I. SAPHIER, Eng.
A. IOSIFESCU
S. MUSTĂȚEA
F. TODOR

Correlation Functions Between Fixed Funds in Industry and Industrial Output

There exists an undeniable causal relation between the fixed funds in industry and the volume of industrial output, as well as between the cumulated investments and the increase of fixed funds.

The plan of the Centre of Economic Computation and Economic Cybernetics for scientific research in the period 1966—1970 includes the problem "Analysis of the degrees of play of the 5-year plan and determination of balance equations for the plan". The solving of this problem has a double aim:

a) To limit the number of exogenous variables of the adopted model to the number strictly necessary and sufficient to determine the whole plan, thus avoiding situations of over- or under-determination, respectively of non-determination or of incompatibility consequently to too large or too small a number of data a priori accepted as initial plan hypotheses*;

b) To find certain relations in the form of so-called balance equations of the plan — or of systems of such equations, in which case a model is obtained — by means of which it should be possible, from the very beginning, to determine the extent to which the values attributed to the exogenous variables are compatible. This is considered necessary in view of avoiding situations where the non-correlation, respectively the unbalance of the plan, only stands out after its detailing, hence after considerable useless work expenditure.

The problem has been divided into several themes whose solving is planned over the entire period 1966—1970. The theme planned for 1966—1967 is "Determination of a balance equation for the plan", this paper representing a component part of this theme.

A first balance equation suggested is the equation resulting when equalling the volume of investments in industry which, on the one hand, proceeds from the requirements arising from the planned increase of industrial output

* An initial plan hypothesis is considered as being determined by the set of values a priori accepted and attributed to the exogeneous variables of the adopted plan model.

to the one which, on the other hand, proceeds from the possibilities of allotting the required amounts out of the national income. These possibilities are considered contingent upon a series of so-called general indicators of economic policy (decision variables), such as: the ratio between the accumulation and the consumption funds, the weight of investments allotted to industry out of total investments for the economy, etc., and the indicators belonging to the set of exogenous variables [1]. Consequently, a first sub-theme appears: "Correlation ratio between investments in industry and the level of industrial output." This correlation ratio based on the results formerly obtained will allow the extrapolation with maximum probability of future realization of the couple: investments in industry — industrial output. A problem of economic prognosis is thus reached.

There results from the specialty literature that at the present stage of knowledge and possibilities only the method of correlation allows a relation of economic prognosis to be determined in the conditions where:

— a sufficiently long and reliable set of statistical data is available;

— a causal, a priori, dependence exists between the result variable and the factorial variable (or variables);

— the results of the extrapolation are plausible.

The first condition may be considered as satisfied by the existence of official statistical records for a period of 16 years (1950—1965).

In view of satisfying the second condition, the use of a causal chain was resorted to, introducing the value of fixed funds in function, in industry, as an intermediate factorial variable. Thus, the imperative necessity arises of determining the following two correlation ratios:

$$f_t = \varphi(p_t) \tag{1}$$

$$\sum_t i_t = \Phi(f_t - A) \tag{2}$$

finally resulting in

$$i_t = \frac{d\sum_t i_t}{dt} = \frac{d\Phi_1[\varphi(p_t)]}{dt} = F(p_t, t) \tag{3}$$

where f_t, p_t, and i_t represent relative values for a year whatever t in comparison with a basic year ($t = 0$), viz.:

f_t = relative value of fixed funds in function in industry;
p_t = relative value of total industrial output;
i_t = relative value of volume of investments in industry.

Relation (3) represents the prognosis function sought. The reasons for which the above procedure was chosen are the following:

a) there certainly exists a causal relation between the fixed funds in function in industry and the volume of industrial output. Likewise, there exists an undeniable causal relation between cumulated investments $(\sum_t i_t)$ and the increase of the fixed funds $(f_t - 1)$, since the increase of the fixed funds occurs as a consequence of the investments achieved;

b) this causal chain being used, one finally obtains an extrapolation function in relation (3), which can faithfully reflect the dependence between

the investments required in industry and the obtaining of a certain level for total industrial output.

