Empirical Analysis of the U.S. Real Estate Market Using Nonlinear Price Functions and 2SLS Regression Analysis

D.F. Vámossy* and D. Hoff *
* Withworth University, Spokane, WA, USA
dvamossy16@my.whitworth.edu

***Abstract*—This paper provides an in depth analysis of the U.S. housing market's dynamics. Five different methods were used to describe and predict the markets behavior. The most useful of the five methods was a model based on a system of three nonlinear equations. The proposed model is only accurate when predicting one month ahead. This model combined with the Time Series prediction can offer useful insights into the residential real estate markets dynamics. An econometric model is set up, where the price of real estate market depends on several components, such as crime rate, total population, GDP growth rate, GDP deflator and other factors. This model explains how certain things influence the housing price ceteris paribus.**

I. Introduction

The housing market has a tremendous impact on the economy on many different levels. Having a home is a necessity, so everyone has a stake in the market. In addition to its important role in our life, the real estate market is the most significant contributor to the Gross Domestic Product (GDP). The combination of the real estate markets intrinsic nature and its integral role in financial prosperity show the value of predicting its future behavior. Due to its significance, the dynamics of the housing market have become one of the most studied subjects by modern economists.

The House Price (HPI) measures the price of the residential real estate market. With the exclusion of the Great Recession, the United States House Price Index has seen steady growth since the 1990s. The index experienced a 125% growth from 1991 to 2006. However, the real estate markets failure in 2008 resulted in a significant depreciation of all residential real estate. In spite of this decrease, the index has seen a trend of growth since January of 2012. Studies concerning aggregate housing prices have highlighted demographics, income trends, and government policy as the fundamental drivers of the House Price Index. However, several other factors can lead to inflation in housing prices: an increase in the price of construction materials, an increase in loan rates as well as permanent increases in land prices [7]. Due to the many determining factors, the housing market appears to be best modeled by a nonlinear function [2] [6].

Goodman and Thibodeau defined the housing markets as geographic areas where the price of housing (per unit of service) is constant and individual housing characteristics are available for purchase [4]. They argued that market segmentation should be taken into account while modeling housing prices. A number of studies dealt with the subprime mortgage crisis of August 2007 and its extreme economic implications [1]. Despite fluctuations, like the subprime mortgage crisis, Lingling [6] argued that in the long run the change of housing prices is constant, and will reach equilibrium. Hilber and Vermeulen [3] claimed that changes in income will affect housing prices more in constrained locations. James Kahn introduced a realistic model that was able to capture the medium-and low-frequency fluctuations of both price and quantity from the residential sector [5].

All of these claims and theories show the true complexity of modeling a market that has as many different determining factors as the real estate market. These complex economic relations drive the real estate markets pricing index [6]. The model Lingling, Jinzhu and Yanyan presented is the basis of our analysis and prediction of future market behavior. We found parameters for this model that best replicated the House Price Index in the United States from 1991 to 2012. Using the specified model a prediction for future market behavior was generated. Despite all of the factors known and accounted for in the dynamics of the housing market, things have been overlooked. The model is imperfect; however it captures several of the significant contributing factors.

II. Procedure

A. Background and Data

The primary data set used was supplied by the Federal Housing Finance Agency (FHFA), a government organization that provides data regarding the liquidity and pricing index of non-commercial, specifically residential real estate. The data set provided by the FHFA is a publicly available standardized electronic record. The data set includes the Land Price Index as well as the House Price Index. The FHFA tracks the price and the location of all residential real estate transactions in the United States. This data set allows us to compare our theoretical models predictions with the actual behavior of the market. This enables us to test the accuracy of the model and in turn, determine its usefulness. A secondary data set used in the econometric analysis was supplied by the World Bank Databank, which is another publicly available standardized electronic record. This data set includes time series data on crime rate, GDP growth, unemployment rate, and inflation rate among others.

B. *Linear Regression*

A linear correlation was found between the housing and land price. The linear regression determined the correlation to be 0.9878. This result justifies that the housing and land price have an endogenous interrelationship; past land price has an effect on future land price, as well as future housing price. This strong relationship creates the possibility for basing housing price on previous land and housing price.

