



PREDICTION IN LOGISTICS USING FUNCTIONS WITH FLEXIBLE STRUCTURE

Olga Fadina¹, Mindaugas Mazūra²

Dept of Transport Engineering, Vilnius Gediminas Technical University, Plytinės g. 27, LT-2040 Vilnius, Lithuania
E-mail: ¹Olga@lineka.lt, ²Mindaugas.Mazura@ti.vtu.lt

Received 2002 11 16; accepted 2003 04 30

Abstract. The economic forecasts are aimed at determining major trends of logistics development as well as defining the particular aims and objectives of their functioning. Therefore, the prediction in logistics should reveal major changes in micro and macro factors to obtain the probability criteria of quantitative and qualitative evaluation of the dynamics of logistics activities required by the firm manager. Some methods of predicting the development of the logistics system based on the application of functions with flexible structure and of evaluating the accuracy and truth (reliability) of forecasts are offered.

Keywords: prediction; logistics; management; function with flexible structure.

1. Predicting

In management of logistics the availability of forecasts about the aggregate index (e. i. volume of sales of manufactured products), depending on one or more factors referring to the capacity and output of production equipment, the use of basic funds and working capital, labour force, etc. are of particular importance. In these cases single – and multicriteria correlation – regressive models are often applied, which have a common limitation of not being directly used in prediction [1, 2].

Moreover, the relations between the factors may be non – linear and stochastic causing additional problems for calculation [1, 3–7]. To predict logistics indices based on the assumption of the existence of the statistical relation between two indices y and x , a mathematical approach relying on the functions of flexible structure (FFS) may be effectively used. The form of FFS may change, being automatically determined in the prediction operation, taking into account not only the statistical dependence $y = f(x)$, but the trends of transformation of each factor in time as well. The application of FFS may also be helpful in solving inverse problems encountered in logistics, for instance, to determine extensive and intensive production factors and parameters for planned sales volume.

2. The Application of Functions of Flexible Structure

The function of flexible structure is of the form:

$$F(x) = A_0 + \sum_{j=1}^n A_j \frac{\delta_j(x)}{D}, \quad (1)$$

where n is a fixed natural number;

x_0 – the initial value of the factor – argument in the interval considered;

A_0, A_1, A_n – parameters;

D – Van der Mond determinant of n -th order.

$$D = \begin{vmatrix} 1 & 1 & 1 & \dots & 1 \\ r_1 & r_2 & r_3 & \dots & r_n \\ r_1^2 & r_2^2 & r_3^2 & \dots & r_n^2 \\ \dots & \dots & \dots & \dots & \dots \\ r_1^{n-1} & r_2^{n-1} & r_3^{n-1} & \dots & r_n^{n-1} \end{vmatrix}$$

r_1, r_n – real – valued or complex conjugate pairs;

$\delta_j(x - x_0)$ – functions obtained from the determinant D substituting the j -th row by the exponents of the form:

$$\frac{1}{r_v} \left(r_v^{(x-x_0)} - 1 \right), \quad v = \overline{1, n}. \quad (2)$$

The essence of FFS approach is to represent the initial process as

$$y(x) = F(x) + R(x), \quad (3)$$

where $R(x)$ – the approximation function (remainder);
 $F(x)$ – FFS.

The structure of FFS is mostly dependent on the values of r_j ($j = \overline{1, n}$). If all of them are real – valued, different and not equal to zero, then FFS will be represented by a linear combination of exponents. If the numbers r_j are complex, FFS is a sum of exponents multiplied by the harmonic (sin–cos) components. When we

have $r_j = 0 (j = 1, n)$, FFS is transformed into a polynomial of some degree.

Hence, the main problem is to find optimal values of $r_j, (j = 1, n)$.

