Learning from Past Prices: Evidence from the UK Housing Market

Simona Risteska

The London School of Economics *

November 8, 2020 – This Version
June 11, 2020 – First Version

Abstract

Imperfect knowledge about the structure of information flows can cause pricing mistakes in environments characterised by social learning. Using a natural experiment from the UK housing market that allows me to identify regular shocks to the information set of prospective sellers, I show that they overweight stale information when setting prices. I argue that this is inconsistent with Bayesian rationality and propose a model of naïve learning. I use the model, calibrated to the data, to measure the economic magnitude of the long-run effects arising from pricing mistakes. The results indicate that noisy signals bearing little information about future demand can have a long-lasting effect on aggregate prices when the dynamics of demand are highly persistent.

Keywords: Learning, Housing Markets, Behavioural Biases, Bounded Rationality, Herding.

JEL Classification: D83, D84, D91, G40, R20, R31.

*I benefited from helpful comments from Cynthia Balloch, Daniel Ferreira, Dirk Jenter, Christian Julliard, Francesco Nicolai, Daniel Paravisini, Marco Pelosi, Cameron Peng, Walker Ray, Huan Tang, Michela Verardo, and the participants at the LSE seminar. The paper contains HM Land Registry data © Crown copyright and database right 2019. The data is licensed under the Open Government Licence v3.0. 1. I thank the University of Glasgow - Urban Big Data Centre for providing Zoopla property data. Zoopla Limited, © 2019. Zoopla Limited. Economic and Social Research Council. Zoopla Property Data, 2019 [data collection]. University of Glasgow - Urban Big Data Centre. Simona Risteska, Department of Finance, The London School of Economics and Political Science, Houghton St, London, WC2A 2AE, UK, email: s.risteska@lse.ac.uk.

1 Introduction

The main objective of this paper is to investigate how economic agents make inference from the observable actions of others. In the face of uncertainty and imperfect knowledge about the state of the world, social learning allows individuals to learn private information that is embedded in other people's actions. However, in settings where individuals have limited knowledge about the structure of information flows, this may lead to inferential mistakes whenever agents fail to account for the repetitive use of stale information. A particular form of social learning commonly used in financial markets is referred to as pricing by comparables. When determining the value of an asset, individuals frequently draw on past observations of similar transactions to guide their decision-making. This valuation method is employed in a variety of settings, from firm valuation in corporate finance to the pricing of illiquid assets such as corporate bonds and loans and, perhaps most notably, in the housing market when assessing the value of commercial and residential properties.

In this paper, I show that sellers in the housing market overweight old information when setting prices. In particular, I use the UK market for residential housing as a laboratory to investigate how agents make inference when exposed to the release of new information. The institutional setting in the UK serves as an ideally suited natural experiment in these regards: beginning in March 2012, the UK Land Registry has been regularly publishing, on the twentieth working day of each month, data on all the transactions of residential properties that have taken place in the previous month. This provides me with a shock to the information set of prospective sellers around the latest price data publication date. Combined with rich data on listings from a large property website in the UK and data on house characteristics, this allows me to analyse the causal effect of recent transactions on property listings. I first supply empirical evidence that prospective sellers use data on past transactions to inform their decisions. Transactions from the previous month have a significantly larger effect on listings posted in the period after these have been made publicly available. I, therefore, confirm the well-known fact that pricing by comparables is widely used by sellers as one should expect given the fact that this is an approach openly recommended by real-estate agents, property websites and other housing market professionals. Any given comparable transaction has about 0.45% incremental effect on listings that observe it relative to those that do not, even though the latter are closer in time. I then proceed to show that this is a lower bound on the true effect by conducting a difference-in-differences analysis where I benchmark the incremental effect described above to data before the first publication date in March 2012: the results indicate that the actual response is almost twice as large than the previously estimated one.

I then proceed to the main results of the paper: by looking at how the influence of any given transaction evolves with its repeated use, I demonstrate that sellers in the housing market fail to recognize potential duplication of information and are, therefore, prone to overweight stale news at the expense of more recent information. Specifically, I find that the effect of recently published transaction prices increases monotonically with the number of redundant channels of influence. The ability to observe the date and price of new property listings (hereafter also referred to as quote) allows for a detailed analysis of the way information flows from past prices to subsequent listings, potentially also via intermediate comparable listings. The results show that the incremental effect that recent transactions have on future quotes can be more than 3% when the number of intermediate comparables grows beyond three relative to the case where no such redundant channels are present. This incremental effect is added to the baseline influence that recent prices have on future listings of about 83-84% implying that housing market fundamentals are quite persistent and, consequently, even small pricing mistakes can have significant long-run effects. The above findings cannot be squared with Bayesian inference. In particular, a Bayesian agent would take into account the fact that recent comparables have been influenced by earlier ones and should, therefore, adjust the relative weights placed on observables accordingly in order to avoid double-counting stale information. This implies that the effect of a given comparable cannot increase as the number of interim listings grows. I, therefore, reject the null hypothesis that agents in the housing market behave in a Bayesian fashion.

The above results can be reconciled with a different learning model where agents fail to recognise potential duplication of information in prior observables, practice known in the theoretical literature on social learning as naïve herding or persuasion bias (Eyster and Rabin, 2010; DeMarzo et al., 2003). These papers have shown that, in order to make correct inference as implied by Bayesian updating, one needs to engage in a very complex process of discerning all the channels through which a given signal might have already exercised an indirect influence on his actions or else he would be inclined to overweight that piece of information. Specifically, the agent needs to be able to disentangle all determinants of a given observation, namely: (a) the private signal of that individual; (b) the part that is influenced by the observability of prior actions and; (c) the public information about fundamentals observed by everyone. This is not an easy task even in a setting with fully rational agents and common knowledge of the structure of the network, nevertheless introducing uncertainty about information flows significantly exacerbates the problem. Agents who instead, due to bounded rationality, attempt to make approximate inference from past actions by assuming that these are driven purely by distinct signals will be subject to naïve learning and risk placing too much weight on stale information.

Since sellers in the housing market have difficulty recognising potential duplication of information, it would be interesting to study whether this leads to inferior market outcomes. In particular, one could analyse if part of the mistakes made by sellers are corrected upon matching with buyers. In the final set of tests, I provide some suggestive evidence that this is indeed the case by showing that sellers who eventually sell their properties at the largest percentage difference to listed price are those that have been most highly influenced by past prices.

The empirical results provide solid evidence that learning and pricing behaviour in the real-estate market cannot be reconciled with Bayesian inference. They are, however, unable to demonstrate what the economic impact of such behaviour would be on housing market dynamics in the long run. For this reason, I finally develop a simple model of learning to simulate the response of naïvely formed prices to various shocks and benchmark this to the rational case. The results indicate that in a world with naïve agents, prices are much more sensitive to noisy signals about demand as they overreact to this information for a long time. The deviation from fundamental values can be very large at 35% of the shock at a twenty-year horizon. On the other hand, naïve prices exhibit underreaction to true changes in the value of the underlying state due to the fact that real shocks get suppressed by stale news. These results are of particular importance once we consider that the decision to purchase a (new) home is typically one of the biggest financial decisions households need to make and, therefore, pricing mistakes in the housing market can have large effects on household welfare.

In this paper, I provide empirical evidence of the extrapolative behaviour of sellers in the housing market. I, therefore, contribute to the literature on the behaviour of real-estate market participants and the way it affects pricing dynamics (Merlo and Ortalo-Magne, 2004; Piazzesi and Schneider, 2009; Head et al., 2014; Brunnermeier and Julliard, 2008; Ngai and Tenreyro, 2014; Merlo et al., 2015; Burnside et al., 2016; Anenberg, 2016; Glaeser and Nathanson, 2017; Guren, 2018; Giacoletti and Parsons, 2019; Bracke and Tenreyro, 2020). In particular, I expand on the results of Glaeser and Nathanson (2017) who calibrate a model where house market participants extrapolate from past prices by failing to adjust for the fact that past actions reflect beliefs about future demand. More broadly, this paper relates to the work of Murfin and Pratt (2019) who show that lenders in the market for corporate loans similarly overweight old information by treating past transactions as independent signals.

The paper proceeds as follows: in Section 2 I provide a theoretical foundation of naïve learning and outline the natural experiment that will guide the empirical analysis; Section 3 provides a survey of the existing literature on naïve learning and housing market dynamics; Section 4 describes the data and shows summary statistics; Section 5 presents the results of the empirical analysis; Section 6 develops a model in order to convey the economic magnitude of the long-run effects arising from pricing mistakes, and; Section 7 concludes.

2 Theoretical Motivation and Methodology

When there is uncertainty about the state of the world and the amount of knowledge that other actors possess, agents are naturally inclined to use observable actions and outcomes as a way to make better informed decisions. One of the most obvious examples of this is the widespread use of comparables for pricing financial assets. Under this approach, agents looking to determine the value of a given asset make use of available data on prices and transactions of similar securities[1]. When agents have less than full information regarding the path through which information propagates, they are likely to incur in mistakes if they apply the comparables approach blindly. In particular, if agents do not account for common drivers among the set of observed comparables and, instead, treat these observations as independent from each other, they might overweight some signals at the expense of others. This practice is known in the literature on social learning as naïve herding. The theoretical literature on this topic, pioneered by DeMarzo et al. (2003) and Eyster and Rabin (2010), has shown that agents that fail to account for common signals embedded in past actions are likely to make suboptimal choices and even herd on the wrong decision in the long run with positive probability. Even more surprisingly, Eyster and Rabin (2014) show that agents are required to anti-imitate, i.e., apply negative weight on the observable actions of some agents, in order to perform correct inference. My goal is to provide empirical evidence of the way that economic agents learn from past information and examine whether there is any indication of naïve herd-

[1]Consider a simple asset with a periodic cash flow C, growth rate g and discount rate r. Its price P is then determined by the standard Gordon growth formula: $P = \frac{C}{r-g}$. Re-arranging, we obtain the price-to-cash flow ratio: $\frac{P}{C} = \frac{1}{r-g}$. This formula implies that assets with the same discount rate and growth rate (or difference thereof) should have the same value multiple. The approach of pricing by comparables thus relies on the availability of assets with similar risk and growth characteristics to the asset in question.

ing. I will use the housing market as the setting for my analysis as this is one of the areas where the use of the comparables is most common. Moreover, the market for residential properties is, unlike most other financial markets, largely populated by households which might be less sophisticated compared to major actors in other security markets. As a result, the challenge of extracting the correct signals and avoiding any learning mistakes could be more difficult to overcome in the housing market where agents are present only temporarily and with possibly limited time and information resources.

To motivate the empirical analysis of this paper, I provide a simple stylized model that illustrates the key features of naïve learning and contrasts it with Bayesian updating. Consider an environment where agents learn about the state of demand denoted by D. Prospective sellers looking to determine the listing price for their property receive a noisy signal s_n with a normally distributed error, $\varepsilon_n \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$ identically and independently distributed across agents and time:

$$s_n = D + \varepsilon_n \tag{1}$$

Agents act sequentially, and every period, each seller observes the entire history of prices: $I_n = \{p_{n-1}, p_{n-2}, ..., p_0\}$. For simplicity, suppose that sellers set prices equal to their best estimate of the state of demand: $p_n(D) = E[D|s_n, I_n]$. The first agent $n = 0$ receives a signal s_0 and, not observing any prior actions, sets the price equal to the signal:

$$p_0 = E[D|s_0, I_0] = s_0 \tag{2}$$

Agent 1 receives a signal and observes the action of agent 0 and, given the equal precision, assigns the same weight to both signals:

$$p_1 = E[D|s_1, I_1] = \left(1 - \frac{\sigma^{-2}}{(\sigma^{-2} + \sigma^{-2})}\right) \times \underbrace{E[D|I_1]}_{=p_0} + \frac{\sigma^{-2}}{(\sigma^{-2} + \sigma^{-2})} \times s_1 = \frac{1}{2} \times (s_0 + s_1) \tag{3}$$

The difference between Bayesian and naïve updating arises with the arrival of the third agent. A Bayesian approach would require the agent to calculate the posterior belief as the average of his prior and the new signal weighted by the signal precision, where the prior is equal to the expectation of demand given the history of observed actions:

$$p_2 = E[D|s_2, I_2] = \left(1 - \frac{\sigma^{-2}}{(\sigma^{-2} + \sigma^{-2} + \sigma^{-2})}\right) \times \underbrace{E[D|I_2]}_{=p_1} + \frac{\sigma^{-2}}{(\sigma^{-2} + \sigma^{-2} + \sigma^{-2})} \times s_2 \quad (4)$$

Note that agent 2's prior belief is equal to agent 1's posterior, i.e., under common knowledge of rationality and the informational structure, agent 2 will simply set his prior equal to agent 1's best estimate. Plugging in the expression for the posterior of agent 1 into equation (4), we obtain the following expression:

$$p_2 = \frac{2}{3} \times p_1 + \frac{1}{3} \times s_2 = \frac{1}{3} \times (s_0 + s_1 + s_2) \quad (5)$$

The key take-away from the last equation is that under Bayesian updating, agent 3 will assign appropriate weights to all previous signals in proportion to their respective precisions. More generally, the posterior of agent n is equal to the precision-weighted average of all n available signals, i.e., for any $n \geq 1$:

$$p_n = \frac{1}{n+1} \times (s_0 + s_1 + ... + s_n) \quad (6)$$

Crucially, note that a simpler way to achieve this is by using the previous agent's posterior belief and disregarding all prior actions, i.e., the posterior belief of agent $n-1$ is a sufficient statistic for all previously observed information[2]:

$$p_n = \frac{n}{n+1} \times p_{n-1} + \frac{1}{n+1} \times s_n \quad (7)$$

[2]Note that the last result holds only in the case where agents act sequentially and there is only one agent per period. If, instead, there are multiple agents in a given period, say k of them, who are unable to observe each other's actions, the prior of the subsequent set of agents will not be equal to any of those agents' posterior beliefs or the average thereof. This is because the former would imply failure to absorb the private information of the remaining k-1 agents from period n-1, while the latter would lead to overweighting of commonly observed signals relative to the private signals from period n-1. Nevertheless, the result that the posterior of a given agent equals the weighted average of all available signals still holds.

The approach above, however, implies knowledge of the full history of actions and, most importantly, the way they have influenced each other. This might not be feasible in many real-world scenarios and might therefore lead to suboptimal decision-making in environments with social learning. To see this, consider a naïve learner in the third period, defined as one who fails to account for the redundancy of previous signals, treating them as independent instead. In other words, a naïve learner assumes that previous agents have not taken into account any prior information, forming prices based solely on their respective private signals. Agent 3's posterior will then be given by:

$$\begin{aligned} \widetilde{p}_2 = \widetilde{E}[D|s_2, I_2] &= \left(1 - \frac{\sigma^{-2}}{(\sigma^{-2} + \sigma^{-2} + \sigma^{-2})}\right) \times \widetilde{E}[D|I_2] + \frac{\sigma^{-2}}{(\sigma^{-2} + \sigma^{-2} + \sigma^{-2})} \times s_2 \\ &= \frac{2}{3} \times \left(\frac{1}{2} \times p_0 + \frac{1}{2} \times p_1\right) + \frac{1}{3} \times s_2 = \frac{1}{3} \times p_0 + \frac{1}{3} \times p_1 + \frac{1}{3} \times s_2 \end{aligned} \tag{8}$$

The above equation states that our naïve learner would use a wrong prior given by the precision-weighted average of the previous agents' *posteriors* as opposed to their signals: $\widetilde{E}[D|s_2, I_2] = \frac{\sigma^{-2}}{(\sigma^{-2}+\sigma^{-2})} \times (p_0 + p_1)$. Plugging in the expressions for p_0 and p_1 from equations (2) and (3) above, we obtain:

$$\widetilde{p}_2 = \frac{1}{2} \times s_0 + \frac{1}{6} \times s_1 + \frac{1}{3} \times s_2 \tag{9}$$

The last equation shows that naïve updating leads agents to assign wrong weights on prior signals. In particular, by overlooking the influence that signals further in the past have had on more recent actions, agents end up overweighting stale news. In contrast to Bayesian learners, naïve agents treat signals coming from early actions as distinct sources of information. This mistake gives rise to multiple channels of influence of early news: the direct channel arising from the placement of an explicit weight on past signals and the indirect one that emerges from their effect on intermediate observations. For a general

$n > 1$, the price looks as follows:

$$\widetilde{p}_n = \frac{1}{1 \times 2} \times s_0 + \frac{1}{2 \times 3} \times s_1 + ... + \frac{1}{n \times (n+1)} \times s_{n-1} + \frac{1}{n+1} \times s_n \tag{10}$$

Comparing equations (6) and (10) we see that naïve learning is an issue of relative over- and under-weighting of the signals coming from previous periods: notice that both Bayesian and naïve learners assign the same weight to their private information. It therefore arises even if agents are more confident about their own signals, as long as they learn from past data to some extent. It can further be noted that weighting mistakes would be present regardless of whether the signal precisions are equal across agents or not: naïve learning implies over-weighting of old signals relative to optimal weights even with heterogeneity in signal precisions.

The key distinctions between a Bayesian and naïve approach to learning outlined in equations (4)-(10) will guide my empirical analysis going forward. Specifically, the core of the paper will seek to benchmark the way that prices in the housing market influence each other against the two learning models described above. Note that deviations from Bayesian learning occur once early observable information gets embedded into intermediate actions. Bayesian learning implies that the effect of a given signal should not increase with the number of subsequent uses, rather we should expect it to decline with the arrival of more news as each individual piece of information gets progressively lower weight. To see this, we can fix a price from a given period $k \geq 0$ and compute its covariance with prices from subsequent periods $n \geq k+1$ under the rational and the naïve models. We can then compare the evolution of covariance functions as the number of intermediate observations grows. For the rational model, we have:

$$Cov(p_n, p_k) = \frac{1}{n+1} \times \sigma^2 \tag{11}$$

The above expression shows that, as we increase n or the amount of interim prices observed by agent n but not by agent k, the covariance between p_n and p_k monotonically decreases. This is intuitive since optimal learning implies that agent n will assign proportionally lower weights to the information embedded in p_k as he observes more and more recent news. On the other hand, the covariance of the same two prices in the naïve model would be as follows:

$$Cov(\widetilde{p}_n, \widetilde{p}_k) = \sum_{0 \leq i \leq k} \left(\frac{1}{i \times (i+1)} \right)^2 \times \sigma^2 + \frac{1}{(k+1)^2 \times (k+2)} \times \sigma^2 \tag{12}$$

Note that the covariance in the naïve case is no longer decreasing with n. In particular, in this simple setting covariances do not depend on n and, as a result, prices in all subsequent periods will comove with $\widetilde{p}_k$ by the same amount[3]. To better explain the empirical results of Section 5 below, I develop a dynamic model with evolving state of demand and introduce an additional commonly observed public signal which introduces correlation in the signals: the details of the model are presented in Section A of the Appendix.

