Dr. J. LÖRINCZI

Programme for Estimating Medium and Short Term Forecasts (PROGNO—1)

An application of the Brown-Meyer method for estimating medium term ($C = 1$) and short term ($C \neq 1$) quantitative forecasts is presented. The calculation formula (4a) proposed in the paper allows for a rapid computation of the values forecast for the past and present. The adjustment parameters necessary for the calculation are optimized by a simulation after the data of the dynamic series considered. The programme elaborated and listed for the Felix C 256 system allows for the automatic carrying out of the forecast calculations, the display on the line printer — according to the option specified on the parameter card — of the forecast data or of the experimental and forecast data ($C = 1$ or $C \neq 1$), of the medium variability coefficients of the seasonal proportions ($C \neq 1$) and of the variation diagram of the experimental and of the forecast data.

The study of the evolution in time of natural or social phenomena which are met with in current productive activities in industry, in scientific research or in other fields of social activity is closely linked with the processing of data collections illustrating the dynamic evolution in time of the considered phenomenon. As an eloquent example in this direction we mention the medical field, in which the chronological series of the number of sick persons suffering from a certain disease provides a significant image of the effect- and harmful proportions of that disease within a certain time period. Naturally, it would be very useful if one could establish with anticipation, with a corresponding approximation the probable trend of the sickness in future, namely the number of cases of sicknesses in a coming period in order to take in due time the necessary measures for an efficacious control of the

respective sickness. The present paper aims at establishing approximately the trend of the future evolution of the studied phenomenon, described by chronological experimental data. The computation method presented can be used equally in the medical field, in industry and in the social field.

The evolution in time of the states of a concrete physical system can be described by a series of numerical data x_i obtained experimentally at consecutive time intervals $t_i (i = 1, 2, ..., k, ...; t_1 < t_2 < ... < t_k < ...)$. The series of data characterizing the evolution of the state of the examined system is called *"time-series"* and is denoted [2] by the symbol: $\{x_i, t_i; i = 1, 2, ..., ..., k, ...\}$. The dependence of data x_i on values t_i is, generally, unknown and as a rule cannot be expressed analytically with sufficient exactitude, because some of the conditions determining the evolution of the system are developing in a deterministic way (systematically) and other ones in a random way. Consequently, size x should be conceived as a superposition of two elements: a systematic component $f(t)$ and another random $\varepsilon(t)$. The systematic component renders the *trend* of the process which can be conceived as a time function.

The following work hypotheses are assumed [2], [8]:

1°. The trend of the evolution of the process is determined by the variation in time of the systematic component $f(t)$.

2°. The successive time intervals are equally distanced between them, namely $t_{i+1} - t_i = h =$ constant $(i = 1, 2, ..., k, ...)$.

3°. The time-series with seasonal variations contain the same number of observation periods in each seasonal cycle.

Hypothesis 1°. is formulated on the basis of the practical finding that in most of the cases, the physical systems have, from the quantitative point of view "an inert evolution in the light of which the future appears as an extension proved by arguments of the present on general lines which can be detached from the past evolution" [2]. The time-series to be found in the quantitative forecasting computations contain a finite number n of terms corresponding to the finite number of observation periods. By taking this aspect into consideration, the *forecast problem* is enunciated as follows [2]:

Knowing the time-series

$$x_i = f(t_i) + \varepsilon(t_i), \qquad (i = 1, 2, ..., n)$$

build some *predictors* which should estimate the terms of the series at intervals of the size $t_n + j$ moments subsequent to moment t_n where j represents the *forecast horizon.* The last term of the time-series (x_n) refers to the relative present to which the data of the future periods comprised within the limits of the forecast horizon are related.

The estimation of the systematic components $f(t_i)$ of the dynamic series could be made by the method of the least squares, if the parameters of the states determining the systematic component would remain constant in the time interval covered by variable t_i. In most of the cases, however, the parameters of the states vary in time. Moreover, the value of the systematic component $f(t_i)$ is extremely difficult to grasp by a solely defined relation for the entire set of the consecutive periods, by using the method of the least squares.

