Empirical Analysis of Housing Prices in Chinese Market

Jieqiong Wang and Dejun Xie

***Abstract*—This work constitutes, in part, our recent efforts in exploring more realistic and appropriate approaches for valuing mortgage backed securities as well as facilitating the financial management of market participants in China's housing market. Unlike structural analysis where usually only interest rate effect is considered in valuing properties and relevant securities, this paper focuses on the housing price dynamics in relation to the house price changes in time and several key contributing factors in China's real estate market. The data used for the study are the monthly housing price indices from July 2005 to December 2010, collected and maintained by the National Bureau of Statistic of China. Several autoregressive models are proposed, analyzed, and compared to best identify these relations. The study shows that house price changes in current month are influenced by the house prices in previous months, with the optimal number of price lags inferable using regression procedures. Land price and consumer price index are also contributing factors for housing price changes.An autoregressive model in the form of fractional polynomial of contributing factors is shown to have best forecasting power. Numerical experiments and simulation are carried out to test the effectiveness and robustness of our model. In particular, the accuracy and convergence of the parameter estimation are tested with different sampling processes.**

***Index Terms*—Housing price, mortgage security, autoregressive model, parameter estimation.**

I. Introduction

Housing market has grown rapidly in China since early 1990s, particularly after 1998 when the country's historical welfare housing distribution system was largely replaced by a market oriented housing system where majority population have to purchase houses from property developers (see *Notice on the Deepening of Housing Reform and Fasten Housing Construction,* Gov. [1998] No.23).According to the National Bureau of Statistics of China, the average price of commercialized real estates in China has increased by 127% from 1998 to 2009.The fast increasing house price impacts not only the development of a city and its financial credit, but also living cost of commonpeople. A sound understanding of the price dynamics of housing market is of great significance not only to market participators such as individual buyers and developers, but also to regulators and policy makers.

Price dynamics of China's housing market has been studied by considerable literatures. However, there does not seem to have a unified and generalized model for explaining the market movement in sound accuracy as calibrated by both in sample and out of sample analysis, as desired by many scholars and industry practitioners. Given the validity of such a quest from various parties and the importance of a model to adjacent fields such as commercialization and valuation of mortgage securities, it is natural to render reasonable efforts in studying the housing price mechanism from system evolution perspective.This would contrast the usual literatures with structural analysis in nature such as [6], [7] and [8] where only interest effect is considered in addressing house buyer's market strategies.A successful attempt in this respect may solve thefinancial queries from many property investors, and the applications of the method may stretch beyond pricing of commercial housing to a large spectrum of asset pricing and modeling problems.In the next section two, three main variables which may impact house prices are identified and discussed. Section three and four explain the data and methodology used for the study. Section fiveprovides numerical examples and results of our models. Concluding remarks and comments are given in section six.

Manuscript received July 16, 2012; revised August 28, 2012.

Jieqiong Wang is with School of Mathematics, the University of Edinburgh, Scotland, SC005336 UK. (e-mail: J.Wang-45@sms.ed.ac.uk).

DejunXie is with Department of Mathematical Sciences, Xi'an Jiaotong Liverpool University, Suzhou, China. (e-mail: Dejun.Xie@xjtlu.edu.cn).

II. Factor Analysis

A. House Price Lags

Under the hypothesis of efficient market, all the market information is reflected in the currently priced value. According to [9] and[10], houses are often viewed as an investment product whose price changes are influenced by the expectation of future return, especially when this expectation has a property of 'self-actualization'.Investors whose financial decisions are influenced by such a mechanism tend to regard price increase as a positive signal of the real estate market that leads to a further price increase.Similar mechanism may also persuade new buyers to join the market. This shows that the housing prices in the recent past may positively impact the housing price trend until other factors come into play and dilute such an effect. As an example, [5] uses vector autoregressive model and finds that one month delay on China Real Estate Index of Shanghai has a large impact on the index itself. Similar models can be found in

B. Land Price

For real estate industry, the relationship between house price and land price is very close because land cost is an important component of housing price and houses are the main products of land.The supply side of the market is low in elasticity because of the nonrenewable property of land, and the land price is mainly determined by its demand. In recent years, the large demand of land boosts the land prices, especially after an important notice, *Provision for the transferring state-owned land by bids, auction or listing,* was issued in 2003, where a more market oriented land allocation method was adopted. The provision enables land prices to be quoted and negotiated in a more competitive market. The increasing land price results in a high house price.