In the rest of this paper, I make use of an ideally-suited setting for analysing the comovement in housing prices with the arrival of intermediate observations that potentially contain the same signals. Namely, starting from March 2012, the UK Land Registry has been publishing monthly housing transaction data on a regular basis on the twentieth working day of the subsequent month. Consequently, on this date prospective sellers receive an information shock due to the release of house price data from the previous month. Prior to March 2012, the data was available for purchase under contract and there was no such a sharp and regular discontinuity in the information set of sellers. Figure 1 gives a graphical representation of the environment. Suppose, for instance, that the twentieth working day of March of a given year is March 28th: this is the date when February transactions data is made publicly available. Sellers who list their properties

[3] It is important to note that throughout my empirical analysis, I consistently compare the effect of prices from a given period on subsequent listings based on the amount of intermediate information by fixing k and varying n. In particular, if we were to fix n and vary k, i.e., analyse the effect of prices from different periods on the same quote, the covariance implications would be different. To see this, note that equation (11) shows that under the rational model the covariance of p_n with different past prices p_k does not depend on the amount of intermediate observations. On the other hand, under the naïve model it can be shown from (12) that the same covariance is decreasing in k, i.e., increasing with the number of intermediate observations.

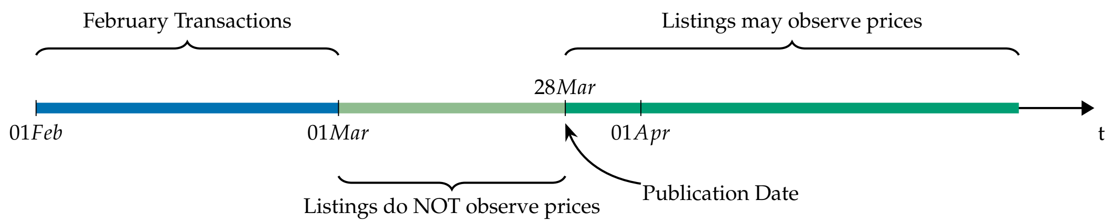


Figure 1 Empirical setting: Example of a transaction data publication date and listings posted in the surrounding period. The figure presents the natural experiment that generates shocks to the information set of sellers: beginning in March 2012, the Land Registry publishes regular monthly data on transactions on the twentieth working day of the subsequent month. For example, transaction prices, depicted in blue, from February, are published on the twentieth working day of March which is 28th March in this case. The property listings published at the beginning of March and before the publication date, depicted in light green, do not observe the data on February transactions, while those published after this date, depicted in dark green, may observe February price data and therefore can use this to make inference about market demand.

after this date can thus make use of the latest set of price data to inform their decisions. Sellers who have listed their properties just a few days before, however, are not able to observe the data on February transactions and thus cannot infer any private signals. Comparing the correlation of February transactions with properties listed just before and just after the publication date will, therefore, give us an idea of the effect of pricing by comparables in the housing market. Any incremental effect on quotes posted in the post-publication period shows evidence that sellers use information on newly published prices to learn about the current state of demand. The results of this exercise are shown in Section 5.1 and can be interpreted as the direct effect of pricing by comparables: they, however, fall short of explaining whether sellers incorporate new information in an optimal or naïve way. For this reason, after having established the baseline effect, in Section 5.2 I look at the way that the comovement of quotes with a given price evolves through chains of influence from subsequent listings. Specifically, using data on property listings and their timing in addition to the price paid data, we can get a good approximation of each agent's information set at the time of setting the quote. We can then compare the covariance between a given price p and a subsequent quote q based on the number of intermediate quotes that are observable by q which may or may not contain information also embedded in p. Varying the number of intermediate observations, the pattern in the covariance estimates between p and q can be benchmarked against equations (11) and (12) to determine if learning in the housing market takes place in an optimal way. Note that the regularity in the price publishing dates provides a good setting for estimating covariances of prices with subsequent quotes by taking into account the amount of interim information that sellers possess. To the extent that the listings posted on the two sides of the price publication date do not differ in a systematic way, comparing the evolution of influence from recent prices through sequences of listings around this date allows us to benchmark the estimated covariance coefficients to the Bayesian and naïve models described above. In other words, the changes in the covariance estimates between quotes around the publication date as the number of intermediate observations grows will allow us to determine if sellers are able to correctly extract the real news from a given price or if they instead end up overweighting commonly contained signals due to their inability to understand the duplication of information.

Throughout the analysis, I investigate the impact of transaction prices on listings of properties with similar characteristics in order to avoid any selection on observables. Moreover, I minimise concerns regarding the evolution of fundamentals by looking at a very tight window of listings posted in the two weeks surrounding the publication date. Similarly, I compare the effect of prices of properties sold that have at least one comparable listing before and one after the publication date to make sure that the results are not driven by systematic differences in the independent variable. In the next section, I provide a review of some of the existing literature that relates to this paper in order to outline its main contributions.

3 Previous Literature

The present paper relates to two broad strands of literature. First, it provides empirical evidence that complements the large body of theoretical literature on social learning beginning with the models of herd behaviour and informational cascades by Banerjee (1992) and Bikhchandani et al. (1992) in a setting with rational agents. Both papers show that when agents move sequentially and everyone observes all prior actions, using past observations to learn the information other agents might have had can lead to so-called herd behaviour where Bayesian agents stop listening to their own signals and follow everyone else. This in turns makes each agent's action less informative about their own signal and thus less useful to others. Banerjee (1992) demonstrates that the welfare implications of this type of behaviour can be significant to the extent that agents might gain by constraining information sharing. The type of positive feedback effects present in this setting implies that outcomes can be very different across game repetitions and that this might lead to excess volatility in asset markets. Bikhchandani et al. (1992) further go on to show that this type of cascades are fragile in the sense that they can seemingly break down in a drastic manner with the arrival of a small amount of information or a slight possibility of a value change. They also demonstrate that the gradual release of public information once a cascade has started can reverse this and eventually lead to individuals settling into the correct cascade. The above papers, inspect only herding effects that result in social settings with rational inference. Convergence on the wrong action with fully rational agents is, nonetheless, rare, it occurs primarily in cases where agents are not confident in their beliefs and, as Ho (1993) and Smith and Sorensen (2000) show, it arises in situations with coarse action or signal spaces. One of the early papers that study bounded rationality in social-learning environments is DeMarzo et al. (2003) who introduce the concept of "persuasion bias" defined as the failure to adjust for possible repetition of information coming either from one source over time or multiple sources connected through a network. In their paper they emphasize that the key issue causing this type of behaviour is the intractability of the path that led all prior individuals to form their beliefs. Theoretical papers that most tightly relate to the present article are Eyster and Rabin (2010) and Eyster and Rabin (2014) who study observational learning in rich-informaton settings with naïve agents. Specifically, Eyster and Rabin (2010) describe a form of so-called "inferential naïvety" whereby players learning from the observable actions of others fail to account for the influence of early actions on interim players' choices and, instead, treat all observations as purely driven by each player's private information. Just like in the simple model presented in Section 2 above, agents in their model move sequentially after receiving a private signal and observing the full history of past actions. They demonstrate that this type of behaviour can lead agents to converge to the wrong beliefs with full confidence to the point that they are made worse off by being able to observe the actions of previous movers. Perhaps most crucially for the subsequent tests on the network effects of learning from comparables, Eyster and Rabin (2014) prove that rational learning implies that in environments where agents share common observations, they should either never imitate more than one predecessor or rather engage in anti-imitating behaviour as well.

In terms of empirical literature, a closely-related paper that studies naïve learning is Murfin and Pratt (2019) who look at the market for corporate loans. They exploit the date on which a given loan is reported in Refinitiv's Dealscan database to identify the effect of new additions to the dataset on the pricing of subsequent loans. They find strong evidence of comparables pricing in credit markets and naïve inference whereby the effect of a given comparable increases by three to five percentage points in the presence of redundant channels of influence, up from a baseline effect of about 6-10%. The benefit of the present paper is that it makes use of a more cleanly defined shock to the information set of agents to identify the direct effect of pricing by comparables. Moreover, while Murfin and Pratt (2019) study the behaviour of investment bank professionals, I primarily look at households operating in the residential housing market who might be less sophisticated and, consequently, more prone to be influenced and to commit pricing errors. Furthermore, the purchase of a home most often is the biggest financial decision that households make which emphasizes the importance of any pricing mistakes. Less related to the present study, numerous other papers analyse the use of the comparables pricing method in corporate finance[4]. Papers that study herding behaviour in wider financial markets are, among others: Lakonishok et al. (1992), Grinblatt et al. (1995), Hong et al. (2005) and Dasgupta et al. (2011) who analyse institutional herding among money managers; Alevy et al. (2007), Ivković and Weisbenner (2007) and Wang and Wang (2018) who look at portfolio choices of retail and professional investors; Hong et al. (2004) and Brown et al. (2008) who study stock market participation among neighbours; Fracassi (2017) who looks at peer-effects among corporate managers; Bailey et al. (2018) who study social network effects on individual housing decisions.

Perhaps most relevant to the present work is Glaeser and Nathanson (2017)'s research on suboptimal learning behaviour in the housing market that looks at homebuyers who extrapolate from past transaction data by assuming that past prices are pure manifestation of contemporaneous demand. They develop and calibrate a model of house prices and demonstrate that it matches fairly well the short-term autocorrelation, as well as the medium-term reversal and excess volatility of house prices observed in the data. Most interestingly, they find that bubble-like features are most severe when buyers have decent amount of data about past prices but limited information about fundamentals. Although I similarly look at naïve inference in the housing market, the key distinction between my paper and the one by Glaeser and Nathanson (2017) is that I study the implications of pricing biases on the part of homesellers. Furthermore, I provide more detailed micro-evidence on the pricing patterns that result from naïve learning by employing a rich dataset of house prices and characteristics. Specifically, the ability to observe a good proxy for the information set of prospective sellers allows me to identify chains of influence and obtain empirical estimates for the indirect effects of past prices on future listings that arise under naïve learning. I subsequently use these estimates to calibrate the structural parameters and show the effect of various shocks to agents' information

[4]See, for instance, Baker and Ruback (1999), Bhojraj and Lee (2002) and Liu et al. (2002) for the study of the performance of this approach in equity valuation, Kim and Ritter (1999) and Purnanandam and Swaminathan (2004) for evidence on the use of the comparable firms multiples approach in initial public offerings, and DeAngelo (1990) and Kaplan and Ruback (1995) on its use in the market for corporate control.

sets on aggregate pricing dynamics.

The paper also relates to the broader literature on housing markets trying to explain the behaviour of market participants and aggregate market dynamics[5]. Piazzesi and Schneider (2009) present a model where a small number of irrationally optimistic individuals can have a large price impact without the need to obtain a large market share. Head et al. (2014) develop and calibrate a dynamic search model that generates close to half of the serial correlation in house price growth. Burnside et al. (2016) propose a model with heterogeneous beliefs and social interactions to study the boom-bust cycles prevalent in housing markets. Anenberg (2016) presents a micro-search model where sellers facing information frictions update their beliefs about house values with the arrival of buyers. His model is able to match many of the micro features present in the data and can explain half of the short-term persistence in aggregate price dynamics. Guren (2018) proposes a mechanism that amplifies frictions through strategic complementarity, i.e., the willingness of sellers to set listing price close to the cross-sectional average in order to optimise the trade-off between selling price and time on the market. He shows that this mechanism causes sluggish price adjustment by sellers and can magnify momentum by a factor of two to three.

Finally, this paper touches on the literature on extrapolative expectations and behavioural biases. Fuster et al. (2010) propose a dynamic model where agents form expectations that overestimate the persistence of economic shocks. Similarly, Barberis et al. (2015) study a consumption asset-pricing model where only a group of agents form beliefs by extrapolating from past returns. Both papers find that the model fits the data on aggregate economic and financial variables well. In a similar vein, Kuchler and Zafar (2019) use survey data to show that individuals extrapolate from personal experience when forming beliefs about aggregate outcomes such as house price changes and unemployment levels.

[5]For a survey of the literature on the microstructure of housing markets, see Han and Strange (2015). For a review of the theoretical and empirical literature on house price dynamics, see Cho (1996).

4 Data and Summary Statistics

In this section I will describe the data I use for the empirical analysis of the effect of using comparables in the housing market and the learning behaviour of agents. The data on house prices comes from the Price Paid dataset published by the HM Land Registry. This data contains information on transactions of residential houses in England and Wales starting from 1995 to the present. Apart from some exemptions[6] all transactions of residential properties that have been sold for full market value are recorded and made publicly available by the UK Land Registry. The Price Paid dataset provides information on the date of the transaction[7], the transfer price, the full address of the property, as well as some additional characteristics about the property such as: the age of the property, i.e., whether the property is a new construction or an existing building; the duration of the lease (freehold or leasehold)[8]; and the property type categorised as either a detached, semi-detached, terraced house, or a flat.

Data on listed properties and listing prices comes from the Zoopla Property data[9]. Zoopla is the second largest provider of property data for consumers and property professionals in the UK, having access to over 27,000,000 residential property records and 15 years of price data. The full dataset available for research purposes contains over 5,000,000 records of properties listed for sale and over 3,000,000 records of properties advertised for rent. Zoopla's website is one of the most commonly used in the UK for listing properties for sale, second only to Right Move but expanding in market coverage. The data mainly covers the period between 2009 through 2018 for properties located in

[6]Transactions that are excluded from the Price Paid dataset include commercial transactions, property transactions that have not been lodged with the HM Land Registry and properties sold below market value. For more details on the property sales not included in the dataset the reader can visit the HM Land Registry website: https://landregistry.data.gov.uk.

[7]This is the completion date of the sale as stated on the transfer deed.

[8]Note that first registration of leases for seven years or less are not recorded in the dataset.

[9]The access to the dataset has been provided by the University of Glasgow - Urban Big Data Centre. Access to the dataset for research purposes can be obtained directly through the Urban Big Data Centre. The data has been collected by Zoopla. Zoopla Limited, © 2019. Zoopla Limited. Economic and Social Research Council. Zoopla Property Data, 2019 [data collection]. University of Glasgow - Urban Big Data Centre.

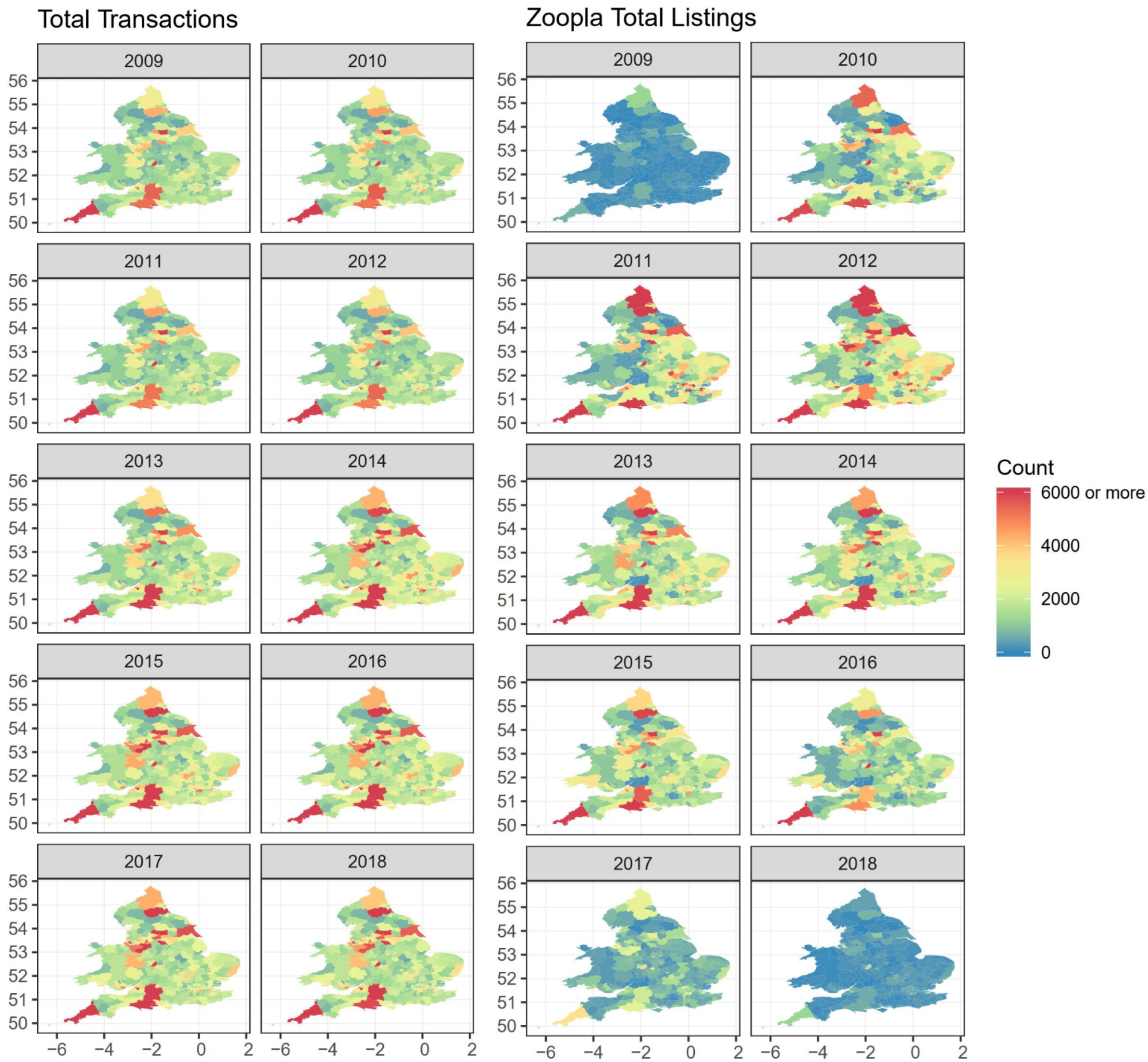
Great Britain, with partial coverage from 2005. The key variables for my empirical work are the quoted prices along with the dates at which these have been updated. The data also gives information about the date on which the property has been initially listed and the date on which it has been withdrawn from the market. The above information will be crucial to my empirical analysis as the goal of investigating the impact of newly available prices on new listings requires me to have a precise idea of the moment in time when listing prices are set/updated and the information set of the prospective sellers at that time. In addition, this dataset contains other property characteristics such as property location, property type[10], whether the property has been categorised as residential or commercial[11], number of bedrooms, number of reception rooms, number of bathrooms, number of floors and whether the property is listed for sale or for rent[12].

The final piece of data I use in my empirical analysis is the Domestic Energy Performance Certificates dataset from the Ministry of Housing, Communities and Local Government. Before 2008, the Energy Performance Certificates (EPC) for domestic properties could be lodged on a voluntary basis. From 2008 onwards, however, it has become mandatory for accredited energy assessors to lodge the energy certificates. Consequently, the data coverage drastically improves around that time, as does my ability to match these with the Price Paid and the Zoopla data. More specifically, the matching rate goes from a little over 50 percent in 1995 to over 90 percent around 2008. The dataset contains information on the address, property type, total floor area, number of storeys, number of rooms, floor level and height, along with many indicators of energy efficiency and quality of glazed surfaces.

