UDC 519.6
DOI: 10.15587/1729-4061.2023.289292

THE POLYNOMIAL FORECASTS IMPROVEMENT BASED ON THE ALGORITHM OF OPTIMAL POLYNOMIAL DEGREE SELECTING

Yurii Turbal
Corresponding author
Doctor of Technical Sciences, Professor*
E-mail: turbaly@gmail.com

Ganna Shlikhta
PhD
Department of Information and Communication Technologies
Rivne State University of Humanities
S. Bandery str., 12, Rivne, Ukraine, 33000

Mariana Turbal*
Bogdan Turbal
Departament of Mathematical Informatics
Taras Shevchenko National University of Kyiv
Volodymyrska str., 60, Kyiv, Ukraine, 01033
*Department of Computer Sciences and Applied Mathematics
National University of Water and Environmental Engineering
Soborna str., 11, Rivne, Ukraine, 33028

The object of research in the paper is extrapolation problems based on interpolation polynomials. Polynomial-based prediction methods are well known. However, the problem is that such methods often give very large errors in practice. The permissible error of extrapolation even by one grid step is not ensured by the high accuracy of interpolation using polynomials.

The paper proposes an algorithm that allows to significantly improve polynomial forecasts by optimizing the procedure for choosing the power of the polynomial, on the basis of which the forecast is built.

The algorithm is based on the procedure for building all polynomial forecasts according to known data and analysis of these forecasts. In particular, the presence of monotonicity and a tendency to convergence allows determining the optimal degree of the polynomial. In the absence of monotonicity, provided that certain ratios are met, the forecast can be constructed as the arithmetic average of all polynomial forecasts. An important result is the estimation of the error of the forecasting method by averaging polynomial forecasts.

The development of the algorithm became possible due to the use of a special method of constructing a one-step polynomial forecast. The method differs in that it allows to build a forecast without using the cumbersome procedure of calculating the unknown coefficients of the polynomial.

The numerical results presented in the work demonstrate the effectiveness of the forecasting technique based on the average of polynomial forecasts. In particular, for the test functions, the relative error was about 2–5 %, while polynomials of different degrees in the worst case yielded more than 50 %.

The obtained results can be useful for building short-term forecasts of series of economic dynamics, forecasting the behavior of arbitrary processes with a dominant deterministic component

Keywords: prediction algorithm, problem extrapolation, time series, split differences net, Newton's polynomials, Pascal's triangle, convergence of predictions, binomial coefficients, extrapolation error

Received date 21.08.2023
Accepted date 23.10.2023
Published date 30.10.2023

How to Cite: *Turbal, Y., Shlikhta, G., Turbal, M., Turbal, B. (2023). The polynomial forecasts improvement based on the algorithm of optimal polynomial degree selecting. Eastern-European Journal of Enterprise Technologies, 5 (4 (125)), 34–42. doi: https://doi.org/10.15587/1729-4061.2023.289292*

1. Introduction

The problems of forecasting based on experimental data are extremely important and relevant [1], despite the presence of a large number of known methods and approaches. Known mathematical methods used to build forecasts and based on interpolation dependencies [1, 2] are limited by the problem of the adequacy of the mathematical model, the degree of stochasticity of the process under investigation. After all, it is impossible to predict a stochastic process with one experiment at all [3] and then to talk about certain averaged characteristics. If there is an adequate mathematical model of some deterministic process, it is easy to build good forecasts.

However, in practice, there are often small time series of observations for which it is problematic to determine an adequate mathematical model, in particular, due to insufficient information [4, 5].

This applies, for example, to various series of economic dynamics, when the degree of stochasticity depends on the intervals between observations: at large intervals, the process can be considered stochastic, at small intervals, it is conditionally deterministic, if to take into account a certain inertia of economic systems. In such cases, it is possible to use trend curves, in particular, polynomials. However, the polynomial extrapolation of tabulated functions has a number of peculiarities and problems, which, in particular, is indicated in the paper [6]. Therefore, the development of new extrapolation algorithms that will allow to improve polynomial forecasts is an urgent task.

2. Literature review and problem statement

The issue of approximation and extrapolation of functions based on interpolation polynomials is discussed in detail in [1]. However, due attention is not paid here to estimates of the extrapolation error, and attention is also not paid to the fact that increasing the degree of the polynomial to increase the interpolation accuracy leads to an increase in the extrapolation error.

The paper [2] deals with approximation theorems on the approximation of continuous functions by polynomials for Lebesgue spaces, which is, in a certain sense, a generalization of the well-known theorems of Weierstrass, Dziadyk, and Tiemann. These and a number of similar results apply, as a rule, to a finite interval. If the function is defined on an infinite interval, then the conditions of the corresponding theorems are not fulfilled. Applying relevant results for extrapolation is also quite difficult, because there is no certainty in the stationarity of real processes.

The paper [3] suggests the use of generalized Chebyshev polynomials for extrapolation, but nothing is said about the error. This is due to the fact that the very concept of extrapolation error cannot be correctly introduced without additional conditions, since the function is then not defined outside the time interval where the observations are known. The expression for the error contains an unknown value that requires at least some interpretation, for example, a probabilistic one.

Trigonometric polynomials for approximating functions in specific spaces are considered in [4]. The problem here is similar – if a time series is known, which is well approximated by trigonometric polynomials on a finite interval, then it is quite difficult to use such polynomials for extrapolation, if there is no additional information about the properties of the function.

In [5] certain seasonal fluctuations were considered, which also limits the class of problems. After all, then certain stochastic shifts in seasonality arise, which will affect the error.

