UDC 338.27
Doi: 10.31772/2587-6066-2020-21-1-34-40

For citation: Shiryaeva T. A., Khlupichev V. A., Shlepkin A. K., Melnikova O. L. The use of the inverse transformation method for time series analysis. *Siberian Journal of Science and Technology.* 2020, Vol. 21, No. 1, P. 34–40. Doi: 10.31772/2587-6066-2020-21-1-34-40

Для цитирования: Ширяева Т. А., Хлупичев В. А., Шлепкин А. К., Мельникова О. Л. Метод обратного преобразования для анализа временных рядов // Сибирский журнал науки и технологий. 2020. Т. 21, № 1. С. 34–40. Doi: 10.31772/2587-6066-2020-21-1-34-40

THE USE OF THE INVERSE TRANSFORMATION METHOD FOR TIME SERIES ANALYSIS

T. A. Shiryaeva[1], V. A. Khlupichev[1], A. K. Shlepkin[1*], O. L. Melnikova[2]

[1]Krasnoyarsk State Agrarian University
90, Mira Av., Krasnoyarsk, 660001, Russian Federation
[2]Khakas State University
90, Lenin Av., Abakan, 655017, Russian Federation
[*]E-mail:ak_kgau@mail.ru

In modern conditions of technology development, signs of systemacity are manifested to one degree or another in all areas, so the use of system analysis is an urgent task. In this case, the main factors in this situation are data processing and prediction of the state of a system. Mathematical modeling is used as a prediction method for a given subject area. A mathematical model is a universal tool for describing complex systems representing the approximate description of the class of phenomena of the external world expressed by mathematical concepts and language. The mathematical model can be represented as a set of systematic components and a random component. In this article, the object of prediction is the irregular random component of a model, which reflects the impact of numerous random factors. The origin, nature and laws of variation of the random variable are known, therefore, to simulate its behavior or predict its future value, one needs high degree of certainty to establish the form of continuous distribution function of the random variable. The empirical distribution function is calculated using the sample of random variable values. This empirical function is close to the values of the desired unknown function of distribution. The resulting empirical function is discrete, therefore it is necessary to apply piecewise linear interpolation to obtain a continuous distribution function. The predicted random component of time series has been included in the initial regression model. In order to compare augmented and initial regression models, several values were excluded from the time series and new prediction was built. The value of the average approximation error for assessing the quality of the model is calculated. The augmented regression model proved to be more effective than the original one.

Keywords: forecasting, time series analysis, inverse transformation, system analysis.

МЕТОД ОБРАТНОГО ПРЕОБРАЗОВАНИЯ ДЛЯ АНАЛИЗА ВРЕМЕННЫХ РЯДОВ

Т. А. Ширяева[1], В. А. Хлупичев[1], А. К. Шлепкин[1*], О. Л. Мельникова[2]

[1]Красноярский государственный аграрный университет
Российская Федерация, 660001, Красноярск, просп. Мира, 90
[2]Хакасский государственный университет имени М. Ф. Катанова
Российская Федерация, 655017, Абакан, просп. Ленина, 90
[*]E-mail:ak_kgau@mail.ru

В современных условиях развития технологий признаки системности проявляются в той или иной степени во всех областях, поэтому использование системного анализа является актуальной задачей. При этом главными факторами в данной ситуации являются обработка данных и прогнозирование состояния системы. Для заданного объекта в качестве способа прогнозирования в данной работе применяется моделирование, а точнее математическое моделирование. Математическая модель – это универсальное средство исследования сложных систем, представляющее собой приближенное описание какого-либо класса явлений внешнего мира, выраженное с помощью математической символики.

Математическую модель можно представить как совокупность систематических компонентов и случайной составляющей. В данной статье регрессионная модель уже определена, а в качестве объекта прогнозирования рассмотрена остаточная нерегулярная компонента модели, которая отражает воздействие многочисленных факторов случайного характера.

Происхождение, природа и законы изменения данной случайной величины нам неизвестны, поэтому для моделирования ее поведения или предсказания ее будущих значений, необходимо с высокой степенью достоверности установить вид непрерывной функции распределения данной случайной величины.

