	\bar{y}_4	↓	\bar{y}_5	-	-
Final mobile means				$\bar{\bar{y}}_1$		$\bar{\bar{y}}_2$		$\bar{\bar{y}}_3$		$\bar{\bar{y}}_4$			
$\bar{\bar{y}}_i$	-	-		$\bar{\bar{y}}_1$		$\bar{\bar{y}}_2$		$\bar{\bar{y}}_3$		$\bar{\bar{y}}_4$		-	-

In the case in which the adjustment is made on the basis of mobile means calculated out of even number of terms, the mobile means are obtained in two steps:

- 1) provisional mobile means (\bar{y}_i), which are placed between the series terms;
- 2) final or adjusted mobile means ($\bar{\bar{y}}_i$), which are placed right in front of the series terms and by which the adjustment of the initial series terms is made.

• *Adjustment through the graphic method*

The series of empirical data is represented graphically and the line or curve which connects the two extreme points of the chronological series is

visually drawn up so that it shows minimum deviations from the position of the real values in the graph. This visual adjustment is based on the hypothesis that the action of all causes should have been constant all over the period, implying the same absolute or relative form of growth for all the terms and which can be interpreted on the basis of the line/curve of the real values taken depending on time.

The graph used for representing a dynamic series (chronogram) is based on the system of rectangular coordinates, whereas the time is set on the abscissas axis (OX), the forecasted indices being used to set up the representation scale on the ordinates axis (OY)[1].

It is resulting that the adjustment on the graphic representation basis constitutes an instrument of evaluation of the growth tendency, depending on which the method (proceeding) itself to be used as far as the estimation of long and short term tendency may be chosen.

Generally speaking, it is accepted that the best mean of adjustment is the proceeding which, when applied to the empirical data series, allows that the obtained theoretical terms giving minimum deviations from the corresponding real values. In the case presented, taking into account the graph aspect, it is to be expected that the theoretical values which correspond to the adjustment in form of an arithmetical progression (average growth and the linear equation) is giving the smaller deviations. To check this allegation, several calculation proceedings can be used out of which the one satisfying to the best the minimum condition is chosen.

- *Adjustment on the basis of the average growth*

It is used when by processing the data series there are individual growths with growths with chain basis resulting, of close values to each other. In fact, this is corresponding to an increase of the levels of the analyzed characteristic under the form of an arithmetical progression having the ratio equal to the average growth.

The adjustment on the basis of the average growth is based on the relationship existing between the first term (growths with chain basis) and the last term:

$$y_n = y_o + \Delta_{1/o} + \Delta_{2/1} + \Delta_{3/2} + \dots + \Delta_{n/n-1}$$

If admitting that the plus or minus deviations of the individual growths as against the average growth are minimal and get mutually compensated, it becomes:

$$y_n = y_o + \underbrace{\bar{\Delta} + \bar{\Delta} + \bar{\Delta} + \dots + \bar{\Delta}}_{ntimes}$$

namely:

$$y_n = y_o + n\bar{\Delta}$$

Based on the respective relationships, all the terms of the arithmetical progression can be written one by one. The time between the two extreme terms of the series has to be considered as a statistical variable, with values from $t_1 \dots t_n$.

In the statistical theory and practice is it customary that the adjusted terms are noted by Y in order to make the difference with the terms of the empirical series which are noted by y .

By using this proceeding it results:

$$Y_i = y_o + t_i \Delta$$

where:

y_o = the term considered as adjusting basis;

t_i = time variable considered in connection with the adjusting basis being used.

Consequently, in order to find out an adjusted term based on the average growth the term is to be selected as adjusting basis, adding to it the average growth taken from a number of time units equal to the position had by the respective term as against the term chosen as base.

