Research on Modeling and Simulation System of Economic Factors Based on Swarm Platform

Yuanrui Zhao

School of Maritime Economics and Management, Dalian Maritime University, Dalian, 116016, China

Abstract. Based on the VAR model, this paper uses the stepwise regression analysis method to analyze the main influencing factors of the economy, and empirically analyzes the impact of Nanning's fixed asset investment on urban and rural economic development in the past ten years. Use eviews statistical analysis software to perform regression analysis on the data to obtain a regression model, and build a VAR model according to investment and income, and finally find out the correlation of variables. On the basis of the existing theory (genetic algorithm, cellular automata and classification system), according to the agent-based modeling and simulation method Swarm, according to the steps of agent-based distributed simulation of complex system, modeling and simulation are carried out. The abstraction level of simulation is established on the big data computing level, and the final simulation results are briefly explained and explained.

Keywords: Swarm, modeling, simulation, multiple linear regression; VAR model.

1. Introduction

Swarm's original development team defined Swarm as a "multi-agent platform for the simulation of complex adaptive systems". Researchers at the Santa Fe Institute hope that Swarm will become a standard research tool for constructing discrete event simulations of complex systems with heterogeneous agents, and for studying complex adaptive systems (CAS). Swarm simulation modeling adopts the mode of independent agents interacting through discrete events. In this framework, Swarm does not make any assumptions or restrictions on the interaction between the model world or model elements. The modeling idea of Swarm is to let a series of independent agents interact through independent events to help study the behavior of complex adaptive systems composed of multiple agents. Through these class libraries, many reusable classes are included to support the analysis, display and control of simulation experiments, that is, users can use the class libraries provided by Swarm to build simulation systems, so that the subjects and elements in the system interact through discrete events. Since Swarm does not impose any constraints on the model and the interaction between model elements, Swarm can simulate any physical system, economic system or social system. Swarm is actually being programmed in a wide variety of research fields, including biology, economics, physics, chemistry, and ecology.

Considering the representativeness of the data, this paper selects Nanning as the research object. From the perspective of econometrics, the gross domestic product of Nanning is studied to find out the main factors affecting the development of its national economy, and through this study, it can find the law of its development and predict its future development trend, so as to understand the future development of Northwest China. The direction of the economy, and put forward policy guidance and planning guidelines for the regional economic construction. Therefore, this paper takes the gross domestic product of Nanning as the explained variable, and the industrial production and other indicators as the explanatory variables. By constructing a multiple linear regression model and VAR, the main factors affecting the national economy of Nanning are tested, and effective methods for the future development of Nanning are proposed. Suggestion [1].

2. Analysis of technology platform for modeling

This section introduces the relevant knowledge of Swarm software, including its development background and history, the development tools of the Swarm model, and how to construct a Swarm model, including the modeling ideas of the Swarm model, the main building blocks of the Swarm model, and the relationships and connections between them [2]. Swarm's class library main functions, etc.

2.1 Analysis of Swam Platform

Based on the CAS theory, SF Engineering of the Santa Fe Institute has developed a multi-agent software tool set, SWARM. For the multi-agent toolset SWARM, it is a general-purpose software platform that "can be used in biology, ecology, culture and anthropology, computer science and industry, political science, geography, defense, and more". A Swarm is an object that implements memory allocation and event planning [3]. A basic SWARM simulation program generally includes an observer Swarm and a model Swarm. The model Swarm is a necessary part of a SWARM simulation program, where we can create all kinds of subjects required in the model, the subject has attributes and methods, the attributes are used to indicate the state of the subject, and the methods are used to indicate the behavior of the subject. The schematic diagram of the Swam platform is shown in Figure 1.