C. *Data Selection*

Parameters were chosen in order to minimize the RMS error in the prediction of the Land and House Price Indexes. Based on the characteristics of the data set, two different periods were categorized. Between 1991 and 2005, housing and land prices showed steady growth, following expected market behavior.

However, between 2006 and 2012 the subprime mortgage crisis occurred and caused unpredictable dynamics. Despite their inherent differences, the market also needed to be modeled in a comprehensive way from 1991 to 2012. The parameters were determined for these three different periods.

D. *Curve Fitting*

In general, the price is characterized by the nonlinear inverse demand functions $P = a\,e^{-Q}$. and $P = a - b\sqrt{Q}$ where a and b are positive constants and Q is the total quantity in the market (6). Based on this concept and the claims made by Lingling, second and fourth order equations were determined to be potential fits of the Land and House Price Indexes.

E. *Second Order Fitting*

The assumption that price index and time have a quadratic correlation was made when performing this data fitting. In the second order equation price index was de fined to be only dependent on time. Despite our initial hypothesis, the second order fitting cannot imitate market conditions. Due to its nature, the second order fitting is limited at following long term market trends. The market is too complex for its full dynamics to be captured with this simplistic fitting.

F. *Fourth Order Fitting*

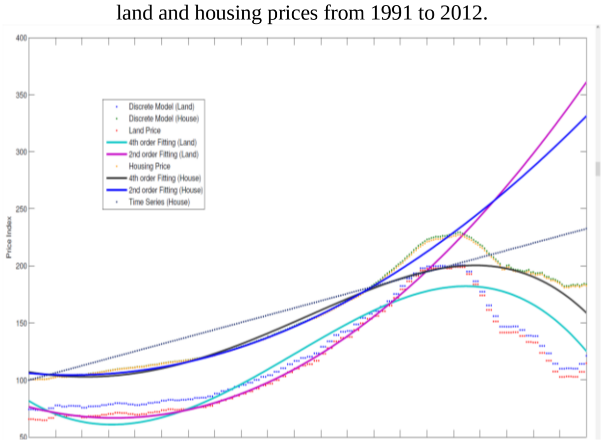
Reflecting on historical data, an assumption that time and price index may have a fourth order correlation was reached. The parameters obtained seemed to be good for short term behavior; however it was limited at predicting long term fluctuations. Due to its fourth order characteristics and nature of the economic crisis the model fits the indexes from 2006–2012 well. Despite our limited success, improvements on this model are feasible through the inclusion of additional factors.

G. *Limitations of Polynomial Fittings*

As a result of the complex nature of the market, these polynomial models could not fully capture the dynamics of the market. These fittings only took time and initial conditions into account, however time is not the only driving factor of the market's behavior. Additionally, the fittings only worked for a short amount of time, before needing to be reset. If left unaltered, both fittings show unprecedented growth in a finite time. The inability to determine when the models should be reset severely diminishes their value. These complications and shortcomings showed the inadequacies of a time based polynomial fitting.

H. *Time Series*

Figure 1. This figure shows the four different methods of modeling the land and housing prices from 1991 to 2012.



The wide usage of the time series model for prediction of future market behavior warranted its testing. Since the time series model is deseasonalized, it is superior to traditional linear predictions. This time series model cannot predict the fluctuations of the economy, only the long term trend. The model did not predict the subprime mortgage crisis, which discredits its reliability. However, it seemingly predicts long term behavior well. Its long term prediction capabilities may prove to be useful; however its inability to effectively predict crucial fluctuations limits its usefulness.

