UNITED ARAB REPUBLIC

THE INSTITUTE OF NATIONAL PLANNING

التخطيط القومي

Memo. No. 1100

The Representative Model Of Economic Time-Series

By

Dr. Josef Kozak

May 1975

CAIRO
SALAH SALEM St.-NASR CITY

Josef KOZÁK

THE REPRESENTATIVE MODEL OF ECONOMIC TIME-SERIES

1. INTRODUCTION

In the last few years I have been working with my colleagues at the Institute of Statistics of the Prague School of Economics in the field of the time-series analysis of selected economic indicators of the post-war Czechoslovak economy. In this connection we use the well-known techniques of decomposition of time-series into trend, seasonal and random components in order to construct short-term forecasts of related economic indicators.

The experience of these studies has shown that in the contemporary stage of development of our economy we have very often to deal with the so-called representative time-series.

In this paper I should like to explain the basic features of this not yet considered concept of time-series of economic indicators. The arrangement of this paper will be as follows:

(a) The concept of the representative model of time-series.

(b) The properties of unweighted least-squares estimation.

(c) The loss-function of the naïve forecast and one way of estimation of the representative constant.

2. CONCEPT OF THE REPRESENTATIVE TIME-SERIES

2.1 The Regression Model

The traditional methods of decomposition of time-series into components (trend, seasonal and random components) are based on the following model

$$Y_{t-k} = X_k' \; a + e_{t-k} \quad \text{for given } t \text{ and } k=0,1,2,\dots , \qquad (1)$$

where for the given time-point t=...,-1,0,1,... and age of observation k=0,1,2,... the meaning of the used symbols is the following.

(a) The Y_{t-k} means for given t and $k=0,1,...,D$ a set of known empirical observations of a given economic indicator; this set constitutes the dependent variable of the model.

(b) The e_{t-k} for given t and $k=0,1,2,...$ means a set of unobservable random errors (disturbances). We assume this set to be a realization of a "white noise" random process with a parameter σ^2 and therefore the errors have zero expectations, variances $\sigma^2 > 0$ and are mutually uncorelated ; written in symbols

$$E(e_{t-k}) = 0$$

$$E(e_{t-k} \cdot e_{t-j}) = \sigma^2 \quad \text{for } k=j \qquad (2)$$

$$= 0 \quad \text{for } k \neq j$$

for $k,j=0,1,2,...$, where $E(.)$ means the expectation operator in probabilistic sense and the stochastic parameter σ^2 is considered to be unknown.

(c) The X_k for $k=0,1,2,...$ is a chosen column vector of (K+1) elements with constants x_{ik} for $i=0,1,...,K$, i.e. after transposition

$$X_k' = [x_{0k}, x_{1k}, ..., x_{Kk}] \text{ for } k=0,1,2,..., \qquad (3)$$

where $K \geq 0$ is an integer. The constants x_{ik} are various functions of age of observation and fulfill the following conditions:

- firstly $x_{0k} = 1$ for $k=0,1,2,...$;
- secondly

$$X_k = L.X_{k-1} \text{ for } k=1,2,... ,$$

where L is a (K+1)-square matris with constant coefficients.

The last condition means that the chosen functions fulfill the conditions of the so-called translational invariance in the sense of R.G. Brown and the matrix L is the shift matrix. It could be shown that under these conditions the vector (3) is composed from polynomial, trigonometric and exponetial functions of age of observation.

(d) In the model (1) a means a (K+1)-element column vector of unknown structural parameters of the model, i.e. after transposition

$$a' = [a_0, a_1, \ldots, a_K]$$

Under the given conditions we can rewrite the model (1) in the form

$$Y_{t-k} = a_0 + \sum_{i=1}^{K} a_i . x_{ik} + e_{t-k} \quad \text{for given } t \text{ and } k=0,1,2,\ldots,$$

where the term

$$X_k' \, a = a_0 + \sum_{i=1}^{K} a_i . x_{ik} \quad \text{for } k=0,1,2,\ldots \tag{4}$$

means the deterministic component of the time-series (the sum of the trend and sesonal components of the model) and the e_{t-k} is the random element of the model. The model (1) is therefore the simplest regression model of a time-series.