In practice, forecasts based on polynomials often give very poor results for extrapolation, as indicated in [6]. The reason for this is the behavior of polynomials of high degrees at the boundaries of the study interval, where they grow or fall too quickly. The results of work [6] became the basis of numerical algorithms for calculating polynomials. In this paper, a formula for a fast polynomial forecast was proposed, however, the theoretical result contained the limit of polynomial forecasts, which made it very difficult to use it in practice.

Methods based on radial basis functions and extrapolation methods based on splines are known [7]. In particular, the work [7] describes the method of extrapolation to new cubic splines, where the parameters are chosen from the condition of minimizing the quadratic functional. However, such a method also assumed certain boundary conditions for the rightmost prognostic spline, which is a certain limitation of the method. In addition, the algorithm has a rather high computational complexity due to the need to solve systems of high-dimensional equations to find parameters. Similarly, cubic splines were considered in [8].

Various statistical approaches described in works [9] and [10] are widely known, in particular, the geometric method, which consists in taking into account the arithmetic mean of the ratios of the elements of the time series, Bayesian methods.

In [11], the Poly-Weibull method is proposed, in which the error has a special class of distribution. However, such an assumption requires additional research for specific time series and narrows the possibilities of the methodology in a certain sense.

The problems of extrapolation using interpolation polynomials explain the emergence of extrapolation methods that do not use the principles of interpolation, in particular, the use of trend curves that do not pass directly through the interpolation points [12]. Such methods include, for example, the use of Bezier curves. In [12], a pyramidal method of extrapolation is proposed, which is based on the search for regularities in the data. This method can be attributed to intelligent methods of analysis and forecasting. The disadvantage of this approach is large errors in the case when the process contains a stochastic component as well as dependence on the uncertainty interval. Machine learning methods and artificial neural networks are important in forecasting [13, 14].

The analysis of literary sources shows that many methods are based on an assumption about the type of function that models the behavior of the investigated process. And this is a drawback of the corresponding approaches, especially under conditions of information limitation. After all, it is obvious [15] that for any given set of points there are many curves that pass through them or somehow bring them closer together. And that is why it is difficult to claim that any one curve (model) is exactly the law that comprehensively describes the phenomenon and will allow to effectively predict its behavior in the future [15]. The functional dependence to which the point observations correspond is effective if it allows predicting the corresponding process with acceptable accuracy starting from some observation step. But in practice, it is far from always possible to find such a dependence.

All this gives reason to assert that it is expedient to conduct research and develop a new combined method of extrapolation, which would allow combining the advantages of polynomial-based forecasting methods with statistical approaches and methods of intelligent data analysis.

3. The aim and objectives of the study

The aim of the study is to improve polynomial forecasts by developing an algorithm that allows determining the optimal degree of the polynomial for forecasting. This will make it possible to significantly improve the accuracy of forecasting using polynomials.

To achieve the aim, the following objectives were set:

– theoretically substantiate the method of directly finding the predictive value based on a polynomial of arbitrary degree without solving systems of linear algebraic equations for finding the coefficients of the polynomial;

– determine the conditions under which the extrapolation value for one step can be constructed as the arithmetic mean of polynomial forecasts and find an estimate of the extrapolation error;

– formulate a general algorithm for constructing a forecast value based on the procedure for finding the optimal degree of an extrapolation polynomial.

4. Materials and methods of the study

The object of the study is extrapolation methods based on interpolation polynomials. The problem of extrapolation is classically formulated. Let $f_i=f(i\Delta)$, $i=n, n-1, ..., n-m$ be the known values of some function, which are determined at the points $i\Delta$, respectively. It is necessary to evaluate the value of this function f_{n+1}, defined at some point $x=x_{n+1}>x_n$. Here let's consider the point $x=(n+1)\Delta$.

To solve such a problem in practice, in particular, Newton's interpolation polynomials of the second kind of different powers are used under the condition that the point x is close to the point x_n [2].

The main hypothesis of the study is that the effective forecast of the function at the point $x=(n+1)\Delta$ can be constructed as an arithmetic average of forecasts based on polynomials of different degrees or a polynomial of the maximum-highest degree in the presence of convergence of polynomial forecasts. However, in order to implement such an approach, it is important to determine the conditions under which such a forecast will be effective. A separate important issue is the error of extrapolation.

In the process of research and development of the forecasting algorithm, the method of finding the forecast value for one step based on a polynomial of any degree without finding the coefficients of the polynomial, which was proposed in [15], was used. This method makes it much more efficient to find the predictive value per grid step compared to the classical approach. Indeed, in order to build a forecast based on a polynomial of a certain degree, it is necessary to determine the coefficients of the polynomial by solving a system of linear algebraic equations. If to sequentially consider forecasts based on 2, 3, 4, etc. consecutive points $f_i=f(i\Delta)$, $i=n, n-1, ..., n-m$, then it is possible to notice that the forecast result based on m points it is impossible to use for $m+1$ points – it is still necessary to solve a new system of equations. Therefore, in order to optimize the computational procedure, it is better to abandon the problem of solving a system of linear algebraic equations for building a one-step forecast, and use the result proposed in [6].

Due to the simplicity of the corresponding calculation procedure, the predictive value can be easily found. Therefore, MS Excel was used in the study to check the effectiveness of the forecasting algorithm and to illustrate the methodology. If polynomials of higher degrees are used, it is easy to use any high-level programming language to construct a forecast. At the same time, the main element of implementation will be the calculation of binomial coefficients. Let's give some examples.