[10]Property types include: barn conversion, block of flats, bungalow, business park, chalet, château, cottage, country house, detached bungalow, detached house, end-terrace house, equestrian property, farm, farm house, finca, flat, hotel/guest house, houseboat, industrial, land, leisure/hospitality, light industrial, link-detached house, lodge, longère, maisonette, mews house, mobile/park home, office, parking/garage, pub/bar, restaurant/cafe, retail premises, riad, semi-detached bungalow, semi-detached house, studio, terraced bungalow, terraced house, town house, unknown, villa and warehouse. For my analysis, I focus on the following property types: detached house, end-terrace house, flat, link-detached house, maisonette, mews house, semi-detached house, terraced house, town house, studio and villa.

[11]I keep only properties categorised as residential.

[12]I exclude properties listed for rent from my sample.



(a) Transactions **(b)** Listings

Figure 2 Geographic Coverage

The figure plots heat maps of the geographic coverage of the transaction and listing data between 2009-2018 by year across England and Wales, computed as the total number of observations by local authority district. Figure 2a displays the total number of transactions, while figure 2b the total number of unique listings in the sample.

Figure 2 displays heat maps of the spatial coverage of the data across England and Wales. Figures 2a and 2b show, for every year, the total number of transactions and listings, respectively, by local authority district. Comparing the two sets of maps, it can be noted that they display similar patterns and thus the listings data closely matches the true sales activity in the UK housing market. However, it can be seen that the Zoopla sample mainly covers the period between 2010 and 2017, with very few observations in 2009 and 2018. Figures 1a and 1b in the Appendix confirm this by showing the ratio of listings-to-transactions and the fraction of transactions whose listing information could be found in the Zoopla data across areas as a way of demonstrating the relative Zoopla coverage. As can be seen, between twenty and eighty percent of transactions have been matched to their respective listing in the Zoopla dataset across most regions in the period between 2010 and 2017 with the coverage peaking between 2011 and 2015. Nonetheless, it is reassuring to know that the data is well-dispersed across space and time as this reduces the probability that the results presented later in the paper are driven by a small subsample unrepresentative of the aggregate dynamics of the UK housing market.

As I am looking to investigate the effect of using the comparables pricing method in the housing market, most of my empirical work requires me to match listings with recent transactions of properties with similar characteristics. The goal is to replicate the natural approach that a prospective seller would take when deciding at what price to list their property. For this purpose, I match listings to recently sold houses based on four criteria: (1) the property location measured using the first half of the postcode[13]; (2) a rural/urban indicator from the 2011 Census classification of Output Areas; (3) property type divided in four categories, these beeing a detached house, semi-detached house, terraced house

[13]Postcodes in the UK are formed of five to seven alphanumeric characters and are typically split into two parts: the outward code and the inward code. In my work, I compare properties that have the same outward code which corresponds to properties that belong to the same subdistrict.

and a flat, and; (4) number of rooms in the property[14,15].

Table 1 provides summary statistics for the sample of listings and transactions that have at least one match and, therefore, will form part of the empirical analysis. The main sample covers the period after March 2012, the date when the Land Registry began publishing monthly Price Paid data on a regular basis. However, the data before March 2012 will be used in some of the robustness checks later and thus I separately present summary statistics for this part of the sample for comparison. I remove observations where the listing or transaction price is below £10,000 or above £25,000,000 to make sure that outliers do not drive the results. I also eliminate properties that have more than twenty rooms as well as observations with no rooms. For the final sample, I end up with 1,983,528 listings and 2,521,505 transactions deemed comparable post March 2012; prior to March 2012, there are 1,007,942 such listings matched with 986,287 recent transactions. Looking at the price statistics, we can observe that quoted prices tend to be larger than transaction prices: post March 2012 the average price at which a property is listed equals £268,402, while the average price paid for a property is £256,734. In the earlier part of the sample both are slightly lower at £233,497 and £220,134, respectively, which is natural given that real-estate prices normally exhibit a positive trend. The data also confirms the positive skewness in house prices with the median listing and transaction prices being significantly lower than the average at £194,950 and £189,995, respectively in the sample from 2012 onwards. The above results are consistent with findings from the previous literature (Merlo and Ortalo-Magne, 2004; Carrillo, 2012; Han and Strange, 2016; Guren, 2018). Second, I divide the data based on time-invariant property characteristics in order to show that the sample is well-balanced both across sets (listings and transactions) as well as across sample periods. In particular, about 15% to 19% of the properties in my

[14]The number of rooms variable of choice comes from the EPC dataset and it includes any living room, sitting room, dining room, bedroom, study and similar, a non-separated conservatory with an internal quality door and a kitchen/diner with a discrete sitting area. Excluded from the count are rooms used solely as a kitchen, utility room, bathroom, cloakroom, en-suite accommodation and similar, any rooms not having a window and any hallway, stairs or landing.

[15]I group into one room category properties having between six and ten rooms. Similarly, all properties with more than ten rooms are also considered comparable to each other.

Table 1 Listings and Recent Comparable Transactions - Summary Statistics
The table presents summary statistics for the set of listings and comparable transactions that have at least one match. Summary statistics are presented separately for the sample of data before March 2012 and post March 2012. Nb. of Observations refers to the total number of unique listings and transactions, respectively. Listing price is the first quote at which a property has been listed, while transaction price is the final agreed price between the buyer and the seller. Property type refers to the built-form of the property which can be one of four possible categories: detached, semi-detached, terraced house or a flat. Number of rooms refers to the total number of habitable rooms in the property. I report the following statistics on the distribution of prices and number of rooms: mean is the average value, min is the lowest value, p25, median and p75 are the 25-th, 50th and 75-th percentile of the distributions, respectively, and max is the highest value observed in the sample. For property type, I report the fraction of observations that are of a given type.

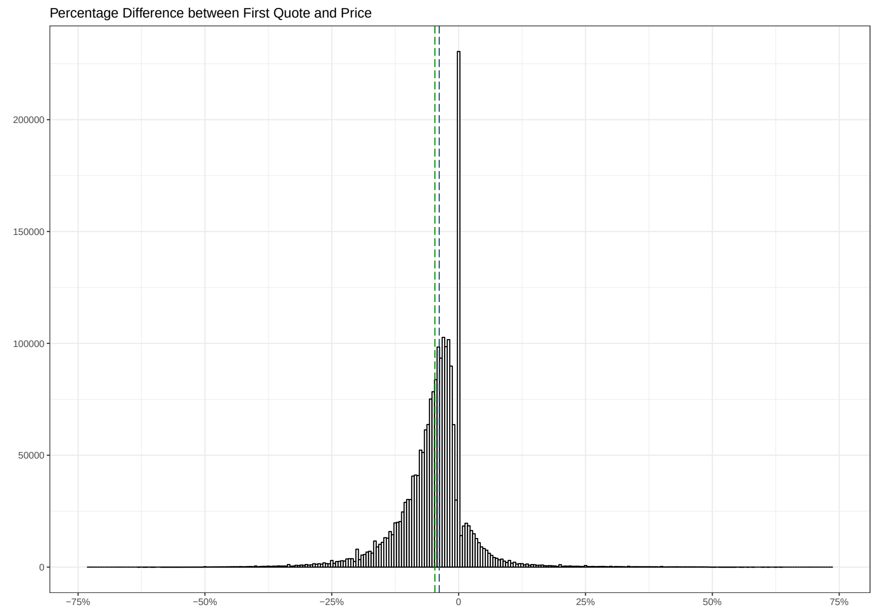
	Listings		Transactions	
	Pre March 2012	Post March 2012	Pre March 2012	Post March 2012
Nb. of Observations	1,007,942	1,983,528	986,287	2,521,505
Listing/Transaction Price				
Mean	£233,497	£268,402	£220,134	£256,734
Min	£10,500	£10,500	£10,300	£10,018
P25	£125,000	£129,995	£119,995	£125,000
Median	£178,500	£194,950	£170,000	£189,995
P75	£265,000	£310,000	£250,000	£300,000
Max	£17,500,000	£25,000,000	£19,250,000	£18,500,000
Property type (%)				
Detached	16.00	15.23	17.34	18.67
Semi-detached	28.11	28.84	27.84	29.12
Terraced	31.66	34.15	31.27	31.72
Flat	24.23	21.77	23.55	20.49
Number of rooms				
Mean	4.47	4.53	4.49	4.59
Min	1	1	1	1
P25	3	3	3	4
Median	4	4	4	5
P75	5	5	5	5
Max	18	19	18	19

sample are detached houses, 28% to 29% semi-detached houses, 31% to 34% terraced houses and the remaining 20% to 24% are flats. The average property has between four and five rooms and this is consistent across sample periods.

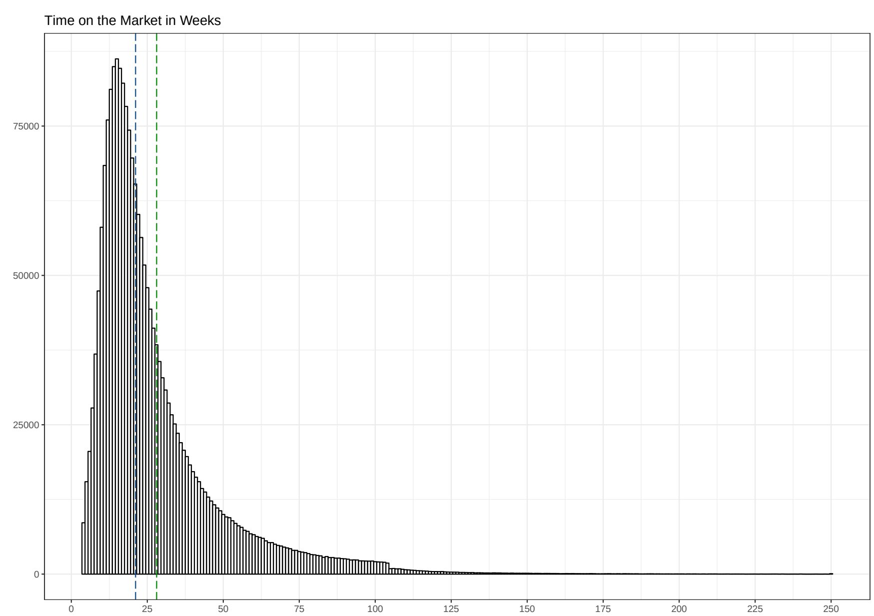
As part of my analysis will focus on ways that any potential mistakes made by sellers when setting quotes could be rectified by buyers at the selling stage, I also attempt to match listings to their respective ex-post transactions. To achieve this, I first match the data from Zoopla with the Price Paid data by property address; I then keep only the matches for which the transactions occurs at least four weeks[16] and no more than five years after the property has been listed on the market; I finally eliminate cases where the sale price is more than 50% above or below the final quote for that listing. This procedure leaves me with a sample of 2,086,462 listings matched to their transactions for the period between 2009 and 2018. Figure 3 displays the distributions of the price differential and time on the market (TOM) for the set of matched properties. The mean and median values of the two distributions are represented by the green and blue vertical lines, respectively. Looking at Figure 3a, it is notable that the distribution of the percentage difference between the listed price and sale price exhibits a large spike at zero, namely, over 10% of the matched transactions occur at the ask price, consistent with the findings in Merlo and Ortalo-Magne (2004) and Guren (2018), among others. This suggests that the process of determining the listing price is very important given that, although buyers can negotiate the final price with sellers, this one often ends up being equal or very close to the quoted price. Unsurprisingly, we see that the price discount distribution is very asymmetric around zero with most of the properties being sold below the listed price and only about 12% being sold at a premium. The average and median properties in the sample sell 4.68% and 3.83% below listed price, respectively[17]. With regard to TOM, Figure 3b

[16]Discussions with real-estate agents and Zoopla information suggests that, due to the lengthy conveyancing process, it on average takes about six weeks to complete a freehold sale and eight to ten weeks a leasehold one, but that this can go down to as little as a couple of weeks. As the Price Paid data contains the date when the sale has been completed, I take a conservative approach and remove occurrences with less than four weeks between listing and completion.

[17]The average and median discount with respect to the final quote equal -3.06% and -2.73%, respectively, which suggests that most price changes are likely to be downward revisions.



(a) Discount with respect to First Quote



(b) Time on the Market (weeks)

Figure 3 Histograms of Price Discount and Time on the Market for Matched Listings
The figure displays the distributions of price discount and time on the market (TOM) for the set of property listings that were matched to their respective ex-post transactions in the sample from 2009 onward. Figure 3a plots the histogram of the percentage difference between the first listed price and the final transaction price, while Figure 3b shows the histogram of time on the market measured as the number of weeks since the property was first listed. The mean and median values of the two distributions are represented by the green and blue vertical lines, respectively.

shows that properties sell within 28 weeks on average, with the median property selling within 21 weeks since listing[18]. It is also reassuring to see that 99% of the properties in the sample sell within no more than two years which confirms the matching quality. In Figure 2 of the Appendix I plot the time-series of the average and median price discount and TOM. We can note that there is a positive correlation between discount and time spent on the market although it seems that TOM is less sensitive to market conditions.

Table 2 provides summary statistics for the set of listings and their respective subsequent transactions in the restricted sample used in the later regressions, i.e., listings post March 2012 that have at least one comparable transaction in the prior month. Contrasted to the full sample of listings and transactions in Table 1, the matched sample is pretty similar across all characteristics, although it contains around 3% more houses, and consequently larger properties, at the expense of flats compared to the post 2012 sample in Table 1. The mean and median price discount equal -3.78% and -3.13%, with the average and median TOM being 26 and 21 weeks, respectively. Coupled with the results in Figure 2 in the Appendix, we can conclude that the majority of the observations in the test sample come from periods of hot housing markets. Thus, it would be interesting to see how agents behave in response to new information in times of moderate to good market conditions and contrast this with findings of previous papers that have focused mostly on times of depressed housing markets (Anenberg, 2016).

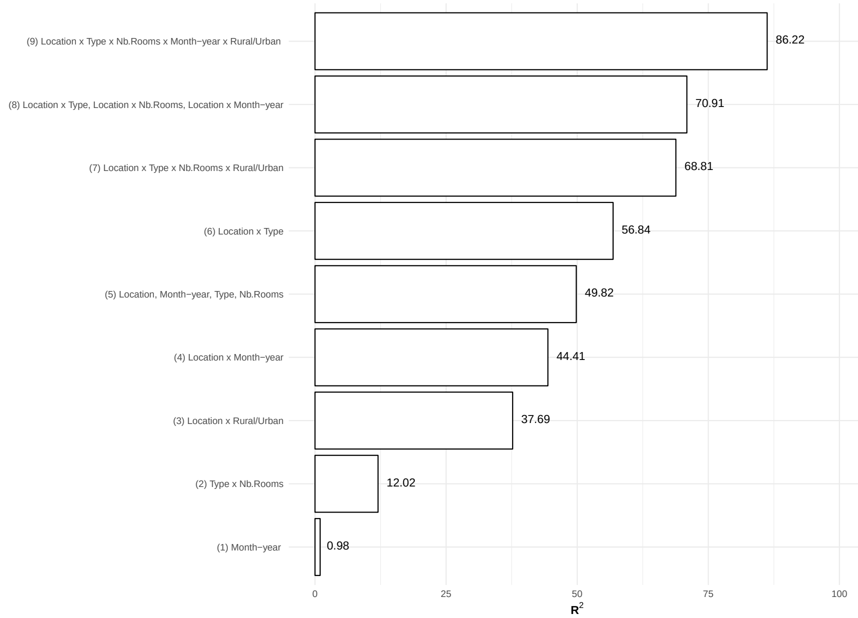
To provide further evidence on the effectiveness of looking for comparable transactions that match across location, property type, number of rooms and time, I next show the fraction of the variation in prices that is explained by various characteristics. Figure 4 displays the R-squared obtained by regressing prices on the above fixed effects, separately for listings and transactions. I sequentially increase the number of fixed effects in the regressions in order to discern the incremental improvement in explanatory power. Comparing Figures 4a and 4b, we can note that the explanatory power of the various

[18]It is important to bear in mind that this is the time difference between listing and sale completion. Taking into account the average time it takes to finalise a sale, we can conclude that the average (median) seller finds a buyer in about 21 (14) weeks.

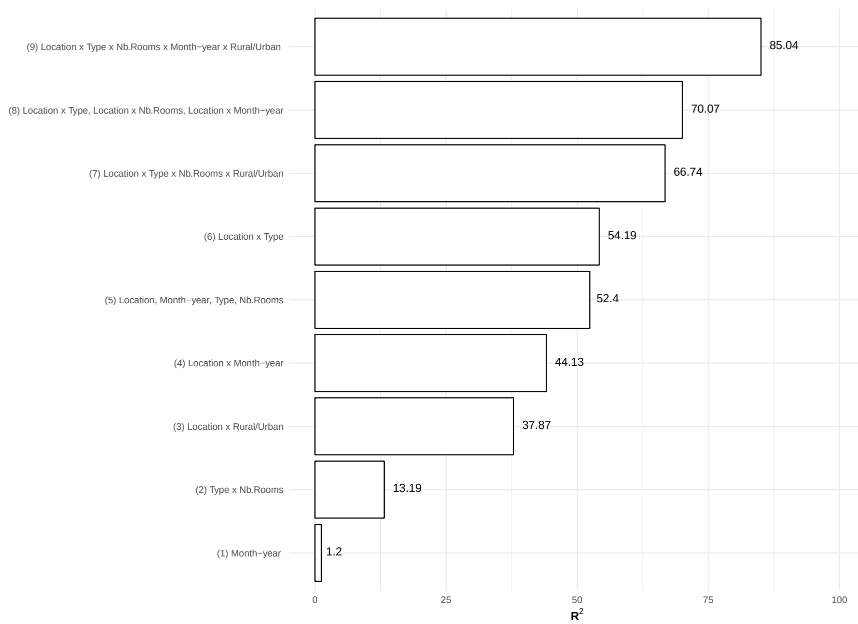
Table 2 Listings Matched to Subsequent Transactions - Summary Statistics

The table presents summary statistics for the set of listings that have been matched to their respective ex-post transactions in the period from March 2012. Listing price is the first quote at which a property has been listed, while transaction price is the final agreed price between the buyer and the seller. Price discount is the percentage difference between the listed price and the final transaction price. TOM is time on the market measured in weeks. Property type refers to the built-form of the property which can be one of four possible categories: detached, semi-detached, terraced house or a flat. Number of rooms refers to the total number of habitable rooms in the property. I report the following statistics on the distribution of prices, price discounts, time on the market and number of rooms: mean is the average value, min is the lowest value, p25, median and p75 are the 25-th, 50th and 75-th percentile of the distributions, respectively, and max is the highest value observed in the sample. For property type, I report the fraction of observations that are of a given type. Nb. of Observations refers to the total number of transactions in the matched sample.