2.2 The Representative Model

As has been mentioned the regression model is typical of the traditional methods of decomposition time-series into components; there is a linear regression model with the regressor - an economic indicator Y and regressands - polynomial, trigonometric and exponential functions of time. The application of this kind of model is very easy while we can use all methods of linear regression analysis (including analysis of variance, testing hypotheses etc.). On the other hand we must keep in mind the limitations of this approach , especially in the field of time-series analysis of economic indicators and question whether this simple approach can be adequate in this connection. We mean such circumstances as complicated development of the trend component, unstable seasonal pattern of the time-series, which are very often observed phenomena of a lot of economic time-series in the post-war economics of Czechoslovakia.

The first typical feature of the mentioned regression model are very simple assumptions about the random component. In the statistical literature of the period of the last two decades it has been proved without any doubt that this assumptions are mostly unrealistic in the area of economic time-series analysis. In this connection there have been made different proposals in order to improve this part of the regression model. In my opinion the culminating work in this connection is the well-known book of Box and Jenkins; the basic idea of this book is that every time-series can be thought as a realization of an unknown ARIMA random process (abreviation of "autoregressive-integrated-movin-average" random process); the logical consequence of this approach is then to recover the character of this process, to estimate its parameters and to use them for instance for constructing naive forecasts of the future level of the dependent variable.

I should not underestimate the described kind of the assumptions about the random component of the regression model. But in my opinion in the case of the time-series of economic indicators the kernel of the problem does not lie in this part of assumptions. We must ask why the "classical" assumptions of the regression model (4) are in the case of most Czechoslovak economic time-series in the post-war period unsatisfactory and therefore we have rather to improve the assumptions about the deterministic component of the time-series. In this connection I tried to formulate another concept of the so-called representative model of the time-series.

This model could be described as follows. As in the case of the regression model (1) we consider that

$$Y_{t-k} = Z_{t-k} + e_{t-k} \quad \text{for given } t \text{ and } k=0,1,2,\ldots, \tag{5}$$

where:

(a) as in the foregoing paragraph Y_{t-k} for given t and $k=0,1,\ldots,D$ is a set of known empirical observations and e_{t-k} for $k=0,1,\ldots$ is a set of unobservable random errors with properties (2);

(b) Z_{t-k} constitutes a set of values of the deterministic component (i.e. the sum of the trend and the seasonal) and has an unstochastic character.

In comparison with the foregoing considerations let us assume that there exists such an integer $H \geq K$ (the so-called representative constant of the given time-series), which allows us to describe the first (H+1) values of the deterministic component by means of the appropriate linear regression model, i.e.

$$Z_{t-k} = X_k' \; a_t \quad \text{for given } t \text{ and } k=0,1,\ldots,H, \tag{6}$$

where X_k for $k=0,1,\ldots$ is the known vector (3) and

$$a_t' = [a_{0t}, a_{1t}, \ldots, a_{Kt}] \text{ for given } t \tag{7}$$

is the unknown vector of structural parameters which are changing in time. The basic concept of the representative model in comparison with the regression model is that the description of the following part of the deterministic component by means of the linear regression model is impossible, i.e.

$$Z_{t-k} \neq X_k' a_t \quad \text{for given } t \text{ and } k=H+1, H+2, \ldots \tag{8}$$

In the concept of the representative model in comparison with the regression model is the existence of the representative constant H . This constant could be in connection with (6) considered as a time-characteristic of a "good approximation" of the first values of the unknown deterministic component. In my opinion we can interpret this constant under the conditions of stabilized socialist economy also in a somewhat different way. In this conditions for a given time-point t and ages of observation $k=0,1,\ldots,H$ the chosen linear regression function $X_k' a_t$ has to "discover" the relative stabilized regularities of the given indicator of production, consumption etc. For the given t and higher ages $k \geqq H+1$ these regularities could not be described with the help of this function because the vector of structural parameters a_t has changed in time or the used vector X_k is for $k \geqq H+1$ not appropriate any more.

At the end of this sedtion let us describe the circumstances under which there is a possibility to substitute the representative model of an economic time-series with the regression one. The first condition is that the elements of the vector (7) are relatively stable in time, i.e. there are on a set of time-points t invariant and therefore for these points a_{it} converges to a constant a_i for $i=0,1,\ldots,K$. The second condition is that the representative constant H is relatively high (theoretically H tends to infinity). In our experience in the case of Czechoslovak post-war economy both these conditions are fulfilled very rarely.