It must be underlined that the basic year of the series cannot be used as adjusting year for all the situations, as in this case there is only the difference of size between the first and the last term of the series which is considered. In order to increase the degree of accuracy of the adjustment according to this proceeding, it is recommended that the selection of the adjustment basis is done after the visual adjustment. Out of the graph, it is chosen the specific term which, by its position, is the closest one to the theoretical right line connecting the two extreme points of the series. It is estimated that in the respective point the arithmetical progression relationship between the first term, the annual growths with chain basis and the last term is achieved to the best. As to the analyzed series which, generally speaking, follows the form of an arithmetical progression, it is not necessary to change the adjusting basis. When a new adjusting basis will be selected, the results will show time values of negative sense for the terms positioned before the respective term as well as positive time values for the subsequent terms.

The adjusting equation will be:

$$Y_i = y_o \pm t_i \bar{\Delta}$$

Two arithmetical progressions will be obtained: a down warding one, calculated from the adjusting basis towards the previous terms and an up warding one, calculated towards the last terms of the series.

- *Adjustment on the basis of the growth average index*

It is used when the terms have an increasing tendency in form of a geometrical progression, where the ratio may be considered as equal to the growth average index (\bar{I}) .

The adjustment is based on the relationship between the first term, the indices of growth with chain basis and the last term. If the last term is written depending on the first one, then will equal the first term successively multiplied by the indices with chain basis:

$$y_n = y_0 \cdot I_{1/0} \cdot I_{2/1} \cdot I_{3/2} \dots I_{n/n-1}$$

If the indices with chain basis have closed values between them, the values may be substituted with the value of the growth average index. This is based on the relationship between the indices with chain basis and their average value. It results:

$$y_n = y_0 \cdot \underbrace{\bar{I} \cdot \bar{I} \cdot \bar{I} \dots \bar{I}}_{n \text{ times}}$$

where from:

$$y_n = y_0 \cdot \bar{I}^n$$

In this case as well, it is possible to determine all the terms of the geometrical progression, calculated depending on the adjusting basis and ratio. A certain adjusted term equals the term selected as basis, multiplied by the growth average index amounted to a power equal to the number of time units corresponding to its position as against the basic term. If the term selected as adjusting basis is within the interior of the series, the equations for the adjusted terms become:

$$Y_{t_i} = y_0 \cdot \bar{I}^{\pm t_i}$$

- *Adjustment through the Method of the least squares*

The adjustment of the chronological series through the submitted proceedings is based on the relationship existing between the first and the last term of the series considered depending the number of terms and expressed as an absolute size – the average growth – or as a relative size – the growth average index. The respective procedures are also known by the specialized literature as mechanical proceedings. It is known that the form of variation is also influenced by the intermediate terms of a dynamic series, which may present deviations from the systematic modification generated by the

governing law. In this respect, the chronological series can be considered as a time variable which forms as a function such as:

$$y = f(t_i)$$

where:

t = the values of the independent variable (time);

y = the values of the dependent variable (phenomena presented by the chronological series)

In terms of statistical practice, we are meeting the most frequently those phenomena which, by their dynamics, appear as linear, exponential and parabolic or hyperbolic functions. Further on, we present the **adjustment on the basis of the linear, exponential and parabolic functions, as being the most frequently utilized for adjusting the chronological series.**

The selection of the function corresponding to the best to the real form of the phenomena evolution is done also on the basis of a careful analysis of the graph as well as the absolute and relative indicators which feature the empirical series of data.

In the case that the graph shows a tendency of absolute steady increase, verified also by a small variation of the growths with chain basis, one can evaluate that the phenomenon is increasing according to a linear function by an estimation equation:

$$Y_{t_i} = \alpha + bt_i$$

where:

Y_{t_i} = the theoretical values of the characteristic which has to be adjusted, calculated depending on the values of the factorial characteristic (t);

α = the parameter having the meaning of an average size which shows the value the resulting characteristic y should have reached if the influence of all factors, except the recorded one, were constant all over the period;

b = the parameter which synthesizes the influence of the factorial characteristic (t) only which, geometrical speaking shows the degree of sloping of the adjusting line/curve. For the adjusting equations of a chronological series b cannot be otherwise than positive if considering the series with objective tendency of steady increase and negative when the series presents a tendency of continuous decrease;

t_i = the values of the factorial characteristic which, in the case of the dynamic series, is time.

