Development of a proactive project monitoring model

T A Averina¹, A Yu Glushkov¹, S I Moiseev¹,² and L V Shevchenko¹

¹ Voronezh State Technical University, 84, XX-letiya Oktyabrya str., Voronezh, 394006, Russian Federation
² Plekhanov Russian University of Economics (Voronezh branch), 67a, Karl Marx str., Voronezh, 394030, Russian Federation

E-mail: ta_averina@mail.ru

Abstract. The article proposes a model of proactive monitoring of the project performance consisting of a set of works. This model will allow you to predict in advance the values of the principal features of the project, the main of which is its duration. The use of this model in project management tasks will make it possible to make decisions at earlier stages. In the case of predicting the excess of their critical values, it becomes possible to take measures to eliminate the problem in advance. The mathematical model is based on the methods of regression analysis, namely, on modeling and forecasting the trends of the time series, consisting of the project deadlines at various points in time. The results of computational experiments are presented, which showed that the developed forecasting model provides adequate estimates and the forecasting results are in good agreement with empirical data. In addition, the model depends on a small number of parameters, the influence of which is studied in this paper.

1. Introduction

One of the main distinguishing features of any project, consisting of a sequence of work related to each other, is its limited time, that is, the presence of a start and end point. Based on the amount of work, as well as the number of processes that must be completed to execute the project as a whole, it is possible to calculate the approximate dates of the project task [1].

In order for the project to be completed on time, constant monitoring of the execution time of individual works is required, as a rule those that are responsible for the duration of the entire project, that is, lying on the critical path of the network schedule [2]. If there is a lag in the deadlines for their implementation, measures must be taken to reduce this lag, for example, transfer additional resources to critical work from those that have non-zero time reserves. At the same time, resource management should be proactive, since with significant delays in the project schedule, it becomes more difficult to reduce this lag, as more and more additional resources are required, and their number is always limited [3].

In this paper, we propose a model of proactive monitoring of the execution time of a project, consisting of a set of works that is based on the theory of time series, and more precisely, on modeling and predicting the trend of a time series, consisting of the project deadlines at different points in time.

2. Setting a problem

Consider a project containing a sequence of interrelated work. The methods of network planning and management highlight critical work that allows you to calculate the project time. At each current point
in time \( t \) it is possible to evaluate the degree of lag of the real time of the project at this stage \( w(t) \) from the planned time \( w_p(t) \), that is, the value: \( u(t) = w(t) - w_p(t) \), which can be and negative. Given the stochastic nature of most works, the quantity \( u(t) \) can be considered as some random process [2, 5]. For small lags \( u(t) \) surgical intervention in the work package is usually not required, since it is caused by the influence of random factors and is natural for random processes.

However, a significant lag behind the project implementation schedule at the current moment of time with a high probability can lead to the fact that in subsequent periods of time this lag will grow even more. This is due to the fact that when you deviate from the schedule, various costs increase when performing work, and if you do not carry out operational management to eliminate them, the backlog from the schedule will continue to grow. We assume that there is some critical value of the lag \( u_{cr} \), upon reaching which it is necessary to apply operational control.

The task is to predict in advance the passage of the lag for a critical value in order to take a measure in advance to eliminate it. This can be done by modeling the time lag trend. Figure 1 is a graphical interpretation of proactive monitoring.

3. Mathematical forecasting model
We will directly consider a mathematical model that allows us to predict the values of the time lag from the project schedule for several time periods.

The forecasting model will be based on the theory of time series [6, 7]. Let there be a time series of the indicator \( u(t) \), which was measured over \( k \) time periods: \( u_1, u_2, \ldots, u_k \), where \( u_k \) is the last (current) value of the indicator. Let us find a forecast of the value of this indicator in \( p \) periods, that is, at time moment with the number \( k+p \). To do this, we will use the levels of the time series from the period from \( (k-l+1) \) to \( k \), where \( l \) is the lag of the series or its maximum lag. The forecasting scheme is shown in figure 2.

