Implicit Hedonic Pricing Using Mortgage Payment Information

R. Kelley Pace
LREC Endowed Chair of Real Estate
Department of Finance
E.J. Ourso College of Business Administration
Louisiana State University
Baton Rouge, LA 70803-6308
OFF: (225)-578-6256, FAX: (225)-578-9065
kelley@spatial.us
and
Shuang Zhu
Assistant Professor
Department of Finance
Kansas State University
Manhattan, KS 66506
shuangzhu@ksu.edu

October 12, 2015

Abstract

Hedonic pricing, estimating the value of housing characteristics through the use of transactions data, is one of the most common tasks performed by real estate researchers. However, transaction frequency varies over time, locations, and other factors. In contrast, two-thirds of homeowners have a mortgage that they pay each month. This manuscript explores the use of mortgage payment data as an alternative or supplemental means to derive the value of housing characteristics. Insofar as price affects default and housing characteristic values determine price, housing characteristic values affect default. The implicit housing characteristic values come from parameter values that fit observed patterns of payments and defaults. We find a strong correlation $(0.74 - 0.92)$ between the explicit and implicit approaches. In addition, the Monte Carlo simulation indicates that the mortgage based pricing model has potential to help reveal housing market information especially during low sales activity and/or high default rate situations.

KEYWORDS: mortgage, mortgage default, latent index, implicit prices, hedonic house price

1 Introduction

A wide variety of fields use hedonic pricing, regression of housing prices on characteristics, to obtain information of interest. For example, a transportation expert might wish to know the effects of accessibility (or lack of it) on housing prices. An environmental economist may use house price data to estimate the loss of value associated with chemical contamination at various sites. Researchers interested in ramifications of educational quality can obtain indicators of this through house prices (after adjusting these for the mix of housing characteristics). A macroeconomist would also need to adjust for house characteristics to find a constant quality house price index (hedonic price index) over time.[1]

Historically, all of these groups relied on house prices to obtain this information. However, only a small proportion of houses sell in any particular year and the proportion selling may vary from urban to rural areas, dry up near new environmental hazards, and fluctuate over the macroeconomic cycle. This raises questions about data availability and potential sample selection bias.

On the other hand, two thirds of homeowners have mortgages and each of these homeowners makes a decision each month on whether to pay or default. All else equal, higher housing prices lead to lower default, as suggested by the option pricing theory regarding mortgages (e.g., Deng et al., 2000; Epperson et al., 1985; and Kau et al., 1995). Lower default, all else equal, indicates higher house prices. Therefore, payment or default reveals information about house prices. Essentially, two-thirds of the homeowners participate in a "Grand Survey" every month and either cast a vote of confidence (pay) or a vote of no-confidence (default) in their house price.

[1]See Malpezzi (2003) for a general review of hedonic pricing as well as Goodman (1998) for a review of the early history of hedonic pricing involving Court (1939) and others. See Harrison and Rubinfeld (1978) for a hedonic pricing study involving pollution, Bateman et al. (2001) for a discussion of traffic and house prices as well as Brasington and Haurin (2006) for the relation between educational variables and house prices.

If mortgage payment or default decisions depend upon house prices and house prices depend upon the value of housing characteristics, payment decisions will also depend upon the value of housing characteristics. Given the large quantity of mortgage observations, we can estimate the value of the housing characteristics by choosing parameters that best explain observed patterns of payments and defaults.[2] We term this approach as implicit or latent hedonic pricing because replacing price with the characteristics and associated parameters means price does not explicitly appear in the model, but is defined implicitly. To make this more concrete, suppose $\pi = F(Z, P)$ where π is the probability of payment or default for each observation, P is the price for each observation, Z are other variables, and $F(\cdot, \cdot)$ is some default function. If we know that $P = X\beta$ where X is a matrix of house characteristics and β is a vector of their values, we can substitute the price equation into the mortgage payment model to obtain $\pi = F(Z, X\beta)$ and solve for the β vector that leads to the best agreement with the data.

The purpose of this manuscript is to explore the use of mortgage data for this form of implicit hedonic pricing and compare the implicit hedonic pricing approach using mortgage data with the explicit approach based on house prices.

As a summary, we find agreement between explicit and implicit hedonic pricing. In particular, we find estimates of locational factors (zip code dummies) show close agreement with a correlation between the two approaches ranging between 0.75 to 0.93. The Monte Carlo simulation indicates that the mortgage based pricing model could be especially helpful in revealing housing market information during low sales activity and/or high default rate situations.

The manuscript provides the theory behind the implicit hedonic pricing in Section 2, presents the empirical evidence in Section 3, conducts the Monte

[2] Pace and Zhu (forthcoming) showed evidence that mortgage behavior reflects housing price information in a repeated sales setting.

Carlo simulation in Section 4, and summarizes the key findings as well as suggestions for future research in Section 5.