Listing Price		*Transaction Price*	
Mean	£255,038	mean	£245,569
Min	£12,000	min	£11,000
P25	£134,950	p25	£127,000
Median	£190,000	median	£186,500
P75	£299,950	p75	£290,000
Max	£15,000,000	max	£16,200,000
Price Discount (%)		*TOM (weeks)*	
Mean	-3.78	mean	26.48
Min	-63.52	min	4.00
P25	-6.41	p25	14.43
Median	-3.13	median	20.57
P75	0.00	p75	31.43
Max	1.16	max	256.86
Number of rooms		*Property type (%)*	
Mean	4.62	Detached	15.14
Min	1	Semi-detached	32.16
P25	4	Terraced	35.28
Median	5	Flat	17.42
P75	5		
Max	16		
Nb. of Observations		1,067,282	



(a) Listing Prices



(b) Transaction Prices

Figure 4 Fraction of Explained Variation in Prices

The figure displays the percentage of the variation in the price data that is explained by observable characteristics, measured as the R-squared from a regression of prices on various fixed effects. Figure 4a shows the variation explained for the listing data and Figure 4b for the transaction data. Fixed effects included are: month-year of the listing or transaction; property type (detached, semi-detached, terraced house or a flat); number of rooms in the property, where properties with between 6 and 10 rooms are placed in one bucket and properties with more than 10 rooms in another; location, measured as the address outcode, and; a rural/urban area indicator from the 2011 Census classification of Output Areas.

property characteristics and time effects is very similar for listing and transaction prices. Starting from the bottom, the date of the listing/transaction measured in months explains about 1% of the price variation - this is not surprising as we are comparing houses of very different types, size and location across the entire England. The combination of house innate characteristics such as type and number of rooms has a considerably larger explanatory power of about 12-13%. Location is by far the most important determinant of house prices, explaining close to 38% of the variation in prices alone, close to 45% of this variation when coupled with time effects and about 55% when combined with property type effects. This is in line with the well-known fact that location is the key feature driving property values. As we increase the number of fixed effects the R-squared gradually increases, reaching over 85% for the full set of characteristics used in the matching exercise. This result gives validity to the comparables search method I employ in my empirical work by re-affirming the assertion that the pairs of listing and transaction prices I use in the regressions are indeed largely driven by common variables. Consequently, any incremental effect of recent transaction prices on listings in the post-publication period that I find in the data would arise mainly on the account of changes to the informaton set of prospective sellers. For space reasons, Figure 3 in the Appendix shows the explanatory power that the same set of characteristics have for the variation in the absolute and percentage price discount for the sample of transactions matched to their respective listings. We can note that the R-squared is considerably lower across most specifications and, in particular, time-invariant house effects explain just 25% of the variation in the dollar differential and less than 13% of the variation in the percentage price discount. Location is a much less important factor for price discounts explaining less than 10% of the variation. On the other hand, in contrast to the variation in price levels, the variation in price differences is much more significantly driven by time effects, re-asserting the conclusion that the price discount is an aggregate feature of the housing market that evolves similarly across properties of different types and location. The combination of all fixed effects achieves an R-squared of 78% and 68% for the dollar and percentage price difference, respectively.

Before I proceed to present the results of my empirical analysis, I will briefly investigate any potential selection biases that we might have to be aware of. First of all, it is important to note that the final sample used in the regressions below considers only transactions that serve as comparables to at least one listing posted before and one after the date when the price data is released. In this way, we can avoid potential concerns regarding systematic differences in the independent variable between the sets of treated and untreated listings. The results in the next section should therefore be interpreted as the incremental effect that the same set of prices have on subsequent listings following the Price Paid data publication dates. Accordingly, the only reason why the effect might be different comes from the discontinuity in the information set of prospective sellers around these dates.

In Figure 5 I provide some evidence on the similarity across the sets of treated and untreated listings. Namely, I regress listed prices on a dummy for the signed number of days between the listing and the publication date of the latest price data, by adding the usual fixed effects used in the matching process[19]. The figure shows that there is little variation in the prices at which properties are listed, controlling for house characteristics. On all days but two we cannot reject the null hypothesis that prices insignificantly differ relative to those of properties listed on the publishing date. Even on the two days where this difference is significant, it is never larger than £2,000. Furthermore, in untabulated analysis I regress listing prices on a dummy for treated, that is, I compare the price levels of listings occurring before and after the publication date[20]. The results shows that there is no significant trend in listing prices around publishing dates: the coefficient on treated suggests that listing prices in the post publication periods are about £300 pounds larger,

[19]Figure 4 in the Appendix plots the results of the same test for the sample period before March 2012 used in some of the robustness checks below.

[20]Specifically, I run the following regression: $q_i = \alpha + \beta \times Treated_i + FE + \varepsilon_i$, where the fixed-effects correspond to the characteristics the matching is based on, i.e., location, property type, number of rooms and month-year, and $Treated_i$ is a dummy that equals zero for listings that occur in the two weeks before the new transaction data is published and one for those that are posted in the two weeks after.

however, the p-value is above 0.15.

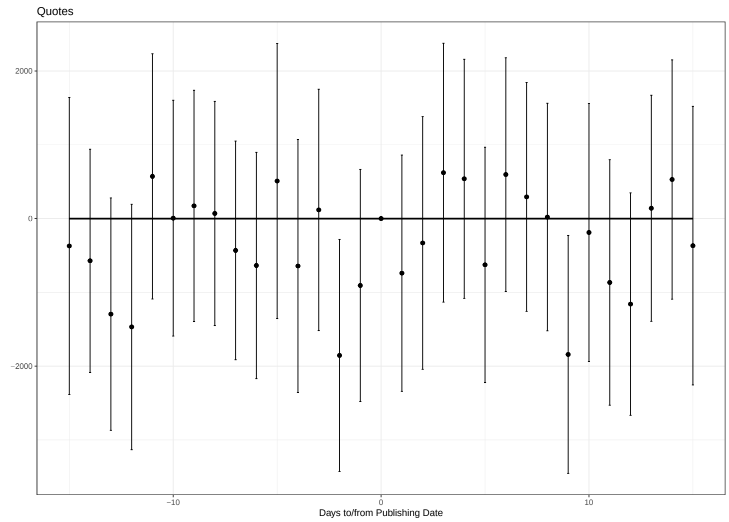


Figure 5 Variation In Listing Prices Around Publishing Dates
The figure plots the results from a regression of listing prices on dummies for the signed number of days between the listing date and the price data publication date for the sample after March 2012. The regression is specified as follows: $q_i = \alpha + \sum_{\Delta=-15}^{15} \gamma_\Delta + FE + \varepsilon_i$, where the fixed-effects correspond to the characteristics the matching is based on, i.e., location, property type, number of rooms and month-year, and Δ is a dummy for the signed difference in days between the date on which a listing is posted and the publication date. The baseline coefficient is the one for listings posted exactly on the publication date. The vertical lines represent the 95% confidence bounds of the point estimates for the average listing prices.

Finally, I investigate any potential selection into treatment by running a density test. Specifically, I test for any potential discontinuity in the density of observations in the days around price publishing dates. Examining the Zoopla data in more detail, however, shows that listings exhibit a strong pattern in terms of week days, with a lot of activity in the middle of the week (Tuesday through Friday) and significantly less listings being posted on weekends and Mondays. For this reason, I first regress the count of observations per date on days of the week dummies and conduct a McCrary test (McCrary, 2008) using the residuals from this regression. The results of the density test are shown in Figure 6. As is evident from the picture, there does not seem to be manipulation of the running variable around price data publication dates. The shape of the density function is fairly smooth without exhibiting a jump on the treatment day. This is confirmed by the

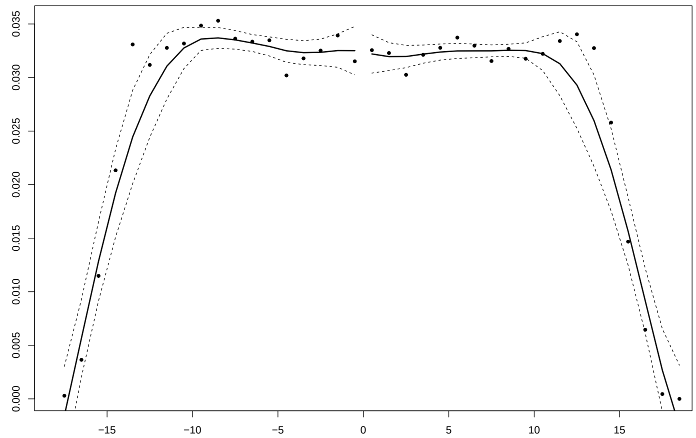


Figure 6 Density of Listing Observations Around Publishing Dates
The figure displays the smoothed density of the number of listing observations per day around price data publication dates for the sample after March 2012, where I fit two polynomials on each side of the publication date. The total daily count is first regressed on day-of-the-week dummies and the residuals of this regression are used for the density test.

p-value of the test which is equal to 0.638, well above the significance threshold.

Equipped with the above reassuring evidence, I will now proceed to the next section where I present the empirical results of my study of the impact of comparables on the price behaviour of agents in the housing market.

5 Results

The main goal of this paper is to investigate if prospective sellers in the housing market are able to extract and use information from past prices in the optimal way or, perhaps due to the complexity of the chains of inter-influence among recent comparables, they are prone to double counting repeated information at the expense of real news. For this reason, I will first provide some evidence that the comparables pricing approach is indeed used in this market by exploiting the shock to sellers' information sets that occurs on each Price Paid data publication date. I will then present results on the indirect effect of past prices on future listings that occurs due to repeated use of this approach by a sequence of sellers. Finally, I will address the question of whether any mistakes made due to suboptimal learning in such an imperfect environment are corrected at the selling stage by looking at the sample of listings matched to their respective subsequent transactions.

5.1 Evidence of Pricing by Comparables in the Housing Market

Before analysing how agents process newly released information when setting house prices we need to make sure that past prices of similar properties significantly affect their decisions. The housing market in the UK provides a natural experiment for testing this hypothesis, namely, whether agents behave differently after they have been exposed to the most recent set of transactions in their market of interest. Recall from Section 2 and Figure 1 that, starting from March 2012, the Land Registry publishes monthly transaction data for the previous month on the twentieth working day of each month. On these dates, prospective sellers receive a shock to their information set. Specifically, in the days leading to the publication date, sellers, real-estate agents and other property professionals have access to pricing information only if they have been directly involved in the transaction or if they have access to other sources of private information. After the publication date, everybody can potentially observe the full set of transactions that have taken place in the previous month. In other words, individuals who list their properties before the twentieth working day of the month may not directly observe the prices at which similar properties have been sold in the past month, while those who do so after the publishing date will have access to this information. Notice that while sellers may not be aware of the release of information, recent prices are usually immediately incorporated in the statistics available on common property platforms such as Zoopla. This implies that the seller becomes inadvertently a user of the newly released data as long as he is guided by the information on these platforms. The sellers' lack of knowledge regarding the publishing dates makes the discontinuity in the information set less sharp but alleviates the concern that sellers strategically select when to list which further explains the results in

Section 4. Together with the fact that sellers might have access to private information about recent transactions, this consideration implies that the findings in this section will represent a lower bound of the true effect of newly released information on prices.

I will start by comparing the effect of transaction prices from the past month on listings around the publication date. In particular, I match each listing price by the date on which it has been posted to its closest price publication date. In this way, listing prices posted in the roughly two weeks before the closest publication date do not observe the newest set of data and are thus untreated. Those posted in the days after the most recent publication date are, by contrast, able to observe the latest set of pricing data and are therefore treated. The matching is done by following a natural approach mirroring that of a prospective seller, i.e., looking at prices at which properties comparable in location (measured by the first half of the postcode and by an indicator from the 2011 Census rural-urban classification of Output Areas), property type (flat, detached, semi-detached or a terraced house) and number of rooms have sold in the past month[21].

Table 3 displays the results of the following regressions:

$$log(q_i^{pre}) = \alpha + \beta^{pre} \times log(p_j) + Controls + \varepsilon_i \tag{13}$$

$$log(q_i^{post}) = \alpha + \beta^{post} \times log(p_j) + Controls + \varepsilon_i \tag{14}$$

$$log(q_i) = \alpha + \beta_0 \times log(p_j) + \beta_1 \times log(p_j) \times Treated_i + \gamma Treated_i + Controls + \varepsilon_i \tag{15}$$

where q_i is the initial listed price for property i, p_j is a transaction price for a comparable property j that has been or will be published on the closest publication date and $Treated_i$ is a dummy that turns on if the listing has been posted after the price publication date of the given month. The data runs from March 2012 to May 2018 as this is the sample period during which the Land Registry has been publishing price data monthly on a regular basis. I keep only comparable prices that have at least one treated and one

[21] The same approach of matching comparable properties is followed throughout the rest of the analysis.

Table 3 Effect of Transaction Prices on Quotes Around Publishing Dates

The table presents the results of the following regression: $log(q_i) = \alpha + \beta \times log(p_j) + Controls + \varepsilon_i$, where q_i is the listed price for property i and p_j is the transaction price for a comparable property j sold in the previous month. Columns (1) and (2) present the results of running separate regressions for the set of untreated and treated listings, where treated listings are those that are able to directly observe the most recent price data. Columns (3)-(6) combine the two samples in a single regression of the following form: $log(q_i) = \alpha + \beta_0 \times log(p_j) + \beta_1 \times log(p_j) \times Treated_i + \gamma Treated_i + Controls + \varepsilon_i$, where Treated is a dummy that turns on when the listing price has been set in the period following the price data publication date. Controls for the time distance between the listing and the comparable transaction measured in days and its interaction with log price are included in all columns but (5). Column (5) instead includes time distance (measured in weeks) dummies and their interaction with log price. Fixed-effects included are: listing month-year dummies in columns (4)-(6), and; transaction ID dummies in column (6). Standard errors double-clustered at the transaction and listing ID levels are reported in parentheses.

	Untreated	Treated	Full Sample			
	(1)	(2)	(3)	(4)	(5)	(6)
Price × Treated			0.0045***	0.0045***	0.0039**	0.0034***
			(0.0015)	(0.0015)	(0.0015)	(0.0011)
Price	0.8401***	0.8446***	0.8401***	0.8402***	0.8367***	
	(0.0026)	(0.0036)	(0.0023)	(0.0023)	(0.0030)	
Treated			-0.0550***	-0.0548***	-0.0459**	-0.0387***
			(0.0185)	(0.0185)	(0.0183)	(0.0134)
Controls						
Price x Time distance	Yes	Yes	Yes	Yes	No	Yes
Price x Time distance dummies	No	No	No	No	Yes	No
Fixed-Effects						
Month-year	No	No	No	Yes	Yes	Yes
Transaction ID	No	No	No	No	No	Yes
Observations	3,698,564	3,768,386	7,466,950	7,466,950	7,466,950	7,466,950
R^2	0.7028	0.7056	0.7043	0.7050	0.7050	0.8689
Within R^2	–	–	–	0.7032	0.7032	0.0000

Two-way (Transaction ID & Listing ID) standard-errors in parentheses

*Signif Codes: ***: 0.01, **: 0.05, *: 0.1*

untreated match to make sure that the set of prices affecting listings before and after publication dates is similar. Each listing has an average of 5.67 and a median of 4 comparable prices. Transaction prices on the other hand have an average of 3.09 treated and 3.03 untreated comparable listings, while the median number is 2 across the two sets. In column (1) I regress only the set of untreated quotes on the transaction prices of the previous month, while in column (2) I repeat the procedure for the set of treated listings. In both regressions, I control for the distance in days between the date on which the comparable transaction took place and the date of the subsequent listing and for the interaction between this distance and the transaction price in order to account for any trends in housing prices. The difference in the coefficients on the price variable then gives an indication of the extent to which past prices affect future listing price decisions once they become publicly available. Specifically, comparing the two coefficients allows us to isolate the correlation between past prices and future quotes that arises due to the evolution of common fundamentals from the effect of deliberately using information contained in past prices to learn about the state of the housing market and inform future decisions. In particular, the first regression shows that the baseline effect of past transaction prices on listed prices in the following month is 84%. The magnitude of this coefficients confirms the well-established fact that prices in the residential property market exhibit high persistence. The coefficient from the second regression, however, suggests that the mere fact of being able to directly observe the latest set of transaction price data increases this effect by additional 0.45%. This incremental effect is statistically significant at the one percent level with an F-statistic of 7.6052.

Columns (3)-(6) provide additional evidence of this result by running the regression specified in equation (15) on the full sample of treated and untreated listings. Column (3) controls for the time distance and the potential differences in the way that prices affect future listings across different distances, similarly to the regressions in the first two columns; column (4) also adds a month-year fixed effect to account for the average level of listed prices across different periods; column (5) introduces a different control for time distance, namely, it allows for non-linear effects of prices on future listings by interacting the price with dummies for time distance measured in weeks, and finally; in column (6) I add a transaction ID fixed effect, in addition to the previous controls, which allows me to account for common unobservables across listings matched to the same transaction. The results are robust across all specifications, namely, the incremental effect of prices on future quotes set after the data becomes publicly available remains at 0.45% and statistically significant. Even when comparing the effect by controlling for average levels of quotes matched to the same transaction to alleviate concerns that some transactions might be matched to disproportionately more treated or untreated listings, the effect retains both its statistical and economic significance at 0.34%. Tables 1-4 in the Appendix provide additional evidence of the direct effect of newly published transactions: Table 1 refines the sample by varying the interval of time around publication dates in which quotes are considered and by limiting the number of comparables in order to make sure that the results are not driven by a small number of listings with too many comparable transactions; Table 2 shows robustness to the timing of publication dates, i.e., it controls for day-of-the-week of publications and whether these have occured at the end of the month or the beginning of the subsequent month; Table 3 focuses on existing properties only and investigates if the effect is different across properties of different price range; Table 4 finally adds real-estate agent fixed effects to make sure that the comparables pricing effect is not absorbed by varying business practices across agents. All these refinements largely confirm the economic and statistical significance of the effect of news release on future listings.

The results so far display evidence of the use of the comparables method in the residential housing market and give sense of the magnitude of its influence on future quotes. One might, however, be worried that the set of prospective sellers who choose to list their properties in the days following a publication date is different from the set of sellers who do so in the days before or that properties are systematically different along some unobservable dimension. To alleviate this concern, I next include in the sample all price updates in addition to the original quote for each listing and the dates at which these have occurred. This allows me to control for any potential unobservable differences in the immutable characteristics of prospective sellers or their properties by adding a listing ID fixed effect. Table 4 displays the results of this exercise which follows the regression specified in equation (15). In columns (1) and (2) I rerun the regression specifications from columns (4) and (6) of Table 3, respectively, without listing ID fixed effects in order to show that the treatment effect is similar in this extended sample: being able to directly observe the latest set of transaction data increases their effect on subsequent listings by about 0.43% and 0.30% once we add comparable fixed effects. Columns (3)-(6) introduce listing ID fixed effects to the regressions - in this way, the effect of the treatment is estimated solely by using the set of listings that have had at least one price change. As before, I control for the time distance in days between the transaction and the subsequent quote update and for its interaction with price in column (3); I add a quote month-year fixed effects in column (4); in column (5) I replace the usual time distance control with dummies for time distance measured in weeks, and; in column (6) I include transaction ID fixed effects in addition to month-year and listing ID fixed effects. The magnitude of the coefficients of interest naturally decreases as most of the variation is explained by the listing ID fixed effects, however, they remain statistically significant at about 0.08%. Table 5 in the Appendix displays the results of the same regression on the restricted sample that includes only listings with more than one quote available. The incremental effect of transactions after the publication date is this time even larger at about 0.11% to 0.33% depending on the specification, with listing ID fixed effects included.