C. Inflation Represented by Consumer Price Index

Inflation constitutes another factor in house price dynamics. As shown by the empirical studied in [1] and [3], real estate may have hedging abilityon inflation. In [2], two opposing effects of inflation on housing demand were discussed. The first effect refers to that inflationreduces the size of the mortgage loan a household can obtain, thus dampens the housing demand. On the other hand, an increase of inflation rate declines the after-tax cost of house purchasing, which results in an upward move of housing demand.In the absence of supply and demand pressures, the price of goods tends to remain the same. If the only change introduced to the economy is the addition of money, then the price of goods will rise. When the influence of other factors is small, more money input and fast money circulation will lead to inflation and result in housing price increase.

III. Data

The data used in this paper are collected and maintained by the National Bureau of Statistic of China. They are available at the Bureau's official website and the National Statistic Yearbook of China. However, only monthly indices from July 2005 to December 2010 of each variable instead of the actual amount in RMB are available. This is mainly because the existing statistic regulations of real estate market in China do not require the disclosure of housing prices in further details. We use the national average figure of each variable to give a comprehensive picture of the overall market of China.

The first data set is the month-to-month average house price index of China during the period from July 2005 to December 2010. The original data is delivered in the relative values given the last month of the same year is 100 (See Fig.1). These values do not reflect the actual index movement through time. Therefore, we first recalculate the index in each month by setting the value of June 2005 being 100, and multiplying it by the house price index of July 2005 to obtain the actual index value for July 2005 (See Fig.2).

The second data set used is the month to month residential land price index of China. It should be noted here that we only analyze lands for residential construction purposes. Landsused for other categories are not considered in the residential land price index. The third data set in the paper is the monthto month consumer price index of China based on housing prices only. Existing literatures usually use consumer price index to represent inflation.In our paper, we also consider housing consumer price index as a good proxy of inflation.

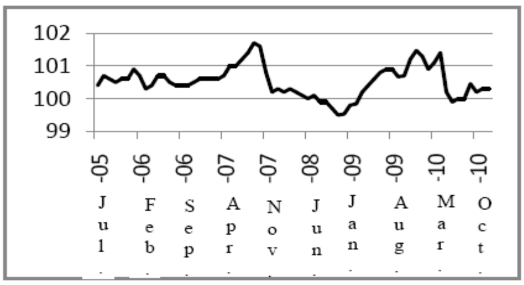

Fig. 1 Original montly data of average house price index in china

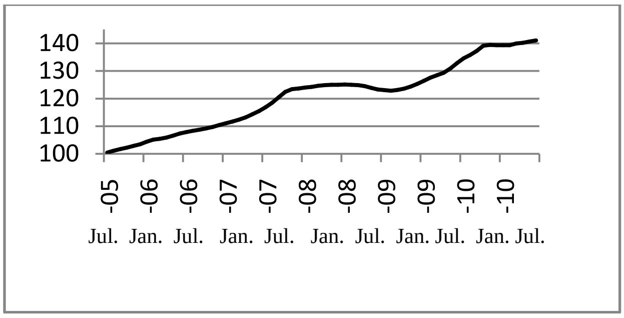

Fig. 2 Recalculated Montly Data of Average House Price Index in China

IV. Methodology

One of the simplest ways to model relations between consecutive observations is the autoregressive model, which can be found in, for instance, [4].

$$Y_t = \delta + \theta Y_{t-1} + \varepsilon_t \tag{1}$$

where Y_{t-1} is the dependent random variable in time period $t-1$, also referred as the first lagged value of $Y_t$or first lag of Y_t ; and u_t is the random error which is independently and identically distributed with mean 0 and variance σ^2.

