INTERNATIONAL REAL ESTATE REVIEW

2019 Vol. 22 No. 3: pp. 333 – 357

An Adorable Housing Paper: The Informational Content of Agent Remarks

Sean Brunson
Department of Finance UNC-Charlotte. Email: sbrunso2@uncc.edu

Richard J. Buttimer Jr.
Department of Finance UNC-Charlotte. Email: buttimer@uncc.edu

Steve Swidler
Department of Economics Lafayette College. Email: swidlers@lafayette.edu

This paper considers the information content of Multiple Listing Service (MLS) descriptions and employs a significantly larger data set than previous studies. The analysis first catalogs the most frequently used terms by real estate agents in MLS descriptions. Using hedonic modeling, we estimate the effect of this qualitative information on transaction price and days on the market. Finally, we extend earlier empirical work by utilizing our larger MLS data set to forecast the probability that a house will sell after it is listed. This last contribution further sheds light on the role of qualitative information to infer property condition or circumstances that surround the sale of the property.

Keywords

Textual Analysis, Agent Remarks, MLS, Hedonic Pricing

1. Introduction

Hedonic modeling of house prices identifies the attributes of a property that contribute to its value. The attributes may be internal to the home such as square footage, and number of bedrooms and bathrooms, or external to the property, including neighborhood environment, school district and access to public transportation. One source of property information is the Multiple Listing Service (MLS) which contains quantitative data for both internal and environmental services that affect housing prices. Additionally, MLS entries include descriptive information that provides the qualitative attributes of the property and its surroundings.

MLS property descriptions tend to be relatively short (250 words or less) and frequently informal. They may contain abbreviations (e.g., “SS appl” for stainless steel appliances), symbols (&, #) and add exclamation marks to generate excitement!!! Given the brevity and informal writing style of MLS descriptions, a full-blown textual analysis that examines readability or sentiment or structure is likely excessive and misses the purpose of the text. Instead, agents wish to supplement the quantitative statistics with qualitative information that paints a more complete picture of the property itself.

This paper builds on earlier research that considers the information content of MLS descriptions and adds to the discussion in several ways. First, it documents the frequency of descriptive words and phrases by using a significantly larger data set than previous studies. A challenge with doing textual analysis is that some of the words and phrases may correlate strongly with physical or otherwise objective attributes of the property. The resulting collinearity, while not an issue for the predictive results or the overall economic analysis, can cause individual parameter estimates to be relatively unstable. However, by using a very large data set of more than 700,000 observations, the collinearity effects will be greatly mitigated. Second, the analysis examines the context of word appearance in MLS descriptions. Third, following the previous literature, we use a hedonic model to estimate the effect of qualitative information on transaction price and days on the market. Finally, we extend the empirical work on qualitative information by exploiting our larger MLS data set to forecast the probability that a house will sell after it is listed. This last contribution also sheds light on the use of qualitative information to infer property condition or circumstances that surround the sale of the property.

Rosen (1974) first discusses the theoretical foundations of hedonic modeling. Subsequently, researchers have used hedonic models to quantify the impact that certain attributes have on the transaction price of a home. In addition to area, structure age, number of rooms and location, empirical work shows that attributes such as air quality (Smith and Huang, 1995), school quality (Black, 1999; Figlio and Lucas, 2004), and nearby foreclosures (Campbell, Giglio, and

Pathak, 2011; Anenberg and Kung, 2014) all have a significant impact on the market value of a home.

An extension of this research is the work that examines agent remarks and the effect that qualitative attributes have on price in the hedonic model. One of the earliest studies is Haag, Rutherford, and Thomson (2000) who classify words into two categories: factually verifiable and opinion. In the analysis, they identify certain words that lead to lower transaction prices. Goodwin, Waller, and Weeks (2014) also find that MLS descriptions have a significant impact on market transactions. Specifically, they examine 16,373 sold and unsold MLS properties and find that positive opinions increase the price and time-on-the-market of a home as well as increase the probability of selling the home. After correcting for self-selection, signal variables (bring offer, motivated, price reduced and vacant) tend to increase the sale price and also days on the market. In a follow-up paper, Goodwin, Waller, and Weeks (2018) analyze text to determine the favorability of descriptive real estate terms. Finally, Knight (2002) exploits the comments section to identify “motivated” sellers.

To properly measure the effect of qualitative information, it is important to recognize the trade-off between price and time on the market. Sellers may be willing to wait longer and sell at a higher price or sell quickly by accepting a lower price. Thus, to accurately measure the influence of qualitative information on the market transaction, the econometric modeling must address the simultaneous endogeneity issue.

Benefield, Cain, and Johnson (2014) examine nearly 200 studies that use a hedonic model with time-on-the-market as the independent variable. Out of those articles, the time-on-market variable is negative in 100, statistically insignificant in 73 and positive in the remainder. Similarly, the authors find mixed results in over 200 studies that estimate time-on-the-market with sale price as an independent variable. The authors attribute these ambiguities to the modelling choice taken by the different studies as well as the different definition used for the time-on-the-market variable. They note that some researchers use a hazard model while others use a two-stage least squares (2SLS) approach.

In the following analysis, we further consider the effect of qualitative information in the hedonic model and examine a data set that is more than a magnitude greater than the one in Goodwin, Waller, and Weeks (2014). We adjust for endogeneity by estimating a 2SLS model and acknowledge up front that previous work has failed to isolate the exact relation between price and days on the market. Nevertheless, our focus is the effect of qualitative information in the hedonic model including how certain descriptors impact the likelihood that a property will sell.

2. Data

The paper examines MLS data provided by the Charlotte Regional Realtor Association in the USA. The listings cover the period of 2001 to 2018 and include properties located in eight counties in the Charlotte area (state of North Carolina), six in the state of North Carolina and two in the state of South Carolina. During this period, there were a total of 711,188 listings. Of these properties, 426,816 sold, which is slightly more than 60 percent of the total listings.

