S. COSTIN
S. BORDEANU

Econometric Methods in Predicting Electric Power Consumption

The process of accelerated development of national economy is unconceivable without meeting the ever increasing power requirements. Concomitantly, facts confirm the constant tendency to an ever increasing weight of electric power consumption within the general structure of power requirements.

INTRODUCTION

From the economic standpoint, the branch of electric power production is characterized by specific high costs and by a relatively long cycle for putting new installations in operation. Both factors determine important freezing of investment funds in this branch. The curtailment of the freezing of funds is closely related to the forecasting of electric power requirements, that is to the estimation of the way in which the demands for electric power in all the branches of national economy will develop.

The negative economic effect which overestimation of the requirements in power to be installed may have is made clear if one notes that in 1966, a year in which new power installed in the power system exceeded 1,200 MW, the effort made in investments almost reached 4,000,000,000 lei.

Concomitantly, the underestimation of the actual requirements of users has disturbing negative effects on the development of national economy as a whole, effects whose importance is stressed by the fact that at the instant that the underestimation is noticed, no efficient measures can practically be taken, in view of straightening up the situation (since the cycle of electric

power installation is longer than the other cycles required for the installation of industrial capacities). All this spotlights the importance of elaborating predictions for electric power consumption, the econometric methods proving their usefulness in this domain too.

METHODS USED IN PREDICTING ELECTRIC POWER CONSUMPTION

The determination of electric power consumption for a country or for a sufficiently extensive geographic zone in the respective country is dependent on several factors, both of an economic nature — present and future development of production factors — and of a geographic, demographic, social, etc. nature.

The development of electric power consumption is a random process based on particularly complex connections among various determined factors. Consequently, the preliminary problem of this development could not be treated on causal bases and its solution was directed along probabilistic paths of various degrees of complexity.

Among the paths used all over the world to solve the problem of predicting electric power consumption, the following directions can be noticed:

1° Analytic — by determining consumption development for various classes of consumption and summing them up;

2° Synthetic — relating the development of electric power consumption to the evolution of other parameters considered as being determining factors.

Although the analytic method should apparently lead to more accurate results since it is based on the analysis of the evolution of each class of consumption, it can practically only respond for predictions dealing with relatively short terms for which there exist data in connection with the evolution of the various consumption classes. Outside these boundaries the determination of consumption by classes can only be carried out by synthetic methods, so that the use of the global methods (which no longer require division) becomes at least as interesting.

Solving the problem can be carried out, in principle, in two ways:

— empirically: establishing, by estimation, relations corresponding to the situations already achieved and which could be projected into the future;

— using probabilistic statistics: by means of the correlation and regression methods, adjusted according to the methods generally known in mathematical statistics.

The a priori advantage of the second way is that it offers the reliability of prediction to be estimated.

In Romania, the endeavours to determine fundamental relations for predicting electric power consumption have practically covered the entire domain of possibilities mentioned above, both with regard to empiric predictions [5] and to predictions by extrapolation or correlation with economic indices [7], [8].

From the studies elaborated more recently in Romania and in other countries, a tendency can be noticed towards the ever wider use, in predictions, of the methods of the correlation theory. The grounds for this trend is the existence of causal relations between the predicted phenomena and the characteristic indicators of economic life. Included in the same trend, this paper presents some elements which may contribute to elucidate the correlative relations existing between electric power consumption, on the one hand, and some characteristic indicators of economy, on the other hand.

In all the studies elaborated, the resultant variable has been considered as being the total gross electric power consuption in a given country, whether it be Romania or other countries. This, since:

a) the estimation of total consumption is the subject of research whose aim is to determine the requirement for new production capacities;

b) by estimating consumption by sections the situation is reached where various classes of consumption can be predicted with a different degree of accuracy, the prediction of some consumption classes being particularly uncertain;

c) the estimation of electric power production required by export is not subjected to the laws of random phenomena since export contracts are included in the class of decision phenomena.

In this connection, we should mention the difficulties which have sprung up on the occasion of the works carried out to make the data proceeding from various sources agree, owing to the absence on an international scale of a unitary manner of treatment and presentation of the statistic data.

