A SVR BASED FORECASTING APPROACH FOR REAL ESTATE PRICE PREDICTION

DA-YING LI[1], WEI XU[2], HONG ZHAO[3], RONG-QIU CHEN[1]

[1]School of Management, Huazhong University of Science and Technology, Wuhan, Hubei, 430074, China
[2]School of Information, Renmin University of China, Beijing, 100872, China
[3]School of Management, Graduate University of Chinese Academy of Sciences, Chinese Academy of Sciences, Beijing, 100190, China
E-MAIL: lidaying@vip.sina.com, xuwcas@hotmail.com, zhaohong@gucas.ac.cn, rongqiu@mail.hust.edu.cn

Abstract:

The support vector machine (SVM), proposed by Vapnik (1995), has been successfully applied to classification, cluster, and forecast. This study proposes support vector regression (SVR) to forecast real estate prices in China. The aim of this paper is to examine the feasibility of SVR in real estate price prediction. To achieve the aim, five indicators are selected as the input variables and real estate price is used as output variable of the SVR. The quarterly data during 1998-2008 are employed as the data set to construct the SVR model. With the scenarios, real estate prices in future are forecasted and analyzed. The forecasting performance of SVR model was also compared with BPNN model. The experimental results demonstrate that based on the mean absolute error (MAE), the mean absolute percentage error (MAPE) and the root mean squared error (RMSE), the SVR model outperforms the BPNN model and the SVR based approach was an efficient tool to forecast real estate prices.

Keywords:

Support vector machine (SVM); real estate price; forecast

1. Introduction

Real estate industry is one of the most important industries in China, and it makes a strong impact on China's economy development. In recent years, a number of factors (e.g. China's economy development, Consumer Price Index, and financial crisis) affect the real estate prices, and the salience of each individual factor varies from time to time, so real estate prices show large fluctuation. Real estate prices in China, especially in some big cities of China, have continued to rise since 2003, and real estate prices increased 9.7% in 2004, and 7.6% in 2005 [1]. However, in the face of the global financial crisis, real estate prices have slowed down and even shut down in 2008.

Real estate market in China has recently received considerable attention from the government, real estate companies, and academicians. So far, a number of studies have focus on real estate market. Grebler [2] pointed out that some indicators (e.g., income, the Consumer Price Index (CPI), seasonal factors, vacancy rate, and previous real estate prices) can be used to predict the real estate price. Nellis and Longbottom [3] found that the determinants of real estate price were real disposable income, loan interest rates, and total loans. Reichert [4] identified that regional real estate prices reacted uniformly to certain national economic factors such as mortgage rates, and local factors (e.g., population shifts, employment, and income trends) often have a unique impact on real estate prices. Clapp and Giaccotto [5] studied the influence of economic variables on local real estate price dynamics and found that some variables such as population, employment and income had a considerable forecasting ability for housing price. To identify the factors influencing yearly urban real estate prices, Potepan [6] used a number of indicators including the privately owned dwelling price index, monthly rent based on the hedonic model, land price, medium income, population, quality of public services, crime rate, air pollution, non-dwelling consumable price, mortgage rate, construction cost, farm land price, land restriction and so on. Baffoe-Bonnie [7] used a nonstructural estimation technique to analyze the dynamic effects of four key macroeconomic variables on real estate prices and the stock of houses sold. The impulse response functions derived from the VAR suggested that the real estate market was very sensitive to shocks in the employment growth and mortgage rate at both national and regional levels. Malpezzi [8] used a time-series cross-section regression to find that real estate prices did not change according to a random walk, at least could be partially forecasted.

Among these studies, most of them have focus on the determinants of real estate prices. However, few focused on real estate price prediction, particularly these in China. Seko [9] thought that there was a strong correlation between Japanese dwelling prices and economic fundamentals in some areas, and used a time series model to forecast housing prices with some indicators, including average sales price of private ownership dwelling, annual household income, population, new-started dwelling, the CPI and the vacancy rate. Case and Shiller[10] forecasted the real estate price and excess returns in the real estate market using the percentage change in real per capita income, real construction costs, adult population, marginal tax rate and housing starts. Anglin [11] set up a VAR model to forecast Toronto real estate prices, with the predictors' three lags for the average real estate price growth rate, CPI, mortgage rates and unemployment rate. Yan et al. [12] pointed out the limitations and shortcomings in the previous research and proposed an integrated method for real estate price forecast based on TEI@I.