2 Implicit Hedonic House Price Approach

Every homeowner has a variety of variables they consider in their decision of whether to pay or not to pay their mortgage. A linear combination of these variable constitutes the latent index. We let L_i represent the latent index for the ith individual and if the latent index is positive, the individual pays their mortgage ($y_i = 1$) and if the latent index is negative, the individual defaults ($y_i = 0$) as shown in (1) and (2).

$$L_i > 0 \rightarrow y_i = 1 \tag{1}$$

$$L_i < 0 \rightarrow y_i = 0 \tag{2}$$

To give this more structure, we assume that the latent index depends on a variety of characteristics z which could include their FICO score, documentation of their assets and income, whether they have an adjustable rate mortgage (ARM) or a fixed rate mortgage (FRM), and the current balance of their loan. In addition, we assume that the latent index depends on P, is the current market value of the house, as shown in (3).

$$L_i = z_i\theta + \ln(P_i)\,\delta \tag{3}$$

Hedonic pricing models often use the semi-log specification in (4) where the log of the house price for the ith observation P_i depends linearly on the 1 by k vector of housing and neighborhood characteristics x_i with associated valuations of the characteristics in the k by 1 vector β as shown in (4).

$$\ln(P_i) = x_i\beta \tag{4}$$

Substitution of (4) into (3) yields (5).

$$L_i = z_i\theta + x_i\beta\delta = z_i\theta + x_i\gamma \tag{5}$$

$$\gamma = \beta\delta \tag{6}$$

Note, we estimate γ which is proportional to β since $\delta > 0$ (higher prices increase chance of payment). However, probit or logit are scaled to have a disturbance variance equal to 1 or $\pi^2/3$ to achieve identification which affects the magnitude of the estimated coefficients. Therefore, one can not directly compare estimates in linear regression with those from a non-linear technique like probit and logit. In fact, one cannot directly compare probit and logit coefficients directly due to the scaling. Moreover, if there is constant vector in z and in x, the parameter estimate associated with the constant term (intercept) can not be identified. Nonetheless, the coefficients from probit and logit should have the right signs and there should be a strong correlation between the estimates from the explicit hedonic pricing model and the implicit hedonic pricing model. Finally, it should be possible to set an intercept and scaling of the coefficients after estimation given some transactions data or empirical moments of house prices.

3 Empirical Evidence

This section first introduces the data sources, samples and variables in Section 3.1, then presents the empirical results in Section 3.2.

3.1 Data

We used several data sources for our analysis. The mortgage-level data comes from Blackbox Logic's BBx database. This dataset covers over ninety-five percent of the residential non-agency securitized mortgages in the United States. BBx data contains comprehensive loan origination, and monthly mortgage performance information. Our implicit hedonic price analysis requires both the mortgage-level and the property-level characteristics. However, other than property locations, few detailed property characteristics are included in the BBx data. Thus, we supplemented the BBx data with real estate public records from Clark county in Nevada. The public records data also allows us to conduct the traditional observed sales based hedonic house price model, which is then used as the benchmark for housing market information and compared with the newly proposed latent or implicit hedonic price model. The public record data contains two datasets: the property assessor data from the Clark County Assessor's Office, and the historical real estate property sales data from the Clark County Recorder's Office. The national average 30-year fixed rate mortgage interest rates come from Freddie Mac's national mortgage survey.

The main purpose of this paper is to propose an implicit, mortgage performance based hedonic price model, and investigate the feasibility of the implicit approach by comparing it with the sales based hedonic method. Two sets of samples are needed to estimate separately the implicit model and the traditional sales based explicit hedonic model. The implicit model is a hedonic price enhanced mortgage default model. Both the mortgage information, and the housing characteristics are needed for the analysis. To have the implicit sample, we merged the BBx data with the public record data by requiring the same zip codes, same interest rate types, similar original loan amounts, and did not allow much of a difference between the loan origination and the public record transaction dates. The merge procedure and criteria of these two datasets is discussed in detail in Zhu and Pace (2014).[3] The historical sales data was merged with the assessor data by a common identifier to form the sales sample for traditional hedonic price analysis.

Our analysis focuses on the residential properties in Clark County of Nevada. Both the implicit and the sales samples require observations with valid input for each variable. To rule out potential commercial properties, we limit the total living area of the property from 500 to 7000 square feet, and the original house price up to one and half million dollars. As for the implicit sample, only first lien mortgages with no piggyback loans are included. Mortgages with piggyback loans are taken out since our data does not have updated information on the second liens, while our analysis needs the updated total loan amount. To ensure our implicit sample consists only of first lien mortgages, mortgages with the original loan-to-value ratio below 0.4 are excluded as these might be home equity loans or second liens. The upper limit of the loan-to-value ratio is set to be 1.2 to rule out potential data errors. As for the sales sample, we adopted two different sales samples. The first sales sample uses the Clark county public record and the sales are identified by the arms length resale transactions, which exclude quit claim transfers, construction and time share transactions, as well as the distressed sales. However, since the composition of the underlying properties for the implicit sample and the Clark County sales sample might not be the same, we constructed the second sales sample using the new purchase loan origination information from BBx data. The summary statistics of the housing statistics from different samples are presented in Table 1. Year 2010 results are presented through the paper. Year 2008 and 2009 results are presented as additional evidence. Table 1 shows that the sales sample from Clark county public record seems to have on average a smaller living area and less variation in the housing characteristics. While the sales sample from the BBx data are

[3]The impacts with different merge criteria are investigated, and the results are not sensitive to some specific merge criteria.

quite comparable with the mortgage sample.