I. *Discrete Model*

Due to the shortcomings of the Time Series model and polynomial fittings of the data, a different nonlinear model was a necessity. The model we used came from Lingling [6]. Their model was based on nonlinear Cobweb Theory, which states that the price of land and housing is determined by prior period price and the relationship of supply and demand. Based on this assumption a system of discrete equations was generated to model the dynamics of the real estate market:

$$p_1(t) = p_1(t-1) + \gamma[c + c_1 \ln p_1(t-1) - a_1 p_1(t-1)], \tag{1}$$

$$p_2(t) = p_2(t-1) + \beta[d_0 + b_1 p_1(t-1) - b p_2(t-1) + d p_2^2(t-1)], \tag{2}$$

$$p_3(t+1) = \omega p_3(t) + (1-\omega) p_2(t), \tag{3}$$

where p_1 represents land price, p_2 represents housing price, and the prediction of the next month's housing price is p_3. The parameters which minimized RMS error are in Table 1. Due to its iterative nature, the accuracy of the previous month's prices is critical for predicting future prices. Because of this, we used the actual price indexes as the previous price instead of the price generated by the model. This increased the accuracy significantly. When predictions were made using prices generated by the model, the model showed strange market behavior (Fig. 2).

This shows the system's sensitivity to initial conditions. Potential implications for chaotic behavior are a bi-product of the system's dependency on initial conditions. Due to its seemingly limited scope of prediction, the model can successfully predict market behavior a month in advance. However, its limited ability to generate a reliable long term prediction diminishes its value.

J. Local Stability

The non-negative equilibrium points of the system were examined. These equilibrium points were obtained using equations (4) and (5).

$$c_1 \ln p_1 = a_1 p_1^* - c, \tag{4}$$

$$p_2^* = \frac{b \pm \sqrt{b^2 - 4d(d_0 + b_0 p_1^*)}}{2d}, \tag{5}$$

which confirms the results of Lingling [6]. This system provided the Nash Equilibrium points of (226.9976, 8.536) and (0.7944, 9.57895). The Jacobian matrix takes the form (6).

$$\begin{vmatrix} \frac{1-\gamma a_1+\gamma c_1}{p_1^*} & 0 \\ 0 & \frac{1-\beta b+2\beta d}{p_2^*} \end{vmatrix} \tag{6}$$

The eigenvalues of the Jacobian matrix of Nash equilibrium point given the initial parameters are (-0.0095, 8.5138), which is a saddle, and (-2.7175, 9.554) which is an unstable node. The saddle follows the behavior seen in Fig 2.

III. Econometric Analysis

To progress with real estate market modeling, several other factors need to be considered. A supposition that

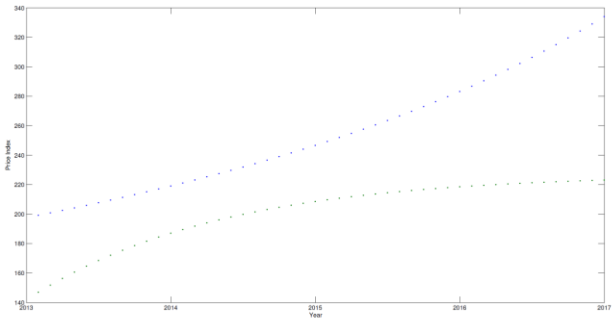


Figure 2. This figure shows the predicted behavior of the real estate market using the discrete model.

housing price may not only be dependent on land price, and land price may not be only dependent on past land price was proposed. A. Buck, T. Harrison, L. Johnson, and H. Sontag [8] conducted a research across states in the U.S. where they included thirteen variables, such as unemployment rate, population density, average credit score, median housing price in order to decide where to invest. In this section, we statistically analyze data from the United States to determine whether there is a relationship between crime rate, unemployment rate, population, population growth rate, income distribution and housing price and other factors. Econometric models can be written of the form:

A. Factors that affect the House Price Index negatively

There exists a negative relationship between crime rate and House Price Index. 67% of the changes in the House Price Index are caused by changes in crime rate, as indicated by the R^2 value. An increase of 1,000,000 crime leads to decrease of 28.4 in the House Price Index. Analysis of the regression results indicates that the slope parameter is significantly differen t from zero at the 99% level. Therefore, there is a significant relationship between crime rate and HPI (House Price Index). The 95% confidence interval around our slope parameter suggests that a