![Figure 1. Proactive monitoring scheme.](image)

![Figure 2. Forecasting scheme.](image)

We will use the linear trend model constructed by the analytical alignment method [7]. To do this, it is necessary to build a linear regression model according to:

\[
\begin{align*}
    u & \quad u_{k+l+1} & \quad u_{k+l+2} & \quad \ldots & \quad u_{k-1} & \quad u_k 
\end{align*}
\]
In accordance with [6, 8], it will have the form:

\[ \tilde{u}_k = a_k (k + p) + b_k, \tag{1} \]

where

\[
\begin{align*}
a_k &= l \left( \sum_{j=k-l+1}^{k} j \cdot u_j \right) - \left( \sum_{j=k-l+1}^{k} j \right) \sum_{j=k-l+1}^{k} u_j = \left( \sum_{j=k-l+1}^{k} j \cdot u_j \right) - \left( k - l - 1 \right) \left( \sum_{j=k-l+1}^{k} u_j \right), \\
b_k &= \frac{1}{l} \sum_{j=k-l+1}^{k} u_j - a_k \left( \sum_{j=k-l+1}^{k} j \right) = \frac{1}{l} \sum_{j=k-l+1}^{k} u_j - a_k \left( k - l - 1 \right). 
\end{align*}
\]

Based on the obtained model, it is possible to evaluate the accuracy of forecasting. At the first stage, it is necessary to calculate the standard regression error, which estimates the standard deviation of the results of observations of the indicator around the regression line \( S_k \). This indicator can be interpreted as the average scatter of the real values of the predicted indicator around the trend line, that is, in the observation area, and not for the predicted values:

\[ S_k = \sqrt{\frac{1}{l-2} \sum_{j=k-l+1}^{k} (u_j - \tilde{u}_j)^2}. \tag{2} \]

Next, we find the average forecast error \( S_n \), that is, the average value by which the forecast value deviates from the real values of the indicator when forecasting:

\[ \bar{S}_k = S_k \sqrt{1 + \frac{12 ((l-1)/2 + p)^2}{l^3 - l}}. \tag{3} \]

Based on this, we find the maximum error by which the true value of the parameter and its forecast will differ with probability \( p \):

\[ \Delta u_k = |u_k - \tilde{u}_k| = \bar{S}_k \cdot t_{l-\alpha} (l-2) = S_k t_{l-\alpha} (l-2) \sqrt{1 + \frac{12 ((l-1)/2 + p)^2}{l^3 - l}}, \tag{4} \]

where \( t_{l,\alpha}(m) \) – Student’s quantile of distribution [9] with significance level \( \alpha \) and with degrees of freedom \( m \). The confidence interval of forecasting in this way will be equal to \( (\tilde{u}_k - \Delta u_k; \tilde{u}_k + \Delta u_k) \).

Thus, preparation for operational management in the case of high reliability against a false forecast should occur when the condition \( u_k < \tilde{u}_k - \Delta u_k \). In this case, the probability of a false forecast is \( \alpha \). In this case, the preliminary readiness mode for the onset of operational management can be set already when the condition \( u_k < \tilde{u}_k \).

Next, we consider some properties of the forecasting model presented in the work.

4. The study of the forecasting model properties

To study the properties of the model, a computational experiment based on simulation was carried out, which was as follows. A time series was randomly generated, which until the 30th time period did not have a directed trend, but had a random component distributed according to the normal law. Starting from the 30th to the 70th time period, a positive trend was connected. The simulation model made it possible to control the dispersion (dispersion), the magnitude of the trend, and the parameters \( l \) and \( p \).
The time series had autocorrelation (aftereffect), which is typical of real random processes of this kind.

Examples of simulation results are presented in figure 3. The figures along with the graphs of the series and the forecast also show the confidence interval calculated by (4). In these figures and all subsequent ones, the significance level in calculating the confidence interval is taken $\alpha=0.05$.

In figure 3, a small value of the variance of the levels of the series around the trend was used, when the variance of reproducibility $\frac{1}{n} \sum_{i=1}^{n} (\bar{u}_i - \bar{u})^2$ was equal to the variance of adequacy $\frac{1}{n} \sum_{i=1}^{n} (\tilde{u}_i - u_i)^2$. It should also be noted that for clarity, the forecast graph is depicted on the same time periods as the series graph, although it was obtained on $p$ rows earlier. Since a series of length $l$, is initially necessary to obtain a forecast, the forecast schedule does not start from the first period, but from the period $l+p$. 

