EESTI NSV TEADUSTE AKADEEMIA TOIMETISED. 30. KÖIDE
ÜHISKONNATEADUSED. 1981, NR. 1

ИЗВЕСТИЯ АКАДЕМИИ НАУК ЭСТОНСКОЙ ССР. ТОМ 30
ОБЩЕСТВЕННЫЕ НАУКИ. 1981, № 1

https://doi.org/10.3176/hum.soc.sci.1981.1.03

M. MEEL

ON A MEAN-VARIANCE MODEL FOR OPTIMAL PLANNING OF THE EDUCATIONAL LEVEL OF A REGION'S POPULATION

Presented by K. Habicht

First, the paper presents a linear dynamic input-output model for optimal planning of the educational level of a region's population. In principle, the model is an optimal planning problem determined on the mean values of random variables. However, probable deviation of some parameters from their mean values may prove to be significant. For this reason, the model is complemented with the dispersions (variances) of these parameters. Though more complicated to solve, the mean-variance model approximates reality better than the one compiled with the help of mean values alone. When complementing the problem it is attempted not to turn it too complicated for a practical solution.

1. A model determined on the mean values of parameters

Let us first present a linear dynamic input-output model for the optimal planning of the educational level of the population of a region in the form worked out by the author earlier [1]:

$$\max x_0=\left\{\sum_{t=1}^{w}\left[u^t\,\frac{\sum\limits_{g=1}^{v}k_g x_g^t}{N^t}:\frac{\sum\limits_{g=1}^{v}k_g x_g^0}{N^0}+\right.\right.$$

$$\left.\left.+(1-u^t)\sum_{g=1}^{v}c_g^t\,\frac{x_g^t}{v_g^t}\right]\right\}:w \qquad (1)$$

with the constraints

$$\left.\begin{array}{ll} x_g^{0t}-\sum\limits_{j=1}^{n}r_{gj}^t y_j^t=0, \quad g=1,\ldots,v & (2)\\ \sum\limits_{j=1}^{n}q_{gj}^t y_j^t-x_g^t=0, \quad g=1,\ldots,v & (3)\\ \sum\limits_{j=1}^{n}a_{ij}^t y_j^t\leqslant b_i^t, \quad i=1,\ldots,m & (4)\end{array}\right\}\ t=1,\ldots,w$$

where

a) the indices are denoted by:

$g=1, \ldots, v$ — the index of a population (workers) group according to its educational level;
$i=1, \ldots, m$ — the index of an economic resource;
$j=1, \ldots, n$ — the index of educational activity;
$t=1, \ldots, w$ — the index of an interval (year);

b) the parameters are denoted by:

a^t_{ij} — the input coefficient of the resource i into the activity j during the year t;
b^t_i — the limit of resource i;
r^t_{gj} — the share of the members of group g in the activity j;

$$0 \leqslant r^t_{gj} \leqslant 1 \quad \text{and} \quad \sum_g r^t_{gj} = 1;$$

g^t_{gj} — the output coefficient of the members of the group g from the activity j;

$$0 \leqslant q^t_{gj} \leqslant 1 \quad \text{and} \quad \sum_g q^t_{gj} = 1;$$

v^t_g — the national economic (social) requirement for the group g;
x^{0t}_g — the number of the members of the group g at the beginning of the interval t; $x^{0t}_g = x^{t-1}_g + s^{t-1}_g$, where s^{t-1}_g is the migration balance of the population of the group g;
x^0_g — the number of the representatives of the group g during the base year;
N^t — the region's population number at the end of the interval t; $N^t = \sum_{g=1}^{v} x^t_g$;
N^0 — the region's population number during the base year; $N^0 = \sum_{g=1}^{v} x^0_g$;
c^t_g — weighing coefficient of the importance of the educational level g in the objective function from the standpoint of its national economic necessity;
k_g — the number of grades in the group g (by grade we refer to the stage in the educational system passed, as a rule, during one interval);
u^t — weighing coefficient of the importance of the index of the rise of the population's educational level as compared to the index of the rate of meeting the national economic demand for the population with different levels of education;

c) the variables to be determined are denoted by;

y^t_j — the intensity of the educational activity (number of students);
x^t_g — the number of the representatives of the group g at the end of the interval t.

The above is a model determined on the mean values of the parameters. The advantage of such a model is the simplicity of its solutions: it is linear. If the planning or forecasting period is short, the probable deviation of the values of the parameters from their means is usually small and the error resulting from the use of the means will be negligible. However, in case of a long forecasting period, the fidelity of the forecasts of the parameters is reduced, and the probable error of the solution will increase. Therefore, in case of a long forecasting period the problem should be turned stochastic at least for the parameters whose divergence from the mean values are estimated to be considerably significant. Below the model is analyzed in this aspect.

