ISSN:2229-6247

Housing Price Forecasting based on Stochastic Time Series Model

Jianhua GUO
Department of Mathematics, Hunan University of Science and Technology, Xiangtan, China
jhguo888@163.com

Abstract:
Whether the real estate industry robustly develops or not largely affects national macroeconomic development and national quality of life. By analyzing monthly averaged prices of commercial residential building in Changsha City from Jan, 2002 to Dec, 2011, this paper aims to reasonably construct a forecasting model to predict short-term housing price trend and affords reference to homebuyer and investors, what's more, affords technical support to government's policy making. First, how to select rational forecasting model is discussed, and then a price forecasting ARMA model is constructed, lastly, empirical analysis is put forward.
Keywords: housing price, forecasting, stochastic time series, ARMA model

1. INTRODUCTION

Commercial housing price affects national macroeconomic development and national quality of life in a large extent. Variation of commercial housing price also deeply impact the running of national economy, so, study on housing price is a significant job.

So far, many scholars studied influencing factor of housing price and acquired plenty of achievement, but the quantitative research about housing price is in the infant stage. Among which, Malpezzi(1999) studied 133 American cities' commercial housing price index and concluded that the housing price is not a random walk but partly predictable. Clapp J.H. and Giaccotto C.(2002), by means of autoregression model and relying on the past quarterly housing prices during 1976 and 1997, forecasted the next quarterly price. Seko(2003) studied the relationship between dwelling price and macro economy in Japan. Wedyawati W. and Meiliu L.(2004) made a price for commercial residential building by constructing data mining system between homebuyer and seller and by aid of regression technique. Okmyung B.(2004) compared the parametric and nonparametric price forecasting methods. Bradford C. et al(2004) compared 4 housing price forecasting methods including Dubinrsquos -krigining local forecasting model, Clapprsquou local regression model, local instance forecasting and nearest-neighbor method. Anglin(2006),by introducing the averaged growth rate of lagged 3-term housing prices, CPI, interest rate of mortgage lending and rate of unemployment and constructing VAR model, forecasted the housing price variation in Toronto. Liu,J.G. and Zhang,X.L. et al(2006) estimated property price making use of fuzzy-neural network technique. Zhou,Wei-Xing and Didier,S.(2008) analyzed the housing price indexes of Las Vegas during June,1983 and March, 2005 and found out the seasonal trend, then, they forecasted the next year's housing price index by moving average method. Min Hwang John M. Quigley(2010), based on Singaporean housing prices during 1990 and 2000, constructed a price forecasting model and found that the housing price submitted to mean-reversion model and was in connection with region.

2. MARKET MODELS

Time series analysis is a widely used quantitative analyzing method and time series model, including linear model and nonlinear model, is an important kind of economic model to depicting economic data's evolving process. Linear model comprise of AR model, MA model and ARMA model, etc., as to any time series, we can ascertain whether it is stationary time series or not by unit-root testing, i.e., the time series is stationary when all characteristic roots are bigger than 1.what's more, if the autocorrelation coefficients with order bigger than 3 all located in confidence interval and converging to zero, we also consider the time series to be stationary, and vice verse.

2.1 AR model

As to time series $\{X_t\}$, if there is

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \varepsilon_t \qquad (1)$$

Then, $\{X_t\}$ is called an autoregression time series, and (1) is called an AR(p) model. Where $\phi_1, \phi_2, \cdots, \phi_p$ are autoregression coefficients, $\varepsilon_t \sim N(0, \sigma^2)$ and $E(\varepsilon_t X_{t-i}) = 0, i = 1, \cdots, p$. If lagged operator B is introduced, then, $X_{t-1} = BX_t$, $X_{t-2} = B^2 X_t$, ……, $X_{t-p} = B^p X_t$,and (1) can be transferred into

$$X_t = \phi_1 BX_t + \phi_2 B^2 X_t + \cdots + \phi_p B^p X_t + \varepsilon_t \qquad (2)$$

Further more, denoting $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$, then

$$\phi(B) X_t = \varepsilon_t \qquad (3)$$

We call $\phi(B) = 0$ is the characteristic equation of the autoregression model AR(p), which is called a stationary autoregression model, if all of characteristic roots $\lambda_i, i = 1, \cdots, p$ of $\phi(B) = 0$ are located outside the unit circle, i.e.,

$$|\lambda_i| > 1, i = 1, \cdots, p \qquad (4)$$

A time series evolving as a stationary model AR(p) is called a stationary AR(p) time series. The necessary and sufficient conditions for a stationary time series to be an AR(p) time series is that the partial correlation function is p-step clipped, i.e., $\{X_t\}$ only has p nonzero partial correlation coefficients ψ_{kk}.

