Individual Housing Price Dynamics:
Evidence from the American Housing Survey

Yu Zhou, Wenping Hong
(Peking University HSBC Business School, yuzhou@phbs.pku.edu.cn)

Abstract: Unlike most previous researches using macroscopic data, this work investigates the price dynamics of individual housing. The American Housing Survey (AHS) 1985-2009 data is applied to a two-step procedure in which the AHS-recorded individual homeowner's self-reported house evaluation is proxied for that house's unobserved market price. In the first step, a hedonic is estimated in each AHS survey year to calculate the market average evaluation for each house in that year. Then the deviation of observed individual housing price from its calculated market average evaluation is established. In the second step, an auto regression and mean reversion (ARMR) dynamic panel data model is applied to the individual housing price sequence. We confirm that individual housing price has significant auto regression and mean reversion property. Further research on this topic can lead players on the real estate market to understand more thoroughly the law of individual property's pricing.

Keywords: individual housing price, auto regression and mean reversion, AHS

1. Introduction

Previous researches concerning the auto regressive and mean regression (ARMR) mechanism of housing prices mainly apply data from country or city levels with control variables on fundamental factors, for example, Abraham and Hendershott (1966), Zhou (2005), Zhang and Jia (2007), Wang et al. (2009), Qiu and Li (2011), Lu and Miao (2011), among others. As a result of limitation of microscopic data, few researches study the formation and adjustment of individual housing price dynamics using this ARMR mechanism. The macroscopic findings may not be applicable when viewed from the microscopic perspective; rational individuals maximize their benefit, which might not benefit the whole society. This paper aims to fix this gap, which can help players on the real estate market to understand more thoroughly the law of individual property's pricing.

In this paper, we define individual housing prices following the ARMR mechanism as the following: the current individual housing price growth is affected by its previous housing price growths, and when its previous housing price deviated from its long term equilibrium one, its current housing price growth will be reversely adjusted towards its long run equilibrium.

We apply the American Housing Survey (AHS) biennial survey data between 1985 and 2009. Each time, the AHS asked the sampled residents to self-report their home values. The individual housing value estimation is different from the market transaction price; it is individual's short term evaluation based on available market information. This paper takes individual resident's self-reported house evaluation as proxy for that house's unobserved market price.

We first set up the market average evaluations of a property as its fundamental prices at each AHS survey year based on the hedonic modeling approach. We then calculate deviations of observed individual housing prices from its calculated market average evaluations. In the second step, an auto regression and mean reversion (ARMR) dynamic panel data model is applied to the individual housing price sequence, trying to confirm that whether individual housing price dynamics satisfies the characteristics of auto regression and mean reversion. Results of this paper help players on the real estate market to understand more thoroughly the law of individual property's pricing.

The first part of the paper introduces the data of AHS 1985-2009, as well as variables used in this paper. The second part introduced models, including the hedonic model and the ARMR dynamic panel data model. The third and the fourth parts present the empirical results. The last part concludes.

2. Data and variables

At each survey year, the AHS visited the same sampled homes, recording information on the residents, properties and communities. Especially, the AHS asked each resident to evaluate his or her home at each survey year. The data period in this paper between 1985 and 2009 contains 13 biennial surveys in total. There are about 40000 properties in the AHS sample; however, not all of them are selected in our research. Firstly, we select only those homes that have no basic structural change since 1985, which can help ensure individual valuation change is not mainly caused by its fundamental structural alteration. We define a home without structural change as the number of bedrooms and the number of restrooms kept unchanged. Secondly, we select only owner occupied homes. From the perspective of information asymmetry, homeowners usually have more complete information on their homes compared to renters, so homeowners' housing valuation are believed to be more accurate. Lastly, in order to maintain continuity of data, we select only those properties which have records on all 13 surveys. After this process, we construct our panel data which have 624 properties in 13 survey years.

