

# Research on Investment Rate of P2P Platform Based on PSO-ARIMA Model

Xin Wen

School of Management Science and Information Engineering, Jilin University of Finance and Economics, Changchun130117, China  
578522834@qq.com

Mengli Li

School of Management Science and Information Engineering, Jilin University of Finance and Economics, Changchun130117, China  
295184854@qq.com

ZhuoxiYu\*

School of Management Science and Information Engineering, Jilin University of Finance and Economics, Changchun130117, China

\*Corresponding author: 719229899@qq.com

Lu Qin

School of Management Science and Information Engineering, Jilin University of Finance and Economics, Changchun130117, China  
1346984231@qq.com

**Abstract**—P2P Internet lending has attracted a lot of attention as the most popular and controversial part of Internet finance. As a financial time series, the interest rate of P2P platform shows obvious non-stationary characteristics. So, in order to study the investment interest rate of P2P platform and predict the investment rate of P2P net lending platform, it is very difficult to obtain an accurate and simple model adopting traditional linear ARMA model modeling method. The PSO-ARIMA model proposed in this paper can overcome the disadvantages of exhaustive method. The optimal parameters of ARIMA can be obtained by introducing PSO algorithm. In the end, the time series of investment interest rates of P2P platform can be used for modeling, analyzing and forecasting based on PSO-ARIMA. The results show that the PSO-ARIMA model has good prediction effect on the investment rate of P2P platform in a short time.

**Keywords**—P2P platform; Investment interest rate; PSO-ARIMA model

## I. INTRODUCTION

As the style of finance management has undergone tremendous changes, P2P net lending attract a large number of investors with its low cost, high efficiency, rapid mobility, high income and other unique advantages [1].

Evolution characteristics of aP2P net lending platform can be reflected directly and sensitively by investment rate sequence [2]. Investor should take full account of the potential risks before formulate corresponding investment strategies. Similarly, borrowers also need to pay attention to the interest rate in order to make suitable P2P net lending platform. Modeling, analyzing and forecasting the interest rate of P2P net lending platform has important significance for investors and borrowers to avoid risks and make reasonable decisions.

Considering market factors, people prefer a safe way of investing. Therefore, this paper choose a platform called “e

Road concentric” which owns national capital. Experimental data set is composed of historical data from May 25, 2016 to July 22, 2016. First, a model combining ARIMA with PSO (Particle Swarm Optimization, PSO) algorithm is proposed. And then, using sample data to carry out empirical research. In the end, contrast with the classical model. Results of experiment show that PSO-ARIMA model has higher efficiency and more accurate prediction. It also proved that the ARIMA model has a good effect on short-term prediction.

As a classical model of time series analysis, ARIMA has a wide range of applications. Yang Yingmei (2015) forecast consumer price index of Beijing based on ARIMA model [3]. Hu Xiaoyuan (2016) used ARIMA model and GRNN model to predict the incidence of tuberculosis [4].

## II. PROCESS OF METHODS

### A. Autoregressive Integrated Moving Average (ARIMA)

Time series mainly studies the process of occurrence and the following definitions are used for the ARIMA model:

$$\begin{cases} \Phi(B)\nabla^d x_t = \Theta(B)\varepsilon_t \\ E(\varepsilon_t) = 0, \text{Var}(\varepsilon_t) = \sigma_\varepsilon^2, E(x_t \varepsilon_s) = 0, s \neq t \\ E(x_t \varepsilon_s) = 0, \forall s < t \end{cases}$$

$$\begin{cases} \nabla^d = (1-B)^d \\ \Phi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p \\ \Theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_p B^p \end{cases}$$

Meanwhile,  $P$  means the order of Autoregressive;  $q$  means the order of Moving Average;  $d$  is the differential order

**B. Particle swarm optimization algorithm**

Particle swarm optimization (PSO) is derived from the study of foraging behavior of birds. It is an evolutionary algorithm called [5]. According to the flight experience, the particle dynamically adjusts the flight speed and follows the better particle position in the swarm, thus the optimal solution of the optimization problem is obtained. This parallel algorithm has attracted the attention of the academic world for its advantages.

