

## SELECTED METHODS FOR THE ESTIMATION OF THE LOGISTIC FUNCTION PARAMETERS

The logistic function can be employed as a model of record for a number of processes occurring in the management of technical objects and in logistics. The methods used for function parameter estimation include the analytical methods by Hotelling and Tinter and the numerical procedure of optimization in Excel with the Theil index as the optimization criterion. A numerical example is presented to illustrate the estimation accuracy of the methods discussed.

**Keywords:** estimation, logistic function, Hotelling's method, Tintner's method, Theil's index, optimization in Excel.

### 1. Introduction

In the management of technical objects, there are a number of phenomena with a trend of the function in which the rate of increase accelerates in the initial range of the independent variable until it reaches a maximum at the turning point and then tails off and the function tends asymptotically to a certain dependent variable. The trend model with an increasing and then decreasing rate of changes can be written as many exponential functions called S-functions because of their shape [4]. Examples of the exponential functions are given in Table 1. The considerations will focus on one of the functions, i.e. a logistic function, and the estimation of its parameters. As there are many interesting applications of the logistic function, whether in forecasts [3, 4] or economic models based on empirical data, one may hope to find more applications, for instance, in logistics and management of technical objects.

### 2. Logistic function

The logistic function, also known as logistic curve [3, 4], is described by Eq. (4) given in Table 1. The basic mathematical properties of the function can be established by determining its first (7) and second (8) derivatives. The function is increasing for  $x \geq 0$  and has a horizontal asymptote with equation  $y = a$  and a turning point with coordinates  $x_p = (1/c) \ln b$  and  $y_p = a/2$ .

$$y' = \frac{abc \exp(-cx)}{(1 + b \exp(-cx))^2} > 0 \tag{7}$$

$$y'' = abc^2 \exp(-cx) \frac{b \exp(-cx) - 1}{(1 + b \exp(-cx))^3} \tag{8}$$

The logistic function is convex for  $0 \leq x < (1/c) \ln b$  and concave for  $x > (1/c) \ln b$ . Figure 1 shows examples of logistic functions for four series of data assuming that parameters  $a$  and  $b$  are constant and they are:  $a = 100$  and  $b = 4$ . Parameter  $c$  is different for different series: for series 1  $c = 0.0462$ , for series 2  $c = 0.2773$ , for series 3  $c = 0.0924$ , and for series 4  $c = 0.0231$ .

The logistic function is nonlinear with respect to variable  $x$ . With respect to parameters  $a$ ,  $b$ , and  $c$ , it is not linear either, so their values may be difficult to determine. Gąsiorowski and Kuszewski [4] suggest that a special case of the logistic function (so called Pearl's function), when  $c = 1$ , should be analyzed. The function is defined by the formula:

$$y_x = \frac{a}{1 + b \exp(-x)} \tag{9}$$

where:  $a > 0, b > 1$ .

Tab. 1. Selected exponential and related functions corresponding to S-shaped trends (source: Ref. [3])

	function TYPE	function form	COMMENTS
1	exponential	$y = \exp[a + (b/x)]$ $x \geq 0, b < 0$ (1)	turning point at $(-b/2, e^{a-2})$ $\lim_{x \rightarrow \infty} y = e^a$
2	exponential	$y = \exp[a + (b/\sqrt{x})]$ $x \geq 0, b < 0$ (2)	turning point at $(-b^2/9, e^{a-3})$ $\lim_{x \rightarrow \infty} y = e^a$
3	exponential	$y = \exp[a + (b/x^2)]$ $x \geq 0, b < 0$ (3)	turning point at $(\sqrt{-\frac{2}{3}b}, e^{a-3/2})$ $\lim_{x \rightarrow \infty} y = e^a$
4	logistic	$y = \frac{a}{1 + b \exp(-cx)}$ $a > 0, c > 0, b > 1$ (4)	turning point at $[(\ln b)/c, a/2]$ $\lim_{x \rightarrow \infty} y = a$
5	generalized logistic	$y = d + \frac{a-d}{1 + b \exp(-ct)}$ $c > 0, b > 1, 0 \leq d < a$ (5)	turning point at $[(\ln b)/c, d + (a/2)]$ $\lim_{x \rightarrow \infty} y = d + a$
6	Gompertz's	$y = ab^{e^{-x}}$ $a > 0, 0 < b < 1, 0 < c < 1$ (6)	turning point at $[(\ln(-\ln b))/\ln c, a/e]$ $\lim_{x \rightarrow \infty} y = a$