Let's have the function $f(x)=x^6\sin(x)$, defined at points 1, 2, 3, 4, ..., 10. In the Table 2, the first line shows the values of the function at points 13, 14, 15, 16, 17, 18, 19, 20, and the following lines show the values of interpolation polynomials of the corresponding degree constructed from the previous points at the same point. Then it is possible to determine which degree of the polynomial is best to calculate the predicted value and find the error. In the Table 1 in the first highlighted line, the exact values are given and the best predicted values are highlighted. As a result, let's obtain get that for the value of the function at point 13, the best prediction is given by a polynomial of degree 8, 14–6.15–9.16–3.17–10.18–4.19–6.20–10. Thus, it can be seen that looking at the values of an arbitrary function at different consecutive points, there are different degrees of polynomials that give a minimum error.

Table 2 shows the values of the arithmetic mean of 7 interpolation forecasts, the minimum and maximum values that determine the interval of polynomial uncertainty, the length of the corresponding interval and its middle.

Fig. 1 shows forecast points for the fifth column of the Table 1.

As can be seen from this example, the degree of the interpolation polynomial best for forecasting is the one whose forecast value deviates minimally from the average value of all forecasts. In other words, the average of all polynomial predictions gives the best prediction.

The next example is the value of the same function in points 7, 7.5, 8, 8.5, 9, 9.5, 10, 10.5, respectively. Forecast values for this case are given in the Table 3, where the cells with the best forecasts are similarly highlighted.

Table 1

Illustration of choosing the best power of the interpolating polynomial for extrapolation

2028066	7458814	7407185	–4830216	–23205798	–25542746		7051105
–1432853	5658330	12889561	7355556	–17067617	–41581380		–27879695
–35985	9119249	14690045	1873180	–29253389	–47719562		–11841061
1825368	11183300	13029610	–5409680	–35956785	–41671970		10335754
3428506	11385997	9305124	–11032105	–35377321	–28920984		26464978
4273126	9985558	5377940	–12930044	–29175432	–16749461		29843216
4173055	7740498	2851196	–10900800	–21075603	–10779828		21049930
3259596	5595508	2569512	–6344812	–15005019	–12910023		6287012
1910319	4363978	4432818	–1507139	–13490424	–21110802		–6345712
625813	4481726	7527653	1467228	–16813501	–30826176		–10777657
–117658	5883979	10504741	1346760	–23110946	–37218473		–5494227

Table 2

Additional parameters in the analysis of forecast values

Min	–1432853	4363978	2569512	–12930044	–35956785	–47719562
Max	4273126	11385997	14690045	7355556	–13490424	–10779828
average	1420137	7874988	8629779	–2787244	–24723605	–29249695
interval	5705979	7022019	12120533	20285601	22466361	36939734

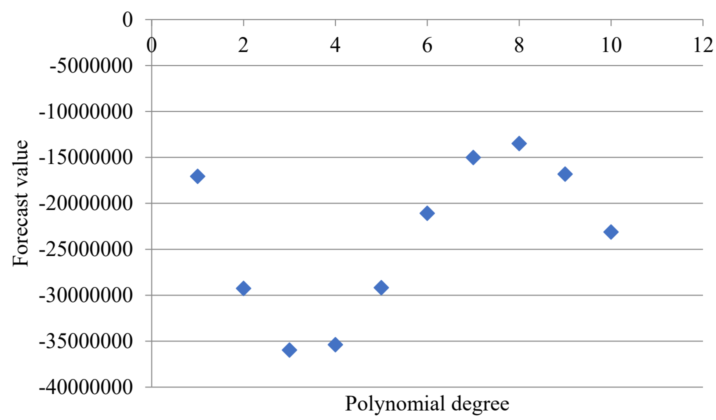

Fig. 1. Extrapolation points based on interpolation polynomials of degrees 1, 2, ... respectively (value of column 5 of Table 1)

Table 3

Illustration of choosing the best power of the interpolating polynomial

77294	166944	259354	301149	219017	–55243	–544021
45485	138364	256594	351765	342944	136884	–329503
68252	170173	285174	354525	292328	12957	–521630
79979	179215	281945	328705	238952	–60354	–589830
82985	176530	269674	306114	211395	–80289	–584719
81439	170838	260089	295794	206430	–72668	–559673
78795	166693	256194	295060	211785	–60081	–542248
76939	165191	256445	298220	217874	–52849	–537410
76336	165546	258198	301129	220802	–51706	–539804
76608	166504	259596	302286	220822	–53491	–543341
77140	167190	260035	302045	219685	–55297	–545093

Table 3 shows a significant shift of the best forecasts to polynomials of higher degrees. At the same time, there is a certain degree of convergence starting from the 7th degree polynomial, the corresponding forecast values differ relatively little. Obviously, in such a situation, a polynomial of the maximum degree should be chosen. At the same time, average values will also give a good forecast with a large number of observations.

The graph for the data of the first column of the Table 3 is shown in Fig. 2. In this figure, it is possible to see a clear tendency towards convergence. Therefore, the optimal value of the exponent is chosen as the maximum for the available data.

A similar situation is observed for the function $f(x)=\exp(x)$, defined in points 7, 7.5, 8, 8.5, 9, 9.5, 10. The corresponding forecast values are given in the Table 4. It is easy to see that there is a tendency to convergence, the sequence of forecast values is monotonous and limited, the distance between neighboring points goes to 0. In such a situation, let's choose a polynomial of the maximum degree.

In Fig. 3, to illustrate the convergence, a graph is given based on the seventh column from the Table 4.