Для этого была рассчитана эмпирическая функция распределения с помощью выборки из значений случайной величины. Данная эмпирическая функция в определенной степени приближена к значениям искомой неизвестной функции распределения. Полученная эмпирическая функция носит дискретный характер, поэтому необходимо применить кусочно-линейную интерполяцию и, таким образом, получить непрерывную функцию распределения.

В исходную регрессионную модель была включена спрогнозированная случайная компонента временного ряда. Для того чтобы сравнить дополненную и исходную регрессионные модели, из динамического ряда были исключены несколько значений и построен новый прогноз. Рассчитано значение средней ошибки аппроксимации для оценки качества модели. Дополненная регрессионная модель показала себя эффективнее исходной.

Ключевые слова: прогнозирование, анализ временных рядов, обратное преобразование, системный анализ.

Introduction. For the specialists involved in data analysis, in most cases prediction may be said to be the main goal and task. Modern methods of statistical forecasting are often able to predict almost any possible indicators with high accuracy [1].

Forecasting is a system of scientifically based ideas about the possible conditions of an object in the future and alternative ways of its development [2]. There are no universal prediction methods for all occasions. Any practical forecasting problem can be satisfactorily solved only by a limited number of methods [3]. The choice of a forecasting method and its effectiveness depend on many conditions: the purpose of the forecast, the period of its lead prediction, the level of detailing and the availability of initial information [4]. The most commonly used forecasting method is mathematical modeling. A mathematical model is an approximate description of a specific process or phenomenon of the external world, expressed using a mathematical apparatus [5].

The components of the time series. Commonly, when studying a time (dynamic) series, it is depicted in the form of the following mathematical model:

$$Y_t = \hat{Y}_t + E_t$$

where Y_t – time series value; $\hat{Y}_t$ – systematic (deterministic) component of the time series; E_t – random component of the time series [6].

The systematic component of the time series Y_t is a result of the influence of constantly acting factors on the process being analyzed. Two main systematic components of the time series can be distinguished:

1. The trend of the time series.
2. The cyclic oscillations of the time series.

A trend is a general pattern of change in the indicators of the time series, stable and observed over a long period of time. A trend is described using some function, usually monotonic. This function is called 'trend function', or simply “trend” [7].

Among the factors that form the cyclical oscillations of the series, in turn, two components can be distinguished:

1) seasonality;
2) cyclicity.

Seasonality is a result of the influence of factors acting at a predetermined periodicity. These are regular fluctuations that are periodic in nature ending within a year. The cyclic component is a nonrandom function that describes long (more than a year) periods of rise and fall [8; 9]

The random component of the time series E_t is the component of the time series remaining after the allocation of systematic components. It reflects the effects of numerous random factors; it is a random, irregular component.

Random variables are diverse in nature and origin, although the distribution law can be written in a uniform universal form, namely, in the form of a distribution function that is equally suitable for discrete and continuous random variables [10].

Inverse Transformation Method. For forecasting purposes, as well as simulation, one may need a method for generating the random component of a time series. For this purpose, we use the inverse transformation method.

Let the random variable X have the distribution function $F(x)$. We assume that $F^{-1}(x)$ is the inverse function of $F(x)$. Then the algorithm for generating the random variable X with the distribution function $F(x)$ will be the following:

1. Generate the value U having the uniform distribution over the interval (0;1);
2. Return $X = F^{-1}(U)$.

Fig. 2 depicts this algorithm graphically; the random variable corresponding to this distribution function can take either positive or negative values; it depends on the specific value of U. In Fig. 1, the random number U_1 gives the positive value of the random variable X_1 as a result, while the random number U_2 gives the negative value of the random variable X_2 as a result [11].

Estimating the distribution function of a random variable. Let us consider the time series as a sequence $x_1, x_2, \ldots, x_n$ of independent and equally distributed, according to a certain law, random variables; this sequence is called the sample of volume n. Each $x_t (t = 1, 2, ..., n)$ is called variation. Having the sample, we do not have information about the form of the distribution function $F(x)$. It is required to construct an estimate (approximation) for this unknown function.

The most preferred estimate of the function $F(x)$ will be the empirical distribution function $F_n(x)$.