2.3 The Naive Prognostic Model

The main target in analysing the economic time-series is to construct an idea of the unknown level of the economic indicator in the future, i.e. to construct its forecast (prediction). It is well-known that this problem can be rationally solved only as follows.

(a) In the present time t=T we must formulate our notion about the future level or the related economic indicator, i.e. to formulate the prognostic model.

(b) Then the related forecast could be constructed in such a way that the unknown parameters of the prognostic model are substituted by its estimates reached at the present time t=T.

In this paper I shall restrict myself on the simplest naïve prognostic model

$$Y^{*}_{T+j} = X'_{-j} \; a_T + e_{T+j} \quad \text{for given } j \, , \tag{9}$$

where j=1,2,...,H is the given horizon of prediction. In this model we assume therefore the continued development of the analysed time-series as observed in the past, i.e.:

(a) for the description of the deterministic component we use the same linear regression model as in the representative model (6) for t=T but for a "future" age k=-j ;

(b) we assume the existence of the random disturbance term fulfilling conditions (2).

3. UNWEIGTED LEAST-SQUARES ESTIMATION

3.1 ULS-Estimation of Structural Parameters

In the last paragraph we tried to show that in time-series analysis of economic indicators we have to prefer the representative model to the regression one. Therefore we have now to solve several questions connected with processing this kind of time-series.

The first of these problems is the estimation of elements of the vector a_t for given t . Because we hold the representative constant H to be unknown as a rule this problem is far from trivial.

Let us restrict ourselves for a moment to the case that we have based these estimates on empirical observations Y_{T-k} for given present time t=T and k=0,1,...,D under the assumption that D is any integer from a set D=K,K+1,...,H. Under these conditions it follows from (5) and (6) that

$$Y_{T-k} = X_k' a_T + e_{T-k} \quad \text{for } k=0,1,\dots,D$$

If the hypotheses (2) about the random element of the model are fulfilled we are in an analogical situation as in the case of the regression model and we can obtain the best linear unbiased estimators of structural parameters a_{iT} for i=0,1,...,K under the criterion

$$F_D(a_t) = \sum_{k=0}^{D} (Y_{T-k} - X_k' a_T)^2 \dots \text{min. for } D=K,K+1,\dots,H, \tag{10}$$

i.e. under the criterion of unweighted least squares (ULS).

Denoting as

$$\hat{a}_{DT}' = [\hat{a}_{D,0T}, \hat{a}_{D,1T}, \dots, \hat{a}_{D,KT}]$$

the estimated vector a_T', we obtain the ULS-estimator in the well-known form

$$\hat{a}_{DT} = A_D^{-1} N_D Y_{DT} \quad \text{for } D=K,K+1,\dots, \tag{11}$$

where

$$N_D = [X_0, X_1, \dots, X_D] \quad \text{for } D=K,K+1,\dots$$

is a (K+1) x (D+1) matrix of given constants x_{ik} for i=0,1,...,K and k=0,1,...,D,

$$Y_{DT}' = [Y_T, Y_{T-1}, \dots, Y_{T-D}] \quad \text{for } D=K,K+1,\dots$$

is the known observation vector,

$$A_D = N_D N_D' \quad \text{for } D=K,K+1,\dots$$

is a symmetrical matrix (K+1) x (K+1) of full rank and A_D^{-1} denotes its inverse matrix.

If we use formula (11) for estimating the vector a_T in the case D=K,K+1,...,H, we can prove that we obtained the best linear unbiased estimates; it can be shown that the mean-square error matrix

$$M(\hat{a}_{DT}) = E\left[(\hat{a}_{DT}-a_T)(\hat{a}_{DT}-a_T)'\right]$$

equals in this case the matrix

$$M(\hat{a}_{DT}) = \sigma^2 A_D^{-1} \quad \text{for } D=K,K+1,\ldots,H \tag{12}$$

In the situation D=H+1,H+2,... it could be shown that these optimal properties do not hold anymore. We could prove that the properties of the estimator (11) differ from the foregoing ones in two directions:

(a) such an estimator is now biased, i.e.

$$E(\hat{a}_{DT}) = a_T + C_{DT} \quad \text{for } D=H+1,H+2,\ldots,$$

where

$$C_{DT} = A_D^{-1} \sum_{k=H+1}^{D} X_k \left(Z_{T-k} - X_k' a_T \right) \text{ for } D=H+1,H+2,\ldots \tag{13}$$

is a (K+1) column vector (the bias-vector) with non-zero elements only because under the assumption (8)

$$Z_{T-k} - X_k' a_T \neq 0 \text{ for } k=H+1,H+2,\ldots ;$$

(b) therefore the relating mean-square error matrix equals now

$$M(\hat{a}_{DT}) = \sigma^2 A_D^{-1} + C_{DT} C_{DT}' \quad \text{for } D=H+1,H+2,\ldots, \tag{14}$$

i.e. the bias-vector affects also the mean-square error matrix.