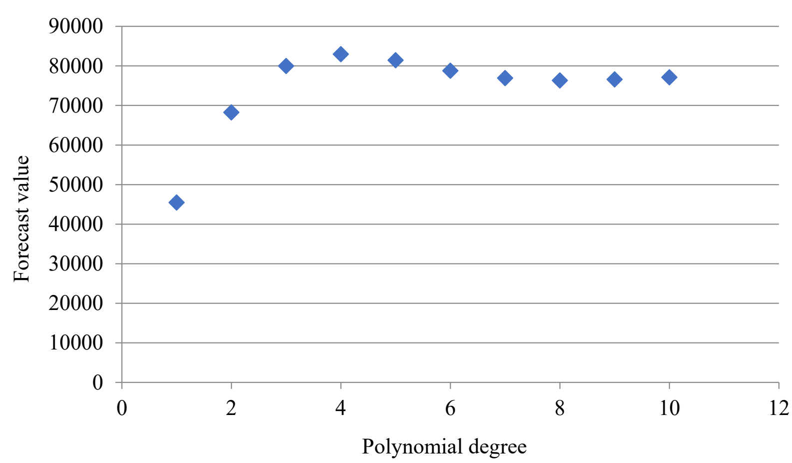

Fig. 2. Extrapolation points based on interpolation polynomials of degrees 1, 2, ... respectively (values of the first column of Table 3)

Table 4

Illustration of choosing the best power of the interpolating polynomial

Degree	1097	1808	2981	4915	8103	13360	22026	36316
1	927	1528	2519	4154	6849	11291	18616	30693
2	1030	1698	2799	4615	7609	12546	20685	34103
3	1070	1765	2910	4797	7909	13040	21499	35445
4	1086	1791	2953	4868	8027	13234	21819	35973
5	1093	1801	2970	4897	8073	13310	21945	36181
6	1095	1805	2977	4908	8091	13340	21994	36262
7	1096	1807	2979	4912	8098	13352	22014	36295
8	1096	1808	2980	4914	8101	13357	22021	36307
9	1097	1808	2981	4914	8102	13359	22025	36312
10	1097	1808	2981	4915	8103	13359	22026	36314

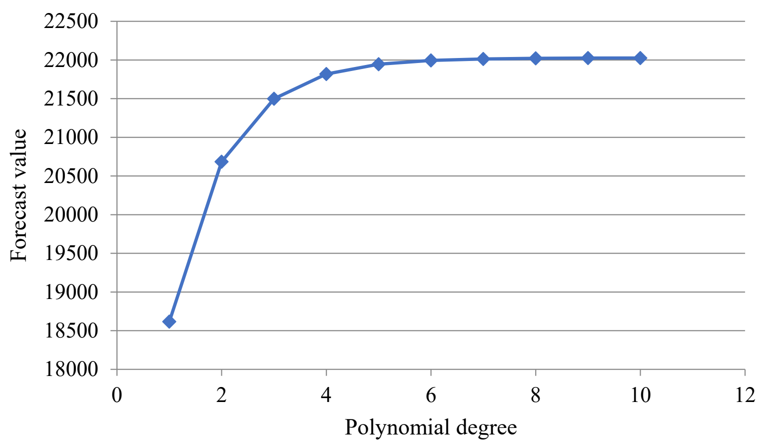

Fig. 3. Extrapolation points based on interpolation polynomials of degrees 1, 2, ..., respectively (values of the seventh column of Table 4)

5. Research results of the forecasting algorithm based on the optimal choice of the power of the polynomial

5. 1. The method of finding the predictive value for one step based on a polynomial of arbitrary degree without finding the polynomial coefficients

Statement 1. Let's have a uniform grid with a step Δ, the values of some function $f_i=f(i\Delta)$, $i=n, n-1, ..., n-m$ are known. Then the predictive value of the function constructed on the basis of the interpolation Newton polynomial of the second form $P_2^N(x)$ at the point $(n+1)\Delta$ can be found on the basis of the relation:

$$P_2^m\left((n+1)\Delta\right)=\sum_{k=1}^{m}(-1)^{k-1}C_m^k f_{n-k+1}. \qquad (1)$$

Thus, only the sign-changing sum of products of binomial coefficients and time series values is used to calculate the forecast value for one step for any polynomial.

Newton's interpolation polynomials of the second kind [2] are written in the form:

$$P_2^N(x)=f_n+\Delta^1 f_{n-1}(x-x_n)+$$
$$+\Delta^2 f_{n-2}(x-x_{n-1})(x-x_n)+\ldots+$$
$$+\Delta^N f_{n-N}(x-x_{n-N+1})\ldots(x-x_n).$$

where

$$\Delta^j f_i=\frac{\Delta^{j-1}f_{i+1}-\Delta^{j-1}f_i}{x_{i+j}-x_i}, \qquad (2)$$

$$\Delta^0 f_i=f_i.$$

The corresponding result can be obtained by analyzing the table of divided differences. Let's suppose that for some k the condition is fulfilled:

$$\Delta^k f_{n-k+1}=\Delta^k f_{n-k}. \qquad (3)$$

In this case, the predicted value can be found by the following obvious calculation procedure:

$$\Delta^{i-1}f_{n-i+2}=\Delta^i f_{n-i+1}(x_{n+1}-x_{n+1-i})+\Delta^i f_{n-i}, \qquad (4)$$

$$i=\overline{k,0}.$$

Let $x=x_{n+1}$. Taking (1) into account:

$$\Delta^k f_{n-k}=\Delta^k f_{n-k+1}=\frac{\Delta^{k-1}f_{n-k+2}-\Delta^{k-1}f_{n-k+1}}{x-x_{n-k+1}}.$$

From here it is easy to get the ratio:

$$f_{n+1}=\Delta^k f_{n-k+1}\left(x-x_{n+k+1}\right)\left(x-x_{n-k+2}\right)...\left(x-x_n\right)+$$
$$+\Delta^{k-1} f_{n-k+1}\left(x-x_{n-k+2}\right)\left(x-x_{n-k+3}\right)...\left(x-x_n\right)+...+f_n .$$

Then, under condition (2), the relation holds:

$$f_{n+1}=\Delta^k f_{n-k}\left(x-x_{n+k+1}\right)\left(x-x_{n-k+2}\right)...\left(x-x_n\right)+$$
$$+\Delta^{k-1} f_{n-k+1}\left(x-x_{n-k+2}\right)\left(x-x_{n-k+3}\right)...\left(x-x_n\right)+...+f_n . \qquad (5)$$

Thus, when condition (3) is fulfilled, the predictive value of the function f_{n+1} at some point $x=x_{n+1}$ is determined as the value of the Newton polynomial of the second form (backward interpolation formula) of degree i, which is equivalent to the calculation procedure (4).