In fact, because of the randomicity of sample, all ψ_{kk} will not equal to zero though k>p, instead, $\psi_{kk} \sim N(0, 1/n)$, where n denotes the sample length, and as for enough large n, there is $P\{|\psi_{kk}| < 2/\sqrt{n}\} = 0.955$, thus, we generally recur to interval-testing method to ascertain whether a time series is an AR(p) time series or not.

Note, if $\{X_t\}$ is an AR(p) time series, the partial correlation function ψ_{kk} is p-step clipped, but its autocorrelation function ρ_k is trailing, i.e., attenuating according to a negative exponential function.

2.2 MA model

As to time series $\{X_t\}$, if

$$X_t = \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \cdots - \theta_q\varepsilon_{t-q} \qquad (5)$$

Then, $\{X_t\}$ is called a moving average time series, and (5) is called a MA(q) model. Where $\theta_1, \theta_2, \cdots, \theta_p$ are moving average coefficients. Similarly, if lagged operator B is introduced, and denoting $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$, (5) can be transferred into

$$X_t = \theta(B)\varepsilon_t \qquad (6)$$

We call $\theta(B) = 0$ is the characteristic equation of the moving average model MA(q), which is called a reversible moving average model, if all of characteristic roots $\lambda_i, i = 1, \cdots, p$ of $\theta(B) = 0$ are located outside the unit circle, i.e., $|\lambda_i| > 1, i = 1, \cdots, p$

A time series evolving as a reversible model MA(q) is called a reversible MA(q) time series. The necessary and sufficient conditions for a stationary time series to be an MA(q) time series is that the auto correlation function is q-step clipped, i.e., $\{X_t\}$ only has q nonzero auto correlation coefficients ρ_k. It means that, as to a MA(q) time series, X_t and X_s are irrelevant when $|t - s| > q$.

In fact, because of the randomicity of sample, all ρ_k will not equal to zero though k>q, instead, $\rho_k \sim N(0, \frac{1}{n}(1 + 2\sum_{i=1}^{q}\rho_i^2))$, where n denotes the sample length, and as for enough large n, there is $P\{|\rho_k| < \frac{2}{\sqrt{n}}\sqrt{1 + 2\sum_{i=1}^{q}\rho_i^2}\} = 0.955$, thus, we generally recur to interval-testing method to ascertain whether a time series is an MA(q) time series or not.

Similarly as AR(p) time series, if $\{X_t\}$ is a MA(q) time series, the auto correlation function ρ_k is q-step clipped, but its partial correlation function ψ_{kk} is trailing, i.e., attenuating according to a negative exponential function.

2.3 ARMA model

If time series $\{X_t\}$ satisfies

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \cdots - \theta_q\varepsilon_{t-q} \quad (7)$$

Then, we call a autoregression-moving average time series, and call (7) a autoregression-moving average model, i.e., ARMA(p,q) model.

Where $\phi_1, \phi_2, \cdots, \phi_p$, $\theta_1, \theta_2, \cdots, \theta_p$ and $\varepsilon_{t-1}, \cdots, \varepsilon_{t-q}$ are same as in subsection 2.1 and subsection 2.2.

Also, when introducing lagged operator B, (7) become into

$$\phi(B)X_t = \theta(B)\varepsilon_t \qquad (8)$$

Obviously, AR(p) model and MA(q) model are special cases of ARMA(p,q) model.

If a stationary time series $\{X_t\}$ has clipped partial correlation function and trailing autocorrelation function, $\{X_t\}$ is an AR time series; If a stationary time series $\{X_t\}$ has clipped autocorrelation function and trailing partial correlation function, $\{X_t\}$ is a MA time series; If the partial correlation function and autocorrelation function of $\{X_t\}$ all are trailing, it is an ARMA time series.

Note

1) In reality, many time series are not stationary time series, then, we can translate them into stationary ones by difference method.

2) We can make use of the residual error testing method or fitting testing method to decide whether an AR(p) model, MA(q) model or an ARMA(p,q) model is appropriate to depict a time series or not.