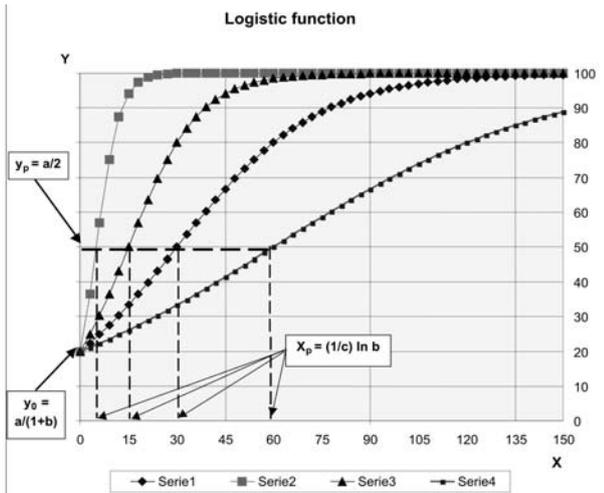


Fig. 1. Logistic functions with different parameter  $c$

Two parameters of Pearl's function can be determined using the following substitutions:

$$\hat{y}_x = \frac{1}{y_x} \quad \square = \frac{1}{a} \quad \hat{b} = \frac{b}{a} \quad \hat{x} = \exp(-x)$$

which will result in the linear relationship:

$$\hat{y}_x = \hat{a} + \hat{b}\hat{x} \tag{10}$$

Basing on the transformed empirical data, i.e. the reciprocal of the original data, it is possible to estimate the parameters of function (10) with the least squares method and then calculate the parameters of Pearl's function (9).

Estimating parameters of a logistic function requires applying a more complex methodology. In Ref. [3] Stanisz provides an interesting discussion of three analytical methods of estimation of the logistic function parameters:

- a) Hotelling's method,
- b) Tintner's method,
- c) Bonus's method.

The author states that only the first method has been popular with Polish scientists. Tintner's and Bonus's methods are generally less known. Stanisz analyzes the three methods thoroughly and proves that Tintner's method is the most accurate. He includes a complete derivation of formulas so that each method of estimation can be understood and applied properly. In this paper, the first two methods, i.e. Hotelling's and Tintner's, are discussed.

### 3. Estimation of the logistic function parameters with Hotelling's method

The idea of this method is to transform the logistic function into a linear one, with the new parameters being simple functions of parameters  $a$ ,  $b$  and  $c$ . Thus, it is essential to transform the formula into the derivative of logistic function (7):

$$\frac{y'}{y} = c - \frac{c}{a}y \tag{11}$$

where:  $y, y'$  – the logistic function and its first derivative, respectively,  $a, c$  – parameters of the logistic function.

Hotelling writes the left side of Eq. (11) as a differential quotient assuming that the increment of the argument is equal to 1. It is also assumed that increments of the left side of the equation have values close to those of quotient  $y'/y$  for the consecutive  $x=1, \dots, n-1$ , where  $n$  is the number of data of the time series. Equation (11) is written as:

$$\frac{y_{x+1} - y_x}{y_x} = c - \frac{c}{a}y_x \quad (x=1, \dots, n-1) \tag{12}$$

Basing on the empirical data in the form of the time series  $(1, y_1), \dots, (n, y_n)$ , one can calculate the relative increments

$$u_x = \frac{y_{x+1} - y_x}{y_x} \quad (x=1, \dots, n-1) \tag{13}$$

and create a new time series  $(1, u_1), \dots, (n, u_{n-1})$ . Differential equation (12) can be written in the linear form

$$u_x = c - \frac{c}{a}y_x \quad (x=1, \dots, n-1) \tag{14}$$

Now applying the least squares method to the above equation and the new series  $(y_1, u_1), \dots, (y_{n-1}, u_{n-1})$ , one obtains the estimates of parameters  $a$  and  $c$ . Parameter  $b$  can be determined from the formula:

$$\hat{b} = \frac{1}{n} \sum_{x=1}^n \left( \frac{a}{y_x} - 1 \right) \cdot e^{c \cdot x} \tag{15}$$

The following twenty-element time series:  $(x, y_x) = \{(1, 3), (2, 3.5), (3, 5.5), (4, 6), (5, 9.5), (6, 12.7), (7, 15), (8, 16), (9, 20), (10, 24), (11, 26.5), (12, 28), (13, 29.5), (14, 36), (15, 37), (16, 38), (17, 40), (18, 44), (19, 46), (20, 47)\}$  was used in the calculations as the empirical data to obtain the differential equation in the linear form (14) (see its graphical representation in Fig. 2).