We first set up each survey year's market average valuation for each selected home using the hedonic pricing, where the independent variable (*VALUE*) is an owner's survey year home evaluation divided by that year's *CPI*. Factors on the right hand side of a hedonic are generally in the following categories: physical characteristics such as home size, facilities and etc.; community characteristics such as neighborhood environment, transportation convenience, crime rate and etc.; residents' characteristics such as gender, age, income, occupation and etc. The AHS provides varieties of information on properties' structure, community environment and residents. This paper borrows from previous literatures, coupled with the AHS data availability, and selects the following factors as independent variables in the hedonic regression. (1) The central area dummy variable (*CENTRAL*) refers to whether a home is located in the city central or in the suburbs. If it is located in urban areas, the *CENTRAL*=1, otherwise, the *CENTRAL*=0. (2) The region dummy variables (*NORTHEAST*, *MIDWEST*, *SOUTH*, and *WEST*, where *WEST* is used as the reference) refer to four US regions: Northeast, Midwest, South, and West. (3) The home age variable (*BUILT*) refers to how many years passed of a home since its date of completion. (4) The land size variable (*LOT*) refers to how many square feet of the land. (5) The construction area variable (*UNITSF*) refers to how many square feet of a home. (6) The number of bathroom variable (*BATHS*) refers to how many bathrooms. (7) The number of bedroom variable (*BEDRMS*) refers to how many of bedrooms of a home. (8) The garage dummy variable (*GARAGE*) refers to whether the home has a garage; if so, then *GARAGE*=1, otherwise *GARAGE*=0. (9) The heating facilities dummy variable (*HEAT*) refers to whether the heating equipment is installed; if so, then *HEAT*=1, otherwise *HEAT*=0. (10) The resident's age (*AGE*) refers to how old the respondent is. (11) The education variable (*EDU*) refers to how many years' education the respondent has received. (12) The marital status dummy variable (*MAR*) refers to the respondent's marital status; if the respondent is married, then *MAR*=1, otherwise, *MAR*=0. (13) The gender dummy variable (*MALE*) refers to the gender of the respondent; if the respondent is male, then *MALE*=1. (14)The Spanish origin dummy variable (*SPAN*) refers to whether the respondent is of Spanish origin; if so, then *SPAN*=1, otherwise *SPAN*=0. (15) The Caucasian ethnicity dummy variable (*WHITE*) refers to whether the respondent is white or not; if so, *WHITE*=1, otherwise *WHITE*=0. (16)The income variable (*INCOME*) refers to the annual income of all members of the surveyed household divided by that year's *CPI*.

3. The models

We use the logarithm for the dependent variable (*VALUE*), and independent variables of *LOT*, *UNITSF*, and *INCOME* in our hedonic regression. We run a separate regression for each one of the 13 survey year. An implicit assumption of separated regressions is that the unit prices of each kind of features change with time and market conditions. The hedonic price regression at each survey year is as follows:

$$\ln VALUE_i = \alpha_0 + \sum_{p=1}^{3} \alpha_k \ln x_{k,i} + \sum_{q=4}^{Q} \alpha_q x_{q,i} + v_i \tag{1}$$

The $\ln x_1$ to $\ln x_3$ are logarithmic argument for *LOT*, *UNITSF*, and *INCOME*. The x_4 to x_Q are level formats of other independent variables. The v is a random error term. The subscript i refer to the i^{th} property observation.

The hedonic model returns a market average evaluation of the i^{th} property as $\ln VALUE_i^*$, expressed as:

$$\ln VALUE_i{*} = \hat{\alpha}_0 + \sum_{p=1}^{3}\hat{\alpha}_k \ln x_{k,i} + \sum_{q=4}^{Q}\hat{\alpha}_q x_{q,i} \tag{2}$$

Where a series of $\hat{\alpha}$ are estimates of coefficients in the equation (1). The market average evaluation calculated from the equation (2) is used as benchmark for individual homeowners to adjust their previous housing price evaluation. On the other hand, people's estimation of current housing prices should be affected by historical experience. We next use the ARMR model to investigate the individual housing price sequence. Since the explanatory variables in an ARMA model include the lagged explanatory variables, displaying the dynamic effect, we create the following ARMR dynamic panel data model:

$$g_{i,t} = \alpha + \sum_{k=1}^{K}\beta_k g_{i,t-k} + \gamma(\ln VALUE_{i,t-1} - \ln VALUE^{*}_{i,t-1}) + \delta_i + \varepsilon_{i,t} \tag{3}$$

Where the individual housing price growth rate $g_{i,t}$=(ln*VALUEi,t*-ln$VALUE_{i,t-1}$). The K is the order of the lagged dependent variable, describing the autoregressive effect of individual housing price growth. The (ln$VALUE_{i,t-1}$-ln$VALUE^{*}_{i,t-1}$) measures deviation of individual housing price from its market equilibrium evaluation. The γ is expected to be negative, showing the mean reversion effect. The δ_i captures individual fixed effect because the dynamic panel model assumes individual homeowner is influenced by individual habits, preferences, and etc. The $\varepsilon_{i,t}$ is a random error term.

