Why Disclose *Less* Information?
Toward Resolving a Disclosure Puzzle in the Housing Market

Xun Bian
Associate Professor of Finance & Real Estate
Longwood University
bianx@longwood.edu

Justin C. Contat
Assistant Professor of Economics
Longwood University
contatjc@longwood.edu

Bennie D. Waller
Professor of Finance & Real Estate
Longwood University
wallerbd@longwood.edu

Scott A. Wentland*+
Research Economist
Bureau of Economic Analysis
scott.wentland@bea.gov

Last Updated: December 1, 2017

Abstract

We examine the role of information in the housing market, offering both theory and evidence for why we observe variability in information disclosure among property listings. Our initial empirical findings suggest a positive link between sellers disclosing more information about their properties and marketing outcomes (i.e. home prices and liquidity). While intuitive and consistent with the foundational literature on the economics of information disclosure, it raises an interesting puzzle: why do some agents disclose less information, or why do we not observe all properties disclosing the maximum amount of information possible? Analytically, we show that it may be optimal to disclose less information in some circumstances in this market, particularly when homes are more heterogeneous along certain dimensions. Empirically, the data support the prediction that less information disclosure yields positive returns for a sizable subsample of properties. We utilize a unique aspect of real estate microdata, where we can analyze the marketing strategies of real estate agents (when they market their own properties) as compared to their clients, finding that on average agents disclose less information than their clients, particularly when their homes are likely to be more horizontally differentiated. Finally, this study unearths a peculiar variation of information asymmetry stemming from this principal-agent relationship, where the principal can actually easily observe and monitor a dimension of agent effort (i.e. information disclosure), but the relevant asymmetry arises out of a divergence in the knowledge of the optimal marketing strategy for which the expert was hired and possibly the erroneous conflation of effort and information disclosure by clients.

Keywords: residential real estate, information disclosure, brokerage, agent, sale price, time on market, probability of sale, principal-agent, agency cost, asymmetric information, photos
JEL Codes: R30, J41, J31, D82, D83

*Contact author
+Disclaimer: Any views expressed here are those of the authors and not necessarily those of the Bureau of Economic Analysis or U.S. Department of Commerce.

Why Disclose *Less* Information?
Toward Resolving a Disclosure Puzzle in the Housing Market

1. Introduction

Despite the fact that, as an industry, real estate (and related activity) makes up approximately 13% of the U.S. economy and represents 47% of all private fixed assets,[1] the average homeowner is involved in a real estate transaction only a few times in his/her lifetime (for many, as few as once or twice). Unlike routine and repetitive transactions such as grocery shopping, real estate transactions are more apt to involve inexperienced buyers and sellers with relatively little knowledge of the market or the process. Hence, the typical homeowner is likely to seek out the expertise of a real estate professional (i.e. a broker or agent) to assist in the transaction and to devise a strategy to market the property. As part of the marketing effort, an agent will disclose information (descriptors, photos, etc.) about the property to the market (typically on a multiple listing service (MLS)). Using detailed micro-data from a medium-sized MLS, at first glance we observe an intuitive phenomenon: better information disclosure generally leads to better market outcomes (e.g. higher price, higher probability of sale, and higher liquidity[2]) for the homeowner. Yet, given the evident advantages of information disclosure, it remains a puzzle why a substantial part of the market provides *less* information to the market about their listings (e.g. few photos, few descriptors in the comments of the MLS listing). Why do *some* agents and sellers rationally choose to provide *less* information? Upon closer inspection and careful analysis we argue that this lack of information can be deliberate, and that only sellers of some types of homes find it profitable to adopt this low-information strategy.

In this paper we argue that partial information disclosure in the housing market is present only because homes are differentiated goods. In particular, homes are both *vertically* differentiated (i.e. some home characteristics nearly all consumers agree are better) and also *horizontally* differentiated (i.e. buyers express different tastes for

[1] Sources: U.S. Bureau of Economic Analysis, "GDP by Industry / VA, GO, II" (2016) and "Current-Cost Net Stock of Fixed Assets and Consumer Durable Goods" (2017).
[2] In this paper, we measure liquidity by time on market, measured in days.

different attributes). For goods that are only vertically differentiated, the microeconomics literature suggests that it may be a dominant strategy to reveal more information, particularly if that information shows the seller's product to be of a high quality. For example, it seems plausible that a home with located in a highly rated school district would want to reveal this to potential buyers. Additionally, omitting any information about the school district may lead potential buyers to infer that the home is located in a poorly rated school district.[3]

Broadly, this disclosure example is the logic of the "unraveling" literature,[4] where vertically differentiated sellers have information about their quality and choose how much information to disclose about this quality.[5] If sellers can costlessly disclose their quality, then they will find it in their own best interest to disclose their own quality. Any seller with quality higher than the average will want to disclose quality because not doing so will lead consumers to expect a lower (average) quality. This implies that in equilibrium consumers rationally infer that only sellers with the lowest possible quality do not disclose their quality. Thus, information in these markets tend to "unravel" as *all* sellers eventually reveal their quality to avoid being branded as the worse (even if they are below average).

Generally, the data refutes any complete "unraveling" of information in housing markets. In fact, we observe a distribution of information being revealed to the market that departs dramatically from the prediction of the unraveling disclosure literature. We posit that the real estate market does not "unravel" because homes are horizontally differentiated, and that this dimension may be particularly pronounced for certain sub-sections of the market. For example, consumers may not agree on the best style kitchen

---

[3] Carrillo, Riegg Cellini, and Green (2013) provide evidence of strategic information disclosure for schools, where sellers are more likely to omit information about the school district if the home is assigned to a lower quality school.

[4] See Grossman (1981), Grossman Hart (1980), and Milgrom (1981) for seminal contributions for the monopoly case, Stivers (2004) for the case of competition, and Okuno-Fujiwara et al (1990) for the case of oligopoly.

[5] Here, and throughout this paper, disclosure will mean that the seller can make any true statement about a property's quality, and the market largely prevents or deters them from telling outright falsehoods. While this may not be 100% true in any market, this definition of disclosure follows more closely with the theoretical disclosure literature we discuss below.

(e.g. a Tuscan-style vs. French Country-style kitchen), so disclosing the particular kitchen type of a given home may alienate some potential buyers. We build off of the theoretical work of Hotz and Xiao (2013), Sun (2011), and Celik (2014), who argue that partial disclosure of information in equilibrium can be optimal for some sellers who sell differentiated products. The surprising result of this literature is that it is sometimes the *high-quality* seller who chooses not to reveal information, while the low-quality seller will fully reveal information. Our results empirically document the existence of this phenomenon in the housing market.

Our main result shows that some agents deliberately choose to reveal less information (or, at least, they benefit from revealing less), and that this effect is significantly magnified for higher-end homes, a proxy for more heterogeneous (more horizontally differentiated) homes. We posit that sellers of higher priced homes may not want to reveal information because their potential buyers' preferences are likely to be more heterogeneous than the consumer base for lower-priced homes. As a result, omission of information of a higher-end can be rationally inferred as a signal that the home has a distinctive (preference-specific, not necessarily quality-related) characteristic from which buyers may shy away. In contrast, omission of information for a lower-priced home is a signal that the home is of low quality because the consumer base is less differentiated. Thus the degree of horizontal differentiation seems to be a key driver of our results, as predicted by our model and the literature more broadly.

The literature has suggested several explanations for why full disclosure might not be optimal for sellers. One common explanation is that information is not fully revealed because it is costly to do so[6]. If an agent must expend costly effort to disclose information, then in equilibrium the agent chooses the level of information that equalizes the marginal costs of effort with the marginal benefits of (expected) revenues, which is not necessarily the maximal amount of information. However in the real estate market, costly disclosure does not seem to be a reasonable assumption. Agents can very quickly and easily upload photos to a local MLS and enter in basic information about houses. Yet

[6] See for example Jovanovic (1982), Verrecchia (1983), Dye (1986), and Farrell (1985)

this is precisely the type of information that is typically not disclosed to its fullest extent. Other common explanations include limited rationality of buyers[7], search costs of buyers[8], imperfect disclosure technology[9], or even the amount of competition in the market[10]. We provide an alternate explanation consistent with recent theoretical literature, namely that the heterogeneity of the consumer base for the different sub-markets in our data is an important driving force of partial information disclosure.

Using a data set that initially consists of more than 300,000 listings from central Virginia MLS listings from 2001-2013, we find that, on average, more information provided by the listing agent on the MLS is associated with both a higher selling price and liquidity. Of course, this relationship may not be causal, as more photogenic properties may garner more photos, for example. However, we find the opposite results for more heterogeneous or higher-end properties, where *less* information is actually associated with higher selling prices and lower time on market. The bias in this case should only make the magnitude of this result more conservative. Our results are robust to numerous alternative specifications and distributional assumptions, as well as instrumenting for information disclosure.

As part of our identification strategy, like Levitt and Syverson (2008) we utilize differences in behavior for agents representing others and agents representing themselves (i.e. when the agent is also the owner of the property he or she is marketing). If more information always yields better market outcomes, then when agents market their own properties we would expect to see that more information is disclosed. Agents are the hired experts after all, so the more optimal strategy to market a home should be exemplified by an agent's decision to sell his or her own property on average. Consistent with a principal-agent conflict, we also find that agents provide significantly less information (e.g. providing few photos of the property; providing very few words describing the property in the remarks, etc.) for properties that they own themselves.

[7] For example Fishman Hagerty (2003) show that if some consumers don't understand the disclosure process, then disclosure will be too low. Hirshleifer, et al (2003) shows limited attention spans
[8] Anderson Renault (2006,2009) provide excellent examples.
[9] For example see Caldieraro et al (2011).
[10] See Levin et al (2009), Jansen (2015), and Guo Zhao (2009).

Further, we find that agents provide even less information for higher-end or more heterogeneous properties, which is consistent with the more optimal strategy for these types of properties.

In contrast with traditional moral hazard problems which emphasize imperfect monitoring of the agent, the issue in the housing market is that the owner of the home may not know the correct *strategy* to pursue in terms of how much and which type of information to disclose for a his or her property. After all, it is easy for a seller to check his or her own listing on the MLS (or a website like Zillow that often shows MLS listings) to see how much information is being disclosed. Expert agents may not be able to convince their clients, however, that the optimal strategy to market the home involves smaller amounts of information. As we speculate, the owner may mistakenly compel an agent to reveal more information than is optimal, given that providing less information may seem like shirking. Or, the rule of thumb that "more information is better" may put an agent in a quandary as they worry about unsatisfied customers damaging their reputation if lower information is perceived as shirking.