The effect caused by the introduction of additional factorial variables and particularly of the time (t), the only factorial variable that can be safely forecast and that can have quite a large influence, will be analyzed by means of these functions. A multiple correlation ratio is thus obtained, which takes into account several factors having a preponderant influence on the result variable, respectively on the investigated phenomenon.

A first problem within the above sub-theme is therefore to find a suitable correlation ratio between the value of the fixed funds in function in industry and global industrial output (1).

In view of finding a suitable correlation ratio between the value of the fixed funds in function in industry and total industrial output, four forms of functions were analyzed and were subjected to several conditions, a relatively large number of correlation ratios resulting therefrom, as will be shown hereunder (by testing, one of them was chosen). Both simple correlation and multiple correlation were resorted to. In the latter case, besides the main factorial variable — total industrial output — other factorial variables were also used:

— the time (t), taking into account the fact that the factor time can play an important part; total industrial output, by means of fixed funds in function, of the same value, may vary in time as a result of technical progress;

— labour productivity (m), taking into account the fact that, for higher labour productivity a larger output can be obtained with the same fixed funds.

The statistical data used are:

Year	Level of		
	Total industrial output	Fixed funds in industry	Labour productivity
1950	100.0	100.0	100.0
1951	123.7	109.3	115
1952	145.7	118.3	130
1953	167.5	129.7	140
1954	178.1	138.0	143
1955	202.5	152.5	160
1956	224.6	167.0	172
1957	243.2	181.2	187
1958	266.5	192.5	198
1959	293.4	205.6	213
1960	340.3	225.1	237
1961	393.0	242.7	253
1962	447.2	263.0	270
1963	502.2	299.1	291
1964	574.0	326.7	320
1965	648.6	357.0	343

These data were used to compute the coefficients, respectively the correlation ratios as well as the determining coefficients for the factorial variables taken into consideration, the following results being obtained:

— For the simple correlation: $f_t = F(p_t)$:

	linear form	non-linear form
correlation coefficient	0.98	—
correlation ratio	—	0.996
determining coefficient	0.96	0.992

— For the double correlation: $f_t = F(p_t, t)$;
partial correlation (selection) coefficients:

$$r_{fp \cdot t} = 0.245; \qquad r_{ft \cdot p} = 0.701; \qquad r_{tp \cdot f} = 0.518$$

multiple correlation coefficients:

$$r_{f \cdot pt} = 0.998; \qquad r_{p \cdot ft} = 0.997; \qquad r_{t \cdot fp} = 0.998$$

— For the triple correlation: $f_t = F(p_t, m_t, ,t)$: partial correlation coefficients:

$$r_{fp \cdot mt} = -0.015; \qquad r_{fm \cdot pt} = 0.575; \qquad r_{ft \cdot pm} = 0.2763$$
$$r_{pm \cdot ft} = -0.1934; \qquad r_{pt \cdot fm} = 0.683; \qquad r_{mt \cdot fp} = 0.2674$$
$$r_{fp \cdot tm} = -0.017; \qquad r_{fm \cdot tp} = 0.576; \qquad r_{ft \cdot mp} = 0.2761$$
$$r_{mp \cdot tf} = 0.183; \qquad r_{pt \cdot mf} = 0.6827; \qquad r_{mt \cdot pf} = 0.267$$

multiple correlation coefficients:

$$r_{f \cdot pmt} = 0.9487; \qquad r_{p \cdot fmt} = 0.9608$$
$$r_{m \cdot fpt} = 0.9733; \qquad r_{t \cdot fpm} = 0.9921$$

There results from the above data that there is a sufficiently close correlation between the resulting variable and the factorial variables taken into consideration. With respect to the determining coefficients there results that in the case of double correlation the factorial variables p and t determine the value of the fixed funds in a proportion of 99 per cent, the share of industrial production being 26 per cent, and that of time 73 per cent. The large value of the determining coefficient for time could be explained by the fact that total industrial output is also, in turn, a time function, and 23 per cent only reflects its influence independently of the time factor.