The results presented above point to the idea that the monthly publication of transaction prices is a salient feature of the UK residential housing market that significantly affects prospective sellers' behaviour. To provide further evidence that the effect of prices around the publication dates is indeed systematic and not coincidental, I will next conduct a few robustness tests that are meant to rule out alternative hypotheses. For instance, one might think that we would observe a similar pattern in listing price

Table 4 Dirrect Effect of Transaction Prices on Quotes - Including Listing Price Updates

The table presents the results of the following regression: $log(q_i) = \alpha + \beta_0 \times log(p_j) + \beta_1 \times log(p_j) \times Treated_i + \gamma Treated_i + Controls + \varepsilon_i$, where q_i is the listed price for property i, p_j is the transaction price for a comparable property j sold in the previous month and Treated is a dummy that turns on when the listing price has been set/updated in the period following the most recent price publication date. Controls for the time distance between the listing and the comparable transaction measured in days and its interaction with log price are included in all columns but (5). Column (5) instead includes time distance (measured in weeks) dummies and their interaction with log price. Listing ID fixed effects are included in specifications (3)-(6). Additional fixed-effects include: listing month-year dummies in all columns but (3) and; transaction ID dummies in columns (2) and (6). Standard errors double-clustered at the transaction and listing ID levels are reported in parentheses.

	(1)	(2)	(3)	(4)	(5)	(6)
Price × Treated	0.0043***	0.0030***	0.0008***	0.0008***	0.0008***	0.0025***
	(0.0012)	(0.0009)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Price	0.8418***		-0.0012***	−0.0001		
	(0.0020)		(0.0002)	(0.0001)		
Treated	-0.0521***	-0.0342***	-0.0135***	-0.0080***	-0.0080***	-0.0331***
	(0.0150)	(0.0103)	(0.0027)	(0.0024)	(0.0024)	(0.0023)
Controls						
Price × Time distance	Yes	Yes	Yes	Yes	No	Yes
Price × Time distance dummies	No	No	No	No	Yes	No
Fixed-Effects						
Listing ID	No	No	Yes	Yes	Yes	Yes
Transaction ID	No	Yes	No	No	No	Yes
Month-year	Yes	Yes	No	Yes	Yes	Yes
Observations	11,410,244	11,410,244	11,410,244	11,410,244	11,410,244	11,410,244
R^2	0.7080	0.8695	0.9985	0.9989	0.9989	0.9994
Within R^2	0.7053	0.0000	0.0011	0.0002	0.0002	0.0156

Two-way (Transaction ID & Listing ID) standard-errors in parentheses

*Signif Codes: ***: 0.01, **: 0.05, *: 0.1*

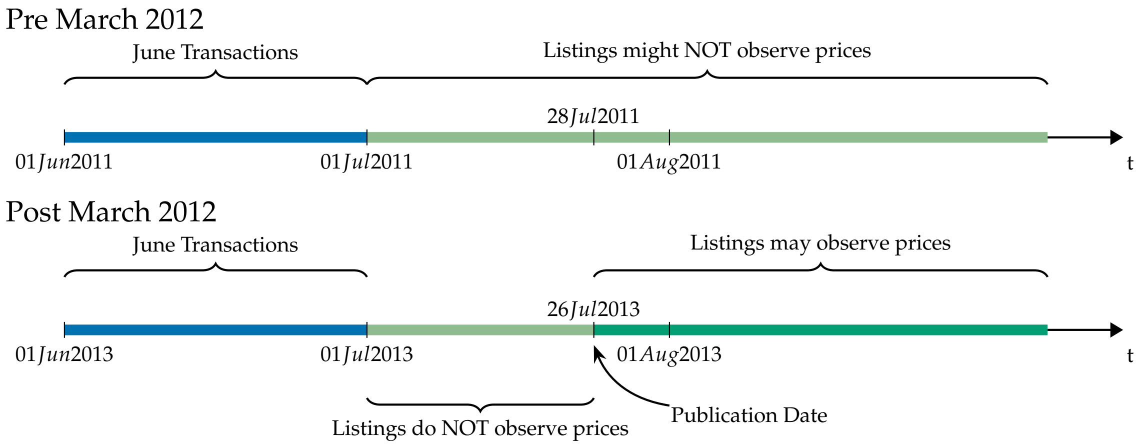


Figure 7 Difference-in-difference analysis: Example of a change to the information set of sellers around publication dates before and after March 2012
The figure presents the change to the institutional setting that occurred in March 2012: beginning in March 2012, the Land Registry publishes regular monthly data on transactions on the twentieth working day of the subsequent month. The second figure shows that transaction prices, depicted in blue, from June 2013, are published on the twentieth working day of July which is 26th July in this case. The property listings published at the beginning of June and before the publication date, depicted in light green, do not observe the data on June transactions, while those published after this date, depicted in dark green, may observe June price data and therefore can use this to make inference about market demand. The first figure shows that in the period before March 2012, there was no such regular shock to the information set. For example, although the twentieth working day of July 2011 was 28th July, the June transactions were not made publicly available on this date. The listings that occurred throughout the months of July and August 2011, therefore, are depicted in light green as they might not observe the June transaction data or, at least, not have this information arrive at the same regular intervals.

behaviour if housing prices were very persistent and had a trend, even if publication dates did not matter. Although the time controls should account for this possibility, to reject this hypothesis with more confidence, I will employ two strategies: first, I will conduct a difference-in-differences analysis whereby I take advantage of the large sample of data available and compare the effect of transactions on listings around publication dates prior to March 2012 and thereafter; second, I will show that the effect is no longer present if I shift the publication date to a few days before or after the actual one. The results of this analysis are presented in Tables 5 and 6. The sample I use for these tests includes quote updates, however, for robustness reasons, I provide the results of the same analysis when only the original quote for each listing is included in Tables 6 and 7 of the Appendix.

Figure 7 illustrates the idea behind the first of these tests through an example. Although the Price Paid data was available to purchase under a licence from the Land Registry prior to March 2012, this was done at the discretion of the real estate agencies and other property data providers. This means that firms could get access to the data at varying dates that would most likely not always coincide with the twentieth working day of each month. As a result, there should not be a significant increase in the effect of past prices on future listings around the hypothetical publishing dates in the period before March 2012. The regressions in Table 5 make use of this fact by comparing the effects of prices on listings around publication dates before and after March 2012 via a difference-in-differences approach. Take, for instance, the twentieth working days of July 2011 and July 2013: these fell on July 28th in 2011 and July 26th in 2013. Transaction data from June of the same year were made publicly available in 2013 but not in 2011. If we compare the effect of June prices on listings before and after July 26th 2013 to that on listings before and after July 28th 2011, we would be able to eliminate any systematic variation between listings around different periods in a given month, assuming these do not dramatically change in the years after 2012. The results of the following regression are presented in Table 5:

$$\begin{aligned} log(q_i) =& \alpha + \beta_0 \times log(p_j) + \beta_1 \times log(p_j) \times Post\ March\ 2012_i + \beta_2 \times log(p_j) \times Treated_i + \\ & \beta_3 \times log(p_j) \times Treated_i \times Post\ March\ 2012_i + \gamma_1 Post\ March\ 2012_i + \gamma_2 Treated_i + \\ & \gamma_3 \times Treated_i \times Post\ March\ 2012_i + Controls + \varepsilon_i \end{aligned} \tag{16}$$

where q_i is the listing price for property i, p_j is the price at which a comparable property has been transacted in the previous month, $Treated_i$ is a dummy that turns on for listings posted on or after the twentieth working day of the month and $Post\ March\ 2012_i$ is a dummy that turns on starting from March 2012. As usual, column (1) includes controls for the time distance between the listing and the comparable transaction and its

Table 5 Direct Effect of Transaction Prices on Quotes - Before vs After March 2012
The table presents the results of the following regression: $log(q_i) = \alpha + \beta_0 \times log(p_j) + \beta_1 \times log(p_j) \times Post\ March\ 2012_i + \beta_2 \times log(p_j) \times Treated_i + \beta_3 \times log(p_j) \times Treated_i \times Post\ March\ 2012_i + \gamma_1 Post\ March\ 2012_i + \gamma_2 Treated_i + \gamma_3 \times Treated_i \times Post\ March\ 2012_i + Controls + \varepsilon_i$, where q_i is the listed price for property i, p_j is the transaction price for a comparable property j sold in the previous month, Post March 2012 is a dummy that equals one for listings published starting from March 2012 and Treated is a dummy that turns on when the listing price has been set/updated in the period following the most recent price publication date. Controls for the time distance between the listing and the comparable transaction measured in days and its interaction with log price are included in all columns but (3). Column (3) instead includes time distance (measured in weeks) dummies and their interaction with log price. Fixed-effects included are: listing month-year dummies in all columns but (1) and; transaction ID dummies in column (4). Standard errors double-clustered at the transaction and listing ID levels are reported in parentheses.

	(1)	(2)	(3)	(4)
Price × Treated × Post March 2012	0.0075***	0.0072***	0.0072***	0.0040**
	(0.0021)	(0.0021)	(0.0021)	(0.0018)
Price × Treated	-0.0041**	-0.0038*	-0.0043**	-0.0010
	(0.0020)	(0.0020)	(0.0020)	(0.0016)
Price × Post March 2012	0.0513***	0.0508***	0.0508***	
	(0.0020)	(0.0020)	(0.0020)	
Price	0.7889***	0.7893***	0.7875***	
	(0.0024)	(0.0024)	(0.0028)	
Treated	0.0459*	0.0417*	0.0492**	0.0102
	(0.0236)	(0.0235)	(0.0235)	(0.0190)
Post March 2012	-0.6188***	-0.6381***	-0.6382***	
	(0.0241)	(0.0247)	(0.0248)	
Treated × Post March 2012	-0.0878***	-0.0839***	-0.0838***	-0.0443**
	(0.0246)	(0.0245)	(0.0245)	(0.0216)
Controls				
Price × Time distance	Yes	Yes	No	Yes
Price × Time distance dummies	No	No	Yes	No
Fixed-Effects				
Month-year	No	Yes	Yes	Yes
Transaction ID	No	No	No	Yes
Observations	16,367,900	16,367,900	16,367,900	16,367,900
R^2	0.6805	0.6814	0.6814	0.8565
Within R^2	–	0.6782	0.6782	0.0000

Two-way (Transaction ID & Listing ID) standard-errors in parentheses
*Signif Codes: ***: 0.01, **: 0.05, *: 0.1*

interaction with price; column (2) adds month-year fixed effects; column (3) replaces the linear time distance control with dummies for time distance between the quote and the price measured in weeks, and; column (4) adds transaction ID fixed effects. It is interesting to see that the coefficient of interest in row 1 almost doubles compared to Table 3 and is now around 0.75%. Coupled with the second row coefficients that display the effect of prices on listings after hypothetical publication dates before March 2012, we can conclude that the net effect of recently published prices on future listings is close to 0.35%. It is important to emphasise that the correlation between recent transaction prices and quotes decreases after the twentieth working day of the month before March 2012 as evidenced by the negative coefficients in the second row. This is not surprising as the ″treated″ listings naturally come after the ″untreated″ ones and we thus should expect that quotes closer to recent transactions have more correlated fundamentals than those further in the future. This result further strengthens the conclusion that publication dates provide a salient enrichment of the information set of prospective sellers that they incorporate into their listing behaviour. In other words, the results presented in Table 3 above can be thought of as the lower bound of the direct effect of past prices on future listings that arises due to comparables pricing. The third row coefficients show that prices are generally more correlated in the period post March 2012 which suggests that agents might now have a more frequent access to new data than before. Anecdotal evidence suggests that real-estate agencies used to purchase new Price Paid data less regularly such as every quarter or half a year. The shift to monthly price updates then presents an important increase in the frequency at which they would update their price forecasts and client advice. Finally, when we compare the effect within listings matched to the same transaction by adding transaction ID fixed effects in column (4), we see that the effect of prices on listings does not significantly change around the hypothetical publishing dates before March 2012, however, it does significantly increase post March 2012 by additional 0.4%.

Moving on to Table 6, I now conduct a second type of robustness checks of the im-

Table 6 Effect of Transaction Prices on Quotes Around Placebo Publishing Dates

The table presents the results of the following regression: $log(q_i) = \alpha + \beta_0 \times log(p_j) + \beta_1 \times log(p_j) \times Treated_i + \gamma Treated_i + Controls + \varepsilon_i$, where q_i is the listed price for property i, p_j is the transaction price for a comparable property j sold in the previous month and Treated is a dummy that turns on when the listing price has been set/updated in the week before (first four columns) or one week after (last four columns) the closest price publication date. Controls for the time distance between the listing and the comparable transaction measured in days and its interaction with log price are included in all columns but (3) and (7). Columns (3) and (7) instead include time distance (measured in weeks) dummies and their interaction with log price. Additional fixed-effects include: listing month-year dummies in all columns but (1) and (5) and; transaction ID dummies in columns (4) and (8). Standard errors double-clustered at the transaction and listing ID levels are reported in parentheses.

	7 days before				7 days after			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Price × Treated	0.0014	0.0012	0.0006	0.0000	0.0016	0.0012	0.0007	-0.0007
	(0.0016)	(0.0016)	(0.0016)	(0.0014)	(0.0016)	(0.0016)	(0.0016)	(0.0013)
Price	0.8457***	0.8451***	0.8428***		0.8436***	0.8435***	0.8429***	
	(0.0026)	(0.0026)	(0.0028)		(0.0035)	(0.0035)	(0.0033)	
Treated	-0.0208	-0.0192	-0.0106	0.0008	-0.0222	-0.0164	-0.0100	0.0086
	(0.0194)	(0.0194)	(0.0194)	(0.0162)	(0.0191)	(0.0190)	(0.0191)	(0.0160)
Controls								
Price × Time distance	Yes	Yes	No	Yes	Yes	Yes	No	Yes
Price × Time distance dummies	No	No	Yes	No	No	No	Yes	No
Fixed-Effects								
Month-year	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Transaction ID	No	No	No	Yes	No	No	No	Yes
Observations	4,569,583	4,569,583	4,569,583	4,569,583	4,705,129	4,705,129	4,705,129	4,705,129
R^2	0.7061	0.7069	0.7069	0.8785	0.7104	0.7113	0.7113	0.8811
Within R^2	–	0.7038	0.7038	0.0000	–	0.7085	0.7085	0.0000

Two-way (Transaction ID & Listing ID) standard-errors in parentheses

*Signif Codes: ***: 0.01, **: 0.05, *: 0.1*

portance of publishing dates. This time, I limit the analysis to the sample period starting from March 2012 and I vary the publishing dates by seven days back and forth from the actual ones. I then look at the difference in the effect of prices from the previous month in the two weeks around these placebo publication dates using the regression specified in equation (15). The first four columns show the results when I consider as treated the listings that are posted at most seven days before the actual publication date, while in the remaining four columns I consider as untreated the listings that occur in the first week after the actual publication date. I add the usual controls for time distance and its interaction with price as well as month-year and transaction ID fixed effects to make the tests comparable to those in Tables 3 and 4. Looking at row one of Table 6, we can note that the price effect does not significantly change after each of these two sets of placebo publication days. In particular, although listings are still strongly correlated with transaction prices from the previous month, this correlation does not increase in the week following the hypothetical publishing dates. This finding corroborates our previous conclusion that days on which new Price Paid data is made publicly available by the Land Registry do matter and they significantly affect the behaviour of prospective sellers.

I have so far provided evidence of the baseline effect of recent prices on subsequent quotes for properties up for sale. One might be led to think that this effect seems too small to be of any economic significance. It is worth remembering, however, that: (a) this is the effect of one single transaction, while most prospective sellers would look at multiple similar properties before making a pricing decision: to the extent that these comparables are driven by correlated signals their common component can be much more heavily over-weighted by future sellers; (b) I have been very conservative in my comparables search strategy by matching prices to listings only if they match across type, location (measured by the first half of the postcode), size proxied by number of rooms, a rural/urban area indicator, and if they occur in the roughly two weeks around the publication date, i.e., up to two months and a half following the sale; (c) the above estimates can be considered a lower bound for the true effect of comparables pricing due to the fact that the discontinuity around publication dates might not be perfectly sharp, but also due to the results from Table 5 where we see that the incremental effect of prices on future quotes is negative before March 2012. For all these reasons, the actual impact of past prices that results from the use of the comparables method is probably considerably larger. Moreover, I have not yet examined the way that the price effect changes with the number of interim channels of influence. To get a better idea of the total effect and its evolution, therefore, in the next section I will investigate the manner in which the direct influence from recent prices gets amplified through the sequential use of past observable data on listings and transactions. I will then use the obtained estimates to investigate the long-run effects of any mistakes on aggregate prices via a simple model of learning in Section 6.

5.2 Indirect Effect of Past Prices Through Intermediate Channels of Influence

In this section, I will explore the way that prospective sellers in the housing market process information they receive from other comparables when the information sets across comparables might not necessarily be independent. More specifically, as transaction and listing price data become available, new prospective sellers use them in order to learn private information about the state of housing demand. Suboptimal pricing behaviour can arise, however, if sellers do not appropriately account for the potential duplication of information: if everyone else takes the same approach, then the most recent prices contain information that is also embedded in older ones. Optimal learning requires sellers to distinguish between the new signal coming from most recent data and the part that has been influenced by previous prices and other commonly observed information in order to make correct inference. This issue is exacerbated by the fact that being able to appropriately extract all the different pieces of information that drive recent prices entails knowledge of the connections among past observations and the full structure of information flows.