2. Related Literature and Contribution

Our paper contributes to a literature on information disclosure that started with the seminal contributions of Grossman (1981), Grossman Hart (1980), and Milgrom (1981). These papers consider a single seller with private information about the quality of the good he or she is selling. The main result of these papers is that if the seller can (costlessly) disclose verifiable information about the quality of the good, then there will be a full "unraveling" where each seller will reveal as much information as possible to distinguish his type from lower quality types. The low cost of revealing information in the real estate market conforms closely, though not exactly, to this assumption. Okuno-Fujiwara Postlewaite Suzumura (1990) generalize this result and provide sufficient conditions for full information revelation to take place, which they call "weakly-positive-monotone in beliefs". This condition holds if for any subset of types (in our setting qualities), a seller would prefer to be thought of as the highest type. The key assumption in this literature is that any information withheld from consumers is inferred to be the worst possible signal of quality. In the housing market, for example, this would imply that if the flooring of the house is not mentioned in the ad, then it is likely that the flooring is of poor quality. Stivers (2004) shows that the unraveling will continue to take place under competition between vertically differentiated sellers. Notably absent from this strand of the literature is horizontal differentiation, which we argue is a defining characteristic of the housing market, particularly for higher-end homes.

Several papers have been able to demonstrate that, under different conditions, partial equilibrium disclosure may be optimal for sellers in equilibrium, even with only vertical differentiation. In the case that the buyer also has private information, and that information is correlated with the seller's private information, then Cella (2008) shows that full information revelation might not take place. The seller's form of a contract, including the information structure, allows the buyer to make inferences about the quality of the seller. In equilibrium, the seller might not want this information to be known. However, the model cannot explain our results, namely that high quality sellers tend to disclose less information.

Our model of the housing market is most closely related to the theoretical models of Sun (2011), Celik (2014), and Hotz and Xiao (2013). These papers consider sellers whose goods are both vertically and horizontally differentiated. Under different conditions, these papers illustrate the counter-intuitive notion that only higher-quality firms would not want to reveal information about their products. We postpone a more detailed discussion of these models until the next section.

The closest known empirical work to ours is Allen, Dare, and Li (2016), who show that property listings with a larger number of views tend to sell more often, at higher prices, and stay on the market for a shorter period of time. Their model and empirical results suggest that every seller provides full information, yet they do not offer any explanation of why many sellers often provide less than full information. In contrast, we provide a justification for why some sellers, but not all sellers, provide full information: there are strategic concerns about which differentiated market in which the house will compete. A higher-end home, for example, does not want to reveal some of the details about its high quality kitchen because consumers willing to pay for high quality might not like its horizontal characteristics. In addition, a number of other papers have examined information disclosure in the housing market regarding disclosure of school districts (e.g. Carrillo, Cellini, and Green (2013)) and marketing strategies based on how agents use words in the comments section of an MLS listing (e.g. Nowak and Smith (2016); Goodwin, Waller, and Weeks (2014)). Thus, while some papers have examined the role of information in the housing market, we are the first to take the next step to investigate the sources of its variability and alternative marketing strategies that may provide a rationale for less information disclosure under particular circumstances. Central to our solution of this puzzle in the housing market is the fact that homes are horizontally differentiated, and the rationale for this flows from the theoretical literature on information disclosure.

3. **Modeling the Economics of Information Disclosure**

In this section we summarize some relevant and competing models of information disclosure before presenting our model. We discuss several other models to highlight the relevance of both horizontal differentiation and also partial disclosure in our model. We first present the model of Grossman (1981), who considers a monopolist selling a vertically, but not horizontally, differentiated good. Consumers do not know the quality and so the monopolist chooses how much information to disclose[11] about quality. We present the familiar unraveling argument more formally, whereby in there is full information disclosure. Next, we move onto Hotz and Xiao (2013), who consider a horizontally and vertically differentiated firm that also makes disclosure decisions about the vertical characteristic. Introducing the horizontal dimension drastically changes the results, as now the highest quality firms are the ones *less* likely to reveal information. Finally we present the models of Sun (2011) and Celik (2014), who also consider a vertically and differentiated firm, but this time the firm discloses its *horizontal* information. By contrasting the models within the context of a real estate market, we hope to illustrate the importance of horizontal differentiation, and also the surprising irrelevance of which type of information is disclosed. In the real estate setting, we believe

[11] Note that our use of disclosure entails both information revelation, and also that the discloser cannot tell a lie. The discloser simply chooses how coarsely to reveal true quality.

the information not being disclosed is mostly information about horizontal characteristics.

3.A. Canonical Model of Unraveling Literature

The seminal model of Grossman (1981) can be extended to include oligopoly settings[12] where firms compete by selling differentiated goods, though here we only consider a single seller and a single buyer. We adapt the model and use real estate terminology for ease of exposition.

Suppose there is a home for sale, but only the seller can observe its quality. Note that with a single consumer it is without loss of generality to consider only vertically differentiated homes since horizontal differentiation requires different willingness to pay for the same characteristic. There is no discrepancy between ranking homes with a single buyer. The seller has the ability to make costless and verifiable statements about the good's quality q, which is a value in the interval $[q_{min}, q_{max}]$. The consumer forms beliefs about q, and we denote these beliefs by r. Thus r_i is the probability that the home has quality q_i. The consumer maximizes expected utility, whose form is common knowledge between both the consumer and the monopolist.

To influence the buyer's beliefs r_i, the seller can make costless and verifiable statements about q. Since the seller cannot lie, the seller effectively is announcing a set of possible qualities, one of which must be the actual quality. Thus the seller essentially conveys that "my quality is a member of the set Q," and then announces the members of the set Q. By making this set larger, the seller deliberately obscures the true quality of the home from the consumer.

The unique equilibrium is easily characterized. The seller always truthfully reveals his quality q. In other words, the set Q is a singleton in equilibrium for each true quality q_i. Each consumer updates her beliefs perfectly using Bayes rule so that $r_i = 1$ if the true quality is q_i and $r_i = 0$ otherwise. Prices are chosen to make consumers indifferent between buying and not buying.

[12] See Stivers (2004) for a detailed treatment.

To see that this is an equilibrium, note that since consumers are willing to pay more for higher quality, a seller of the highest quality would always like to perfectly reveal his quality, since any other beliefs would necessarily be of lower quality. This lower quality would translate into lower prices and profits. Thus the highest quality seller will have perfect information disclosure. Consider a seller of the second highest quality. The consumer can never be convinced that he is of the highest quality, since they know the highest quality seller will reveal truthfully, i.e. they will observe $Q = \{q_{max}\}$. The second highest seller cannot lie, so the best quality they can hope for is their actual quality. Any other belief held by the consumer would generate lower average quality. Hence the second highest seller will fully disclose as well. The argument then repeats for all possible remaining types. Hence, once we acknowledge the highest type will disclose fully, then an "unraveling" occurs where all lower types do so as well.

The key feature of this equilibrium is that if the consumer does not observe information about the quality, then in equilibrium they rationally infer that the quality must be the lowest. By process of elimination, the seller of the lowest quality q_{min} need not reveal his quality as consumers will automatically and correctly infer that he is of the lowest quality. Thus in equilibrium the consumer forms skeptical beliefs about the omission of information: any information not disclosed is assumed to be the worst possible information. Hence if the seller does not reveal information about the flooring, it must be because the seller knows the quality of the flooring is the lowest quality possible.

Prediction 1*: Given the trivial cost of information revelation in the housing market, we expect more information disclosure to be associated with more optimal marketing strategies (i.e. higher home prices, lower time on market, higher probability of sale) for homes that are broadly more substitutable, i.e. less horizontally differentiated.*

We note here that, for the purposes of the model, it may make more sense to think of horizontal differentiation that is more binding. Lower-end homes may have plenty of scope for horizontal differentiation (e.g. old green shag carpet vs. old orange shag carpet in the basement); but, if these are more likely to be replaced (or ignored) by potential buyers anyway, then their tastes for orange versus green may not be particularly relevant, while the presence of the old shag carpet itself (i.e. vertical dimension) becomes the more binding choice. Generally, a lower-end home has more scope for a new buyer to customize it to his/her own tastes. So, while superficially there may be horizontal differentiation in lower-end homes, it is likely to matter less than, for example, a taste-specific new bamboo floor that is more likely found in a higher-end home.

As we will see, the unraveling result breaks down when consumers have heterogeneous preferences for housing characteristics. There will exist equilibria where rational consumers respond to omission of information in non-skeptical ways. Nonetheless we believe the Grossman (1981) model explains why lower priced homes disclose: they are less horizontally differentiated, conforming more closely to the assumptions of the Grossman (1981) model. Hence, in this submarket, omission of information is viewed with suspicion by potential buyers.

3.B Disclosure of Vertical Information for seller with vertical and horizontal differentiation (Hotz and Xiao (2013))

We now generalize the previous model by allowing the good to be both horizontally and vertically differentiated. Suppose there are two home-owners trying to sell their homes. Each hires a real estate agent to sell his home[13] on the housing market. To do so, the agent must disclose information about the house on a public directory (like the M.L.S.). Though the agent was hired for her expertise in how to market the property, the owner ultimately has the authority to choose which information goes into the advertisement. In the market we consider, we say that there is asymmetric information regarding marketing strategy because only the agent knows the optimal amount of information to disclose in the ad. Our paper departs from the traditional principal-agent framework in that we allow the owner to observe all information in the ad at all times. This is a reasonable assumption because the M.L.S. is a publicly observable website, and the information is usually reflected in other easy to use websites like Zillow and Trulia. Thus our focus is not that the agent chooses some amount of information and the owner does not get to observe the amount or quality or quantity of information, but rather that

[13] For ease of exposition, we arbitrarily assign pronouns, making the homeowners male and the agent female.

the agent has private knowledge about the optimal *strategy* of how to market the home. The owner does not know which amount (and what type) of information would be optimal to display in the ad. Additionally we abstract away from different preferences of the agent and owner and just assume that each cares about maximizing the revenues collecting from selling the house.[14] The novelty of the results is that even with costless disclosure, full information may not be optimal.