An example of the calculation of split differences is shown in Fig. 4. Here, the grid step is equal to 1.

It is easy to see that in the case when the grid step is equal to 1, the relation holds:

$$\Delta^i f_j=\frac{\sum_{k=0}^{i}(-1)^k C_i^k f_{j+i-k}}{i!}. \qquad (6)$$

In the case of an arbitrary uniform grid with step Δ:

$$\Delta^i f_j=\frac{\sum_{k=0}^{i}(-1)^k C_i^k f_{j+i-k}}{i!\delta^i}. \qquad (7)$$

The corresponding ratio is proved by the method of mathematical induction on the index i. Indeed, for $i=1$, relation (6) is obvious. Let it be fulfilled for some n.

Then:

$$\Delta^{n+1} f_j=\frac{\Delta^n f_{j+1}-\Delta^n f_j}{n+1}\Delta^i f_j=$$
$$=\frac{1}{n+1}\left(\frac{\sum_{k=0}^{n}(-1)^k C_n^k f_{j+1+n-k}}{n!}-\frac{\sum_{k=0}^{n}(-1)^k C_n^k f_{j+n-k}}{n!}\right)=$$
$$=\frac{1}{(n+1)!}\left(\sum_{k=0}^{n}(-1)^k C_n^k f_{j+1+n-k}-\sum_{k=0}^{n}(-1)^k C_n^k f_{j+n-k}\right)=$$
$$=\frac{1}{(n+1)!}\left(\sum_{k=0}^{n}(-1)^k C_n^k f_{j+1+n-k}-\sum_{k=0}^{n}(-1)^k C_n^k f_{j+n-k}\right)=$$
$$=\frac{1}{(n+1)!}\left(\begin{array}{l} f_{j+1+n}+\sum_{k=1}^{n}(-1)^k C_n^k f_{j+1+n-k}- \\ -\sum_{k=0}^{n}(-1)^k C_n^k f_{j+n-k}\end{array}\right)=$$
$$=\frac{1}{(n+1)!}\left(\begin{array}{l} f_{j+1+n}+\sum_{k=0}^{n-1}(-1)^{k+1} C_n^{k+1} f_{j+n-k}- \\ -\sum_{k=0}^{n-1}(-1)^k C_n^k f_{j+n-k}-(-1)^n f_j\end{array}\right)=$$
$$\frac{1}{(n+1)!}\left(\begin{array}{l} f_{j+1+n}-\sum_{k=0}^{n-1}(-1)^{k} C_n^{k+1} f_{j+n-k}- \\ -\sum_{k=0}^{n-1}(-1)^k C_n^k f_{j+n-k}-(-1)^n f_j\end{array}\right)=$$
$$=\frac{1}{(n+1)!}\left(f_{j+1+n}-\sum_{k=0}^{n-1}(-1)^k C_{n+1}^{k+1} f_{j+n-k}-(-1)^n f_j\right)=$$
$$=\frac{1}{(n+1)!}\left(f_{j+1+n}-\sum_{k=1}^{n}(-1)^{k-1} C_{n+1}^{k} f_{j+n-k+1}-(-1)^n f_j\right)=$$
$$=\sum_{k=0}^{n+1}(-1)^k C_{n+1}^k f_{j+n+1-k}.$$

6	f_6				
		$\Delta^1 f_5$			
5	f_5		$\Delta^2 f_4$		
		$\frac{f_5-f_4}{1}$		$\Delta^3 f_3$	
4	f_4		$\frac{f_5-2f_4+f_3}{1*2}$		$\Delta^4 f_2 = =\frac{f_5-4f_4+6f_3-4f_2+f_1}{1*2*3*4}$
		$\frac{f_4-f_3}{1}$		$\frac{f_5-3f_4+3f_3-f_2}{1*2*3}$	
3	f_3		$\frac{f_4-2f_3+f_2}{1*2}$		$\frac{f_5-4f_4+6f_3-4f_2+f_1}{1*2*3*4}$
		$\frac{f_3-f_2}{1}$		$\frac{f_4-3f_3+3f_2-f_1}{1*2*3}$	
2	f_2		$\frac{f_3-2f_2+f_1}{1*2}$		
		$\frac{f_2-f_1}{1}$			
1	f_1				

Fig. 4. Illustration of the construction of split differences

If to use assumption (3) and construct the predicted value by procedure (4), it is easy to see that the predicted value is simply the sum of the corresponding finite differences from the previous diagonal of the table, multiplied by the corresponding values. In particular, for the corresponding values from the Table 2:

$$\Delta^3 f_3=4\frac{f_5-4f_4+6f_3-4f_2+f_1}{2*3*4}+$$
$$+\frac{f_5-3f_4+3f_3-f_2}{2*3}=$$
$$=\frac{2f_5-7f_4+9f_3-5f_2+f_1}{2*3},$$

$$\Delta^2 f_4=3\frac{2f_5-7f_4+9f_3-5f_2+f_1}{2*3}+$$
$$+\frac{f_5-2f_4+f_3}{2}=$$
$$=\frac{3f_5-9f_4+10f_3-5f_2+f_1}{2},$$

$$\Delta^1 f_5=2\frac{3f_5-9f_4+10f_3-5f_2+f_1}{2}+$$
$$+f_5-f_4=4f_5-10f_4+10f_3-5f_2+f_1,$$

$$f_6=f_5+\Delta^1 f_5=f_5+4f_5-10f_4+10f_3-$$
$$-5f_2+f_1=5f_5-10f_4+10f_3-5f_2+f_1.$$

Or, on another note,

$$f_6 = 1\left(2\left(3\left(4\Delta^4 f_1 + \Delta^3 f_2\right) + \Delta^2 f_3\right) + \Delta^1 f_4\right) + f_5 =$$
$$= 5f_5 - 10f_4 + 10f_3 - 5f_2 + f_1.$$

As can be seen, the forecast value is the sum of all rows of the «significant variable» Pascal triangle (Fig. 5), the elements of which are multiplied by the corresponding values of the function.