Variable	Mean	STD	Min	Max
Sales Sample 1				
Bath	2.32	0.68	1	10
SQFT	1764.44	756.88	500	6979
House Age	15.00	12.95	0	109
Sales Sample 2				
Bath	2.40	0.72	1	10
SQFT	1926.63	845.73	520	6918
House Age	16.44	13.56	1	107
Mortgage Sample				
Bath	2.40	0.72	1	10
SQFT	1926.72	845.77	520	6918
House Age	18.44	13.56	1	109

Table 1: Housing Characteristics Summary Statistics for Different Samples

The implicit method follows a standard logistic mortgage default model by replacing the HPI updated house price with the hedonic specification. Our investigation is in a cross-sectional setting. The dependent variable for the implicit model is a binary mortgage payment status variable, which equals one if a mortgage falls into 90+ days delinquency status in a certain year, and otherwise equals zero. Independent variables used to explain the mortgage payment performances include housing characteristics together with zip code dummies, and borrower and loan information. Housing characteristics include total living area in logged square footage (SQFT), house age, square of house age (Age-Square) to control for the nonlinearity of the house age effect, and the bathroom dummy which equals to one if the property has more than three bathrooms. The loan and borrower characteristics include logged loan amount, borrower credit score (FICO scaled by 100), full documentation status (Full Doc) which equals to one for full documentation status, exotic adjustable rate mortgage (Exotic ARM) which includes non fully amortized adjustable rate mortgages such as interest only, balloon payment and negative amortization loans, hybrid adjustable rate mortgage (Hybrid ARM) that have an initial fixed rate time period followed by later adjustable rates (no exotic features), exotic FRM dummy, regular ARM dummy, and loan age. The mortgage rate difference of the prevailing 30-year loans between the current time period and the loan origination time period (Rate Diff) is included to control for the prepayment risk. The coefficient estimates from the zip code dummies from the mortgage performance equation reflect the relative price levels across different zip code areas in a certain year.

The sales based hedonic estimation model follows the standard hedonic price model. The dependent variable is the logged transaction amount. The independent variables include the same housing characteristics as in the mortgage default model and the zip code dummies. The coefficient estimates from the hedonic regression reflect the estimate of relative price level across the different zip codes as indicated by the observed real estate transactions.

3.2 Empirical Estimates

This section presents empirical evidence to investigate whether the implicit approach is potentially feasible. The regression results of the implicit method and the sales based explicit method are reported, the correlation coefficients between the two price level estimates are calculated, and the coefficient estimates of the housing variables are compared with different methods. Because of the scaling issue for logit model as discussed in Section 2, the comparison between the explicit and the implicit methods focuses on the correlation of the relative price levels and the signs of the estimates.

Since the composition of the underlying properties for the implicit sample and the Clark County sales sample might not be the same, two different actual sales samples are investigated. The first analysis uses the Clark County public records to derive the sales based hedonic model. The second analysis uses the new purchase loan origination information from BBx data to derive the sales based hedonic model.

We conduct our analysis in a cross-sectional setting. The implicit approach is estimated using Equation (7), which is the estimation model for mortgage default where the various characteristics have values contained in the parameter vectors or scalars $\phi_1 \dots \phi_4$. There are n observations on the dependent variable y and the Housing, Borrower, and Location characteristics. The function $\Lambda(\cdot)$ represents the logistic cumulative density function. The sales based house price estimation equation appears in Equation (8), which is the standard hedonic pricing model where the various characteristics have values contained in the parameter vectors or scalars $\varphi_1 \dots \varphi_3$ and ϵ represents a n_s by 1 vector of disturbances where n_s represents the number of sold properties. Year 2010 results are presented throughout the paper. Year 2008 and 2009 results are presented as additional evidence.

$$Pr(y_i = 1) = \Lambda(\phi_1 + \text{Housing}_i \cdot \phi_2 + \text{Borrower}_i \cdot \phi_3 + \text{Location}_i \cdot \phi_4) \quad (7)$$

$$\ln(P_i) = \varphi_1 + \text{Housing}_i \cdot \varphi_2 + \text{Location}_i \cdot \varphi_3 + \epsilon_i \quad (8)$$

Table 2 to 4 present the first set of results. The Clark county public sales data is used for the sales based hedonic model for this set of results. Table 2 reports the first half of the regression results: the coefficient estimates and the statistical inferences of zip code dummies for both the implicit and the explicit hedonic models for year 2010. We require that each zip code contains at least 300 mortgage observations. The coefficient estimates for the zip code dummies reflect the relative price level across different zip codes. The results show that the relative price levels from the implicit estimates and the explicit estimates have the same signs for all zip codes, although not all the statistical significance levels matched exactly with each other.

For example, both methods turn out to have five zip codes with positive coefficients. Figure 1 plots the sales based hedonic price versus the implicit price from Table refestable1.

Next, to investigate whether the price information from the implicit method reflects housing market information, we calculated the Pearson and the Spearman correlation coefficients between the implicit and the sales based explicit prices. Table 3 reports the correlation coefficients and the corresponding p-values from year 2008 to year 2010. As an additional check for the usefulness of the implicit method, we set different thresholds of the minimum number of loans in each zip code to one hundred, two hundred and three hundred as in Panel A, B and C correspondingly. Overall, the correlation between the implicit and the sales based prices are consistently high through the sample time period and with different thresholds of the minimum number of loans included in each zip code. For example, the Pearson correlations range from 0.7472 to 0.9256, and the Spearman correlations range from 0.7792 to 0.9006. The p-values are significant and lower than 0.0001 in all the cases investigated. The results provide evidence that the implicit prices are highly correlated with the explicit prices and help reveal housing market information.