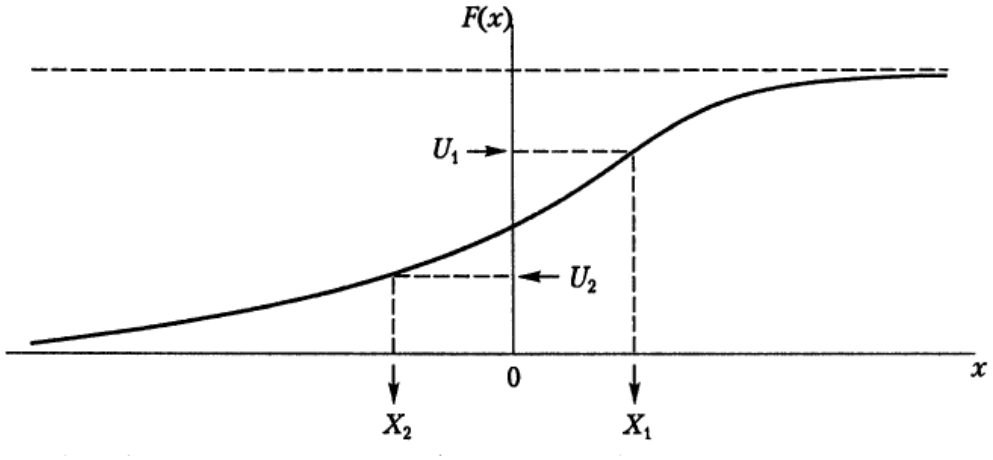

Fig. 1. Using the Inverse Transformation Method to generate a random variable

Рис. 1. Использование метода обратного преобразования
для генерирования случайной величины

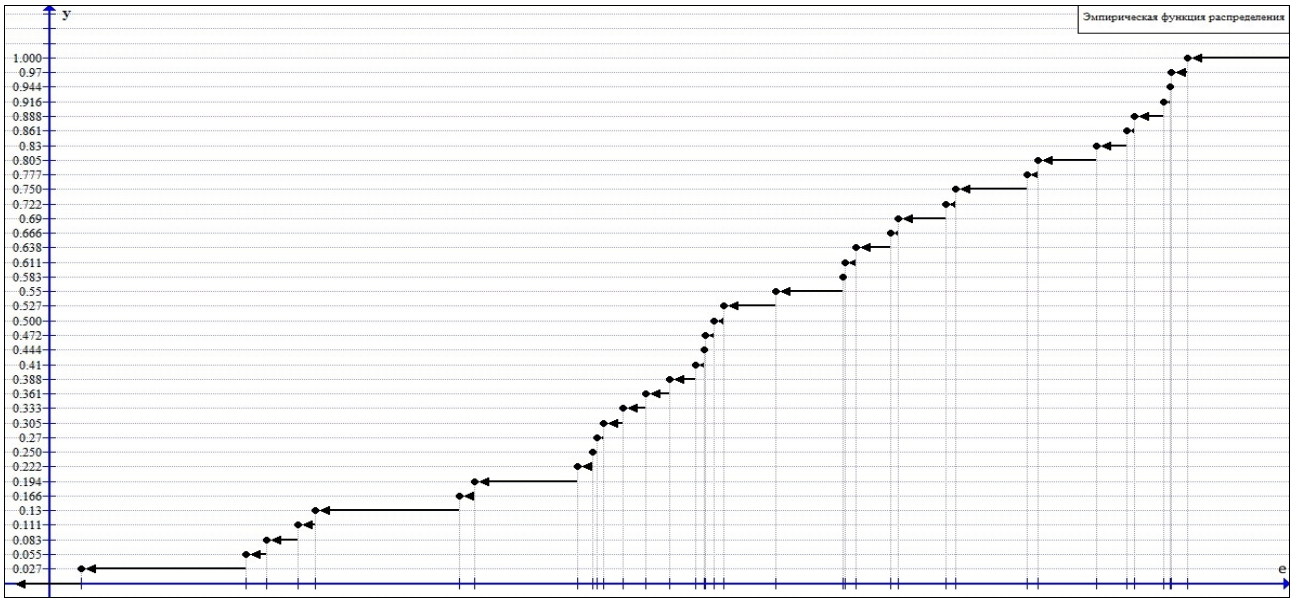

Fig. 2. Graph of the empirical distribution function $F_n(e)$

Рис. 2. График эмпирическая функция распределения $F_n(e)$

The empirical distribution function (sampling distribution function) is the function $F_n(x)$, which determines the relative frequency of the event $X < x$ for each x value, i. e.

$$F_n(x) = \frac{n_x}{n},$$

where n_x – the number of x_t values, less than x; n – sample size.