4. The regression result of the equation (1)

At each survey year t, we use a cross-sectional data to return a market average evaluation of the i^{th} property as ln$VALUE^{*}_{i,t}$. We apply the Wright Test for the potential heteroscedasticity problem. The test result rejects the null hypothesis, that is to say, heteroscedasticity exists in the equation (1). As a result, this paper uses the Weighted Least Squares to estimate the equation (1) and uses the Heteroskedasticity Robust Inference to calculate the estimate standard errors, so as to construct more effective t statistics. The equation (1) estimation results as presented in the table 2.

Most explanatory variables have expected effect signs; for example, the home size (*LNUNITSF*), the number of bathrooms (*BATHS*), the number of bedrooms (*BEDRMS*), the garage presence (*GARAGE*), and the heating facilities presence (*HEAT*) all have positive and statistically significant contribution to housing prices. Homeowners who are older, more educated, and/or earning more income tend to evaluate their homes higher, which is most likely due to their more wealth helping secure more expensive dwellings.

Table 1: The Equation (1) Estimation Results

Variable/Year	1985	1987	1989	1991	1993	1995
Constant	6.6937***	6.5093***	7.3793***	6.7548***	7.2080***	6.8998***
CENTRAL	-0.0728**	-0.0869**	-0.1109**	-0.1268**	-0.1182**	-0.1000**
NORTHEAST	-0.0156	0.1595***	0.1851***	0.0509***	-0.0020	-0.1278**
MIDWEST	-0.1789**	-0.1735**	-0.1686**	-0.1962**	-0.2268**	-0.2816**
SOUTH	-0.1284**	-0.1393**	-0.1771**	-0.2922**	-0.3304**	-0.3947**
BUILT	0.0001	-0.0012**	-0.0011**	0.0002	0.0007***	0.0015***
LNLOT	0.0129***	0.0053	-0.0142	-0.0157	0.0210***	0.0322***
LNUNITSF	0.2005***	0.1770***	0.1848***	0.2624***	0.2504***	0.1922***
BATHS	0.2317***	0.2057***	0.2243***	0.2381***	0.2223***	0.2160***
BEDRMS	0.0357***	0.0884***	0.0921***	0.0777***	0.0974***	0.1136***
GARAGE	0.1352***	0.0966***	0.0746***	0.0388***	0.1422***	0.1080***
HEAT	0.1071***	0.0403**	0.2178***	0.0786***	0.0814***	0.0754***
AGE	0.0034***	0.0038***	0.0039***	0.0038***	0.0029***	0.0025***
EDU	0.0068***	0.0047***	0.0062***	0.0084***	0.0087***	0.0141***
MAR	-0.0335**	-0.1087**	-0.0791**	0.0534***	-0.0126	0.0008
MALE	0.0445***	0.0996***	0.0801***	-0.0038	0.0810***	0.0403***
SPAN	0.0710***	-0.0041	-0.0197	-0.0158	-0.0191	-0.1249**
WHITE	0.1849***	0.1480***	0.1071***	0.0414**	0.0954***	0.0366**
LNINCOME	0.1693***	0.2157***	0.1396***	0.1477***	0.0650***	0.0993***
Adj R^2	0.8814	0.8440	0.8602	0.9295	0.8896	0.8071
DW	1.7991	1.8715	1.9852	1.8078	1.9089	1.8762

1997	1999	2001	2003	2005	2007	2009
7.5449***	6.8596***	6.8352***	7.4206***	6.9549***	7.7866***	7.1553***
-0.1383***	-0.0900**	-0.1151**	-0.1020**	-0.1556**	-0.1529**	-0.1052**
-0.1569***	-0.1637**	-0.2168**	-0.2970**	-0.3366**	-0.3866**	-0.2881**
-0.1869***	-0.1685**	-0.1654**	-0.2786**	-0.4077**	-0.5488**	-0.4660**
-0.3451***	-0.3688**	-0.4071**	-0.5400**	-0.5761**	-0.6654**	-0.5066**
-0.0004	0.0010***	0.0024***	0.0015***	0.0025***	0.0010***	-0.0003
-0.0069	0.0178***	0.0469***	0.0423***	0.0281***	0.0109*	0.0200***
0.1625***	0.2165***	0.2025***	0.1825***	0.1573***	0.2066***	0.2993***
0.2114***	0.1572***	0.2363***	0.1956***	0.2868***	0.2447***	0.2338***
0.1279***	0.1557***	0.1169***	0.1043***	0.0580***	0.1076***	0.0992***
0.0428***	0.0461***	0.0439***	0.0386***	0.0508***	0.0289**	0.1043***
0.0750***	0.1113***	0.0850***	0.2309***	0.0457***	0.0212***	0.0228***
0.0025***	0.0023***	0.0001	0.0001	0.0031***	-0.0004	-0.0001
0.0142***	0.0183***	0.0256***	0.0193***	0.0258***	0.0162***	0.0144***
0.0687***	0.0343***	0.0467***	0.0796***	0.0606***	0.0331***	-0.0061
0.0159	-0.0169	-0.0063	-0.0542**	-0.0772**	-0.0547**	-0.0664**
-0.2425***	-0.1041**	0.0146	-0.1189	-0.0992**	0.0008	-0.0853**
0.0570***	0.0754***	0.1015***	0.0105	0.0215***	0.0281***	0.0280***
0.0929***	0.0865***	0.0482***	0.0712***	0.1205***	0.0864***	0.0623***
0.8094	0.8318	0.8264	0.8489	0.8236	0.8907	0.7601
1.7993	1.8101	1.8941	1.8719	1.9162	1.7559	1.9661