3. Sampling And Modeling

3.1 Sampling and stationarity test

In this paper, we select monthly housing prices of Changsha City from Jan,2009 to Dec,2011(shown in figure 1) as within sample data for modeling, while selecting monthly housing prices of January and February, 2012 to detect price forecasting precision.

First, as to figure 1, we can see that there is visible uptrend component in price series and the price series is non-periodic, so, it is reasonable to decide the price series be non-stationary. In addition, the autocorrelation function is attenuating as a smooth exponential function while the partial correlation function is clipped with positive peak value, which, just as unit-root testing, shows the price series is non-stationary, see figure 2 and figure 3, respectively. Therefore, we make differential operation for the original housing price series (see figure 4), according to its autocorrelation function and partial correlation function, as well as unit-root testing (see figure 5 and figure 6, respetively), we can ascertain that the differenced series is stationary.

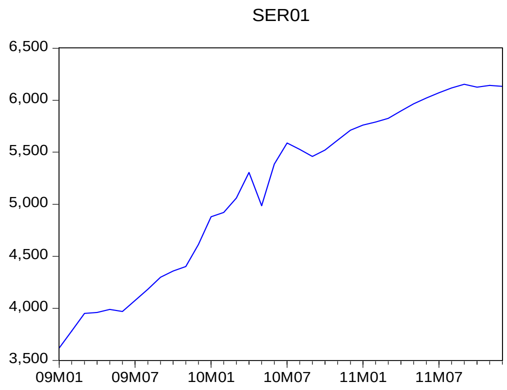


Figure 1 The housing price series of Changsha City during Jan, 2009 and Dec, 2011
(Unit: CNY per square meter)

Date: 03/12/12 Time: 22:58
Sample: 2009M01 2011M12
Included observations: 36

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.919	0.919	32.990	0.000
		2	0.844	0.001	61.663	0.000
		3	0.777	0.012	86.722	0.000
		4	0.700	-0.105	107.65	0.000
		5	0.618	-0.071	124.52	0.000
		6	0.530	-0.105	137.32	0.000
		7	0.439	-0.074	146.43	0.000
		8	0.352	-0.045	152.50	0.000
		9	0.272	-0.012	156.25	0.000
		10	0.193	-0.048	158.21	0.000
		11	0.110	-0.085	158.86	0.000
		12	0.033	-0.037	158.92	0.000
		13	-0.029	0.022	158.97	0.000
		14	-0.089	-0.048	159.47	0.000
		15	-0.140	0.002	160.74	0.000
		16	-0.176	0.028	162.87	0.000
		17	-0.242	-0.245	167.10	0.000
		18	-0.290	0.011	173.49	0.000
		19	-0.318	0.035	181.60	0.000
		20	-0.335	0.045	191.18	0.000
		21	-0.372	-0.187	203.81	0.000
		22	-0.398	0.002	219.28	0.000
		23	-0.414	-0.028	237.32	0.000
		24	-0.424	-0.008	257.79	0.000
		25	-0.421	0.014	279.80	0.000
		26	-0.408	0.043	302.59	0.000
		27	-0.393	0.007	326.04	0.000
		28	-0.371	-0.014	349.62	0.000
		29	-0.342	-0.023	372.44	0.000
		30	-0.305	0.045	393.62	0.000
		31	-0.261	0.045	412.24	0.000
		32	-0.215	-0.014	428.05	0.000
		33	-0.166	0.062	440.58	0.000
		34	-0.118	-0.022	450.09	0.000
		35	-0.062	0.030	455.35	0.000

Figure 2 Figure of autocorrelation function and partial correlation function of price series

Null Hypothesis: SER01 has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 2 (Automatic based on AIC, MAXLAG=9)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-0.105300	0.9926
Test critical values:	1% level	-4.262735	
	5% level	-3.552973	
	10% level	-3.209642	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(SER01)
Method: Least Squares
Date: 03/12/12 Time: 23:00
Sample (adjusted): 2009M04 2011M12
Included observations: 33 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
SER01(-1)	-0.015266	0.144974	-0.105300	0.9169
D(SER01(-1))	-0.343628	0.209131	-1.643121	0.1115
D(SER01(-2))	-0.308079	0.193387	-1.593068	0.1224
C	239.9401	520.4076	0.461062	0.6483
@TREND(2009M01)	-2.441900	11.93560	-0.204590	0.8394