Using the parameters of the linear function (14), it was possible to calculate the parameters of the logistic function:  $a = 44.02; b = 19.54; c = 0.35$ .

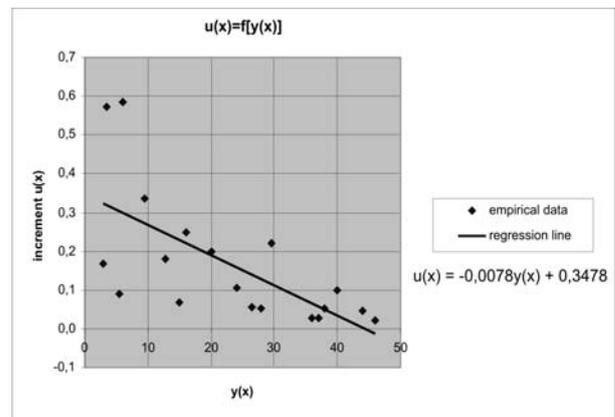


Fig. 2. Relationship (14) for the assumed empirical data

### 4. Estimation of the logistic function parameters with Tintner's method

The first operation is the transformation of the empirical data of the series  $(1, y_1), \dots, (n, y_n)$  into a new series given by the formula:  $(z_x, z_{x+1}) = (1/y_x, 1/y_{x+1}), (x=1, \dots, n-1)$ ,

where:

$$z_x = \frac{1}{y_x} = \frac{1 + b \exp(-cx)}{a} \quad (16)$$

$$z_{x+1} = \frac{1}{y_{x+1}} = \frac{1 + b \exp(-c(x+1))}{a} \quad (17)$$

Then, the equations are transformed into a differential equation:

$$z_{x+1} = (\exp(-c))z_x + \frac{1 - \exp(-c)}{a} \quad (18)$$

which can be represented as a linear equation with respect to the new parameters:

$$z_{x+1} = \delta z_x + \Gamma \quad (19)$$

where:

$$\delta = \exp(-c) \quad \Gamma = \frac{1 - \exp(-c)}{a}$$

The least squares method was used for the differential equation (19) and the data of the new series ( $z_x, z_{x+1}$ ). It was possible to establish, first, the estimates of the parameters of the differential equation, and then the estimates of parameters  $a, b$  and  $c$  of the logistic function. The estimates of the parameters can be determined from the following formulas:

- for parameter  $a$       $\alpha = \frac{1-d}{g} \quad (20)$

- for parameter  $c$       $\gamma = -\ln d \quad (21)$

- for parameter  $b$       $\beta = \frac{1}{n} \sum_{x=1}^n \left( \frac{a}{y_x} - 1 \right) \cdot \exp(\gamma x) \quad (22)$

where:

$$g = \frac{U_0}{U} \quad d = \frac{U_1}{U} \quad U = (n-1) \sum_{x=1}^{n-1} z_x^2 - \left( \sum_{x=1}^{n-1} z_x \right)^2$$

$$U_0 = \sum_{x=1}^{n-1} z_{x+1} \sum_{x=1}^{n-1} z_x^2 - \sum_{x=1}^{n-1} z_x \sum_{x=1}^{n-1} z_x z_{x+1}$$

$$U_1 = (n-1) \sum_{x=1}^{n-1} z_x z_{x+1} - \sum_{x=1}^{n-1} z_{x+1} \sum_{x=1}^{n-1} z_x$$

The parameters of the logistic function were calculated basing on the same empirical data as those in Section 3. Tintner's method was applied and the parameters of the logistic function estimated from the values of the linear function parameters:  $a = 50.83; b = 23.71; c = 0.28$ . The graphical representation of relationship (19) is the diagram in Fig. 3.