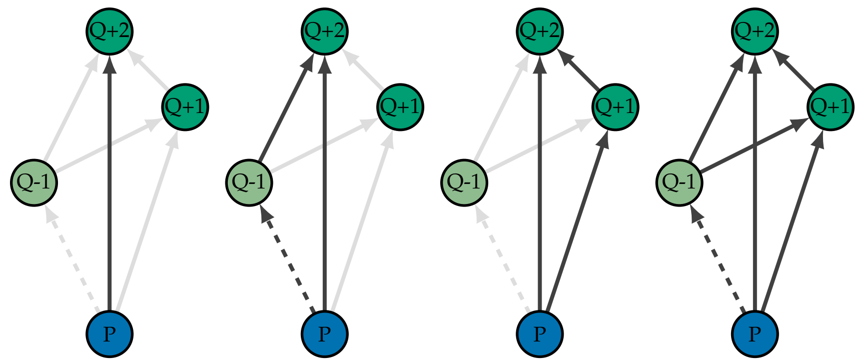


Figure 8 Indirect price effects I: multiple channels of influence from a given price to subsequent quotes through intermediate comparable listings
The figure provides an example of the possible channels through which a given observation might have an influence on subsequent sellers. The blue circles represent transaction prices from a given month, the light green ones are listings that were posted in the week before the price data becomes available (Q-1), while the dark green circles are listings posted in the first or second week following the publication date, Q+1 and Q+2, respectively. Focusing on the listings posted in week two after publication and their link to the transaction prices from the previous month, I show the four possible cases that can arise. Going from left to right, there may be: (a) no comparable listing posted in any of the two weeks surrounding the Price Paid data publication date; (b) comparable listings only in the week before but not the week after; (c) comparable listings only in the week after but not the week before, and; (d) at least one similar listing in both weeks.

For this reason, I next investigate the importance of indirect effects whereby past directly observable prices potentially affect future seller behaviour also through other intermediate comparables. The evolution of the effect that prices have on future listings as the number of intermediate channels of influence grows can be benchmarked against the Bayesian and naïve learning models presented in Section 2 in the main body of the paper and Section A in the Appendix in order to gain understanding about the way agents process information.

For the first set of tests, I focus on listings occurring in the second week after the publication of the latest transaction data that have at least one match within the set of sold properties. I then check the number of listings that are comparable to this pair in the week before and the week after the publication. Figure 8 depicts the four possible cases that can arise. Specifically, going from left to right, the matched pair might have: (a) no comparable listing posted in any of the two weeks surrounding the Price Paid data publication date; (b) comparable listings only in the week before but not the week after; (c) comparable listings only in the week after but not the week before, and; (d) at least one similar listing in both weeks. Note that, although listings in the week before the data is published do not directly observe the recent transaction prices, they might still be correlated due to commonly observed fundamentals. Prospective sellers in the week after, on the other hand, are able to directly observe the latest transaction data and so they have a second channel of influence. Looking at listings posted in the second week following the price publication date, therefore, allows us to test if agents are able to disentangle the different pieces of information embedded in a new observation and thus avoid double-counting redundant news. In particular, Section A in the Appendix shows that, conditional on agents directly observing a given transaction and its price, Bayesian updating implies that its effect on future quotes should be monotonically decreasing with the number of intermediate links between the two, i.e., the covariance between the two is expected to decline with the arrival of new information as agents optimally place lower weight on each individual signal. Conversely, if agents are unaware of or unable to discern the different channels of influence, then we might see the effect of that same transaction increase with the number of intermediaries relative to the Bayesian case.

Table 7 displays the results of this analysis. Across the four columns, I divide the sample of listings posted in week two post publication into four groups corresponding to the cases in Figure 8 above and I regress the quotes on the prices of comparable transactions which have just been made available:

$$log(q_i^s) = \alpha_s + \beta_s \times log(p_j^s) + Controls + \varepsilon_i \tag{17}$$

where q_i is the listed price, p_j is the price at which a comparable property has been transacted the month before and s is an index that captures whether the quote-price pair has a comparable listing in the the week before, the week after or in both weeks around the publication date. In all regressions, I control for the number of comparable matches in each week and for the time distance between the price and the quote in question. Going

Table 7 Indirect Price Effects Through Intermediate Listings

The table presents the results of the following regression: $log(q_i^s) = \alpha_s + \beta_s \times log(p_j^s) + Controls + \varepsilon_i^s$, where q_i is the listed price for property i and p_j is the transaction price for a comparable property j sold in the previous month. The sample includes quotes that have been set/updated during the second week following the most recent price data publication date. Column (1) considers quotes with no comparable listings in the previous two weeks; column (2) quotes with comparable listings only in the week before the publication date; column (3) quotes with comparable listings only in the week after the publication date, and; column (4) quotes with comparable listings in both weeks. Controls for the time distance between the listing and the comparable transaction measured in days and its interaction with log price are included in all columns, as well as controls for the number of comparables in the current week and each of the two previous weeks. Standard errors double-clustered at the transaction and listing ID levels are reported in parentheses.

	No prior comps	Comps in wk -1	Comps in wk +1	Comps in all wks
	(1)	(2)	(3)	(4)
Price	0.8166***	0.8328***	0.8402***	0.8479***
	(0.0094)	(0.0103)	(0.0084)	(0.0054)
Controls				
Price x Time distance	Yes	Yes	Yes	Yes
Nb. of comps per week	Yes	Yes	Yes	Yes
Observations	384,735	332,801	486,689	1,728,033
R^2	0.6927	0.6947	0.7092	0.7110
Adjusted R^2	0.6927	0.6947	0.7092	0.7110

Two-way (Transaction ID & Listing ID) standard-errors in parentheses
*Signif Codes: ***: 0.01, **: 0.05, *: 0.1*

from column 1 to column 4, we can see that the price coefficient is monotonically increasing. Column 1 gives the effect of recently published prices on listings in the second week following the publication for the case where there are no comparable listings in the two weeks surrounding it. In particular, the effect of recent prices on listings of comparable properties two weeks post publication is about 81.66%. This effect can be thought of as the sum of the correlation that arises due to commonly observed signals and the direct effect of price j on quote i that results from the use of the comparables method. The following three columns then indicate the incremental effect that comes from the existence of additional links between the two. From column 2, we can infer that having a comparable listing in the week before the price data is published already increses this effect by additional 1.62%. That is, although the matched listing of week -1 does not directly observe price j, the fact that it is highly correlated to it because of commonly observed news makes agent i overweight this common signal when using both to inform his decisions. Column 3 shows an additional increase of 0.74% if the intermediate comparable is instead in the week post publication. Intuitively, the incremental effect here arises because, on top of being driven by common fundamentals, the intermediate listing now is able to directly observe transaction j and thus its price also embeds the private and common signals coming from price j. So a prospective seller who uses both transaction j and the intermediate listing of the week before as two independent pieces of information is likely to overweight old information such as the signal coming from transaction j and the common news. Finally, column 4 shows that having all possible chains of influence present raises the coefficient on price j to 84.79%, a 0.77% higher than in Column 3 and a striking 3.13% larger than the baseline effect in Column 1. This result further underscores our previous assertion that the direct effect estimated in Section 5.1 above is likely to be the lower bound of the overall impact that past prices have on future ones. Table 8 in the Appendix confirms these findings by presenting the results from running a single regression on the full set of week two listings by interacting the effect of price with a dummy for whether the quote-price pair has comparable listings in the two weeks surrounding the publication date. As evidenced by the coefficients in row 4 of this table, prices have on average more than 2.32% larger effect on future listings when multiple chains of influence are present. The coefficients in rows 2, 3 and 4 are statistically different from each other at the 5% level or higher as shown by the p-values obtained by running linear hypothesis tests. Table 9 in the Appendix provides some evidence that this effect does not evolve in the same way in the sample before March 2012. The difference in coefficients is not significant and, if something, the coefficient declines in row 4 where we have listings with the largest information set.

Before proceeding to an alternative experiment, I examine more closely the way that the effect from past prices might evolve with the number of intermediate comparables in a given week. Table 8 presents the results of the following regression:

$$\begin{aligned} log(q_i) = &\alpha + \beta_0 \times log(p_j) + \sum_k \beta_k \times log(p_j) \times k\ Comps\ in\ week\ n + \\ &\sum_k \gamma_k \times k\ Comps\ in\ week\ n + Controls + \varepsilon_i \end{aligned} \tag{18}$$

where q_i is the listed price for property i, p_j is the transaction price for a comparable property j sold in the previous month and k Comps in week n is a dummy that turns on when quote i has k comparable listings in week n, n being either the week before or the week following the publication date. The first three columns show how the price effect changes as the number of comparable quotes in the week before the publication date increases. Going from row 2 to row 5, we can see that the coefficients exhibits an upward trend; the difference between the effect when having one comparable and that on quotes having more than 3 is significant at about 1.05% in all specifications but (1) which does not control for the number of comparables in the other weeks. The last three columns show the effect as the number of comparables in the week post publication increases. This time the effect on the quotes in week two is much larger and it monotonically grows with the number of intermediate quotes as these directly observe the price and thus create additional redundant links between the two. The difference between the price effect

Table 8 Indirect Price Effects by Number of Intermediate Comparable Listings
The table presents the results of the following regression: $log(q_i) = \alpha + \beta_0 \times log(p_j) + \sum_k \beta_k \times log(p_j) \times$ k *Comps in week* $n + \sum_k \gamma_k \times k$ *Comps in week* n + *Controls* $+ \varepsilon_i$, where q_i is the listed price for property i, p_j is the transaction price for a comparable property j sold in the previous month and k Comps in week n is a dummy that turns on when quote i has k comparable listings in week n. The sample includes quotes that have been set/updated during the second week following the most recent price data publication date. Columns (1)-(3) investigate the indirect effect of past transactions on quotes based on the number of comparable listings in the week before the price data publication date, while columns (4)-(6) consider the indirect price effect based on the number of comparable listings in the first week after the publication date. Controls for the time distance between the listing and the comparable transaction measured in days and its interaction with log price are included in all columns, as well as controls for the number of comparables in each of the other weeks and their interaction with the log price in all columns but (1) and (4). Columns (3) and (6) also include listing month-year fixed effects. Standard errors double-clustered at the transaction and listing ID levels are reported in parentheses.

	$n=-1$			$n=+1$		
	(1)	(2)	(3)	(4)	(5)	(6)
Price	0.8330***	0.8367***	0.8358***	0.8277***	0.8333***	0.8323***
	(0.0041)	(0.0043)	(0.0043)	(0.0041)	(0.0043)	(0.0043)
Price × 1 comp in week n (1)	0.0080***	0.0090***	0.0089***	0.0121***	0.0137***	0.0135***
	(0.0024)	(0.0024)	(0.0024)	(0.0025)	(0.0025)	(0.0025)
Price × 2 comps in week n	0.0137***	0.0157***	0.0157***	0.0153***	0.0187***	0.0186***
	(0.0029)	(0.0030)	(0.0030)	(0.0028)	(0.0029)	(0.0029)
Price × 3 comps in week n	0.0123***	0.0154***	0.0146***	0.0199***	0.0252***	0.0248***
	(0.0036)	(0.0038)	(0.0038)	(0.0034)	(0.0036)	(0.0036)
Price × >3 comps in week n (2)	0.0132***	0.0199***	0.0194***	0.0233***	0.0359***	0.0350***
	(0.0031)	(0.0037)	(0.0037)	(0.0029)	(0.0035)	(0.0035)
1 comp in week n	-0.0990***	-0.1117***	-0.1122***	-0.1441***	-0.1636***	-0.1627***
	(0.0290)	(0.0294)	(0.0293)	(0.0297)	(0.0299)	(0.0299)
2 comps in week n	-0.1678***	-0.1930***	-0.1922***	-0.1835***	-0.2253***	-0.2256***
	(0.0350)	(0.0361)	(0.0360)	(0.0341)	(0.0350)	(0.0349)
3 comps in week n	-0.1460***	-0.1847***	-0.1764***	-0.2358***	-0.3009***	-0.2970***
	(0.0440)	(0.0459)	(0.0458)	(0.0414)	(0.0431)	(0.0429)
>3 comps in week n	-0.1490***	-0.2354***	-0.2275***	-0.2695***	-0.4238***	-0.4131***
	(0.0380)	(0.0447)	(0.0446)	(0.0349)	(0.0418)	(0.0417)
(2)−(1)	0.0052	0.0109***	0.0105***	0.0112***	0.0222***	0.0215***
p-value	(0.1193)	(0.0029)	(0.0046)	(0.0002)	(0.0000)	(0.0000)
Controls						
Price x Time distance	Yes	Yes	Yes	Yes	Yes	Yes
Price x Nb. of comps per week	No	Yes	Yes	No	Yes	Yes
Fixed-Effects						
Month-year	No	No	Yes	No	No	Yes
Observations	2,932,258	2,932,258	2,932,258	2,932,258	2,932,258	2,932,258
R^2	0.7067	0.7068	0.7077	0.7068	0.7069	0.7078
Within R^2	–	–	0.7050	–	–	0.7051

Two-way (Transaction ID & Listing ID) standard-errors in parentheses
*Signif Codes: ***: 0.01, **: 0.05, *: 0.1*

on quotes with only one comparable and those with more than 3 is between 1.12-2.22% and is significant at the 1% level across all specifications.

To further corroborate the above conclusions, I now propose a second approach to examining the way that baseline effects get overweighted by naïve learners over a sequence of listing prices. Figure 9, gives a graphic representation of the chain of interactions from a given transaction to subsequent comparable listings. The selling price is depicted in blue, listings matched to it that occur prior to its publication date are depicted in light green and those occurring after its publication date are in dark green. In an environment such as the housing market where agents act sequentially, observing a sequence of past prices requires agents to disentangle between various sources of information that drive recent actions. Recall from the first set of tests and Section A in the Appendix that rationality and full knowledge of the links across observations implies that agents should be able to extract the private signal coming from every new observation and avoid double counting information already embedded in prior actions. In other words, fully rational agents would not be disproportionately affected by past news based solely on the number of intermediate observations; we would thus expect to see the effect of a given transaction monotonically subside with the increase in the amount of new observations as a growing information set implies that each individual component gets a proportionally lower weight. Naïve agents, on the other hand, might fall in the trap of treating newly observed quotes as pure representations of independent information about demand; as the number of in-between links increases, naïve learners would therefore keep overweighting the information embedded in early prices relative to the Bayesian framework. To test this hypothesis, I next compare the effect of a given transaction on listed prices occurring in the month around its publishing date by order of match, i.e., for every quote, I control for the number of comparable quotes that happen before it and for whether the quote in question occurs before or after the publication date.

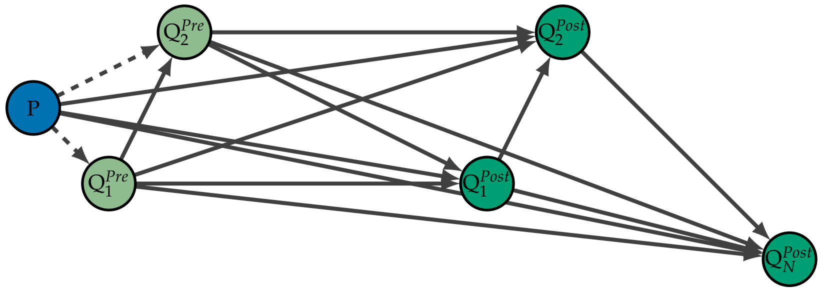


Figure 9 Indirect price effects II: the evolution of the effect of prices on future quotes as the number of comparables grows
The figure shows how a given observation can have more channels of indirect influence on future listings as the number of interim comparables grows. The blue circle repesents a given transaction price, the light green circles are listings posted before its publication date and the dark green circles are listings posted after. The listings are indexed in order to capture their chronological arrival in the market and provide an idea of the information set of all subsequent sellers.

Table 9 presents the results of this analysis. Specifically, I run the following regression:

$$\begin{aligned} log(q_i) =& \alpha + \sum_{k=1}^{10} \beta_k^{pre} \times log(p_j) \times Comp\ Order\ k\ Pre + \sum_{k=1}^{10} \gamma_k^{pre} Comp\ Order\ k\ Pre + \\ & \sum_{k=1}^{10} \beta_k^{post} \times log(p_j) \times Comp\ Order\ k\ Post + \sum_{k=1}^{10} \gamma_k^{post} Comp\ Order\ k\ Post + Controls + \varepsilon_i \end{aligned} \tag{19}$$

where q_i is the listing price, p_j is the price at which a similar property has been sold in the month before, *Comp Order k Pre* is a dummy that turns on if the listing occurs in the period before the publication date and it is the k-th chronological match to transaction j, and *Comp Order k Post* is a dummy that equals one if the listing occurs in the period after the price data is published and it is the k-th chronological post-publication match to transaction j. Due to its length, only the coefficients of interest are presented in Table 9 while the full table can be found in the Appendix (Table 10).

Table 9 Chain Effects of Prices on Future Listings per Order of Match

The table presents the results of the following regression: $log(q_i) = \alpha + \sum_{k=1}^{10} \beta_k^{pre} \times log(p_j) \times Comp\ Order\ k\ Pre + \sum_{k=1}^{10} \gamma_k^{pre} Comp\ Order\ k\ Pre + \sum_{k=1}^{10} \beta_k^{post} \times log(p_j) \times Comp\ Order\ k\ Post + \sum_{k=1}^{10} \gamma_k^{post} Comp\ Order\ k\ Post + Controls + \varepsilon_i$, where q_i is the listed price for property i, p_j is the transaction price for a comparable property j sold in the previous month and Comp Order k Pre (Post) is a dummy that turns on when quote i is the k-th sequential match to transaction j in the period before (after) the price data publication date. The sample includes listings in the one-month period surrounding the publication date that have a comparable transaction which has at least one treated and one untreated match. Controls for the time distance between the listing and the comparable transaction measured in days and its interaction with log price are included in all columns but (1). Column (3) also includes listing month-year fixed effects. Standard errors double-clustered at the transaction and listing ID levels are reported in parentheses.