R-squared	0.192827	Mean dependent var	66.18182
Adjusted R-squared	0.077516	S.D. dependent var	118.8632
S.E. of regression	114.1634	Akaike info criterion	12.45187
Sum squared resid	364931.8	Schwarz criterion	12.67861
Log likelihood	-200.4558	Hannan-Quinn criter.	12.52816
F-statistic	1.672240	Durbin-Watson stat	1.950402
Prob(F-statistic)	0.184368		

Figure 3 Unit-root test results

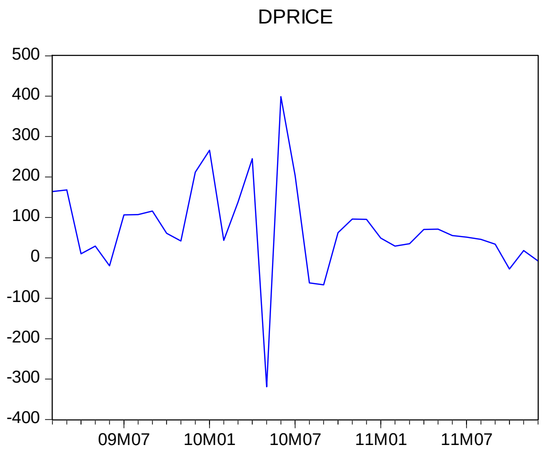


Figure 4 The differenced price series of the original housing price series

Date: 03/12/12 Time: 23:10
Sample: 2009M02 2011M12
Included observations: 35

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.205	-0.205	1.6010	0.206
		2	-0.178	-0.230	2.8493	0.241
		3	0.185	0.102	4.2294	0.238
		4	-0.027	-0.000	4.2601	0.372
		5	0.027	0.085	4.2908	0.508
		6	0.190	0.212	5.8970	0.435
		7	-0.058	0.062	6.0532	0.534
		8	-0.165	-0.132	7.3567	0.499
		9	-0.018	-0.177	7.3722	0.598
		10	0.082	-0.036	7.7163	0.657
		11	0.114	0.145	8.4167	0.676
		12	-0.067	0.031	8.6664	0.731
		13	0.028	0.112	8.7126	0.794
		14	-0.082	-0.045	9.1247	0.823
		15	0.003	-0.014	9.1254	0.871
		16	0.085	-0.047	9.6180	0.886
		17	0.009	-0.037	9.6244	0.919
		18	-0.133	-0.118	10.972	0.896
		19	-0.017	-0.038	10.995	0.924
		20	-0.045	-0.064	11.169	0.942
		21	-0.036	-0.053	11.292	0.957
		22	-0.038	-0.132	11.437	0.968
		23	-0.055	-0.099	11.763	0.974
		24	-0.037	-0.042	11.927	0.981
		25	-0.010	0.036	11.939	0.987
		26	-0.005	0.008	11.943	0.991
		27	-0.010	0.005	11.958	0.994
		28	0.009	0.017	11.973	0.996
		29	0.009	0.048	11.992	0.998
		30	0.021	0.031	12.105	0.998
		31	-0.013	0.004	12.163	0.999
		32	-0.020	-0.037	12.336	0.999
		33	-0.027	-0.033	12.803	0.999
		34	-0.016	-0.045	13.119	1.000

Figure 5 Figure of autocorrelation and partial correlation functions of the differenced price series

Null Hypothesis: DPRICE has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 1 (Automatic based on AIC, MAXLAG=8)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-6.006053	0.0001
Test critical values:	1% level	-4.262735	
	5% level	-3.552973	
	10% level	-3.209642	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(DPRICE)
Method: Least Squares
Date: 03/12/12 Time: 23:11
Sample (adjusted): 2009M04 2011M12
Included observations: 33 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DPRICE(-1)	-1.671236	0.278259	-6.006053	0.0000
D(DPRICE(-1))	0.316093	0.174724	1.809097	0.0808
C	181.7769	52.74650	3.446237	0.0018
@TREND(2009M02)	-3.677396	2.151592	-1.709151	0.0981

R-squared	0.673774	Mean dependent var	-5.333333
Adjusted R-squared	0.640026	S.D. dependent var	187.0068
S.E. of regression	112.2000	Akaike info criterion	12.39166
Sum squared resid	365076.3	Schwarz criterion	12.57305
Log likelihood	-200.4623	Hannan-Quinn criter.	12.45269
F-statistic	19.96512	Durbin-Watson stat	1.956295
Prob(F-statistic)	0.000000		