### 5. Numerical estimation of the logistic function parameters

State-of-the-art mathematical computer programs can generate a nonlinear function optimally fitted to the given data. Also, the popular spreadsheet Microsoft Excel can be used to fit the function to the data within the line trend determination procedure. First, open an XY-type point diagram. Then, by clicking with the right button of the mouse several data series, select from the menu "add a trend line". The program offers six types of the trend function: *linear, moving average, logarithmic, polynomial, power and exponential*. To fit another model, for instance, the logistic function, which is of interest to us, one needs to use an additional program - *Solver* - compatible with Excel. The program selected from the pull-down menu of Excel is employed to optimize the calculations.

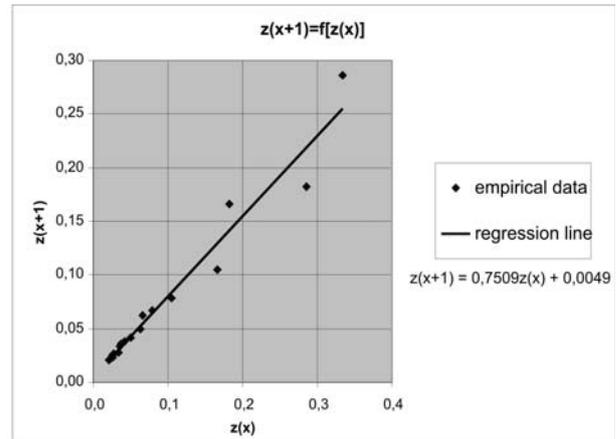


Fig. 3. Relationship (19) for the same empirical data as those in Hottelling's method

The information about the program included in the spreadsheet help guide is not sufficient. Similarly, a very brief description is given in Ref. [1]. Bourg writes that Solver uses the generalized reduced gradient algorithm developed by Leon Lasdon and Allan Waren for optimizing nonlinear problems. The algorithm will be employed to approximate the time series data ( $x, y_x$ ). It is assumed that in the example the approximation function  $\hat{y}_x$  will be the logistic function. The determination of parameters of the function  $\hat{y}_x$  will be reduced to an optimization problem, in which the value of the measure selected to fit the function  $\hat{y}_i$  to the empirical data is minimized. The optimization involves the following operations [1]:

- in *Solver*, specify the optimization criterion or the target function; enter the address of the target cell containing the formula that will be modified by the program until it reaches a maximum, minimum or another desired value;
- specify the constraints that define the relationships between the variables; in the program only one-side constraints, e.g.  $h_i(x) \geq k_i$ , can be specified in the "Constraint conditions" window;
- specify the boundary conditions that provide information about the ranges in which the required variables can be found in the optimal solution; in the program the boundary conditions are also specified in the "Constraint conditions" window;
- define the variables; in the program we enter the addresses of the cells being changed.

Analyzing Bourg's comments [1], one will notice that the target cell may contain numerous references to other cells with subsequent formulas. The cells with the formulas may have further references, etc. Thus, there seem to be a great many possibilities of expansion of the target function. In the example, it was assumed that the optimization criterion is the minimization of the Theil index, which is described by Eq. (23). In the original record (see Ref. [2]) the Theil index is employed to assess a forecast error. For the purpose of the example, the formula for the Theil index will be converted into the following form:

$$I^2 = \frac{\sum_{x=1}^n (y_x - \hat{y}_x)^2}{\sum_{x=1}^n y_x^2} \quad (23)$$

where:  $y_x$  – empirical values of the time series,  $\hat{y}_x$  – values of the logistic function.

The index is assumed to be equal to zero, if the fitting of the logistic curve to the empirical data is accurate. In the optimization problem, the Theil index is the target function because it is a sum of three components, each with a different estimate of fitting.

$$I^2 = I_1^2 + I_2^2 + I_3^2 \quad (24)$$

In forecasts, the first component  $I_1^2$  determines the size of errors relating to the forecast biasability. The errors result from the fact that the average value of the forecasting variable has not been defined. In the example, the component should be used as a measure of loadability of approximation. It can be determined according to the formula:

$$I_1^2 = \frac{(\bar{y} - \bar{\hat{y}})^2}{\frac{1}{n} \sum_{x=1}^n y_x^2} \quad (25)$$

where:  $\bar{y}$  - arithmetical mean of the empirical values of the time series,  $\bar{\hat{y}}$  - arithmetical mean of the values of the logistic function.

