http://www.moderntechno.de/index.php/meit/article/view/meit31-00-083
DOI: 10.30890/2567-5273.2024-31-00-083

UDC 658.51

METHOD OF THE STRICTLY STATIONARY TIME SERIES PREDICTION OF THE TYPE "WHITE NOISE"

Borysov Ye. M.
c.ph.m.s., as.prof.
ORCID: 0000-0001-8273-8655
Melnyk O.O.
c.ph.m.s., as.prof.
ORCID: 0000-0002-4399-176X
Lutsyshyna Zh.V.
ph.d, as.prof.
ORCID: 0000-0002-2678-3221
Kyiv National Economic University named after Vadym Hetman,
Kyiv, Beresteiska Avenue 54/1, 03057

Abstract. *A strictly stationary time series obtained by the random number generator was considered in the article. The method of finding predictive values for this series is proposed. It is based on the fact that the average value for a strictly stationary series is the constant quantity. The method of prediction based on the combination of moving average and regression functions is used. In the initial series two other series were allocated between which a close correlation was found. It was found that the prediction accuracy with increasing prediction interval first increases and then gradually decreases. Predicted values are average values for several future values of the time series. Prediction is based on linear or nonlinear regression functions.*

Key words*: strictly stationary time series, white noise, the method of prediction, regression functions.*

Introduction.

Methods of prediction based on combination of moving average and autoregression are widely used and described in detail in the literature [1]. In the article the method of prediction based on the combination of moving average and regression functions [2] is used. Using this method in the work [2] nonstationary time series allowing the existence of a trend were considered. Statistically significant positive correlation was found, on the base of which the prediction was built.

Main text. In the paper we consider a strictly stationary time series of the type "white noise." For this series prediction a certain method [2] is used the feature of which is that the prediction is an average (mean) value for several future values of the time series.

As an example the series obtained by the random number generator in length of 120 values was considered. It was found that the prediction accuracy with increasing prediction interval first increases and then gradually decreases.

Constructing the model. Let us consider a discrete strictly stationary time series of length N:

$$X_N(t) = x(t_1),\ x(t_2),\ldots, x(t_N)\,. \quad (1)$$

And let $E_N = \dfrac{1}{N}\sum_{i=1}^{N} x_i$ be the mean value of this series.

Let us single out in this series the other two series with lengths n and p respectively that are placed consecutively:

$$X_n(t) = x(t_{k+1}),\ x(t_{k+2}),..., x(t_{k+n}). \quad (2)$$

$$X_p(t) = x(t_{k+n+1}),\ x(t_{k+n+2}),..., x(t_{k+n+p}). \quad (3)$$

If in the series (2) we assume that $k = 0$, we shall get a series, the beginning of which coincides with the beginning of the main series (1). For simplicity we will further assume that $k = 0$.

Let $E_n = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $E_p = \frac{1}{p}\sum_{i=n+1}^{n+1+p} x_i$, the mean value for series (2) and (3) respectively.

It is known that for stationary series predictable value is constant. Then we can say

$$\lim_{n+p\to N}(E_n + E_p) = E_N$$

Let us put into the idea of the method the fact that the average value for the stationary series is the value constant. Then it can be stated that, for example, if the average value is $E_n \prec E_N$, than the average value for a series (3) in general, should be greater than the average for a series (1) ($E_p \succ E_N$) and vice versa.

On the basis of the series (2), (3) let us create the new two series: $Y_s(t) = y(t_1),\ y(t_2),..., y(t_s)$, $Z_s(t) = z(t_1),\ z(t_2),..., z(t_s)$ with length of s using consistently formula for finding the average values

$$y(t_m) = \frac{1}{n}\sum_{i=m}^{n+m-1} x(t_i), \qquad m = 1,\ 2,..., s. \quad (4)$$

$$z(t_m) = \frac{1}{p}\sum_{i=n+m}^{n+m+p-1} x(t_i), \qquad m = 1,\ 2,..., s. \quad (5)$$

The length of the series s should be of such size that the inequality $n + p + s \le N$ is performed.

The hypothesis of the model: The future average value of the time series depends on the past average values of the time series. In other words – between time series formed by formulas (4), (5) there exists the correlation. Or, the changes of the series (5) (dependent variable) are explained by the changes of series (4) (independent variable). The calculations carried out below confirm the correctness of the hypothesis.

On the basis of the suggested hypothesis it is proposed to find future (predicted) values using the equation of linear (or nonlinear) regression constructed by formed series (4) and (5). Then, as the prediction we will get the average value of series of the length p. By changing the value of p we can obtain the predicted values for future mean value series different lengths including medium and long-term predictions.