With a sufficiently large sample size, the functions $F_n(x)$ and $F(x) = P(X < x)$ differ insignificantly from each other.

The difference between the empirical distribution function and the theoretical one is that the theoretical distribution function determines the probability of the event $X<x$, and the empirical function determines the relative frequency of the same event [12].

The empirical distribution function has all the properties of the integral distribution function:

1) the values of the empirical distribution function belong to the interval [0; 1];

2) $F_n(x)$ is non-decreasing function;

3) $F_n(x) = 0$ at $x \le x_{\min}$, if $x_{\min}$ is the smallest variation;

$F_n(x) = 1$ at $x > x_{\min}$, if $x_{\min}$ is the largest variation [6].

However, to use the inverse transformation method, it is convenient to have a continuous distribution function; therefore, it is necessary to interpolate the obtained empirical function.

Building a predicted model. We give an example of using the inverse transformation method when constructing a predicted model. As the initial data, we use the average monthly indicators of power consumption in the Krasnoyarsk Territory for 3 years from January 2009 to December 2011 [13].

Using some regression model, the forecast values of the time series were calculated. Actual Y_t and predicted $\hat{Y}_t$ values of the time series are presented in tab. 1

We build the empirical distribution function of the values of the deviations e_t of the predicted values $\hat{Y}_t$ from the actual values Y_t of the time series $e_t = Y_t - \hat{Y}_t$. To do this, it is necessary to rank the sample $\{e_t\}$, thus obtaining the sample $\{e_{(t)} = \{e_{(1)} < e_{(2)} < ... < e_{(n)}\}$ (tab. 2).

Since the frequency of each variation equals 1, the empirical function will have the following form:

$$F_n(e) = \begin{cases} 0, & e \in \left[-\infty, e_{(1)}\right); \\ ...; & \\ \dfrac{t}{n}, & e \in \left[e_{(t)}, e_{(t+1)}\right); \quad t = 0,1,...,n; \\ ...; & \\ 1, & e \in \left[e_{(n)}, +\infty\right). \end{cases}$$

The graph of the function $F_n(e)$ is shown in fig. 2.

The obtained empirical function $F_n(e)$ has a discrete form. We use piecewise linear interpolation to obtain the continuous distribution function of the random variable $F_n^*(e)$. To do this, we use the equation of a straight line passing through two points:

$$y = (x - x_1) \times \left(\frac{y_2 - y_1}{x_2 - x_1}\right) + y_1.$$

The continuous distribution function of the random variable $F_n^*(e)$ will have the following form:

$$F_n(e) = \begin{cases} 0, & e \in \left(-\infty, e_{(1)}\right); \\ ...; & \\ (e' - e_{(t)}) \times \left(\dfrac{1/(n-1)}{e_{(t+1)} - e_{(t)}}\right) + \dfrac{t-1}{n-1}, & e \in \left[e_{(t)}, e_{(t+1)}\right); \quad t = 0, 1, ..., n; \\ ...; & \\ 1, & e \in \left[e_{(n)}, +\infty\right). \end{cases}$$