					1					
				1		-1				
			1		-2		1			
		1		-3		3		-1		
	1		-4		6		-4		1	
1		5		-10		10		-5		1

Fig. 5. A fragment of Pascal's modified triangle

In the general case, the following recurrence relation can be written:

$$\phi_1 = \Delta^{n-1} f_1,$$

$$\phi_i = (n-i+1)\phi_{i-1} + \Delta^{n-i} f_i, i = 2,3,...n,$$

$$f_{n+1} = \phi_n.$$

Then, taking into account the known combinatorial relation, Proposition 1 is obtained.

5. 2. Condition of efficiency of extrapolation based on the arithmetic mean of polynomial forecasts

According to relation (1):

$$P_2^m\left((n+1)\Delta\right) = \sum_{k=1}^{m}(-1)^{k-1} C_m^k f_{n-k+1},$$

$$f_{n+1}^m = \sum_{k=1}^{m}(-1)^{k-1} C_m^k f_{n-k+1} = C_m^1 f_n + \sum_{k=2}^{m}(-1)^{k-1} C_m^k f_{n-k+1} =$$
$$= C_m^1 f_n + \sum_{k=2}^{m}(-1)^{k-1} C_m^k f_{n-k+1} -$$
$$-\sum_{k=2}^{m}(-1)^{k} C_m^k f_{n-k+2} + \sum_{k=2}^{m}(-1)^{k} C_m^k f_{n-k+2} =$$
$$= C_m^1 f_n + \sum_{k=2}^{m}(-1)^{k-1} C_m^k\left(f_{n-k+1} - f_{n-k+2}\right) - \sum_{i=1}^{m-1} f_{n+1}^i.$$

From here:

$$0 = C_m^1 f_n + \sum_{k=2}^{m}(-1)^{k-1} C_m^k\left(f_{n-k+1} - f_{n-k+2}\right) - \sum_{i=1}^{m} f_{n+1}^i,$$

$$0 = C_m^1 f_{n+1} + \sum_{k=1}^{m}(-1)^{k-1} C_m^k\left(f_{n-k+1} - f_{n-k+2}\right) - \sum_{i=1}^{m} f_{n+1}^i,$$

$$f_n = \frac{1}{m}\sum_{i=1}^{m} f_{n+1}^i - \frac{1}{m}\sum_{k=2}^{m}(-1)^{k-1} C_m^k\left(f_{n-k+1} - f_{n-k+2}\right),$$

$$f_{n+1} = \frac{1}{m}\sum_{i=1}^{m} f_{n+1}^i - \frac{1}{m}\sum_{k=1}^{m}(-1)^{k-1} C_m^k\left(f_{n-k+1} - f_{n-k+2}\right) =$$
$$= \frac{1}{m}\sum_{i=1}^{m} f_{n+1}^i - \frac{1}{m}\sum_{k=2}^{m}(-1)^{k-1} C_m^k\left(f_{n-k+1} - f_{n-k+2}\right) - \left(f_n - f_{n+1}\right) =$$
$$= \frac{1}{m}\sum_{i=1}^{m} f_{n+1}^i - \frac{1}{m}\sum_{k=2}^{m}(-1)^{k-1} C_m^k\left(f_{n-k+1} - f_{n-k+2}\right) - \left(f_n - f_{n+1}\right).$$

The following extrapolation method directly follows from the last relation. Let $f_{n+1}^i = P_2^i\left((n+1)\Delta\right)$. Then let's find the forecast in the form:

$$f_{n+1} = \frac{1}{\widehat{m}}\sum_{i=1}^{\widehat{m}} f_{n+1}^i, \tag{8}$$

where:

$$\widehat{m} = \arg\min_m \left| \frac{1}{m}\left(\begin{array}{l} \sum_{k=2}^{m}(-1)^{k-1} C_m^k\left(f_{n-k+1} - f_{n-k+2}\right) + \\ + \sum_{i=1}^{m} f_{n+1}^i \end{array} \right) - f_n \right|. \tag{9}$$

At the same time, the estimate of the absolute error of extrapolation is determined by the expression:

$$error = \left| \frac{1}{m}\left(\begin{array}{l} \sum_{k=2}^{m}(-1)^{k-1} C_m^k\left(f_{n-k+1} - f_{n-k+2}\right) + \\ + \sum_{i=1}^{m} f_{n+1}^i \end{array} \right) - f_n \right|. \tag{10}$$

At the same time, in practice, it is also possible to use a relative error of the form:

$$verror =$$
$$= \left| \frac{1}{m}\left(\begin{array}{l} \sum_{k=2}^{m}(-1)^{k-1} C_m^k\left(\begin{array}{l} f_{n-k+1} - \\ -f_{n-k+2} \end{array} \right) + \\ + \sum_{i=1}^{m} f_{n+1}^i \end{array} \right) - f_n \right| / \frac{1}{m}\sum_{i=1}^{m} f_{n+1}^i. \tag{11}$$

Expression (11) allows to estimate the applicability limits of this method. The method is effective only for small values of expression (11).