3.2 <u>ULS-Naïve Forecast</u>

As we have said before we are interested in the vector a_T in connection with the naïve prognostic model (9) only. We can then construct a ULS-naïve forecast in the form

$$P_{DT}(j) = X_{-j}' \hat{a}_{DT} \quad \text{for given } j=1,2,\ldots,H,$$

i.e. we substitute the unknown vector a_T in (9) by its estimate

(11) and the unobservable disturbance e_{T+j} by its zero expectation (2). For computing the ULS-naïve forecast we can therefore use the formula

$$P_{DT}(j) = X'_{-j}\, A_D^{-1}\, N_D\, Y_{DT} \quad \text{for given } j \text{ and } D=K,K+1,\ldots \qquad (15)$$

Before analysing the properties of this forecast we must describe the concept of such analysis. First we can define the forecast error

$$\Delta_{DT}(j) = P_{DT}(j) - Y_{T+j} \quad \text{for given } j \text{ and } D=K,K+1,\ldots,$$

where Y_{T+j} means the predicted future value of the economic indicator. But this quantity is in the present time t=T unknown and we must restrict our analysis to the conditional forecast error

$$\Delta^{*}_{DT}(j) = P_{DT}(j) - Y^{*}_{T+j} \quad \text{for given } j \text{ and } D=K,K+1,\ldots$$

only, where Y^{*}_{T+j} means the level of the naïve prognostic model (9). Comparing $\Delta_{DT}(j)$ with $\Delta^{*}_{DT}(j)$ we can find that these quantities differ in the error of the naïve prognostic model, i.e. in the nonstochastic quantity $(Y^{*}_{T+j} - Y_{T+j})$ for given j.

In analysing the properties of the ULS-naïve forecast we have to investigate the expectation $E\{\Delta^{*}_{DT}(j)\}$ and the loss-function defined by Theil as

$$M\{\Delta^{*}_{DT}(j)\} = E\{[\Delta^{*}_{DT}(j)]^2\}$$

for given j and different D=K,K+1,... .

Let us first consider the case D=K,K+1,...,H again. Under the foregoing assumptions it could be shown that the expectation $E\{\Delta^{*}_{DT}(j)\}$ equals zero and

$$M\{\Delta^{*}_{DT}(j)\} = \sigma^2 (1 + X'_{-j}\, A_D^{-1}\, X_{-j}) \quad \text{for given } j \text{ and } D=K,K+1,\ldots,H \qquad (16)$$

Therefore the ULS-naïve forecast for a given D from a set D=K,K+1, ...,H is a best unbiased linear forecast in the Goldberger sense.

These relations do not apply in the case D=H+1,H+2,... . It could be shown that now:

(a) the ULS-naïve forecast is biased in the sense that

$$E\{\Delta^*_{DT}(j)\} = X'_{-j}\, C_{DT} \quad \text{for given } j \text{ and } D=H+1,H+2,\dots ,$$

where C_{DT} means the bias-vector (13) again;

(b) the relating loss-function equals now

$$M\{\Delta^*_{DT}(j)\} = \sigma^2\,(1 + X'_{-j}\,A_D^{-1}\,X_{-j}) + X'_{-j}\,C_{DT}\,C'_{DT}\,X_{-j} \qquad (17)$$

$$\text{for given } j \text{ and } D=H+1,H+2,\dots,$$

i.e. in comparison with (16) it differs in the quantity $X'_{-j}\,C_{DT}\,C'_{DT}\,X_{-j}$ which can be interpreted as a square of the bias of the ULS-naïve forecast.