Table 1 summarizes several important quantitative or objective characteristics of both all listings and sold properties. The differences in means between the two groups are almost always statistically significant which is due to the large samples. Practically speaking, the differences in the typical house from each category are virtually the same. The median age is 12 years, with 2 full baths, 1 half bath and 3 bedrooms. All listings are slightly larger than sold houses (median: 2087 square feet (SF) vs. 2050 SF), have similar lot sizes (median: 0.31 vs. 0.30 acres) and marginally higher listing prices (median: US$199,900 vs. US$195,000). One noticeable difference is that sold homes are more frequently of new construction compared to all listings (17.61% vs. 15.34%). If we compare the sold to unsold homes, properties that closed tend to be smaller, less expensive and newer than those that did not sell.

While Table 1 summarizes the quantitative attributes of Charlotte homes, we next consider the descriptive text in the MLS listing. Rather than impose arbitrary rules, we let the data do the talking by parsing the data and performing a simple frequency analysis. From the MLS “Agent Comments” field, we extract the 500 most common words in the sample data. Removing nouns and words that describe areas in the house (kitchen, bedroom, bathroom, etc.) along with articles, pronouns and auxiliary verbs (a, an, the, we, they, would, could, and should), left 44 adjectives. The list appears in Table 2.

Two of the three most frequently observed words in Table 2 concern size. “Large” appears in 35.25% of all the agent remarks, whereas “spacious” occurs in 15.00% of the MLS descriptions. The second most popular word, “beautiful”, appears in 20.59% of the comments.

The collection of words includes terms that express emotions, tones or feelings such as “fantastic”, “incredible”, and “awesome”. Some words may serve as an attribute or condition of the property. Examples include a “green” home which means one with sustainable features or the property may be described as “immaculate”, “clean”, or “refinished”. Still other terms can have very different meanings depending on the context. A home can be a “short” distance from shopping and schools or the MLS description may inform potential buyers that this is a “short” sale where the listing price is below the outstanding loan amount on the property.

Table 1 Summary Statistics

Variable	All Listings Mean	All Listings Median	All Listings Min	All Listings Max	Sold Houses Mean	Sold Houses Median	Sold Houses Min	Sold Houses Max	Difference in Means
Total Listings	711,188								
Age	20.37	12	0	218	19.89	12	0	210	0.4758***
Baths Full	2.27	2	0	10	2.24	2	0	10	0.0251***
Baths Half	0.56	1	0	10	0.55	1	0	10	0.0124***
Beds Total	3.54	3	0	10	3.53	3	0	10	0.0172***
Closed	60.01%								
Distressed Listing	6.11%								
Exterior Construction: Brick	44.38%				43.66%				0.0071***
Exterior Construction: Siding	10.15%				10.31%				-0.0017***
Flooring: Carpet	76.97%				76.72%				0.0026***
Flooring: Tile	6.93%				6.92%				1.00E-04
Flooring: Wood	3.68%				3.57%				0.0011***
Green Certification	7.00%				7.54%				-0.0054***
Heating: Central	69.47%				70.10%				-0.0062***
Heating: Furnace	1.34%				1.43%				-9e-04***
Heating: Pump	20.34%				19.60%				0.0075***
Heating: Window	1.24%				1.03%				0.0021***
List Price	$276,556.86	$199,900.00	$775.00	$12,000,000.00	$254,460.31	$195,000.00	$2,500.00	$6,500,000.00	22096.5514***

(*Continued…*)

(Table 1 Continued)

	All Listings				Sold Houses				Difference in Means
Variable	Mean	Median	Min	Max	Mean	Median	Min	Max	
Lot Size Area (Acres)	0.69	0.31	0.01	200	0.58	0.3	0.01	200	0.1074***
New Construction	15.34%				17.61%				-0.0227***
Parking: Carport	5.14%				5.10%				4.00E-04
Parking: Garage	73.26%				74.29%				-0.0103***
Septic	16.76%				15.15%				0.0161***
Square Feet Total	2,333.67	2,087	104	19,939	2,266.76	2,050	104	19,842	66.9088***
Square Feet Total (Unheated)	300.69	204	0	9,000	290.36	210	0	7,900	10.3338***
Total Sold					426,816				
Close Price					$247,055.63	$190,000.00	$1,000.00	$6,320,000.00	
Distressed Sale					7.35%				
DOM					119.63	90	1	1,494	

Also found in Table 2 are words that frequently connote circumstances that detract from the value of a property. Specifically, homes that are a good “investment”, have “reduced” prices, or “motivated” sellers often elicit offers that result in lower transaction prices. If that is the case, one question of interest is why do realtors use these terms in their remarks?

Table 2 Relevant Word Appearances

Word	Rank	Total MLS Appearances	Percentage MLS Appearances
large	3	250,665	35.25%
beautiful	7	146,451	20.59%
spacious	22	106,673	15.00%
wooded	62	53,661	7.55%
gorgeous	74	46,644	6.56%
convenient	102	34,860	4.90%
gourmet	104	34,189	4.81%
quiet	105	34,182	4.81%
lovely	114	32,042	4.51%
charming	149	25,081	3.53%
beautifully	154	24,559	3.45%
immaculate	159	24,002	3.37%
stunning	172	21,963	3.09%
country	174	21,720	3.05%
fabulous	193	19,419	2.73%
excellent	208	17,801	2.50%
buyers	210	17,666	2.48%
short	226	15,934	2.24%
fantastic	227	15,932	2.24%
popular	246	14,367	2.02%
investment	249	14,163	1.99%
reduced	253	14,002	1.97%
luxury	266	13,410	1.89%
motivated	269	13,318	1.87%
incredible	271	13,201	1.86%
refinished	279	12,727	1.79%
awesome	290	12,039	1.69%
established	294	11,823	1.66%
small	338	9,510	1.34%
luxurious	339	9,500	1.34%
clean	344	9,378	1.32%
welcome	350	9,119	1.28%
starter	357	8,807	1.24%

(***Continued…***)

(Table 2 Continued)

Word	Rank	Total MLS Appearances	Percentage MLS Appearances
inviting	362	8,547	1.20%
adorable	368	8,399	1.18%
conveniently	372	8,306	1.17%
soaring	379	8,027	1.13%
relaxing	386	7,875	1.11%
nicely	416	6,798	0.96%
historic	434	6,397	0.90%
ideal	439	6,350	0.89%
green	456	6,064	0.85%
rental	462	5,954	0.84%
tenant	487	5,591	0.79%

Note: From the MLS "Agent Comments" field, we extract the 500 most common words in the sample data. Removing nouns and words that describe areas in the house (kitchen, bedroom, bathroom, etc.) along with articles, pronouns and auxiliary verbs (a, an, the, we, they, would, could, and should), left 44 adjectives.