Figure 6 Unit-root test results of the differenced price series

3.2 Modeling and Forecasting

According to figure 5, all autocorrelation coefficients and partial correlation coefficients locate in confidence interval when k>1, and convergence to zero, so, we can model the original price series with one of the ARI(1,1),MAI(1,1) or ARIMA(1,1,1) models. Relying on the EVIEWS software, the AIC values of three models are 12.44865, 12.38843, 12.44639, respectively, and by means of theory of minimizing criterion function, we first select MAI(1,1) as our candidate time series model. However, the AIC value is not the sufficient condition for selecting the optimal model. Here, we do significance test for parameters of the selected model or do test for randomness of residual error, if pass the test, the selected model with least AIC value is the optimal model, or else, select model with minor AIC value and decide whether parametric test or residual error test passed, and so on, until the appropriate model is decided.

In this paper, after trial and error, the ARIMA(1,1,1) is identified as the optimal one. Come next, we estimate model's parameters. Among all favorite parameter estimation methods, including moment estimation, maximum likelihood estimation and nonlinear least-square(NLS) estimation, etc, here, in EVIEWS software, we adopt NLS method to estimate parameters, results as shown in figure 7.

Dependent Variable: DPRICE
Method: Least Squares
Date: 03/13/12 Time: 17:24
Sample (adjusted): 2009M03 2011M12
Included observations: 34 after adjustments
Convergence achieved after 16 iterations
MA Backcast: 2009M02

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	69.06684	14.00293	4.932315	0.0000
AR(1)	0.112233	0.535655	0.209525	0.8354
MA(1)	-0.395474	0.510031	-0.775392	0.4440

R-squared	0.081758	Mean dependent var	69.17647
Adjusted R-squared	0.022516	S.D. dependent var	118.3437
S.E. of regression	117.0038	Akaike info criterion	12.44639
Sum squared resid	424386.8	Schwarz criterion	12.58107
Log likelihood	-208.5886	Hannan-Quinn criter.	12.49232
F-statistic	1.380074	Durbin-Watson stat	1.954560
Prob(F-statistic)	0.266587		

Inverted AR Roots	.11
Inverted MA Roots	.40

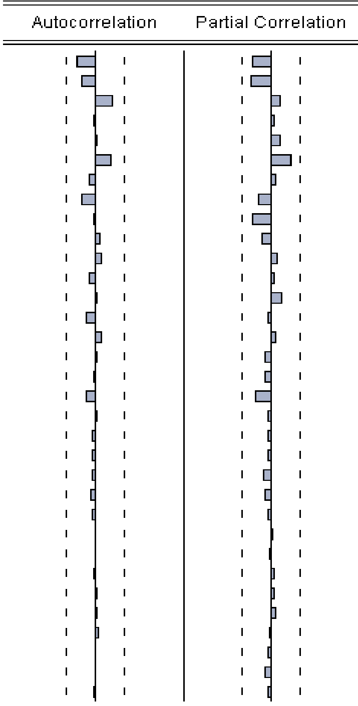
Figure 7 estimation of ARMA model

Thus, the expressions of 1-order difference of original housing price series is as following

$$\Delta price_t = 69.0668 + 0.11223\Delta price_{t-1} + \varepsilon_t - 0.395474\varepsilon_{t-1} \quad (9)$$

The randomicity of residual error series, i.e., all autocorrelation coefficients come close to zero when K>1, is an important indicator to decide a model be appropriate or not. Shown as in figure 8, all residual errors are minute and autocorrelation coefficients and partial correlation coefficients all locate in confidence interval, as well as, the P-values of Q statistic are much bigger than 0.05, so, it's reasonable to consider the residual error series as a white noise series and the ARIMA(1,1,1) model is appropriate to depict the housing price series.