4. TWO FACTORS OF THE LOSS-FUNCTION AND AN ESTIMATE OF THE REPRESENTATIVE CONSTANT BASED ON THE ANALYSIS OF PSEUDO-FORECASTS

4.1 Two Factors of the Loss-Function

As shown in (6) and (8) in the concept of the representative model we can describe the unknown deterministic component in the present time t=T by means of a chosen linear regression function $X'_k\,a_T$ in such a way that

$$Z_{T-k} - X'_k\,a_T \;\; = 0 \quad \text{for } k=0,1,\dots,H \text{ and}$$

$$\neq 0 \quad \text{for } k=H+1,H+2,\dots$$

Without loss of generality we can therefore define the bias-vector (13) for an arbitrary D=K,K+1,... as

$$C_{DT} = A_D^{-1} \sum_{k=0}^{D} X_k\,(Z_{T-k} - X'_k\,a_T) \text{ for } D=K,K+1,\dots, \qquad (18)$$

because this vector contains for D=K,K+1,...,H zeros only.

Let us now for a given T,j and D=K,K+1,... define two functions:

(a) the factor of generalization

$$F_1(j;D) = \sigma^2 \,(\, 1 + X'_{-j}\, A_D^{-1}\, X_{-j} \,) \text{ for given } j \text{ and } D=K,K+1,\ldots,$$

which depends on the stochastic parameter σ^2 of the model;

(b) the factor of biasedness

$$F_2(T,j;D) = X'_{-j}\, C_{DT}\, C'_{DT}\, X_{-j} \quad \text{for given } T,j \text{ and } D=K,K+1,\ldots$$

By inspection of (16) and (17) it can be seen that the loss-function of the ULS-naïve forecast is composed of both of these functions, i.e.

$$M\{\Delta^*_{DT}(j)\} = F_1(j;D) + F_2(T,j;D) \text{ for given } T,j \text{ and } D=K,K+1,\ldots \quad (19)$$

The following two theorems can then be proved:

<u>Theorem 1:</u> The factor of generalization forms for a given j and increasing D=K,K+1,... a monotonic decreasing function with $\lim_{D \to \infty} F_1(j;D) = \sigma^2$.

<u>Theorem 2:</u> The factor of biasedness is for given T,j and D=K, K+1,... a function with properties

$$F_2(T,j;D) \begin{cases} = 0 & \text{for } D=K,K+1,\ldots,H \text{ and} \\ > 0 & \text{for } D=H+1,H+2,\ldots \end{cases}$$

The consequence of these two theorems is that the loss-function of the ULS-naïve forecast reaches the minimum at D=H, i.e.

$$\min\left[M\{\Delta^*_{DT}(j)\}\right] = M\{\Delta^*_{HT}(j)\} = F_1(j;H) \quad (20)$$

4.2 The Analysis of the Mean-Square Error of Pseudoforecasts

The main difficulty of applying the foregoing theory in practice is that the representative constant H is, as a rule, unknown. Therefore if we chose an integer $D > K$ we never know if we are not in the situation $D > H$, i.e. if the unknown representative constant H has not been overestimated by the chosen integer D . Therefore we have to propose some simple procedure for estimating the representative constant.

In this connection we use the concept of the so-called ULS-naïve pseudoforecasts. Let us aasume we have a set of known observations of an economic time-series Y_t for the time-points $t=1,2,\ldots,T$, where as before $t=T$ denotes the present time. For given j we can then compute ULS-naïve pseudoforecasts

$$P_{D,t-j}(j) = X'_{-j}\,\hat{a}_{D,t-j} \quad \text{for given } j,D \text{ and } t-j=D+1,\ldots,T-j$$