Table1

Actual Y_t and predicted $\hat{Y}_t$ values of the time series

t	Y_t	$\hat{Y}_t$	t	Y_t	$\hat{Y}_t$	t	Y_t	$\hat{Y}_t$	t	Y_t	$\hat{Y}_t$
1	51.0123	53.09764	**11**	35.5809	43.42324	**21**	37.8690	44.58914	**31**	17.1468	15.15964
2	38.2345	38.37535	**12**	53.2584	54.26158	**22**	63.4957	57.61664	**32**	20.8548	21.45395
3	40.0023	35.24303	**13**	52.3887	52.69219	**23**	72.9843	72.28322	**33**	29.3791	31.55531
4	25.1288	32.13879	**14**	39.9125	41.30390	**24**	88.0214	83.09426	**34**	51.1710	44.09931
5	22.9338	27.08163	**15**	39.2113	32.14284	**25**	82.6095	79.01463	**35**	61.5869	59.79590
6	27.0146	20.64426	**16**	31.3420	24.38189	**26**	62.7282	63.15378	**36**	71.2594	73.02318
7	25.1154	17.77792	**17**	26.0102	19.52312	**27**	50.0250	48.20592			
8	16.6987	19.20278	**18**	20.5578	17.87433	**28**	29.6211	34.02474			
9	27.3114	23.86244	**19**	12.1214	22,62140	**29**	22.2954	22.75450			
10	29.2400	31.48983	**20**	24.9374	32.44863	**30**	17.8092	15.25792			

Table 2

Range of Values e_t and $e_{(t)}$

t	e_t	$e_{(t)}$	t	e_t	$e_{(t)}$	t	e_t	$e_{(t)}$	t	e_t	$e_{(t)}$
1	–2.085	–10.500	**11**	–7.842	–2,085	**21**	–6.720	1.791	**31**	1.987	6.370
2	–0.141	–7.842	**12**	–1.003	–1.764	**22**	5.879	1.819	**32**	–0.599	6.487
3	4.759	–7.511	**13**	–0.303	–1.391	**23**	0.701	1.987	**33**	–2.176	6.960
4	–7.010	–7.010	**14**	–1.391	–1.003	**24**	4.927	2.551	**34**	7.072	7.068
5	–4.148	–6.720	**15**	7.068	–0.599	**25**	3.595	2.683	**35**	1.791	7.072
6	6.370	–4.404	**16**	6.960	–0.459	**26**	–0.426	3.449	**36**	–1.764	7.337
7	7.337	–4.148	**17**	6.487	–0.426	**27**	1.819	3.595			
8	–2.504	–2.504	**18**	2.683	–0.303	**28**	–4.404	4.759			
9	3.449	–2.250	**19**	–10.500	–0.141	**29**	–0.459	4.927			
10	–2.250	–2.176	**20**	–7.511	0.701	**30**	2.551	5.879			

The graph of the function $F_{36}^{*}(e)$ is shown in fig. 3.

Evaluating the predicted model. For further analysis of this method, let us consider several forecast models [14]:

1. We take the value Y_t of the time series as a completely deterministic process, to carry out the forecast we use the values $\widehat{Y}_t$ calculated using the regression model;

2. The value Y_t of the time series will be taken as a random variable for which we construct the distribution function $F_n^{*}(e)$ and calculate the predicted values Y_t';

3. We will take the value Y_t of the time series as a set of values $\widehat{Y}_t$ calculated using the regression model and the random component e_t, for which we construct the distribution function $F_n^{*}(e)$ and calculate the predicted values e_t'.

Let us make the operational prediction of energy consumption levels. For this purpose, we exclude from consideration the last 5 observations from the sample and we calculate new estimates of the parameters of the regression model, as well as the new distribution functions $F_{31}^{*}(x)$ and $F_{31}^{*}(e)$.

We apply the inverse transformation algorithm to the obtained functions $F_{31}^{*}(x)$ and $F_{31}^{*}(e)$. To this end, we generate the sample $\{u_t\}$ of random numbers having a uniform distribution in the interval [0; 1], and return $Y't = F_{31}^{*-1}(u_t)$ and $e't = F_{31}^{*-1}(u_t)$. The calculation results are presented in tab. 3.

When considering the obtained results, it is clear that the sum $Y_t + e_t'$ is closer to the actual data than the predicted values calculated using the regression model. Thus, the predicted values smoothed out the predicted error to an extent (fig. 4).

As a criterion for assessing the quality of the model, we determine the value of the average approximation error, which is calculated using the formula [15]:

$$A = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{\left(Y_{act} - Y_{pr}\right)}{Y_{act}}\right| \cdot 100\ \%,$$

where Y_{pr} – predicted value of time series; Y_{act} – actual value of time series; n – size of time series [10].