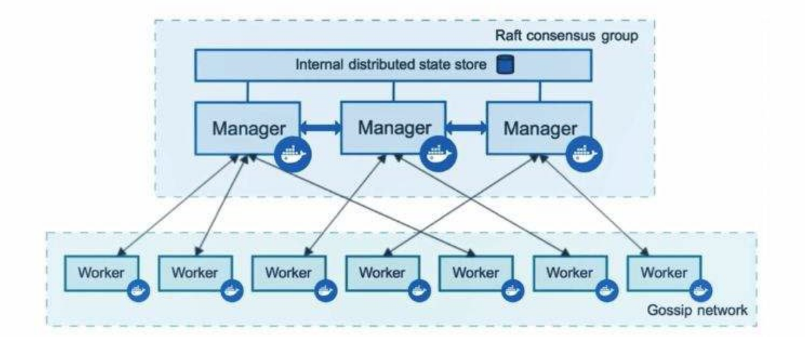

Figure 1. Schematic diagram of Swam platform

2.2 Simulation simulation structure

The simulation flow of the model is as follows: The simulation clock advances in a discrete manner, and an object is randomly selected at time t. We call this object the meta-client, that is, the evaluation of the operator by other clients in this group is affected by the meta-object [3]. The influence probability of the meta object is P0, and 8 customers are distributed on a two-dimensional grid around him, each customer has a different state S and latency (normal distribution with beta function).

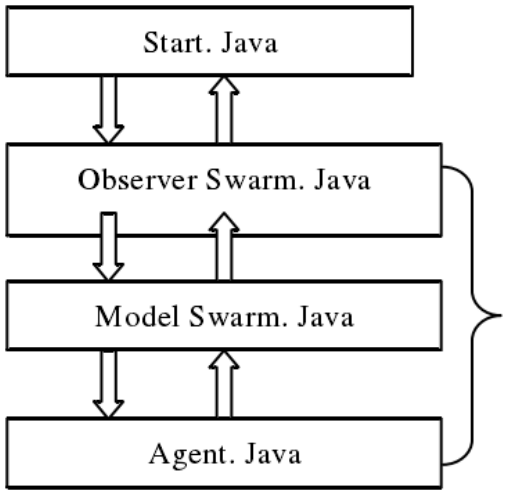

Figure 2. Basic hierarchy of the model based on Swarm

3. Comprehensive consideration and analysis of the economic impact of independent variables

3.1 Correlation analysis

The independent variables and dependent variables have been determined above, so Pearson correlation analysis can be performed on the data first to obtain the following operating results:

Table 1. Correlation coefficients between variables

	X 1	X 2	X 3	X 4	X 5	X 6	X 7	Y
X 1	1.00	0.99	0.88	0.92	0.99	0.91	0.97	0.84
X 2	0.99	1.00	0.92	0.96	0.99	0.88	0.98	0.89
X 3	0.88	0.92	1.00	0.99	0.93	0.82	0.94	0.98
X 4	0.92	0.96	0.99	1.00	0.96	0.86	0.97	0.98
X 5	0.99	0.99	0.93	0.96	1.00	0.91	0.99	0.89
X 6	0.91	0.88	0.82	0.86	0.91	1.00	0.95	0.80
X 7	0.97	0.98	0.94	0.97	0.99	0.95	1.00	0.91
Y	0.84	0.89	0.98	0.98	0.89	0.80	0.91	1.00

We consider these two variables to be highly correlated when the Pearson correlation coefficient is greater than 0.7. We consider these two variables to be moderately correlated when the Pearson correlation coefficient is between 0.3 and 0.7. When it is below 0.3, we consider the relationship between these two variables to be very weak. The premise we require is that the significance is less than the significance level of 0.01 or 0.05, and we consider these two variables to be significantly related. Such results are statistically significant [4]. From the table, we can use the Pearson correlation coefficient to see the degree of correlation between variables. As shown in the data, it can be inferred that an industry can drive the common prosperity of other industries by investing in the comprehensive development of one industry [5].

3.2 Building a stepwise regression model

We found that there is a strong correlation between the variables, and there may be multicollinearity. We will remedy the multicollinearity below and use stepwise regression.

Table 2. Stepwise regression model

Variable	Coefficient	Std. Error	t-Statistic	Prob.
X3	10.65047	0.680623	15.64812	0.0000
C	33845.76	25065.45	1.350295	0.2139

R-squared	0.968362	Mean dependent var	388548.9
Adjusted R-squared	0.964408	S.D. dependent var	179330.3
S.E. of regression	33832.33	Akaike info criterion	23.87308
Sum squared resid	9.16E+09	Schwarz criterion	23.93359
Log likelihood	-117.3654	Hannan-Quinn criter.	23.80669
F-statistic	244.8636	Durbin-Watson stat	2.171300
Prob(F-statistic)	0.000000		

According to the above results, it can be seen that the larger value of the regression coefficient of determination is 0.968, and the corrected coefficient of determination is 0.964. Such data indicate that the regression equation has a good relationship to the data. It can also be guided by the model that we do to fit the sample better, that is, the independent variable has a good correlation with the factor. According to the ANOVA results, we get an F value of 244.8636. By querying the F value distribution table, if it is greater than the critical value, it is considered to pass the F test, and its representative significance is 0.000. When the F value is less than the significance level of 0.05, the overall significance of the equation is considered. It can be seen from the above that the model collinearity statistic VIF is less than 10, which indicates that the model designed this time does not have the characteristic of multicollinearity. According to the significance of the independent variables of the model, the significance of its added value to regional GDP is less than the significance level of 0.01. When the null hypothesis is rejected, we know that the model coefficients are considered significant [6].