Therefore, in the practice of quantitative forecast one resorts to special methods which succeed to grasp to a satisfactory extent both the effects of the long-and-short term elements on the trend and of the accidental factors.

Hypothesis 2°, also fulfilled in most of the cases of the quantitative evolution of the physical systems leads to the obtaining of simple computation formulas, easily programmable on computers, the complete automation of the quantitative forecast computations, with a relatively short running length, being thus possible. In case of the time-series which satisfy this hypothesis too, the structure of the series is preserved, in exchange, the calculations become considerably simpler. This is due to the fact that property $t_{i+1} - t_i = \text{constant}$ allows for an advantageous variable change according to which the time-series appears in the form: $x_i = f(i) + \varepsilon_i$. The estimation by calculus of the systematic components $f(i)$ in keeping with the data of the past periods $x_1, x_2, ..., x_n$ leads us to estimative values $\hat{x}_i$ and the term of the time-series for a certain present period will be

$$x_i = \hat{x}_i + \hat{\varepsilon}_i \qquad (i = 1, 2, ..., n),$$

where $\hat{\varepsilon}_i$ represents the estimative value of the random component $\hat{\varepsilon}_i$. The estimating predictors will be built in this case for intervals of size $n + j$ subsequent to the relative present n.

Consider

$$x_1, x_2, ..., x_i, ... x_n \tag{1}$$

a time-series which satisfies the three above working hypotheses. In the acceptation of the present paper, the denomination of *medium term forecast* refers to a time-series without a seasonal variation for which the forecast horizon m is considered

$$m = \begin{cases} n/5, \text{ if } n \leq 50 \\ 15, \text{ if } n > 50. \end{cases}$$

By *short term forecast* we mean the forecast established on the basis of a time-series with a seasonal variation for which the forecast horizon m is considered:

$$m = C,$$

where C represents the number of periods comprised in a seasonal cycle.

Consequently, $C \neq 1$ means: short-term forecast and $C = 1$ means medium term forecast.

The present paper has adopted the computation method of quantitative forecast by exponential adjustment after the quadratic model of the Brown type elaborated on the basis of the Brown-Meyer fundamental theorem [3]. The method allows the calculation of medium and short term forecasts on the basis of the time-series witha sufficiently big number of terms. On the basis of the data of the time-series (1) the respective calculus relations allow for the establishment of the values forecast for the past and present moments:

$$\hat{x}_1, \hat{x}_2, ..., \hat{x}_i, ... \hat{x}_n \tag{2}$$

and of the values forecast for the future moments:

$$\hat{x}_{n+1}, \hat{x}_{n+2}, ..., \hat{x}_{n+j}, ..., \hat{x}_{n+m}. \tag{3}$$

In the series of values (2) and (3) we have [1], [2], [3], [4], [6], [7], [8]:

$$\hat{x}_i = \begin{cases} x_i & i \leq i \leq C \\ a_{i-1} + b_{i-1} + c_{i-1} & \text{if } C = 1 \\ (a_{i-1} + b_{i-1} + c_{i-1})\hat{Z}_{i-c} & \text{if } C \neq 1 \end{cases} \Big\} \; C < i \leq n \tag{4}$$

$$\hat{x}_{n+j} = \begin{cases} \hat{a}_n + \hat{b}_n \cdot j + \hat{c}_n \cdot j^2 & \text{if } C = 1 \\ (\hat{a}_n + \hat{b}_n \cdot j + \hat{c}_n \cdot j^2) \cdot \hat{z}_{n-C+j} & \text{if } C \neq 1, \\ (j = 1, 2, ..., m) \end{cases} \tag{5}$$