CORRELATION BETWEEN ELECTRIC POWER CONSUMPTION AND GROSS INDUSTRIAL OUTPUT

The index of gross industrial output is the synthetic index which practically characterizes the level of social production in the sphere of commodities. Although the negative sides and the determining imperfections for this index are known, it nevertheless remains the most comprehensive and characteristic index for defining the dynamic conditions of evolution of industrial production. We believe it is obvious, even without further proof, that the variations of the index of industrial production influence directly the consumption of electric power in industry. Concomitantly, the index of industrial output influences the entire economic and social life, so that the hypothesis that the total consumption of electric power depends to a great extent on the index of industrial output appears to be justified.

The above have been the main arguments which guided our studies, in the first instance towards the relation between the variation in electric power consumption and the index of industrial output. Investigation has been carried out by several means, the most promising of which have proved to be: a) the correlation between the mean yearly rates of the factors considered; b) the correlation between the absolute values (or the relative values, compared to a fixed basis) of the same factors.

CORRELATION BETWEEN MEAN YEARLY RATES OF GROWTH

The computation of the correlation between the mean yearly rates of growth of total electric power consumption and of gross industrial output has been carried out on the basis of the data obtained for an as large as possible number of countries, both socialist and capitalist. The data used cover a period of about 18 years, thus taking into account the possibility for both the value of correlation and that of the regression coefficients to be variable within longer time periods. Likewise, the respective values have not been taken into consideration for countries whose level of industrialization is very low, even if their economy presents a rapid rate of increase, since the large variations in the data of those countries from one year to the next may bear a negative influence on the determination of the general correlation.

The data gathered and retained for use have first been treated in view of removing some insignificant leaps and deviations, by computing the mean yearly rate of increase for every indicator with a five year span.

The correlation has been computed in agreement with data extracted from the statistics of 23 countries among which 9 socialist countries. These countries are: Albania, Austria, Belgium, Brazil, Bulgaria, Canada, Czechoslovakia, Denmark, France, German Democratic Republic, Federal Republic of Germany, Great Britain, Greece, Hungary, Italy, Japan, Norway, Poland, Romania, Sweden, U.S.A., U.S.S.R. and Yugoslavia.

The computations were carried out in 4 variants:

— for all the data retained (103)

— for the data pertaining to the socialist countries (48)

— for the data pertaining to the capitalist countries (55)

— for the data pertaining to Romania (12).

The number in brackets indicates the pairs of data taken into consideration for each correlation.

Table I shows the results of the computations carried out by the regression method, according to the known relations and applying the least squares method.

The form of the linear regression equation is:

$$Y = a + bx$$

where:

Y — mean yearly rate of growth of electric power consumption;

x — mean yearly rate of growth of gross industrial output.

Table I

Group of countries	r	a	b
All the data retained	0.818	4.341	0.643
Data from socialist countries	0.649	4.765	0.624
Data from capitalist countries	0.681	4.855	0.526
Data from Romania	0.590	5.560	0.704

In table I r is understood to be the correlation coefficient

$$r = \frac{\Sigma(x - \bar{x})\,(y - \bar{y})^*}{\sqrt{\Sigma(x - \bar{x})^2 \Sigma(y - \bar{y})^2}}.$$

Before subjecting the results obtained to a basic analysis it has been considered necessary to verify the significance of the regression coeficient b and of the correlation coefficient r as well as to determine the reliable boundaries corresponding to a pre-determined threshold of significance for the values acquired by these coefficients. The domain of reliability of the free term a in the regression equation has been determined as well.

The significance of the regression coefficient b has been verified by means of the criterion t^{**} defined by the relation:

$$t = \frac{b}{S_b} > t_q$$

where S_b is the selection error of the regression coefficient and is determined from the expression

$$S_b = \sqrt{\frac{\sum_{i=1}^{N} (y_i - Y_i)^2}{(N-2)\sum_{i=1}^{N} (x_i - \bar{x})^2}}$$

where

$$\sum_{i=1}^{N} (y_i - Y_i)^2 = \sum_{i=1}^{N} y_i^2 - N\bar{y}^2 - b^2 \sum_{i=1}^{N} (x_i - \bar{x})^2$$

and

$$\sum_{i=1}^{N} (x_i - \bar{x})^2 = \sum_{i=1}^{N} x_i^2 - \frac{\left(\sum_{i=1}^{N} x_i\right)^2}{N}$$

and t_q is the table value of Student's function t with (chosen) significance threshold q; in our case this significance threshold is 0.05.