When the graph shows a tendency of a relative, steady increase, respectively absolute increases higher and higher, verified also by obtaining close values of the indices with chain basis, then it can be estimated that the phenomenon is increasing as an *exponential function* having the estimating equation:

$$Y_{it} = ab^{it}$$

When on the graph there is a curve occurring, having either a point of maximum, or a point of minimum, it can be stated out that the investigated phenomenon is changing in time under the form of a *second degree parabola*. There are frequent situation when on the graph there is only a section of up warding or down warding parabola, typical to certain stages. In this case it is necessary to evaluate, by comparison with another previous stage, which is the long term tendency and, consequently, to proceed to the series adjustment.

The estimating equation of a second degree parabola, expressed depending on time:

$$Y_{it} = \alpha + bt_i + ct_i^2$$

Similar to the correlation case, in order to find out the parameters of the regression function needed by the series adjustment, the Method of the least squares is to be apply, namely:

$$\Sigma[(y_i - (\alpha + bt_i))]^2 = \min$$

In order to satisfy the respective condition it is necessary to determine the values of the two parameters a and b . The system of normal equations will be used, which are measuring the linear connection between the independent variable x and the dependent variable y . Replacing x_i by t_i , we have:

$$\left\{ \begin{array}{l} na + b\Sigma t_i = \Sigma y_i \\ a\Sigma t_i + b\Sigma t_i^2 = \Sigma t_i y_i \end{array} \right\}$$

Analyzing the variation of the variable y_i depending on time, it has been stated out that it is not depending on time but on other endogenous and exogenous factors.

When analyzing the degree of dependence between the time characteristic t and the volume of production, either for one product or on an overall basis, at the micro or macroeconomic level for which we have to calculate the adjusted values Y , we can notice that the production is not a time function. The production volume of electric energy – as any other social

and economic phenomenon, generally speaking – is depending on a series of other factors which influence is present within all time units. The production depends, among other factors, on the degree of technical endowment, the degree of labor qualification, on the mode of organizing the labor process etc. If these factors do not act, the mere fact that another year passed does not lead to its modification. Time does nothing else but synthesizes, depending on the series periodicity, the combined influence of all the factors of influence, systematic and casual, essential and non-essential, objective and subjective. Due to this reason, in order to annihilate the influence of the time variation, there is a condition to observe:

$$\sum t_i = 0$$

For $\sum t_i = 0$, the system of normal equations previously presented becomes:

$$\text{where from: } \left\{ \begin{array}{l} na = \sum y_i \\ b \sum t_i^2 = \sum t_i y_i \end{array} \right\}, \quad a = \frac{\sum y_i}{n}, b = \frac{\sum t_i y_i}{\sum t_i^2}$$

On the basis of solving the equation system as well, one can demonstrate that the value of a is equal with just the series mean, namely with the value formed on the basis of the constant influence of the essential causes only, since:

$$a = \frac{\sum y_i}{n} = \bar{y}$$

The condition of applying the adjusting proceeding is $\sum t_i = 0$. In order to meet this condition it is necessary to consider the origin of the time values as being the series centre. The series will be divided into two parts, in which the number of negative time units must be equal to the number of the positive ones. In case that the series is formed by an uneven number of terms, the origin of the variation of the time values will be just in front of the central term and the time variation will be measured as whole intervals: 0; $\pm 1; \pm 2; \pm 3$ etc.

In the case of a dynamic series formed by an even number of terms, the origin of the time values will drop between the two central terms and the time variations will be measured as half time intervals: $\pm 1; \pm 3; \pm 5$ etc., 0 being at $\frac{1}{2}$ of the distance -1 and +1[2].