The demand side of the market for homes is heterogeneous. Consumers differ both in their willingness to pay for quality (vertical differentiation) and also their willingness to pay for particular housing characteristics (horizontal differentiation). For simplicity, each home has one horizontal characteristic and one vertical characteristic. The horizontal characteristic can be described by the Hotelling line with linear transport/dis-utility costs, where the sellers (*A* and *B*) are at different ends of the line. Suppose that consumers are located uniformly on the unit interval. More importantly, assume that consumers closer to one of the sellers, say seller *B*, are willing to pay more for quality. Thus there is correlation between the two dimensions of consumers' tastes. Let this degree of correlation be denoted as β.

Hotz and Xiao (2013) present such a model and find that seller B's expected revenues are strictly increasing in the expected difference in quality. In other words, if seller *B* is of high quality it would prefer to be seen as high quality. Additionally, if seller *B* is of high quality and if β is sufficiently large enough, then seller *A* would prefer to be viewed as a home with medium quality to be able to charge a higher price, *even if* it actually has a high quality product. Intuitively, if seller *A* revealed that its true quality was high, then it also means it alienates consumers along the horizontal dimension. The paper illustrates this by showing the elasticity of demand is larger when high quality information is revealed. In short, home A would find it in its own best interest to not reveal *any* information about quality if it faces competition in the high quality market. This leads us to our second prediction.

[14] A more complete treatment would allow consumers to explicitly trade off liquidity (time on market) with price.

Prediction 2*: We expect less information disclosure to be associated with more optimal marketing strategies (i.e. higher home prices and lower time on market) for homes that are more heterogeneous along horizontal dimensions.*

3.C Disclosure of Horizontal Information for Sellers with Vertical and Horizontal Differentiation – (Sun (2011) and Celik (2014))

Now we move to a class of models where disclosure of horizontal dimensions matter. As before the seller's product is vertically and horizontally differentiated, but now it is the *horizontal* information that is disclosed and becomes more relevant. First consider the model of Sun (2011). A seller with quality q is located on the Hotelling line with location $l \in [0,1]$. The joint distribution of q and l is arbitrary. Unlike Hotz and Xiao (2013), this paper's results do not depend upon the degree of correlation between the two dimensions, thus providing a new justification for partial disclosure with firms that are both vertically and horizontally differentiated. Consumers are uniformly distributed on the Hotelling line. In our context the Hotelling line represents the preferences for a particular housing characteristics (e.g. the style of a kitchen). If either q or l is unknown to the consumer, the seller has the option of either fully revealing (q,l) or revealing nothing at all. Thus partial disclosure of information is not permitted.

The paper has two relevant results for our model. First, they show (Propositions 3 and 5) that is quality q is known, then there exists a partially revealing equilibrium where only firms located in the middle of the Hotelling line reveal their location. Additionally, the set of the types of firms that disclose information *decreases* as quality q is increased. In other words, the higher quality the firm is, the *less* likely the firm is to disclose. Hence if consumers know the quality of a home, but are unsure of the particular horizontal characteristics, their model suggests that non-disclosure of information should be interpreted as a sign of unusual or uncommon horizontal characteristics. Next the paper shows that if on the other hand quality is not known, then we get exactly the opposite results: firms with higher qualities are *more* likely to reveal their horizontal information. This result suggests that if a buyer is unsure of the quality, then non-disclosure signals to the buyer that the home must be of a lower quality. We collect the results of this paper into Prediction 3:

Prediction 3: *Suppose that only full disclosure or no disclosure are possible. If the quality of a home is known, higher quality sellers are* less *likely to disclose information. If the quality of a home is unknown, higher quality sellers are* more *likely to disclose information.*

Finally, we consider the model of Celik (2014), which is largely an extension of Sun (2011) that allows for partial disclosure of horizontal information. The model is the same as that of Sun (2011), except that transportation costs of consumers are now quadratic. It is not clear how much the differences in results between Sun (2011) and Celik (2014) are due to the different functional form for transportation costs or because of the information structure.[15] However the main result of the paper is that if partial disclosure is permitted, then in equilibrium the only strategies that survive the "undefeated" criteria of Malilath et al. (1993) entail full information disclosure.

Contrasting this with Sun (2011) we see that the information technology available to our home seller matters for optimal information disclosure. Note that the differences between Predictions 2 and 3 may even allow us to determine the type of information disclosure (i.e. partial or full) present in our model. On the one hand, choosing whether or not to reveal a home has a pool seems to be a binary decision, suggesting that partial disclosure may not seem realistic. On the other hand, a pool is just one characteristic of a home. Choosing not to reveal information about the pool but revealing information about other characteristics suggests that partial disclosure is in fact taking place. Our results suggest the former intuition. Potential buyers are more sensitive to some characteristics but not others. We leave it to future work to explore this in more detail.

3.D A Model of Search and Information Disclosure in the Housing Market

We now present our own model that builds on the literature above. Let q represent the degree of vertical differentiation (e.g. quality) of the home and IQ the quantity of information the home discloses about its horizontal characteristic. Using the results of

[15] Consider the standard Hotelling model of product differentiation. With linear transportation costs the unique equilibrium has firms choosing *maximal* differentiation while with quadratic transportation costs the unique equilibrium has firms choosing *minimal* differentiation.

Sun (2011), we assume that the optimal amount of information, *IQ**, is decreasing in *q*. To embed this into a search model, we assume that the demand for a product can be encapsulated into the arrival rate of buyers in a search model we formally introduce shortly. The seller's aim is to attract as many potential buyers as possible. As the results of Sun (2011) suggest, this is done by choosing the correct amount of information disclosure, which is not necessarily maximal information. If *R* is the arrival rate, we note that this arrival rate of consumers depends upon the degree of vertical differentiation of the property *q* and also the information quantity *IQ*. Hence we write *R(IQ,q)* to emphasize this fact. The novelty here is that *R(IQ,q)* is not always increasing in *IQ*. The figure below illustrates the assumption.

[Figure 1 here]

Time is discrete and each period (day) a seller meets a buyer with probability *R(IQ,q)*. For simplicity we assume that at most one buyer can arrive in a given day. If a buyer arrives, the buyer makes an offer of *p*. If the seller accepts, the seller receives *p* and the game is over. If the seller rejects, he receives a flow payoff of *b*. Offers are drawn from an i.i.d. distribution *F* that depends upon the housing characteristics, so that this distribution may be very different for different sellers. Sellers cannot recall past offers, though this assumption can be relaxed. In equilibrium the seller's strategy will be to follow a cutoff strategy. Let *V(p)=p* be the (present value) of accepting an offer of *p*. Next, let *U* be the continuation payoff from declining an offer of *p*, which is necessarily the same for each *p*. Now, *U* can be written as

$$U = b + \beta[R(IQ,q)E[Max\{V(\cdot),U\}] + \big(1 - R(IQ,q)\big)U]$$

We can rewrite this as

$$U = \frac{b}{1-\beta(1-R)} + \frac{R\beta}{1-\beta(1-R)}\int_0^\infty \max\{V(x),U\}dF(x)$$

Analogously, the solution to the maximization problem is the familiar "reservation wage" solution from the labor search literature: the seller should accept offers above some threshold $\hat{P}$ and reject all other (lower) offers. Setting $V(\hat{P}) = U$, we get

$$\hat{P} = b + \frac{\beta R}{1-\beta}\int_{\hat{P}}^{\infty} 1 - F(x)dx.$$

This implicitly defines $\hat{P}$ in terms of the exogenous features of the model: $b, \beta, R,$ and *F*. A straightforward application of the implicit function theorem shows that $\frac{\partial \hat{P}}{\partial R} > 0$. So, agents choose information quantity *IQ*, which in turn determined the arrival rate *q*. Thus *R* is endogenously determined by *IQ*. If *R* were instead a choice variable, then all agents would set it equal to 1. Instead, in our model agents are trying to maximize *R* by choosing the correct level of *IQ*.

Applying the logic of the information disclosure literature we model *R* as a function of both *IQ* and *q*, where again *q* is the degree of vertical differentiation (i.e. quality). The main idea is that the optimal amount of information *IQ* (optimal in that it generates the largest *R*) may differ for different values of *q*. Recall that the optimal amount of information was decreasing in D, i.e. $IQ^*(q)$, is a decreasing function. Thus, a seller who wants to maximize the net present value of selling his home will choose the value of *IQ* that maximizes the arrival rate of offers.

Note that the stationarity of the reservation price immediately implies that the hazard rate is constant. The hazard rate is just $\lambda = R(1 - F(\hat{P})$, so that the probability the house stays on the market until time *t* is just $e^{-\lambda t}$ from elementary identities of the hazard function. As mentioned earlier, $\hat{P}$ is increasing in *R*. Viewing λ as the product of two terms, the first term *q* is clearly increasing in *R* while the second is clearly decreasing in *q*. Thus whether or not the hazard rate (and hence TOM) increases or decreases with information depends upon the shape of *F* (which is exogenously determined) and the derivative of $\hat{P}$ with respect to *IQ* (which may be positive or negative depending upon the sign of $\frac{ðR(IQ,D)}{ðIQ}$). This leads us to the main prediction of our paper.

Prediction 4*:* Sellers of homes of higher quality, more heterogeneous homes will choose to disclose less information, and as a result have shorter time on market.

As we will discuss more in section 4, we posit that higher-end homes tend to be more heterogeneous along horizontal dimensions, more consistent with the assumptions of our model as opposed to Grossman (1981).

3. Data

We use residential real estate data from an MLS located in central Virginia, including Richmond and parts of surrounding areas over the sample period of 2001-2013. Our initial sample consists of more than 300,000 listings in the residential real estate market between 2001 and 2013, making up the vast majority of the market for homes in this area. Among others, Levitt and Syverson (2008) point out that MLS data may be incorrect or incomplete because they are entered by real estate agents who represent the listing. As a result, the data were carefully examined. One pertinent aspect of the data was that up to 2006 there was relatively little variation in information quantity provided in listings, at least in our data set, because listings most commonly had only one photo (usually the front exterior of the home).[16] In our data, it was not until 2007 when MLS listings commonly included many pictures, which is why we then restrict our analysis to 2007-2013. After culling for incomplete, missing or illogical data that suggest data entry errors, the data set then consists of approximately 192,939 (sold and unsold) homes on the market, where 88,746 were sold.[17]

While information disclosure norms change over time (usually as information becomes cheaper and easier to disclose), we include time fixed effects in all of our regressions below to control for these types of time varying factors. By 2013, we observe more than 10% of properties listed with as many as 24 photos and more than 75% with at least 6 photos, for example. As we observe in Figure 2, the distribution of photos tends to be bimodal. For most years, the most frequent category in the histogram is near the max

[16] As of 2005, 95% of our listings had only one photo, where such little variation in the variable of interest necessitates dropping these years from the sample.