The following definitions are used for the PSO algorithm:

The dimensional solution space is  $d$ ,  $N$  represents the amount of particles in the population. Vector  $X_i = (x_{i1}, x_{i2}, \dots, x_{id})$  means the location of particle  $i$  in the  $d$  dimensional solution space. And the vector  $P_i = (p_{i1}, p_{i2}, \dots, p_{id})$  represents the historical optimal location. The vector  $V_i = (v_{i1}, v_{i2}, \dots, v_{id})$  is the velocity. The location of the particle  $i$  is updated as follows:

$$V^{k+1} = \omega V^k + c_1 r_1 (P_i^k - X^k) + c_2 r_2 (P_g^k - X^k)$$

$$X^{k+1} = X^k + V^{k+1}$$

In this formula,  $t$  is the iteration frequency of updated particle.  $P_i^k$  is the optimal location that the particle  $i$  got in the  $d$  dimensional solution space.  $P_g^k$  is the best location of the particles population.  $c_1$  and  $c_2$  mean the particle acceleration coefficient, and  $\omega$  is the inertia coefficient, meanwhile, the value of these variables are all depend on the experience.  $r_1$  and  $r_2$  are independent random numbers, which are evenly distributed in  $[0, 1]$ .

**C. PSO-ARIMA Model**

In this paper, a new model called ‘‘PSO-ARIMA Model’’ is created which is based on ARIMA model and PSO algorithm. Reciprocal of the value of Akaike information criterion (AIC) will be as a fitness function. During the process of operation, utilizing PSO algorithm fit parameters of the ARIMA model constantly. And the value of the fitness function is calculated, compared with the previous numerical, updated and preserved every time.

TABLE I The Modeling step

Input: Data set	
Output: Model estimation	
Step 1	Assign values to three parameters p,d,q at random as initialization parameters
Step 2	Calculates the value of the fitness function; if the time series after difference operation meet the requirements, store results.
Step 3	Updates the speed and information of particle
Step 4	Determine whether the iterations reached the upper limit. If the maximum value is reached, exit algorithm and output results. If else, proceed to the next step
Step 5	Compare the existing location function value with the previous generation best function value (AIC Etc.), and change the current optimal value if it is better.
Step 6	Determine whether the accuracy meets the requirements. If satisfied, save the results and exit the program. Instead, go back to step 2.

It can be seen from the above chart that proposed PSO-ARIMA model can make up for the traditional method. According to the fitness function, starting from the three random parameters, calculating, comparing and update the parameter values until output result with low cost, high computational efficiency.

PSO-ARIMA model has predominance such as efficiency, operability and so on. And that, PSO-ARIMA model can achieve a better result than traditional ARIMA and also can provide a reference for the study of high order ARIMA model

**III. PSO-ARIMA**

**A. PSO-ARIMA model parameter estimate**

As the financial time series, the investment rate of P2P net lending platform shows obvious non-stationary characteristics. According to traditional ARIMA model, it is necessary to do the differential smoothing of the original sequence and then try and compare with the exhaustive method to find the optimal model. As can be seen from figure 1-2, the time series of the investment interest rate is obviously non-stationary.

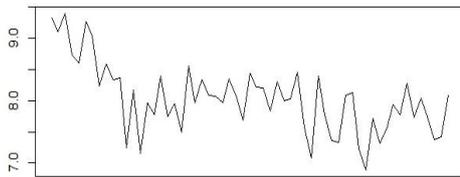


Fig. 1 Series of investment rate of P2P net lending platform

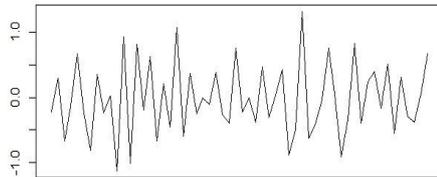


Fig. 2 Series after first order differential.

In order to enhance the accuracy and reduce the running time of the PSO algorithm, assume that the parameter value is not 0. In the R software, the hydroPSO function provided by the package called “hydroPSO” can complete the experiment. The initialization parameters are as follows:

Here is the illustration of partly parameters. The range of parameter  $p$  and  $q$  are the same, which is  $[0, 5]$ . And the range of parameter  $d$  is  $[1, 3]$ . The amount of particles population is  $npart = 40$ .  $max\ it = 1000$  is the maximum number of iterations. The optimize approach is  $method = spso2011$ . The value of topology is  $topolog\ y = vonNeumann$ , and the limitations of accuracy are  $reitol = 1e - 20$ ,  $c_1 = 0.5 + \log(2)$ ,  $c_2 = 0.5 + \log(2)$ ,  $\omega = 1 / (2 \times \log(2))$ .

