The Regression Analysis and Prediction of Real Estate Added Value Based on Genetic Algorithm

Yanli Zhao [1]
School of Management
Harbin University of Commerce
Harbin, China
E-mail: yanlizhao1964@sina.com

Shuangshuang Jia [2]
School of Management
Harbin University of Commerce
Harbin, China
E-mail:jiashuangshuang1222@126.com

***Abstract*—Total productive value of Real estate is a crucial part of the Service Industry, which directly affects the value of GDP. It is significant to predict the added value of the total productive value of Real estate, by historical observed data and dynamic regression equations. Compare dynamic regression equations with genetic algorithm to regression equation from exponential regression and linear regression. The predicted results through genetic algorithm method get closer to the true value than the other two methods. Meanwhile, the result also points out the added value of the total productive value of Real estate with genetic algorithms in the next years.**

***Index Terms*—real estate added value, genetic algorithms, egression equation**

I. INTRODUCTION

In recent years, the total value of GDP is great increase in real estate, which brings large effect to the total value of GDP in the overall Service Industry. The real estate added values in 1978 -2007 (Table I) showed a rising trend .According to this trend make some further predicted in the coming years that will be benefit for the relevant government departments and real estate investment related industries in scientific decision-making. Through statistic the real estate increase value show a clear dispersion characteristic, according to this discrete data characteristic, use the genetic algorithms fit the real estate added value and get a relatively high degree regression equation, then use the regression equation forecast the added value of real estate in the future years.

II. THE PROOCESS OF GENETIC ALGORITHM TO AUTOMATICALLY SIMULATION EQUATION

A. Representation of Simulation Program That Equation

Genetic Algorithms,(GA) is a high degree of genetic mechanisms reference biosphere parallel, randomized, adaptive search algorithm for global optimization of probability, it uses a structured combination of the random switching structure of the group all the best of survival factors, create the best code. It is widely used, De Jong primary apply to the function optimization problem[1], with the constant development of its application in the optimal combination of production scheduling [2-3] and machine learning[4] and other fields. Computer programming for solving optimization problems using the program, said its a different approach, such as are currently used programming language source code expression representation, the binary string that law, but the most commonly used method of procedure that is used to represent the syntax tree, the group composed of all individuals to adopt a dynamic tree structure to represent. Node of this tree by the end point, the composition of the original functions and operators, the root node and intermediate nodes collectively referred to as internal nodes, which are combinations of these original variables function. For example: $y=\ln x+e^{\cos x+2}$ Evolution of the genetic algorithm in automatic programming of the computer can use a syntax tree, as shown in Figure1:

B. The Design of Genetic Operators

According to the evolution of genetic algorithm representation of computer programs to design appropriate crossover operator and mutation operator, this paper studied only operator syntax tree design.

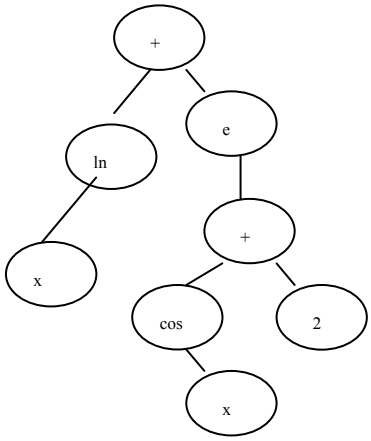


Figure 1 ： Syntax tree

1) Operational Design of Crossover operator

Crossover operator is select from two parent bodies , and randomly exchange their two sub-trees, which are new populations of the two new individuals.

2) Operational Design of Mutation operator

Mutation operator is randomly selected from the population of a parent of a subtree with another subtree to replace it and produce the next generation of offspring.

C. The minimum cost function

Cost function [5] is an evaluation function of genetic algorithm. It determines the limits of genetic algorithm to determine the evolution of computer programming iteration

termination condition. Therefore, in theory give the genetic algorithm smaller cost function value, the greater the number of iterations. Papers in accordance with the principle of least squares using the Euclid norm of the evolution process as a cost function, that:

$$d = \min(\|x\|^2) = \min(\sum_{i=1}^{N}(y(i) - \sum_{k=1}^{M} a_k F_k(x_1(i), \quad, x_L(i))))^2$$

That $x_1(i), x_2(i) \quad x_L(i), i=1,2 \quad N$ is the original problem of the dependent variable, N is the number of data points, $y = (x_1, x_2, \quad, x_L) = \sum_{k=1}^{N} a_k F_k(x_1, x_2, \quad, x_L)$ is the fitting function, Here a_k is a constant. Suppose:

$$A = \begin{pmatrix} F_1(x_1(1), \quad x_L(1)) & & F_M(x_1(1), \quad x_L(1)) \\ & & \\ F_1(x_1(N), \quad x_L(N)) & & F_M(x_1(N), \quad x_L(N)) \end{pmatrix}$$