where

$$\begin{cases} \hat{a}_i = 3S_i^{(1)} - 3S_i^{(2)} + 3S_i^{(3)} \\ \hat{b}_i = \dfrac{\alpha}{2(1-\alpha)^2}[(6 - 5\alpha)\, S_i^{(1)} - 2(5 - 4\alpha)\, S_i^{(2)} + (4 - 3\alpha) S_i^{(3)}] \\ \hat{c}_i = \dfrac{\alpha^2}{2(1-\alpha)^2}[S_i^{(1)} - 2\, S_i^{(2)} + S_i^{(3)}] \quad (i = 2, 3, ..., n) \end{cases} \tag{6}$$

$$\begin{cases} S_i^{(1)} = \begin{cases} x_1 & x = 1 \\ \alpha x_i + (1-\alpha) S_{i-1}^{(1)} & C = 1 \\ \alpha \dfrac{x_i}{z_{i-C}} + (1-\alpha)\, S_{i-1}^{(1)} & C \neq 1 \end{cases} \Big\} \; 2 \leqslant i \leqslant n \\ S_i^{(2)} = \begin{cases} x_1 & i = 1 \\ \alpha S_i^{(1)} + (1-\alpha) \cdot S_{i-1}^{(2)} & 2 \leqslant i \leqslant n \end{cases} \\ S_i^{(3)} = \begin{cases} x_1 & i = 1 \\ \alpha S_i^{(2)} + (1-\alpha) \cdot S_{i-c}^{(3)} & 2 \leqslant i \leqslant n. \end{cases} \end{cases} \tag{7}$$

$$\hat{z}_i = \begin{cases} z_i & 1 \leqslant i \leqslant C \\ \beta z_i + (1-\beta)\, \hat{z}_{i-C}, & C < i \leqslant n. \end{cases} \tag{8}$$

The notations used in formulas (6), (7), (8) have the following significance:

α — adjustment coefficient of the dynamic series (1) $(0 < \alpha < 1)$;

$S_i^{(k)}$ — the value of the exponential adjustment operator of order k $(k = 1, 2, 3)$ associated to the period with the current index i $(i = 1,, 2, ..., n)$;

z_i — the coefficient of nonadjusted seasonal variation associated with the experimental value x_i corresponding to period i $(i = 1, 2, ..., n)$;

β — the adjustment coefficient of the dynamic series made up of the coefficients of nonadjusted seasonal variation z_i $(0 < \beta < 1)$;

$\hat{\bar{z}}_i$ — coefficient of adjusted seasonal variation.
The mean square error of the values forecast for the future is given by formula [6]:

$$e_{n,j} = \varepsilon \sqrt{\alpha(2 + 3\alpha + 3(\alpha j)^2)} \qquad (j = 1, 2, ..., m) \tag{9}$$

where ε represents the mean square deviation of the values forecast for the past and present from the experimental ones:

$$\varepsilon = \left(\sum_{i=1}^{n} (\hat{x}_i - x_i)^2 / (n-3) \right)^{1/2}. \tag{10}$$

The forecast computation algorithm also includes (for case $C \neq 1$) the calculus of the mean variability coefficient of the seasonal sizes v_k [9]:

$$v_k = 100\, a_k / m_k, \ \%, \qquad (k = 1, 2, ..., C) \tag{11}$$

where

a_k = mean square deviation of the seasonal size k;

m_k = arithmetic mean for the entire time-series (1) of the seasonal sizes corresponding to season k. We have [9]:

$$m_k = \frac{C}{n} \sum_{j=0}^{\left(\frac{n}{C}-1\right)} x_{C.j+k^i} \qquad a_k = \left(\frac{C}{n} \sum_{j=0}^{\left(\frac{n}{C}-1\right)} (X_{Cj+k} - m_k)^2 \right)^{1/2}. \tag{12}$$