Level of particle can be reflected by fitness function and the reciprocal of AIC value is used as fitness value. The minimum fitness value is the optimal values. If fitness value is lower, the particle is reserved. Otherwise, it would be eliminated. Three parameters of PSO-ARIMA model are obtained as shown in the following figure:

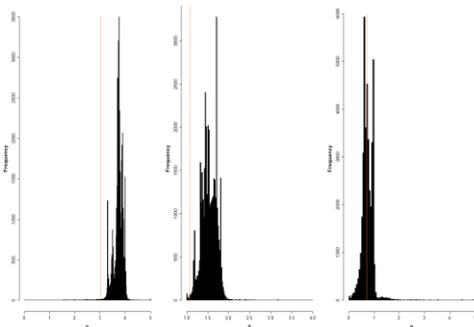


Fig. 3 Params Histograms

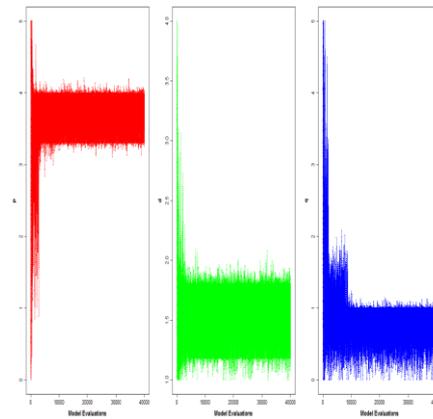


Fig. 4 Params Values Per Run

As can be seen from figures 3 and 4, the parameters are optimized by the PSO algorithm. The interval of parameter  $p$  estimation is  $[3, 4]$ ; the interval of parameter  $d$  is  $[1, 2]$ ; and the interval of parameter  $q$  is  $[0, 1]$ . To provide reference for higher order financial time series analysis, the maximum number of iterations is limited to 1000. Iteration operations of the relevant legend of PSO-ARIMA model are following:

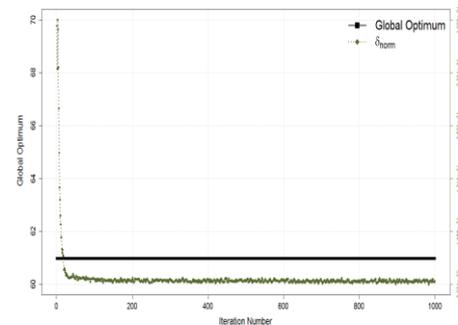


Fig. 5 Convergence Measures

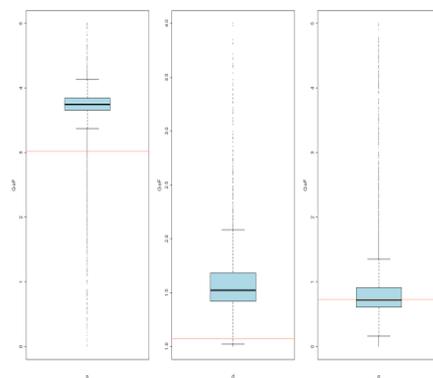


Fig. 6 Params Boxplot

From Figure 5, it can be found that when the number of iterations reaches about 100 times, value of the optimization parameters tends to be stable and is no longer updated in general. From the boxplot in Figure 6, we can see the distribution of each parameter. It further explains the optimum range of parameters.

According to the experimental results, the parameters of the PSO-ARIMA model were 3.72, 1.51, and 0.75. The experimental were handled with R. When fitting PSO-ARIMA

model, it can simply take the integer portion of each parameter. In other words, the optimal model is PSO-ARIMA(3, 1, 0).

*B. Model test*

Based on the above experimental results, the optimal model of investment rate of P2P net lending platform is obtained, and then some operations should be done.

*(a) Test for Significance of Parameters*

TABLE I RESULTS OF PARAMETER SIGNIFICANCE TEST

		Estimated Value	SE	t	Sig
	Lag 1	-0.7706	0.1218	6.3268	.000
AR	Lag 2	-0.5513	0.1398	3.9435	.001
	Lag 3	-0.3882	0.1207	3.2162	.004

It can be seen from table 2 that the Sig value is far less than 0.05, so the relevant parameters of the model are considered effectively by significance test.