On basis of the above simplest autoregressive model, we would like to construct models more applicable to the current problem. In particular, we apply the following p^{th} order autoregressive equation andcharacterize the relationship between proportional house price changes in period t, $\Delta Ln(hp_t)$ and its distant values, $\Delta Ln(hp_{t-j})$ as the following model, containing information of more distant past periods.

$$\Delta Ln(hp_t) = \beta_0 + \beta_1 \Delta Ln(hp_{t-1}) + \beta_2 \Delta Ln(hp_{t-2}) + \cdots + \beta_p \Delta Ln(hp_{t-p}) + \varepsilon_t \tag{2}$$

where

1) $\Delta Ln(hp_t) = Ln(hp_t) - Ln(hp_{t-1})$is the difference of logarithm house prices in period t and $t-1$, which denotes the percentage change of house price;
2) $\Delta Ln(hp_{t-j})$is the j^{th}lagged value of logarithm house price, corresponding to the value of j months ago;
3) $\varepsilon_t$is the random error.

We can use $\Delta Ln(hp_t)$ to define the proportional change of logarithm house price index, the appropriateness of which can be proved by the following simple analysis.

$$Ln(hp_t) = Ln(hp_t) - Ln(hp_{t-1})$$
$$= Ln(hp_{t-1} + \Delta hp_t) - Ln(hp_{t-1})$$
$$= Ln(1 + \frac{\Delta hp_t}{hp_{t-1}})$$

$$\lim_{\Delta hp_t \to 0} \Delta Ln(hp_t) = \lim_{\Delta hp_t \to 0} Ln(1 + \frac{\Delta hp_t}{hp_{t-1}}) = \frac{\Delta hp_t}{hp_{t-1}} \tag{3}$$

Similarly, we use $\Delta Ln(A_t)$ as the appropriate form of time series data setA_t, representing the proportional change of each variable, where A_t represents land price index, residential land price index and consumer price index.

One assumption of (2) is that the time series variable should be stationary. The stationary property ensures that the time series data are distributed in the same pattern so that models based on previous observations may be predictive for the future. In this paper, we use the Augmented Dickey-Fuller (ADF) statistic to test the stationary of data. The ADF regression can be represented by

$$\Delta A_t = \theta + \delta A_{t-1} + \gamma_1 \Delta A_{t-1} + \gamma_2 \Delta A_{t-2} + \ldots + + \gamma_p \Delta A_{t-p} + \varepsilon_t \quad (4)$$

TABLE I: HYPOTHESIS TEST OF (4)

Null hypothesis	$H_0: \delta = 0$	Nonstationary
Alternative hypothesis	$H_1: \delta < 0$	Stationary

TABLE II: CRITICAL VALUE OF ADF TEST

Significant level	10%	5%	1%
Critical value	-2.57	-2.86	-3.43

If one can reject the null hypothesis at α% significance level, as represented by the observed ADF value < Critical value, then there is a probability of $(100-\alpha)$% that the data set is stationary.

Moreover, we use the following Schwarz Information Criterion (SIC) to determine the optimal lag length in (2).

$$SIC(p) = Ln\left(\frac{SSR(p)}{T}\right) + (p+1)\frac{LnT}{T} \quad (5)$$

where $SSR(p)$ is the sum of squared residuals of (2) and T is the total number of observations (See in [6]). The model with the lowest value of SIC is preferred.

As discussed in section two, house price dynamics may be influenced by other variables such as land price and consumer price index. Therefore, if one can test and verify that there does have some relationships between them, then these other variables can be added into the regression model as predictors for house price dynamics forecasting. In this paper, we use Granger causality test to verify if land price and consumer price index contain predictive content in forecasting house price dynamics.

V. NUMERICAL RESULTS AND DATA ANALYSIS

TABLE III: STATIONARY TEST OF (2)

Null Hypothesis: DLNHP has a unit root		
Lag Length: 1 (Automatic based on SIC, MAXLAG=10)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-2.844086	0.0579
1% level	-3.538362	
5% level	2.908420	
10% level	-2.591799	
*MacKinnon (1996) one-sided p-values.		