For the triple correlation a total determining of only 85 per cent was reached, the influence of total industrial output being 74 per cent, that of labour productivity 7 per cent and that of time 4 per cent. It seems that the additional data introduced here for labour productivity have no common basis of computation with those referring to total industrial output this being reflected in the fact that, although from the logical standpoint the determining coefficient should improve, it actually becomes worse. As it will also be seen further, the triple correlation also yields unfavourable results when tested.

With respect to the correlation function the following forms were analyzed:

$$f_t = a_1 + b_1 p_t \tag{4}$$

$$f_t = a_2 p_t^{b_2} \tag{5}$$

$$f_t = a_3 p_t^{b_3} e^{r_3 t} \tag{6}$$

$$f_t = a_4 p_t^{b_4} m_t^{c_4} e^{r_4 t} \tag{7}$$

The constants a, b, c and r were determined by the least squares method. The functions (5), (6), and (7) were linearized by logarithms, and the functions (6) and (7) by derivation. Likewise, for the functions (5), (6) and (7) the constants were also computed for the following supplementary conditions:

1) the curve must necessarily pass through the point corresponding to the basic year ($a = 1$);

2) the curve must necessarily pass through the point corresponding to the last year of the known period taken into consideration.

The latter condition was set particularly in view of finding out if, taking into account the fact that the future is more powerfully influenced by the recent past than by the more remote, a function more adequate for extrapolation is not obtainable.

In view of choosing the most suitable correlation ratio, the only possible text was resorted to, admitting that, from the set of available data, the 1950—1960 period should be considered known and the constants a, b, c and r be computed by means of these data and, that, by continuing in 1961—1965 the curves thus resulting, the achievements of these latest years be checked.

Based on the above, 14 correlation functionsl resulted. They were computed with the data in the 1950—1960 period, the folowing relations resulting:

— For the function of the form (4):

$$f_t = 0.613 + 0.457\, p_t \tag{8}$$

— For the function of the form (5), linearized by logarithms

$$f_t = 0.9406\, p_t^{0.7038} \tag{9}$$

with the above condition 1):

$$f_t^{(1)} = p_t^{0.6388} \tag{10}$$

with the above condition 2):

$$f_t = 0.939\, p_t^{0.714} \tag{11}$$

— For the function of the form (6)

a) linearized by logarithms:

$$f_t = 0.993\, p_t^{0.154} e^{0.0635\, t} \tag{12}$$

$$f_t^{(1)} = p_t^{0.038} e^{0.0697\, t} \tag{13}$$

$$f_t^{(2)} = 0.999\, p_t^{0.121} e^{0.0664\, t} \tag{14}$$

b) linearized by derivation:

$$f_t^{(1)} = p_t^{0.157} e^{0.0537\,t} \tag{15}$$

$$f_t^{(2)} = 2.067\, p_t^{0.0207} e^{0.00596\,t} \tag{16}$$

— For the function of the form (7)

a) linearized by logarithms:

$$f_t = 0.997\, p_t^{0.058} m_t^{0.179} e^{0.6576\,t} \tag{17}$$

$$f_t^{(1)} = p_t^{0.041} m_t^{0.108} e^{0.0675\,t} \tag{18}$$

$$f_t^{(2)} = 0.999\, p_t^{0.115} m_t^{0.012} e^{0.066\,t} \tag{19}$$

b) linearized by derivation:

$$f_t^{(1)} = p_t^{0.332} m_t^{0.0074} e^{0.0054\,t} \tag{20}$$

$$f_t^{(2)} = 2.251\, p_t^{-0.000006} m_t^{0.000036} e^{-0.000006\,t} \tag{21}$$

Function (21) leads to unplausible results, the fixed funds in function in industry being unable to remain practically constant whatever the level of industrial production. For the other 14 functions remaining under further investigation, the mean (ε) and standard (σ) deviations were computed for the period 1950—1960, considered known, as well as for the extrapolation of the curve in the period 1961—1965. The results obtained are shown in the table hereunder.