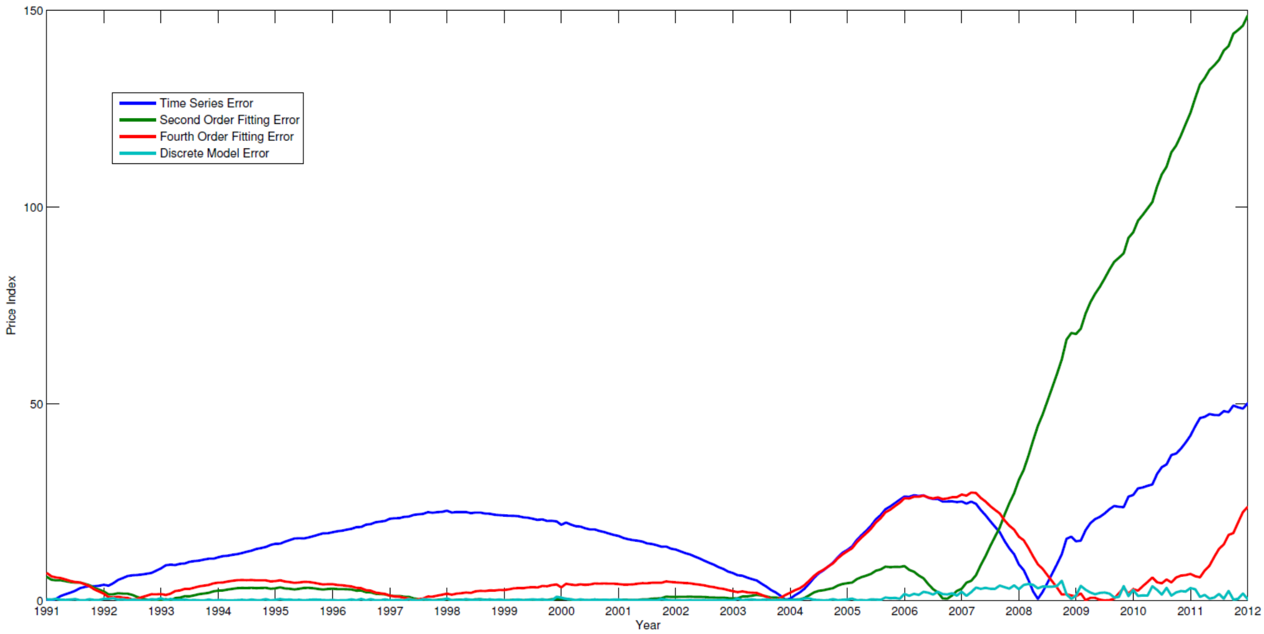


Figure 3. This figure shows RMS error of each of the different models.

1,000,000 increase in crime rate can decrease HPI anywhere from 17.5 to 39.4.

The simple linear regression shows that there exists a negative relationship between unemployment, interest rate and House Price Index. 20.1% of the changes in the House Price Index are caused by changes in unemployment rate, and 8% of the changes are due to changes in interest rate. An increase of 1% unemployment leads to a decrease of 19.09 unit in the House Price Index, whereas a 1% increase in interest rate decreases HPI by 7.51 unit. Analysis of the regression results indicates that the slope parameter of unemployment rate is significantly different from zero at the 90% level. Therefore, there is a significant relationship between unemployment and HPI. However, interest rate is statistically insignificant, and therefore there is not enough evidence that changes in interest rate affects the House Price Index. The 95% confidence interval around our slope parameter suggests that a 1% increase in unemployment can decrease HPI anywhere from -40.04 to 1.85.

