INTERNATIONAL REAL ESTATE REVIEW

2025 Vol. 28 No. 2: pp. 117 – 148

Comparative Cheap Talk: Evidence from the Housing Market

Lingxiao Li
College of Business and Economics, California State University–Fullerton
E-mail: lingli@fullerton.edu

Jia Xie*
College of Business and Economics, California State University–Fullerton
E-mail: jxie@fullerton.edu

Using real estate transaction data, we provide the first empirical test of the comparative cheap talk model in Chakraborty and Harbaugh (2010), which predicts that a real estate agent can credibly reveal information of a house by making comparative statements that make the house more appealing in some areas but less along others. Consistent with the prediction, we find that comparative statements are associated with a 0.8% price premium, with all else being equal. The premium is larger for houses with more potential buyers, but switches sign and becomes a discount when there are too few buyers.

Keywords

Cheap talk, Comparative statement, Information asymmetry, House price

*Corresponding author

1. Introduction

Institutions and individuals rely greatly on the advice of experts to make decisions because they are privy to particular information. Experts, however, often have state-independent (and therefore biased) preferences so that they may want the decision maker to make the same decision regardless of the state of the nature. For instance, a financial adviser is very incentivized to recommend the stock or option that pays him/her the most whether or not the investments are really the best for the client. A biased salesperson always wants a customer to buy a product regardless of the quality.

The question of how the expert can credibly communicate important information for better decision making has been extensively studied in the literature (Crawford and Sobel, 1982; Ottaviani and Sørensen, 2006; Inderst and Ottaviani, 2009). In particular, Chakraborty and Harbaugh (2010) show that when information is multidimensional, a very biased expert who has state-independent preferences can still influence decision making by making comparative statements that help the expert in certain areas but hurt him/her in other areas.

In this paper, we focus on cheap talk communication between a real estate agent who wants to sell a house and the potential buyers, and we will test the comparative cheap talk model in Chakraborty and Harbaugh (2010) in real estate transactions.

Empirical testing of the comparative cheap talk model—or any information communication model—is very challenging given the difficulty in observing and measuring the communication between market participants. Nevertheless, real estate transactions provide a very attractive setting to test the impact of comparative cheap talk by experts, for two main reasons. First, the real estate market is characterized by heterogeneous products, unsophisticated buyers/sellers and the significant role of agents. A typical real estate consumer engages in a limited number of transactions during his or her lifetime. Due to imperfect information on the market value of properties and the location of potential buyers, sellers often seek the services of a real estate agent for assistance in the home selling process: 88% of buyers and 92% of sellers use a real estate agent when buying or selling their home (National Realtor Association, 2016), respectively. Since real estate agents play such an important role in real estate transactions, their communication strategies can potentially affect the final sales outcome to the extent that is testable with data.

Second, the listing agent often discloses information about house characteristics that are hard to observe and quality, by making public remarks in the multiple listing service (MLS). More importantly, these public remarks can be observed not only by buyers, but also researchers. For each real estate transaction, we say that the agent makes a comparative statement if s/he use both positive and negative words in the remarks, where the words are chosen from a dictionary in the real estate literature (Goodwin et al., 2014; 2018, and Haag et al., 2000).

Following Chakraborty and Harbaugh (2010), we examine cheap talk communication of multiple characteristics of a listed house between the listing agent (the expert) and potential buyers (decision makers). We show that the agent can credibly reveal information by making comparative statements that make the house more appealing in some areas but less in others. We also show that comparative statements can increase the expected house sale price when there are enough (i.e., more than 3) potential buyers, but hurt the seller by reducing the expected sale price when there are few buyers. This is because the comparative statements provide a better match of the house with the buyer who values it the most, but also reduce competition among potential buyers. When there are many buyers, the positive effect from better matching dominates. However, when there are few buyers, the negative effect from less competition dominates.

Next, we proceed to test these predictions by using the MLS data of 14,285 home sales in Indiana, USA. To the best of our knowledge, this paper is the first to provide direct empirical evidence of the comparative cheap talk model. Using the public remarks for each house, we identify whether the listing agent makes a comparative statement in promoting the listed property and how this affects the transaction outcomes.

To minimize the impacts from the different property types and different types of owners, we focus on single family houses owned by non-agent individuals, by omitting condominiums and agent-, bank- and government-owned houses. We also drop foreclosure and short sales, to minimize the impact of motivated sellers.

Indeed, we find that transactions where the listing agents made comparative statements are associated with a 0.80% price premium, even after controlling for a wide range of house characteristics, effort put forth by the agent, contractual terms, location, time of sale, and agent and broker characteristics. The results are robust to the endogeneity of the use of comparative statements, endogeneity of days-on-market (DOM), sample selection bias, and model misspecification.

This basic result is not only consistent with the prediction of the comparative cheap talk model, but also other competing hypotheses. For instance, homes sold with comparative statements may have unobservable characteristics that facilitate their high selling prices. However, this is unlikely a valid explanation, as negative comments are likely to represent undesirable characteristics, which will make homes sell for lower (rather than higher) prices.

Aside from our basic result, we find additional evidence in support of the comparative cheap talk explanation, but this could not be explained by unobservables. That is, the price premium of comparative statements is higher when there are more buyers. In particular, the premium switches sign and becomes a discount when there are few potential buyers. We find such evidence along three dimensions. First, we classify the full sample into the boom (2000-2006) and bust (2008-2010) periods. The former tends to have a seller's market, where each listing has more potential buyers, while the bust period is more likely to have a buyer's market. Indeed, we find that the price impact of comparative statements is positive in the boom period, but negative for the bust period.

The second dimension is about the number of competing houses. We define competing houses as other houses that are located in the same school district as the subject house and actively listed at the time when the subject house is sold. Houses that face competition from more competing houses are expected to have fewer potential buyers. We, therefore, classify our data into two exclusive and exhaustive subsamples according to the number of competing houses and then re-estimate the price premium of the comparative statements in these two subsamples. As expected, the price impact is positive for houses with little competition, but becomes negative for houses with more competition.

The third dimension is about the atypicality index of the house. Following Haurin (1988), we calculate the atypicality index for each house and classify our full data into two exclusive and exhaustive subsamples according to this index. A house with high atypicality is more likely to have unusual characteristics, and therefore may have fewer potential buyers. Indeed, we find that the price premium of comparative statements is higher for houses with higher atypicality.

This paper is related to two strands of the literature. First, the work is built on the theoretical literature of the communication of non-verifiable information, i.e., cheap talk. Crawford and Sobel (1982) first find that information can be partially communicated in the equilibrium of a cheap talk model where the information space is one-dimensional and the agent has state-dependent preferences. Chakraborty and Harbaugh (2007; 2010) further show that informative communication can be achieved in multidimensional models even if the preferences of the agent are state independent. More recently, Malenko and Tsoy (2019) show that information can be fully delivered in a dynamic setting, such as in an ascending-price (English) auction. Bouvard et al. (2015) build a stylized model to study the optimal level of information disclosure by regulators of the finance system. This paper is the first, to the best of our knowledge, to provide empirical evidence of cheap talk theories.

Second, the paper contributes to the empirical literature on information asymmetry in the real estate market. For instance, Garmaise and Moskowitz (2004) focus on the information asymmetry between sellers and buyers in the commercial real estate market. By using property tax assessment quality as a measure of information asymmetry, they find strong evidence of information asymmetry. Rutherford et al. (2005) and Levitt and Syverson (2008) study the information asymmetry between real estate agents and their seller clients. By using MLS data, they find that agents sell the house of their clients cheaper and faster than their own house, and interpret this finding as evidence of the informational disadvantage of the sellers relative to agents. More recently, Agarwal et al. (2019) and Allen et al. (2019) focus on the information asymmetry of the buy-side between real estate agents and buyers in the housing market. They find that real estate agents enjoy a 2.45%-4% discount when buying houses for themselves versus for their clients. Kurlat and Stroebel (2015) find that neighborhood characteristics provide a significant source of information asymmetry in housing markets. In this paper, we look instead at the information asymmetry between the listing agent and buyer. This paper is also closely related to two recent papers - Shen and Ross (2021) and Sing and Zhou (2024), who study the effectiveness of various marketing strategies used by real estate agents. Shen and Ross (2021) quantify the value of "soft" information contained in real estate property descriptions, and find that a one standard deviation increase in the uniqueness of a property based on this "soft" information leads to a 10%-15% increase in property sale price. Sing and Zhou (2024) study the housing market in Singapore, and find that houses sold through the online channel are sold for 4.21% higher than those sold through the traditional channel.