Function	Period 1950—1960				Extrapolation in 1961—1965			
	Deviation (%)				Deviation (%)			
	maximum	minimum	mean	standard	maximum	minimum	mean	standard
(8)	8.131	0.905	1.232	5.37	−2.75	0.22	+0.63	1.419
(9)	−5.94	0.024	0.038	2.39	3.56	−0.47	0.74	2.01
(10)	7.49	0.0	1.02	4.22	−7.53	−1.015	−4.62	5.34
(11)	−6.11	0.0	0.138	2.937	4.004	0.089	1.234	2.216
(12)	−1.931	0.057	0.175	1.02	−3.738	1.654	−1.19	2.75
(13)	−1.827	0.0	0.04	0.97	−4.204	1.50	−1.50	2.98
(14)	−4.364	0.0	−0.397	1.657	−5.246	1.013	−2.142	3.455
(15)	10.09	−4.0	0.1597	4.801	42.0	17.42	28.40	29.06
(16)	89.3	0.04	14.286	17.55	−34.12	−6.43	37.01	37.30
(17)	1.522	−0.164	0.26	1.012	−4.353	1.643	−1.56	3.06
(18)	−1.76	0.0	0.04	0.97	−4.88	1.197	2.02	3.354
(19)	−2.257	0.0	−0.087	1.031	−5.075	0.818	−2.29	3.486
(20)	−89.14	−1.19	36.71	47.30	−49.2	−30.6	20.9	23.2

Functions (15), (16) and (20), obtained by using the derivation method for their linearization, must be eliminated since in the form thus obtained their deviations are unacceptable.

There has resulted from a first analysis that by linearization and derivation, and particularly by passing from differentials to finite differences, supplementary errors are introduced, which may act in a same sense and this

would finally lead to abnormally large deviations. Thus, the fact that for function (20) all the deviations are of the same sign confirms this hypothesis. It seems that in similar cases a (linear, parabolic or exponential) adjustment of the function could be resorted to, thus leading, at least formally, to very favourable results. The further investigation of these functions will be effected in a separate study.

Of the 10 functions remaining under discussion we notice that function (9) satisfies to the greatest extent the conditions set for acceptance, respectively it may be considered as fulfilling to the highest degree the adopted criteria of choice, viz.:

— mean and standard deviations are smaller than the maximum values considered acceptable ($\varepsilon < 1\%$; $\tau < 3\%$), both in the period considered known and in the extrapolated period;

— the scattering of the points corresponding to realizations in the period 1961—1965 (2.01; 2.39) is reduced in comparison with the extrapolated curve.

Consequently, in the present stage of the work, a function of the form (5), without any supplementary condition, is chosen as being the most suitable correlation ratio between the fixed funds in function and total industrial output. This form represents a simple correlation.

It is noteworthy mentioning that neither the bringing of several factorial variables into the computation, nor the setting of supplementary conditions (particularly the second supplementary condition shown above) have led to more suitable correlation functions, as might have been expected. This could be explained, on the one hand, by the fact that the factorial variables p and m are not strictly independent, and both may, in their turn, be considered functions of time. However, when resorting to a multiple correlation, it is known that it is generally advisable for the factorial variables to be independent from one another. On the other hand, as already shown before, statistic recording of labour productivity does not seem to permit its connection with total industrial output, since they have no strictly common basis.

The form of the prognosis function once established, the recomputation of the constants a_2 and b_2 was undertaken, the entire set of available statistical data (1950—1965) being used.

The following results were obtained:

$$f_t = 0.94\, p_t^{0.7075} \tag{22}$$

maximum deviation: — 5.935% for 1950
minimum deviation: — 0.009% for 1951
mean deviation: 0.1424 1%
standard deviation: 2.595 3%

with which the work for the sub-theme discussed will be continued.

Mention is made of the fact that this study has required a great volume of computations and that for some of the more voluminous works, such as the reversal of some 4th order matrices, the CIFA Economist 102 electronic computer at the Centre of Economic Computation and Economic Cybernetics was also used.