---

* For simplicity of computation, the correlation coefficient is computed according to a transform of the above formula [2], viz.

$$r = \frac{n\Sigma xy - (\Sigma x)\,(\Sigma y)}{\sqrt{[n\Sigma x^2 - (\Sigma x)^2]\,[n\Sigma y^2 - (\Sigma y)^2]}}.$$

** b is considered significant when the value of t computed in agreement with the investigation data is larger than the table value corresponding to the significance level q, i.e. to the probability $p = 1 - \frac{q}{2}$ and the number of degrees of freedom, $F = N - 2$[4].

Table II shows the results of the verifying thus carried out.

Table II

Group of countries	Regression coefficient b	S_b	$t = \frac{b}{S_b}$	Freedom degrees $F = N-2$	$t_{0.05}$	Reliable boundaries for β
All countries	0.643	0.0458	14.05	101	1.98	$0.553 \div 0.733$
Socialist countries	0.624	0.1058	5.9	46	2.015	$0.416 \div 0.831$
Capitalist countries	0.526	0.0757	6.95	53	2.00	$0.378 \div 0.674$
Romania	0.704	0.304	2.316	10	2.23	—

There results from this table that the regression coefficient for the 4-th variant is not significant ($t \approx t_q$), so that — after this test — it is necessary to give up this correlation.

The significance of the correlation coefficient has been verified for large volume selections ($N > 30$) according to the criterion t. In fact, the criterion determined whether the theoretic value of the correlation coefficient is significantly different from zero, or not. If the hypothesis that the theoretic value of the correlation coefficient is zero is invalidated by computation*, r is considered approximately equal to ρ, and the range of reliability is computed from

$$S_r = \frac{1 - r^2}{\sqrt{N - 1}}$$

with the boundaries

$$r \pm t_p S_r.$$

For a reduced volume selection Fischer's "transformed correlation method" is used [4], introducing the variable

$$Z = \operatorname{arg\,th} r = \frac{1}{2} \log_e \frac{1 + r}{1 - r},$$

having an almost normal distribution, with the standard deviation

$$S_z = \frac{1}{\sqrt{N - 3}}.$$

The significance condition for r is

$$\frac{Z}{S_z} > 1$$

and the boundaries corresponding to a given probability P are

$$Z \pm t_p S_z.$$

* If $t = \frac{r}{S'_r} > 1$, where $S'_r = \frac{1}{\sqrt{N - 1}}$.

Table III shows the results of the verifying computation for the significance of the correlation coefficient in the four variants studied.

Table III

Group of countries	r	S'_r	$\frac{r}{S'_r}$	S_r	f	t_p	Reliable boundaries for ρ
All countries	0.818	0.099	8.26	0.033	102	2.364	0.740÷0.896
Socialist countries	0.649	0.146	4.45	0.085	47	2.409	0.444÷0.854
Capitalist countries	0.681	0.136	5.00	0.073	54	2.398	0.506÷0.856

Since the selection for Romania is of reduced volume, the computations have been carried out by the transformed correlation method. In agreement with the relations mentioned it has been established that for $r = 0.590$, the transform Z has the value $Z = 0.678$, and the confidential range of the variable ρ is

$$0.025 \leqslant \rho \leqslant 0.869$$

If, according to [1] the correlation for Z is applied:

$$Z' = Z + \Delta Z = \frac{1}{2} \log_e \frac{1+r}{1-r} + \frac{r}{2(N-1)}$$

a new range is obtained after the necessary transformations:

$$0.052 \leqslant \rho \leqslant 0.876$$

The determination of the reliable range for the free term α of the regression equation, computed on basis of the values of a resulting from the various selections studied, was carried out taking into account the fact that the variable

$$t = \frac{a-\alpha}{S_a}$$

follows Student's law with $K = N - 2$ degrees of freedom. Consequently, the reliable range sought is

$$a - tS_a < \alpha < a + tS_a$$

where

$$S_a = \sqrt{\frac{\sum_{i=1}^{N} (y_i - Y_i)^2}{(N-2) \cdot N}}$$

Y_i — value corresponding to x_i in the regression equation.