Remarks

1. After a brief survey of the known formulas (4) — (12), we realize that most of the forecasting calculations are formed by the determination of the values forecast for the past and present according to formula (4). Any shortening of the camputation cycle for obtaining values $\hat{x}_i$ has repercussions on the length of the forecasting computations as a whole. In this case, by changing values $\hat{a}_{i-1}$, $\hat{b}_{i-1}$ and $\hat{c}_{i-1}$ given by formulas (6) into formulas (4), resulting are:

$$\hat{x}_i = \begin{cases} x_i & \\ S_1 \cdot S^{(1)}_{(i-1)} - S_2 \cdot S^{(2)}_{(i-1)} + S_3 \cdot S^{(3)}_{(i-1)} & \text{if } C = 1 \\ (S_1 \cdot S^{(1)}_{(i-1)} - S_2 \cdot S^{(2)}_{(i-1)} + S_3 \cdot S^{(3)}_{(i-1)}) \cdot \hat{z}_{(i-C)} & \text{if } C \neq 1 \end{cases} \quad \begin{matrix} 1 \leqslant i \leqslant C \\ C < i \leqslant n \end{matrix} \tag{4a}$$

where

$$\begin{aligned} S_1 &= (3 - \alpha(3-\alpha))/(1-\alpha)^2 \\ S_2 &= (3-\alpha)/(1-\alpha)^2 \\ S_3 &= 1/(1-\alpha)^2. \end{aligned} \tag{13}$$

The adjustment coefficient α being a constant value, it results that sizes S_1, S_2 and S_3 are also constant. Consequently, through formulas (4a) one can directly pass to the expression of $\hat{x}_i (i = 1, 2, ..., n)$ in keeping with the exponential adjustment operators $S_i^{(k)} (k = 1, 2, 3)$ without the explicit calculation of sizes $\hat{a}_{(i-1)}$, $\hat{b}_{(i-1)}$ and $\hat{c}_{(i-1)}$ according to formula (6) being to necessary.

2. Formula (8) expresses the adjusted seasonal variation coefficient (with reference to case $C \neq 1$) for an adjustment coefficient known β and the nonadjusted seasonal variation coefficient z_i. To determine the first elements of series $\{z_i\}_{i=1,...,n}$ various ways of initialization are recommended in literature. Supposing that the volume of the dynamic series (1), n, is an integer multiple of the number of periods contained in a cycle, C, one has proceeded as follows in the present paper:

a) attached to each experimental value x_i was a value y_i obtained by an initial adjustment achieved by a linear regression containing a series of $2C$ consecutive values (two consecutive cycles) [7];

b) the nonadjusted seasonal variation coefficients $z_i (i = 1, 2, ..., n)$ by Shiskin's method [4], [5];

c) for a β known, the adjusted seasonal variation coefficient $\hat{z}_i$ were determined through formula (8).

By proceeding to the way defined at a) and b), the following iterative relations were obtained:

$$ZC_{j+k} = C \frac{xC_{j+k}}{yC_{j+k}} \Big/ \sum_{k=1}^{C} \frac{xCj_{+h}}{yCj_{+h}} \tag{14}$$

$$\begin{cases} y_i = \begin{cases} a - \dfrac{(2C-1)}{2} \cdot b, & \text{if } i = 1 \\ y_{i-1} + b, & \text{if } 2 \leq i \leq 2C \\ a_i & \text{if } 2C < i \leq n \end{cases} \\ a_i = \begin{cases} a & \text{if } i = 2C \\ a_{i-1} + (x_i - x_{i-2C})/2C & \text{if } 2C < i \leq n \end{cases} \\ b_i = \begin{cases} b & \text{if } i = 2C \\ b_{i-1} + 3((2C-1)\, x_i + (2C+1)\, x_{i-2C} - 4Ca_{i-1})/(4C^2-1)C, & \text{if } 2C < i \leq n \end{cases} \\ a = \dfrac{1}{2C} \sum_{k=1}^{2c} x_k \\ b = \dfrac{3}{(4C^2-1) \cdot C} \sum_{k=1}^{2C} (2k - 2C - 1) \cdot x_k \end{cases} \tag{15}$$

3. In the computation formulas presented, the adjustment coefficients α and β appear as unknown parameters. For their determination, (as far as we know) no general computation method valid for all dynamic series con-

sidered was elaborated so far. Therefore, in this paper, the optimal values of parameters α and β are determined through a deterministic simulation by using expressions

$$A(\alpha) = \sum_{i=1}^{n} (x_i - \hat{x}_i)^2, \quad B(\beta) = \sum_{i=1}^{n} (z_i - \hat{z}_i)^2.$$

(Obviously, size $B(\beta)$ is used only in case $C \neq 1$).