Our discussion suggests that context determines whether a word describes a qualitative feature that adds or detracts from the value of the property. To that end, further insights may be drawn by examining word combinations that appear within a given MLS description. Table 3 lists pairwise frequencies of the more salient words found in Table 2.[1] The entries represent conditional probabilities and equal the number of MLS descriptions where both words appear as a percentage of the number of listings of the less frequently found word. Thus, for example, in 37.22% of the MLS descriptions that contain “awesome”, the word “large” also appears.

Notable in Table 3 is the relatively high conditional frequencies for pairs that contain the word “large”. In the total sample, “large” occurs in 35.25% of all listings. However, in all but one of the cases in Table 3, the conditional frequencies exceed 35%. In nearly 40% of all “adorable” homes, “large” appears in the listing. The highest conditional frequency in the table is for the pair “large” and “spacious” at 48.31%. Both words denote size and further augment any quantitative dimensions included in the MLS.

Table 4 shows that “large” and “spacious” most frequently precede “master” and “kitchen”. While the MLS includes the area of a home (in square feet) and number of rooms, only the average room size can be inferred. By using “large” and “spacious” to describe a master suite and kitchen, the remarks of the listing signal to the buyer a presumably desirable characteristic that concerns two important areas in the house. Other common words that frequently follow

[1] A complete list of the 946 word combinations may be obtained from the authors.

Table 3 **Conditional Pairwise Frequencies**

word	adorable	awesome	gorgeous	historic	investment	large	luxurious	motivated	reduced	spacious
adorable										
awesome	0.0177									
gorgeous	0.0566	0.1010								
historic	0.0255	0.0122	0.0746							
investment	0.0218	0.0118	0.0077	0.0286						
large	0.3958	0.3722	0.4019	0.3475	0.2091					
luxurious	0.0018	0.0241	0.1681	0.0081	0.0009	0.4043				
motivated	0.0187	0.0189	0.0481	0.0150	0.0408	0.3584	0.0131			
reduced	0.0150	0.0190	0.0650	0.0208	0.0256	0.3562	0.0189	0.1142		
spacious	0.1680	0.1650	0.2062	0.1227	0.0563	0.4831	0.2886	0.1398	0.1400	

Note: The entries represent conditional probabilities and equal the number of MLS descriptions where both words appear as a percentage of the number of listings of the less frequently found word.

"large" and "spacious" include "bedrooms", "great" and "open". In general, Table 4 suggests that both words mainly modify the size of a room or area of the house or property.

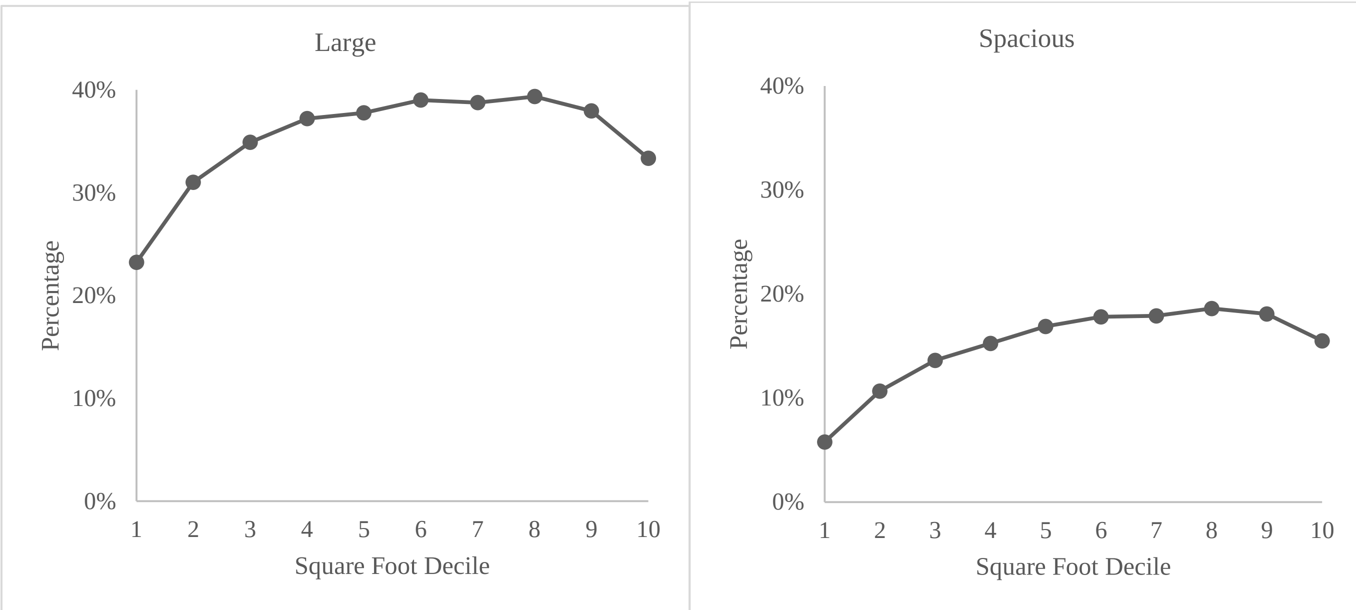
Table 4 Words After "Large" and "Spacious"

After Large	N	Percentage	After Spacious	N	Percentage
master	28,888	11.52%	**kitchen**	12,207	11.44%
kitchen	19,548	7.80%	**master**	10,913	10.23%
deck	17,335	6.92%	**bedrooms**	7,129	6.68%
walk	13,688	5.46%	home	5,307	4.98%
bonus	12,863	5.13%	rooms	4,239	3.97%
lot	12,325	4.92%	**great**	4,202	3.94%
bedrooms	10,649	4.25%	**open**	3,985	3.74%
fenced	9,921	3.96%	living	3,272	3.07%
great	8,367	3.34%	secondary	2,585	2.42%
open	7,173	2.86%	family	2,412	2.26%