$b=(y(1), y(2), \quad y(N))$, $a=(a_1, a_2 \quad a_M)$

Under the Least squares method requirements, to make the minimum cost function, only need to meet the following conditions:

$$(A^T A)a = A^T b$$

D. Genetic Algorithm for Automatic Evolution of Programming

- The first step: determine the predictive model operator sets and data sets.
- Step two: initialization expression tree species.
- Step three: assessment of species groups in each of the individual.
- Step four: the population of the parent tree of the expression of individual hybridization and mutation, produce the next generation of offspring population of individuals.
- Step five: the new individual produced offspring replace the parent generation of individuals with poor fitness.
- Step Six: If the termination conditions not met, then switch to the second step.
- Step seven: Test the model. The model generation of the return of the original problem, the obtained model data were compared with the original problem data to establish the model is reasonable and practical. If you find that the model is unreasonable, then back to the first step to re-modeling.
- Step eight: Application of model output.

As can be seen from the above steps, Genetic algorithm evolution of computer programming automatic modeling method and the traditional method of mathematical modeling, compared with the two main characteristics:

- Evolutionary modeling method need not determine in advance a fixed model structure, only the question of the characteristics identified under the model structure of some of the basic unit,but the traditional mathematical modeling methods, such as: linear regression and exponential regression prediction model such as the economy forecasting model should be determined structure before the model;
- Evolutionary modeling is not using the classical method for solving mathematical problems, but using a simulation of the evolution of evolution algorithms to search the optimal solution, which is the traditional modeling methods in terms of ideology, methods are essential difference.

III. THE ADDEDVALUE OF REAL ESTATE REGRESSION ANALYSIS AND FORECASTING

A. The Added Value of the Real Estate Regression Analysis

The added value of Service Industry is a gross increase production value between this period settlement and the previous settlement cycle in retail industry and service industry. The Service Industry include transportation, wholesale and retail, accommodation and catering industry, the financial sector and real estate industry. In recent years, the total value of GDP is great increase in real estate, which brings large effect to the total value of GDP in the overall Service Industry. First set: function types: complex functions, the tree depth: 9, the maximum number of breeding: 10000, the maximum power times: 6, pop-size = 60, $-100 < a < 100$

$$d = \min(\|x\|^2) = \min(\sum_{i=1}^{N}(y(i) - \sum_{k=1}^{M} a_k F_k(x_1(i), \quad, x_L(i))))^2 < 10^{-5}$$

Automatically generated based on genetic algorithm model of the final steps in the simulation equation is:

$$y = 25007.58 + 2617.24x - \frac{356687.42}{x} - 615.45x^2 + \frac{1778282.28}{x^2} + 45.16x^3$$
$$- \frac{4311442.41}{x^3} - 1.45x^4 + \frac{4966595.34}{x^4} + 0.018x^5 - \frac{2103720.99}{x^5}$$

And with the linear regression equation $y = -2352.71 + 326.54x$ and exponential regression equation $y = 251.976 + e^{4.51411 + 0.16307x}$ Simulation equations were derived from actual results and fitting results are compared in TableI.From the TableI it is obvious to see automatically generated using genetic algorithms simulation equation than the other two model, the average error is small lot, and the fitted value and actual value of the same. At the same time the forecast for 2008 to be significantly higher than the exponential regression and linear regression to be accurate, with the regression curve can be expressed clearly shown the superiority of genetic algorithm look at the Figure2. The Figure 3 and Figure 4 is the exponential regression curve and linear regression curve.

B. The Future Prediction of Real Estate Added Value

In order to obtain the better reference to the related real estate businesses or government, the following table Ⅱ predicts the gross added value of real estate using genetic algorithms for 2010-2013. In that, we can see that the added value of real estate have large proportion.

Whether this change trend is fit to the development of the Service Industry and the actual economic development of our country or not, we should pay much attention to the issue in the future research..