Date: 03/13/12 Time: 17:57
Sample: 2009M02 2011M12
Included observations: 34



	AC	PAC	Q-Stat	Prob
1	-0.228	-0.228	1.9276	0.165
2	-0.168	-0.232	3.0124	0.222
3	0.203	0.114	4.6334	0.201
4	-0.009	0.039	4.6367	0.327
5	0.021	0.099	4.6548	0.459
6	0.187	0.223	6.1913	0.402
7	-0.066	0.057	6.3915	0.495
8	-0.165	-0.150	7.6712	0.466
9	-0.017	-0.218	7.6852	0.566
10	0.049	-0.114	7.8060	0.648
11	0.079	0.081	8.1347	0.701
12	-0.066	0.033	8.3741	0.755
13	0.010	0.126	8.3794	0.818
14	-0.106	-0.040	9.0682	0.827
15	0.069	0.057	9.3748	0.857
16	0.016	-0.081	9.3933	0.896
17	-0.011	-0.074	9.4019	0.927
18	-0.104	-0.178	10.234	0.924
19	0.009	-0.036	10.240	0.947
20	-0.044	-0.036	10.410	0.960
21	-0.043	-0.029	10.588	0.970
22	-0.044	-0.094	10.782	0.978
23	-0.051	-0.068	11.066	0.983
24	-0.030	-0.036	11.177	0.988
25	-0.005	0.013	11.180	0.992
26	-0.007	-0.015	11.187	0.995
27	-0.011	0.033	11.210	0.997
28	0.011	0.025	11.234	0.998
29	0.012	0.057	11.273	0.999
30	0.026	-0.015	11.477	0.999
31	-0.004	-0.033	11.482	0.999
32	0.001	-0.068	11.483	1.000
33	-0.013	-0.033	11.705	1.000

Figure 8 white noise test of the residual error series

According to the constructed ARIMA(1,1,1) model, we forecast the next two months' housing prices in table 1 as follows

Table 1 the forecasted housing price and relative error to actual price

Month	***Actual housing price***	***Forecasted value***	***Relative error***
Jan,2012	6154	6202	0.78％
Feb,2012	6129	6225	1.56％

4. CONCLUSION

The ARMA model forecasts next values based on preceding time series data and it can be used to forecast short-term housing price. The empirical analysis indicates that this model's forecasting precision is satisfactory, and only with some simple modification, this model may be used in other industries. Generally speaking, housing price affected by many factors is nonlinear changing, but there few of literatures study on housing price series from nonlinear point of view, so, we can say that the ARMA model is a comparatively excellent forecasting model for its convenience of application and its strict mathematical assurance, besides, differ from other models such as regression model relying on great deal of related data , the ARMA model only rely on historical data of some certain variable, which make its unique predominance in short-term forecast.

REFERENCE

[1] Malpezzi S. A simple error correction model of housing prices[J]. Journal of Housing Economics, 1999, 8(1):27–62.

[2] Clapp J M, Giaccotto C. Evaluating House Price Forecasts[J]. Journal of Real Estate Research, 2002, 24(1):1–26. available at http://ideas.repec.org/a/jre/issued/v24n12002p1-26.html.

[3] Seko M. Housing prices and economic cycles[C]//The International Conference on Housing Market and the Macro Economy. Hong Kong: [s.n.] , 2003.

[4] Wedyawati W, Lu M. Mining real estate listings using ORACLE data warehousing and predictive regression[C]//. United States: Las Vegas, NV, 2004:296 –301. Predictive data mining;Regression;Real estate multiple listing service; Datamining technology

[5] Bin O. A prediction comparison of housing sales prices by parametric versus semi-parametric regressions[J]. Journal of Housing Economics, 2004, 13(1):68–84.

[6] Case B, Clapp J, Dubin R, et al. Modeling Spatial and Temporal House Price Patterns: A Comparison of Four Models[J]. The Journal of Real Estate Finance and Economics, 2004, 29(2):167–191.

[7] Anglin P. Local dynamics and contagion in real estate markets[C]//The International Conference on Real Estates and Macro Economy. Beijing: [s.n.] , 2006.

[8] Liu J G, Zhang X L, Wu W P. Application of fuzzy neural network for real estate prediction[J]. Advance in Neural Networksisnn 2006, PT 3, PROCEEDINGS, 2006, 3973:1187–1191.

[9] Zhou W X, Sornette D. Analysis of the real estate market in Las Vegas: Bubble, seasonal patterns, and prediction of the CSW indices[J]. Physica A: Statistical Mechanics and its Applications, 2008, 387(1):243 – 260. http: //dx.doi.org/10.1016/j.physa.2007.08.059. Econophysics;Real estate market;Power law

[10] Min Hwang John M. Quigley. Housing Price Dynamics in Time and Space: Predictability, Liquidity and Investor Returns[J]. The Journal of Real Estate Finance and Economics, 2010,41(1):3-23