To provide a better idea of qualitative vs. quantitative measures of size, Figure 1 segments homes by square foot decile and graphs the percentage of homes within the decile that uses the word "large" or "spacious" in the remarks of the agent. Among the smallest homes in the sample, 23.23% use the word "large", and the percentage increases and reaches a peak of 39.35% in the 80th percentile. After that, the homes with the largest area use "large" in the MLS description 33.34% of the time. A similar pattern is found for "spacious" although at lower levels. The use of the word "spacious" again peaks at the 80th percentile, but only 18.60% of the remarks use this word. In comparison, the word "adorable" is most used for smaller homes. Slightly less than 5% of the smallest decile use adorable and that tapers off to near 0 for the largest homes in the sample. Finally, the use of "beautiful" increases with home size and the MLS descriptions of the biggest homes include this word more than 25% of the time.

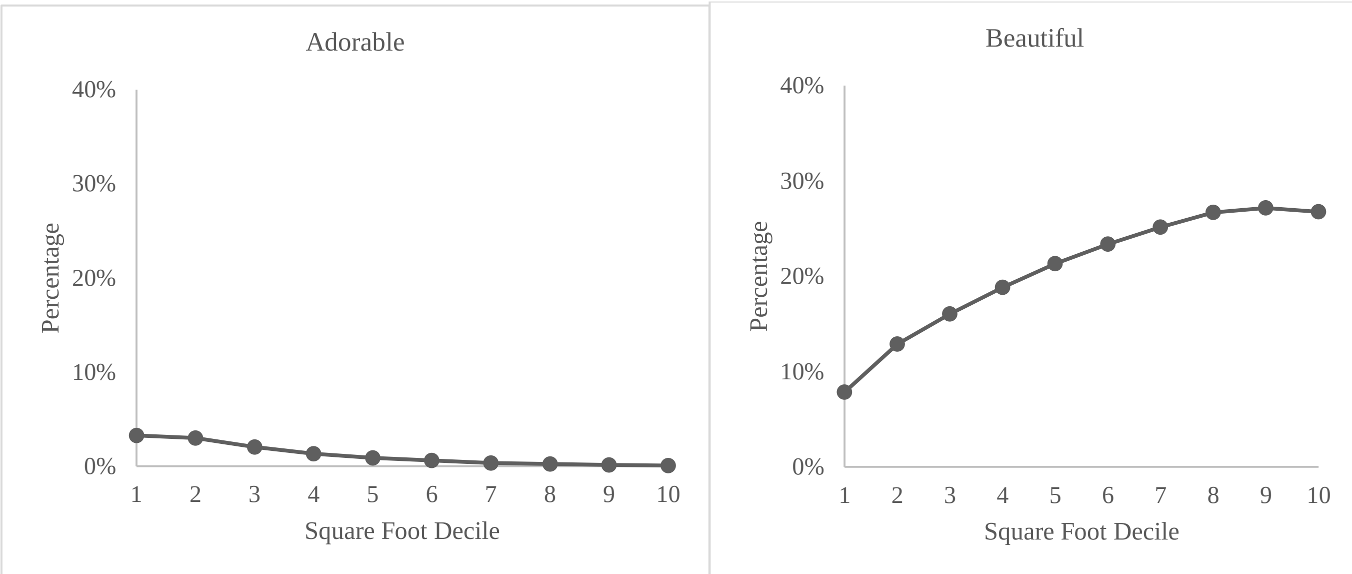
In the next section, we describe the methodology that forms the basis of the hedonic model. The analysis augments the standard hedonics by expanding the attributes to include qualitative descriptions that enhance the information from the standard quantitative measures found in MLS listings or county records. By including descriptive terms, listing agents hope to attract potential buyers who are searching for key features in a home.

Figure 1 Word Appearance by Square Foot Decile



(Continued…)

(Figure 1 Continued)



3. Methodology to Estimate Home Price and Days on the Market

To examine how qualitative information in the MLS data can influence the market transaction of a home, we employ a standard hedonic model. However, a simultaneity issue arises between the sales price and days on the market. This creates biased estimates in our specification, so we use an instrumental variable approach to address this endogeneity issue. In this section, we lay out the main methodology and the instruments used in our analysis.

We closely follow the literature and estimate the dependent variable, the log sales price of transaction *i* at time *t*, as:

$$\log(P_{i,t}) = \alpha_1 + \delta_1 \log(DOM_{i,t}) + \beta_1 \boldsymbol{X}_{i,t} + \lambda_1 \boldsymbol{Words}_{i,t} + QY_t + L_C + \epsilon_{i,t} \tag{1}$$

where $log(DOM_{i,t})$ is the log days on the market, $\boldsymbol{X}_{i,t}$ is a matrix of the housing characteristics, $\boldsymbol{Words}_{i,t}$ is an array of dummies that equal 1 if a specific word is included in the MLS description and 0 otherwise, QY_t is the quarter-year fixed effects, L_C is the county fixed effects, and $\epsilon_{i,t}$ is the error term.

Similarly, the model for the log days on the market takes the form:

$$\log(DOM_{i,t}) = \alpha_2 + \delta_2 \log(P_{i,t}) + \beta_2 \boldsymbol{X}_{i,t} + \lambda_2 \boldsymbol{Words}_{i,t} + QY_t + L_C + \xi_{i,t} \tag{2}$$

From the standpoint of measuring the effect of qualitative information on market transactions, our interest is in the estimated coefficients in vectors λ_1 **and** λ_2. However, in both specifications, there exists an endogeneity issue between the sales price and the days on the market.