The next section of the paper presents the predictions of the theoretical model. Section 3 provides an overview of the data. In Section 4, we present a discussion on the estimation of the models. Section 5 shows the results of the alternative robustness check. Section 6 provides the empirical results on how the impact of comparative statements varies with the number of buyers. Section 7 summarizes the results and offers some concluding remarks.

2. The Model

The model is mostly based on Section II. C of Chakraborty and Harbaugh (2010). A real estate agent lists a house for sale on the MLS so that all potential buyers can view the listing. The agent privately knows that there are a number of characteristics that affect the house value but these cannot be easily measured and quantified, such as the quality of the workmanship and the material used to build the house, and amount of maintenance done. Without loss of generality, we focus on two characteristics denoted by $\theta = (\theta_1, \theta_2)$.

The agent can communicate information of θ to all potential buyers through public remarks on the MLS listing that do not incur any cost, $m \in M$. None of the potential buyers know the true value of θ, but all of them have the same previous belief on the distribution of θ which—without loss of generality—can be assumed to be uniformly distributed on $[0,1] \times [0,1]$. Indeed, even if θ does not follow a joint uniform distribution, the monotonic transformation

$(\eta 1, \eta 2) = \left(F_1^{-1}(\theta_1), F_2^{-1}(\theta_2)\right)$follows a uniform distribution, where F_1 and F_2 are the commutative distribution functions of θ_1 and θ_2, respectively. Moreover, we can treat (η_1, η_2) as a new state variable.

A communication strategy of the agent specifies $m \in M$ as a function of the house characteristics θ.[1] Each buyer estimates the expected value of θ, given his/her previous belief, the communication strategy of the agent, and remarks of the agent. The updated estimate of each buyer is denoted by $e = E[\theta|m]$.

Given the updated estimate, the valuation of the house by each potential buyer is $v_i(e) = \alpha_i e_1 + (1 - \alpha_i) e_2$,where $\alpha_i \in [0,1]$ measures how much Buyer i cares about θ_1 relative to θ_2. We assume that each α_i is independent and uniformly distributed identically on $[0,1]$, and independent of θ. Buyer i privately knows his/her own α_i.

The listing agent is paid a commission that is a fixed percentage of the final sale price. The utility of the agent can be written as:

$$u = r \cdot P \tag{1}$$

where r is the commission rate and P is the final sale price. We assume that the agent sells the house in a first-price auction. Since valuation of the house by each agent is private and independently distributed (given e_1 and e_2), according to the revenue equivalence theorem, the first-price and second-price auctions yield the same expected sale price (Menezes and Monteiro, 2008). In addition, the Bayesian Nash equilibrium bidding strategy of each buyer in a second-price auction is to bid his/her valuation. Therefore, the expected sale price of the house is:

$$E[P] = E[v_{2:n}] \tag{2}$$

where $v_{2:n}$ is the second highest valuation among n potential buyers.

To derive the formula of $v_{2:n}$ we need to distinguish between two cases: (i) $e_1 \geq e_2$ and (ii) $e_1 < e_2$ In particular, we have:

$$v_{2:n} = \begin{cases} \alpha_{2:n} \cdot e_1 + (1 - \alpha_{2:n}) \cdot e_2, & if\ e_1 \geq e_2 \\ \alpha_{n-1:n} \cdot e_1 + (1 - \alpha_{n-1:n}) \cdot e_2, & if\ e_1 < e_2 \end{cases} \tag{3}$$

where $\alpha_{j:n}$ is the jth highest value of α among all the n buyers. That is, $\alpha_{2:n}$is the second highest value of α, and $\alpha_{n-1:n}$ is the second lowest, among all of the n buyers.

Substituting Equation (3) into Equation (2) and then Equation (2) into Equation (1), we get:

[1] We focus on the pure strategy of the agent.

$$E[u(e)] = \begin{cases} r \cdot [E[\alpha_{2:n}] \cdot e_1 + (1 - E[\alpha_{2:n}]) \cdot e_2], & if\ e_1 \geq e_2 \\ r \cdot [E[\alpha_{n-1:n}] \cdot e_1 + (1 - E[\alpha_{n-1:n}]) \cdot e_2], & if\ e_1 < e_2 \end{cases} \quad (4)$$

Since α follows a uniform distribution on $[0,1]$, $E[\alpha_{j:n}] = (n - j + 1)/(n + 1)$. Then Equation (3) can be rewritten as:

$$E[u(e)] = \begin{cases} \frac{r}{n+1} \cdot [(n-1) \cdot e_1 + 2 \cdot e_2], & if \quad e_1 \geq e_2 \\ \frac{r}{n+1} \cdot [2 \cdot e_1 + (n-1) \cdot e_2], & if \quad e_1 < e_2 \end{cases} \quad (5)$$

Note that $E[u(e)]$ is strictly increasing in both e_1 and e_2. That is, the expected sale price increases with the estimate of the buyer of the two characteristics of the house. Therefore, the listing agent is incentivized to misreport the information of θ through exaggeration and puffery, which are not credible.

A pure-strategy perfect Bayesian equilibrium of this cheap talk game is fully specified by the communication strategy of the listing agent. There is always a "babbling" equilibrium where there is no communication. That is, the listing agent describes the two characteristics of the house in a flattering manner and the buyers simply ignore what the agent says.

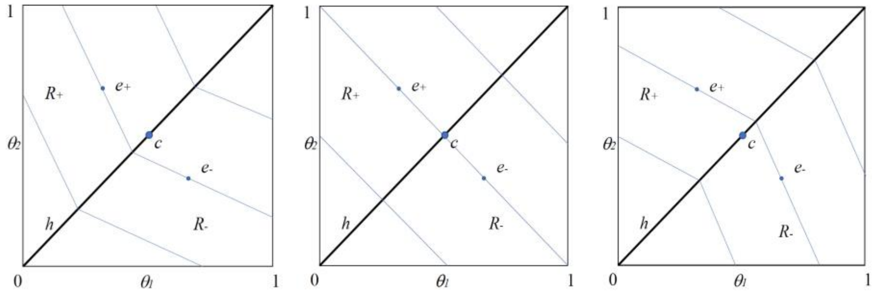
There might be, however, other equilibria—what we call responsive equilibria—in which communication is both informative and influential. A communication strategy is informative if the remarks of the agent vary with house characteristics. That is, m is not constant with respect to θ on the equilibrium path. A communication strategy is influential if the belief of the potential buyers is affected by the remarks of the agent, that is, e is not a constant on the equilibrium path.

We now set to find a responsive equilibrium. We distinguish among three cases: (i) $n = 2$, (ii) $n = 3$, and (iii) $n > 3$. If $n = 2$, i.e., there are two buyers, then Equation (5) becomes:

$$E[u(e)] = \begin{cases} \frac{r}{3} \cdot [e_1 + 2e_2], & if\ e_1 \geq e_2 \\ \frac{r}{3} \cdot [2e_1 + e_2], & if\ e_1 < e_2 \end{cases} \quad (6)$$

In this case, $E[u(e)]$ is a concave function of e, and its indifference curves are bowed in as shown in Figure 1a. Suppose the space is partitioned by a line h—from $(0,0)$ to $(1,1)$ — into two subspaces $R+$ and $R-$, and the listing agent indicates the region that θ falls into. The estimates $e^+ = E[\theta|R+]$ and $e^- = E[\theta|R-]$ fall on the same indifference curve, thus implying that the listing agent has no incentive to misreport the region that θ falls into.

Figure 1 Equilibrium Communication Strategies for Different Values of n



(a) $n = 2$ (b) $n = 3$ (c) $n > 3$

Note that in the babbling equilibrium, where messages convey no information, the estimate of the buyers is on the center point $c = (1/2, 1/2)$ in Figure 1a. Equation (6) is an increasing and concave function of *e*, thus implying that *c* is on an indifference curve that has a higher expected sale price than that where e^+ and e^- are located. That is, the equilibrium estimates $E[\theta|m]$ lead to lower expected agent utility than $E[\theta]$. In other words, when there are only two buyers, there is a responsive equilibrium that has a lower expected sale price than the babbling equilibrium.

Next, we examine the three-buyer case where $n = 3$, then we have:

$$E[u(e)] = \frac{r}{2}(e_1 + e_2) \tag{7}$$

That is, the expected utility of the listing agent is increasing and linear in *e*. In this case, the indifference curve of the utility of the agent is shown in Figure 1b. Again, we can draw a line h— from $(0,0)$ to $(1,1)$ — that partitions the space into two subspaces: $R+$ and $R-$. If the listing agent indicates the region that θ falls into, then the estimate $e^+ = E[\theta|R+]$ and $e^- = E[\theta|R-]$ will fall on the same indifference curve, thus implying that the listing agent has no incentive to misreport the region that θ falls into. In addition, c, e^+ and e^- are all on the same indifference curve, thus implying that responsive equilibrium has the same expected utility (sale price) as the babbling equilibrium, when there are three buyers.