We select as optimal values for α and β those values for which we have:

$$A(\alpha_{optimal}) = \min_{0<\alpha<1} A(\alpha)$$

$$B(\beta_{optimal}) = \min_{0<\alpha<1} B(\beta).$$

The estimation of the forecast is made with parameter $\alpha_{optimal}$ in the case of medium term forecasting ($C = 1$), respectively with parameters $\alpha_{optimal}$ and $\beta_{optimal}$ in case of the short-term forecast ($C \neq 1$).

The results of the forecast (4) and (5) are substantially influenced by the values of parameters α and β. In order to obtain results as probable as possible, the steps of simulation were chosen as follows:

$$\Delta\alpha = 0.002, \quad \Delta\beta = 0.02.$$

The number of simulations necessary in this case is:

$$499, \text{ if } C = 1,$$

$$548, \text{ if } C \neq 1.$$

Naturally, if we have solid reasons, the steps of simulation can be increased correspondingly, which leads to a decrease in the number of the necessary simulations.

* * *

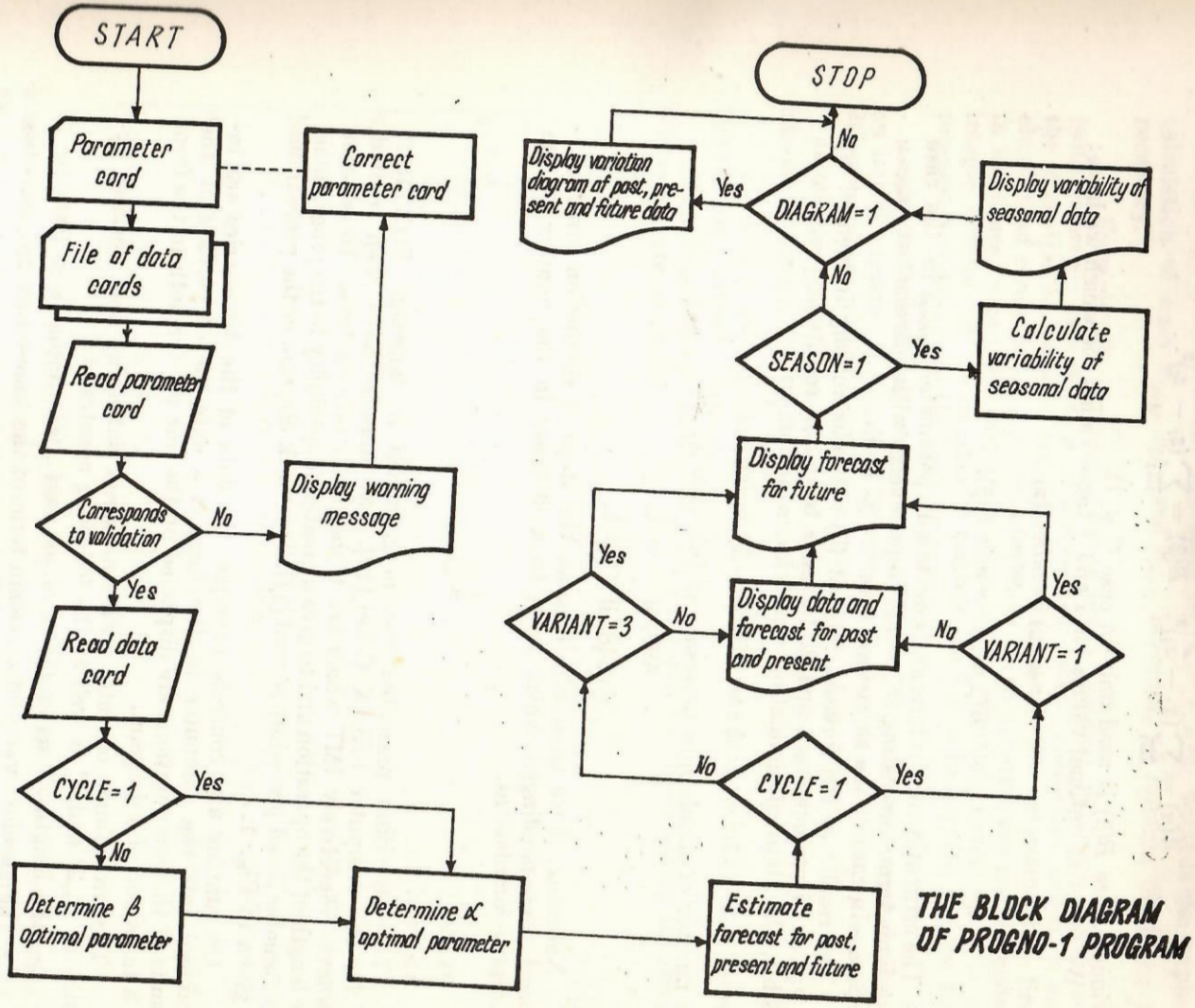
The algorithm presented was programmed in language FORTRAN IV for the computer FELIX C-256 [10]. Two libraries have been created, a source respectively IMT which are to be found safe on tape. To shorten the length of the operation and to give a greater adaptability to the programme, a parameter listed procedure is used [11]. The block diagram of the programme is given in Fig. 1.