If the function $y(x)$ may be differentiated, then the respective derivatives of $y(x)$ may be taken as the parameters A_0, A_1, \dots, A_j . Then (1) may be written as:

$$F(x) = y(x_0) + \sum_{j=1}^n y^{(j)}(x_0) \frac{\delta_j(x-x_0)}{D} \quad (4)$$

In this case the numbers r_1, \dots, r_n are considered the roots of the original equation:

$$r^n + a_{n-1}r^{n-1} + a_{n-2}r^{n-2} + \dots + a_1r + a_0 = 0. \quad (5)$$

The analytical expression of the remainder in formula (3) is as follows:

$$R(x) = \int_{x_0}^x \int_{x_0}^t \eta(\tau) \frac{\Delta_n(t-\tau)}{D} d\tau dt, \quad (6)$$

where $\Delta_n(t-\tau)$ is obtained from the determinant D substituting the n -th (the last) row by the exponent row of the form:

$$1^{r_\nu(t-\tau)}, (\nu = 1, n). \quad (7)$$

The function $\eta(\tau)$ of the equation (6) is obtained from the equation:

$$\eta(\tau) = y^{(n+1)}(\tau) + a_{n-1}y^{(n)}(\tau) + \dots + a_1y^{(2)}(\tau) + a_0y^{(1)}(\tau), \quad (8)$$

where $y^{(j)}$ is the j -th derivative of y .

The idea of optimal approximation in applying FFS is the minimization of the remainder $R(x)$ and specifying such values to the parameters a_0, a_1, \dots, a_{n-1} , that the remainder value at any point of the approximation segment would not be higher than the particular specified approximation error, which, in fact, determines the form of FFS.

Let us consider a simple case, when $n = 1$. Then the expression (4) may be written as

$$F(x) = y(x_0) + y'(x_0) \left[\begin{matrix} 1^{r(x-x_0)} \\ r \end{matrix} \right]. \quad (9)$$

The main equation (5) for r will be of the form:

$$r + a_0 = 0. \quad (10)$$

The function $\eta(\tau) = \eta(x)$ for the remainder $R(x)$ from the equation (8) will be transformed into:

$$\eta(\tau) = y''(x) + a_0y'(x). \quad (11)$$

The condition of the minimum remainder may be written as $\eta(x) = 0$, which will result in a zero approximation error. Then, to get rid of derivatives in the equa-

tion (11), let us integrate it:

$$y'(x) + a_0y(x) = c_1. \quad (12)$$

Substituting the value of the initial point for the argument $x = x_0$ into the above expression, we get

$$y'(x_0) = c_1 - a_0y(x_0). \quad (13)$$

Let us integrate the equation (12) once more:

$$y(x) = a_0 \int_{x_0}^x y(x) dx = c_1(x-x_0) = c_0. \quad (14)$$

When $x = x_0$, we have $c_0 = y(x_0)$, then the equation (14) will be of the form:

$$c_1(x-x_0) - a_0 \int_{x_0}^x y(x) dx = y(x) - y(x_0). \quad (15)$$

From equation (15) one can see that it contains two unknown constants a_0, c_1 and an integral. The integral is assigned by the tabular function $y(x)$ and can be graphically calculated.

To calculate c_1 and a_0 , let us form a system of equations with two unknowns based on the equation (15). This may be achieved if we substitute into (15) the values of two more points taken from the time series $y(x)$. Then

$$\begin{cases} c_1(x_1 - x_0) - a_0 \int_{x_0}^{x_1} y(x) dx = y(x_1) - y(x_0), \\ c_1(x_2 - x_0) - a_0 \int_{x_0}^{x_2} y(x) dx = y(x_2) - y(x_0). \end{cases} \quad (16)$$

When the integrals are calculated, let us find the unknown coefficients c_1 and a_0 . Then the value of the first derivative $y'(x_0)$ should be found from the equation (13). The root of the original equation $r = a_0$ has a negative sign. The parameters obtained are then substituted into the formula for FFS (9) and the analytic expression of $y(x)$ is obtained.