To correct for endogeneity, we use a 2SLS approach. Within the literature, there is no consensus as to which instruments are best to use when controlling for endogeneity between sales price and time on the market. Indeed, as previously noted, Benefield, Cain, and Johnson (2014), in an exhaustive survey of the price and days on the market literature, find wide variation in methodologies, instruments, and results. For our purposes, we need to identify two instruments that are available to us in our data set: one that drives the days on the market but not the sales price, and one that drives the sales price but not the days on the market. From the MLS descriptions, we find two instruments that satisfy this requirement: “country” and whether a home is certified as “green”.[2]

[2] We believe that this approach is robust enough to address our main concern of controlling for any endogeneity, but also note that instrument selection is certainly an area ripe for further research.

If the word "country" is included in the description, then the house may be *perceived* to be in a more rural area or in the style of homes frequently found out in the country. Either way, this will attract a subset of potential buyers interested in a country home. Since there are fewer potential buyers, the number of days on the market should increase. However, including the word "country" in the description may not necessarily have any direct effect on the ultimate sales price. The remaining bidders self-select and are the ones who are interested in purchasing a country home.

Similarly, it is reasonable to assume that if a home is certified "green", it is more energy efficient and has lower utility costs. Given green certification, sellers will ask for a higher price and buyers will be willing to pay a premium. At the same time, we should not expect any real effect of being certified green on the number of days that a home is on the market.

With these two instruments, we estimate price and days on the market in a two-stage process. Instead of Price equation (1), we first estimate the days on the market (*DOM*) variable on the right-hand side as:

$$\log\left(DOM_{i,t}\right) = \pi_1 + \theta_1 country_{i,t} + \Lambda_1 \boldsymbol{X}_{i,t} + \Gamma_1 \boldsymbol{Words}_{i,t} + QY_t + L_C + \eta_{i,t} \tag{3}$$

where $country_{i,t}$ is an indicator variable equal to 1 if the word "country" is used in the MLS description of the property and 0 otherwise. We then substitute the estimated $\log(DOM_{i,t})$ into the right-hand side of the Price equation:

$$\log\left(P_{i,t}\right) = \alpha_1 + \delta_1 \log\left(\widehat{DOM}_{i,t}\right) + \beta_1 \boldsymbol{X}_{i,t} + \lambda_1 \boldsymbol{Words}_{i,t} + QY_t + L_C + \epsilon_{i,t} \tag{4}$$

Similarly, instead of Equation (2), we first estimate $log(P_{i,t})$ by using *GreenCert* as the instrumental variable:

$$\log\left(P_{i,t}\right) = \pi_2 + \theta_2 GreenCert_{i,t} + \Lambda_2 \boldsymbol{X}_{i,t} + \Gamma_2 \boldsymbol{Words}_{i,t} + QY_t + L_C + v_{i,t} \tag{5}$$

Then we use the estimated price as an explanatory variable so that Equation (2) becomes:

$$\log\left(DOM_{i,t}\right) = \alpha_2 + \delta_2 \log\left(\widehat{P_{i,t}}\right) + \beta_2 \boldsymbol{X}_{i,t} + \lambda_2 \boldsymbol{Words}_{i,t} + QY_t + L_C + \epsilon_{i,t} \tag{6}$$

4. Empirical Results

Table 5 displays the results of estimating both the price of the home and days on the market before the house is sold. We report the first stage regressions for completeness, keeping in mind that the instrument variables for Ln Dom and Ln Price estimates are “country” and “green” certified homes respectively. However, it is the second stage estimates that are of interest to determine the effects of both the quantitative and qualitative measures.

Table 5 Instrument Variables Results

Word	First Stage IV Dependent Variable: Log DOM	Second Stage IV Dependent Variable: Log Price	Marginal Effects
Instrument	0.049* (7.196)		
Endogenous Variable		0.068 (1.039)	$141.23
Age	-0.001* (-8.185)	-0.001* (-6.008)	-$193.69
Age^2	0* (14.616)	0* (-4.746)	-$2.38
Distressed Sale	0.059* (12.628)	-0.385* (-86.090)	-$95,048.20
Ln Sqft Total	0.156* (27.699)	0.990* (93.736)	$107.92
Baths Full	0.043* (18.04)	0.154* (50.044)	$38,003.96
Baths Half	0.027* (10.893)	0.038* (18.187)	$9,487.47
Beds Total	-0.018* (-8.814)	-0.087* (-55.642)	-$21,502.41
Ln Lot Size	0.053* (25.653)	0.062* (16.687)	$26,266.15
New Construction	0.318* (87.876)	0.035*** (1.674)	$8,692.77
Adorable	-0.050* (-5.237)	0.175* (31.616)	$43,357.90
Awesome	-0.04* (-4.838)	0.034* (7.096)	$8,309.63
Gorgeous	-0.023* (-5.296)	0.059* (22.569)	$14,503.41

(***Continued…***)

(Table 5 Continued)

Word	First Stage IV: Dependent Variable: Log DOM	Second Stage IV: Dependent Variable: Log Price	Marginal Effects
Historic	0.057*	0.153*	$37,753.12
	(4.553)	(21.918)	
Investment	0.012	-0.246*	-$60,769.96
	(1.345)	(-58.583)	
Large	0.016*	-0.011*	-$2,601.79
	(6.780)	(-6.991)	
Luxurious	0.056*	0.026*	$6,385.51
	(5.710)	(4.369)	
Motivated	0.280*	-0.081*	-$20,046.74
	(29.306)	(-4.272)	
Reduced	0.411*	-0.088*	-$21,667.52
	(47.555)	(-3.209)	
Spacious	0.017*	-0.030*	-$7,351.26
	(5.567)	(-16.263)	
One Exclamation	0.006***	0.027*	$6,649.62
	(1.938)	(18.853)	
Two Exclamations	0.002	0.017*	$4,295.19
	(0.515)	(10.653)	
Three Exclamations	0.016*	0.009*	$2,265.23
	(3.744)	(3.971)	
Four Exclamations	0.020*	0.003	$683.11
	(3.590)	(0.926)	
Five Exclamations	0.019*	-0.001	-$320.94
	(3.886)	(-0.499)	
(Intercept)	1.856*	3.913*	
	(13.319)	(28.097)	
Quarter-Year Dummies	Yes	Yes	
County Dummies	Yes	Yes	
Additional Words	Yes	Yes	
Additional Physical Property Attributes	Yes	Yes	
R^2	0.125	0.779	
F-Statistic	406.551	10,121.90	
N	426,816	426,816	