	(1)	(2)	(3)
Price	0.8388***	0.8404***	0.8401***
	(0.0011)	(0.0020)	(0.0020)
Price × 2nd Untreated	0.0010	0.0011	0.0011
	(0.0016)	(0.0016)	(0.0016)
Price × 3rd Untreated	0.0043**	0.0045**	0.0044**
	(0.0020)	(0.0020)	(0.0020)
Price × 4th Untreated	0.0054**	0.0057**	0.0056**
	(0.0023)	(0.0023)	(0.0023)
Price × 5th Untreated	0.0060**	0.0063**	0.0062**
	(0.0027)	(0.0027)	(0.0027)
Price × 6th Untreated	0.0039	0.0042	0.0041
	(0.0033)	(0.0033)	(0.0033)
Price × 7th Untreated	0.0068*	0.0071*	0.0070*
	(0.0039)	(0.0039)	(0.0039)
Price × 8th Untreated	0.0021	0.0024	0.0025
	(0.0044)	(0.0044)	(0.0044)
Price × 9th Untreated	0.0036	0.0040	0.0041
	(0.0055)	(0.0055)	(0.0055)
Price × 10th or more Untreated	0.0000	0.0004	0.0003
	(0.0038)	(0.0038)	(0.0038)
Price × 1st Treated	0.0015	0.0023	0.0025

Continued on next page

Table 9 – *Continued from previous page*

	(1)	(2)	(3)
	(0.0014)	(0.0017)	(0.0017)
Price × 2nd Treated	0.0039**	0.0048**	0.0050***
	(0.0016)	(0.0019)	(0.0019)
Price × 3rd Treated	0.0054***	0.0064***	0.0065***
	(0.0019)	(0.0022)	(0.0022)
Price × 4th Treated	0.0106***	0.0116***	0.0116***
	(0.0022)	(0.0025)	(0.0025)
Price × 5th Treated	0.0082***	0.0093***	0.0092***
	(0.0026)	(0.0029)	(0.0029)
Price × 6th Treated	0.0098***	0.0109***	0.0107***
	(0.0031)	(0.0033)	(0.0033)
Price × 7th Treated	0.0117***	0.0129***	0.0126***
	(0.0037)	(0.0039)	(0.0039)
Price × 8th Treated	0.0177***	0.0188***	0.0186***
	(0.0044)	(0.0046)	(0.0046)
Price × 9th Treated	0.0099*	0.0110**	0.0108**
	(0.0051)	(0.0052)	(0.0052)
Price × 10th or more Treated	0.0101***	0.0113***	0.0112***
	(0.0036)	(0.0039)	(0.0039)
Controls			
Price x Time distance	No	Yes	Yes
Fixed-Effects			
Month-year	No	No	Yes
Observations	11,292,009	11,292,009	11,292,009
R^2	0.7081	0.7081	0.7089
Within R^2	–	–	0.7063

Two-way (Transaction ID & Listing ID) standard-errors in parentheses
*Signif Codes: ***: 0.01, **: 0.05, *: 0.1*

Column (1) contains the baseline regression, column (2) adds the controls for distance in days between the listing and the matched transaction and the interaction with price, while column (3) also includes listing year-month fixed effects. I limit the number of comparables in the pre- and post-publication period to twenty as there are very few transactions with more than twenty comparables which makes the estimates very noisy. I estimate separate coefficients per order of match for the first nine matches while the tenth coefficient groups all matched listings of order ten or higher. To better understand the results, I also plot the price coefficients along with their 95% confidence bounds from the specification in column (3) in Figure 10 below, while the coefficients for the other two specifications are graphically depicted in Figure 5 of the Appendix. The results are virtually unchanged across all specifications. We can notice that the effect of past prices tends to slightly increase up until the seventh match in the period before publication, however, the magnitude of the incremental effect is not very large peaking at around 0.7% and losing statistical significance thereafter. The positive correlation can be attributed to the fact that fundamentals in the housing market are persistent which means that, although listings in the pre-publication period do not directly observe the previous-month prices, they tend to co-move due to this underlying persistence and commonly observed public information. This will be a key feature of the model presented in Section 6. Moreover, the slight upward trend in this co-movement suggests that prospective sellers are unable to properly isolate the private signals from recent observable listings which in turn leads to placing too much weight on stale news. Moving forward to the tenth or higher-order matched listings, however, the incremental effect starts to wane, dropping down to below 0.04%, as new information begins to dominate. What is interesting is that this trends gets completely turned around in the post-publication period: once transaction prices become publicly available, their influence on future listings sees a significant jump and a large upward trend as the order of match increases. More specifically, the incremental effect on the first match post-publication goes back up to around 0.15-0.25%

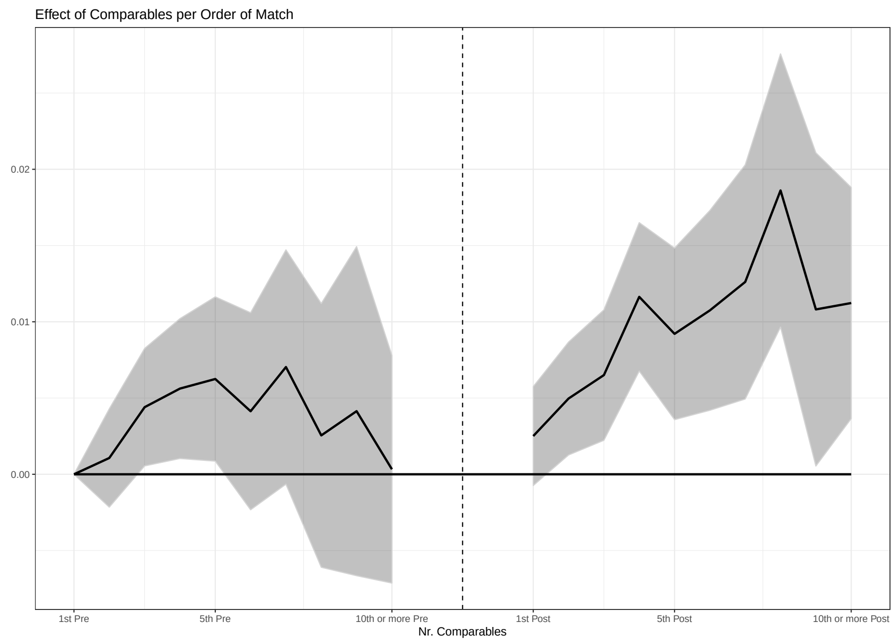


Figure 10 Coefficients on Comparable Transaction Prices by Order of Match
The figure plots the price coefficients from the following regression along with their 95% confidence bounds: $log(q_i) = \alpha + \sum_{k=1}^{10} \beta_k^{pre} \times log(p_j) \times Comp\ Order\ k\ Pre + \sum_{k=1}^{10} \gamma_k^{pre} Comp\ Order\ k\ Pre + \sum_{k=1}^{10} \beta_k^{post} \times log(p_j) \times Comp\ Order\ k\ Post + \sum_{k=1}^{10} \gamma_k^{post} Comp\ Order\ k\ Post + Controls + \varepsilon_i$, where q_i is the listed price for property i, p_j is the transaction price for a comparable property j sold in the previous month and Comp Order k Pre (Post) is a dummy that turns on when quote i is the k-th sequential match to transaction j in the period before (after) the price data publication date. The sample includes listings in the one-month period surrounding the publication date that have a comparable transaction which has at least one treated and one untreated match. Controls for the time distance between the listing and the comparable transaction measured in days and its interaction with log price, as well as listing month-year fixed effects are included in this specification.

percentage points relative to the first pre-publication match; it then steadily climbs to a striking 1.8% at the 8th match where it starts to level off. The coefficients for the post-publication period are both economically and statistically more significant that those in the pre-publication period. Although the effect on the early order of listings might be rationalised by claiming that prospective sellers now learn about the private signals embedded in recently published prices, it would be very difficult to justify the upward trend as the number of intermediate comparables increases. To put it differently, once the news becomes public, we should see an immediate jump in the coefficient as the new information gets embedded into prices which should then remain flat for all future listings or even exhibit a downward trend with the arrival of new information from previous quotes. The fact that the effect is gradually increasing thus provides further proof that agents in the housing market have trouble discerning different drivers of past actions; instead, they treat new observations as independent of previous commonly observed ones. For additional evidence, Table 11 in the Appendix presents the linear trend in the influence of past prices as the number of intermediate comparables grows. The results show that there is no significant change in the comovement between quotes and prices in the pre-publication period, however, the effect increases by about 0.07%-0.08% with each additional comparable in the post-publication period. Finally, Table 12 and Figure 6 in the Appendix provide evidence that the comovement patterns are very different before March 2012: in this sample, the influence from past prices remains largely flat as the number of interim quotes increases and we can even see that the incremental effect relative to the earliest match becomes negative for matches of order 10 or higher across both pre- and post-publication periods. The inability to use information contained in past prices, or the absence of regularity in the frequency of doing so, means that prospective sellers in the following month have less correlated information sets before March 2012 which in turn implies that their price decisions will be less impacted by past transactions.

For the final set of tests in this section, I focus on those listings which have seen price updates, i.e., I search for listings for which I have at least two available prices posted on two different dates. This would allow me to test for any amplification effects that arise due to the repeated use of comparables by taking into account any unobservable property and owner characteristics. Specifically, I now analyse the impact that transaction prices published just before a listing has been posted have on its subsequent price updates. This would allow me to investigate propagation effects within a given listing. To better see this, Figure 11 describes potential ways that information available prior to the very first time a property has been listed could have an increasing influence on later price changes. A prospective seller who has his property on the market for a while could still make use of new sources of information that arrive with new listings. However, if these new listings are using data that is also observable to our seller from the very beginning, then he would again be faced with the challenging task of distinguishing between what is truly new information and stale news.

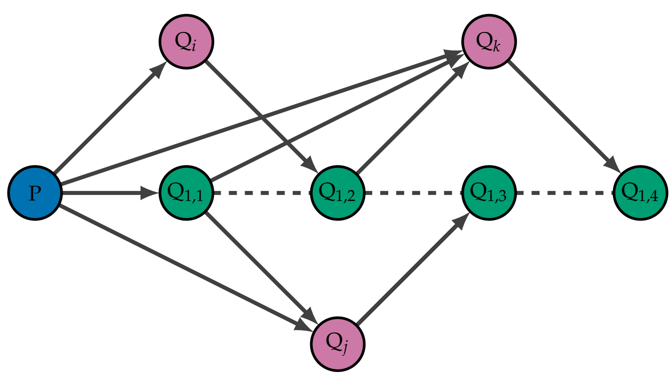


Figure 11 Indirect price effects
III: the evolution of the effect of past prices on a given listing and its quote updates
The figure shows the way that a given observation can exercise an increasing influence on a given listing via its effect on other observable listings. The blue circle depicts a given transaction, the green circles are the ordered quote changes for a listing that has been first posted after the release of the transaction data from the previous month and the pink circles represent other listings posted while the property of interest is still on the market.

Figure 7 in the Appendix shows a histogram of the number of price changes per listing. In my sample, there are around 520,000 listings with one or more price changes that I am able to match with at least one prior transaction. However, most listings have very few price updates, with the vast majority having only one price change (358,939) and only 4,494 listings having five or more price updates. For this reason, I combine all price changes of order four or higher into one category in the regressions. Figure 8 in the Appendix displays a histogram of the number of days between the date the first listing price was set and the subsequent price changes. I limit the analysis to changes that have occurred within up to two years after the property was first listed. It can be seen that most of the price changes occur within two to three months of listing, with the modal number of days between the marketed date and the price change date being 28 (10,666 observations). However, sellers frequently tend to update prices on the day following the listing date (7,529 observations). Nonetheless, it is also often the case that price changes occur even after 200 days have passed since the property was first listed (74,299 observations).

Table 10 presents the results of the following regressions:

$$\begin{aligned} log(q_i^n) =&\alpha + \beta_1 \times log(p_j) + \sum_{n=2}^{5} \beta_n \times log(p_j) \times Update\ Number\ n + \\ &\sum_{n=2}^{5} \gamma_n Update\ Number\ n + Controls + \varepsilon_i^n \end{aligned} \tag{20}$$

where q_i^n is the n-th listed price for listing i, p_j is the price at which a matched property was sold in the month before property i was initially listed and *Update Number n* is a dummy that turns on when quote i is the n-th change to the listing price for property i. The first column shows the results of the baseline regression, column (2) adds the time distance control and the interaction of price and time distance, column (3) adds month-year fixed effects and column (4) adds month-year and listing ID fixed effects. The coefficients and their 95% confidence bounds for the specification in column (3) are displayed in Figure 12 for visual inspection; Figure 9 in the Appendix shows the coefficients for the other three specifications. The results from the baseline regression in column (1) show that the incremental effect on the first price update of transaction prices published just before the property was listed on the market is about 0.97%. This effect gradually increases for subsequent price changes to reach 2.44% for updates of order four

Table 10 Effect of Prices on Listing Price Updates

The table presents the results of the following regression: $log(q_i^n) = \alpha + \beta_1 \times log(p_j) + \sum_{n=2}^{5} \beta_n \times log(p_j) \times Update\ Number\ n + \sum_{n=2}^{5} \gamma_n Update\ Number\ n + Controls + \varepsilon_i^n$, where q_i^n is the n-th listed price update for property i, p_j is the transaction price for a comparable property j sold in the month before property i was initially listed and Update Number n is a dummy that turns on when quote i is the n-th change to the listing price for property i. The sample includes listings in the post March 2012 period that have at least one price change and a comparable transaction that has been published just before the listing has been first posted. Controls for the time distance between the listing and the comparable transaction measured in days and its interaction with log price are included in all columns but (1). Column (3) also includes listing month-year fixed effects and column (4) includes listing ID as well as month-year fixed effects. Standard errors double-clustered at the transaction and listing ID levels are reported in parentheses.

	(1)	(2)	(3)	(4)
Price	0.8320***	0.8345***	0.8310***	-0.0105***
	(0.0010)	(0.0012)	(0.0013)	(0.0002)
Price × 1st Change	0.0097***	0.0132***	0.0129***	0.0070***
	(0.0002)	(0.0009)	(0.0009)	(0.0002)
Price × 2nd Change	0.0160***	0.0224***	0.0217***	0.0131***
	(0.0014)	(0.0021)	(0.0021)	(0.0004)
Price × 3rd Change	0.0194***	0.0282***	0.0272***	0.0184***
	(0.0028)	(0.0036)	(0.0036)	(0.0007)
Price × ≥ 4th Change	0.0244***	0.0359***	0.0349***	0.0225***
	(0.0074)	(0.0080)	(0.0079)	(0.0014)
1st Price Change	-0.1560***	-0.2037***	-0.1990***	-0.1224***
	(0.0019)	(0.0105)	(0.0105)	(0.0025)
2nd Price Change	-0.2542***	-0.3407***	-0.3306***	-0.2299***
	(0.0164)	(0.0250)	(0.0250)	(0.0049)
3rd Price Change	-0.3099***	-0.4294***	-0.4146***	-0.3204***
	(0.0340)	(0.0429)	(0.0428)	(0.0091)
≥ 4th Price Change	-0.3793***	-0.5362***	-0.5226***	-0.3972***
	(0.0894)	(0.0959)	(0.0950)	(0.0171)
Controls				
Price x Time distance	No	Yes	Yes	Yes
Fixed-Effects				
Month-year	No	No	Yes	Yes
Listing ID	No	No	No	Yes
Observations	5,868,384	5,868,384	5,868,384	5,868,384
R^2	0.7160	0.7161	0.7170	0.9981
Within R^2	–	–	0.7101	0.2670

Two-way (Transaction ID & Listing ID) standard-errors in parentheses
*Signif Codes: ***: 0.01, **: 0.05, *: 0.1*

or higher. Moving on to column (2) we see that controlling for the fact that the relationship between the listing price and the independent variable naturally decreases with the passage of time makes the magnitude of the effect even larger. Specifically, the effect on the first price update now goes up to 1.32% while that on the fourth update is striking 3.59% larger than on the initially set price. Adding month-year fixed effects in column (3) leaves the results largely unchanged which suggests that the result is not driven by a mere trend in prices. In particular, aggregate market dynamics do not absorb the increasing effect that past prices have on quote changes of higher order, i.e., controlling for average price levels confirms the finding that listing prices are more heavily influenced by early observable information, which are likely to be heterogeneous across sellers who update their quotes in the same month. Finally, the addition of listing ID fixed effects in column (4) slightly reduces the magnitude of the coefficients, however, they retain their statistical and economic significance across all specifications. Note that the present findings are not inconsistent with the well-known fact that most price updates tend to be downward changes. For example, even among price downgrades, the above results show that sellers who happened to observe a lot of positive news at the moment of listing tend to reduce their quote by less relative to others, and vice versa.

The above results provide remarkable evidence of the notable influence of past information on future seller behaviour both in the short as well as the medium-to-long term. Taking into account the fact that a considerable fraction of price updates in the data occurs even after six months since listing points to the strong stickiness of pricing behaviour among prospective sellers. This could potentially explain a significant amount of the price persistence we observe in the housing market in the aggregate.

I have hitherto provided convincing evidence of the challenges that agents in the housing market face due to the complex structure of connections among sequentially-moving actors. I have described how this intricate environment coupled with sub-perfect knowledge about its structure can lead sellers to place disproportionate weight on stale news at the expense of truly new information by failing to account for commonly ob-

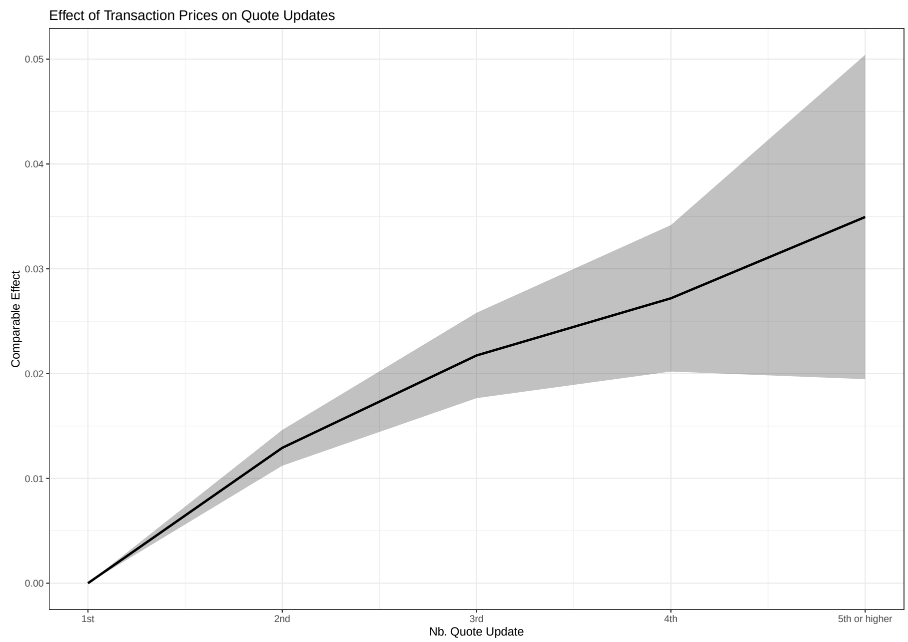


Figure 12 Effect of Comparable Transaction Prices on Quote Updates - Coefficients
The figure displays the price coefficients from the following regression along with their 95% confidence bounds: $log(q_i^n) = \alpha + \beta_1 \times log(p_j) + \sum_{n=2}^{5} \beta_n \times log(p_j) \times Update\ Number\ n + \sum_{n=2}^{5} \gamma_n Update\ Number\ n + Controls + \varepsilon_i^n$, where q_i^n is the n-th listed price update for property i, p_j is the transaction price for a comparable property j sold in the month before property i was initially listed and Update Number n is a dummy that turns on when quote i is the n-th change to the listing price for property i. The sample includes listings in the post March 2012 period that have at least one price change and a comparable transaction that has been published just before the listing has been first posted. Controls for the time distance between the listing and the comparable transaction measured in days and its interaction with log price as well as month-year fixed effects are included in this specification.

served drivers of recent actions. Having said that, however, I have so far only considered one side of the housing market. In particular, it would be interesting to know if the potential mistakes that sellers make when trying to learn the state of demand are partially or fully corrected at the selling stage. In the next section I will therefore look at the sample of listings that I have managed to match to subsequent transactions in order to answer this question.