Table 2: Regression Results for Explicit and Implicit Approach Y2010 - I

	Explicit Estimate			Implicit Estimate		
Variable	Estimate	StdErr	t	Estimate	StdErr	t
zip_2	−0.2145	0.0184	−11.6854	−0.1803	0.0798	−2.2586
zip_3	−0.6210	0.0196	−31.6066	−0.5424	0.0874	−6.2068
zip_4	−0.3530	0.0144	−24.5805	−0.2491	0.0683	−3.6465
zip_5	−0.3907	0.0167	−23.3889	−0.2897	0.0767	−3.7754
zip_6	0.1380	0.0156	8.8269	0.0786	0.0745	1.0547
zip_7	−0.0180	0.0164	−1.0975	−0.0066	0.0758	−0.0871
zip_8	−0.4228	0.0169	−24.9708	−0.5902	0.0966	−6.1101
zip_9	−0.2763	0.0180	−15.3537	−0.3978	0.0967	−4.1147
zip_10	−0.4872	0.0232	−20.9847	−0.5170	0.1123	−4.6052
zip_11	−0.1259	0.0232	−5.4378	−0.2041	0.1182	−1.7265
zip_12	−0.0193	0.0169	−1.1377	−0.3532	0.0885	−3.9895
zip_13	−0.4053	0.0218	−18.5659	−0.4647	0.0933	−4.9805
zip_14	−0.6371	0.0220	−28.9502	−0.4240	0.1042	−4.0690
zip_15	−0.3585	0.0191	−18.7620	−0.3716	0.0888	−4.1868
zip_16	−0.3957	0.0153	−25.8797	−0.3614	0.0724	−4.9942
zip_17	−0.4857	0.0158	−30.7455	−0.3954	0.0717	−5.5177
zip_18	−0.0825	0.0178	−4.6341	−0.0125	0.0924	−0.1356
zip_19	−0.6546	0.0173	−37.9328	−0.5244	0.0858	−6.1139
zip_20	−0.0834	0.0158	−5.2734	−0.1492	0.0759	−1.9659
zip_21	−0.1638	0.0203	−8.0725	−0.0767	0.1089	−0.7047
zip_22	−0.3201	0.0216	−14.8439	−0.3804	0.1044	−3.6436
zip_23	−0.3215	0.0218	−14.7577	−0.2707	0.1027	−2.6356
zip_24	−0.4081	0.0168	−24.2701	−0.4190	0.0799	−5.2450
zip_25	−0.3998	0.0163	−24.5065	−0.4805	0.0860	−5.5895
zip_26	−0.0163	0.0153	−1.0710	−0.1327	0.0706	−1.8805
zip_27	−0.2134	0.0170	−12.5857	−0.3015	0.0790	−3.8173
zip_28	−0.2213	0.0157	−14.0915	−0.1483	0.0724	−2.0477
zip_29	−0.2038	0.0187	−10.8776	−0.1647	0.0835	−1.9722
zip_30	−0.1744	0.0162	−10.7775	−0.2439	0.0740	−3.2970
zip_31	0.2660	0.0204	13.0627	0.0844	0.1001	0.8428
zip_32	0.2618	0.0192	13.6212	0.1686	0.0919	1.8357
zip_33	0.1284	0.0224	5.7296	0.1188	0.1129	1.0522
zip_34	−0.1406	0.0163	−8.6047	−0.2854	0.0861	−3.3141
zip_35	−0.0797	0.0176	−4.5232	−0.1948	0.0881	−2.2117
zip_36	−0.3463	0.0182	−19.0296	−0.4453	0.0861	−5.1716
zip_37	−0.2681	0.0242	−11.0547	−0.1607	0.1107	−1.4512
zip_38	0.1387	0.0209	6.6389	0.0765	0.0961	0.7963
zip_39	−0.1385	0.0191	−7.2636	−0.2628	0.0906	−2.9011
zip_40	−0.2161	0.0256	−8.4345	−0.1962	0.1205	−1.6288
zip_41	−0.0586	0.0159	−3.6791	−0.1580	0.0766	−2.0625
zip_42	−0.1655	0.0154	−10.7202	−0.2570	0.0800	−3.2117
zip_43	−0.2254	0.0163	−13.8344	−0.1969	0.0843	−2.3354
zip_44	−0.5056	0.0187	−26.9812	−0.5189	0.0928	−5.5941
zip_45	−0.1658	0.0162	−10.2152	−0.3785	0.1034	−3.6604