An Example

There is a problem of determining the relationship between the coefficient of renewal of basic funds of a transportation firm (ρ) and its freight turnover (Q). The initial data are given in Table 1.

First, the value of three basic points should be chosen and one of them, representing the initial value, should be in the middle of the row for better approximation. This implies that the function obtained will equally approximate the data found at the beginning and at the end of the row.

Let us take: $x_0 = 34,3 \quad y_0 = 0,618 = (y(x_0))$;

$x_1 = 31,2 \quad y_1 = 0,597$;

$$x_2 = 37,8 \quad y_2 = 0,631.$$

Let us now find the coefficients of the equations (16):

$$x_1 - x_0 = -3,1; \quad x_2 - x_0 = 3,5;$$

$$y_1 - y_0 = -0,021; \quad y_2 - y_0 = 0,013.$$

Table 1. The Coefficient of renewal of basic funds and freight turnover of a transportation firm

Parameter	1996	1997	1998	1999	2000	2001	2002
$\rho, (y)$	0,597	0,597	0,608	0,618	0,615	0,618	0,631
Q, million arbitrary units (x)	31,2	32,3	33,4	34,3	34,5	35,5	37,8

Let us now find the coefficients of the equations (16):

$$x_1 - x_0 = -3,1; \quad x_2 - x_0 = 3,5;$$

$$y_1 - y_0 = -0,021; \quad y_2 - y_0 = 0,013.$$

Let us calculate the integrals of the system (16).

The integral $\int_{x_0}^{x_1} y(x)dx$, the area restricted by the graph of the (empirical) $y(x)$ and the values of x equal to 34,3 and 31,2. have a negative value because the upper limit of the integral is smaller than the lower limit. The area restricted by x values 34,3 and 31,2 will be obtained summing the areas of three trapezoids:

$$\int_{34,3}^{31,2} y(x)dx = - \left[\begin{aligned} &0,597(32,3-32,1) + \frac{0,597+0,608}{2} \cdot (33,4-32,1) + \\ &\frac{0,608+0,618}{2} \cdot (34,3-33,4) \end{aligned} \right] = -1,871.$$

The value of the integral $\int_{34,8}^{37,8} y(x)dx$ will be:

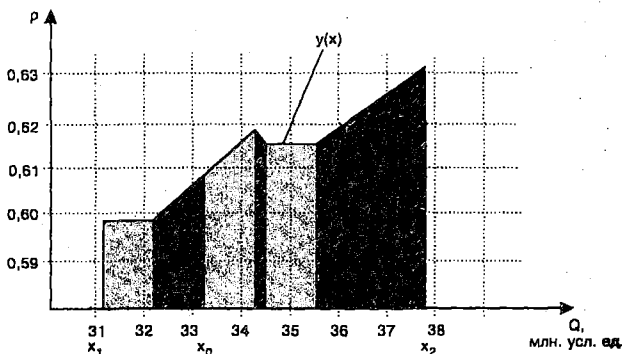


Fig 1. Graphical evaluation of integrals (an example)

$$\int_{34,3}^{37,8} y(x)dx = - \frac{0,618+0,615}{2} (34,5-34,3) + \frac{0,615+0,618}{2} \cdot (35,5-34,5) + \frac{0,618+0,631}{2} (37,8-35,5) = 2,176.$$

Graphical evaluation of integrals is shown in Fig 1.

Let us substitute the values of coefficients and integrals into the system equation (16). We will get:

$$\begin{cases} -3,1c_1 + 1,871a_0 = -0,021, \\ 3,5c_1 - 2,176a_0 = 0,013. \end{cases}$$