This paper proposes a support vector regression (SVR) based method to forecast real estate prices in China. Firstly, five indicators are selected as the input variables and real estate price is used as output variable of the SVR. Then the quarterly data during 1998-2008 are employed as the data set to construct the SVR model. With the scenarios, real estate prices in future are forecasted and analyzed. Finally, the forecasting performance of SVR model was also compared with BPNN model.

2. The SVR based forecasting approach

SVM was originally used for classification but its principle was extended to the task of regression and forecast as well [13-16]. In this paper, we focus on support vector regression (SVR) for regression and forecast. SVR is a learning machine which estimates a function according to a given data set $G=\{(x_i, y_i)\}^n$, where x_i denotes the input vector, y_i denotes the output value and n denotes the total number of the data. Initially, a SVR model performs a regression by an ε-insensitive loss, and Scholkopf [17] developed the $\varepsilon-$SVR model and proposed a $v-$SVR model which is a modification of the $\varepsilon-$SVR model. It minimizes ε automatically, and adjusts the accuracy level according to the data set. The $v-$SVR model is expressed as follows:

$$\min\ R_{SVR}(C) = R_{emp} + \frac{1}{2}\|\omega\|^2 = C\left(v\varepsilon + \frac{1}{n}\sum_{i=1}^{n}(\xi_i + \xi_i^*)\right) + \frac{1}{2}\|\omega\|^2$$

$$\text{Subject to: } \left((\omega\cdot x_i)+b\right) - y_i \le \varepsilon + \xi_i$$
$$y_i - \left((\omega\cdot x_i)+b\right) \le \varepsilon + \xi_i^*$$
$$\xi_i^{(*)} \ge 0,\ \varepsilon \ge 0,\ v \ge 0 \qquad (1)$$
$$i = 1,\ \ldots,\ n$$

where R_{SVR} and R_{emp} represent the regression and empirical risks, respectively, $\|\omega\|^2/2$ denotes the Euclidean norm and C denotes a cost function measuring the empirical risk.

Based on the $v-$SVR model, the SVR-based forecasting method consists of five steps as follows (See Figure 1).

Step 1: Data sampling. Data can be collected from different resources and different types. There are also various inconsistencies and missing points for the market. So, the most reliable data should be selected.

Step 2: Data preprocessing. The selected data may need to be transformed into certain appropriate range for the network learning by logarithmic transformation, difference or other methods. Then the data set should be divided into the training set and the testing set.

Step 3: Training. The parameters of the SVM are learned using the training set.

Step 4: Testing. The SVM is validated using the testing set and the final network architecture of SVM is determined.

Step 5: Forecasting. With the scenarios, the future values of time series can be forecasted by using the SVR based forecasting approach.

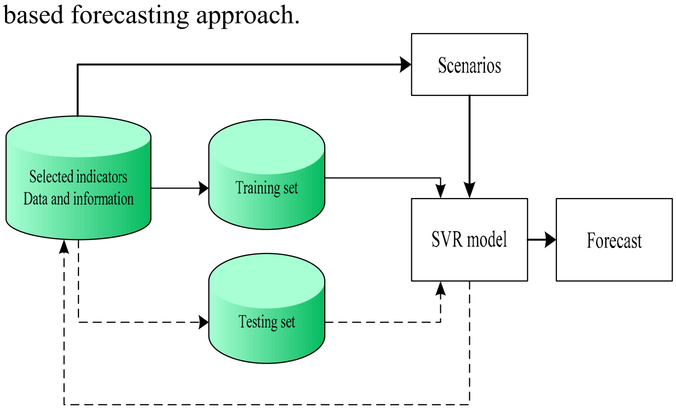


Figure 1. ***The main process of the SVR method***

3. Application of the SVR based forecasting method

In this section, the SVR based forecasting method applies to real estate price prediction in China. Firstly, using the selected indicators, the SVR based forecasting method is constructed based on the history data. Then some scenarios of real estate market are set. Finally, with the scenarios, real estate prices in future are forecasted and analyzed.

3.1. Construction of the SVR based method

Based on the literatures [2-12] and our analysis, five indicators including disposable income, Consumer Price Index (CPI), investment in real estate development (including commercial housing, commercial dwelling, office building and commercial business building), loan interest rates, and lagged real estate price are selected as the input variables, and real estate price is used as output variable of the SVR. The basic structure of the SVR based forecasting model is illustrated in Figure 2.