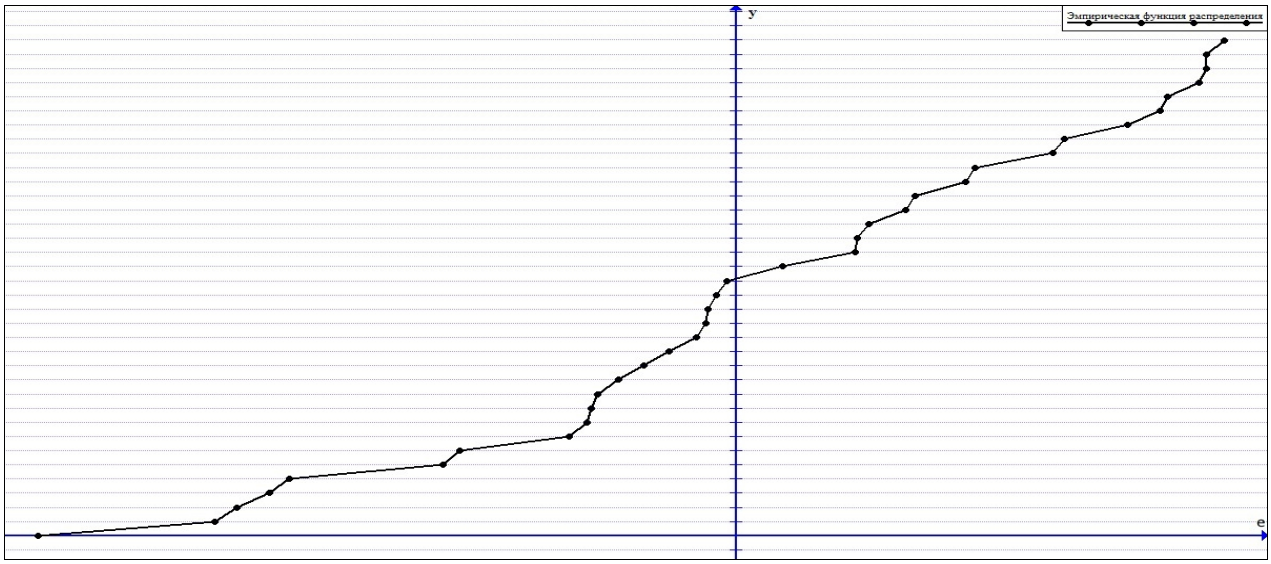

Fig. 3. Graph of the continuous distribution function $F_{36}^{*}(e)$

Рис. 3. График непрерывной функции распределения $F_{36}^{*}(e)$

Table 3

Results of applying the inverse transformation algorithm

№	t	$\widehat{Y}_t$	u_t	e_t'	$Y_t + e_t'$	Y_t'
1	32	25.1456	0.0608	–7.1360	18.0096	6.9693
2	33	37.1619	0.6514	1.9654	39.1274	14.9283
3	34	52.0351	0.6577	2.1085	54.1436	15.2489
4	35	70.5459	0.0515	–7.1507	63.3952	6.8725
5	36	86.8325	0.5448	0.3808	87.2133	13.4902