Table IV shows the results of computations thus made:

Table IV

Group of countries	a	S_a	N	$t_{0.05}$	Reliable boundaries for α
All countries	4.341	0.22	103	1.98	3.906—4.776
Socialist countries	4.765	0.397	48	2.015	3.965—5.565
Capitalist countries	4.855	0.2034	55	2.00	4.448—5.262
Romania	5.56	0.49	12	2.23	4.47 —6.65

There results from the analyses made and shown above, that, according to significance criteria, three of the four correlations analyzed may be accepted for use. The fourth correlation for Romania must be rejected since the significance criterion for the regression coefficient b does not confirm the significance of this magnitude.

Further analyzing the results obtained and the concrete possibilities offered by each of the correlations analyzed, the following is found:

— The correlation for Romania is based on a small volume of data referring to a short period. The volume of data can explain the results obtained practically from computation, which do not confirm the possibility of using the respective regression. However, the fact that the data refer to a short period makes, at any rate, doubtful, the possibility of using regression for predictions, since it is well known that this extrapolation is based on the hypothesis of the invariance of the mean general conditions, an invariance which cannot be maintained when passing from one short period to the next.

— The regression computed from capitalist countries is based on data extracted for industrially developed countries. Consequently, the use of this regression could become timely in the case where predictions are made for more distant time periods where it could be maintained that the general economic conditions become similar.

— The regression computed for the socialist countries is based on data collected from countries having different stages of economic development, from levels of intense industrialization to levels of industrially advanced countries. Taking this into consideration, the use of the regression obtained seems justified.

— The regression computed on the basis of a combination of all the available data (only the countries having a degree of industrialization below that of our country have been excluded from all computations since their data would not supply information useful for predictions in Romania) therefore presents a combination of data proceeding from quite varied economic conditions whose average could correspond to the conditions liable to appear in Romania in the near or not too distant future. Concomitantly, by combining a relatively important number of data the possibility is removed of letting the peculiarities of a certain economy put its stamp on the results.

There follows from the above that for common predictions two regressions could be used: the one for socialist countries and the one for all the data used. A comparison of these two possibilities shows relatively important

advantages in favour of the regression based on all the data, which presents a superior correlation coefficient, and the data of regression with noticeably more restricted ranges of reliability. Consequently, the use of regression computed for all the available data is recommended for predictions.

The practical use of the regression equation mentioned, whose mean value is of the form:

$$Y = 4.341 + 0.643x$$

requires some factors specific to the branch of electric power to be taken into consideration, which determine certain changes in the manner this equation is applied, as compared to the generally known manners. Thus:

a) The cycle for putting into operation installations to produce electric power is longer than the cycles for putting into operation industrial capacities so that, at the instant when practically the necessity is ascertained to install additional electric power in order to satisfy new consumption in course of instalment, it is already too late to be able to take such measures efficiently.

b) The shortage of installed power in the system leading to the impossibility of satisfying the simultaneous demands of users is a serious shortcoming, which can be alleviated by influencing the users, technology (shifting of some consumption to hours outside the peak).

The lack of electric power is, however, a shortcoming that can no longer be alleviated, the users being obliged to restrict their consumption requirements and thus — implicitly, to restrict their production. It is evident that this situation is, economically, much more serious than the lack of power.