[17] We culled our data in accordance with much of the literature, dropping outliers beyond the 99th or 95th percentile (or natural cutoffs where appropriate) for key variables in our analysis. For transparency and replicability, we trimmed the data as follows: keeping listings with a list price greater than $39,950 or less than $1,000,000; age is greater than or equal to zero and less than 150 (note that Richmond is an older American city, so we limit homes to post-Civil War built homes); square footage is greater than 700 and less than 5,639; the number of bedrooms is greater than zero and less than 7; number of bathrooms is greater than or equal to 1 and less than 6.5; acreage is less than 40; and, time on market is less than 580 days (i.e. its 99th percentile).

number of photos, yet this often only encompasses a minority of listings. At most, 30-40% of the listings have near the maximum number of photos disclosed, while in some years this number is much smaller. The central question of this paper can be derived by observing Figure 2: why does a substantial portion of listings disclose *less* information?

Given that agents sometimes leave key fields blank in their listings, we were also able to merge tax data into a number of listings with missing fields in key areas (i.e. square footage, bedrooms, bathrooms, and age). Note that without this supplementation to the data, these observations would be dropped from our analysis because the control variables would be missing. Whether the information is there, in a binary sense, may be an important aspect of information quantity that we will explore in this paper; and, at the very least, retaining these observations will reduce potential bias insofar as the decision (or accident) to leave the field blank may be related to information quantity revealed in the listing.

The coverage of our MLS data represents a typical housing market that includes urban, suburban, and rural sales. Richmond is a medium-sized city located in the eastern part of central Virginia and the MLS covers much of the "Greater Richmond" area. The average property in our sample period has a selling and listing price of $231,711.20 and $250,247.70 respectively. The average listed property was 29 years of age, with 2,038 square feet, 3.5 bedrooms, and 2.3 bathrooms, and was on the market for 109 days. See Table I for additional descriptive statistics for other variables used in our analysis below.

[Table 1 here]

4. Methodology

4.A Baseline Hedonic Price Model Using OLS

We begin our analysis by using a standard hedonic OLS model or cross-sectional approach to estimate the association between information quantity and home prices, utilizing a traditional model that accounts for heterogeneous characteristics of both homes and their locations. We estimate the following linear regression:

$$SP_{ij} = \alpha + \beta\, IQ_i + \sum \gamma\, X_i + \varphi\, LOC_j + \varepsilon \qquad (1)$$

where SP_i is the sale price (logged) and IQ_i, the variable of interest, is a measure of disclosure or information quantity (see below) of listing *i*, X_i is a vector of property specific characteristics and listing year fixed effects,[18] LOC_i is a vector for location control (i.e. Census block group fixed effects in our baseline regressions), and ε, is an error term. In addition to clustering at the block group level, we use robust, heteroscedastic-consistent standard errors.[19] The purpose of this regression is to show the cross-sectional relationship between information revealed in an MLS listing and its sale price, controlling for the observable characteristics of the home and location; and, the direction in which it is correlated with marketing outcomes (like price and liquidity) will provide insight into the optimal marketing strategy.

Since information takes a variety of forms, we use a couple measures for the information quantity IQ_i. We begin with simply the number of photos and the number of words in the comments section. These are easily quantifiable measures that convey how much information is being diffused through the MLS listing. When users access the MLS, the primary way to obtain additional detail about the property (beyond the searchable fields of bedrooms, bathrooms, etc.) is through information communicated by photos or the comments/remarks section. However, this information may also be substitutable for one another. A realtor can communicate, for example, whether a home has an unfinished basement either by saying so in the comments or using a picture. The old adage of "a picture is worth a thousand words" implicitly describes some tradeoff between the two kinds of information. Hence, we then combine these variables to obtain an overall measure of information quantity in the following way:

- *Information Quantity (Public)* = (number of words in the public comments/remarks section) + (number of photos*10).
- *Information Quantity (All)* = Information Quantity (Public) + (number of words in the "agent-only" comments)

[18] The property characteristics we use as controls include: number of bedrooms, square footage, number of bathrooms, acreage, age, time on market (measured in days and logged), and dummy variables for whether the home: is a condo/townhouse, is vacant, was an owner-agent listing, a dual agency transaction, a dual brokerage transaction, or was a "quick sale," defined as selling in the first day it was on the market.
[19] Other geographical controls were explored (zip code, census tract, and census block). Overall, the results are not particularly sensitive to the choice of alternative spatial fixed effects.

- *Low Information Quantity (All)* = a dummy variable equaling one if the Information Quantity (All) variable is less than the 10th percentile of all (unculled) listings within a particular year.[20]
- *Key Field Left Blanks* = the number of key fields (square footage, bedrooms, bathrooms, year built (or age)) that was missing in the MLS listing.

The first measure of information quantity compiles the number of words that are publically visible in the comment/remarks section and the number of photos into an index, weighing each photo as 10 words. This weight is somewhat arbitrary, but was constructed such that it is in line with the ratio of the means of these variables. The average listing has 73 words written in the comments section and 8 photos per home (an average ratio of 9.125 words per photo, but we round up to 10 ease of interpreting the index). Our second measure incorporates a word count from the comments that only MLS member agents can view when they pull up a listing. These words are not available on the public listing, but can often provide potentially cooperating agents with important information about the listing.

Our third and fourth measures of information quantity are actually measuring *low* information quantity. The third variable is a dummy variable that is a measure of an agent providing low information quantity relative to other listings in a particular year. Regarding the fourth measure, this is an extreme case of low information: the case of missing information. The MLS does not require all fields to be filled in by the agent, but there is an expectation that certain fields are essential for actually finding a home on the MLS for buyers. Buyers (or their agents) may search based on the number of bedrooms or square footage, for example. When these values are missing, the listing may be more difficult to find on the MLS or less visible to potential buyers.

4.B. Home Liquidity and Parametric (Weibull) Hazard Models

We now turn to our analysis to liquidity, which we follow the literature in measuring as time on market. In order to estimate the effect of information quantity on liquidity, we use both sold and unsold properties within a parametric hazard model that

[20] Note that information disclosure norms and capacity change over time, so this variable is capturing whether the agent provided low information quantity relative to other listings within the particular year it was listed. The determination of the 10th percentile thresholds by year were determined by the full, unculled sample, which is why the summary statistics in Table 1 show the mean being less than 0.1.

assumes the baseline distribution follows a Weibull distribution (e.g. see Rutherford and Yavas (2012); Rutherford, Springer, and Yavas (2005); Bian, Waller, and Wentland (2016)). The key assumption of this specification is that the hazard rate, i.e. the probability the home sells at time *t* given that it has not yet sold, is decreasing with time[21]. In other words, a home that stays on the market is less likely to sell each additional day. We write the hazard rate as

$$\lambda(t;\, X_i, IQ) = \alpha t^{\alpha-1}\exp(X_i\beta + \gamma_1 IQ_i + \gamma_2 LP_i) \quad (2)$$

The first term in the hazard rate, $\alpha t^{\alpha-1}$, represents the baseline hazard that is common to all homes, where α is a parameter we estimate. The monotonicity of the Weibull hazard rate in our data set shows up here in the form of $0 < \alpha < 1$. The second term, $\exp(X_i\beta)$, represents home i's idiosyncratic differences away from the common hazard rate shared by all homes. Any differences between homes that affect the time on the market enter multiplicatively, and are the same at each point in time. For example, if a home is twice as likely to sell than another today, then tomorrow the same house will continue to be twice as likely to sell, provided both homes are still on the market. As with our other baseline hedonic specification, X_i represents housing characteristics and IQ_i represents the information quality associated with home i. Since we are allowing unsold homes, we use list price (LP_i) instead of sale price.

To make interpretation of the coefficients easier, we represent our results in the Accelerated Failure Time (AFT) metric, which essentially allows us to write the (natural log of) time on market in the form

$$\ln(TOM_i) = X_i\beta + \gamma_1 IQ_i + \gamma_2 LP_i + u_i, \quad (3)$$

where u_i is the appropriate error term. We have abused notation somewhat as the coefficients in this metric will differ from those given in the hazard rate specification, but we leave them unchanged for ease of exposition. Conventional wisdom suggests that $\gamma_1 < 0$, so that homes with more information sell faster. When considering the full

[21] Strictly speaking the Weibull distribution also allows for an increasing hazard rate. All that is required is monotonicity. However, as empirical demonstrated, we do see strong evidence of a decreasing hazard rate. For example, this could be due to buyer's perception of longer time on market as a signal of lower quality, as Chen and Rutherford (2012) point out.

sample, we do find this is the case. However, this only begs the question why so many homes decide to use less than maximal information. By stratifying the data along price lines, we later show that the relationship between information and liquidity, i.e. γ_1, differs for higher-priced and lower-priced homes. Specifically, only lower-priced homes experience benefits (in the form of lower time on market) from providing lots of information.

An advantage of using hazard estimation is that it avoids problems of sample selection bias and censoring. Hazard estimation allows us to consider both sold and unsold homes. Besides increasing the sample size, this also allows us to determine whether information can increase the probability of sale. To see this, note that if we restrict attention only to homes that sold, then by definition we are ruling out the possibility of information affecting the probability of sale as the probability is already one for these homes. Hazard estimation allows us to circumvent these problems.

Although many real estate researchers utilize a Weibull distribution of the hazard function, in untabulated tests we employ exponential and gamma distributions to explore the degree to which results are robust to the specified distribution. While untabulated, all hazard models estimated here are robust across these alternative distribution assumptions. The purpose of this analysis is to investigate the relation between information and a home's hazard (i.e. probability of sale) or the time it takes to sell.