B. Factors that affect the House Price Index positively

Population Density, Fertility Rate, Internet Use all have a positive effect on the House Price Index. Population Density and Internet Usage causes 88%, while Fertility Rate causes 49.5% of the changes in HPI. An increase of 1 person per square kilometer in Population Density leads to a 24.32 unit increase, a 1% increase in Internet Usage causes a 1.4844 unit increase in the House Price Index, whereas an increase of 1 children in the Fertility Rate leads to a 714.235 unit change in the HPI. Analysis of the regression results indicates that all of these slope parameters are significantly different from zero at the 99% level. The 95% confidence interval shows that a unit increase in population density can lead to a 19.53 up to a 29.11 unit, a 1% increase in Internet Usage can cause from 1.19 up to 1.77 increase in the HPI, whereas a unit increase in fertility rate can cause an increase of 317.27 up to 1111.195 in the House Price Index. Furthermore, a positive relationship between in ation rate, life expectancy and House Price Index are shown, however only 17.5% changes in the HPI are caused by in ation rate, and a whopping 88.8% by life expectancy rate. An increase of 1% in in ation rate causes a 28.43 unit increase, while a year increase in life expectancy leads to 50.82 unit increase in the House Price Index. Analysis of the regression results indicates that the slope parameter s are significantly different from zero at the 99% level. Therefore, there is a significant relationship between in ation rate, life expectancy and the House Price Index. The 95% confidence interval around our slope parameter suggests that a 1% increase in in ation can increase HPI anywhere from -5.49 to 62.36, and a year increase in life expectancy can lead from 40.55 to 61.07 unit increase in the House Price Index.

In conclusion, other variables should be explored in lieu of or in combination with crime rate, unemployment rate, in ation rate, population density, internet usage, fertility rate and life expectancy to improve the predictability of the HPI.

C. 2SLS Regression Analysis

A 2SLS Regression Analysis using GDP per capita and the total population as instrumente d variables was performed. It is shown that there exists a positive relationship between GDP per capita, interest rate and the HPI; an inverse relationship between crime rate, populatio n and House Price Index. 95.55% of the changes in HPI are caused by changes in GDP per capita, population, crime rate, and interest rate, as indicated by the R2 value.

A 1% increase in interest rate leads to a 4.38 unit increase in the HPI. Analysis of the regression results indicates that the slope parameter is significantly different from zero at the 95% level. Therefore, there is a significant relationship between interest rate and the House Price Index. The 95% confidence interval around our slope parameter implies that a 1% increase in interest rate can increase the House Price Index anywhere from 0.93 to 7.84 units.

Similarly to interest rate, GDPP (GDP per capita) has a positive relationship with the House Price Index. Analysis of the regression results shows that the slope parameter is significantly different from zero at the 99% level; thus there is a significant relationship between GDPP and HPI. The 95% confidence interval around our slope parameter suggests that a 1000 dollar increase in GDPP can increase HPI anywhere from 12.47 to 26.5 units.

Moreover, crime rate and the HPI are inversely related. An increase of 100,000 crimes committed decreases HPI by approximately 3.16 units. Analysis of the regression results shows that the slope parameter is significantly different from zero at the 99% level; thus there is a statistically significant relationship between crime rate and HPI. The result is economically significant, since increasing crime rate should decrease the House Price Index.

IV. Conclusion

Nonlinear models have been presented in order to examine the real estate markets behavior. Due to its nonlinear behavior and broad spectrum of driving factors the real estate market is very hard to predict. These models presented are by no means perfect; however, they serve a purpose for predictions of market behavior. A reliable model would prove useful tool in the prediction of the real estate markets future dynamics. To progress with real estate market modeling, several other factors need to be considered. A supposition that housing price may not only be dependent on land price, and land price may not be only dependent on past land price was proposed.

Other factors mentioned such as: crime rate, loan interest rate, population, population growth rate, literacy rate, average and medium income could have an important role in determining land and housing price. The discrete model proposed is useful for short term prediction, and the time series model may provide some insight in long term market behavior. While these models may be utilized, neither of them can accurately predict both short and long term market behavior. In spite of the progress made, there is room for further research in this complex system.

REFERENCES

[1] C.L.Foot, K. Gerardi, L. Goette, P. S. Wilson, "Just the facts: An initial analysis of subprime's role in the housing crisis", *Journal of Housing Economics*, vol. 19, 2008, pp. 291–305.