(*Continued…*)

(Table 5 Continued)

Word	First Stage IV Dependent Variable: Log Price	Second Stage IV Dependent Variable: Log DOM	Marginal Effects
Instrument	0.013* (6.405)		
Endogenous Variable		-0.617*** (-1.786)	0
Age	-0.001* (-10.472)	-0.002* (-5.633)	-0.243
Age^2	0* (-8.789)	0* (6.904)	0.003
Distressed Sale	-0.381* (-173.948)	-0.176 (-1.331)	-21.035
Ln Sqft Total	1.001* (380.840)	0.774** (2.235)	0.041
Baths Full	0.157* (139.834)	0.140* (2.581)	16.761
Baths Half	0.040* (35.136)	0.052* (3.650)	6.166
Beds Total	-0.088* (-90.813)	-0.073** (-2.382)	-8.715
Ln Lot Size	0.065* (68.267)	0.093* (4.101)	19.139
New Construction	0.057* (33.682)	0.353* (17.267)	42.257
Adorable	0.172* (38.824)	0.057 (0.937)	6.763
Awesome	0.031* (7.969)	-0.021 (-1.542)	-2.523
Gorgeous	0.057* (27.648)	0.012 (0.585)	1.415
Historic	0.157* (26.920)	0.153* (2.760)	18.361
Investment	-0.245* (-60.196)	-0.140 (-1.639)	-16.703
Large	-0.009* (-8.754)	0.010** (2.434)	1.178
Luxurious	0.030* (6.488)	0.074* (5.147)	8.88

(Continued…)

(Table 5 Continued)

Word	First Stage IV Dependent Variable: Log Price	Second Stage IV Dependent Variable: Log DOM	Marginal Effects
Motivated	-0.062*	0.242*	28.979
	(-13.862)	(10.288)	
Reduced	-0.060*	0.374*	44.704
	(-14.792)	(16.624)	
Spacious	-0.029*	-0.001	-0.074
	(-20.014)	(-0.060)	
One Exclamation	0.027*	0.022**	2.689
	(20.134)	(2.272)	
Two Exclamations	0.018*	0.013***	1.505
	(10.913)	(1.793)	
Three Exclamations	0.010*	0.022*	2.69
	(5.103)	(3.948)	
Four Exclamations	0.004	0.023*	2.724
	(1.578)	(3.802)	
Five Exclamations	0	0.019*	2.24
	(-0.008)	(3.756)	
(Intercept)	4.040*	4.350*	
	(62.093)	(3.100)	
Quarter-Year Dummies	Yes	Yes	
County Dummies	Yes	Yes	
Additional Words	Yes	Yes	
Additional Physical Property Attributes	Yes	Yes	
R^2	0.785	0.064	
F-Statistic	10,409.74	380.186	
N	426,816	426,816	

Notes: * indicates significance at 1% level, ** indicates significance at 5% level, and *** indicates significance at 10% level.

[a] Additional Physical Property Attributes: Dummy for heating system (central unit, heating pump, window unit, furnace, or other), unheated square footage, dummy for septic tank, dummy for green certification, dummy for exterior (brick, siding, or other), dummy for floor type (carpet, tile, or wood), and dummy for parking type (garage, carport, or other).

[b] The instrument in the first stage IV with dependent variable Ln DOM is equal to 1 if the description says the word "country", 0 otherwise. The endogenous variable in the second stage IV with dependent variable Ln Price is the predicted value of Ln DOM from the first stage IV.

[c] The instrument in the first stage IV with dependent variable Ln Price is equal to 1 if the home is green certified, 0 otherwise. The endogenous variable in the second stage IV with dependent variable Ln DOM is the predicted value of Ln Price from the first stage IV.

We first consider the second stage estimation of log price. The regression includes the estimated days on the market and the standard hedonic measures of age, size and fixed effects that denote sales quarter and county. The estimated coefficient for the endogenous variable, DOM, shows the trade-off between price and time, and suggests that an additional day will increase the sale price of a home by US$141.23. For Age, we find the estimated coefficient of -0.001 implies that a one year increase will decrease the value of the home by US$193.69. Age^2, however, has a negligible effect on the value of a home and may reflect the relatively young housing stock in the growing Charlotte area.

Turning to size, the Ln Sqft coefficient suggests that increasing the structure by one square foot will increase the mean home value by US$107.92. Additional Baths Full and Baths Half also add value, but this is not true for Beds Total. The last result is consistent with many previous hedonic studies that find additional bedrooms, holding home size constant, decrease the price, as more rooms decrease the average room size. Regarding Ln Lot Size, an increase of an acre of land implies an additional property value of US$26,266.15.

Two flag variables, Distressed Sale and New Construction, also significantly affect home value. Distressed Sale, an identifier marked by the agent, decreases home value by more than US$95,000. Economically, this is a significant drop in value and warrants further thought later in our analysis. New Construction, on the other hand, adds US$8,692.77. For the average home in our sample, this is approximately a 3.50% premium.

Of particular interest is the information content of certain descriptors frequently found in the MLS listings. Table 5 again lists the more salient qualitative variables, although a complete set of estimated coefficients are available from the authors. The second stage Price equation finds positive coefficients for Adorable, Awesome, Gorgeous, Historic and Luxurious. Of this group, Adorable exhibits the largest coefficient and adds more than US$43,000 to the value of the house. Recalling that adorable homes are mostly small and therefore modest in price, the premium is especially large relative to the value of the home. Remember that Adorable supplements size information including square footage, lot size and number of rooms, and potentially distinguishes the home from others in the neighborhood.

Five words have negative coefficients in the hedonic model and suggest attributes that detract from the value of the home. Three of the words, Investment, Motivated and Reduced, likely describe the property conditions or circumstances that surround the sale of the property. Inclusion in the MLS listing results in a US$20,000 to US$60,000 loss in value. Recall that Goodwin, Waller, and Weeks (2014) include a “signal” variable for Motivated and Reduced in one form of their hedonic model and obtain a positive effect on the price of the home.[3] Instead, our results are more in line with Springer (1996)

---

[3] We note, however, that in their Exhibit 7 Price equation, Goodwin, Waller, and Weeks

who finds that motivated sellers agree to lower transaction prices. We continue a discussion of the signal variables in the analysis below.