The risk of predicting an electric power consumption lower than the one which will actually occur is — as shown above — very serious. Consequently, it is considered necessary to compute predictions with as high as possible a safety margin. This leads to take into consideration, in the regression equation, the upper boundaries of the reliable ranges for the coefficients of the given probability. Therefore, the equation for practical use acquires the form

$$Y = a + tS_a + (b + tS_b)\, x$$

and in numerical form, as it results from Tables II and IV:

$$Y = 4.776 + 0.733 \text{ x}.$$

CORRELATION BETWEEN RELATIVE VALUES
(reported to the same basic year)

When computing the correlation between the absolute values of total gross electric power consumption and of the industrial output index the same data were used as a starting point as the ones shown above. The difference between the two ways of processing only consists in the fact that here, instead of the mean yearly variation rates, the relative values of the magnitudes mentioned were adopted, as they result when reporting the yearly achieve-

ments of the two factors to the same fixed basis: the value reached by each of the two magnitudes in 1958 *.

In this case, the number of pairs of data retained for computing the correlation was higher than the one in the previous case since there was no intermediary processing (computation of mean rates).

The computations were carried out in the same four variants as above, each variant being based on the following number of pairs of data: for all countries — 197 pairs of data; for socialist countries — 84; for capitalist countries — 113; for Romania — 17.

The computations were carried out in agreement with the relations shown in the preceding chapter. The complete results of these computations are shown in Tables V and VI.

The analysis of the results obtained in the four variants studied for the correlation between the relative values (with fixed basis) of gross industrial output and of the total electric power consumption makes a few main elements stand out:

a) The correlation established only for Romania, and hence significant as such, shows an important discrepancy in the possible values of the correlation coefficient, thus requiring this correlation to be treated whith caution.

b) The other three correlations studied show high values of the correlation coefficient, thus making them advisable for use. It is nevertheless mentioned that, in this case too, the general-economic considerations made above when analyzing the correlation of increase rates are valid. Thus, among the three correlations the one for capitalist countries, based on data referring to countries whose industrial development is considerably different from that of Romania, seems more difficult to apply.

The considerations that are specific to the electric power industry, mentioned in the preceding paragraph, referring to the necessity of choosing a high margin of reliability for predictions also remain valid in the case of correlations based on relative values (with a fixed basis).

Table V

Group of countries	Correlation coefficient				Coefficient a				Regression coefficient b			
	r	$t=\frac{r}{S'_r}$	S_r	$f=N-1$	t_p	a	S_a	$F=N-2$	t_q	b	S_b	$t=\frac{b}{S_b}$
All countries	0.972	13.6	0.0039	196	2.345	— 6.00	0.604	195	1.96	1.055	0.0177	59.6
Socialist countries	0.988	9.0	0.00262	83	2.372	—12.61	0.820	82	1.99	1.128	0.0187	60.3
Capitalist countries	0.940	9.95	0.011	112	2.363	+12.11	0.749	111	1.98	0.873	0.0306	28.5
Romania	0.993	3.97	0.0558	16	2.583	—20.82	2.464	15	2.13	1.182	0.0355	33.3

* The year 1958 was chosen as a basic year for computation since the data known for the gross industrial output in most countries are subsequent to this date. It should be mentioned that this did not prevent data preceding 1958 from being used. These data were likewise reported to 1958 considered as 100.

ЭКОНОМЕТРИЧЕСКИЕ МЕТОДЫ В ПРОГНОЗЕ ПОТРЕБЛЕНИЯ ЭЛЕКТРИЧЕСКОЙ ЭНЕРГИИ

Предыдущие исследования показали, что эффективный метод предварительного определения потребления электрической энергии основывается на его соотношении с ростом валовой промышленной продукции. С целью тщательности прогнозов на средние периоды (до 5—7 лет) были исследованы связи между ростом потребления электрической энергии и объемом вложений государства, а также влияние, которое может иметь в условиях нашей страны многообразие демографических показателей.

Проведенные анализы показали, что существуют определенные соотносительные статистические связи между ростом потребления электрической энергии и объемом вложений в предыдущие периоды. Расчеты простой и увеличенной регрессии показывают соотношение, дающее самые благоприятные результаты: простое соотношение с вложениями, произведенными с разрывом в 3 года. Одновременно подчеркивается, что в демографических условиях, характерных для нашей страны в период 1955 — 1965 годов (единственный период, дающий необходимые статистические данные) не может быть определено какое-нибудь соотношение влияния демографического фактора на рост потребления электрической энергии в Румынии.