[2] J. Zhang, Q. Da, Y. Wang, "Analysis of nonlinear duopoly game with heterogenous players", *Economic Modelling*, vol. 24. (1) 2007, pp. 138 149.

[3] C.A.L. Hilber, W. Vermeulen, *Supply Constraints and House Price Dynamics: Panel Data Evidence from England*, Spatial Economics Research Centre (SERC), 2009, p. 49.

[4] A. C. Goodman, T. G. Thibodeau, "Housing Market Segmentation", *Journal of Housing Economics*, 1998, pp. 121–143.

[5] J. A. Kahn, "What Drives Housing Prices", *Federal Reserve Bank of New York Staff Reports*, 2008, pp. 1–44.

[6] M. Lingling, L. Ping, L. Yanyan, Z. Jinzhu, "Complexity of a real estate game model with a nonlinear demand function", *International Journal of Bifurcation and Chaos*, vol. 21 (11) 2010, pp. 3171–3194.

[7] J. Ma, L. Mu, "Complex Dynamics in a Nonlinear Cobweb Model for Real Estate Market", *Discrete Dynamics Nature and Society*, 2007, ID 29207, p. 15.

[8] A. Buck, T. Harrison, L. Johnson, H. Sontag, "Using Real Figures to Invest in Real Estate: A Multivariate Statistical Analysis of the US Housing Market", *University of Miami*, 2008, p. 18.

APPENDICES

TABLE I.
THIS TABLE SHOWS THE DIFFERENT VALUES OF THE PARAMETERS USED IN THE DISCRETE MODEL AND THE CORRESPONDING TIME PERIODS.

Parameter	1991-2005	2006-2012	1991-2012
γ	0.77	0.77	0.77
c	1	1	1
c_1	4	4	4
a_1	0.1	0.1	0.1
β	0.0020	0.0004	0.0010
d_0	0.800	0.870	0.037
b_1	1.004	0.886	1.012
b	1.78	6.37	2.65
d	0.022	0.011	0.037
ω	0.75	0.75	0.75

TABLE II.
THIS TABLE SHOWS THE RESULTS OF THE OLS REGRESSIONS

Parameter	Coefficient	t	R^2
α_0	506.3979	7.76	0.67
β_0	-0.0000284	-5.54	0.67
α_1	249.69	4.61	0.201
β_1	-19.09	-1.94	0.201
α_2	82.53	2.24	0.175
β_2	28.43	1.79	0.175
α_3	-592.34	-8.68	0.887
β_3	24.32	10.83	0.887
α_4	199.68	4.19	0.08
β_4	-7.51	-1.15	0.08
α_5	-3741.287	-10.16	0.88
β_5	50.815	10.56	0.88
α_6	-1304.496	-3.45	0.495
β_6	714.235	3.84	0.495
α_7	92.65	15.31	0.888
β_7	1.4844	10.93	0.888

TABLE III.
THIS TABLE SHOWS THE RESULTS OF THE 2SLS REGRESSIONS

Instrumental variables (2SLS) regression

Source	SS	df	MS
Model	35493.5431	5	7098.70862
Residual	1661.9344	15	110.795627
Total	37155.4775	20	1857.77387

Number of obs =	21
F(5, 15) =	12.58
Prob > F =	0.0001
R-squared =	0.9553
Adj R-squared =	0.9404
Root MSE =	10.526

HPI	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
GDPP	.019492	.0032912	5.92	0.000	.0124769	.0265071
PTq	-2.56e-32	1.09e-32	-2.34	0.034	-4.89e-32	-2.27e-33
Crime	-.0000316	6.57e-06	-4.80	0.000	-.0000456	-.0000175
Crimesq	2.39e-12	5.16e-13	4.64	0.000	1.29e-12	3.49e-12
Interest	4.387662	1.623543	2.70	0.016	.9271605	7.848163
_cons	-823.4214	135.5818	-6.07	0.000	-1112.407	-534.4356

Instrumented: GDPP PTq
Instruments: Crime Crimesq Interest Unemployment Inflation Life Fertilityratetotalbirthsper