Note: *, ** and *** indicate 90%, 95% and 99% significance levels, respectively

It should be noted of possible high collinearity between variables in a hedonic price model. How to handle collinearity depends on the research purpose. If the aim of study is to estimate the marginal contribution from each kind of housing characteristics, we then need to solve this problem because collinearity can at least result in insignificant and inaccurate estimates of regression coefficients. However, if we aim at a market average evaluation as a whole, not the individual contribution from each explanatory variable, we then can ignore the collinearity between the explanatory variables. The latter is our research purpose in the article, so we choose not to take care of the possible collinearity problem.

5. The regression result of the equation (3)

The AHS is conducted every two years, so the ln*VALUE* and ln*VALUE** are only observed every other year. In order to ensure the continuity of the data, we apply the interpolation approach assuming a uniform distribution to calculate ln*VALUE* and ln*VALUE** in AHS non-survey years. We take the ln*VALUE* as example (ln *VALUE* * uses the same procedure).

We assume: (1) the individual housing growth is uniformly distributed in a two-year interval of year t-1 and year t+1; (2) the cumulative growth during these two years for the i^{th} property is $h_{i,t+1}$; (3) the sample standard deviation of all housing growth is σ_{t+1} during these two years; (4) from year t-1 to year t, housing growth $g_{i,t}$ is uniformly distributed in $[0.5h_{i,t+1}-0.25\sigma_{t+1},\ 0.5h_{i,t+1}+0.25\sigma_{t+1}]$, then we can infer that housing growth $g_{i,t+1}$ from year t to year t+1 satisfy $(1+g_{i,t})(1+g_{i,t+1})=(1+h_{i,t+1})$ and $VALUE_{i,t}=VALUE_{i,t-1}\times(1+g_{i,t})$.

Sample statistics of housing price growth before and after the interpolation is presented in the table 2. The homogeneity of variance test and the sample mean test cannot reject the null hypothesis that the mean and standard deviation before and after the interpolation is not significantly different, respectively, showing that the interpolation well retains the distribution of the original data.

Table 2: Sample statistics of housing price growth before and after interpolation

	Mean	S.D.	Skewness	Kurtosis
Before interpolation	0.0215	0.2378	10.48	256.79
After interpolation	0.0221	0.2403	10.18	246.96
Homogeneity of variance test (*F* Test)	0.9900	P value	0.1695	
Sample mean test (*t* Test)	-0.1713	P value	0.8640	

In order to avoid spurious regression results, we apply the unit root tests to time series variables in the equation (3). The test results are shown in the table 3, which reject the null hypothesis that unit roots exist, in line with the requirements of model construction.

Table 3: Unit Root Test

Variable/Test	LLC	IPS	Fisher-ADF	Fisher-PP
$g_{i,t}$	-60.9587***	-50.7410***	6079.54 ***	5137.61***
$lnVALUE_{i,t-1}$-$lnVALUE^{*}_{i,t-1}$	-15.1111***	-21.2735***	2774.57 ***	1484.98***
Note: *, ** and *** indicate 90%, 95% and 99% significance levels, respectively				

Next, we estimate the equation (3). We use the AIC criteria to determine out K=4, the order of the lagged dependent variables on the right hand side of the equation (3). Since the explanatory variables in the model contain the lagged dependent variables, the exogenous assumption is no long satisfied and using OLS for estimation usually leads to biased and inconsistent coefficient estimates. To overcome this problem, we use the Generalized Method of Moments-Difference (GMM-DIFF) to estimate the equation (3). The key of GMM-DIFF is to find out a set of instrumental variables orthogonal to the random disturbance terms. In practices, lagged original variables are selected as the instrumental variables.