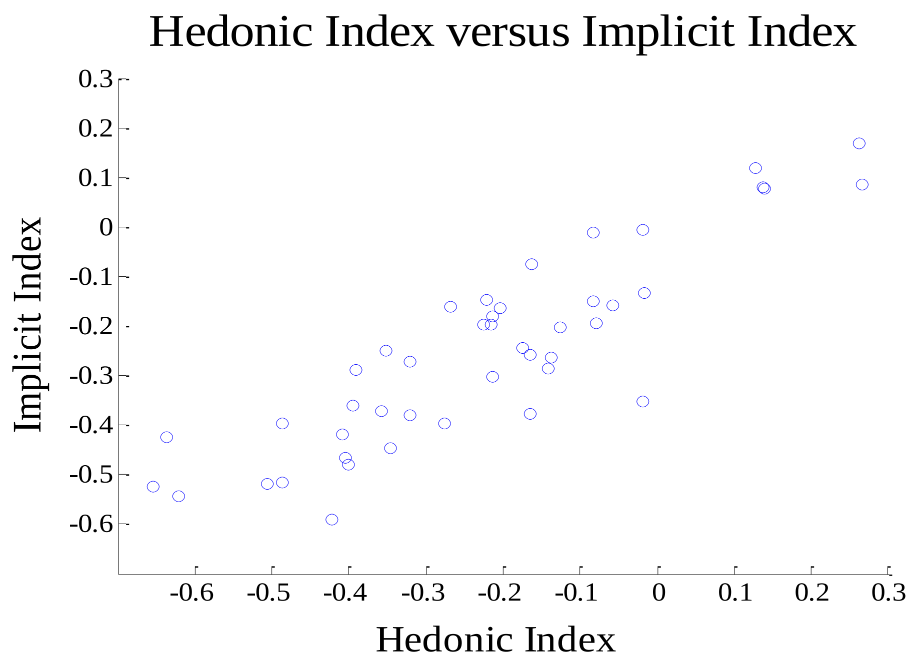

Figure 1: Implicit versus Explicit Hedonic Prices

Table 3: Correlation Coefficients between Implicit and Explicit Prices (Using Public Record Sales for Explicit Prices)

Panel A: *N=100*				
	Pearson Correlation		Spearman Correlation	
Year	Correlation Coefficient	*P-Value*	Correlation Coefficient	*P-Value*
2008	0.8540	<.0001	0.7792	<.0001
2009	0.9256	<.0001	0.9006	<.0001
2010	0.8740	<.0001	0.8784	<.0001
Panel B: *N=200*				
	Pearson Correlation		Spearman Correlation	
Year	Correlation Coefficient	*P-Value*	Correlation Coefficient	*P-Value*
2008	0.7472	<.0001	0.7806	<.0001
2009	0.8910	<.0001	0.8790	<.0001
2010	0.8974	<.0001	0.8736	<.0001
Panel C: *N=300*				
	Pearson Correlation		Spearman Correlation	
Year	Correlation Coefficient	*P-Value*	Correlation Coefficient	*P-Value*
2008	0.7496	<.0001	0.7854	<.0001
2009	0.8854	<.0001	0.8828	<.0001
2010	0.8909	<.0001	0.8763	<.0001

Table 4 Panel A reports the second half of the year 2010 regression: coefficient estimates and statistical inferences of the housing, borrower, and loan variables. Table 4 Panel B and C report the housing coefficient estimates for year 2009 and year 2008 as robustness checks. The interest here is whether the coefficient estimates on housing variables from the implicit approach carry the same signs as those from the explicit approach. In the context of mortgage default, all else equal, larger living area and higher number of bathrooms are associated with higher house valuation and lower loan-to-value ratio, which in turn leads to lower default. Both the sales based method and the implicit approach show positive valuation of the total living area which

has the highest t-value among the housing variables. The pattern is consistent for different years. The bathroom variables exhibit opposite signs for the different approaches with the implicit approach giving a positive value for more than three baths while the explicit hedonic approach has a negative value. The age of the house is negative and statistically significant for the sales based method but becomes insignificant for the implicit approach. Other borrower and loan variables have the expected signs. Borrowers with higher credit scores, full documentation status, and less exotic loans are less likely to default.

Since the implicit mortgage sample contains only properties with privately securitized mortgages, the properties underlying the implicit sample and the public sales sample may not have the same composition. Differences between the samples could affect the coefficient estimates from the implicit and the sales based hedonic approach. To address the sample issue, we conduct the second set of analysis.

This second set of results derives the sales transactions from the BBx data new purchase mortgage origination information, then compares the implicit approach with the sales based approach. The advantage of deriving sales from the BBx data is that the underlying properties should be more comparable for the implicit and the sales samples. We used the new purchases until 2010 to estimate the historical average sales based prices equation.[4]

Table 5 reports the estimates for housing characteristics for year 2010 whole sample, full documentation versus non full documentation sub samples, and fixed rate mortgage versus adjustable rate mortgage sub samples. The signs of all the housing variables except for one (House Age FRM) are the same from both sets of estimates. Table 6 reports the correlation coefficients of price levels from the explicit approach and the implicit approach. The two price estimates are significantly correlated with each other in all cases, with

[4]The BBx data has very few new purchase loans in recent years. In order to have enough sales, we need to combine all previous sales data together. In doing so, there could be a mismatch of sample periods for the implicit and the sales based samples.

Table 4: Regression Results for Explicit and Implicit Approach - II

Panel A: Y2010	Explicit Estimate			Implicit Estimate		
Variable	Estimate	StdErr	t	Estimate	StdErr	t
Intercept	3.4261	0.0489	70.0963	4.1709	0.5900	7.0688
SQFT	1.1526	0.0063	182.2916	0.6378	0.0483	13.2148
Bath	−0.0277	0.0095	−2.9146	0.2176	0.0409	5.3264
House Age	−0.0078	0.0006	−12.2694	0.0007	0.0032	0.2271
Age-Square	−0.0000	0.0000	−1.3017	−0.0000	0.0001	−0.2049
Loan Amount				−0.9936	0.0464	−21.4318
Fico				0.5472	0.0167	32.7068
Full Doc				0.1509	0.0218	6.9256
Exotic ARM				−0.4351	0.0244	−17.8227
Hybrid ARM				−0.3011	0.0339	−8.8739
Exotic FRM				−0.1486	0.0355	−4.1887
Reg ARM				−0.6303	0.0782	−8.0602
Loan Age				0.2448	0.0576	4.2476
Rate Diff				−0.0731	0.0368	−1.9866
Panel B: Y2009	**Explicit Estimate**			**Implicit Estimate**		
Variable	Estimate	StdErr	t	Estimate	StdErr	t
SQFT	1.1381	0.0057	198.6787	0.8204	0.0502	16.3271
Bath	−0.0512	0.0085	−6.0151	0.1153	0.0433	2.6609
House Age	−0.0092	0.0006	−16.3049	0.0047	0.0032	1.4829
Age-Square	−0.0000	0.0000	−2.6742	−0.0001	0.0001	−1.3225
Panel C: Y2008	**Explicit Estimate**			**Implicit Estimate**		
Variable	Estimate	StdErr	t	Estimate	StdErr	t
SQFT	1.1526	0.0063	182.2916	0.6378	0.0483	13.2148
Bath	−0.0277	0.0095	−2.9146	0.2176	0.0409	5.3264
House Age	−0.0078	0.0006	−12.2694	0.0007	0.0032	0.2271
Age-Square	−0.0000	0.0000	−1.3017	−0.0000	0.0001	−0.2049