Two other words, Large and Spacious, also have negative coefficients in the Log Price equation. Use of the word Large decreases the price of a home by US$2601.79, whereas Spacious negatively affects the value by US$7351.26. Given that the model already adjusts for size in several ways, a further narrative about roominess appears to detract from the value of the home. It may be that the buyer believes the realtor is trying to conceal the lack of space by claiming that the house or a room shows bigger than its actual size. From Figure 1, we already know that all but the smallest houses use the words “large” and “spacious” in roughly the same proportion as the homes with the greatest dimensions.

Finally, we consider the use of exclamation marks in the remarks of the agents. One exclamation mark appears to generate extra excitement about the property and raises its value by US$6649.62. However, two exclamation marks raise the value of the home by US$4295.19 and more exclamation marks used result in less added value. In fact, if five exclamation marks are used, the marginal effect is negative, although the coefficient itself is statistically insignificant. The final result is that overuse (abuse) of exclamation marks has virtually no effect on buyer enthusiasm for the property.

We next examine the second stage estimate for DOM. The results appear in the last column of Table 5. With respect to the quantitative variables, two items are of note. First, distressed properties appear to sell 21 days faster than transactions between a willing buyer and seller. However, the coefficient is statistically insignificant. This suggests high volatility in DOM for distressed sales with some transactions completed more quickly and others taking longer than otherwise ordinary sales. The second item is the statistically significant effect of New Construction. New homes take, on average, 42.257 days longer to sell than older homes. This is more than 35.30% longer than the average home sold in our sample.

Concerning the information effects of qualitative variables, only a few words are statistically significant. Historic and Luxurious homes appear more difficult to sell, with the marginal effects of 18.361 and 8.88 days, respectively. Similarly, properties with the descriptions Motivated or Reduced also take longer to close. Both coefficients are statistically significant and suggest remarks that mention Motivated need an additional 28.979 days to sell, on average, whereas Reduced homes take 44.704 more days to close. These results are consistent with both Springer (1996) and Goodwin, Waller, and Weeks (2014) who submit that these properties may be more difficult to sell. Lastly,

---

(2014) include the signal dummy variable as their only qualitative variable. If signal variables appear in MLS remarks along with positive descriptors, it is not clear what net effect is being captured in the coefficients.

the use of exclamation market uniformly hurts the sale of a property. Depending on the number of entries, exclamation marks lead to an additional 1.5 to 2.7 days to close.

Related to Time on the Market is the likelihood that the property sells at all. Table 6 presents the logit results for the sale of the property along with the marginal effects of including any salient word in the MLS description. Included in this regression is the variable, Ratio, which is equal to the listing price of the property divided by its hedonic value. The hedonic value depends only on quantitative MLS measures.

The negative Ratio coefficient implies that a higher listing price relative to the hedonic value of a home means a lower probability that the home will sell. Intuitively, setting a relatively high listing price discourages potential buyers from making offers. If the listing price appears too high relative to the market, buyers may question whether the owner is even serious about selling the home.

Generally, the estimated coefficients suggest that qualitative variables that decrease the time of a home on the market or increase its value are more likely to sell, ceteris paribus. Thus, Adorable, Awesome and Gorgeous increase the probability of a sale by 5.09%, 1.02%, and 1.44% respectively. On the other hand, variables that increase the time of a home on the market or decrease its value tend to decrease the likelihood of a successful closing. This includes Historic and Luxurious; size variables - Large and Spacious; and signal variables - Investment, Motivated and Reduced. The range of marginal effects is a decrease in the probability of a sale from 0.23% (Historic) to 12.78% (Motivated). Once more, the effect of exclamation marks is small, and five exclamation marks tend to reduce the probability of a sale by 1.35%.

Finally, the Ratio variable in the Table 6 logit regression assumes a hedonic price that does not reflect information conveyed in the remarks of the agent. Specifically, the denominator of Ratio does not capture the impact on price of descriptions that include the terms Investment, Motivated or Reduced. From Table 6, it seems that listing price relative to perceived market value is important in determining whether the property ultimately sells.

Table 7 reports the distributions of Ratio for the full sample as well as homes that have MLS comments with the words Investment, Motivated or Reduced. For a typical house in our sample, the listing price is 6.94% above its hedonic price, the latter reflecting the perceived market value of the property. In the center of the distribution, i.e. between the 25th and 75th percentiles, Ratio varies from 0.827 to 1.188. For homes that the agent has marked distressed, the distribution of relative prices is above and skewed more to the right. Note, Distressed is a box that the real estate agent can check off, and considered a quantitative variable that is included in our hedonic estimate. However, the last three variables are qualitative measures and not explicitly accounted for in the hedonic estimate for market value. For these variables, the distribution of Ratio

Table 6 Probability of Home Sold

Word	Estimate	Marginal Effects
Ratio [a]	-0.414*	-0.087*
	(-67.482)	(-68.351)
Adorable	0.249*	0.051*
	(9.900)	(10.179)
Awesome	0.049**	0.010**
	(2.412)	(2.422)
Gorgeous	0.069*	0.014*
	(6.373)	(6.410)
Historic	-0.011	-0.002
	(-0.389)	(-0.389)
Investment	-0.344*	-0.074*
	(-17.541)	(-17.266)
Large	-0.050*	-0.011*
	(-9.048)	(-9.039)
Luxurious	-0.145*	-0.031*
	(-6.372)	(-6.309)
Motivated	-0.590*	-0.128*
	(-31.178)	(-30.808)
Reduced	-0.203*	-0.043*
	(-10.953)	(-10.814)
Spacious	-0.066*	-0.014*
	(-8.809)	(-8.780)
One Exclamation	0.022*	0.005*
	(3.167)	-3.170
Two Exclamations	0.035*	0.007*
	(4.175)	(4.183)
Three Exclamations	-0.017***	-0.004***
	(-1.670)	(-1.669)
Four Exclamations	-0.016	-0.003
	(-1.168)	(-1.167)
Five Exclamations	-0.064*	-0.014*
	(-5.591)	(-5.569)
(Intercept)	3.316*	
	(35.320)	
Quarter-Year Dummies	Yes	
County Dummies	Yes	
Additional Words	Yes	
Vector of Physical Property Attributes	Yes	
Pseudo R^2	0.096	
N	711,188	

Notes: * indicates significance at 1% level, ** indicates significance at 5% level, and *** indicates significance at 10% level

[a] Ratio is the listing price divided by the estimated sales price of the home. The estimated sales price comes from the hedonic regression.

is shifted roughly 4-9% to the left. Without accounting for the information in the MLS description, these properties look like bargains compared to the full sample. Thus, it is important to convey full information to maximize the likelihood of a closing and realize at or near market value for the home.