In forecasts, the second component  $I_2^2$  determines the size of errors relating to the insufficient flexibility of forecast. The errors result from the fact that the fluctuations of the forecasting variable have not been defined. In the example, the component should be used as a measure of insufficient flexibility of approximation. It can be determined according to the formula:

$$I_2^2 = \frac{(s_r - s_p)^2}{\frac{1}{n} \sum_{x=1}^n y_x^2} \quad (26)$$

where:  $s_r$ ,  $s_p$  – standard deviation of the empirical values of  $y_x$  and standard deviation of the values of the logistic function  $\hat{y}_x$  respectively.

In forecasts, the third component  $I_3^2$  determines the size of errors relating to the insufficient consistency of forecasts with the actual direction of changes in the forecasting variable. In the example, the component should be used as a measure of consistency of approximation. It can be determined according to the formula:

$$I_3^2 = \frac{2s_r s_p (1 - r)}{\frac{1}{n} \sum_{x=1}^n y_x^2} \quad (27)$$

where:  $r$  – coefficient of the linear correlation between  $y_x$  and  $\hat{y}_x$ .

### 6. Example of the estimation of the logistic function parameters in the Solver program

It is assumed that the twenty-element set ( $x=1, 2, \dots, 20$ ) of values forming a time series of quantity  $Y$  with values identical with those in the examples presented in Sections 3 and 4, where  $(x, y_x) = \{(1, 3), (2, 3.5), (3, 5.5), (4, 6), (5, 9.5), (6, 12.7), (7, 15), (8, 16), (9, 20), (10, 24), (11, 26.5), (12, 28), (13, 29.5), (14, 36), (15, 37), (16, 38), (17, 40), (18, 44), (19, 46), (20, 47)\}$  is known. Another assumption is that the model of changes in variable  $Y$  in the function of time is the logistic function given by Eq. (4). The problem requires determining the values of three parameters of the logistic function for which the Theil index (23) reaches a minimum. The calculations were made in the Solver program by performing the following operations:

- the values of  $y_x$  (hereafter called empirical set) were represented in a graphical form; then, the parameter values were calculated for  $a = 50$  and  $b = 9$ , which results from the assumption that the initially determined logistic function intersects the OY axis at point  $(y_0, x_0) = (5, 0)$  (see Fig. 1), and  $c = 0.219722$ , which results from the assumption that the abscissa of the turning point of the logistic function,  $x_0$ , is equal to 10 (see Fig. 1);
- the estimated parameters  $a$ ,  $b$  and  $c$  were used to determine the values of the logistic function and calculate the Theil index;
- the components of the Theil index were employed as measures of:
  - $I_1^2$  – biasability of approximation,
  - $I_2^2$  – flexibility of approximation,
  - $I_3^2$  – consistency of approximation.
- the optimization criterion was the minimization of the Theil index; the values of  $\hat{y}_x$  in the formula of the Theil index (23) were calculated after substituting the estimated parameters  $a$ ,  $b$  and  $c$ ,
- the constraints and boundary conditions were established:  $a >= 0.001$ ;  $b >= 1.001$ ;  $c >= 0.001$  to satisfy the assumptions in formula (4); it was also assumed that the parameters to be changed are quantities  $a$ ,  $b$  and  $c$ .

The optimization results obtained with the Solver program are presented in Table 2 and Fig. 4. After the optimization, the logistic curve is better fitted to the empirical data, the evidence of which is a favorable change in the Theil index. Lower values of the first and second components of the Theil index are attributable to smaller differences between the averages of the empirical data and those of the logistic function. The logistic function takes into account the

Table 2. Comparison of the model parameters before and after optimization and applying Hotelling's and Tintner's methods