We observe from Table III that the stationary test statistics $ADF = -2.844086$, which is smaller than the value -2.591799at 10% significant level. Therefore, time series data $\Delta Ln(hp_t)$ is stationary at 10% level. It is improbable to find a perfect model to accommodate all the information, thus to select an optimal many regressors becomes pivotally important.

Based on the regression results in Table IV, we use two selection methods to determine the optimal number of lags. First, the values of both SIC and AIC in AR (2) arethe smallest among the four models under consideration, indicating thatinclusion of two lagged values in (2) gives the most suitable regression.In addition, the'*' mark in Table IV stands for the level of significance or effectiveness of each variable in the regression equation. Variables with coefficients without '*'should not be included in (2) because their coefficients have high probability to be zero. One can see from the table that AR (2) gives better performance because all the coefficients are significant at least at 5% level. In conclusion, the preferred model can be written as

$$\Delta Ln(hp_t) = 0.0011 + 1.068758 * \Delta Ln(hp_{t-1}) - 0.269206 * \Delta Ln(hp_{t-2}) \quad (6)$$

Equation (6) indicates that the present change in house price index is dependent on house price changes of last month and the month before last. Thisis reasonable because during a short period such as two months, otherfactors relating to house price variation may stay constant. The coefficients of the two variables have opposite signs, corresponding to the market reality where permanent increase or decrease ofhousing prices is impossible. The absolute value of the positive coefficient of house price in last month is bigger than that of the negative coefficient of house price two months ago, which matches the observed increasing trend of house price in real estate market.

According to Table V, the observed F-statistic 4.18721 is greater than the critical value 3.65 at 1% significance level. Therefore, the null hypothesis that $\Delta Ln(cpi_t)$ does not Granger Cause $\Delta Ln(hp_t)$ can be rejected at 1% significance level. So the proportional change of logarithm living land price index contains predictive content to the dependent variable and can be added into the model. Similarly, from Table VI we conclude that $\Delta Ln(cpi_t)$ contains predictive content to $\Delta Ln(hp_t)$ and can be added into (2)at 10% significance level.

What is to be chosen next is which forms of living land price and consumer price index should appear in the model. The common procedure of estimating the relationship between two variables is to check if a linear relationship exists. Through regression test, we find that simply adding more lags of the original variables or the logarithm of these variables do not yield better performance because none of the coefficients are significant even at 10% level (See in Table VII). Therefore, what we try next is tochange the forms of thevariables. In particular, we try the power form of the variable living land price. Because our data set of $\Delta Ln(lp_t)$ contains negative values, square root would not be a suitable choice. When cubic root form is used, we find out that its coefficient is 10% statistically significant. Selected regression results with different model choices are listed in Table VIII.

TABLE IV: REGRESSION RESULTS OF (2)

Dependent Variable: $\boldsymbol{\Delta Ln(hp_t)}$
Method: Least Squares

	AR(1)	AR(2)	AR(3)	AR(4)
Constant	0.000883* (0.000486)	0.0011** (0.000489)	0.001* (0.000513)	0.001034* (0.000533)
$\Delta Ln(hp_{t-1})$	0.842154*** (0.068328)	1.068758*** (0.124467)	1.103694*** (0.129831)	1.127117*** (0.132017)
$\Delta Ln(hp_{t-2})$		−0.269206** (0.124288)	−0.415251** (0.186714)	−0.469356** (0.195547)
$\Delta Ln(hp_{t-3})$			0.135517 (0.129725)	0.266392 (0.195829)
$\Delta Ln(hp_{t-4})$				-0.114717 (0.131794)
R-squared	0.710155	0.730343	0.734859	0.74136
Adjusted R-squared	0.705480	0.721354	0.721145	0.722891
S.E. of regression	0.002651	0.002595	0.002608	0.002620
Sum squared resid	0.000436	0.000404	0.000394	0.000385
Durbin-Watson stat	1.544760	1.922672	1.940089	1.871604
Akaike info criterion	-8.996633	-9.023826	-8.998346	-8.972526
Schwarz criterion	-8.929167	-8.929977	-8.861112	-8.799503

Note: ***, ** and * indicate rejection of null hypothesis at 1%, 5%, and 10% significant level respectively.