*(b) Test for Stationary of Time Series*

Therefore, carry out the tests for stationary of time series after differential operation. In this paper, use ADF test of unit root to test stationary of time series and the experimental results are as follows:

TABLE II RESULTS OF ADF TEST

	Lag	p
ADF test	1	<0.01
	2	<0.01
	3	<0.01

From above table 3, it can be known that value of p is obviously less than 0.01. Therefore, unit root test of lag 3 order for the residual sequence can be used to prove that the value of parameter d is effective.

TABLE III COMPARING ARIMA(0,1,1) WITH PSO-ARIMA(3,1,0)

Date	Real Value	ARIMA(0,1,1)	PSO-ARIMA(3,1,0)
7.23	7.17	7.47807	7.224938
7.24	7.23	7.47087	7.532461

*(c) Test for White Noise*

In this paper, the white noise of the model residual sequence is tested by the Box, test function in R software. The experimental results are shown in Table 4 below:

TABLE IV RESULTS OF WHITE NOISE TEST

	Lag	p
White noise test	1	0.9027
	2	0.9907
	3	0.9758

It can be seen from the table that the value of p is greater than the significance level. So, the assumption was accepted that residual error sequence is random sequence.

IV. PREDICTION AND COMPARISON

According to the result, the best model was ARIMA (0, 1, 1) determined by auto.arima function. The optimal model was PSO-ARIMA (3, 1, 0) confirm by PSO-ARIMA method. Prediction effects were compared. The results are as follows:

Table above indicates that the investment rate of the P2P net lending platform will be 7.225 and 7.532. Errors were 0.07% and 4% respectively.

Defining the root mean square error of the model ARIMA (0, 1, 1) as  $RMSE_1$  and the root mean square error of model PSO-ARIMA as  $RMSE_2$ . It can be known that  $RMSE_1 = 0.279684$ ,  $RMSE_2 = 0.217372$  and  $RMSE_1 > RMSE_2$ . The experiment proved that PSO-ARIMA model can achieve a better prediction.

## V. CONCLUSION

In this paper, PSO-ARIMA model was created and used in studying and analyzing the investment of P2P net lending platform. Combing historical data of investment rate of platform, the rate can be predicted in the next short days.

Due to the ARIMA model based on the traditional mathematical method, it has high efficiency and better short-term prediction of financial time series. But when the model is used to select parameters, the experience factors have a great interference and may lead to the lower efficiency. Particle Swarm Optimization (PSO) algorithm is a random search algorithm. All show that the PSO algorithm with higher superiority in many optimization problems and the some deficiency of parameters can be corrected. Starting with getting random values of the parameters, and data is substituted into the PSO algorithm. Calculating, updating and iterating to find out the optimal solution.

The final optimization results also confirmed the superiority and accuracy of the approach that we combined the PSO algorithm with ARIMA model. Comparing with traditional ARIMA model, PSO-ARIMA can scientifically predict the trends over a period of time and have better accuracy in short-term forecasting and it will have a broaden application in areas of internet financial.

## ACKNOWLEDGMENTS

The research was supported by the National Social Science Foundation of China (No.16BTJ020).

## REFERENCES

- [1] Fang Ling. On the development of Internet financial P2P opportunities and risk control - "PPDAI" as an example [J]. *China Business Theory Research*, 2016 (29): 71-72.
- [2] Zhong Guowei. China's P2P market interest rate of the particularity and influencing factors analysis [J]. *CHINA JOURNAL OF COMMERCE*, 2015 (12): 76-78.
- [3] Yang Yingmei. Study on consumption price index of Beijing residents based on ARIMA model [J]. *Statistics and Decision*, 2015 (4): 76-78.
- [4] Hu Xiaoyuan, Wu Juan, Sun Qingwen and so on. Comparative study of ARIMA model and GRNN model for predicting tuberculosis incidence [J]. *Journal of Second Military Medical University*, 2016, 37 (1): 115-119.
- [5] Chen shan Shen shichang. ARIMA model on the sustainable development of China's tourism industry empirical research [J]. *Journal of Qiqihar University (Natural Science Edition)*, 2017, (02): 84-88.