TABLE I. 1978 -2008 YEARS, THE ADDED VALUE OF REAL ESTATE COMPARISON OF REGRESSION MODEL RESULTS

Years	Actual	Genetic algorithm	Exponential	Linear
1978	***79.9***	79.899353	359.776	-2026.17
1979	86.3	86.437934	378.749	-1699.63
1980	96.4	93.36952	401.061	-1373.08
1981	99.9	117.81646	427.3	-1046.54
1982	110.8	78.132456	458.156	-720
1983	121.8	119.48258	494.444	-393.46
1984	162.3	189.03782	537.118	-66.91
1985	215.2	249.55005	587.302	259.63
1986	298.1	297.18416	646.319	586.17
1987	382.6	346.17023	715.723	912.72
1988	473.8	415.79298	797.342	1239.26
1989	566.2	523.05377	893.325	1565.8
1990	662.2	679.35717	1006.202	1892.34
1991	763.7	889.59783	1138.944	2218.89
1992	1101.3	1152.609	1295.049	2545.43
1993	1379.6	1462.3726	1478.629	2871.97
1994	1909.3	1809.6532	1694.518	3198.52
1995	2354	2183.8694	1948.403	3525.06
1996	2617.6	2575.0948	2246.972	3851.6
1997	2921.1	2976.1321	2598.088	4178.15
1998	3434.5	3384.624	3011	4504.69
1999	3681.8	3805.1849	3496.584	4831.23
2000	4149.1	4251.5397	4067.63	5157.77
2001	4715.1	4748.6663	4739.18	5484.32
2002	5346.4	5334.9371	5528.921	5810.86
2003	6172.7	6064.2586	6457.656	6137.4
2004	7174.1	7008.2082	7549.846	6463.95
2005	8243.8	8258.1676	8834.26	6790.49
2006	9664	9927.4538	10344.73	7117.03
2007	12277.5	12153.447	12121.039	7443.57
	Correlation	0.9999	0.9951	0,8859
	Average error	80.217	131.882	1725.94
2008	12720	12908.5	14209.977	7770.03

IV. CONCLUSION

For the real estate added value regression analysis and forecast through genetic algorithm is automatically generated the regression equation, in order to change the old, traditional fixed prediction model regression analysis and forecast.

Obviously, the genetic algorithm obtain the data of the real estate added value is higher accuracy than other regression analysis. By this relatively high fitting regression equation, forecast some of reliability value that give the relevant government departments and related industries of real estate investment some great reference value in scientific decision-making.

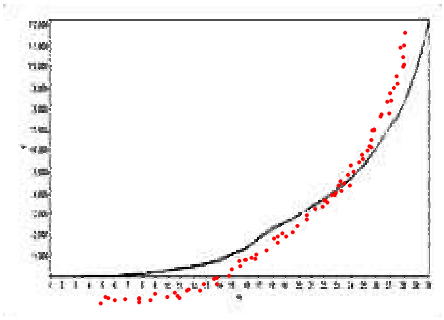
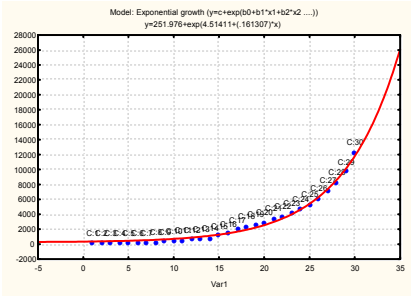


Figure 2： Based on genetic algorithm regression curve.



: Figure 3：Exponential regression curve

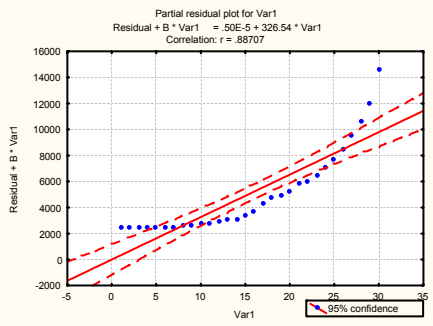


Figure 4：Linear regression curve

TABLE II. THE PREDICTION FROM 2010 TO 2013

number	year	prediction
1	2010	23951.03
2	2011	30333.24
3	2012	38394.32
4	2013	48460.58

REFERENCES

[1] De Jong K A. Anlysis of the Behavior of a Class of Genetic Adaptive Systems. Doctoral Dissertation,University of Michigan,1975.

[2] Hironori Kasahara ,Seinosuke Narita.Practical multiprocessor scheduling algorithms for efficient parallel processing, IEEE Tran on Computers,vol.11,1984.pp,1023-1029

[3] Zhong Qiuxi, Xie Tao .Task matching and scheduling by using genetic algorithms[J].Journal of Computer Research and Development,vol.7,2000 pp,1197-1203

[4] De Jong K. K. Genetic-algorithms-based Learning: An Artificial Intelligence Approach, Vol..Morgan KaufmannPublishers.1990,pp.611-638

[5] Simon Haykin. Neural Networks: A Comprehensive Foundation.China Machine Press,2004.