5. 3. General algorithm for finding the optimal power of an extrapolation polynomial

The above theoretical results and numerical studies make it possible to build an algorithm for improving polynomial forecasts based on the identification of the optimal power of the extrapolation polynomial. The essence of the algorithm is to perform the following steps:

1. An array of experimental data is formed.
2. Based on the experimental data, all possible polynomial predictive values are found based on the ratio (1).
3. The obtained sequence of forecast values is analyzed for the presence of a tendency towards convergence (the distance between neighboring forecast points monotonically decreases or increases).
4. If there is a tendency to convergence, the prediction of the polynomial of the maximum possible degree is selected as the predictive value, and the algorithm is completed.
5. In the absence of monotonicity of the sequence of forecasts, the ratio (10) is calculated and the relative error is estimated (11).
6. If the relative error (11) is small (the level is defined as an expert assessment), a forecast is constructed as an arithmetic average of polynomial forecasts and the algorithm is completed. Otherwise, it is impossible to construct a refinement of the polynomial forecast.

The practical significance of research is confirmed by the following example. The input data are data on the nominal gross domestic product (GDP) of Ukraine from 2012 to 2020 [16]. The corresponding data are given in Table 5.

Table 5

Data of the nominal GDP of Ukraine

Year	2012	2013	2014	2015	2016	2017	2018	2019	2020
GDP (million USD)	175781	183310	131805	90615	93270	112154	130832	153781	155582

According to paragraph 2 of the algorithm, all possible polynomial forecasts are built based on data from 2012 to 2019 using formula (1). At the same time, the following results can be obtained, starting with a polynomial of the 7$^{\text{th}}$ degree: –6725, 153525, 216121, 206390, 185478, 181001, 176730. Checking condition (11) confirms the possibility of calculating the forecast as an arithmetic mean, which is equal to 158931. As it is possible to see, the forecast is minimal deviates from the real value of GDP in 2020, which is equal to 155582, which confirms the effectiveness of the algorithm. At the same time, the forecast based on, for example, a polynomial of the 7$^{\text{th}}$ degree has a significant error and gives a negative value, which has no economic meaning.

6. Discussion of the research results of the automation algorithm for choosing the optimal degree of a polynomial

On the basis of Theorem 1, it is possible to construct predictive values for polynomials of arbitrary degree. At the same time, it is easy to make sure that all the considerations given in Section 5. 1 are valid when the grid step is arbitrary.

In practice, to calculate the predictive value for large degrees of polynomials, the corresponding binomial coefficients should be calculated exactly according to Pascal's triangle in order to avoid the calculation of factorials of large numbers in the classical definition of binomial coefficients.

Obtaining a fast method for calculating the forecast value became possible due to the discovery of new constructive procedures for building a forecast based on Newton polynomials of the second kind (3). Also, the constructive procedure for calculating split differences (6), (7) became the basis of the obtained result. An interesting result of the work is the use of the modified Pascal's triangle to calculate the predictive value shown in Fig. 1, which reveals another still unknown property of Pascal's triangle.

The use of a fast method of finding predictive polynomial values in the general algorithm is a feature of this approach, since with the classical approach, it will be necessary to solve a sequence of systems of linear algebraic equations with a constant increase in dimension to find the coefficients, which will be a very cumbersome computational procedure.

Numerical studies for test functions have been carried out in the work, which allow to confirm the obvious statement: provided there is convergence, the optimal forecast can be chosen based on the polynomial of the maximum degree. Such a situation will arise, for example, when the function has an exponential nature of growth. Confirmation of this statement can be seen, in particular, in Fig. 3, 4. Of course, convergence can also occur in other cases.

The advantage of the proposed method is achieved precisely due to finding the optimal degree of the polynomial in comparison with other forecasting methods based on polynomials. This is illustrated, in particular, in Table 3, which shows how wide the uncertainty interval can be when using forecasts based on polynomials of different degrees. The relative error can reach more than 50 % for the selected test function, while extrapolation based on the averaged polynomial prediction gave about 5 %.

It should be noted that when estimating the error in the corresponding ratio (10), the unknown value was actually replaced by an averaged polynomial forecast. And therefore, there is also a certain error in the performance of the identity, which is not considered in this work. That is, the error estimate (10) is actually approximate. However, this does not exclude the use of the proposed method in practice.

A forecast based on the arithmetic mean of polynomial forecasts is part of the general algorithm for choosing the optimal degree of a polynomial for extrapolation and significantly expands its capabilities. After all, in practice, it is quite difficult to obtain a situation, for example, when there is a convergence of polynomial forecasts due to the presence of a stochastic component in real processes. In the future, it makes sense to further improve the forecasting methodology offered in combination with other methods.

The proposed method differs from the existing polynomial forecasts given, in particular, in works [1–7] in that the optimal degree of the polynomial is selected automatically. However, such a choice of power is not always possible, but only in the cases described above (when the relative error (11) is small or there is convergence). This is a limitation of the method. Among the shortcomings of the proposed method, it is possible to single out insufficient efficiency in the study of processes with a stochastic component, which is inherent in any forecasts based on interpolation. This problem may be the subject of future research and may be solved by a combination of different techniques. It is also advisable to improve the choice of the polynomial of the optimal degree in cases not provided for in this work.

7. Conclusions

1. A detailed analysis and justification of the method for calculating the predictive value based on a polynomial of arbitrary degree was carried out, which is distinguished by the absence of a procedure for solving a system of linear algebraic equations for finding the coefficients of the polynomial. The use of such a method has significant advantages due to the absence of a procedure for finding coefficients and the possibility of building forecasts based on polynomials of large degrees. When making forecasts based on Theorem 1, the main computational complexity is the calculation of binomial coefficients, which can easily be performed on the basis of Pascal's triangle.