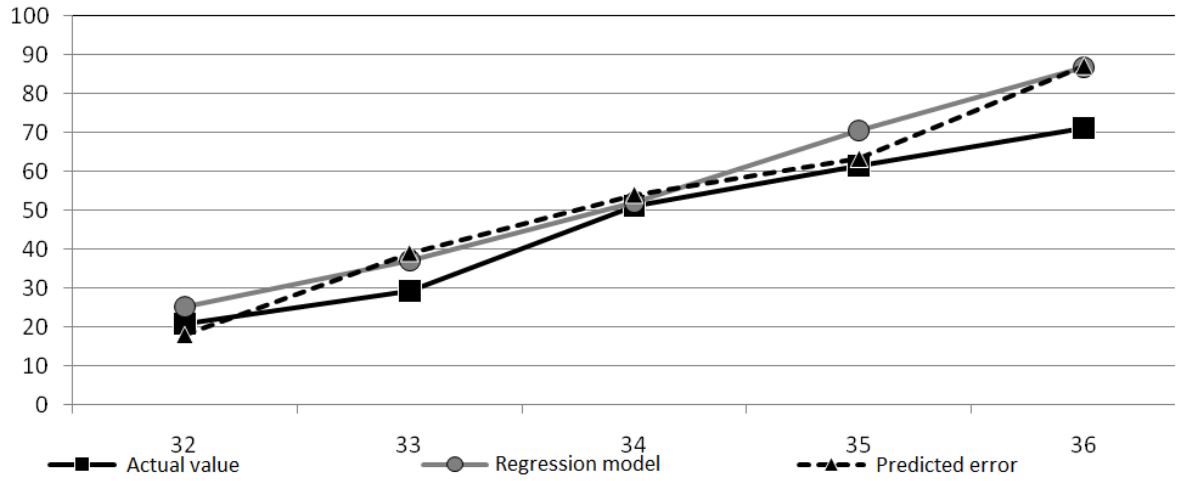

Fig. 4. Graph display of predicted results

Рис. 4. Графическое отображение результатов прогноза

The values of the average approximation error for $\widehat{Y}_t$, Y_t' and $\widehat{Y}_t + e_t'$ are the following:

1) $A(\widehat{Y}_t) \approx 17{,}03\ \%$;

2) $A(Y_t') \approx 71{,}18\ \%$;

3) $A(\widehat{Y}_t + e_t') \approx 15{,}59\ \%$.

The highest indicator of the average approximation error was obtained under the assumption that the time series Y_t is a random variable. The average approximation error for the regression model is 54.15 % less, which tells us that the time series is a determinate value. As a result of including values e_t' in regression, the average approximation error decreased by approximately 1.44 %.

Conclusion. The method presented above can be used to determine the continuous distribution function of a random variable and generate a random variable for predicting and simulation purposes.

References

1. Egorshin A. V. [Statement of the problem of forecasting the time series generated by a dynamic system]. Yoshkarala, Mary State. tech. un-t Publ., 2007, P. 136–140.

2. Urmaev A. S. *Osnovy modelirovaniya na EVM* [Computer modeling basics]. Moscow, Nauka Publ., 1978, 246 p.

3. Ezhova L. N. *Ekonometrika. Nachal'nyykurs s osnovami teorii veroyatnostey i matematicheskoy statistiki*. [Econometrics. Initial course with the basics of probability theory and mathematical statistics. Textbook]. Irkutsk, Baykal'skiy gosudarstvenny universitet Publ., 2008, 287 p.

4. Anisimov A. S., Kononov V. T. [Structural identification of linear discrete dynamic system]. *Vestnik NSTU*, 2005, No. 1, P. 21–36 (In Russ.).

5. Khinchin A. Ya. *Raboty po matematicheskoj teorii massovogo obsluzhivaniya* [Works on the mathematical theory of queuing]. Moscow, Fizmatgiz Publ., 1963, 296 p.

6. Guiders M. A. *Obshchaya teoriya sistem* [General theory of systems]. Moscow, Globus-press Publ., 2005, 201 p.

7. Kondrashov D. V. [Forecasting time series based on the use of Chebyshev polynomials that are least deviated from zero]. *Bulletin of the Samara state. Those. University. Series: Engineering,* 2005, No. 32, P. 49–53 (In Russ.).

8. Pugachev V. S. *Teoriya veroyatnostey i matematicheskaya statistika* [Theory of Probability and Mathematical Statistics]. Moscow, Nauka Publ., 1979, 336 p.

9. Buslenko N. P. *Modelirovanie slozhnyh sistem* [Modeling complex systems]. Moscow, Nauka Publ., 1968, 230 p.

10. Pugachev V. S. *Teoriya sluchajnyh funkcij i ee primenenie k zadacham avtomaticheskogo upravleniya* [The theory of slash functions and its application to the problems of automatic control]. Moscow, Fizmatgiz Publ., 1960, 236 p.

11. Belgorodskiy E. A. [Some discussion problems of forecasting]. *Ural'skiy geologicheskiy zhurnal*. 2000, No. 2, P. 25–32 (In Russ.).