4.C. Determinants of Information Quantity

Finally, we turn to investigating the determinants of information quantity itself (and the choice to provide low information quantity) to test whether agents (i.e. experts) market their properties in ways more consistent with our theoretical construct and whether they market their own properties differently than their clients. Accordingly, we set up a similar estimation framework as above, except that we use an OLS model and linear probability model to estimate the determinants of information quantity and the probability of providing low information quantity as shown the following two equations respectively:

$$IQ_i = \varphi(High_i, X_i, OA_i, LOC_i) + \varepsilon \tag{4}$$

$$LowIQ_i = \varphi(High_i, X_i, OA_i, LOC_i) + \varepsilon \tag{5}$$

where IQ_i and LowIQ_i are the Information Quantity and Low Information Quantity variables defined above, $High_i$ is a proxy for high-end homes where horizontal differentiation is likely to be more binding, OA_i is whether the property was listed by an agent who also owns the property (i.e. "owner-agent"), and the other X_i covariates are defined the same as above.

The $High_i$ proxy takes a couple forms, which we explore for robustness. First, we hypothesized that higher-end homes, which we define as listed in the upper quartile of list price, have a more binding heterogeneity along the horizontal dimensions we discussed in our model section above. That is, otherwise identical homes will be more heterogeneous in unobservable ways if its price is among the top quartile of homes once we control for the observable home and location characteristics. Essentially, this specification is comparing two identical homes in observable ways (i.e. the same number of bedrooms, bathrooms, square footage, etc.), but one is listed for a much higher price. Note that we also control for time on market in this regression, which should account for the seller's time preference and their aggressiveness in setting list price. The unobservables (like whether it has an updated Tuscan-style kitchen with high-end quartz countertops) are likely to be reflected by this proxy when while we are controlling for other home characteristics.

Given that this proxy is imperfect, we look at a related proxy that more specifically designed to measure relevant unobservables. Following Glower, Haurin, and Hendershott (1998), we use a measure that they call "percentage listing error," which is calculated as the ratio of the estimated residuals (from a hedonic list price equation analogous to equation 1 above, though using list price on the left-hand side instead) and the actual list price. In effect, this gives you a measure of how "overpriced" a home is relative to its hedonic expected value based on observables. Note again that we are controlling for time on market in the equations above. So, as Glower et al. (1998) explain, the degree to which this home is priced above its list price based on its estimated fundamentals may reflect *unobservable* aspects of the home like an interesting architectural style, etc. That is, this is a proxy of home heterogeneity along horizontal dimensions above and beyond the (primarily vertical) dimensions that are directly observable with our data. The extent to which these two proxies imperfectly capture the constructs we aim to measure should introduce measurement error, which should make our results somewhat conservative (biasing them toward zero), provided that this error is not correlated with our construct of interest in some counterintuitive way.

Following Levitt and Syverson (2008), we use a similar strategy for identifying the "enlightened" strategy for listing a property by comparing owner-agents to their clients' properties and to determine the nature of the principal-agent problem in this market. The initial estimate on the owner-agent dummy will indicate whether they provide more or less information about their listing than comparable client properties. If agents provide more information relative to comparable client properties, then this may be initial evidence of the Grossman (1981) result that providing more information is the dominant strategy, and that they may be underproviding effort on behalf of their clients (i.e. a more traditional principal-agent problem). However, the opposite result would indicate a different story, more consistent with the framework of our model. If our model and the more recent theoretical literature make predictions that are consistent with this market, we should observe that agents pursuing a more "enlightened" strategy will provide less information for more horizontally differentiated properties, and the information asymmetry responsible for this difference between agents and clients may actually be the knowledge of the more optimal strategy. Indeed, if the typical client is unaware of the possibility that "less is more" for marketing of some properties, and they associated information quantity with effort, then we speculate that we may be observing agents providing *too much* information for some client homes.

5. Empirical Results

5.A. Information Quantity and Sale Price

We first estimate the naïve association between information and marketing outcomes of homes listed on a local MLS. It is clear from the results in Table 2 that higher information quantity provided by the listing agent is associated with a higher sale price. After controlling for property characteristics, time and spatial fixed effects, Column 1 shows that as more photos and more description of the home are provided in the listing, the home generally sells for more, on average. We caution at the outset that these results are not yet identified and a causal interpretation is too strong, an issue we discuss later in the paper. The magnitudes are likely too large because an obvious omitted variable is how photogenic (or otherwise high quality) the home is, which should bias our coefficients upward here. However, it is consistent with the structural model from Grossman (1981) and basic intuition that more information is associated with a more optimal marketing strategy for most homes, on average. Each additional photo, for example, is associated with approximately 1 percentage point higher sale price. The magnitude on the coefficient estimate for additional description in the comments, however, is not quite as large, as an extra 100 words is only associated with a 0.3 percentage point increase in sale price. This is qualitatively consistent with the adage that "a picture is worth 1,000 words," although perhaps not quite as dramatic in magnitude.

Columns 2-5 in Table 2 use alternate measure of information quantity, given that pictures and words may be (imperfect) substitutes for one another and what really matters is the overall level of information provided. When combined in the first information quantity measure or index, where a picture is "worth" ten words, an increase in information quantity is still associated with a higher sale price. While the magnitude is slightly different (given that the measure is weighted differently), this is qualitatively consistent with the measure in column 2 (where a picture is "worth" 10 words) and the third measure in column 3 (which includes the same measure as the first, but summed with comments that only other agents within the MLS can read). A one standard deviation (69 – or approximately 7 pictures) increase in the third measure of information quantity is associated with a 3.5% increase in the sale price of the average home. Finally, when an agent leaves a key field (i.e. bedrooms, bathrooms, sq. ft., and year built) blank when they list a property on the MLS, this is associated with a markedly lower sale price (approximately 21%). Alternatively, if an agent instead provides relatively low information (i.e. within lowest 10th percentile within the year it was listed), then we observe a 4.6% lower sale price for these properties.

[Table 2 here]

5.B. Information Quantity and Home Liquidity

Using a hazard model framework following a Weibull distribution, Tables 3 and 4 show that higher information is associated with greater home liquidity (i.e. lower time on market and higher probability of sale). The key difference between Tables 3 and 4 is that the coefficient estimates from Table 3 are in accelerated failure-time form (corresponding to time on market), while the estimates in Table 4 are in log relative-hazard form (corresponding to the hazard rate, or probability of sale). Using nearly the same covariates as Table 3 (see above), column 1 in Table 4 shows that an additional photo reduces marketing duration by approximately 2.5% on average, which is just less than three days. Descriptive content in the comments has no statistically significant impact on liquidity, however the joint measures of information quantity (from columns 2 and 3) show a consistently negative relation with time on market. The hazard result from column 4 shows that a one standard deviation (69) increase in information quantity (3) is associated with approximately 10% decrease in a home's time on market (about 11 days). As with the price regression, whether a home has key missing information has a significant relation to time on market. Overall, the coefficient estimates for columns 1-5 in Table 4 correspond to the 'other side of the coin' so-to-speak, showing that information quantity consistently increases the probability of sale at a given point in time.

[Table 3 here]

[Table 4 here]

5.C. Information Disclosure for Different Types of Homes: Initial Evidence

It is clear from both the price regressions and liquidity hazard models that a higher quantity information is associated with better outcomes for both sellers *and* the agent representing the listing. Both the sellers and the agent want to sell the home quickly and at a high price. Conversely, given these initial results, it is not clear why agents would supply a *lower* level of information, as their incentives seem to be aligned. However, as we noted above, we do find a substantial number of properties marketed with low information (e.g. one photo and few words in the description) and quite a bit of variation in information quantity across property listings. In fact, it is tempting to say that the least competent agents are likely the ones providing low information quantity, and thus driving this relationship. However, in an untabulated test we control for agent fixed effects (i.e. obtain a "within agent" estimator), and the principal findings above still stand. Given that information disclosure is not very costly in this market, it is unlikely that this is purely shirking, especially since it is easily verifiable by the sellers. So, the results from the first few tables pose an interesting puzzle: given that more information appears to be unambiguously good for both agents and their clients, why might some disclose more information and others disclose less? Why do we not observe a Grossman (1981) unraveling result in the housing market where virtually all properties disclose the maximum amount of information? Or, is there evidence that, in some circumstances, providing less information is actually "more," consistent with the recent theoretical literature and our model above? And, is there a principal-agent problem here that might explain some of the distribution of information in the housing market?

[Table 5 here]

We turn to Table 5 to answer the question about whether there are circumstances where lower information might be associated with a better market result for the seller (namely, a higher price and shorter time on market). Column 1 in Table 5 shows that when agents provide low information (i.e. lowest decile of information within a given year), there is a large negative effect on sale price for homes listed in the bottom three quartiles of list price (with a list price less than $299,950). However, when an agent provides low information for homes with a high list price, we find that the effect for those homes is the *opposite* result, qualitatively. Column 2 shows the coefficient estimates when the sample is stratified in the highest quartile of list price, indicating that these homes sell for approximately 2% more than comparable homes with more information provided. The final two columns in Table 5 shows the upper quartile stratification with a continuous measure of information quantity. The third column indicates the standard positive relation between information quantity and sale price for homes in the lowest three quartiles of list price, while the final column indicates a negative relation between information quantity and sale price for the top subset of homes (albeit statistically significant with only 90% confidence). Table 5 provides evidence that, for a subset of homes, a strategy of *lower* information disclosure appears to be beneficial for sellers. These results are consistent with Predictions 1-4, as it seems reasonable to assume that higher-end properties are more horizontally differentiated, where tastes of potential buyers may be more binding and influential over the arrival rate of buyers.

5.D. Endogeneity of Information Disclosure

The results from Table 5 may also reveal additional information about causality. Note that we cautioned against a causal interpretation of the initial results, because one could speculate that we may actually have reverse causality. Better homes may invite more information to be disclosed. While we control for a number of home characteristics, it is still possible that homes with unobserved high quality dimensions may be "more photogenic," which is why we observe this positive association (analogously, the fashion industry takes many pictures of photogenic models, but it is not the quantity of pictures that make the models more beautiful or marketable). However, this negative relationship between information and home price for higher-end homes runs counter to this reverse causality interpretation. If anything, we would expect these homes to be more photogenic, and despite that, additional information results in a *lower* sale price. When a particular cross-sectional test yields a result consistent with theoretical prediction, but counter to the expected econometric bias, we interpret this as at least some evidence that our estimates are properly identified (at least qualitatively, where the sign of the coefficient is likely to be correct). Or to put this more concretely in terms of the econometrics, if the econometric bias is in the upward direction (given that the omitted "photogenic-ness" is positively correlated with sale price), and our key results are still negative in these specifications, then these particular results are likely to be conservative.