![Simulation model example](image)
Other realizations of the simulation experiment give a similar picture. It can be seen from the analysis of the figures that for small dispersions, a fairly accurate prediction of the levels of the series is observed, and for large dispersions, a random spread, if it is unidirectional and gives some false trend in a certain interval of periods, leads to forecast fluctuations that in amplitude exceed the fluctuations of the series. However, if you use time series without autocorrelation, the forecasting picture becomes more stable, because in such series, the probability of a false trend is much less.

Also, the results of a computational experiment showed that if the trend of the series is significant, or the dispersion is small, then to start the decision-making mechanism for the reallocation of resources, you can use not the lower limit of the confidence interval, but the upper one, which will increase the lead time. However, with small trends and a large random spread of at least one component, this can lead to a false trend.

We now turn to the influence on the forecast parameters \( l \) and \( p \).

The influence of the length of the time interval \( l \) on the basis of which the forecast is built, as expected, is the stability / flexibility of forecasting. With decreasing \( l \) the forecasting flexibility increases and the stabilization interval for the appearance of a trend decreases (time periods from 30 to 40), but it leads to an increase in the spread of forecasting points, the reaction to series fluctuations increases, and the confidence interval widens. An increase in \( l \) leads to an increase in the periods of forecast stabilization with structural changes in the series, but smooths out the series instability, reducing the likelihood of false trends.

As a result of numerous computational experiments, it was noted that the most optimal result is given by models in which the parameters \( l \) and \( p \) are of the same value. The influence of the parameter \( p \) lies in the fact that its decrease leads to an increase in accuracy, a narrowing of the confidence interval, and a decrease in instability, but all this is done by reducing the forecast range.

The influence of one more parameter, the significance level \( \alpha \) consists in the fact that its decrease also reduces the probability of “false positives”, but increases the width of the confidence interval.

As a result of computational experiments, the Pearson pair correlation coefficient \([8, 9]\) between the levels of the series and the forecast was also calculated. The experiments showed that the correlation strongly depends on the variance of the time series, but when the variance of the adequacy and reproducibility is not more than 6 times, the correlation is on average not less than 0.85 and is significant at a significance level of not more than 0.05, which indicates the adequacy of the forecasting model at a given significance level.

5. Conclusion

In general, the results of computational experiments based on simulation methods have shown that the forecasting model presented in this paper gives adequate estimates, the forecasting results are in good agreement with empirical data, the model is easily controlled by a small set of parameters, the influence of which is determined and regulated. All this allows us to conclude that the presented model of exclusive management can be used in practice when solving problems of planning and project management \([10, 11]\).

References

[1] Barkalov S A and Kurochka P N 2017 Model for determining the term of execution of sub-conflicting works Proc. of Tenth International Conference Management of Large-scale System Development (MLSD) (Moscow) 8109598

[2] Barkalov S A, Burkov V and Kurochka P N 2019 Designing systems of group stimulation in the management of energy complex objects Advances in Intelligent Systems and Computing V(983) 55-68

[3] Barkalov S A, Glushkov A Yu and Moiseev S I 2020 Mathematical methods of multicriteria evaluation of attractiveness of projects Bulletin of the South Ural State University. Ser.
Bondarenko Yu V, Sviridova T A and Averina T A 2019 Aggregated multi-criteria model of enterprise management engineering, taking into account the social priorities of the region. *IOP Conf. Series: Materials Science and Engineering* **537** 042045

Azarnova T V, Barkalov S A and Ukhlova V V 2019 Estimation of time characteristics of systems with network topology and stochastic processes of functioning. *Journal of Physics: Conference Series* **1203(1)** 012055

Hamilton J D 1994 *Time Series Analysis* (Princeton: Princeton University Press) 820

Franses P H and van Dijk D 2000 *Time Series Models in Empirical Finance* (Cambridge: Cambridge University Press) 269

Enders W 2004 *Applied Econometric Time Series* (Hoboken, N.Y.: Wiley) 460

Blockwell P J and Davis R A 1991 *Time Series: Theory and Methods* (N.Y.: Wiley-Interscience) 577

Memeti A, Azizi A and Luma-Osman S 2019 Human Resources Management System: SoA Reference Model. *International Journal on Information Technologies and Security* **4(11)** 29-38

Mihaylov D 2019 A Way to Accelerate the Process of Gathering Information for Decision-making. *International Journal on Information Technologies and Security* **4(11)** 39-50