On running any concrete example, the data of the time-series are introduced into the computer in the form of a data file on cards and the results of the computations are displayed on the line printer both in the form of a table and of a diagram.

The programme operates with time-series satisfying the working hypotheses 1°, 2° and 3° as well as the following constraints:

— the length of series (1) is of at most 1000 terms ($n \leq 1000$);

— the absolute value of a certain term of the time-series, x_i, satis-fies inequality $|x_i| < 9999\ ,9999$ ($i = 1, 2, ..., n$).



THE BLOCK DIAGRAM OF PROGNO-1 PROGRAM

Fig. 1

Input Data

The data file on cards operates with records of two types, both types having the length of 80 octets each. The first type of record is that of the parameter card (a single record). On this card, through fields of corresponding data, the end user enters

— the number of terms of the time-series (n);

— if the forecast is on medium term $(C = 1)$ or on short term $(C \neq 1)$;

— if the listing is desired of the values of the forecast for the past and present $(\hat{x}_i)$ and of the forecast for the future $(\hat{x}_{n+j})$ or, only of the forecast for the future $(\hat{x}_{n+j})$;

— if the values of the terms of the time-series (1) are of a real type or of an integer type;

— if (in case $C \neq 1$) the computation and display of the mean variability coefficient of the seasonal sizes (v_k) are desired;

— if the graphic representation is desired in the form of a diagram of the results of the forecasting estimations for the past, present and future;

— the denomination of the phenomenon (action) from which the analysed time-series derives (1).

The second type of records comprises 10 consecutive data each of the dynamic series (1). The number of these records is determined by the number of terms of the dynamic series.

Output data

In keeping with the data contained on the parameter card and on the data cards, we display on the line printer

— the table with the experimental data x_i and with the values forecast for them $\hat{x}_i$ $(i = 1, 2, ..., n)$;

— the table with the results of the forecast for the future $\hat{x}_{n+j}$ and their mean square errors $\varepsilon_{n,\,j}$ $(j = 1, 2, ..., m)$;

— the table with the mean variability coefficients v_k $(k = 1, 2, ..., C)$ — in case of the short-term forecast $(C \neq 1)$;

— the variation diagram of the dynamic series x_i, of the forecast sizes $\hat{x}_i$ and $\hat{x}_{n+j}$ $(i = 1, 2, ..., n;\ j = 1, 2, ..., m)$.

In keeping with the data contained on the parameter card, the PROGNO —1 programme can operate $2^3 = 8$ computation variants in case of medium term prognoses and $2^4 = 16$ computation variants in case of short-term prognoses. In total: 24 computation variants.