Identification of the model. The choice of the model parameters (series lengths n and s) depends primarily on the length of the main series N. It is recommended to choose the length of the series (2) and (3) so that the condition $1 \le p \le 0{,}3 \cdot n$ is performed. The length of series (2) is selected depending on the particular case.

Choosing the necessary prediction interval p on the basis of the obtained correlation coefficients we construct a regression function (linear or non-linear), from which we get the predicted values.

The example. Let us consider a strictly stationary time series of the type "white noise". For this purpose as the initial data we used series created by the random number generator with normal distribution (zero mean value and dispersion equal to 1 (white noise).

Let us apply for this series proposed model and create the new two series based on formulas (4), (5). According to the model hypothesis, correlation between these series is presented. The correlation coefficient as it will be shown below will depend on the length of predictive time series p.

Let us consider two cases.

Case 1. The length of the main series (1) $N = 120$ series (2) $n = 30$ and the length of the series (3) varied in the range from 1 to 6. The length of series (4), (5) $s = 25$. Below in the table (table 1) the dependence of the correlation coefficient on the length of the series p is presented.

Table 1 - Dependence of the correlation coefficient

	P=1	P=2	P=3	P=4	P=5	P=6
Pearson correlation	-,368	-,367	-,454*	-,473*	-,475*	-,413*
significance level	,070	,071	,023	,017	,016	,040

Authoring

Analyzing the table data, it can be concluded:

1. With increasing the prediction interval (the length of prediction series p) correlation coefficient first increases and then decreases.

2. Statistically significant correlation (with the significance level less than 0,05) was observed for series of length 3, 4, 5 and 6.

3. The most accurate prediction will be for a series of lengths $p = 5$.

Case 2. In this case for $N = 120$ was chosen: $n = 60$ and the length of series (3), (4), (5) took the values $3 \le p \le 23$ and $s = 37$ respectively. Below in the table (table 2) the dependence of the correlation coefficient on the length of the series p are presented.

Table 2 - Dependence of the correlation coefficient

	P=10	P=11	P=12	P=13	P=14	P=15	P=16	P=17
Pearson correlation	-,606**	-,647**	-,679**	-,728**	-,789**	-,843**	-,876**	-,906**
significance level	,000	,000	,000	,000	,000	,000	,000	,000

	P=18	P=19	P=20	P=21	P=22	P=23
Pearson correlation	-,931**	-,939**	-,936**	-,931**	-,924**	-,919**
Significance level	,000	,000	,000	,000	,000	,000

Authoring

Analyzing the table data the following conclusions can be done:

1. Increasing the length of the series (2), (4) and (5) in comparison with the previous example, we got a correlation coefficient close to index **one** with the significance level close to index **zero**.

2. With increasing the prediction interval (the length of prediction range) correlation coefficient first increases and then decreases.

3. For the prediction period of 3 and 4 years, the correlation is not significant.

4. The most accurate prediction will be for a series of lengths $p = 19$.

Disadvantages of the model.

- The model enables to find predictions only for average future values.
- The model does not allow predicting with high accuracy for short time periods.

Advantages of the model.

1. The model makes it possible to obtain predicted values for long-term predictions.

Summary and conclusions.

It is known that most models can obtain short-term predictions. At the same time the prediction accuracy decreases sharply with increasing interval of predictions. In this connection, it is recommended to use the proposed model in combination with other known models of prediction.

References:

1. Box J., Jenkins G. (1974). The analysis of time series, prediction and management. – M.: Mir, – p. 406.

2. Borysov Ye., Kuhay N. (2012). About the method of constructing regression model of predicting the time series. - XVIII National Conference "Modern Problems of Applied Mathematics and Computer Science", October 4-5, Lviv, Ukraine.

Анотація. *У статті розглянуто строго стаціонарний часовий ряд, який отримано за допомогою генератора випадкових чисел. Запропоновано метод знаходження прогнозних значень для цього ряду. Він базується на тому, що середнє значення для строго стаціонарного ряду є сталою величиною. Використовується метод прогнозування, заснований на поєднанні ковзної середньої та функцій регресії. Для початкового ряду було виділено два інших ряда, між якими була виявлена тісна кореляція. Виявлено, що точність прогнозу зі збільшенням інтервалу прогнозу спочатку зростає, а потім поступово знижується. Прогнозовані значення – це середні значення для кількох майбутніх значень часового ряду. Прогноз базується на функціях лінійної або нелінійної регресії.*

Ключові слова*: строго стаціонарний часовий ряд, білий шум, метод прогнозування, функції регресії.*

Article sent: 23.02.2024 p.