TABLE V: THE GRANGER CAUSALITY TEST OF $\Delta Ln(hp_t)$ AND $\Delta Ln(lp_t)$

Pairwise Granger Causality Tests
Sample: 2005 2010
Lags: 2

Null Hypothesis:	Observation	F-Statistic	Probability
DLNLP does not Granger Cause DLNHP	63	4.18721	0.02002
DLNHP does not Granger Cause DLNLP		0.72146	0.49035

TABLE VI: THEGRANGER CAUSALITY TEST OF $\Delta Ln(hp_t)$ AND $\Delta Ln(cpi_t)$

Pairwise Granger Causality Tests
Sample: 2005 2010
Lags: 2

Null Hypothesis:	Observation	F-Statistic	Probability
DLNCPI does not Granger Cause DLNHP	63	2.18569	0.12157

Overall, we obtain the following equation.

$$\Delta Ln(hp_t) = 0.00042 + 0.992545 * \Delta Ln(hp_{t-1}) - 0.319813 * \Delta Ln(hp_{t-2}) + 0.007325 * \sqrt[3]{\Delta Ln(lp_t)} \quad (7)$$

Try a similar procedure as above, we obtain the model containing consumer price index.

$$\Delta Ln(hp_t) = 0.000588 + 1.013412 * \Delta Ln(hp_{t-1}) - 0.205312 * \Delta Ln(hp_{t-2}) + 0.002078 * \sqrt[3]{\Delta Ln(cpi_{t-1})} \quad (8)$$

The above two improved models are derived and selected according to statistical criterion. We may also try to evaluate their appropriateness in the context of market mechanism and practice. Notice that in (8), the regressors are all lagged values while in (6), the regressor about land price is the current value in period t. In order to explain this difference, we study the correlation of these three variables.

TABLE VII: SOME TRAILS OF THE REGRESSION

Variable			
Constant	0.000641 (0.001701)	-0.000810 (0.008235)	0.000982 (0.000515)
LNHP_1	1.066704*** (0.125643)	1.067200*** (0.125639)	1.064295*** (0.125042)
LNHP_2	−0.266509** (0.125616)	−0.267056** (0.125620)	−0.298880** (0.130667)
LP_1	3.19E-06 (1.13E-05)		
LNLP_1		0.000386 (0.001660)	
DLNLP_1			0.027002 (0.035450)
R-squared	0.730706	0.730589	0.732969
Adjusted R-squared	0.717014	0.716891	0.719391
S.E. of Regression	0.002615	0.002616	0.002604
Sum squared Resid	0.000404	0.000404	0.000400
Durbin-Watson Stat	1.921629	1.921671	1.918081
Akaike info Criterion	-8.993430	-8.992995	-9.001866
Schwarz Criterion	-8.857357	-8.856923	-8.865794

TABLE VIII: REGRESSION TEST OF CHANGING PATENTS OF VARIABLE

Variable			
Constant	0.001083** (0.000493)	0.000420 (0.000586)	0.000636 (0.000599)
DLNHP_1	1.074039*** (0.126073)	0.992545*** (0.127309)	1.038881*** (0.125731) (0.129934)
DLNHP_2	−0.286383** (0.132597)	−0.319813** (0.123907)	−0.322640**
DLNLP^2	0.326393 (0.831260)		
DLNLP^(1/3)		0.007325* (0.003663)	
DLNLP^(1/3)_1			0.004878 (0.003683)
R-squared	0.731046	0.747457	0.738130
Adjusted R-squared	0.717370	0.734615	0.724815
S.E. of regression	0.002614	0.002533	0.002579
Sum squared resid	0.002614	0.000378	0.000392
Durbin-Watson stat	1.933386	1.946400	1.908081
Akaike info criterion	-8.994698	-9.057648	-9.021385
Schwarz criterion	-8.858617	-8.921576	-8.921576