2. Mean-variance problem

1. Let us supplement the above model so as to get a mean-variance model.

2. First we agree to deal with a fixed-plan problem (x^t_g and y^t_j are determined), i. e., the solution to the optimum problem is a unique decision.

3. Let us first discuss the objective function (1). The total population, number of the representatives of the group g during the base year (x^0_g and N^0) and the number of grades in the group g (k_g) may be considered constant here.

The weighing coefficients c^t_g and u^t are random variables whose values are determined in a subjective way — either a group estimate or the opinion of an individual. As a rule, the estimate is presented as a certain figure considered to be the mean value of the coefficient, or as a deviation interval considered probable by the expert. However, in most cases an expert is incapable of estimating the distribution law and dispersion of these coefficients. Moreover, when setting a concrete problem it is usually not possible to organize a sufficiently large and competent commission of experts whose group estimate would allow us to draw up the distributions and dispersions of these coefficients. Therefore, in the present theoretical formulation of the model it is likewise reasonable to consider these coefficients to be fixed. When solving the problem, these coefficients can be fixed as their supposed mean values and lower and upper deviation intervals. This would yield three different variants of the problem.

The parameter N^t (the region's population number) is a random variable and an object of a demographic forecast. Experience has proved that the component method based on demographic age shift [2] enables to forecast quite a numerous population (as that of a union republic) with considerable accuracy for a rather long period (10—15 years). For that reason, the error resulting from the determination of N^t to its mean value is not considered significant and N^t is treated as a fixed quantity.

The most important parameter of the objective function from the standpoint of the model discussed, yet the one the most difficult to forecast, is the national-economic (social) requirement for the population with different levels of education v^t *. This parameter is affected

* In more exact models, within each educational level speciality groups should be distinguished in addition.

mainly by two groups of factors: the scientific and technological level of the economy and its development rate, and the educational policy of the state. The forecast of the first group of factors should be treated as a random variable since its determination on its mean value would involve a too big simplification. Likewise, the forecast of v_g^t should be treated as a random variable.

Thus, the objective function (1) will also turn out to be random. In case of the objective function determined on mean values, its mean value was maximized. Now, to get a result similar in principle, the probable value of the objective function should be maximized, i. e., its mean value should be maximized and its probable deviation from the mean value should be minimized. The deviation of a random variable α around its mean value is characterized by its dispersion $D(\alpha)$. Thus, the new condition can be presented as a function of the mean value and dispersion of the objective function.

Applying the neoclassical risk theory [3], the condition $\max(x_0)$, where x_0 is a random variable, can be changed to $\max[E(x_0)-pD(x_0)]$, where $E(x_0)$ is the mean value of the random variable x_0, $D(x_0)$ is its dispersion and p is the price of risk (preassigned). The formula (1) is changed correspondingly.

It can be shown [4] that if $f(\alpha)=\frac{a}{\alpha}$ then $E[f(\alpha)]\approx\frac{a}{\mu_\alpha}$ and $D[f(\alpha)]=\frac{a^2\sigma^2_\alpha}{\mu^2_\alpha}$; where a denotes a constant, μ_α is the mathematical expectation of the random variable α and σ^2_α its dispersion.

Let μ_{vgt} denote the mathematical expectation of the random variable v_g^t and σ^2_{vgt}, its dispersion.

Then, after respective changes and relying on the above-said, the objective function can be presented as the following quadratic function:

$$\max x_0' = \left\{ \sum_{t=1}^{w} \left[u \frac{\sum_{g=1}^{v} k_g x_g^t}{N^t} : \frac{\sum_{g=1}^{v} k_g x_g^{0t}}{N^0} + \right.\right.$$

$$\left.\left. + (1-u) \sum_{g=1}^{v} c_g \frac{x_g^t}{\mu_{vgt}} - p(1-u)^2 \sum_{g=1}^{v} c_g^2 \frac{\sigma^2_{vgt}}{\mu^4_{vgt}} \cdot (x_g^t)^2 \right] \right\} : w. \quad (5)$$

A comparison of formula (5) to formula (1) shows that the importance of meeting the requirement (v_g^t) has decreased, i. e., the objective function has become more cautious with regard to this condition. In addition to the primary objectives, i. e., the maximization of the rise in the population's educational level and the meeting of the demand of the national economy for a population with different educational levels, the objective function reduces the probable error in the plan which may arise because of the deviation of the actual requirement from the mean value forecast (in reality the error becomes apparent in the overtautness or underfulfilment of the plan with regard to a certain educational level; this means that either more or less resources than actually needed are allocated to that level).