the Pearson correlation ranging from 0.7854 to 0.8936, and the Spearman correlation ranging from 0.8065 to 0.8874.

Overall, the empirical results indicate that the mortgage performance based implicit hedonic valuation method contains information on the housing market, which could be used as an alternative or supplemental housing market valuation method.

Table 5: Regression Results for Explicit and Implicit Approach (Using BBx Sales Sample for Explicit Prices) Year 2010

Panel A: Full Sample	Explicit Estimate			Implicit Estimate		
Variable	Estimate	StdErr	t	Estimate	StdErr	t
SQFT	0.8499	0.0028	308.3894	0.6061	0.0469	12.9123
Bath	0.1390	0.0036	38.8506	0.2151	0.0399	5.3910
House Age	−0.0029	0.0003	−10.3016	−0.0002	0.0031	−0.0565
Age-Square	0.0000	0.0000	4.7013	0.0000	0.0001	0.3154

Panel B: Full Doc=0	Explicit Estimate			Implicit Estimate		
Variable	Estimate	StdErr	t	Estimate	StdErr	t
SQFT	0.8547	0.0033	258.6604	0.5342	0.0569	9.3969
Bath	0.1280	0.0042	30.6550	0.1942	0.0462	4.2067
House Age	−0.0027	0.0003	−7.9936	−0.0022	0.0037	−0.6051
Age-Square	0.0000	0.0000	3.7257	0.0001	0.0001	1.1986

Panel C: Full Doc=1	Explicit Estimate			Implicit Estimate		
Variable	Estimate	StdErr	t	Estimate	StdErr	t
SQFT	0.8310	0.0050	166.9542	0.8192	0.0862	9.5046
Bath	0.1667	0.0070	23.6652	0.3092	0.0839	3.6863
House Age	−0.0021	0.0005	−4.0829	0.0048	0.0059	0.8204
Age-Square	0.0000	0.0000	0.5151	−0.0001	0.0001	−1.1887

Panel D: FRM	Explicit Estimate			Implicit Estimate		
Variable	Estimate	StdErr	t	Estimate	StdErr	t
SQFT	0.8832	0.0052	171.4829	0.7042	0.0810	8.6932
Bath	0.1656	0.0064	25.9576	0.3060	0.0684	4.4776
House Age	0.0003	0.0005	0.5519	0.0058	0.0053	1.0861
Age-Square	−0.0000	0.0000	−1.8098	−0.0001	0.0001	−1.0591

Panel E: ARM	Explicit Estimate			Implicit Estimate		
Variable	Estimate	StdErr	t	Estimate	StdErr	t
SQFT	0.8344	0.0032	257.3242	0.5698	0.0589	9.6692
Bath	0.1203	0.0043	27.9532	0.1571	0.0501	3.1388
House Age	−0.0037	0.0003	−11.3496	−0.0027	0.0039	−0.7049
Age-Square	0.0000	0.0000	5.6853	0.0001	0.0001	0.8583

Table 6: Correlation Coefficients between Implicit and Explicit Prices (Using BBx Sales Sample for Explicit Prices) Year 2010

	Pearson Correlation		Spearman Correlation	
Sample	Correlation Coefficient	*P-Value*	Correlation Coefficient	*P-Value*
Full Sample	0.8936	<.0001	0.8874	<.0001
Full Doc = 0	0.7854	<.0001	0.8183	<.0001
Full Doc = 1	0.8414	<.0001	0.8363	<.0001
ARM = 0	0.8317	<.0001	0.8065	<.0001
ARM = 1	0.8617	<.0001	0.8864	<.0001

4 Monte Carlo Simulation

To gain some perspective on the information content of the explicit model using actual sales relative to the implicit model using the observed individual mortgage decisions (to pay or default), this section conducts a Monte Carlo simulation experiment. The experiment is intended to illustrate the relative advantages/usefulness of the different models under varying housing market conditions.