The results from Table 6 tell a similar story, but with time on market as the dependent variable and estimated within a hazard model like Table 3. Columns one and three in Table 6 show that low information is associated with longer time on market, and more information is associated with shorter time on market within the first three quartiles of list price. Yet, columns two and four show the opposite result for homes in the upper tier. For higher end homes in the upper quartile of list price, lower information is actually better for a seller's time on market. Specifically, column two shows that homes in the upper tier sell a few weeks sooner when they provided low information relative to otherwise comparable homes. Finally, while untabulated, we also look at the effect of leaving fields blank across these stratifications, and we see no subsample where this strategy makes sense in the regressions in both Tables 5 and 6. Given that it is an extreme variation of information disclosure (i.e. not disclosing something at all), this is likely associated with a mistake or random error by the agent, which unintentionally registers as an unambiguously negative signal to the market. This serves as initial evidence that an extreme strategy of zero information along some dimension is not optimal, which may help explain why this phenomenon is somewhat rare.

We instrument for this potential bias in our Appendix Table A, where we estimate two-stage models of Table 5 with an instrumental variables (IV) model. In an ideal empirical design, we could randomly assign a level of information to be disclosed for a particular home and measure its impact. Absent a randomized control trial (RCT) in the real estate market, we turn to quasi-random variation in the data that could plausibly impact the decision to disclose a given amount of information, but may be unrelated to the property's unobservables in a particular way. We posit that a brokerage office could be a source of this kind of variation.

Like any workplace, real estate brokerages develop workplace cultures and conventions. Often they will use a similar stock listing contract, the same software for generating comparable home prices, and a variety of other day-to-day routines that stem out of firm policy or culture. Office norms do not stop at information disclosure, as the quantity of information across agents within the same brokerage is relatively high. We use the average level of information quantity in a given brokerage office in a given year as an instrument for our information quantity variables, noting that it is unlikely to be correlated with how photogenic a particular house may be (controlling for other observable factors),[22] but will be correlated with information disclosure of a given property. Our results in Table A are consistent with the corresponding regression in Table 5, providing at least initial evidence that this particular bias is not driving the results from Table 5.

Overall, Tables 5 and 6 suggests that different types of houses may call for a different information disclosure strategy. We have just identified one instance, but there is likely to be many more given that distribution of information quantity we see in the market and this may only explain some of this variation. More tellingly, expert behavior (particularly when the principal acts as his or her own agent) may help us better

[22] Controlling for observables here serves an additional role of dampening the likelihood of selection bias that would invalidate the instrument. To a large extent, factors like income are going to be most acutely associated with the size of the house and its location, for which we are controlling. Essentially, we are positing that the horizontal preferences are unlikely to self-select in this way (where sellers with updated Tuscan-style kitchen are not systematically choosing a broker differently than French-country style kitchen owners in the same neighborhood and with the same size home – it is not clear that brokerage offices specialize in *horizontally* relevant dimensions). If that is reasonable, then the brokerage office's influence on an individual listing's disclosure decision becomes the source of quasi-random variation.

understand the optimal strategy and move beyond speculation, which is why we turn to the decision to disclose information in the context of owner-agents and their clients.

5.E. Information Disclosure and Owner-agent Behavior – Lessons from Experts

Table 7 displays the results for when we model the decision to disclose information in the real estate market. Specifically, we estimate equation 4 in columns [1] and [3] and equation 5 in columns [2] and [4]. We find that agents actually disclose less information, on average, for their own homes relative to comparable client homes; or, they are more likely to provide low information quantity on average. This difference provides initial evidence that there may be a departure between the strategy of the owner-agent and a client property. That is, agents may be providing *too much* information than is optimal on their clients' behalf, which is evidence of the aforementioned counterintuitive principal-agent problem. Given the framework from our model, it is easy to see why, in some circumstances, it may be more optimal to provide less information. Indeed, the coefficient on the owner-agent may not be sufficient evidence of a principal-agent problem a la Levitt and Syverson (2008) and others. One alternative explanation of this result might be that agents themselves tend to buy homes that are more heterogeneous or unique along horizontal dimensions, and given the predictions from our model and recent literature, it is correlated with information quantity accordingly. However, this story appears to go against most anecdotes of agents being more concerned with resale value and what we observe in our data as agents selling their own properties quite frequently. Thus, we turn to an interacted effect under circumstances where we have higher confidence in their property's heterogeneity.

For homes priced by the seller in the upper quartile (i.e. above $299,950 in our sample period) or in the upper quartile of price listing error (PLE), which we assume have a higher degree of heterogeneity and more consistent with horizontal differentiation, we find a larger departure from owner-agents and comparable client properties with respect to how much information they disclose. Specifically, columns 1 and 3 in Table 7 show that client homes within the upper quartile of these proxies are listed with more information, while the interacted effect with owner-agents indicates significantly lower information provided. The results in columns 2 and 4, with binary measures of low information quantity, are somewhat weaker, but show the result that owner-agents are more likely to provide low information for their own properties than their clients'. Overall, the results from Table 7 are consistent with our model's predictions that it may be optimal to provide less information for certain types of homes, a result we can glean from how we observe experts marketing their own homes presumably using a more enlightened strategy (as evidenced by Table 6) when their incentives are perfectly aligned. While our proxy for hetereogeneity or horizontal differentiability is one of many types of potential proxies consistent with our model and the recent disclosure literature, it does provide initial evidence that "less is sometimes more" with respect to information disclosure and sheds light on why we do not see uniform information unraveling in this market (a la Grossman (1981)).

This particular principal-agent problem may persist in the housing market, given the inherent asymmetric information between housing market experts (agents) and the sellers regarding the optimal *strategy*. Information content in the listing is something that is easily verifiable by the sellers contracting with a particular agent. Ultimately, less information quantity may appear as if the agent is shirking, even in cases where less information is optimal. Some agents may be able to communicate the optimal (lower information) strategy to their clients when it is appropriate (and given the distribution of information we see in this market, it is clear that some in fact do); but, given the departure in the agent's own listings and their clients', it is clear that a gap may persist if a large enough contingent of sellers are not familiar with alternative information disclosure strategies and believe it is always optimal to disclose more information. Based on these results, we speculate that the distribution of information provided in listings in the real estate market can be explained, at least in part, by the fact that some homes are sufficiently heterogeneous along horizontal dimensions such that it would make sense to provide less information, and that agents are able to convince at least some of these sellers to provide less information (and can certainly convince themselves). We explore limited dimensions of home heterogeneity in this paper, but we leave additional exploration for future research. Further, future research may be able to shed light more light on *how* less information is better, exploring possible mechanisms in greater depth. For example, does information about a preference-specific attribute (say, a Tuscan-style kitchen) in isolation turn away buyers, but when buyers actually visit the home for themselves they are able to discover how the preference-specific attribute fits in with the broader themes or designs of the house (say, a complementary hardwood floor or architectural features that accentuate the Tuscan-style kitchen that are difficult to photograph effectively together).

6. Conclusion

While information often benefits sellers by revealing to the buyer that their product is not of the worst quality, the information disclosure literature suggests that there may be some circumstances when sellers may not want to reveal as much information. We empirically document this possibility in the market for real estate, where detailed microdata about very heterogeneous "products" (i.e. properties) allows us to observe how information can play a role in thousands of actual market outcomes. Our initial results are consistent with the more straightforward hypothesis that "more information is better," in that more information is associated with higher sale prices and lower time on market for otherwise observationally comparable homes. Yet, these results highlight the puzzle implicit in the Grossman (1981) model, which we attempt to shed light on in this paper: if information is so good, why do we observe so much variation in it? Why do some sellers disclose less than the maximum amount of information about their properties?

We document instances where properties tend to do better on the market (from the seller's perspective) when they provide *lower* amounts of information. Namely, we found that homes whose list price was in the top quartile of properties in this market tended to benefit from lower information in the form of a higher selling price and a shorter time on market. Higher-end homes tend to be more heterogeneous and more consistent with the assumptions of our model and recent literature in information disclosure theory, where revealing less information may be a more optimal disclosure strategy. We also find that owner-agents tend to disclose less information about their own homes than their clients; and, when those homes are higher-end, they disclose even less information, consistent with the optimal strategy we outline in our model.

Given the evidence that experts (owner-agents) market their own high-end properties with less information, we can rule out a traditional principal-agent problem where the lack of information generated in this market is simply shirking by agents. In fact, information in the MLS listing is one of the easiest activities to monitor from the standpoint of the seller. Indeed, our model and results suggest that the real information asymmetry lies in the knowledge of the optimal strategy to pursue. It may be difficult for an agent to make the case to their clients that "less is more" in all circumstances that demand it. As a result, some agents may have limited success in convincing their clients they are not shirking or "just being lazy" when the listing contains fewer pictures than the sellers expected. Thus, the market may benefit more broadly in the knowledge that, under some circumstances, "less can be more" when deciding how much information to disclose in real estate markets, consistent with an important theoretical result in economics.

References

Allen, Marcus T. and William Dare and Lingxiao Li. 2016. MLS Information Sharing Intensity and Housing Market Outcomes. *Journal of Real Estate Finance and Economics*. https://doi.org/10.1007/s11146-017-9612-5

Anderson, Simon and Regis Renault. 2006. Advertising Content. *American Economic Review*. 96:1, 93-113.

Anderson, Simon and Regis Renault. 2009. Comparative Advertising: disclosing horizontal match information. *RAND Journal of Economics.* 40:3, 558-581.

Bian, X., B. D. Waller, and S. A. Wentland. 2016. The Role of Transaction Costs in Impeding Market Exchange in Real Estate. *Journal of Housing Research*. 25(2), 115-135.

Caldieraro, Fabio and Dongsoo Shin and Andew Stivers. 2011. Voluntary Quality Disclosure under Price-Signaling Competition. *Managerial and Decision Economics.* 32, 493-504.

Carrillo, P., Cellini, S.R. and Green, R.K., 2013. School quality and information disclosure: Evidence from the housing market. *Economic Inquiry*, *51*(3), pp.1809-1828.

Celik, Levent, 2014, Information Unraveling Revisited: Disclosure of Horizontal Attributes. *The Journal of Industrial Economics*, 62:1, 113-136.

Cella, Michela. 2008. Informed Principle with Correlation. *Games and Economic Behavior*. 64:2, 433-456.

Chen, Jun and Ronald Rutherford. 2012. "Quality and Time on Market, an Empirical Test of Taylor's Theory." *Journal of Real Estate Finance and Economics.* 44:3, 414-428.