3.3 White test

We perform a heteroscedasticity test as follows:

According to the idea of White's test, set up an auxiliary regression model and perform White's test on it:

$$\boldsymbol{\sigma_i^2 = \alpha_0 + \alpha_1 B + \alpha_2 L + \alpha_3 B^2 + \alpha_4 L^2 + \alpha_5 B * L}$$

The inspection output is as follows:

Table 3. White test results

F-statistic	1.821990	Prob. F(2,7)	0.2307
Obs*R-squared	3.423513	Prob. Chi-Square(2)	0.1805
Scaled explained SS	4.453399	Prob. Chi-Square(2)	0.1079

From the test results, it can be concluded that the coefficient p value is 0.2307. Under α=0.05, the distribution table has a critical value $\chi^2_{0.05}$ (5), and the p value is greater than 0.05, so accept the null hypothesis: $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = 0$, indicating that the model has heteroscedasticity. We use weighting to deal with heteroscedasticity as follows:

Table 4. Regression model after eliminating heteroskedasticity

Variable	Coefficient	Std. Error	t-Statistic	Prob.
X3	9.895849	0.799972	12.37025	0.0000
C	56760.72	20918.78	2.713385	0.0265
Weighted Statistics				
R-squared	0.950318	Mean dependent var		332314.5
Adjusted R-squared	0.944107	S.D. dependent var		31627.98
S.E. of regression	24208.19	Akaike info criterion		23.20363
Sum squared resid	4.69E+09	Schwarz criterion		23.26414
Log likelihood	-114.0181	Hannan-Quinn criter.		23.13724
F-statistic	153.0230	Durbin-Watson stat		1.892233
Prob(F-statistic)	0.000002	Weighted mean dep.		299770.2
Unweighted Statistics				
R-squared	0.963331	Mean dependent var		388548.9
Adjusted R-squared	0.958748	S.D. dependent var		179330.3
S.E. of regression	36423.19	Sum squared resid		1.06E+10
Durbin-Watson stat	1.894400			

According to the results shown above, the larger value of the regression coefficient of determination is 0.963, and the corrected coefficient of determination is 0.959. Such results show that the regression equation has a good relationship with the data, and the model fits the samples better [7]. The vast majority of differences in variables as a whole are well explained by them. The F value we got from the ANOVA results was 153.0230. By querying the F value distribution table, if it is greater than the critical value, it is considered to pass the F test. Its significance is 0.000, which is less than the significance level of 0.05, and the overall significance of the equation can be considered. According to the significance of independent variables of the model, the significance of fixed asset investment to regional GDP is less than the significance level of 0.01. We therefore reject the null hypothesis and consider the relationship of the model coefficients to be significant. And when the p-value of the white test is less than the significance level, we can reject the null hypothesis that there is no heteroskedasticity. The Dobin-Watson value of the optimized model is 1.8944, and we believe that there is no serial autocorrelation in the model when the value is around 2 [8].

4. Data analysis and modeling of the impact of investment on the economy

From the correlation analysis above, we know that there is a close relationship between various industries, and investing in one industry to fully develop one industry can lead to the common prosperity of other industries [9]. Studying the impact of fixed asset investment on GDP will help to rationally arrange fixed asset investment projects and promote the improvement of fixed asset investment benefits. In reality, investment is one of the "troikas" that drive economic growth, and investment in fixed assets plays a vital role in the entire production process as an investment in means of production. The value of fixed assets is transferred to the products produced one by one, so investment in fixed assets indirectly affects the output of an economy [10]. From the actual situation of my country's economic operation, when the growth rate of fixed asset investment rises, the growth rate of GDP also rises accordingly; on the other hand, when the growth rate of fixed asset investment declines, the growth rate of GDP also declines accordingly. Next, we will establish a VAR model, focusing on the driving effect of investment on the national economy.