We simulated y, the observed sale price, using (9), and L, the latent index, in (10). The simulations are based on the scalar parameters β, θ, and δ as in (11), along with the n by 1 vectors of random variables x (housing characteristics), z (non-housing explanatory variables for mortgage default), u, and v, which are based on unit normals as in (12). All of the random variables are independent of each other as also shown in (12). As discussed previously, for the latent utility model, if the latent variable L for the ith individual has a positive value, this leads to the borrower paying on their mortgage (current status on mortgage), and otherwise a negative value leads to default as shown in (13).

$$y = x\beta + u \tag{9}$$

$$L = A + B + v \tag{10}$$

$$A = (x\beta)\delta,\ \ B = z\theta \tag{11}$$

$$x,z,u,v \sim N(0,\sigma^2 I_n),\ \ x \perp\!\!\!\perp z \perp\!\!\!\perp u \perp\!\!\!\perp v \tag{12}$$

$$L_i \gtrless 0 \rightarrow \text{current/default} \tag{13}$$

The independence among x and u leads to the relation of variances as in (14) and the expected R_o^2 as in (15). Equation (16) lists the usual form for the variance of the OLS estimates which depends on the actual number of sales n_s and (17) shows the formula for the $t-$statistic for the estimator of

β. We employ the $t-$statistic since this information content measure has a similar meaning across OLS and probit models. The scaling of probit affects both the magnitudes of the estimates and the associated standard errors in the same way. Therefore, the scaling does not affect the ratio of the two.

$$\sigma_y^2 = \sigma_{x\beta}^2 + \sigma_u^2 \tag{14}$$

$$R_o^2 = \sigma_{x\beta}^2/(\sigma_{x\beta}^2 + \sigma_u^2) \tag{15}$$

$$V_{\hat{\beta}} = (\hat{u}'\hat{u}/n_s)(x'x)^{-1} \tag{16}$$

$$t_{ols} = \hat{\beta}/(V_{\hat{\beta}})^{1/2} \tag{17}$$

We now define the measure of the goodness-of-fit statistic that is used to capture the importance of price factors in the latent index L. Based on their mutual independence, the variance of the latent index represents the sum of the individual sources of variation and noise as in (18). Subtracting the variance of the non-housing factors σ_B^2 from both sides of (18) leads to (19), and a partial R_m^2 that reflects the relative roles of price and noise in (20).

$$\sigma_L^2 = \sigma_A^2 + \sigma_B^2 + \sigma_v^2 \tag{18}$$

$$\sigma_L^2 - \sigma_B^2 = \sigma_A^2 + \sigma_v^2 \tag{19}$$

$$R_m^2 = \sigma_A^2/(\sigma_A^2 + \sigma_v^2) \tag{20}$$

We let the stock of all properties equal n = 100,000 and allowed the percentage of the stock that sold to take on values of 0.5, 5, and 10 percent such that a 5 percent level of sales would correspond to n_s = 5,000. A 0.5% turnover rate could exist in a stalled market, a rural area where few arms-length transactions occur, or in some countries with thin or inactive markets. Housing markets vary markedly among countries. For example, Ireland has a turnover rate currently of just under 2% while the UK has a normal turnover rate of 6% (Keenan, 2015). Ignoring new construction, the US, France, and Japan have 5.5%, 2.6%, and 0.38% turnover rates (Koo and Sasaki, 2008) respectively. Japan exhibits preference for new houses which leads to an exceptionally low turnover rate of existing houses. Alternatively, in jurisdictions in or outside the US that do not require disclosure of house transaction price, it may prove difficult to obtain many sales observations. For example, in Idaho, even the assessors do not have access to real estate transaction prices. We examined three levels of default rate of 2.5, 10, and 25 percent so that if d represents the number of defaults, $d/n = 0.025, 0.10, 0.25$. In good times, prime mortgages exhibited around a 2.5% serious delinquency rate whereas in the Great Recession serious delinquency rates rose to around 30% for subprime loans. In terms of goodness-of-fit, many empirical house price models estimated by OLS have R^2 ranging from 0.8 to 0.9. Accordingly, we let R_o^2 (approximately) equal 0.8, 0.9 in the simulation. For the mortgage data in this manuscript, the partial fit with respect to price equals around 0.05. Accordingly, we let the R_m^2 equal (approximately) to 0.10, 0.058, and to 0.038. Table 7 presents the relative $t-$statistics in estimation of the coefficients associated with x (t_{probit}/t_{ols}) for the 2.5%, 10%, and 25% default levels, given the varying proportion of the stock sold, and given the different levels of goodness-of-fit for the mortgage and the direct price models.

As expected, the sales model works well when a higher proportion of the stock sales on the market, and when the sales pricing model has a high goodness-of-fit (high R_o^2). The mortgage-based model does well as the default rate rises, and when the variations in price have higher impact on mortgage payment behavior (high R_m^2). In terms of relative performance, any value in the last three columns above 1 implies that the mortgage payments have more information content than observed prices and any value in the last three columns below 1 implies that observed prices have more information content than mortgage behavior. As extremes, if 10% percent of the housing stock sells and there is only a 2.5% default rate, the relative t statistics

Case	n_s/n	R^2_o	R^2_m	2.5% d/n	10% d/n	25% d/n
1	0.0050	0.8988	0.1005	0.0839	0.8648	1.0750
2	0.0050	0.9018	0.0588	0.0566	0.6586	0.8192
3	0.0050	0.8928	0.0378	0.0568	0.5702	0.7032
4	0.0050	0.8055	0.1004	0.1286	1.2447	1.5946
5	0.0050	0.8274	0.0570	0.0897	0.9311	1.1129
6	0.0050	0.8108	0.0385	0.0784	0.7970	0.9638
7	0.0500	0.8997	0.1008	0.0282	0.2661	0.3425
8	0.0500	0.8972	0.0589	0.0220	0.2160	0.2696
9	0.0500	0.8991	0.0366	0.0122	0.1621	0.2127
10	0.0500	0.8073	0.1016	0.0363	0.3996	0.5029
11	0.0500	0.8057	0.0572	0.0372	0.3081	0.3835
12	0.0500	0.8019	0.0390	0.0282	0.2475	0.3144
13	0.1000	0.9002	0.1010	0.0173	0.1898	0.2404
14	0.1000	0.8996	0.0584	0.0092	0.1461	0.1889
15	0.1000	0.9030	0.0370	0.0091	0.1178	0.1477
16	0.1000	0.7966	0.1022	0.0289	0.2901	0.3692
17	0.1000	0.7972	0.0591	0.0212	0.2218	0.2834
18	0.1000	0.8020	0.0371	0.0219	0.1758	0.2150