Dye, Ronald A. 1986. Proprietary and Nonproprietary Disclosures. *Journal of Business*. 59:2 331-366.

Farrell, Joseph. 1985. Voluntary Disclosure: Robustness of the Unraveling Result, and Comments on Its Importance. *MIT Department of Economics Working Paper* #374.

Fishman, Michael and Kathleen Hagerty. 2003. Mandatory versus Voluntary Disclosure in Markets with informed and uninformed customers. *The Journal of Law, Economics, & Organization*. 19:1, 45-63.

Glower, M., Haurin, D.R. and Hendershott, P.H., 1998. Selling time and selling price: The influence of seller motivation. *Real Estate Economics*, *26*(4), pp.719-740.

Goodwin, K., Waller, B. and Weeks, H.S., 2014. The impact of broker vernacular in residential real estate. *Journal of Housing Research*, *23*(2):143-161.

Grossman, S.J. 1981. The Informational Role of Warranties and Private Disclosure about Product Quality. *Journal of Law and Economics*. 24, 461-483.

Grossman, S.J. and O.D. Hart. 1980. Disclosure Laws and Takeover Bids. *Journal of Finance*. 35:2. 323-334.

Guo, Liang and Ying Zhao. 2009. Voluntary Quality Disclosure and Market Interaction. *Marketing Science*. 28:3, 488-501.

Hirschleifer, David and Siew Hong Teoh, 2003. Limited Attention, Information Disclosure, and Financial Reporting. *Journal of Accounting and Economics*. 36:1-3, 337-386.

Hotz, V.J. and Xiao, M., 2013. Strategic information disclosure: The case of multiattribute products with heterogeneous consumers. *Economic Inquiry*, 51(1), pp.865-881.

Jansen, Jos. 2015. A Note of Quality Disclosure and Competition. *Aarhus University Working Paper Series #2015-15.*

Jovanovic, B., 1982. Truthful disclosure of information. *The Bell Journal of Economics*, pp.36-44.

Levin, Dan and James Peck and Lixin Ye. 2009. Quality Disclosure and Competition. *Journal of Industrial Economics*. 57:1, 167-196.

Levitt, Steven D. and Chad Syverson, 2008. "Market Distortions When Agents Are Better Informed: The Value Of Information In Real Estate Transactions," *The Review of Economics and Statistics*, 90, 599-611.

Mailath, G. and M. Okuno-Fujiwara and A. Postlewaite. 1993, Belief-Based Refinements in Signaling Games. *Journal of Economic Theory.* 60, 241-276.

Milgrom, Paul. 1981. Good News and Bas News: Representation Theorems and Applications. *Bell Journal of Economics*. 12. 380-391.

Nowak, A. and Smith, P., 2017. Textual Analysis in Real Estate. *Journal of Applied Econometrics* 32:4, 896-918.

Okuno-Fujiwara, Mashiro and Andrew Postlewaite and Kotaro Suzumura. 1990. Strategic Information Revelation. *The Review of Economic Studies*. 57:1. 25-47.

Rutherford, R.C., T.M. Springer, and A. Yavas. 2005. Conflicts Between Principals and Agents: Evidence from Residential Brokerage. *Journal of Financial Economics*, 76:3, 627–- 65.

Rutherford, R. and Yavas, A. 2012. Discount Brokerage in Residential Real Estate Markets. *Real Estate Economics*, 40: 508–535.

Stivers, Andrew E. 2004. Unraveling of Information: Competition and Uncertainty. *Topics in Theoretical Economics*. 4:1, 1-18.

Sun, Monic. 2011. Disclosing Multiple Product Attributes. *Journal of Economics & Management Strategy*, 20:1, 195-224.

U.S. Bureau of Economic Analysis. 2016. "Gross Domestic Product (GDP) by Industry Data. https://www.bea.gov/industry/gdpbyind_data.htm

U.S. Bureau of Economic Analysis. 2017. "Current Cost Net Stock of Fixed Assets and Consumers Durable Goods." https"//www.bea.gov/National/FAweb/AllFATables.asp

Verrecchia, Robert E. 1983. Discretionary Disclosure. *Journal of Accounting and Economics*. 5, 179-194.

Figures

Figure 1

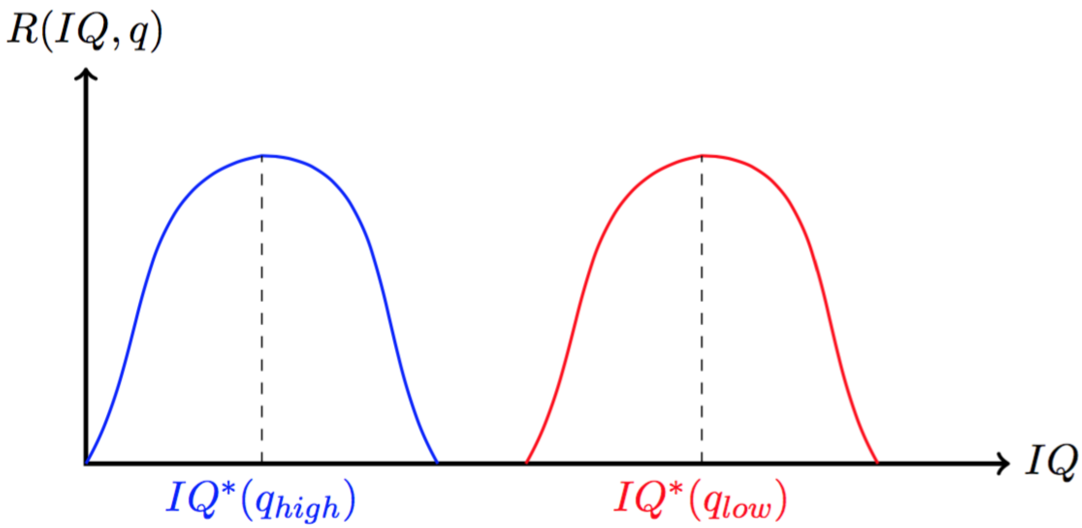
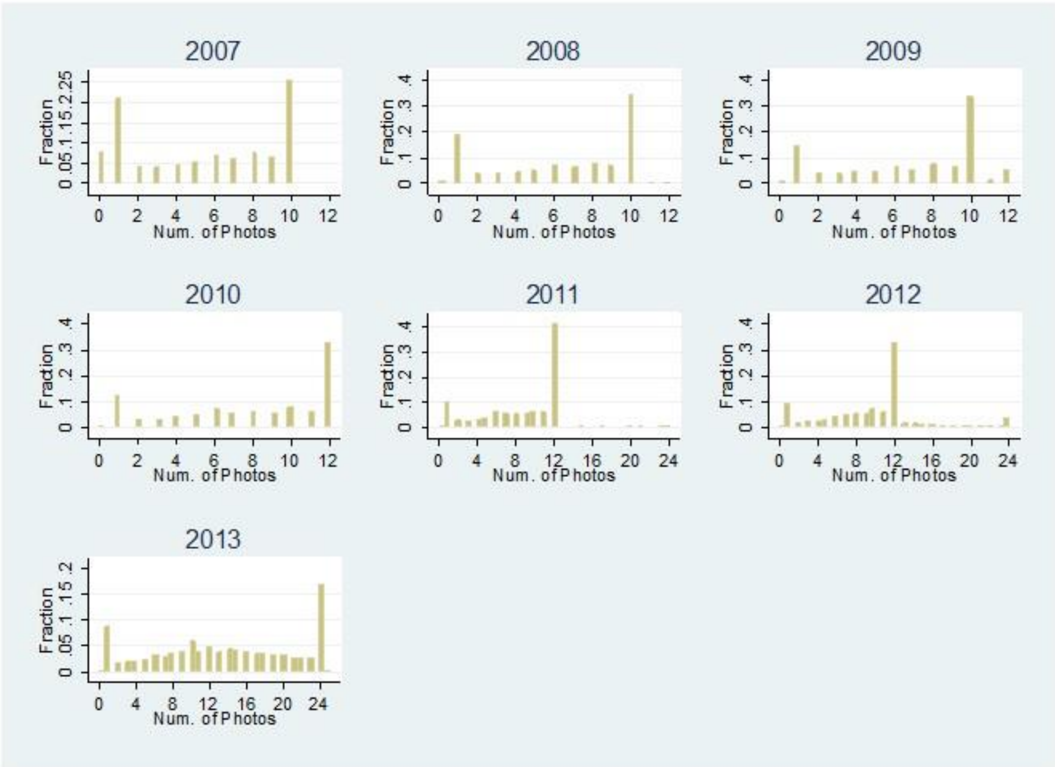


Figure 2



Tables

Table I
Summary Statistics

	Mean	*Std. Dev.*
Sale Price ($)	231,711.20	134,835.10
Time on Market (days)	109.19	91.22
List Price ($)	250,247.70	145,686.70
Sold	0.46	0.50
# of Photos	8.08	5.17
# of Words in the Comments	73.59	39.95
Information Quantity – Public (defined in Section 5)	154.37	66.61
Information Quantity – All	172.46	69.25
Key Field Blank	0.01	0.09
Low Information Quantity – All (dummy)	0.08	0.26
Owner Agent (dummy)	0.04	0.20
Bedrooms	3.45	0.83
Bathrooms	2.34	0.96
Square Feet	2,038.01	876.69
Age (in years)	29.30	27.92
Acreage	1.06	3.04
Condo/Townhouse (dummy)	0.10	0.30
Vacant (dummy)	0.40	0.49
Dual Agency Transaction	0.05	0.22
Dual Brokerage Transaction	0.08	0.26
Quick Sale (1 day)	0.02	0.12

Table 2
The Effect of Information Quantity on a Home's (logged) Sale Price:
OLS Fixed Effects Model

	Dep. Var.: ln(Sale Price)	Dep. Var.: ln(Sale Price)	Dep. Var.: ln(Sale Price)	Dep. Var.: ln(Sale Price)
	[1]	[2]	[3]	[4]
# of Photos	0.0109***			
	(19.79)			
# of Words in the Comments	0.00003*			
	(1.71)			
Information Quantity (Public)		0.0006***		
		(17.92)		
Information Quantity (All)			0.0005***	
			(18.03)	
Low Info. Quantity (All)				-0.0462***
				(-7.04)
Key Field Left Blank				-0.2123***
				(-8.70)
Property Characteristics	✓	✓	✓	✓
Spatial Fixed Effects	✓	✓	✓	✓
Year Fixed Effects	✓	✓	✓	✓
R^2	0.6016	0.5954	0.5948	0.5900
Observations	90,584	90,584	90,584	90,584

Notes. This table shows the effect of different measures of information quantity on a home's logged sale price, where different measures of information quantity/quantity are associated with higher home prices. The last column shows that substantially low information or missing fields in the listing correspond to substantially lower prices.