4.1 Stationarity Test

Before establishing the model, the time series must be stationary first. Generally, the definition of time series stationary is mainly weak stationary, that is, the expectation of a random variable is constant, the variance is constant, and the covariance does not change with time. If a time series is not stationary, it is difficult to predict the future values of the time series. From the definition of stationary, the variance of the random variable will gradually increase to infinity at this time, and the prediction is meaningless [11].

In general, we use the ADF test for stationarity test, which is an extension of the Dickey-Fuller (DF) test. The DF test can only be applied in the case of a first-order AR model. When the series are of high order, there is a lagged correlation, so the more general ADF test can be used. We found that the original series has an obvious upward trend, and we conduct stationarity test for the series of dx3 and y. The specific ADF test of each variable is shown in the following table. After the difference, each variable is in a stationary state.

Table 5. ADF test of explained variable

Y		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-2.904038	0.0993
Test critical values:	1% level	-5.119808	
	5% level	-3.519595	
	10% level	-2.898418	

Table 6. ADF test of explanatory variable X3

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-3.023621	0.0715
Test critical values:	1% level	-5.119808	
	5% level	-3.519595	
	10% level	-2.898418	

4.2 Construction of VAR Model

In order to more strictly explain the dynamic link between the investment in fixed assets and the national economy, a VAR model is established. After inspection, y and dx3 are both stationary time series. First of all, we need to choose the number of VAR lag periods. We selected the second-order lag model through AIC/BC/LR and other tests. The AIC value is 39.79776, which is the smallest in comparison, indicating that the model fitting effect is not bad, as follows:

Table 7. VAR model

	Y	DX3
Y(-1)	0.626322	-0.030332
	(0.75144)	(0.03740)
	[0.83350]	[-0.81092]
Y(-2)	0.904979	0.204693
	(1.20227)	(0.05985)
	[0.75272]	[3.42036]
DX3(-1)	14.17733	-0.385386
	(5.88554)	(0.29296)
	[2.40884]	[-1.31547]
DX3(-2)	-6.293828	-0.955487
	(11.6570)	(0.58025)
	[-0.53992]	[-1.64668]
C	-129480.1	-41634.96
	(178106.)	(8865.59)
	[-0.72698]	[-4.69624]
R-squared	0.988262	0.962870
Adj. R-squared	0.964785	0.888609
Sum sq. resids	2.52E+09	6247992.
S.E. equation	35508.03	1767.483
F-statistic	42.09505	12.96610
Log likelihood	-78.89050	-57.88908
Akaike AIC	23.96871	17.96831
Schwarz SC	23.93008	17.92967
Mean dependent	445054.3	6204.714
S.D. dependent	189217.2	5295.787

4.3 Unit root test

In order to prevent the "pseudo-regression" problem and ensure the stationarity of the series, we carry out the unit root test to judge the stationarity of the series. If the sequence is stationary, you can continue modeling; if the sequence is non-stationary, you need to perform differential processing or use analysis methods such as cointegration. The obtained results are shown in Figure 3. The scattered points fall within the unit circle, and most of the Modulus in the table are around 1, indicating that there is no unit root and they are all stable.