TABLE IX: CORRELATIONS BETWEEN VARIABLES

	$\Delta Ln(hp_t)$	$\sqrt[3]{\Delta Ln(lp_t)}$	$\sqrt[3]{\Delta Ln(cpi_{t-1})}$
$\Delta Ln(hp_t)$	1	0.623007	0.154465
$\sqrt[3]{\Delta Ln(lp_t)}$	0.623007	1	0.089178
$\sqrt[3]{\Delta Ln(cpi_{t-1})}$	0.154465	0.089178	1

From Table VIIII, we observe that the correlation between $\Delta Ln(hp_t)$ and $\sqrt[3]{\Delta Ln(lp_t)}$ is 0.623007, which is relatively high. Therefore, the lagged value of land price may be already explained by the lagged value of house price. There is no need to add these two variables $\Delta Ln(hp_{t-j})$ and $\sqrt[3]{\Delta Ln(lp_{t-j})}$, which add no further predictive content into the regressive model. The current $\sqrt[3]{\Delta Ln(lp_t)}$ is already a complete and predictive enough regressor.

VI. CONCLUSION AND DISCUSSION

This paper attempts to find the relationship between house price changes and three variables, namely house prices of previous months, residential land price and inflation represented by consumer price index in China's real estate market. Autoregressive models and modified regressive models are applied to explore quantitative relationship between those variables. A series of statistical analysis and tests are carried out to obtain the coefficients for our models and to calibrate the accuracy and effectiveness of our model. Our analysis shows that the distribution of house price changes is stationary and we may derive the future value of housing price changes by investigating its previous value. It is observed that the house price changes of current month are influenced by its value of last month and the month before last. Both land price and consumer price index contain predictive contents to house price changes and can help improve the performance of the regressive model.

One limitation of this paper is the data used in the model. The three variables are all explained in the form of index using basis 100. Therefore, the difference between the values for adjacent months is relatively small. For future research, it is desirable to use actual amount in the unit of yuan. Also, how the housing price changes may affect the buyers' propensity for selling the property or settling the mortgage loans remains an interesting and open question, which deserves further study.

ACKNOWLEDGMENT

This work is partially supported by the Natural Science Foundation of Jiangsu Province of China.

REFERENCES

[1] Y. Chu and T. F. Sing, "Inflation Hedging Characteristics ofthe Chinese Real Estate Market," *Journal of Real Estate Portfolio Management*, vol. 10, pp. 145-154, 2004.

[2] R. J. Follain, "Does Inflation Affect Real Behavior: The Case of Housing," *Southern Economic Journal*, vol. 48, no. 3, pp. 570-582, 1982.

[3] W. Wang and C. He, "Housing Price and Inflation Expectations," *Journal of Finance and Economics*, vol. 5, pp. 64-746, 2005.

[4] J. H. Stock and M. W. Watson, *Introduction to Econometrics*, 2nd ed.,Addison-Wesley Series in Economics,2006.

[5] J. H. Tu and J. Zhang, "Why Housing Prices Increase – Evidence from Shanghai Real Estate Market," *International Economy*, vol. 5, pp. 41-151, 2005.

[6] D. Xie, X. Chen, and J. Chadam, "Optimal Payment of Mortgages," *European Journal of Applied Mathematics*, vol. 18, no. 3, pp. 363-388,2007.

[7] D. Xie, "Fixed Rate Mortgage Contract: A Closed Form Approximation," *International Journal of Applied Mathematics*, vol. 1, pp.16-25, 2009.

[8] D. Xie, D. Edwards, and G. Schleiniger, "An Asymptotic Method to a Financial Optimization Problem," *Advances in Machine Learning and Data Analysis*, S. Ao, et al. eds, New York: Springer, pp. 79-94, 2009.

[9] H. Yu, "China's House Price: Affected by Economic Fundamentals or Real Estate Policy?"*Front. Econ. China*, vol. 5, no. 1, pp. 25-51, 2010.

[10] Y. Zhang, "Local Government Factor in Rea Estate Prices in Cities: Reason, Mechanism and Effectiveness," *Urban and Regional Economy*, vol.2, pp. 64-68, 2006.