12. Averill M. L., Kelton D. *Imitacionnoe modelirovanie* [Simulation modeling and analysis. Third edition]. SPb., Piter Publ., 2004, 505 p.

13. Dvoiris L. I. [Forecasting time series based on the analysis of the main components]. *Radiotehnika.* 2007, No. 2, P. 68–71 (In Russ.).

14. Van der Waerden. *Matematicheskaya statistika* [Mathematical statistics]. Moscow, IL Publ., 1960, 436 p.

15. Grenander U. *Sluchajnye processy i statisticheskie vyvody* [Random processes and statistical inferences]. Moscow, IL Publ., 1961, 168 p. (In Russ.).

Библиографические ссылки

1. Егоршин А. В. Постановка задачи прогнозирования временного ряда порождаемого динамической системой. Йошкарала : Марийский гос. техн. ун-т. 2007. С. 136–140.

2. Урмаев А. С. Основы моделирования на ЭВМ. М. : Наука, 1978. 246 с.

3. Ежова Л. Н. Эконометрика. Начальный курс с основами теории вероятностей и математической статистики. Иркутск : Байкальский гос. ун-т, 2008. 287с.

4. Анисимов А. С., Кононов В. Т. Структурная идентификация линейных дискретных динамических систем // Вестник НГТУ. 2005. № 1. С. 21–36.

5. Хинчин А. Я. Работы по математической теории массового обслуживания. М. : Физматгиз, 1963. 296 с.

6. Гайдерс М. А. Общая теория систем. М. : Глобус-пресс, 2005. 201 с.

7. Кондрашов Д. В. Прогнозирование временных рядов на основе использования полиномов Чебышева, наименее уклоняющихся от нуля // Вестник Самарского гос. тех. ун-та. Серия: Технические науки. 2005. № 32. С. 49–53.

8. Пугачев В. С. Теория вероятностей и математическая статистика. М. : Наука, 1979. 336 с.

9. Бусленко Н. П. Моделирование сложных систем. М. : Наука, 1968. 230 с.

10. Пугачев В. С. Теория случайных функций и ее применение к задачам автоматического управления. М. : Физматгиз, 1960. 236 с.

11. Белгородский Е. А. О некоторых дискуссионных проблемах прогнозирования // Уральский геологический журнал. 2000. № 2. С. 25–32.

12. Аверилл М. Л., Кельтон Д. Имитационное моделирование. СПб. : Питер, 2004. 505 с.

13. Двойрис Л. И. Прогнозирование временных рядов на основе анализа главных компонент // Радиотехника. 2007. № 2. С. 68–71.

14. Ван дер Варден. Математическая статистика. М. : ИЛ, 1960. 436 с.

15. Гренандер У. Случайные процессы и статистические выводы. М. : ИЛ, 1961. 168 с.

Shiryaeva Tamara Alekseevna – Cand. Sc., Professor; Krasnoyarsk State Agrarian University. E-mail: info@kgau.ru.

Khlupichev Vladimir Aleksandrovich – Master's Student; Krasnoyarsk State Agrarian University. E-mail: vova.khlp@yandex.ru.

Shlepkin Anatoly Konstantinovich – Dr. Sc., Professor; Krasnoyarsk State Agrarian University. E-mail: ak_kgau@mail.ru.

Melnikova Olga Leonidovna – Cand. Sc., Professor; Khakas State University. E-mail: olga-l-melnikova@yandex.ru.

Ширяева Тамара Алексеевна – кандидат физико-математических наук, доцент, доцент кафедры информационных технологий и математического обеспечения информационных систем; Красноярский государственный аграрный университет. E-mail: info@kgau.ru.

Хлупичев Владимир Александрович – магистрант; Красноярский государственный аграрный университет. E-mail: vova.khlp@yandex.ru.

Шлепкин Анатолий Константинович – доктор физико-математических наук, профессор, профессор кафедры высшей математики и компьютерного моделирования; Красноярский государственный аграрный университет. E-mail: ak_kgau@mail.ru.

Мельникова Ольга Леонидовна – кандидат педагогических наук, доцент кафедры информационных технологий и систем; Хакасский государственный университет имени М. Ф. Катанова. E-mail: olga-l-melnikova@yandex.ru.