We again use the Heteroskedasticity Robust Inference to calculate the estimate standard errors, so as to construct more effective t statistics. The equation (3) estimation results as presented in the table 4. The Wald test shows coefficient estimates were jointly significant; the Hansen test rejects the null hypothesis, indicating there is no over-identification problem; the AR (2) test cannot reject the null hypothesis, indicating the residuals after the first difference have no second-order autocorrelation problem, that is, the instrumental variables satisfy moment conditions; the stationarity test reject the unit root hypothesis, indicating there is no spurious regression problem.

Table 4: The Equation (3) Estimation Results

Variable	Estimates	S.D.	t-statistics	P-value
g_{t-1}	0.5487***	0.0708	7.7500	0.0000
g_{t-2}	-0.4456***	0.0696	-6.4000	0.0000
g_{t-3}	0.3172***	0.0660	4.8000	0.0000
g_{t-4}	-0.0832**	0.0418	-1.9900	0.0460
$lnVALUE_{i,t-1}$-$ln\ VALUE^{*}_{i,t-1}$	-0.4645***	0.1066	-4.3600	0.0000
Note: *, ** and *** indicate 90%, 95% and 99% significance levels, respectively				
Tests		Test statistics	P-value	
Wald test		199.5300	0.0000	
Hansen test		115.6600	0.0740	
AR(2) test		-1.1000	0.2730	
Unit Root test: LLC		-17.2062	0.0000	
Unit Root test: IPS		-20.7726	0.0000	
Unit Root test: Fisher-ADF		2595.84	0.0000	
Unit Root test: Fisher-PP		3853.66	0.0000	

The regression results in the table 4 show that individual housing price is significantly affected by two channels. The first is the autoregressive effect. Current housing price growth is significantly affected by the growth of the previous four months, with the impact magnitudes (0.5487, -0.4456, 0.3172, -0.0832) gradually weakening. In addition, the autoregressive effect from the odd terms is positive, while the autoregressive effect from the even terms is negative. The alternate pattern reduces the fluctuation range of the current housing price growth, consistent to Li and Gao (2012) on the housing price fluctuation. The autoregressive effect reflects that people's housing evaluation behavior is influenced by historical experience. There are both positive feedback effect and negative feedback effect; the positive feedback effect further reinforces the previous evaluation, while the negative feedback effect reversely adjusts the previous evaluation, helping stabilize the housing market. Among the two adjacent opposite impacts, the positive effect is greater than the negative one in magnitude, indicating the presence of inertial effect in the housing evaluation practice.

The second effect is the mean reversion effect. The coefficient of the deviation term ($\ln VALUE_{i,t-1}-\ln VALUE^{*}_{i,t-1}$) was -0.4645, indicating if the previous housing price deviated one percent upward (downward) from its market equilibrium, the current housing price growth will decline (increase) about 0.5 percent. It shows that when individual's past housing evaluation deviated from its market average one, individual will make correction in the opposite direction towards the long term equilibrium.

6. Conclusion

This paper, based on a US micro dataset, applies the autoregressive and mean reversion (ARMR) dynamic panel model to investigate individual housing price evolution. One one hand, the empirical results show individual housing prices have significant ARMR characteristics. Current housing price growth is under the influence of previous four-period moments, with impact direction alternate and magnitude weakening. In addition, the positive impact outweighs the negative one, indicating the presence of inertial effect in housing price evolution. On the other hand, when the past housing price deviated from its long term equilibrium, the current housing price growth will be oppositely adjusted towards that market equilibrium. Further research on this topic can lead people to a more rational understanding of the real estate market.

Reference

[1] Li, Feng and Binghua Gao: The Relationship of Market Expectation and Real Estate Prices, *Price Theory and Practice*, 2012 (7):43-56.

[2] Lu, Jianxin and Jianjun Miao: The Determinants and Dynamic Properties of China's Urban Housing Prices, *Investment Research*, 2011 (6):141-152.

[3] Qiu, Chongming and Huiwen Li: Analysis on China's Real Estate Market Efficiency: Theoretical and Empirical Results, *Fujian Forum*, 2011 (4):10-14.

[4] Wang Zilong, Xiaodi Xu, Haoran Xu and Shilong Luo: Real Estate Price Volatility Model and Empirical Research, *Productivity Research*, 2009 (3):103-104.
[5] Zhang, Lin and Shenghua Jia: Analysis on the Determinants of Urban Housing Prices, *Price Theory and Practice*, 2007 (3):43-44.
[6] Zhou, Jingkui: Real Estate Price Volatility and Speculation: an Empirical Study of China's 14 cities, *Modern Economic Science*, 2005 (4):19-24.
[7] Abraham, Jesse and Patric Hendershott: Bubbles in Metropolitan Housing Markets, *Journal of Housing Research*, 1996 (7):191-207.