When there are more than 3 buyers, i.e., $n > 3$, the indifference curves of Equation (5) are shown in Figure 1c: since the curves of Equation (5) are increasing and convex in *e*, the indifference curves bow outward. Again, we can draw a line $h -$ from $(0,0)$ to $(1,1)$— that partitions the space into two subspaces: $R+$ and $R-$. If the listing agent indicates the region that θ falls into, then the estimate $e^+ = E[\theta|R+]$ and $e^- = E[\theta|R-]$ will fall on the same indifference curve, thus implying that the listing agent has no incentive to misreport the region that θ falls into. In addition, both e^+ and e^- are on an indifference curve that has a higher expected utility than the difference curve where c is located. This implies that the responsive equilibrium has a higher expected utility (sale price) than the babbling equilibrium, when there are more than three buyers.

Finally, by taking derivative of Equation (5) in n, we get:

$$\frac{\partial E[u(e)]}{\partial n} = \begin{cases} \frac{2r(e_1 - e_2)}{(n+1)^2} \geq 0\,, & if \;\; e_1 \geq e_2 \\ \frac{2r(e_2 - e_1)}{(n+1)^2} > 0, & if \;\; e_1 < e_2 \end{cases} \tag{8}$$

Therefore, the expected sale price increases with the number of buyers. We summarize the above results in the following two propositions:

Proposition 1. *A responsive equilibrium where communication is both informative and influential is found in a multidimensional cheap talk model.*

Proposition 2. *The expected sale price in the responsive equilibrium increases with the number of buyers. In particular, the responsive equilibrium is associated with a price premium— relative to the babbling equilibrium—when there are many buyers, but a price discount when there are too few (less than 4) buyers.*

In the following, we test Propositions 1 and 2 with MLS data.

3. Data

We use the MLS data from Indiana, USA, which cover all residential real estate transactions that involve real estate agents in Johnson County from June 1, 2000 to May 31, 2010. Johnson County is one of the largest counties in Indiana and essentially a suburb of the city of Indianapolis in the adjacent county of Marion.

The MLS data employed are unique in several respects. First, the data contain detailed information about each transaction, including sale price, property characteristics, contract term, calendar information (listing and closing dates), and geographic location (school district). Physical characteristics include, but are not limited to, various measures of the scale of the property (number of bathrooms, size of garage, fireplace, pool, and square footage), age of the property, and siding (vinyl, stone, brick, etc.). Contract terms include the duration of the listing contract, buyer agent commission rate, and whether the listing agent has the exclusive-right-to-sell, which gives the agent the right to receive the commission regardless of who brings the buyer. Calendar information is employed to generate property marketing span—i.e., DOM—which is calculated as the number of days from listing date to sold date.

Second, the MLS data contain the public remarks of the listing agent for each listing. By searching for positive and negative words in the public remarks, we can identify the usage of comparative statements. We follow the literature in defining positive and negative words, which are shown in Table 1. (Goodwin et al., 2014, 2018; Haag et al., 2000.) We say the comparative communication strategy is used if at least one positive word and at least one negative word are used in the public remark.

Table 1 Positive and Negative Words

Positive words	#	Positive words	#	Negative words	#
Great	5,337	Convenient	563	As is	934
Beautiful	3,079	Remodeled	528	Motivated	386
Custom	2,094	Replaced/new roof	503	Repair	177
Spacious	1,749	Charming	480	Relocating	99
Updated	1,415	Immaculate	470	Vacant	94
Painted/New paint	1,032	Fabulous	448	Cosmetics	80
View	1,012	Fantastic	414	Transferred	74
New carpet	1,007	Great/ideal/prime /excellent location	371	Price reduced	73
Wonderful	947	Good	361	Quaint	53
Quiet	944	Bright	341	Bring offer	48
Gorgeous	832	Golf	257	Rustic	38
Perfect	822	Tasteful	175	Fixer upper	35
Lake	767	Great/super/fantastic /good buy	133	Foreclosure	28
Well maintained	702	Mint condition	110	Must sell	22
Quality	635	Classic	62	Defect	16
Lovely	579	Modern	53	Selling bonus	9
Cozy	576	Magnificent	52	Poor	3
				Anxious	3
				Fast sale	2

Notes: This table lists the positive and negative words in the remarks of the agent for each listing in the MLS, as well as the number of listings in which each word appears. We use the same dictionary as used in Haag et al. (2000), and Goodwin et al. (2014; 2018). The word "Cute" is omitted because it is inconsistently identified as positive or negative words in the literature.

The original data have 18,895 observations/transactions. To minimize the impacts from different types of properties and different types of owners, we focus on single family houses owned by non-agent individuals, by omitting condominiums and agent-, bank- and government-owned houses. We also drop foreclosure and short sales, to minimize the impact of motivated sellers. Finally, we discard observations with missing values and clearly erroneous data (zero bedroom, zero bathroom, less than 300 square footage, etc.). The final data include 14,285 transactions, among which 6,152 (43%) are associated with comparative statements. This result suggests that the responsive equilibrium where agents use comparative statements is very common in the real estate market.

Table 2 provides the summary statistics for transactions where comparative statements are used and for other transactions, respectively. While measures of the property size (e.g., number of bedrooms, square footage of the living space, and lot size) show no systematical differences across the two types of transactions, non-size-related characteristics indeed vary systematically. For instance, houses sold with comparative statements are more likely to have a fireplace, pool and basement, but less likely to be newly constructed. These systematic differences in the observables highlight the importance of controls in our following analysis.

Table 2 Summary Statistics

	Comparative Cheap Talk		Babbling Talk		Sig. of the difference-in-mean
	Mean	Std. Dev.	Mean	Std. Dev.	
Sale price	156,403	91,388	156,834	101,412	
House characteristics					
Bedroom	3.246	0.703	3.232	0.704	
Bathroom	2.408	0.887	2.406	0.874	
Garage size	1.952	0.777	1.960	0.761	
Square footage	2,100.33	1,047.70	2,097.19	1,076.09	
Acre of lot	0.630	1.947	0.588	2.114	
Fireplace	0.656	0.475	0.572	0.495	***
Pool	0.062	0.242	0.050	0.218	***
Basement	0.276	0.447	0.248	0.432	***
Exterior					
Brick	0.572	0.495	0.513	0.500	***
Vinyl	0.363	0.481	0.381	0.486	**
Stone	0.052	0.222	0.044	0.206	**
Wood	0.080	0.272	0.058	0.234	***
New construction	0.033	0.179	0.172	0.378	***
Age	24.417	30.414	21.553	30.229	***
Contract terms					
Exclusive right to sell	0.988	0.108	0.973	0.161	***
Buyer agent commission	3.372	0.356	3.413	0.661	***
Duration of contract	207.094	935.630	249.894	4052.09	
Effort of agent					
# of images	5.204	4.069	4.908	4.214	***
Open house	0.016	0.139	0.023	0.198	**
Virtual tour	0.071	0.256	0.073	0.260	
N	6,152		8,133		

Notes: * significant at the 10% level, ** significant at the 5% level, and *** significant at the 1% level.

4. Identification of Impact of Comparative Statements

We estimate the impact of comparative statements on the house sale price with the following hedonic model:

$$log(Sale\ Price) = \alpha + \beta * Cheaptalk + \gamma * X + \epsilon \qquad (9)$$

The dependent variable is the logged sale price,log($Sale\ Price$). X is a vector of property and transactional information. The key independent variable is the indicator variable— *Cheaptalk*—which equals 1 if the listing agent used comparative statements in the public remark of the listing, and 0 otherwise.[2]

We estimate the specification in Equation (9) and report the estimated coefficients, robust standard errors, and significance levels in Table 3. Each column in Table 3 represents a separate regression, with the specification gradually saturated from left to right as the set of control variables expand.

In Column (1), we regress the logged sale price on the *Cheaptalk* indicator variable and house characteristics. The house characteristic information includes the number of bedrooms, number of bathrooms, size of garage, etc. After controlling for these variables, we note that the price impact of the comparative statements is 1.73% and statistically significant. Also noteworthy, house characteristic information can explain for almost 80% of the variation in the sale price, thus suggesting that there are important house characteristic differences that affect the sale prices.