REFERENCES

1. SAPHIER, I., *Posibilităţi de folosire în planificare a unor relaţii de corelare dintre unii indicatori economici sintetici*, "Probleme economice", No. 9, 1963.
2. AITKEN, A.C., *Statistical mathematics*, Interescience Publisher, New York, 1947.
3. CHORAFAS, D.N., *La simulation mathémathique et ses applications*, Dunod, Paris, 1966.
4. GERFIN, H., *Langfristige Wirtschaftsprognose*, J.C.B. Mohr, Polygr. Verlag, Tübingen-Zürich, 1964.
5. MIHOC, G. and FIRESCU, E., *Statistica matematică*, Editura Didactică şi Pedagogică, 1966.
6. RANCU, N. and TÖVISSI, L., *Statistica matematică cu aplicaţii în producţie*, Editura Academiei R.P.R. 1963.
7. WORTHING, A. G. and GEFFNER, I., *Prelucrarea datelor experimentale*, Editura Tehnică, 1959.

FONCTIONS DE CORRELATION ENTRE LES FONDS FIXES ET LA PRODUCTION INDUSTRIELLE

Cet article étudie la corrélation entre les fonds fixes en fonction dans l'industrie et la production globale industrielle. On analyse 4 formes de fonction, à savoir:

$$f = a + bp \tag{1}$$

$$f = ap^b \tag{2}$$

$$f = ap^b m^c e^{rt} \tag{3}$$

$$f = ap^b e^{rt} \tag{4}$$

où:

f est le niveau des fonds fixes dans l'industrie;

p est le niveau de la production industrielle globale;

m est le niveau de la productivité du travail dans l'industrie;

t est le temps;

a, b, c sont des constantes.

La linéarisation a été effectuée par les logarithmes, et pour les formes (3) et (4), également par dérivation. Les constantes ont été déterminées par la méthode des moindres carrés. En outre, pour les formes (2), (3) et (4) on a calculé les constantes en posant la condition que la courbe passe obligatoirement par le point correspondant à l'année de base, ainsi que la condition que la courbe passe obligatoirement par le point correspondant à la dernière année de la période considérée connue. On a, de cette façon, obtenu 14 fonctions de corrélation qui ont été testées.

Le test a consisté à considérer la période connue 1950—1960 et on a calculé les déviations moyennes et standard pour les données corespondant aux années 1961—1965, choisissant ainsi la relation la plus convenable pour l'extrapolation.

Après avoir choisi la fonction appropriée, ses constantes ont été recalculées en employant cette fois-ci toutes les données disponibles, respectivement la période 1950—1965 tout entière. On a ainsi établi la fonction de corrélation qui sera employée comme première équation d'équilibre du plan.

ФУНКЦИИ КОРРЕЛЯЦИЙ МЕЖДУ ОСНОВНЫМИ ФОНДАМИ И ПРОМЫШЛЕННЫМ ПРОИЗВОДСТВОМ

Представляемая работа изучает корреляцию между действующими основными фондами в промышленности и валового промышленного производтсва. Были анализированы и формы функций, а именно:

$$f = a + bp \tag{1}$$

$$f = ap^b \tag{2}$$

$$f = ap^b m^c e^{rt} \tag{3}$$

$$f = ap^b e^{rt} \tag{4}$$

где:

f — уровень основных фондов в промышленности;

p — уровень валового промышленного производства;

m — уровень производительности труда в промышленности;

t — время;

a,b,c являются постоянными.

Линиация была произведена по логаритму, а для форм (3) и (4) и по деривации. Постоянные были определены по методу наименьших квадратов. А также, для форм (2), (3) и (4) были подсчитаны постоянные с условием, что кривая пройдет в обязательном порядке через соответствующую точку последнего года периода, считающегося известным. Таким образом были получены 14 функций корреляции и были тестированы.

Тест был составлен с учетом известного периода 1950—1960 гг.; были подсчитаны стандартные и средние отклонения для данных, соответствующих 1961—1965 годам, и было выбрано самое подходящее отношение для экстраполяции.

После выбора подходящей функции, были пересчитаны ее постоянные, причем на этот раз были использованы все имеющиеся данные, а именно весь период 1950—1965 гг. Таким образом была установлена функция корреляции, которая будет использована как первое уравнение равновесия плана.