Table 7: Relative t-statistics (t_{probit}/t_{ols}) Across Proportion of Sales (n_s/n), OLS Fit (R^2_o), Latent Fit (R^2_m), and Default Rates (d/n)

are all under 0.03. In these cases, mortgage behavior would not add much information relative to prices. On the other hand, a high default rate of 25% coupled with a low sales rate of 0.5% leads to the relative $t-$statistics ranging from 0.70 to 1.60. The results indicate that mortgage model has potential to help reveal housing market information under low sales and/or high default market conditions.

5 Conclusion

Only a small proportion of houses sell in a particular year, but the vast majority of homeowners must make a decision of whether to pay or default on their mortgage each month. The requirement to make this decision (to pay or default) does not vary much over the macroeconomic cycle or across locations and therefore may have less selection bias than transactions prices. In effect, most homeowners are required to cast a vote of confidence or no confidence in their house price every month.

Since house prices and therefore payment or default decisions depend upon the value of housing characteristics, these decisions can shed light on the value of housing characteristics. The purpose of this manuscript was to explore whether using such mortgage decisions results in informative estimates of the value of the characteristics. When comparing the estimates from an explicit hedonic pricing model based on transaction data to the estimates from the implicit hedonic housing model based on mortgage payment data, we found a high correlation ($0.74 - 0.92$) between the two approaches.

The explicit and implicit approaches are not mutually exclusive. As shown in the Monte Carlo simulation, the explicit approach performs best when transaction volumes are high such as in a boom and perform worse in a bust. In contrast, the implicit approach works best when default is more prevalent (more equal distribution of zeros and ones) such as a bust or near some new negative externalities (such as a chemical spill) and worse during a boom. This suggests hybridizing these approaches. In addition, the work here suggests the potential for other implicit approaches based on hazard or multinomial models.

References

Bateman, Ian, Brett Day, Iain Lake, and Andrew Lovett (2001), *The Effect of Road Traffic on Residential Property Values: A Literature Review and Hedonic Pricing Study*, Economic and Social Research Council.

Brasington, David, Donald R. Haurin (2006), "Educational Outcomes and House Values: A Test of the value added Approach," *Journal of Regional Science*, Vol 46, pp. 245-268.

Court, Andrew T. (1939), "Hedonic Price Indexes with Automobile Examples," *Dynamics of Automobile Demand*, General Motors, New York, pp. 98-119.

Deng, Yongheng, John M. Quigley, and Robert Order (2000), "Mortgage terminations, heterogeneity and the exercise of mortgage options," *Econometrica*, Vol 68, pp. 275-307.

Epperson, James F., James B. Kau, Donald C. Keenan, and Walter J. Muller (1985), "Pricing default risk in mortgages," *Real Estate Economics*, Vol 13, pp. 261-272.

Goodman, Allen C. (1998), "Andrew Court and Invention of Hedonic Pricing Analysis," *Journal of Urban Economics*, Vol 44, pp. 291-298.

Harrison Jr, David, and Daniel L. Rubinfeld (1978), "Hedonic housing prices and the demand for clean air," *Journal of environmental economics and management*, Vol 5, No. 1, pp. 81-102.

Griliches, Zvi (1961), "Hedonic Prices for Automobiles: An Econometric Analysis of Quality Change," *The Price Statistics of the Federal Government*, Columbia University Press, New York, pp. 137-196.

Harrison Jr, David, and Daniel L. Rubinfeld (1978), "Hedonic Prices and the Demand for Clean Air," *Journal of Environmental Economics and Management*, Vol 5, pp. 81-102.

Kau, James B., Donald C. Keenan, and Taewon Kim (1994), "Default probabilities for mortgages," *Journal of Urban Economics*, Vol 35, pp. 278-296.

Keenan, Mark (2015), "Property sales are running at 50pc 'normal' rate," *Irish Independent*, July 14.

Koo, Richard and Masaya Sasaki (2008), "Obstacles to Affluence: Thoughts on Japanese Housing," *NRI Papers*, No. 137.

Malpezzi, Stephen (2003), "Hedonic Pricing Models: a Selective and Applied Review," *Section in Housing Economics and Public Policy: Essays in Honor of Duncan Maclennan*, eds T. O'Sullivan and K. Gibb, Blackwell Science Ltd, Oxford, UK.

Pace, R. Kelley and Shuang Zhu (forthcoming), "Inferring Price Information from Mortgage Payment Behavior: A Latent Index Approach," *Journal of Real Estate Finance and Economics*.

Zhu, Shuang and R. Kelley Pace (2014), "Modeling Spatially Interdependent Mortgage Decisions," *Journal of Real Estate Finance and Economics*, Vol 49 No.4, pp. 598-620.