Table 7 Listed Price to Hedonic Value Ratio Distributions

Panel A							
Sample	Mean	Median	Std. Dev	Min	Max	25%	75%
Full Sample	1.069	0.987	0.467	0.004	53.977	0.827	1.188
Distressed	1.115	1.054	0.509	0.031	18.324	0.804	1.345
Investment	0.976	0.883	0.614	0.008	32.413	0.681	1.123
Motivated	1.009	0.946	0.380	0.098	6.375	0.798	1.128
Reduced	1.029	0.952	0.418	0.011	11.171	0.789	1.160

Note: Hedonic value is created with the quantitative fields from the MLS (i.e. number of bathrooms, bedrooms, square footage, etc.). In addition, it includes year-quarter and county fixed effects, as well as a dummy for whether the home is a distressed listing.

Panel B							
Sample	Mean	Median	Std. Dev	Min	Max	25%	75%
Full Sample	1.073	0.992	0.466	0.004	74.376	0.833	1.193
Distressed	1.114	1.054	0.497	0.028	18.233	0.809	1.338
Investment	1.265	1.149	0.782	0.012	40.702	0.887	1.462
Motivated	1.096	1.027	0.404	0.103	6.598	0.870	1.226
Reduced	1.104	1.024	0.445	0.010	14.213	0.853	1.245

Note: Hedonic value is created with the quantitative fields from the MLS (i.e. number of bathrooms, bedrooms, square footage, etc.). In addition, it includes year-quarter and county fixed effects, as well as a dummy for whether the home is a distressed listing, one dummy for each of the top 44 words, and one dummy each if there is 0, 1, 2, 3, 4, or 5+ exclamations marks included in the MLS remarks section.

5. Concluding Remarks

This paper extends earlier work that considers the information content of MLS descriptions. Using a superior data set with over 700,000 observations, we document a set of the most frequently used descriptive terms and show that context is important. Following previous research, we use hedonic modeling to estimate the effect of qualitative information on market transactions. The analysis finds several words associated with higher prices with the largest effect attributed to the word “adorable”. As these words lead to more expensive home prices, they frequently take longer to sell. However, positive value words also make it more likely that the home will sell.

Other words including “distressed”, “investment”, “motivated”, and “reduced” tend to decrease home value, take longer to sell, and reduce the probability of a successful closing. The results suggest that in many cases, these words indicate the relatively poor condition of the property and thus lower home value. Future research should investigate whether these provide a signal to other agents that allows for more efficient markets. Whereas these words correlate with lower prices and longer transactions, future work might show whether price would be even lower or days on the market even longer had agent remarks not mentioned these terms.

Finally, two of the most popular descriptors, “large” and “spacious” refer to size. While the area of the home and lot size add to the value of the house, the words “large” and “spacious” decrease the price of the home. Future research might also consider whether use of these words is viewed cynically by potential buyers given that they already know the area of the home and can infer room size. Alternatively, realtors might mistakenly believe that by using these descriptions, they might otherwise interest potential buyers in looking at houses that face size constraints or limitations. In any case, the empirical work warrants additional work on the use of size expressions when home dimensions are already known.

Acknowledgement

The authors thank an anonymous reviewer and the editor for the suggestions and encouragement in the revision of the manuscript.

References

Anenberg, E. and Kung, E. (2014). Estimates of the Size and Source of Price Declines Due to Nearby Foreclosures. *The American Economic Review,* 104(8): 2527–2551.

Benefield, J., Cain, C. and Johnson, K. (2014). A Review of Literature Utilizing Simultaneous Modeling Techniques for Property Price and Time-On-Market. *Journal of Real Estate Literature,* 22(2): 149–175.

Black, S. (1999). Do Better Schools Matter? Parental Valuation of Elementary Education. *The Quarterly Journal of Economics,* 114(2): 577–599.

Campbell, J., Giglio, S. and Pathak, P. (2011). Forced Sales and House Prices. *The American Economic Review,* 101(5): 2108–2131.

Figlio, D. and Lucas, M. (2004). What's in a Grade? School Report Cards and the Housing Market. *The American Economic Review*, 94(3): 591–604.

Goodwin, K., Waller, B. and Weeks, S. (2014). The Impact of Broker Vernacular in Residential Real Estate. *Journal of Housing Research,* 23(2): 143–161.

Goodwin, K., Waller, B. and Weeks, S. (2018). Connotation and Textual Analysis in Real Estate Listings. *Journal of Housing Research,* 27(2): 93–106.

Haag, J., Rutherford, R. and Thomson, T. (2000). Real Estate Agent Remarks: Help or Hype? *Journal of Real Estate Research*, 20(1-2): 205–215.

Knight, J. (2002). Listing Price, Time on Market, and Ultimate Selling Price: Causes and Effects of Listing Price Changes. *Real Estate Economics,* 30(2): 213–37.

Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *The Journal of Political Economy,* 82(1): 34–55.

Smith, K. and Huang J. (1995). Can Markets Value Air Quality? A Meta-Analysis of Hedonic Property Value Models. *The Journal of Political Economy,* 103(1): 209–27.

Springer, T. (1996). Single Family Housing Transactions: Seller Motivation, Price and Marketing Time. *Journal of Real Estate Finance and Economics,* 13(3): 237–54.