4. Next let us discuss constraints (2) and (3).

Constraint (2) comprises, as an auxiliary variable, the population's migration balance s^t_g $(x^{0t}_g = x^{t-1}_g + s^t_g)$. It is a rather unstable variable, and the fidelity of its forecasts is small (its probable divergence from the mean value is considerable). However, its numerical value being negligible as compared to that of the other parameters of the model, and, thus, its effect on the solution unimportant, for simplicity's sake s^t_g may be determined on its prognosticated mean value.

The coefficients r^t_{gj} and q^t_{gj} will likewise be left determined on their mean values — if they were treated as random variables, the variable to be determined x^t_g would also become random. In such a case, however, the problem would not be a fixed plan problem as assumed, and its solution would be too complicated. Moreover, the probable error resulting from the determination of these coefficients on their mean values is considered insignificant for the present model. Namely, on the one hand they are random variables: their values are usually forecast considering their statistical trends and supplementing formalized forecasts with expert estimates; quite often data are available for estimating also their dispersion. On the other hand, however, they have a negative correlation: $\sum_j r_{gj} = 1$, and $\sum_j q_{gj} = 1$, too. Consequently, the deviations of individual members of these sets, r_{gj} or q_{gj}, smooth out one another, and the total probable error resulting from their deviations is smaller than it would be in case of a positive correlation.

5. Let us now discuss constraint (4). Here a^t_{ij} denotes the input coefficient of the resource i in the activity j, and b^t_i, the limit of the resource i.

Like the parameter of the objective function v^t_g, a^t_{ij} and b^t_i are affected mainly by socio-economic factors (the scientific and technological level of national economy, its growth rate, the state's educational policy, i. e., the resources allocated to education, etc.). Therefore, their forecasts should be treated as random variables. Constraint (4) can be met with a certain probability p only, demanding that

$$P\left(\sum_{j=1}^{n} a^t_{ij} y^t_j \leqslant b^t_i\right) \geqslant p. \tag{6}$$

The probability $P(\ldots)$ is expressed as an integral. The solution of a problem with such constraints would be too complicated. Analogically to the objective function, the probability $P(\ldots)$ can be expressed by mean values and variance of random variables.

Denoting the probable upper deviation interval (Q quantile) of the inputs by

$$\mathrm{E}\left(\sum_{j=1}^{n} a^t_{ij} y^t_j\right) + \tau \sqrt{\mathrm{D}\left(\sum_{j=1}^{n} a^t_{ij} y^t_j\right)} = \tilde{u}, \tag{7}$$

where τ is Student's t, and the probable lower deviation interval of the resources (R quantile) is

$$\mathrm{E}(b^t_i) - \tau \sqrt{\mathrm{D}(b^t_i)} = \tilde{w}, \tag{8}$$

it should be required that

$$\tilde{u} \leqslant \tilde{w}. \tag{9}$$

Assuming that there exists no interdependence between the variables a_{ij} (e. g., assuming that the standard number of students per a university lecturer does not depend on the standard number of students per an elementary school teacher), constraint (9) can be presented as the following approximation [5]:

$$\sum_{j=1}^{n} y_j^t \mu_{a_{ijt}} + \tau \sqrt{\sum_{j=1}^{n} (y_j^t)^2 D(a_{ij}^t)} \leq \mu_{b_{it}} - \tau \sigma_{b_{it}}. \tag{10}$$

The need for extraction makes the solution of the problem again too complicated. Formula (10) can be replaced by its approximation:

$$\sum_{j=1}^{n} y_j^t \mu_{a_{ijt}} + \tau \upsilon \sum_{j=1}^{n} y_{jt} \sigma_{a_{ijt}} \leq \mu_{b_{it}} - \tau \sigma_{b_{it}}, \tag{11}$$

where τ is Student's t and

υ is an approximate rooting coefficient found in the course of calculations [6].

The obtained constraint (11) makes the plan more cautious with respect to the spending of resources since here the probability that expenditures might exceed resources is lower than in case of a problem determined on mean values.