In Column (2), we control for the geographic and calendar information of the transaction. In particular, we control for the fixed effects of the school district and month and year of sale. The price impact of comparative statements is slightly reduced to 1.37% and remains significant at the 1% level.

In Column (3), we control for measures of the effort of the agent, such as the number of images of the house that are uploaded online, whether the agent conducted open houses, and whether the agent provides a virtual tour online. The price impact of comparative statements increases slightly to 1.43% and remains significant at the 1% level.

---

[2] We also coded and analyzed the data with more measures that take into consideration both tone (positive/negative) and intensity of comparative statements. We categorized each comparative statement as positive, negative, or neutral. We also considered the number of positive and negative words within each statement as a proxy for intensity. This allowed us to explore potential asymmetries in the effects of positive and negative ”cheap talk”. However, these effects are not consistently statistically significant across all of the specifications. Including tone and intensity measures in the same model leads to multicollinearity issues, thus making it difficult to isolate their individual effects. Given these challenges, and to maintain clarity and focus, we opt for the binary “cheap talk” indicator for our main analysis. This approach allows us to clearly identify the overall impact of using comparative language in MLS listings.

Table 3 Estimated Impact of Persuasive Cheap Talk on Sale Price

	(1)	(2)	(3)	(4)	(5)	(6)
Cheaptalk	0.0173***	0.0137***	0.0143***	0.0124***	0.0119***	0.0080**
	(0.0039)	(0.0037)	(0.0036)	(0.0037)	(0.0038)	(0.0041)
House characteristics						
# of bedrooms						
2	0.146**	0.184***	0.167***	0.173***	0.173***	0.194***
	(0.0664)	(0.0631)	(0.0610)	(0.0612)	(0.0612)	(0.0603)
3	0.136**	0.204***	0.187***	0.193***	0.193***	0.223***
	(0.0661)	(0.0630)	(0.0608)	(0.0610)	(0.0610)	(0.0603)
4+	0.149**	0.213***	0.198***	0.204***	0.204***	0.232***
	(0.0666)	(0.0635)	(0.0613)	(0.0615)	(0.0615)	(0.0608)
# of bathrooms						
2	0.0621***	0.0471***	0.0449***	0.0460***	0.0460***	0.0503***
	(0.0105)	(0.0100)	(0.0098)	(0.0097)	(0.0098)	(0.0092)
3	0.0782***	0.0613***	0.0576***	0.0601***	0.0601***	0.0727***
	(0.0128)	(0.0123)	(0.0119)	(0.0119)	(0.0119)	(0.0113)
4+	0.263***	0.241***	0.236***	0.237***	0.237***	0.249***
	(0.0165)	(0.0159)	(0.0156)	(0.0156)	(0.0156)	(0.0146)
Size of garage						
1 car	0.0656***	0.0708***	0.0673***	0.0677***	0.0677***	0.0595***
	(0.0155)	(0.0149)	(0.0144)	(0.0145)	(0.0145)	(0.0139)
2 cars	0.127***	0.128***	0.123***	0.123***	0.123***	0.111***
	(0.0149)	(0.0143)	(0.0140)	(0.0140)	(0.0140)	(0.0133)
3+ cars	0.293***	0.287***	0.276***	0.275***	0.275***	0.256***
	(0.0159)	(0.0154)	(0.0150)	(0.0150)	(0.0150)	(0.0144)

(Continued···)

(Table 3 Continued)

	(1)	(2)	(3)	(4)	(5)	(6)
Log (square feet)	0.544***	0.541***	0.532***	0.530***	0.530***	0.518***
	(0.0137)	(0.0132)	(0.0129)	(0.0129)	(0.0129)	(0.0128)
Acre of lot	0.0316***	0.0323***	0.0318***	0.0318***	0.0318***	0.0316***
	(0.0026)	(0.0028)	(0.0027)	(0.0027)	(0.0027)	(0.0027)
Fireplace	0.116***	0.100***	0.0919***	0.0902***	0.0902***	0.0811***
	(0.0046)	(0.0045)	(0.0044)	(0.0044)	(0.0044)	(0.0044)
Pool	0.0684***	0.0679***	0.0629***	0.0623***	0.0623***	0.0618***
	(0.0077)	(0.0076)	(0.0075)	(0.0075)	(0.0075)	(0.0077)
Basement	-0.0314***	-0.0427***	-0.0422***	-0.0422***	-0.0422***	-0.0403***
	(0.0081)	(0.0077)	(0.0075)	(0.0075)	(0.0075)	(0.0076)
Exterior material						
Brick	0.133***	0.120***	0.114***	0.114***	0.114***	0.108***
	(0.0041)	(0.0041)	(0.0040)	(0.0040)	(0.0040)	(0.0042)
Vinyl	-0.0936***	-0.0801***	-0.0813***	-0.0802***	-0.0802***	-0.0779***
	(0.0041)	(0.0040)	(0.0039)	(0.0039)	(0.0039)	(0.0039)
Stone	0.0783***	0.0666***	0.0660***	0.0660***	0.0659***	0.0663***
	(0.0102)	(0.0101)	(0.0099)	(0.0099)	(0.0099)	(0.0096)
Wood	-0.0199**	-0.0191**	-0.0185**	-0.0191**	-0.0191**	-0.0229**
	(0.0094)	(0.0091)	(0.0089)	(0.0090)	(0.0090)	(0.0090)
New construction	0.0839***	0.0921***	0.130***	0.128***	0.128***	0.113***
	(0.0060)	(0.0060)	(0.0062)	(0.0063)	(0.0063)	(0.0077)
Age	-0.00208***	-0.00179***	-0.00179***	-0.00180***	-0.00180***	-0.00174***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)

(Continued···)

(Table 3 Continued)

	(1)	(2)	(3)	(4)	(5)	(6)
Effort of agent						
# of images			0.0120***	0.0121***	0.0121***	0.0116***
			(0.0006)	(0.0006)	(0.0006)	(0.0006)
Open house			0.0102	0.0091	0.0092	0.0114
			(0.0072)	(0.0072)	(0.0072)	(0.0083)
Virtual tour			0.0499***	0.0507***	0.0506***	0.0488***
			(0.0069)	(0.0069)	(0.0069)	(0.0075)
Contractual terms						
Exclusive right to sell				0.0418***	0.0410***	0.0449***
				(0.0139)	(0.0140)	(0.0137)
Buyer agent commission				-0.0252***	-0.0252***	-0.0142**
				(0.0057)	(0.0057)	(0.0066)
Duration of contract				3.44e-07**	3.50e-07**	-3.64e-07
				(1.71e-07)	(1.72e-07)	(2.38e-07)
DOM	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects for						
School District	No	Yes	Yes	Yes	Yes	Yes
Year & Month of Sale	No	Yes	Yes	Yes	Yes	Yes
Agent experience	No	No	No	No	Yes	Yes
Broker	No	No	No	No	No	Yes
R^2	0.795	0.812	0.819	0.820	0.820	0.840
N	14,285	14,285	14,285	14,285	14,285	14,285

Notes: * significant at the 10% level, ** significant at the 5% level, and *** significant at the 1% level. Each column in Table 3 represents a separate regression, with the specification gradually saturated from left to right as the set of control variables expand. Numbers in parentheses are standard errors.

Much of the effort of the agent is actually unobservable and affected by the contract terms. For instance, agents with the exclusive right to sell will often have more incentive to put forth effort than the other agents. In Column (4), we control for the listing contractual terms, including the duration of the listing contract, whether the agent has the exclusive right to sell, and buyer agent commission rate. After controlling for these contractual terms, the price impact of comparative statements is reduced to 1.24% and remains significant at the 1% level.

Experience of Unobservable Agent n Comparative Cheap Talk

Even though we control for the detailed characteristics of the property and effort of agents, there is still the possibility that the agents who have ever used comparative statements may have certain characteristics that are systematically different from those who have never used comparative statements. These characteristics may give agents additional convenience or incentive to use comparative statements. For instance, agents who have used comparative statements may have a larger pool of buyers and sellers. To the extent that these agent characteristics can also affect the sale price, our estimations can be subject to biases caused by unobservable agent characteristics.

To proxy for the unobservable agent characteristics, we generate an indicator variable of agents who have ever used comparative statements in the sample period. This indicator is added to our estimation and the results are reported in Column (5) of Table 3. The price premium associated with comparative statements decreases to 1.19% and remains significant at the 1% level.