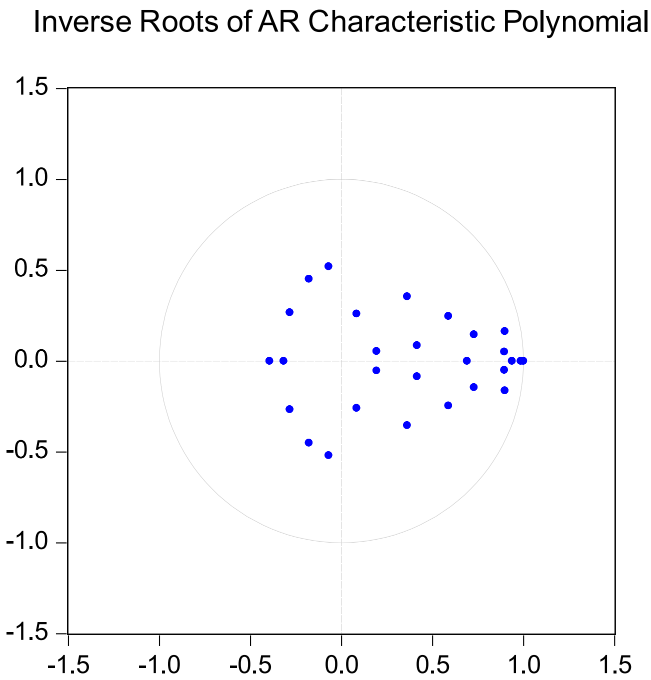

Figure 3. Unit root test results

4.4 Granger causality test

Granger causality can be used to test whether all lagged terms of a variable have an effect on the current value of another variable or variables. It is the premise of whether the impulse function can be established. When a variable is indeed affected by the lag of other variables: if the effect is significant, it means that the variable has Granger causality on another variable or several variables; if the effect is not significant, it means that the variable has a significant effect on another variable or several variables. The variables do not have Granger causality. Table 8 shows that the null hypothesis "investment in fixed assets is not the Granger cause of Nanning's GDP. From Table 8, we can see that the significance level of X3 does not Granger Cause y is 0.000, and the p value is less than the significance level of 0.05. Reject the null hypothesis, indicating that X3 is the Granger cause of Y, the significance level of y does not Granger Cause X3 is 0.0030, and the p-value is less than the significance level of 0.05, indicating that Y is the Granger cause of the Granger of X3, so It can be shown that there is a causal relationship between fixed asset investment and gross national product, and the two interact and influence each other.

Table 8. Granger causality test results

Dependent variable: X3

Excluded	Chi-sq	df	Prob.
Y	114.3240	2	0.0000
All	114.3240	2	0.0000

Dependent variable: Y

Excluded	Chi-sq	df	Prob.
X3	11.60051	2	0.0030
All	11.60051	2	0.0030

4.5 Impulse response and variance decomposition

In the VAR model, the role of impulse response analysis is to analyze the time and magnitude of the influence of one variable on another variable. Study the impact on the entire model system when the disturbance term changes, which is used to describe how a change in one variable affects all other variables in the model. It can be seen from the figure that after giving x3 a positive shock, y tends to be flat in the short term but rises in the long term, indicating that investment in fixed assets cannot promote the economy in the short term, but in the long run, when the time is sufficient When the length and space are large enough, investment in fixed assets has a very positive effect on the economy.