2. On the basis of the optimized procedure for building a polynomial forecast, it is proposed to find the extrapolated value as the arithmetic mean of polynomial forecasts. At the same time, a ratio was found to assess the accuracy of this forecasting method and the limits of its application were determined. Numerical experiments using test functions confirmed the effectiveness of this method. In particular, for the test functions, the relative error of extrapolation using polynomials of arbitrary powers reached 50 %, while

extrapolation based on the averaged polynomial forecast gave a relative error that did not exceed 5 %.

3. A general algorithm for constructing a predictive value based on the procedure for finding the optimal power of an extrapolation polynomial is formulated. The corresponding algorithm includes the analysis of forecast values based on polynomials of different degrees and finding conditions when the optimal forecast can be built on the basis of a polynomial of the maximum degree. Numerical studies have demonstrated cases when polynomial forecasts converge and, accordingly, the choice of a polynomial of the maximum possible degree as the optimal forecast. Also, if the relevant conditions are met, the forecast is constructed as the arithmetic average of polynomial forecasts, which is equivalent to choosing the optimal degree of the polynomial for extrapolation based on the minimum deviation of the extrapolation value and the arithmetic average of all forecasts.

Conflict of interest

The authors declare that they have no conflict of interest in relation to this study, including financial, personal, authorship, or any other, that could affect the study and its results presented in this article.

Financing

The study was conducted without financial support.

Data availability

The manuscript has no associated data.

References

1. Brezinski, C., Redivo-Zaglia, M. (2020). Extrapolation and Rational Approximation: The Works of the Main Contributors. Springer. doi: https://doi.org/10.1007/978-3-030-58418-4
2. Israfilov, D. M., Testici, A. (2018). Some Inverse and Simultaneous Approximation Theorems in Weighted Variable Exponent Lebesgue Spaces. Analysis Mathematica, 44 (4), 475–492. doi: https://doi.org/10.1007/s10476-018-0403-x
3. Bomba, A. Ya., Turbal, Y. V. (2015). Data Analysis Method and Problems of Identification of Trajectories of Solitary Waves. Journal of Automation and Information Sciences, 47 (10), 13–23. doi: https://doi.org/10.1615/jautomatinfscien.v47.i10.20
4. Guliyev, V. S., Ghorbanalizadeh, A., Sawano, Y. (2018). Approximation by trigonometric polynomials in variable exponent Morrey spaces. Analysis and Mathematical Physics, 9 (3), 1265–1285. doi: https://doi.org/10.1007/s13324-018-0231-y
5. Hyndman, R. J., Kostenko, A. V. (2007). Minimum sample size requirements for seasonal forecasting model. FORESIGHT, 6. Available at: https://robjhyndman.com/papers/shortseasonal.pdf
6. Turbal, Y., Bomba, A., Turbal, M., Alkaleg Hsen Drivi, A. (2021). Some aspects of extrapolation based on interpolation polynomials. Physico-Mathematical Modelling and Informational Technologies, 33, 175–180. doi: https://doi.org/10.15407/fmmit2021.33.175
7. Shalaginov, A. V. (2011). Kubicheskaya splayn ekstrapolyatsiya vremennykh ryadov. International conference on System Analysis and Information Technologies SAIT 2011, Institute for Applied System Analysis of National Technical University of Ukraine. Kyiv. Available at: https://cad.kpi.ua/attachments/141_2011_025s.pdf
8. Kostinsky, A. S. (2014). On the principles of a spline extrapolation concerning geophysical data. Reports of the National Academy of Sciences of Ukraine, 2, 111–117. doi: https://doi.org/10.15407/dopovidi2014.02.111
9. Makridakis, S., Bakas, N. (2016). Forecasting and uncertainty: A survey. Risk and Decision Analysis, 6 (1), 37–64. doi: https://doi.org/10.3233/rda-150114
10. Zhan, Z., Fu, Y., Yang, R.-J., Xi, Z., Shi, L. (2012). A Bayesian Inference based Model Interpolation and Extrapolation. SAE International Journal of Materials and Manufacturing, 5 (2), 357–364. doi: https://doi.org/10.4271/2012-01-0223
11. Demiris, N., Lunn, D., Sharples, L. D. (2011). Survival extrapolation using the poly-Weibull model. Statistical Methods in Medical Research, 24 (2), 287–301. doi: https://doi.org/10.1177/0962280211419645
12. Turbal, Y., Bomba, A., Turbal, M., Sokh, A., Radoveniuk, O. (2019). Pyramidal method of extrapolation for short time series. International Journal of Computing Science and Mathematics, 10 (6), 525. doi: https://doi.org/10.1504/ijcsm.2019.104025
13. Monroe, J. I., Hatch, H. W., Mahynski, N. A., Shell, M. S., Shen, V. K. (2020). Extrapolation and interpolation strategies for efficiently estimating structural observables as a function of temperature and density. The Journal of Chemical Physics, 153 (14). doi: https://doi.org/10.1063/5.0014282
14. Wang, L.-Y., Lee, W.-C. (2015). One-step extrapolation of the prediction performance of a gene signature derived from a small study. BMJ Open, 5 (4), e007170–e007170. doi: https://doi.org/10.1136/bmjopen-2014-007170
15. Bakas, N. P. (2019). Numerical Solution for the Extrapolation Problem of Analytic Functions. Research, 2019. doi: https://doi.org/10.34133/2019/3903187
16. Valovyi vnutrishniy produkt (VVP) v Ukraini 2023. Ministerstvo finansiv Ukrainy. Available at: https://index.minfin.com.ua/ua/economy/gdp/