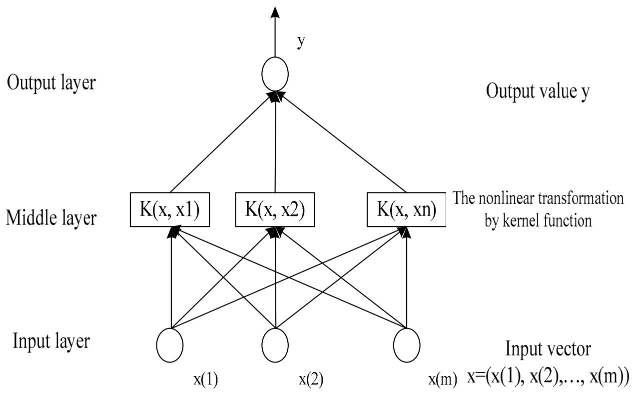


Figure 2. *The based structure of the SVR model*

As can be seen from Figure 2, the structure of SVR model is similar to that of neural networks. Similarly, the SVR model also consists of an input layer, middle layer and output layer. The difference is that every node output of the middle layer is a support vector transformed by the kernel function. Usually, the Gaussian function is used as a kernel function. Note that the SVR could overcome the important drawbacks of the neural network, such as over-fitting and local minima.

During the computing process, the quarterly data during 1998-2006 are employed as the training set to construct the SVR model, and the quarterly data during 2007-2008 are used as the testing set to verify the SVR model.

3.2. Scenarios

Observations showed that within last decade, the growth rate of disposable income fluctuated between 5% and 18%. The growth rate of CPI fluctuated from -3% to 2%. The growth rate of investment in real estate development fluctuated from 9% to 40%. Loan interest rates (one year) fluctuated between 5.31% and 8.48%. Forecasting the real estate price is carried out with three scenarios.

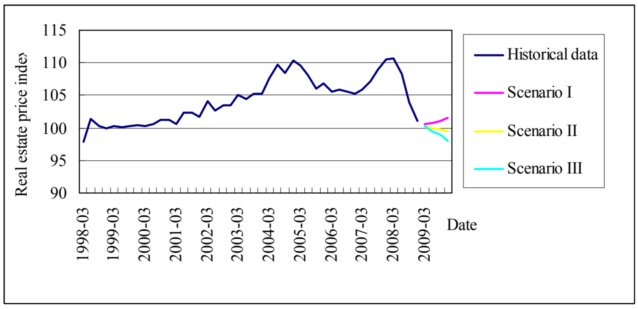
Scenario I. With the governments' effective measures and policy, the growth rate of disposable is projected to 10%, the growth rate of CPI is projected to -0.5%, the growth rate of investment in real estate development is to 30%, and loan interest rates remain 5.31%.

Scenario II. Based on the current situation, the growth rate of disposable is projected to 8%, the growth rate of CPI is projected to -1%, the growth rate of investment in real estate development is to 25%, and loan interest rates remain 5.31%.

Scenario III. If the global financial crisis will last for a long period or even worsen, the growth rate of disposable is projected to 5%, the growth rate of CPI is projected to -2%, the growth rate of investment in real estate development is to 15%, and loan interest rates are declined to 4%.

3.3. Real estate price prediction

With the three scenarios, the real estate prices in future are shown in Figure 3.



As can be seen from Figure 3, the real estate price will slow increase in Scenario I, but in Scenario II, the real estate price will slow decrease, and in Scenario III, the price will accelerative decrease.

4. Evaluation

To validate the proposed approach, we conduct an experiment by using real estate historical data. The experiment compares the proposed approach with BPNN method. The traditional performance indices including the mean absolute error (MAE), the mean absolute percentage error (MAPE) and the root mean squared error (RMSE) can be used as measures of forecast accuracy.

The quarterly data from 1998 to 2006 are employed as the training set, and the data from 2007 to 2008 are used as the testing set. The performances of the SVR based forecasting approach and the BPNN based forecasting method [18, 19] in testing set were compared by using MAE, MAPE and RMSE, and the results were shown in

Table 1.

Table 1. The performances of the SVR and BPNN

	MAE	MAPE	RMSE
SVR	1.363	0.013	1.893
BPNN	1.788	0.017	2.481

The experimental results demonstrate that the SVR model outperforms the BPNN model based on the MAE, the MAPE and the RMSE. The SVR based forecasting approach was an efficient tool to forecast real estate prices.