Unobservable Brokerage Firm Characteristics

According to Indiana state law, each real estate agent has to display his/her license in a real estate brokerage firm, and work under the guidance and legal protection of the brokerage firm. To the extent that brokerage firms may have different guidance and strategies, which may affect both the sale price and decision of the agent to use comparative statements, our previous estimates are subject to potential biases caused by unobservable brokerage firm characteristics.

To correct for the potential biases mentioned above, we control the fixed effects of the brokerage firm. The corresponding results, reported in Column (6) of Table 3, again show that comparative statements are associated with a 0.80% price premium, with all else being equal.

5. Robustness Check of Main Results

In this section, we check whether our main results are robust by studying the following potential issues: (i) the endogeneity of comparative cheap talk, (ii) the endogeneity of DOM, (iii) sample selection bias, and (iv) model misspecification. The results are reported in Table 4.

Table 4 Impact of Comparative Cheap Talk on Sale Price — Robustness Check

	(1)	(2)	(3)	(4)
Problem	Endogenous Cheap Talk	Endogenous Days-on-Market	Sample Selection Bias	Model Misspecification
Solution	Endogenous Switching Regression Model	3-Stage Lease Square	Heckman 2-Step Estimation	Propensity Sore Matching#
Cheaptalk	0.0050***	0.0079***	0.0077***	0.0081
	(0.0009)	(0.0040)	(0.0038)	(0.0105)
R^2	0.8473	0.0304	0.8346	N/A
		(0.0215)		
N	14,285	14,285	29,654	12,029
Other covariates:				
House characteristics	Yes	Yes	Yes	Yes
Effort of agent	Yes	Yes	Yes	Yes
Contract terms	Yes	Yes	Yes	Yes
Days-on-market	Yes	Yes	Yes	Yes
Fixed effects for:	Yes	Yes	Yes	Yes
School district	Yes	Yes	Yes	Yes
Month and year of sale	Yes	Yes	Yes	Yes
Agent experience	Yes	Yes	Yes	Yes
Broker	Yes	Yes	Yes	Yes

Notes: * significant at the 10% level, ** significant at the 5% level, *and ** significant at the 1% level.
#The propensity score matching uses kernel matching algorithm.

5.1 Endogeneity of Comparative Cheap Talk

Comparative cheap talk may be an endogenous decision that is affected by many factors. If there are unobservable factors that affect both the usage of comparative statements and the house sale price, then our previous estimations may be biased. Instrument variables are usually used to solve an endogeneity issue. However, it is often difficult to find a valid instrument variable that affects the usage of comparative statements but not the final sale price.

In this section, we adopt two alternative approaches to study the impact of this potential endogeneity. The results of both approaches suggest that our main results are robust to the endogeneity of using comparative statements.

5.1.1 Endogenous Switching Regression Model

In the first approach, we build and estimate an *endogenous switching regression model*.

Suppose that the sale price is determined by two different equations for two possible regimes (i.e., comparative cheap talk and babbling cheap talk), and selection into one regime is endogenously determined. The model comprises three equations as follows:

$$Cheaptalk_i = \begin{cases} 0, & if\, Cheaptalk_i^* = \mu + \lambda' \cdot Z_i + \xi_i \leq 0 \\ 1, & if\, Cheaptalk_i^* = \mu + \lambda' \cdot Z_i + \xi_i > 0 \end{cases} \tag{10}$$

$$log(sale\_price_{0i}) = \alpha_0 + \gamma_0' X_i + \epsilon_{0i}, \quad \rightarrow\ if\, Cheaptalk_i = 0 \tag{11}$$

$$log(sale\_price_{1i}) = \alpha_1 + \gamma_1' X_i + \epsilon_{1i}, \quad \rightarrow\ if\, Cheaptalk_i = 1 \tag{12}$$

Equation (10) is the selection equation, where $Cheaptalk_i$ is the indicator variable of comparative cheap talk and $Cheaptalk_i^*$ is the corresponding latent variable. Z_i is a vector of observables that affect the usage of comparative cheap talk, and assumed to be the same as X_i, except that the calendar information in $Z_i$is related to the time when the house is listed, instead of the time when the house is sold.

Equation (11) is an outcome equation that determines the sale price in the first regime,i.e., when the listing agent uses comparative cheap talk. Similarly, Equation (12) is the other outcome equation for the second regime, i.e., when the listing agent does not use comparative cheap talk. For observation i, observable outcome $\log(sale\ price_i)$ is either $\log(sale\ price_{0i})$ or $\log(sale\ price_{1i})$. However, the two outcomes cannot be observed simultaneously.

To estimate the endogenous switching regression model, we need the following distributional assumptions:

$$\epsilon_0 \sim N(0, \sigma_0^2) \tag{13}$$

$$\epsilon_1 \sim N(0, \sigma_1^2) \tag{14}$$

$$\xi \sim N(0,1) \tag{15}$$

$$Corr(\epsilon_0, \xi) /= 0 \tag{16}$$

$$Corr(\epsilon_1, \xi) /= 0 \tag{17}$$

where *Corr* stands for the correlation coefficient. Equations (16) and (17) suggest that the error terms of the outcome equations are correlated with the error term in the selection equation. That is, the selection into comparative cheap talk is endogenous.

We implement the maximum likelihood estimation of the endogenous switching regression model, see Equations (10)-(17). Then for each observation, we estimate the fitted values of the log sale price for each regime respectively, as follows:

$$log\left(\widehat{sale\_price_{0i}}\right) = \hat{\alpha}_0 + \hat{\gamma}_0' X_i + E[\epsilon_{0i} \mid Cheaptalk_i = 0] \tag{18}$$

$$log\left(\widehat{sale\_price_{1i}}\right) = \hat{\alpha}_1 + \hat{\gamma}_1' X_i + E[\epsilon_{1i} \mid Cheaptalk_i = 1] \tag{19}$$

Finally, we calculate the average difference in the fitted values of the log sale price across all observations:

$$\tau = \frac{\sum_i^N \left[log\left(\widehat{sale\_price_{1i}}\right) - log\left(\widehat{sale\_price_{0i}}\right)\right]}{N} \tag{20}$$

The estimated coefficient of τ is reported in Column (1) of Table 4, and can be interpreted as a treatment effect of comparative cheap talk on the sale price. The result from the endogenous switching regression model is consistent with our main results: comparative cheap talk is associated with a 0.50% price premium, holding everything else constant, and the premium is significant at the 1% level.

5.1.2 Testing for Selection Based on Unobservables

The use of comparative cheap talk can be based on unobservables. While we have controlled for a detailed set of factors in the estimations, it is possible that a small amount of selection on unobservables could explain for much of the estimated effect of comparative cheap talk. We now explore this possibility by using the relationship between comparative cheap talk and the observables to make inferences on the relationship between selection on the observables and unobservables.

We adopt the approach in Altonji et al. (2005; 2008) which estimates the relative amount of selection of unobservables required to explain the estimated comparative cheap talk effect if the true effect is zero (i.e., the null hypothesis). This technique relies on the following condition:

$$\frac{E(\epsilon \mid Cheaptalk = 1) - E(\epsilon \mid Cheaptalk = 0)}{Var(\epsilon)} = \frac{E(\gamma * X \mid Cheaptalk = 1) - E(\gamma * X \mid Cheaptalk = 0)}{Var(\gamma * X)}. \tag{21}$$

The left-side of Equation (21) represents the selection of unobservables while the right-side represents the selection of observables. This condition assumes that the use of comparative cheap talk relies on unobservables to the same extent as observables. Note that all items in Equation (21) can be estimated from the data, except for $E(\epsilon|Cheaptalk = 1) - E(\epsilon|Cheaptalk = 0)$.

Let $\tilde{\epsilon}$ be the residuals of a regression of *Cheaptalk* on *X* so that $Cheaptalk = \mu * X + \tilde{\epsilon}$.