Robust t-statistics in parentheses (Errors Clustered by Census Block Group)
*** p<0.01, ** p<0.05, * p<0.10

Table 3
The Effect of Information Quantity on a Home's Time on Market:
Parametric Hazard Models – Accelerated Failure-Time Form

	Weibull Coefficients: Dep. Var.: Days on Market	Weibull Coefficients: Dep. Var.: Days on Market	Weibull Coefficients: Dep. Var.: Days on Market	Weibull Coefficients: Dep. Var.: Days on Market
	[1]	[2]	[3]	[4]
# of Photos	-0.0257*** (-21.10)			
# of Words in the Comments	-0.0001 (-0.74)			
Information Quantity (Public)		-0.0014*** (-16.94)		
Information Quantity (All)			-0.0014*** (-17.33)	
Low Info. Quantity (All)				0.1042*** (3.82)
Key Field Left Blank				0.4072*** (5.17)
Property Characteristics	✓	✓	✓	
Spatial Fixed Effects	✓	✓	✓	✓
Year Fixed Effects	✓	✓	✓	✓
Sold Homes ("Failures")	88,746	88,746	88,746	88,746
Observations	192,939	192,939	192,939	192,939

Notes. This table shows the effect of different measures of information quantity on time on market. All coefficients are in accelerated failure-time form (Weibull parametric hazard model), showing that higher information in a listing is associated with shorter time on the market. The last column shows that substantially low information or missing fields in the listing correspond to longer time on market.

Robust z-statistics in parentheses (Errors Clustered by Zip Code)

*** $p<0.01$, ** $p<0.05$, * $p<0.10$

Table 4
The Effect of Information Quantity on a Home's Probability of Sale:
Parametric Hazard Models – Log Relative-Hazard Form

	Weibull Coefficients: Dep. Var.: Days on Market	Weibull Coefficients: Dep. Var.: Days on Market	Weibull Coefficients: Dep. Var.: Days on Market	Weibull Coefficients: Dep. Var.: Days on Market
	[1]	[2]	[3]	[4]
# of Photos	0.0201*** (20.77)			
# of Words in the Comments	0.0001 (0.74)			
Information Quantity (Public)		0.0011*** (16.93)		
Information Quantity (All)			0.0011*** (17.27)	
Low Info. Quantity (All)				-0.0813*** (-3.82)
Key Field Left Blank				-0.3180*** (-5.15)
Property Characteristics	✓	✓	✓	✓
Spatial Fixed Effects	✓	✓	✓	✓
Year Fixed Effects	✓	✓	✓	✓
Sold Homes ("Failures")	88,746	88,746	88,746	88,746
Observations	192,939	192,939	192,939	192,939

Notes. This table shows the effect of different measures of information quantity on a home's probability of sale or hazard rate. These models are identical to Table 3 except that all coefficients are in log relative-hazard form (Weibull parametric hazard model), instead showing that higher information is associated with a higher hazard rate at a given point in time. Substantially low information or missing fields in the listing correspond to a lower hazard rate or probability of sale at a given point in time.

Robust z-statistics in parentheses (Errors Clustered by Zip Code)

*** $p<0.01$, ** $p<0.05$, * $p<0.10$

Table 5
Heterogeneous Effects of Information on Home Prices – List Price Stratifications

	Dep. Var.: ln(Sale Price)	Dep. Var.: ln(Sale Price)	Dep. Var.: ln(Sale Price)	Dep. Var.: ln(Sale Price)
	[1] *Lower 3 Quartiles (<$299,950)*	[2] *Upper Quartile List Price (>$299,950)*	[3] *Lower 3 Quartiles (<$299,950)*	[4] *Upper Quartile List Price (>$299,950)*
Low Info. Quantity (All)	-0.0704*** (-13.48)	0.0176*** (2.58)		
Information Quantity (All)			0.0006*** (22.87)	-0.00004* (-1.89)
Property Characteristics	✓	✓	✓	✓
Spatial Fixed Effects	✓	✓	✓	✓
Year Fixed Effects	✓	✓	✓	✓
Observations	69,271	21,519	69,271	21,519

Notes. This table shows the coefficient estimates of the hedonic price model in Table 2 (column 4 and 3), but stratified by list price, to show that the effect of information quantity is different for high priced vs. lower priced homes, with list price used as a proxy for home heterogeneity. While negative for lower priced homes, the second column shows that the effect of low information is positive for high priced homes. The final two columns shows a continuous measure of information quantity, showing that it is positively (negatively) associated with sale price for homes in the upper quartile (lower three quartiles) of list price. While untabulated, the effect of leaving a key field blank in the MLS listing, however, is substantively unchanged across these stratifications and is always associated with lower prices.
Robust t-statistics in parentheses (Errors Clustered by Census Block Group)
*** p<0.01, ** p<0.05, * p<0.10

Table 6
Heterogeneous Effects of Information on Time on Market– List Price Stratifications

	Weibull Coefficients Dep. Var.: Days on Market	Weibull Coefficients Dep. Var.: Days on Market	Weibull Coefficients Dep. Var.: Days on Market	Weibull Coefficients Dep. Var.: Days on Market
	[1] *Lower 3 Quartiles (<$299,950)*	[2] *Upper Quartile List Price (>$299,950)*	[3] *Lower 3 Quartiles (<$299,950)*	[4] *Upper Quartile List Price (>$299,950)*
Low Info. Quantity (All)	0.1808*** (7.95)	-0.2491*** (-4.67)		
Information Quantity (All)			-0.0017*** (-19.59)	0.0005** (2.35)
Property Characteristics	✓	✓	✓	✓
Spatial Fixed Effects	✓	✓	✓	✓
Year Fixed Effects	✓	✓	✓	✓
Observations	142,252	50,683	142,252	50,683

Notes. This table shows the coefficient estimates of the hedonic hazard model in Table 3 (columns 4 and 3), but stratified by list price, to show the effect of information quantity on time on market for high priced vs. lower priced homes, with list price used as a proxy for potential home heterogeneity along horizontal dimensions. Low information is associated with shorter time on market for high priced homes, and longer time on market for lower priced homes. The final two columns shows a continuous measure of information quantity, indicating that it is positively associated with time on market for homes in the upper quartile of list price, and negatively associated with time on market for the bottom three quartiles.

Robust z-statistics in parentheses (Errors Clustered by Zip Code)

*** p<0.01, ** p<0.05, * p<0.10

Table 7

Determinants of Information Quantity Provided in an MLS Listing

	Dep. Var.: Information Quantity (All)	Dep. Var.: Low Information Quantity (All)	Dep. Var.: Information Quantity (All)	Dep. Var.: Low Information Quantity (All)
	[1]	[2]	[3]	[4]
Upper Quartile List Price	1.5452**	0.0060**		
(=1 if >$299,950)	(2.12)	(2.34)		
Upper Quartile PLE			3.9506***	-0.0047**
(=1 if PLE>15.35%)			(5.70)	(-2.13)
Owner-agent	-4.3836***	0.0165***	-5.2353***	0.0158***
	(-4.34)	(3.46)	(-5.02)	(3.23)
Upper Quartile LP * Owner-agent	-6.8924***	-0.0002		
	(-3.09)	(-0.03)		
Upper Quartile PLE * Owner-agent			-4.6466**	0.0031
			(-2.01)	(0.38)
Total Interacted O-A Effect from Above				
Linear Combination of Estimators	-9.7308***	0.0222***	-5.93135***	0.0142**
(Joint Significance t-test)	(-4.62)	(3.36)	(-2.83)	(2.17)
Property Characteristics	✓	✓	✓	✓
Spatial Fixed Effects	✓	✓	✓	✓
Year Fixed Effects	✓	✓	✓	✓
R^2	0.2132	0.0169	0.2135	0.0169
Observations	198,345	198,345	198,345	198,345

*Note*s. This table shows the effect of proxies for home heterogeneity (along horizontal dimensions) on the how much information is provided in an MLS listing. To summarize, the results show that when agents market their own homes, they tend to provide less information than their clients' or are more likely to provide "low information." In addition, when they market their own more heterogeneous home (proxied by list price or percentage listing error (PLE)), they also tend to provide less information than their clients' properties. The test of joint significance reports the coefficient for the linear combination of estimators for the coefficients of the variable of interest, owner-agent variable, and the interaction among the two (yielding the total effect when an owner-agent markets a heterogeneous home for their own listing relative to a normal client listing).

T-statistics in parentheses (Errors Clustered by Census Block Group)

*** p<0.01, ** p<0.05, * p<0.10

Appendix

Table A
Instrumental Variables Estimation of Heterogeneous Effects on Home Prices –
List Price Stratifications (which are identical to Table 5)

	Dep. Var.: ln(Sale Price)	Dep. Var.: ln(Sale Price)	Dep. Var.: ln(Sale Price)	Dep. Var.: ln(Sale Price)
	[1]	[2]	[3]	[4]
	Lower 3 Quartiles (<$299,950)	*Upper Quartile List Price (>$299,950)*	*Lower 3 Quartiles (<$299,950)*	*Upper Quartile List Price (>$299,950)*
Low Info. Quantity (All)	-0.0704*** (-18.68)	0.0171*** (3.65)		
Information Quantity (All)			0.0006*** (39.80)	-0.00005*** (-2.84)
Property Characteristics	✓	✓	✓	✓
Spatial Fixed Effects	✓	✓	✓	✓
Year Fixed Effects	✓	✓	✓	✓
Observations	69,271	21,519	69,271	21,519

Notes. This table shows the coefficient estimates of an instrumental variables model of the Table 5 regressions on sale price, but stratified by list price, to show that the effect of information quantity is different for high priced vs. lower priced homes, with list price used as a proxy for potential home heterogeneity along horizontal dimensions. Information quantity and low information quantity are instrumented by the average level of information provided by the listings of the same brokerage office in a given year. The results are similar to the OLS results in Table 5.

Robust t-statistics in parentheses (Errors Clustered by Census Block Group)
*** p<0.01, ** p<0.05, * p<0.10