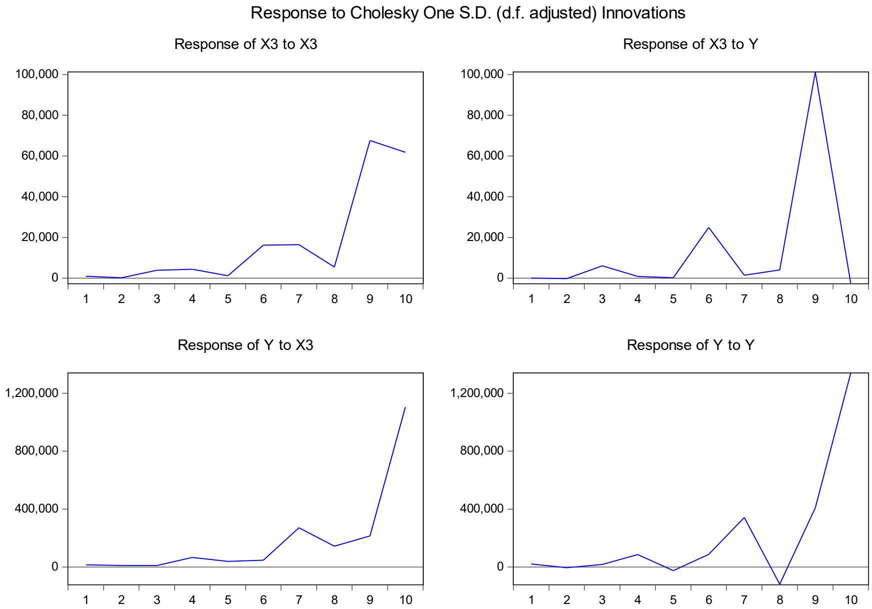

Figure 4. Analysis of investment variables

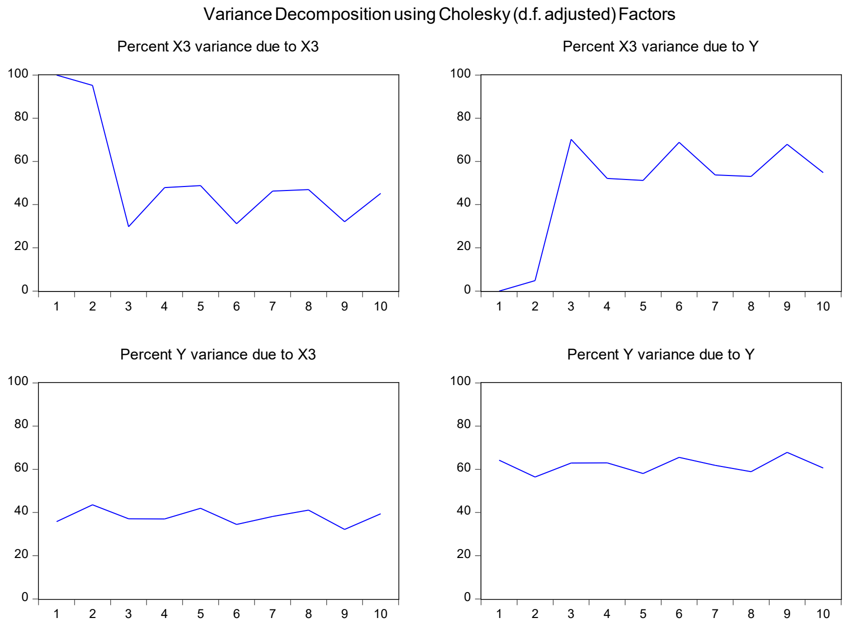

Figure 5. Analysis of investment variables

5. Conclusions and recommendations

By analyzing the data on the influencing factors of Nanning's national economy from 2011 to 2020, a multiple linear regression model and a var model are established, and the conclusions are as follows: In the analysis of the influencing factors including the industrial production value, it is concluded that the Nanning national economy The main influencing factor of the economy is investment in fixed assets. However, the driving effect of agriculture, forestry, fishery and animal husbandry and industry on the national economy cannot be ignored, but there is a phenomenon that the impact on the local economy is not significant in the analysis. Therefore, the local government should strengthen the construction of regional industries on the basis of attaching importance to investment, stimulate the development of related industries with a large amount of investment and policy support, and drive the local economy. The following recommendations are made:

1. Optimize the environment for the coordinated development of industry, construction and the national economy. The relevant departments of the regional government should formulate corresponding development policies to create a favorable market environment for the development of the industry and construction industry, make full use of policy resources in the process of development, and promote the mutual benign development of industries by strengthening market supervision and management. To promote the development of the national economy well;

2. By investing in information technology, optimize and upgrade the industrial structure, integrate informatization and industrialization, take the road of new industrialization characteristics, and promote the healthy and stable development of the industry.

3. In addition to industry and construction, local governments and related enterprises should also make full use of local advantageous resources and vigorously develop other characteristic industries to improve economic development, such as tourism.

4. Vigorously invest in high-tech industries, promote technological progress, and pay attention to the introduction of professional talents;

5. Provide subsidies for local small and micro enterprises with financing difficulties to help them grow rapidly into the backbone of the industry; 6. Reasonable allocation of funds to effectively stimulate coordinated development among various industries and promote common prosperity.

References

[1] Research on the interactive relationship between environmental regulation, FDI and economic growth in Shaanxi - Impulse response analysis based on VAR model Jiang Yonghong [J].

[2] Analysis of the relationship between urbanization and tertiary industry development based on VAR model Zhou Haoran [J].

[3] Empirical Analysis and Prediction of Xinjiang's GDP Based on VAR Model Wei Yanjing, Xie Chengxing, Wang Fengxiao [J].

[4] Empirical analysis of social and economic development in Guangxi Deng Yun [J].

[5] The essence of thinking from object-oriented to agent-oriented [J]. Tang Peihe, Liu Hao, Yang Xinlun. Journal of Guangxi Institute of Technology. 2006(01).

[6] Research on a method of system effectiveness analysis based on multi-agent simulation [J]. Li Honglei, Gan Renchu. Journal of System Simulation. 2006(02).

[7] Comparison between object-oriented and agent-oriented [J]. Zhao Shaodong. Modern Computer. 2005(11).

[8] Development method of reusable conceptual model [J]. Hao Shimin, Tan Dongfeng, Yu Junxin. Intelligence Command and Control System and Simulation Technology. 2005(04).

[9] Overview of Industrial Cluster Classification [J]. Liu Shanqing, Chen Wenhua, Ye Xiaolan. Enterprise Economics. 2005(05).