Then substituting the last equation into Equation (9), one gets:

$$log(Sale\ Price) = \alpha + \beta * \tilde{\epsilon} +\cdot (\beta * \mu + \gamma) *\cdot X + \epsilon \tag{22}$$

Given that both ϵ and $\tilde{\epsilon}$ are orthogonal to *X*, Equation (22) leads to:

$$\begin{aligned} plim(\tilde{\beta}) &= \beta + \frac{Cov(\tilde{\epsilon}, \epsilon)}{Var(\tilde{\epsilon})} = \beta + \frac{Cov(Cheaptalk - \mu * X, \epsilon)}{Var(\tilde{\epsilon})} \\ &= \beta + \frac{Cov(Cheaptalk, \epsilon)}{Var(\tilde{\epsilon})} \\ &= \beta + \frac{Var(Cheaptalk)}{Var(\tilde{\epsilon})} * [E(\epsilon \mid Cheaptalk = 1) \\ &- E(\epsilon \mid Cheaptalk = 0)] \end{aligned}$$

That is, the bias in the estimated comparative cheap talk effect due to the selection of unobservables is:

$$Bias(\beta) = \frac{Var(Cheaptalk)}{Var(\tilde{\epsilon})} * [E(\epsilon \mid Cheaptalk = 1) - E(\epsilon \mid Cheaptalk = 0)] \tag{23}$$

The fraction in Equation (23) can be estimated directly from the data, and the item in the squared brackets can be calculated from Equation (21). As shown in Table 5, following Equation (23), we estimate the bias, *Bias*(β), to be 0.0052. Recall that the estimated comparative cheap talk effect is 0.0080 (Column 6 in Table 3). This suggests that the selection on unobservables needs to be more than 65% of the selection on observables,[3] which is very unlikely given that we have a detailed list of observables. Therefore, we reject the null hypothesis that the effect of comparative cheap talk on sale price is zero.

[3] 0.0052/0.080=65%.

Table 5 Amount of Selection on Unobservables Relative to Selection on Observables Required to Attribute the Entire Effects of Comparative Cheap Talk to Selection Bias

E(ϵ∣Cheaptalk = 1) − E(ϵ∣Cheaptalk = 0)	*Var(Cheaptalk)/Var(˜ϵ)*	**Bias(β)**	*ˆβ*	*ˆβ/Bias(β)*
0.0039	1.3241	0.0052	0.008	0.65

5.2 Endogeneity of Days-on-Market

The sale price and DOM of a house may be endogenously determined. If this is true, then we have a bias due to the endogeneity of the DOM. To solve this potential endogeneity problem, we re-estimate our full specification in Column 2 of Table 4 (as in Column 6 of Table 3) via 3-stage least square (3SLS) regressions, with the first stage being a regression of the log of listing price specified as follows:

$$\log(listing\ price) = \alpha_2 + \beta_2 * Cheaptalk + \gamma_2 * Z + \epsilon_2 \quad (24)$$

where $\log(listing\ price)$ is the natural logarithm of the listing price. Z is a vector of the observables that affect the listing price, and the same as that in Equation (10). In particular, the calendar information in Z is related to the time when the house is listed, but not the time when the house is sold, as we believe that the listing price relies on the listing time, rather than the sold time.

The residual, $\hat{\epsilon}_2 = \log(listing\ price) - \log(listing\backslash price)$, from this first stage regression measures how much the house is listed above its expected listing price, which we think affects the DOM, but not necessarily the sale price. We then control for $\hat{\epsilon}_2$ in the second stage regression of the DOM which is specified as follows:

$$DOM = \alpha_3 + \beta_3 * Cheaptalk + \gamma_3 * Z + \epsilon^2 + \epsilon_3 \quad (25)$$

where *DOM* is the days-on-market.

The estimated DOM, $\widehat{DOM} = E[DOM|Cheaptalk, Z, \epsilon^2]$, is then included in the third-stage regression, which is our previous sale price regression as specified in Equation (9). Column (2) of Table 4 reports the results of this 3SLS estimation. The impact of comparative cheap talk is estimated to be 0.79%, and significant at the 1% level.

5.3 Sample Selection Bias

The outcome variable–$log(sale\ price)$–is only observable for houses that are successfully sold, which may not be representative of the entire housing market. If the subsample of sold houses is not a random sample of the entire housing market, our previous estimators are likely to suffer from sample selection bias.

To correct this potential bias, we employ a bivariate sample selection model, which uses the full sample including both sold and unsold houses. The bivariate sample selection model comprises a selection equation and an outcome equation. The selection equation is:

$$Sold = \begin{cases} 1, & \text{if } Sold^{*} = \lambda' Z + \xi > 0 \\ 0, & \text{if } Sold^{*} = \lambda' Z + \xi \leq 0 \end{cases} \tag{26}$$

where *Sold* is the indicator variable of a house being sold. *Sold** is the corresponding latent variable. *Z* is a list of covariates that affects the probability of a house being sold, which we assume to be the same as in Equation (10).[4] ξ is the error term that follows a normal distribution.

The outcome equation is as follows. In particular, the outcome variable $\log(sale\ price)$ is observable only when $Sold = 1$.

$$log(sale\_price) = \begin{cases} \alpha + \beta\ Cheaptalk + \gamma' X + \epsilon & , \text{if } Sold = 1 \\ & , \text{if } Sold = 0 \end{cases} \tag{27}$$

We estimate the bivariate sample selection model, Equations (26) and (27), by using Heckman's two-step estimation, sometimes also called the Heckman estimator. The Heckman's two-step estimation relies on the assumption that

$$\epsilon = \upsilon \cdot \xi + \omega \tag{28}$$

$$\xi \sim \mathrm{N}(0,1) \tag{29}$$

where the random variable ω is independent of ξ. Heckman (1979) shows that, under Equations (28) and (29), the ordinary least square (OLS) estimation of the following model by using the subsample of sold houses is consistent and robust to sample selection bias.

$$log(sale\ price) = \alpha + \beta * Cheaptalk + \gamma' * \mathrm{X} + \upsilon * IMR(\hat{\lambda}' Z) + \epsilon \tag{30}$$

The only difference between Equations (9) and (30) is the additional covariate, $IMR(\hat{\lambda}' Z)$, that stands for the inverse Mills ratio, and calculated by using:

$$IMR(\hat{\lambda}' Z) = \frac{\phi(\hat{\lambda}' Z)}{\Phi(\hat{\lambda}' Z)} \tag{31}$$

where ϕ and Φ are the standard normal density and distribution functions, respectively. In addition, we have $\upsilon = COV\ (\xi, \epsilon)$. That is, the coefficient of

[4] The Heckman two-step estimator can be identified without exclusive restriction. In particular, the same regressors can appear in the first-step probit regression and second-step OLS regression, as long as the first-step probit model can well discriminate between sold and unsold homes. Indeed, in our first-step probit regression, there is a considerable range in the predicted probabilities of houses to be sold from 1.51% to 99.99%.

the inverse Mills ratio is the covariance of the error terms in the selection and the outcome equations.

Therefore, our estimation of Equation (30) in fact requires two steps: the first step is a probit regression of Equation (26) by using the full sample of both sold and unsold homes. The second step is an OLS regression of Equation (30), which uses only the subsample of sold homes and controlling for the inverse Mills ratio calculated from the first step.

The result from the Heckman's two-step estimation is reported in Column (3) of Table 4. Consistent with our main results, comparative cheap talk is associated with a 0.77% price premium.

5.4 Model Misspecification—Propensity Score Matching

All of the models we have used so far assume linear impacts of the covariates on the dependent variable. If this assumption is invalid, our previous estimators may be biased due to functional misspecification. To deal with this potential issue, we apply the matching method; more specifically, the propensity score matching (PSM) method.

The matching estimation is obtained by simply comparing outcomes among transactions that received the treatment (i.e., the treatment group) versus those that did not (i.e., the comparison group). Using terminology from the matching literature, we define the outcome as the natural logarithm of the sale price; the treatment group is defined as transactions where the listing agents used comparative cheap talk; the comparison group is defined as all of the other transactions.

One advantage of using matching estimation (compared to a regression) is that the key identifying assumption is weaker: the effect of covariates on the outcome need not be linear, as the matching method estimates the effect by matching homes with the same covariates instead of a linear model for the effect of the covariates. Matching, however, cannot solve for any unobservable variable bias. Similar to regressions, matching is based on the assumption that the source of selection bias is the set of observed covariates. Matching estimators are biased if the adoption of comparative cheap talk is based on unobservable variables.

Finding matches that are similar with respect to all of the relevant covariates can be difficult if there is a large number of covariates but the sample is relatively small. Nevertheless, Rosenbaum and Rubin (1983) show that matching the one-dimensional propensity score (which is the estimated probability of comparative cheap talk) suffices to adjust for the differences in the observed covariates. Matching on the propensity score is called PSM, which is the technique we use for the following estimation. The key estimator is called the average treatment effect on the treated (ATT), which has a similar interpretation as the coefficient in an OLS: they measure the difference in the sale price between transactions with comparative cheap talk and other transactions, with everything else being equal.