In the case of the adjustment on the exponential function basis, the condition of minimum:

$$\sum (y_i - Y_{ii})^2 = \min$$

becomes:

$$\Sigma [y_i - (ab^t)]^2 = \min$$

Being an exponential function, the equation of the mean of tendency becomes:

$$\log \bar{Y}_i = \log a + t_i \log b$$

The system of normal equations, for the connection of exponential form, is accommodated for the situation when the factorial variable x is considered a time variable t , namely:

$$n \log a + \log b \Sigma t_i = \Sigma \log y_i$$

$$\log a \Sigma t_i + \log b \Sigma t_i^2 = \Sigma (t_i \log y_i)$$

which, for the same condition binding to the dynamic series, respectively $\Sigma t_i = 0$, makes the system becomes:

$$n \log a = \Sigma \log y$$

$$\log b \Sigma t_i^2 = \Sigma (t_i \log y_i)$$

where:

$$\log a = \frac{\Sigma \log y_i}{n}$$

$$\log b = \frac{\Sigma (t_i \log y_i)}{\Sigma t_i^2}$$

Through logarithms, the system of normal equations is similar to the one resulting for the linear function, exception made by the fact that logarithms are used and the values of the two parameters a and b can be obtained through the anti-logarithms process on the system roots.

In order to verify the objectivity of the used function, the dynamic series of the real and theoretical values calculated on the basis of the regression equations of exponential type will be graphically represented.

If the adjustment of the chronological series is made according to the equation of a function of a second degree parabola form, the system of normal equations accommodated to the time factor is obtained:

$$\left\{ \begin{array}{l} na + b\Sigma t_i + c\Sigma t_i^2 = \Sigma y_i \\ a\Sigma t_i + b\Sigma t_i^2 + c\Sigma t_i^3 = \Sigma t_i y_i \\ a\Sigma t_i^2 + b\Sigma t_i^3 + c\Sigma t_i^4 = \Sigma t_i^2 y_i \end{array} \right\}$$

When putting the condition $\sum t_i = 0$, all the uneven powers are null so that the system of normal equations which are to be solved:

$$\left\{ \begin{array}{l} na + c\sum t_i^2 = \sum y_i \\ b\sum t_i^2 = \sum t_i y_i \\ a\sum t_i^2 + c\sum t_i^4 = \sum t_i^2 y_i \end{array} \right\}$$

Out of the second equation the parameter b can be calculated, this being the same as in the case of the right line equation, but from the presented system there is a new system in a and c which is forming:

$$\left\{ \begin{array}{l} na + c\sum t_i^2 = \sum y_i \\ a\sum t_i^2 + c\sum t_i^4 = \sum t_i^2 y_i \end{array} \right\}$$

By solving the system, the values of the two parameters are obtained and the average equation of tendency is written being used afterwards for calculating the individual equations depending on the values of t followed by the adjustment.

In this case as well, both the real and the adjusted data are represented by a graph and if the deviations between the empirical chronogram and the theoretical one are insignificant this means that the adjustment function is well selected.

Selective bibliography

- [1] Anghelache, C. (2008) - „*Tratat de statistică teoretică și economică*”, Editura Economică, București
- [2] Biji, E. et al. (2012) – “*Statistică pentru economiști*”, Editura Economică, București
- Anghelache, C., Bugudui, E., Deatcu, C. (2011) – “*Elemente de econometrie teoretică și aplicată*”, Editura Artifex, București
- Bardsen, G., Nymagen, R., Jansen. E. – “*The Econometrics of Macroeconomic Modelling*”, Oxford University Press, 2005, ISBN 0-19-924649-1, 978-19-9246496
- Bernardin, C. – “*Hydrodynamics for a System of Harmonic Oscillators Desturbed by a Conservative Noise*”, article published by the review “*Stochastic Processes and their Applications*”, Volume 117, no. 4, April 2007, Orlando USA, ISSN 0304-4149