Name	Value before the optimization (for the estimated a, b, c)	Value after the optimization	Hotelling's method	Tintner's method
Parameter a	50.0	51.1	44.0	50.8
Parameter b	9.0	15.3	19.5	23.7
Parameter c	0.22	0.25	0.35	0.28
Theil index	$7.8 \cdot 10^{-3}$	$1.8 \cdot 10^{-3}$	$12.3 \cdot 10^{-3}$	$3.9 \cdot 10^{-3}$
1st component of the Theil index	$3.3 \cdot 10^{-3}$	$0.0 \cdot 10^{-3}$	$3.5 \cdot 10^{-3}$	$0.1 \cdot 10^{-3}$
2nd component of the Theil index	$3.2 \cdot 10^{-3}$	$0.1 \cdot 10^{-3}$	$0.0 \cdot 10^{-3}$	$1.2 \cdot 10^{-3}$
3rd component of the Theil index	$1.4 \cdot 10^{-3}$	$1.7 \cdot 10^{-3}$	$9.1 \cdot 10^{-3}$	$2.7 \cdot 10^{-3}$

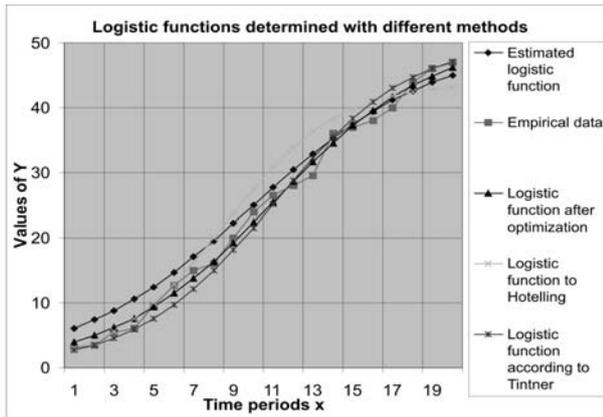


Fig. 4. Result of the estimation of the parameters of the logistic function determined with different methods for the given empirical data

fluctuations of the empirical data. Unfortunately, there is hardly any change in the third component after optimization and its high share in the value of the Theil index shows poor consistency of curve fitting. It may seem necessary to look for another function to record the changeability of a given phenomenon. The results show that the best fitting was obtained for Tintner's method.

## 7. Conclusions

- The paper presents a set of functions for modeling certain phenomena and processes, which may be of use to record trends with an initial increase and then a decrease in the rate

## 8. References

- [1] Bourg D.: *Excel w nauce i technice. Receptury*. Wyd. Helion S. A. 2006. Tłumaczenie z: *Excel Scientific and Engineering Cookbook* 1<sup>st</sup> edition ©2006 O'Reilly Media, Inc.
- [2] Cieślak M. (pod red.) i inni: *Prognozowanie gospodarcze. Metody i zastosowania*. Wyd. Naukowe PWN, Warszawa 1997.
- [3] Stanisław T.: *Funkcje jednej zmiennej w badaniach ekonomicznych*. Państwowe Wydawnictwo Naukowe, Warszawa 1986.
- [4] Szapiro T. (pod red.) i inni: *Decyzje menedżerskie z Excelem*. Polskie Wydawnictwo Ekonomiczne, Warszawa 2000.

of changes. The logistic function was selected for detailed analysis.

- To popularize the complex analytical methods of estimation of nonlinear curves, which include the logistic function, it was necessary to explain in a simple way the mathematical concept of Hotelling's and Tintner's methods. It was also essential to discuss the practical application of the numerical method of nonlinear optimization of the generalized gradient available in the Solver-Excel program.
- The above-mentioned methods were employed to assess three parameters of the logistic function assuming that the example set of empirical data is in the form of a time series.
- It is suggested that the Theil index can be used as a measure of estimation accuracy for the methods discussed above. The index components can be employed to assess the approximation.
- An example shows that the most accurate is the numerical method. A slightly less accurate is Tintner's method, where the Theil index is twice as big and the values of each component are higher. It should be noticed that Hotelling's method is the least accurate method. The Theil index in this case is 0.012, which testifies to good fitting of the logistic curve to the assumed empirical data.
- The result of the method comparison cannot be generalized, as it is based on the solution of one problem only. The considerations require further investigations based on more advanced mathematical methods.

---

**Dr inż. Zbigniew SKROBACKI**

Kielce University of Technology  
 Faculty of Mechatronics and Machine Building  
 Al. 1000-lecia Państwa Polskiego 7  
 25-314 Kielce, Poland  
 e-mail: zbig@tu.kielce.pl

---