REFERENCES

1. BOTEZ, M.C.: *Mathematical Tools in Forecasting Techniques.* Bucharest, 1973.
2. BOTEZ, M.C. and contributors: Curs de prognoză. *A Forecasting Course,* Bucharest, 1973.
3. BROWN, R.G., MEYER, R.F.: The Fundamental Theorem of Exponential Smoothing. Operational Research, Vol. 9, Nr. 5, 1961.
4. IONESCU-SISEŞTI, I.: *Prognoza vinzărilor cu aplicaţie la unele bunuri de folosinţă îndelungată.* (Sales Forecasting with Applications to Long Use Commodities). The Academy of Economic Studies, Bucharest, 1972 (in manuscript).
5. KAUFMAN, H., GROBOILLOT, J. L.: *La prévision économique à court terme. — Méthodes générales, lissage exponentiel;* Dunod. Paris, 1968.
6. KILDIŞEV, G. S., FRÄNKEL, A.A.: *Analiz vremennih riadov i prognozirovanie.* Time-Series T. Analysis and Forecasting Statistika. Moscow, 1973.
7. LEWANDOWSKI, R.: *Modelle und Methoden der Ökonomischen Vorhersage.* Dortmund, 1969.
8. MERTENS, P. and contributors: *Prognoserechnung.* Physica-Verlag, Würzburg-Wien, 1973.
9. MIHOC, G., CIUCU, G., CRAIU, V.: *Teoria probabilităţilor şi statistică matematică.* (The Theory of Probabilities and Mathematical Statistics). Editura Didactică şi Pedagogică, Bucharest, 1970.
10. NICULESCU, S.: *Iniţiere în FORTRAN* (Initiation into FORTRAN). Editura tehnică. Bucharest, 1972.
11. PETEANU ,V. and contributors: *Culegere de probleme de programare în limbajul COBOL.* (Collection of Problems in COBOL Language Programming). Editura Dacia, Cluj-Napoca, 1976.

PROGRAMME POUR LE CALCUL DES PREVISIONS A TERME MOYEN ET A COURT TERME

On présente une application de la méthode Brown-Mayer pour le calcul des prévisions quantitatives à terme moyen ($C = 1$) et à court terme ($C \neq 1$). La formule de calcul (4a) avancée permet le calcul rapide des valeurs prévues pour le passé et le présent. Les paramètres de lissage nécessaires au calcul sont optimisés par une simulation selon les données de la série dynamique envisagée. Le programme élaboré et catalogué pour le système Felix C 256 permet d'effectuer automatiquement les calculs de prévision, l'affichage à l'imprimante—selon l'option spécifiée sur la carte paramètre — des données prévues ou des données expérimentales et de celles prévues ($C = 1$ ou $C \neq 1$), des coefficients de variation moyenne des grandeurs saisonnières ($C \neq 1$) et de la diagramme de variation des données expérimentales et de celles prévues.

ПРОГРАММА ДЛЯ РАСЧЕТА СРЕДНЕ — И КРАТКОСРОЧНЫХ ПРОГНОЗОВ

В работе показывается применение метода Brown—Meyer для разчёта количественных средне ($C = 1$) и краткосрочных ($C \neq 1$) прогнозов. Предлагаемая в работерасчётная формула обеспечивает быстрый расчёт значений прогнозируемых величин для прошлого и настоящего периодов. Параметры сглаживания, необходимые для расчёта оптимизируються посредством симуляции по данным предполагаемого динамического ряда. Программа, разрабонная для ЭВМ Felix C—256 обеспечивает автоматический расчёт прогнозов, печатание результатов, — по указанию специальной карты —, относящихся к прогнозируемым или экспериментальным и прогнозируемым данным ($C = 1$ или $C \neq 1$), коеффициентов среднего варьирования сезонных величин ($C \neq 1$) и диаграммы вариации экспериментальных и прогнозируемых данных.

6. 12. 1978.