Solving the system we will obtain:

$$c_1 = 0,108; \quad a_0 = 0,168.$$

Then the values of the first derivative at the initial point are found substituting the parameters obtained into the equation (13):

$$y'(x_0) = c_1 - a_0 y(x_0) = 0,108 - 0,168 \cdot 0,618 = 0,00417.$$

From the original equation (10) we will get:

$$r = -a_0 = -0,168.$$

Now, having obtained all the required parameters, we can get the expression for FFS:

$$F(x) = 0,618 + 0,00417 \left[\frac{1^{-0,168(x-34,3)} - 1}{-0,168} \right].$$

Opening the brackets and making the necessary calculations, we finally get

$$y = \rho = 0,643 - 0,0248e^{-0,168(Q-34,3)}.$$

Let us find the prediction for 2003. Taking into consideration that the planned freight turnover would be 41 million arbitrary units, we can get the required (predict) coefficient of basic funds renewal:

$$\rho = 0,648 - 0,0248 \cdot 1^{-0,168(41-34,3)} = 0,635.$$

In solving the inverse problem we will get the expression for finding the forecast of freight turnover:

$$Q_1 = 29,057 + 8,24 \exp[3r5(\rho_1 - 0,618)].$$

3. Evaluation the Accuracy and Reliability (truth) of Prediction

Applying various methods based on η models of prediction in logistics management special attention should be paid to the accuracy and reliability of forecasts.

The evaluation of the accuracy and reliability of

prediction, often referred to as forecast verification, is a complicated problem

Verification of prediction models may be carried out by various approaches:

- *direct verification* which consists of obtaining the same results;
- *indirect verification meaning* that prediction is verified by referring to the forecast of the same object in literature;
- *consequential verification* based on obtaining the verifying prediction by deducing consequences from the available forecasts by logical or mathematical operations;
- *inverse verification* – i.e. backward extrapolation. It implies that the data of prediction made by inverse extrapolation are compared with actual data of a retrospective series;
- verification by minimizing systematic errors.

The total error of the problem of prediction may be found in the following way:

$$\varepsilon_n = \varepsilon_u + \varepsilon_M + \varepsilon_\theta + \varepsilon_H, \quad (17)$$

where ε_u – initial data errors;

ε_M – prediction method errors;

ε_θ – calculation error;

ε_H – accidental error caused by unforeseen events affecting the forecast.

In case of inverse verification a major criterion of the adequacy of the type of approximation relationship chosen is mean square deviation of theoretical values of the function from the empirical data of a retrospective series:

$$\sigma_y = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-1}}, \quad (18)$$

where n – a number of observations (members of a series);

y_i – values of an empirical series;

\hat{y}_i – calculated (theoretical) values of a series.

Except for the criterion (18), some more indices of the accuracy of the approximation relationship chosen are used:

- variation coefficient:

$$v = \frac{\sigma_y}{\bar{y}} 100\%, \quad (19)$$

where \bar{y} – an estimate of mathematical expectation of a retrospective series;

- correlation index:

$$R^2 = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (\bar{y} - y_i)^2}, \quad (20)$$

- Fisher's criterion:

$$F = \frac{R^2(n-m)}{(1-R^2)(m-1)}, \quad (21)$$

where m – a number of unknown parameters of the approximation function.

The best approximation function is chosen from the conditions:

$$\sigma, v \rightarrow \min; R^2, F \rightarrow \max. \quad (22)$$

Since the prediction is of probability nature, the parameter predicted should be determined in confidence limits (in terms of intervals).

With the normal distribution of deviations ($y_i - \hat{y}_i$) and σ specified, a confidence interval for a forecast y^* is expressed as:

$$(y^* - \delta, y^* + \delta),$$

where δ – accuracy of estimation obtained from the expression:

$$\delta = \frac{t\sigma}{\sqrt{n}}. \quad (23)$$

If the confidence probability γ is given, the parameter t is found from the relationship:

$$2\Phi(t) = \gamma, \quad (24)$$

where Φ is the Laplace function.

When $\gamma = 0,95$ $\Phi(t) = 0,475$. Referring to the tables of the Laplace function we get that $t = 1,96$.