6. To sum up, a mean-variance model (12)—(15) has been obtained, which considers the random character of some of its most important parameters, such as the population's need (v_g^t), the input coefficients (a_{ij}^t) and the limits (b_i^t) of the resources:

$$\max x_0' = \left\{ \sum_{t=1}^{w} \left[u \frac{\sum_{g=1}^{v} k_g x_g^t}{N^t} : \frac{\sum_{g=1}^{v} k_g x_g^0}{N^0} + \right. \right.$$

$$\left. \left. + (1-u) \sum_{g=1}^{v} c_g \frac{x_g^t}{\mu_{vgt}} - p(1-u)^2 \sum_{g=1}^{v} c_g^2 \frac{\sigma_{vgt}^2}{\mu_{vgt}^4} (x_g^t)^2 \right] \right\} : w, \tag{12}$$

$$x_g^{0t} - \sum_{j=1}^{n} r_{gj}^t y_j^t = 0, \tag{13}$$

$$\sum_{j=1}^{n} q_{gj}^t y_j^t - x_g^t = 0, \tag{14}$$

$$\sum_{j=1}^{n} y_j^t \mu_{a_{ijt}} + \tau \upsilon \sum_{j=1}^{n} y_{jt} \sigma_{a_{ijt}} \leq \mu_{b_{it}} - \tau \sigma_{b_{it}}. \tag{15}$$

As compared to a problem determined on mean values, this model yields a relatively more cautious plan. In the first place, the objective function (12) reduces the probable error in the plan, which may result from the deviation of the actual requirement from its prognosticated mean value. Secondly, the constraint on the resources (15) considering the variance of the input coefficients (a_{ij}^t) and the limits (b_i^t) around their mean values, reduces the probability that expenditures might surpass resources.

The supplemented model (12)—(15) has a strictly convex objective function and linear constraints. Therefore, its practical solution should not be insurmountably complicated.

REFERENCES

1. Мээл М. Об одной динамической линейной матричной модели для прогнозирования оптимального развития системы образования. — In: О моделях нормативного прогнозирования социально-экономического развития региона. Таллин, 1979.
2. Laas, K. Rahvastiku prognoosimise metodoloogilisi probleeme. — ENSV TA Toim. Ühisk., 1976, 3, p. 230—235.
3. Эннусте Ю. О задачах оптимального планирования на основе средних значений и дисперсий. — Изв. АН ЭССР. Обществ. науки, 1978, 2, p. 111—117.
4. Смирнов Н. В., Дунин-Барковский И. В. Курс теории вероятностей и математической статистики. М., 1965.
5. Мээл М. Об одной стохастической модели планирования оптимального распределения приема учащихся в систему специального образования. — Изв. АН ЭССР. Обществ. науки, 1976, 1, p. 3—13.
6. Колбин В., Танская В. Некоторые задачи стохастического линейного программирования и алгоритмы их решения. — In.: Математические программы и вычислительные методы оптимального планирования. М., 1971, p. 391—401.

Academy of Sciences of the Estonian SSR,
Institute of Economics

Received
Dec. 24, 1979

M. MEEL

ÜHEST REGIOONI ELANIKKONNA HARIDUSTASEME OPTIMAALSE ARENGU PLANEERIMISE KESKVÄÄRTUS-DISPERSIOONMUDELIST

Artiklis on esitatud lineaarne dünaamiline maatriksmudel regiooni elanikkonna haridustaseme optimaalse arengu planeerimiseks. Sisuliselt kujutab see mudel endast juhuslike suuruste keskväärtustele determineeritud optimaalse planeerimise ülesannet. Mudeli mõnede parameetrite tõenäolist hälbimist keskväärtusest võib aga pidada oluliseks. Seepärast on mudel nende parameetrite osas täiendatud keskväärtus-dispersioonmudeliks. Kuigi keerulisem, lähendab see tegelikkust paremini kui rangelt keskväärtustele determineeritud mudel. Täiendamisel on püütud silmas pidada, et saadav ülesanne ei osutuks liiga keeruliseks lahendada.

Eesti NSV Teaduste Akadeemia
Majanduse Instituut

Toimetusse saabunud
24. XII 1979

M. МЭЭЛ

ОБ ОДНОЙ МОДЕЛИ ОПТИМАЛЬНОГО ПЛАНИРОВАНИЯ УРОВНЯ ОБРАЗОВАНИЯ НАСЕЛЕНИЯ РЕГИОНА НА ОСНОВЕ СРЕДНИХ ЗНАЧЕНИЙ И ДИСПЕРСИЙ

Анализ модели оптимального планирования уровня образования населения региона, построенной на основе средних значений параметров, показал, что вероятностные отклонения значений некоторых параметров от прогнозируемых средних значений весьма значительны. Поэтому модель следует дополнить дисперсиями этих же параметров, но таким образом, чтобы это не слишком осложнило практическое решение модели. Такая модель, на наш взгляд, гораздо лучше отражает действительность, чем модель, основанная только на средних значениях параметров.

Институт экономики
Академии наук Эстонской ССР

Поступила в редакцию
24/XII 1979