There are various matching algorithms that differ in how the matched single transactions are selected. In this paper, we focus on kernel matching.[5] As in Smith and Todd (2005), we implement the trimming method to determine the region of common support: we drop 10 percent of the treatment observations (i.e., transactions with comparative cheap talk) at which the propensity score density of the comparison observations (other transactions) is the lowest. The ATTs estimated through kernel matching are reported in Column 4 of Table 4. Consistent with our previous estimators, the PSM results show that comparative cheap talk is associated with a 0.81% price premium. Although the estimated ATT is statistically insignificant, it has the same sign and a similar size of the coefficient from the OLS regression (Column 6 of Table 3).

As Heckman et al. (1997), and Dehejia and Wahba (1999) note, the PSM estimator is only defined in the region of common support. Matching incomparable observations could cause evaluation bias. Hence, an important further step is to check the common support of the propensity scores for comparative cheap talk transactions and that for other transactions.

The most straightforward way to verify common support is a visual analysis of the density distributions of the propensity scores for both comparative cheap talk transactions and other transactions. Figure 2 shows the propensity score distribution for transactions with comparative cheap talk (above), and other transactions (below). Figure 2 provides strong evidence of overlapping propensity score distributions, thus suggesting that our PSM estimation is well identified and reliable.

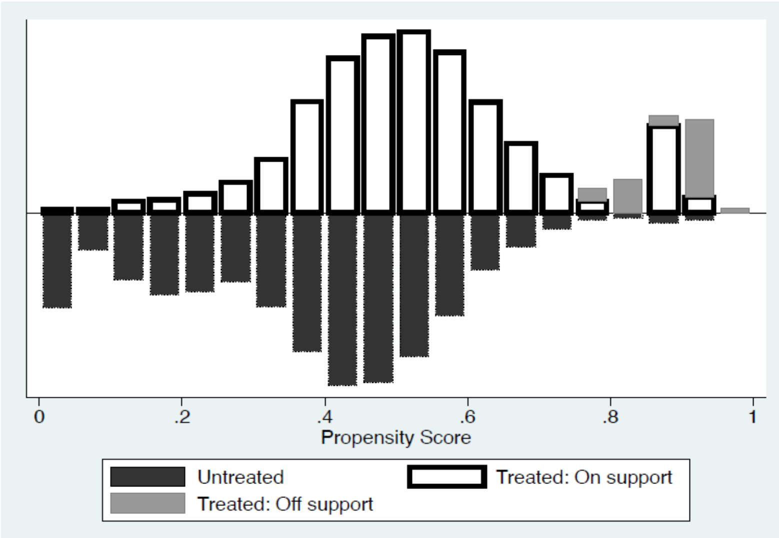
6. Explanations of Price Premium of Comparative Statements

The basic result that we have identified and tested in the two previous sections is consistent with the prediction of the comparative cheap talk model, but also other competing hypotheses. For instance, homes sold with comparative statements may have unobservable characteristics so that they can sell at high prices. However, this is unlikely a valid explanation, as negative comments are likely to represent undesirable characteristics, which will make homes sell for low (rather than high) prices.

---

[5] For the technical detail of each matching algorithm, see Imbens (2004), Smith and Todd (2005), and Caliendo and Kopeinig (2008).

Figure 2 Propensity Score Distributions of Comparative Cheap Talk (Treated) and Babbling Cheap Talk (Untreated)



Besides our basic result, we find additional evidence that is consistent with the comparative cheap talk explanation, but cannot be explained by unobservables. In particular, we find empirical evidence in support of our Proposition 2, that is, the price premium of comparative statements is higher when there are more buyers. In particular, the premium switches sign and become a discount when there are few potential buyers.

We find evidence along three dimensions. (i) the boom period (2001-2006) vs the bust period (2008-2010); (ii) listings with little competition vs those with much competition; and (iii) typical vs atypical houses. All of the results are reported in Table 6, and consistent with the theoretical prediction from the model, thus providing strong empirical support to the theoretical model of comparative cheap talk.

The first dimension is related to the time of the sale. We classify the full sample into the boom (2000-2006) and bust (2008-2010) periods. Compared to the former, the latter tends to have a seller's market, where each listing has more potential buyers. We re-estimate our full specification (as in Column 6 of Table 3) by using the two subsamples, respectively. The results are reported in Columns (1) and (2) of Table 6. Indeed, we find that the price impact of comparative statements is positive in a boom period, but negative in a bust period (1.27% vs -2.53%).

The second dimension that we test is the number of competing houses. For each house, competing houses are defined as other houses that are located in the same school district as the subject house and actively listed at the time when the subject house is sold. A house facing more competition from similar houses tends to have fewer potential buyers. We therefore classify our data into two exclusive and exhaustive subsamples according to the number of competing houses, and then re-estimate the price premium of the comparative statements in these two subsamples. The results are reported in Columns (3) and (4) of Table 6. As expected, the price impact is positive for houses with less competition but negative for houses with much competition (2.13% vs -0.51%), although the negative impact is statistically insignificant.

The third dimension is along the atypicality index of the house. Following Haurin (1988), we calculate the atypicality index of each house as follows, assuming that the house is located in zip code *j*:

$$I = \sum_{i}^{m} \hat{\gamma}_i |X_i - \bar{X}_{i,j}| \tag{32}$$

where *m* is the number of observable characteristics of the house, $| X_i - \overline{\mathrm{X}}_{ij} |$ is the deviation of the observed house characteristics i from the average level of all the houses in zip code j, and $\hat{\gamma}_{\mathrm{j}}$ is the estimated coefficient of characteristic X_i in the following model:

$$p = \sum_{i}^{m} \gamma_i X_i + \epsilon \tag{33}$$

where p is the sale price. Therefore, $\hat{\gamma}_{\mathrm{i}}$ can be interpreted as the implied marginal price of characteristic i. In sum, the atypicality index of a house is the weighted average deviation of the house characteristics from the average level in the zip code where the house is located. The weight of each characteristic is the implied marginal price of that characteristic.

The atypicality index is an aggregated measure of how much the house is different from a typical house in the zip code. A house with a high atypicality index is more likely to have unusual characteristics, and therefore tends to have fewer potential buyers. We thereby classify our full data into two exclusive and exhaustive subsamples according to the atypical index. We then re-estimate the price premium of comparative cheap talk in these two subsamples. The results are reported in Columns (5) and (6) in Table 6. Indeed, we find that the price premium of comparative cheap talk is larger for houses with a lower atypicality index than those with a high atypicality index (1.06% vs 0.50%).

Finally, if comparative statements are an effective marketing strategy to credibly reveal information to potential buyers, then we should expect shorter DOM for houses with comparative statements. We therefore rerun our main specification (as in Column 6 of Table 3), but replace sale price with DOM. The empirical result indeed shows that comparative statements are associated with a three-day reduction in time-on-market, assuming all other factors remain constant.

Table 6 Impact of Comparative Cheap Talk on Sale Price in Subsamples

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Subsample	2001-2006	2008-2010	Few competitions	Many competitions	Typical houses	Atypical houses
# of potential buyers	Many	Few	Many	Few	Many	Few
Cheaptalk	0.0127***	-0.0253**	0.0213***	-0.0051	0.0106**	0.0050
	(0.0046)	(0.0110)	(0.0068)	(0.0043)	(0.0056)	(0.0060)
Other covariates:						
House characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Effort of agent	Yes	Yes	Yes	Yes	Yes	Yes
Contract terms	Yes	Yes	Yes	Yes	Yes	Yes
Days-on-market	Yes	Yes	Yes	Yes	Yes	Yes
Fixed effects for	Yes	Yes	Yes	Yes	Yes	Yes
School district	Yes	Yes	Yes	Yes	Yes	Yes
Month & year of sale	Yes	Yes	Yes	Yes	Yes	Yes
Agent experience	Yes	Yes	Yes	Yes	Yes	Yes
Broker	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.844	0.852	0.862	0.821	0.738	0.886
N	10,350	2,422	7,086	7,096	7,086	7,096

Notes: Each column is from a separate regression with the use of different subsamples. * significant at the 10% level, ** significant at the 5% level, and *** significant at the 1% level.

7. Conclusion

In this paper, we apply the comparative cheap talk model in Chakraborty and Harbaugh (2010) in real estate and provide empirical evidence to show that comparative cheap talk exists in equilibrium and the impact of comparative cheaper talk increases when there are more potential buyers. This is the first study, to the best of our knowledge, to use empirical data to validate a theoretical cheap talk model. The results highlight the value of the comparative cheap talk model in a market where assets are heterogenous with various attributes that are hard to quantify, and transaction of the asset relies heavily on intermediation.