Then:

$$\delta = \frac{t\sigma}{\sqrt{n}} = \frac{1,96\sigma}{\sqrt{n}}.$$

This means that the confidence interval has the following limits:

$$\begin{cases} y_H^* = y^* - \frac{1,96\sigma}{\sqrt{n}}, \\ y_\theta^* = y^* + \frac{1,96\sigma}{\sqrt{n}}. \end{cases}$$

For getting more precise and reliable forecasts in logistics an integrated approach (or complex evaluation) may be used, allowing to combine two or more procedures. A combined scheme (synthesis of forecasts) allowed us to compensate (offset) disadvantages of one group of methods with the advantages of the other, especially if they are based on different databases.

An example of combined forecasting

Assume that the expected demand for freight transportation (turnover) of a particular firm was determined by three methods, the results of which are given in Table 2.

Table 2. The results of prediction demand for freight transportation (turnover)

Method of forecast, Parameter of forecast	1. Interval – based forecast (observation time $t=30$ months)	2. Interval – based forecast by exponential smoothing	3. Point – based forecast by linear interpolation
Mean value of y , thous. units	3,220	3,400	3,900
Mean square deviation σ_y , thous. units	0,082	0,100	-

To increase the truth of a forecast, a complex scheme should be used.

Let us determine the boundary values for methods 1 and 2 defining the confidence intervals.

To make the forecast by the least squares approach, a formula of dispersion for linear dependence ($y = a_0 + a_1t$) is used:

$$\sigma_y^2(t) = \sigma_{a_0}^2 + \sigma_{a_1}^2 t^2 + 2tKa_0a_1,$$

$$\sigma_{a_0}^2 = \sigma_y^2 B_0; \sigma_{a_1}^2 = \sigma_y^2 B_1; K_{a_0a_1} = \sigma_y^2 B_2.$$

$$\left. \begin{aligned} B_0 &= \frac{\sum t_i^2}{\Delta} = \frac{1139}{1075} = 1,06; \\ B_1 &= \frac{N}{\Delta} = \frac{4}{1075} = 0,0037; \\ B_2 &= \frac{\sum t_i}{\Delta} = -\frac{59}{1075} = -0,055. \end{aligned} \right\}$$

The coefficients B_0, B_1, B_2 are obtained from the least squares equations for $N = 4; t = 30$.

$$\sigma_y^2 = (0,082)^2 = 0,0067.$$

Let us find the mean square deviation

$$\sigma_y(t) = \sqrt{\sigma_y^2 (B_0 + B_1 t^2 + 2tB_2)} =$$

$$= \sqrt{0,0071 + 0,000025t^2 + 0,00074t}.$$

When $t = 30$, we will get $\sigma_y(t = 30) = 0,086$ thous. arbitrary units.

Let us define the confidence intervals for confidence probability $\gamma = 0,9$ from the Student's tables, when the number of freedom degrees is:

$$(k = N - m - 1 = 4 - 1 - 1) = 2: t_{\gamma=0,9, k=2} = 2,92.$$

A forecast of the demand in terms of intervals will

be as follows:

$$Y_{MHK}^{(1)} = y_1 \pm t_r \cdot \sigma_{y(t)} = 3,22 \pm 2,92 \cdot 0,086 = 3,22 \pm 0,25$$

thous. units.

The upper bound $y_B^{(1)} = 3,22 + 0,25 = 3,47$ thous. units.

The lower bound $y_B^{(1)} = 3,22 - 0,25 = 2,97$ thous. units.

Let us calculate the confidence limits for the second forecast obtained by means of exponential smoothing for the linear Brown model. They will be respectively equal to:

$$Y_{Bp}^2 = y_2 \pm t_j \cdot \sigma_y = 3,40 + 0,29 = 3,69 \quad \text{thous. units.}$$

Upper bound $y_B^{(2)} = 3,40 + 0,29 = 3,69$ thous. units.

Lower bound $y_B^{(2)} = 3,40 - 0,29 = 3,11$ thous. units.