5 Conclusions

This paper proposes a SVR based forecasting approach for real estate prediction in China. Five indicators including disposable income, CPI, investment in real estate development, loan interest rates, and lagged real estate price are employed to forecast real estate prices. The experimental results demonstrate that the SVR model outperforms the BPNN model, and the SVR based forecasting approach was an efficient tool to forecast real estate prices.

In future work, besides demand and support indicators, the psychological and behavior indicators such as Consumer Sentiment Index (CSI) will be further considered in analysis of real estate market. Analysis and forecast of real estate prices in big cities, especially in Beijing, Shanghai, and Shenzhen, will be further investigated. Finally, an online intelligent support system can be designed and developed for analysis and forecast real estate prices in China.

References

[1] "National Bureau of Statistics", People's Republic of China. China Monthly Economic Indicators. National Bureau of Statistics.

[2] L. Grebler, "The Inflation of Housing Price", Its Extent, Cause and Consequences. Lexington Books, 1979.

[3] J. Nellis, and J. Longbottom, "An empirical analysis of the determination of house prices in the United Kingdom", Urban Studies, Vol. 18, No. 1, pp. 9-21, 1981.

[4] A. K. Reichert, "The impact of interest rate, income and employment upon regional housing prices", Journal of Real Estate Finance and Economics, Vol. 3, No. 4, pp. 373-391, 1990.

[5] J. M. Clapp, and C. Giaccotto, "The influence of economic variables on local housing price dynamics", Journal of Urban Economics, Vol. 36 No. 2, pp. 161-183, 1994.

[6] M. J. Potepan, "Explaining intermetropolitan variation in housing price, rent and land price", Real Estate Economic, Vol. 24, No. 2, pp. 219-245, 1998.

[7] J. Baffoe-Bonnie, "The dynamic impact of macroeconomic aggregates on housing prices and stock of houses: a national and regional analysis", The Journal of Real Estate Finance and Economics, Vol. 17, No. 2, pp. 179-197, 1998.

[8] S. Malpezzi, "A simple error correction model of housing prices", Journal of Housing Economics, Vol. 8, No. 1, pp. 27-62, 1999.

[9] M. Seko, "Housing prices and economic cycles", Paper Presented in The International Conference on Housing Market and the Macro Economy, Hong Kong, 2003.

[10] K. E. Case, and R. J. Shiller, "Forecasting prices and excess returns in the housing market", AREUEA Journal, Vol. 18, No. 3, pp. 253-273, 1990.

[11] P. Anglin, "Local dynamics and contagion in real estate markets", The International Conference on Real Estates and Macro Economy, Beijing, 2006.

[12] Y. Yan, W. Xu, H. Bu, Y. Song, W. Zhang, H. Yuan, and S. Y. Wang, "Method for housing price forecasting based on TEI@I methodology", Systems Engineering-Theory & Practice, Vol. 27, No. 7, pp. 1-9, 2007.

[13] V. N. Vapnik, "The Nature of Statistical Learning Theory". Springer, New York, 1995.

[14] A. Smola, and B. Schoelkopf, "A tutorial on support vector regression", Statistics and Computing, Vol. 14, pp. 199-222, 2004.

[15] B. Schoelkopf, A. Smola, R. Williamson, and P. Bartlett, "New support vector algorithms", Neural Computation, Vol. 12, pp. 1083-1121, 2000.

[16] K. R. Muller, A. J. Smola, G. Ratsch, B. Scholkopf, J. Kohlmorgen, and V. N. Vapnik, "Prediction time series with support vector machines", In W. Gerstner, A. Germond, M. Hasler, and J. D. Nicoud (Eds.), Proceedings of International Conference on Artificial Neural Networks, Lausanne, Switzerland. Lecture Notes in Computer Science, Vol. 1327, pp. 999-1004, 1997.

[17] B. Schoelkopf, A. Smola, and R. Williamson, "Shrinking the tube: a new support vector regression algorithm", In M. S. Kearns, S. A. Solla, and D. A. Cohn (Eds.), Advances in Neural Information Processing Systems, Cambridge, MA. MIT Press Vol. 11, pp. 330-336, 1999.

[18] Y. S. Murat, and H. Ceylan, "Use of artificial neural networks for transport energy demand modeling", Energy Policy, Vol. 34, pp. 3165–3172, 2006.

[19] P. Mandal, T. Senjyu, and T. Funabashi, "Neural networks approach to forecast several hour ahead electricity prices and loads in deregulated market", Energy Conversion and Management, Vol. 47, pp. 2128–2142, 2006.