When describing the characteristics of a listed house, the comparative cheap talk model predicts that a listing agent with state-independent preferences can credibly reveal information to potential buyers, by making comparative statements that make the house more appealing in some areas but less so in other areas. Moreover, the listing agent strictly benefits from comparative statements, if his/her preferences over the estimates of buyers are quasiconvex. In an auction setting, this quasiconvex preference condition is satisfied if there are enough buyers. In other words, comparative statements increase expected house sale price when there are many (i.e., more than 3) potential buyers, but decrease the price when there are too few (less than 4) buyers. The reason is that comparative statements induce a better match of the house with the buyer who values it the most, but also weaken competition among potential buyers. When there are many potential buyers, the positive impact from a better match overweights the negative impact from less competition. When there are few buyers, the negative impact from reduced competition outweights the positive impact.

Housing transactions provide a natural experiment to test the comparative cheap talk model, where real estate products are heterogeneous, and agents play a significant role in the transactions. Using the MLS data, we identify the comparative statements that agents provide about a property in an MLS profile. Indeed, we find that all else equal, transactions where the listing agents have made comparative statements are associated with a 0.80% price premium, relative to the other transactions. The results are robust to the endogeneity of the decision to use comparative statements, endogeneity of DOM, sample selection bias, and model misspecification.

In addition, we test if the magnitude of the price impact of comparative statements is larger for houses with more potential buyers. In particular, is the impact positive when there are many buyers but switches sign and becomes negative when there are few buyers? We test this hypothesis along three dimensions. First, we classify the full sample into the boom period (2000-2006) and the bust period (2008-2010). Compared to the bust period, the boom period tends to be a seller 's market, where each listing has more potential buyers.

Indeed, we find that the price impact of comparative statements is positive in a boom period, but negative in a bust period.

The second dimension that we test is the number of competing houses. For each house, we define the competing houses as other houses that are located in the same school district as the subject house and actively listed at the time when the subject house is sold. A house that faces less competition from similar houses tends to have more potential buyers. We therefore classify our data into two exclusive and exhaustive subsamples according to the number of competing houses, and then re-estimate the price premium of comparative statements in these two subsamples. As expected, the price impact is positive for houses with less competition, but negative for houses with substantial competition.

The third dimension is about the atypicality index of the house. Following Haurin (1988), we calculate the atypicality index for each house, and classify our full data into two exclusive and exhaustive subsamples according to the atypical index. A house with a high atypicality index is likely to have more unusual characteristics, and therefore tends to have fewer potential buyers. Indeed, we find that the price impact of comparative statements is larger for houses with a lower atypicality index.

References

Agarwal, S., He, J., Sing, T.F. and Song, C. (2019). Do Real Estate Agents Have Information Advantages in Housing Markets? *Journal of Financial Economics*, 134(3). http://dx.doi.org/10.1016/j.jfineco.2019.05.008

Allen, M., Rutherford, J., Rutherford, R., and Yavas, A. (2016). *Conflicts of Interest in Residential Real Estate Transactions: New Evidence*. Florida Gulf Coast University, University of South Florida, and University of Wisconsin, Madison Unpublished working paper. Available at: https://www.aeaweb.org/conference/2016/retrieve.php?pdfid=13794&tk=eN2yGGba

Altonji, J., Elder, T., and Taber, C. (2005). Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools. *Journal of Political Economy,* 113(1), 151-184. https://doi.org/10.1086/426036

Altonji, J., Elder, T., and Taber, C. (2008). Using Selection on Observed Variables to Assess Bias from Unobservables When Evaluating Swan-Ganz Catheterization. *The American Economic Review,* 98(2), 345-350. https://www.jstor.org/stable/29730045

Bouvard, M., Chaigneau, P. and De Motta, A. (2015). Transparency in the Financial System: Rollover Risk and Crises. *Journal of Finance*, 70(4), 1805-1837. https://doi.org/10.1111/jofi.12270

Caliendo, M., and Kopeinig, S. (2008). Some Practical Guidance for the Implementation of Propensity Score Matching. *Journal of Economic Surveys*, 22(1), 31-72. https://doi.org/10.1111/j.1467-6419.2007.00527.x

Chakraborty, A. and Harbaugh, R. (2007). Comparative Cheap Talk. *Journal of Economic Theory,* 132(1), 70-94. https://doi.org/10.1016/j.jet.2005.07.003

Chakraborty, A. and Harbaugh, R. (2010). Persuasion by Cheap Talk. *American Economic Review,* 100(5), 2361-2382. DOI: 10.1257/aer.100.5.2361

Crawford, V. P. and Sobel, J. (1982). Strategic Information Transmission. *Econometrica*, 50(6), 1431-1451. https://doi.org/10.2307/1913390

Dehejia, R.H., and Wahba, S. (1999). Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs. *Journal of the American Statistical Association,* 94(448), 1053-1062. https://doi.org/10.2307/2669919

Garmaise, M. J., and Moskowitz, T.J. (2004). Confronting Information Asymmetries: Evidence from Real-Estate Markets. *Review of Financial Studies,* 17(2), 405-437. https://www.jstor.org/stable/1262704

Goodwin, K., Waller, B. and Weeks, HS. (2014). The Impact of Broker Vernacular in Residential Real Estate. *Journal of Housing Research,* 23(2), 143-161. https://doi.org/10.1080/10835547.2014.12092089

Goodwin, K., Waller, B. and Weeks, HS. (2018). Connotation and Textual Analysis in Real Estate Listings. *Journal of Housing Research,* 27(2), 93-106. http://dx.doi.org/10.1080/10835547.2018.12092149

Haag, J., Rutherford, R. and Thomson, T. (2000). Real Estate Agent Remarks: Help or Hype? *Journal of Real Estate Research,* 20(1), 205-215. http://dx.doi.org/10.1080/10835547.2000.12091024

Haurin, D. (1988). The Duration of Marketing Time of Residential Housing. *Real Estate Economics,* 16(4), 396-410. https://doi.org/10.1111/1540-6229.00463

Heckman, J.J., Ichimura, H. and Todd, P. (1997). Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme. *The Review of Economic Studies,* 64(4), 605-654. https://doi.org/10.2307/2971733

Heckman, J. J. (1979). Sample Selection as a Specification Error. *Econometrica*, 47(1), 153-161. DOI 10.3386/w0172

Imbens, G.W. (2004). Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review. *Review of Economics and Statistics,* 86(1), 4-29. https://doi.org/10.1162/003465304323023651

Inderst, R. and Ottaviani, M. (2009). Misselling through Agents. *American Economic Review,* 99(3), 883-908. DOI: 10.1257/aer.99.3.883

Kurlat, P, and Stroebel, J. (2015). Testing for Information Asymmetries in Real Estate Markets. *Review of Financial Studies*, 28(8), 2429-2461. https://doi.org/10.1093/rfs/hhv028

Levitt, S. and Syverson, C. (2008). Market Distortions When Agents are Better Informed: The Value of Information in Real Estate Transactions. *The Review of Economics and Statistics,* 90(4), 599-611. https://www.jstor.org/stable/40043103

Malenko, A. and Tsoy, A. (2019). Selling to Advised Buyers. *American Economic Review*, 109(4), 1323-48. DOI: 10.1257/aer.20170334

Menezes, F.M. and Monteiro, P.K. (2005). *An Introduction to Auction Theory*. Oxford University Press, Oxford, UK.

National Association of Realtors. (2016). *2016 PROFILE OF HOME BUYERS AND SELLERS.* Available at: https://www.nar.realtor/sites/default/files/reports/2016/2016-profile-of-home-buyers-and-sellers-10-31-2016.pdf

Ottaviani, M. and Sørensen, P.N. (2006). Reputational Cheap Talk, *The RAND Journal of Economics*, 37(1), 155-175.

Rosenbaum, P. R., and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41–55. https://doi.org/10.2307/2335942

Rutherford, R.C., Springer, T.M. and Yavas, A. (2005). Conflicts between Principles and Agents: Evidence from Residential Brokerage. *Journal of Financial Economics,* 76, 627-665. https://doi.org/10.1016/j.jfineco.2004.06.006

Sing, T.F. and Zou, Y. (2024). Improving Search Efficiency in Housing Markets Using Online Listings, Real Estate Economics, 52(4), 1020-1044.

Shen, L. and Ross, S. (2021). Information Value of Property Description: A Machine Learning Approach. *Journal of Urban Economics*, 121, 103299. https://doi.org/10.1016/j.jue.2020.103299

Smith, J. and Todd, P. (2005). Does Matching Overcome Lalonde's Critique of Nonexperimental Estimators. *Journal of Econometrics*, 125(1), 305-353.