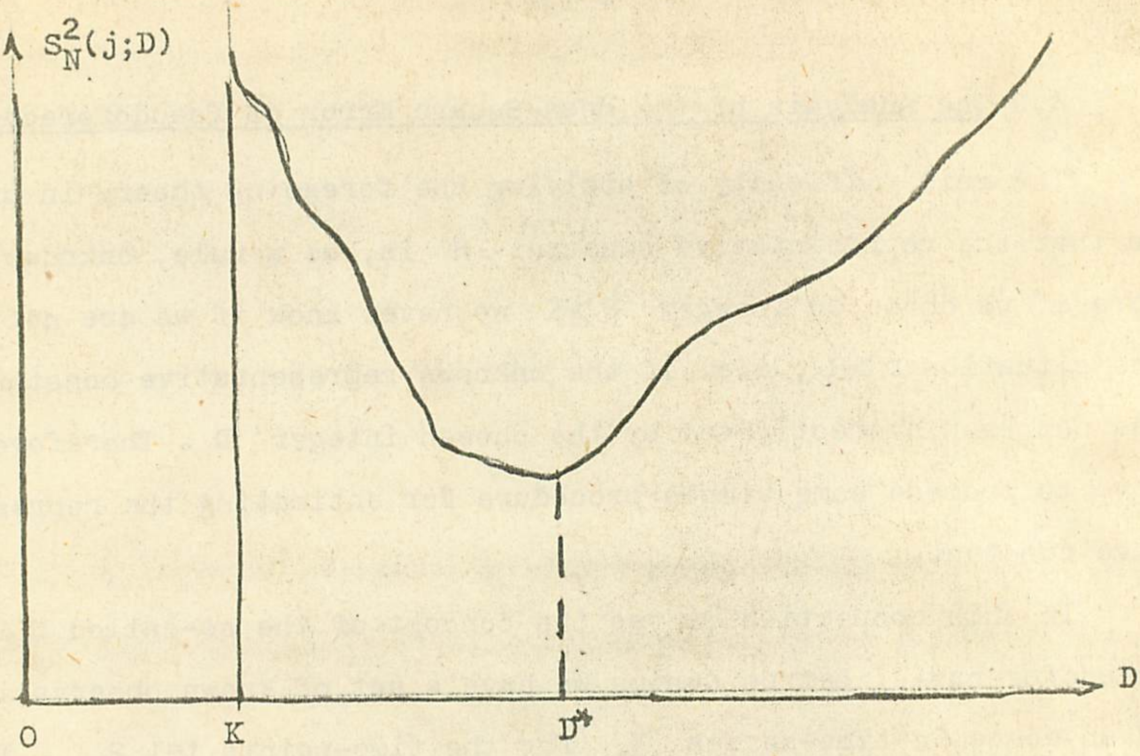
in order to "estimate" the known observations $Y_{(t-j)+j} = Y_t$ for $t=D+j+1,\ldots,T$.

In such a situation we can compute the mean-square error of ULS-naïve pseudoforecasts for given T, j as

$$S^2_N(j;D) = \frac{1}{N} \sum_{t=T-N+1}^{T} \left\{ P_{D,t-j}(j) - Y_t \right\}^2$$

for $D=K,K+1,\ldots,T-N-j$, where N is a chosen integer.

If we used this technique on Czechoslovak economic time-series and plotted the values of $S^2_N(j;D)$ on a graph for different $D=K,K+1,\ldots,T-N-j$, we found that the resulting picture is in the case of stabilized economic time-series as follows .



How to explain such a behaviour of the diagramm of the mean-square error ? If we analyse stabilized economic time-series (i.e. time-series of economic indicators with a stabilized development without substantial jumps) we can prove the following properties of the foregoing graph:

(a) the decreasing tendency for $D=K,K+1,\dots,D^*$ can be motivated by the dominant part of the factor of generalization of the loss-function of ULS-naïve forecasts ;

(b) the increasing tendency for $D=D^*+1,D^*+2,\dots$ can be motivated by the dominant part of the factor of biasedness.

Therefore the unknown representative constant H lies in the neigbourhood of the integer D^* .

5. <u>CONCLUDING REMARK</u>

This paper is a shortened synopsis of one part of a prepared author´s book only. In this book there are, in addition, analysed two interesting quastions connected with the representative model of economic time-series:

(1) Application of weighted LS with weights constituting or the Pascal (negative binomial) or the logistic distribution.

(2) Construction of adaptive forecasts.

References:

BOX,G.E.P.-JENKINS,G.M.: Time-Series Analysis Forecasting and Control; Holden-Day, London-Amsterdam 1970

BROWN,R.G.: Smoothing Forecasting and Prediction of Discrete Time-Series; Prentice-Hall Int., London 1963

GOLDBERGER,A.S.: Best Linear Unbiased Prediction in the Generalized Regression Model; Journal of the American Statistical Association 1962, pages 369-375

THEIL,H.: Applied Economic Forecasting; North-Holland Publ. Comp., Amsterdam 1966

KOZÁK,J.: Representative Time-Series and Their Use in Economic Prognostics (Basic Ideas and Problems); study under preparation, Institute of Statistics of the Prague School of Economics 1975 (written in Czech)

Prague, 5. May 1975

Adress of the author: Doc. Ing. Josef Kozák, Institute of Statistics of the Prague School of Economics; Praha 3, Nám. G. Klimenta 4, Czechoslovakia