For checking the consistency of the integral values obtained in the forecast evaluation the graph of distribution densities $f_i(y)$ of forecast is plotted (Fig 2).

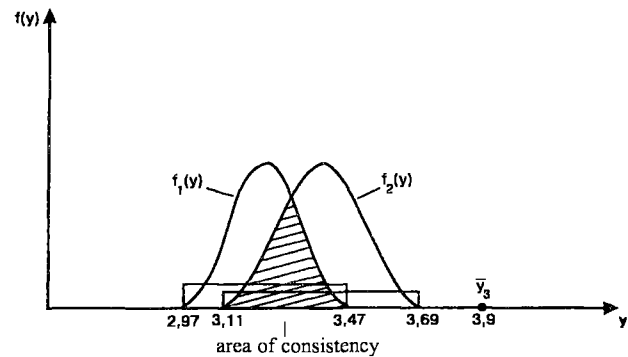


Fig 2. Determination consistency area of forecasts

In Fig 2 one can see that the first two methods have the overlapping area and can be considered consistent. The third method (linear interpolation) is excluded as inconsistent.

Let us calculate weight coefficients for a combined (complex) forecast:

$$\omega_1 = \sigma_2^2 / (\sigma_1^2 + \sigma_2^2) = (0,1)^2 / ((0,086)^2 + (0,1)^2) = 0,57;$$

$$\omega_2 = 1 - 0,57 = 0,43.$$

Let us also define the parameters of a combined (complex) forecast:

$$y^* = \omega_1 y_1 + \omega_2 y_2 = 0,57 \times 3,22 + 0,43 \times 3,4 = 3,3 \quad \text{thous. units.}$$

$$\sigma_{y^*} = \sqrt{\omega_1^2 \sigma_1^2 + \omega_2^2 \sigma_2^2} =$$

$$\sqrt{0,57^2 \times 0,086^2 + 0,43^2 \times 0,1^2} = 0,065 \quad \text{thous. units.}$$

The interval – based forecast is as follows:

$$y_{H,t}^* = 3,3 \pm 2,92 \times 0,065 = 3,3 \pm 0,2 \text{ thous. units.}$$

4. Conclusions

1. To forecast the logistics indices based on the assumption of statistical relationships between two indices y and x , a mathematical approach relying on the functions of flexible structure may be effectively used.

2. The FFS structure is mainly determined by real or complex conjugate pairs. This allows the optimal value to be found.

3. The accuracy and reliability (truth) of a forecast may be determined by direct, indirect, consequential and inverse verification.

4. To obtain more accurate and reliable forecasts in logistics a combined scheme, compensating disadvantages of one group of methods by the advantages of others, is used.

References

1. Mazūra, M. Prediction of Major Trends of Transportation Development. *Transport*, 2002, Vol XVII, No 2, p. 57–59.
2. Baublys, A.; Griškevičienė, D.; Žvaliauskas, A. Prognosis of Transport Activity. Lithuania Academy of Science. Book 22. Vilnius, Lithuania science, 1999, p. 9–61.
3. Yang, W.-H.; Mathur, K.; and Ballou, R. H. Stochastic Vehicle Routing Problem with Restocking. *Transportation Science*, 34, 2000, p. 99–112.
4. Fumero, F.; and Vercellis, C. Synchronized Development of Production, Inventory and Distribution Schedules. *Transportation Science*, 33, 1999, p. 330–340.
5. Reiman, M. I.; Rubio, R.; and Wein, L. M. Heavy Traffic Analysis of the Dynamic Stochastic Inventory-Routing Problem. *Transportation Science*, 33, 1999, p. 361–380.
6. Miller-Hooks, E. D.; and Mahmassani, H. S. Least Expected Time Paths in Stochastic Time – Varying Transportation Networks. *Transportation Science*, 34, 2000, p. 198–215.
7. Yang, H. System Optimum, Stochastic User Equilibrium, and Optimal Link Tolls. *Transportation Science*, 33, 1999, p. 354–360.