5.3 Interaction with Buyers

For the final set of results, I am going to analyse the relationship between varying degrees of influence from past prices among prospective sellers and the ex-post discount that they are faced with at the transaction stage. Specifically, if prospective sellers mis-estimate the state of demand in the housing market, then this should eventually be corrected by buyers making offers that are further from the listed price the more this one does not coincide with current fundamentals. To provide some indicative evidence of this, I now restrict my attention to the listings in the sample that I have matched to subsequent transactions. For each listing, I compute the price differential between the first quote and the transaction price for that property[22]. I then split the sample of matched listings into five buckets corresponding to the five quantiles of the price discount distribution and run the following regression per bucket:

$$log(q_i^k) = \alpha + \beta^k \times log(p_j) + Controls + \varepsilon_i^k \quad (21)$$

where q_i^k is the first quoted price for listing i which is in quantile k of the price discount distribution and p_j is the transaction price for a similar property which has been published in the month before the listing was first posted. I include the usual controls for the time distance between past prices and matched quotes and its interaction with price, as well as listing month-year fixed effects in order to absorb any aggregate pricing dy-

[22]I have also done the same analysis with respect to the last quote available for a given listing, when there are multiple price changes, without any significant change in the results.

namics. The sample used in these tests contains 1,067,282 listings in the post March 2012 period that have been paired with their corresponding subsequent transactions. The mean price differential is -3.78% and the median one equals -3.13% which implied that sellers are willing to accept an average price discount of around 3.5% relative to the listed price. The average property sold stays on the market for about 26 weeks, with the median property sold in 21 weeks since listing[23]. Dividing the sample based on the price discount distribution leads to the following five buckets: (1) properties sold at a price that is more than 7.5% lower than the listed price; (2) properties sold at a discount of between 7.5% and 4.2% to listed price; (3) properties sold at a discount of between 4.2% and 2.2%; (4) properties sold at the listed price or a discount of up to 2.2%, and; (5) properties sold at a premium to quoted price. This division restates the strong skewness present in the distribution of price discount whereby we observe that over 74% of the properties are sold at a discount, with 12% of properties selling at the listed price and only 14% being sold at a premium, consistent with previous findings.

Table 11 displays the results of the regression in equation (21). For a visual representation, I plot the coefficients per bucket along with their 95% confidence bounds in Figure 13. The dashed horizontal line represents the price effect using the full sample of matched listings. The results suggest a strong U-shaped relationship between the effect of recent transactions on listings and the ex-post price discount. In particular, prospective sellers who are more heavily influenced by recent price data tend to subsequently sell their properties at a very large discount or at a premium to listed price. The difference in the price coefficients between the extremes (buckets 1 and 5) and the middle buckets, which contain properties sold closer to listed price, shows an 8% increase in the influence from recently observed data for those cases where the final price deviates the most from the quoted one. This piece of evidence suggests that making ”wrong” inference about demand by over-weighting stale information is indeed somewhat corrected in the final stage of the selling process. Interestingly, the effect goes in both directions, i.e., not only

[23]Recall, again, that this is the length of time between the property listing and the completion of the sale.

Table 11 Relation between Influence from Past Prices and Future Price Discount
The table presents the results of the following regression: $log(q_i^k) = \alpha + \beta^k \times log(p_j) + Controls + \varepsilon_i^k$, where q_i is the first quoted price for listing i which is in quantile k of the price discount distribution and p_j is the transaction price for a similar property which has been published in the month before the listing was originally posted. The sample includes listings in the sample period starting from March 2012 that have been matched to their respective ex-post transactions. The first five columns present the coefficients on recent transaction prices per quantile of price discount: column (1) considers listings sold at the largest price discount while column (5) listings that were sold at a premium to quoted price. The final column includes all matched listings to give an idea of the average price effect. Controls for the time distance between the listing and the comparable transaction measured in days and its interaction with log price are included in all regressions, as well as listing month-year fixed effects. Standard errors double-clustered at the transaction and listing ID levels are reported in parentheses.

	(1)	(2)	(3)	(4)	(5)	Full Sample
Price	0.8671***	0.8198***	0.7851***	0.7921***	0.8737***	0.8254***
	(0.0053)	(0.0053)	(0.0054)	(0.0049)	(0.0073)	(0.0027)
Controls						
Price x Time distance	Yes	Yes	Yes	Yes	Yes	Yes
Fixed-Effects						
Month-year	Yes	Yes	Yes	Yes	Yes	Yes
Observations	956,899	1,000,010	985,789	1,411,376	813,810	5,167,884
R^2	0.7392	0.7209	0.7084	0.6936	0.7041	0.7153
Within R^2	0.7332	0.7183	0.7072	0.6926	0.6984	0.7144

Two-way (Transaction ID & Listing ID) standard-errors in parentheses

*Signif Codes: ***: 0.01, **: 0.05, *: 0.1*

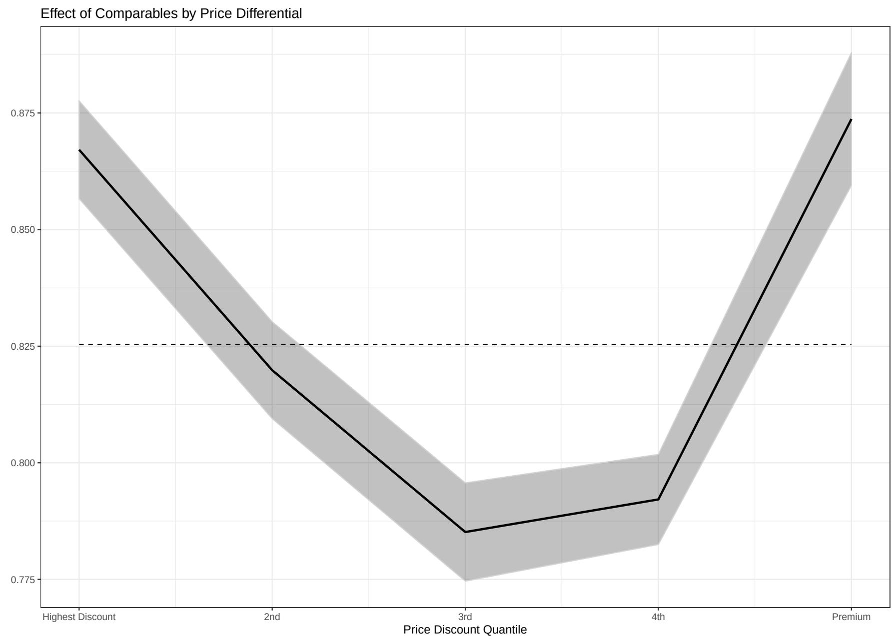


Figure 13 The Effect from Recently Observed Prices by Price Discount - Coefficients
The figure plots the coefficients of the following regressions along with their 95% confidence bounds: $log(q_i^k) = \alpha + \beta^k \times log(p_j) + Controls + \varepsilon_i^k$, where q_i is the first quoted price for listing i which is in quantile k of the price discount distribution and p_j is the transaction price for a similar property which has been published in the month before the listing was originally posted. The sample includes listings in the sample period starting from March 2012 that have been matched to their respective ex-post transactions. Each coefficient comes from a regression of listings from a different quantile of the price discount distribution: the first coefficient is based on listings sold at the largest price discount while the last one on listings that were sold at a premium to quoted price. The horizontal line represents the average price effect across the full sample. Controls for the time distance between the listing and the comparable transaction measured in days and its interaction with log price, as well as listing month-year fixed effects are included in the regressions.

do we see overpriced houses being sold at large discounts, but also properties that have been under-valued by sellers see their final transaction prices rise the most due to strong buyer competition.

Unlike most other financial markets, the market for residential housing clears on two dimensions (Yavas and Yang, 1995; Chen and Rosenthal, 1996; Carrillo, 2012). Specifically, mis-valued properties might take longer to sell if sellers are not willing to accept a significant discount. The relationship between making wrong inference and time spent on the market will, however, not be monotone as it was the case with the absolute price differential. In particular, while an overpriced property would take longer to sell, an underpriced one should sell very quickly due to high demand. As a result, I will now investigate the relationship between the effect of prices and time on the market by dividing the sample into two categories: properties sold at a discount, and properties sold at a premium to listed price. I run the regression specified in equation (21) by splitting the properties into five buckets based on the quantiles of the distribution of time spent on the market. Table 12 displays the results of this exercise: in Panel A I look at properties sold at a discount to listed price (793,203 unique observations), while Panel B considers properties sold at the listed price or above (274,079 distinct listings). Note that the time on the market distribution is quite different for properties sold below the quoted price, compared to those sold at or above it. In particular, while the average property sold at a discount spends about 28 weeks on the market with the median property selling in 22 weeks, the average and median properties sold at premium spend 22.5 and 17 weeks on the market, respectively, which corresponds to approximately five weeks faster selling time for the second sample. Of course, it is a well known fact that volume and prices are highly correlated in the housing market, i.e., houses sell quickly and at higher prices in hot markets, relative to cold markets (Genesove and Mayer, 2001; Glaeser and Nathanson, 2015). This suggests that the time coverage of the properties sold at discount compared to those sold at premium is probably quite different. As a result, it is crucial to introduce time effects to the regression that would capture the average price level in the housing market in a given month. In this way, I now compare relative mistakes that prospective sellers make across TOM buckets after partialling out any aggregate temporal variation of house prices.

Figure 14 plots the effect of recent prices on listings by quantile of the TOM distribution: Figure 14a shows the coefficients for the sample of listings sold at a discount, while Figure 14b provides the coefficients for the listings sold at a premium. It can be observed that the effect is contrasting across the two samples. Namely, for the sample of properties sold below the quoted price, we can see that there is a positive, albeit delicate connection between the amount of time spent on the market and the degree of correla-

Table 12 Relation between Influence from Past Prices and Time on the Market

The table presents the results of the following regression: $log(q_i^k) = \alpha + \beta^k \times log(p_j) + Controls + \varepsilon_i^k$, where q_i is the first quoted price for listing i which is in quantile k of the time-on-the-market (TOM) distribution and p_j is the transaction price for a similar property which has been published in the month before the listing was originally posted. The sample includes listings in the sample period starting from March 2012 that have been matched to their respective ex-post transactions. Panel A considers properties that sold at a discount to listed price, while Panel B focuses on properties that sold at a premium. The first five columns present the coefficients on recent transaction prices per TOM quantile: column (1) looks at listings that took the least time to sell while column (5) listings that had the longest duration. The final column includes all matched listings to give an idea of the average price effect. Controls for the time distance between the listing and the comparable transaction measured in days and its interaction with log price are included in all regressions, as well as listing month-year fixed effects. Standard errors double-clustered at the transaction and listing ID levels are reported in parentheses.

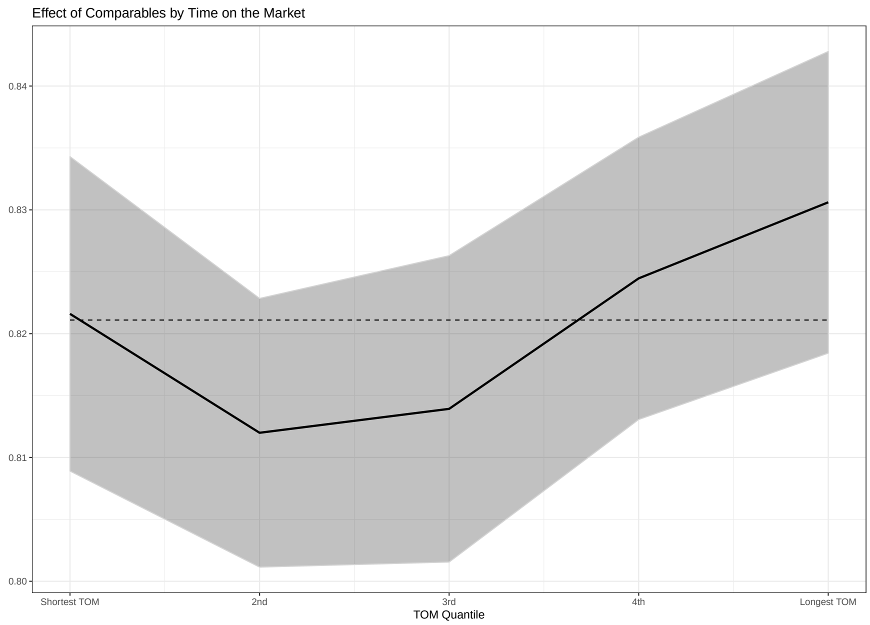
Panel A: Properties Sold at Discount						
	(1)	(2)	(3)	(4)	(5)	Full Sample
Price	0.8216***	0.8120***	0.8139***	0.8245***	0.8306***	0.8211***
	(0.0065)	(0.0055)	(0.0063)	(0.0058)	(0.0062)	(0.0030)
Controls						
Price x Time distance	Yes	Yes	Yes	Yes	Yes	Yes
Fixed-Effects						
Month-year	Yes	Yes	Yes	Yes	Yes	Yes
Observations	736,353	905,252	680,331	752,650	638,851	3,713,437
R^2	0.7230	0.7232	0.7273	0.7276	0.7258	0.7242
Within R^2	0.7220	0.7222	0.7253	0.7243	0.7165	0.7223
Panel B: Properties Sold at Premium						
	(1)	(2)	(3)	(4)	(5)	Full Sample
Price	0.9042***	0.8113***	0.8152***	0.8157***	0.8303***	0.8431***
	(0.0115)	(0.0117)	(0.0109)	(0.0107)	(0.0115)	(0.0054)
Controls						
Price x Time distance	Yes	Yes	Yes	Yes	Yes	Yes
Fixed-Effects						
Month-year	Yes	Yes	Yes	Yes	Yes	Yes
Observations	314,764	253,091	291,296	317,690	277,606	1,454,447
R^2	0.7080	0.7002	0.6939	0.6782	0.6814	0.6933
Within R^2	0.7026	0.6940	0.6887	0.6739	0.6781	0.6895

Two-way (Transaction ID & Listing ID) standard-errors in parentheses

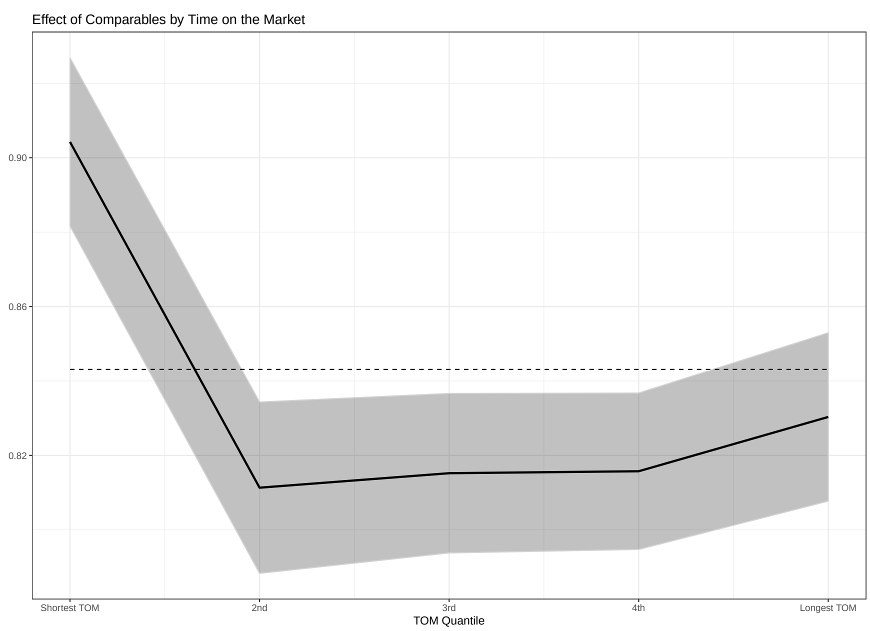
*Signif Codes: ***: 0.01, **: 0.05, *: 0.1*

tion with past price data. It seems that properties that take longer to sell might be more heavily affected by recent comparable transactions, however, the statistical significance of this effect is close to zero. In particular, we cannot rule out the hypothesis that the correlation with recent transactions is equal to the sample average across all TOM buckets. The results are quite different in Panel B of Table 12: it is now evident that properties that have been most influenced by past news take the least time to sell with the difference in this effect being 7-9% larger for properties with the shortest TOM compared to the rest of the sample. On the other hand, we cannot reject the hypothesis that there is no significant variation in the way that sellers in the top four quantiles of the TOM distribution make inference from past available information. The findings above suggest that the housing market is at least to some extent efficient at correcting mistakes made by sellers due to wrong inference resulting from an environment of imperfect information. Furthermore, the analysis conducted in this section hints at the potential asymmetry in the way that the market achieves this rectification. Specifically, while overpricing is typically corrected through sellers accepting discounts to their ask price, undervaluation gets amended through both the price channel (buyers willing to pay a price above the quoted one) and the time channel (significantly underpriced properties take less time to sell).

In this section, I have provided empirical evidence that prospective sellers in the housing market are subject to inference biases that result from the complex nature of the information structure. Namely, the results on the indirect effects from past prices through intermediate comparable listings indicate that stale information might have long-lasting effects on the behaviour of future sellers who are unable to disentangle between common signals and real news. In the penultimate section of this paper, I will therefore investigate the economic impact that pricing mistakes might have on aggregate house market dynamics in the long run.



(a) Properties Sold at Discount



(b) Properties Sold at Premium

Figure 14 The Effect from Recently Observed Prices by Time on the Market - Coefficients
The figure plots the coefficients of the following regressions along with their 95% confidence bounds: $log(q_i^k) = \alpha + \beta^k \times log(p_j) + Controls + \varepsilon_i^k$, where q_i is the first quoted price for listing i which is in quantile k of the TOM distribution and p_j is the transaction price for a similar property which has been published in the month before the listing was originally posted. The sample includes listings after March 2012 that have been matched to their respective ex-post transactions. Figure 14a considers properties that sold at a discount to listed price, while Figure 14b properties that sold at a premium. Each coefficient comes from a regression of listings from a different quantile of the TOM distribution: the first coefficient is based on listings that took the least time to sell while the last one on listings that had the longest duration. The horizontal line represents the average price effect across the full sample. Controls for the time distance between the listing and the comparable transaction measured in days and its interaction with log price, as well as listing month-year fixed effects are included in the regressions.

6 Economic Magnitude of Learning Mistakes

The aim of this section is to provide some indications of what the effect of individual pricing mistakes that result from naïve learning as demonstrated in Section 5.2 might be on the aggregate picture, once we consider that the majority of house market agents are subject to the same learning biases. This can further help us gain understanding about some recent market phenomena, such as the effect of the Brexit referendum vote and various stamp duty holidays on housing market dynamics. For this purpose, I will sketch a simple model whose goal is to capture some of the key features of real-estate markets and the information structure established in the rest of the text.

Let us assume that the log of house prices are determined by a fundamental δ_t which follows an AR(1) process with persistence parameter ρ and mean a:

$$\delta_t = a + \rho\delta_{t-1} + \varepsilon_t \text{ , } \varepsilon_t \overset{iid}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2) \tag{22}$$

The above assumption is for simplicity and is meant to capture, in reduced-form, the excess demand that prospective sellers face. This implies that prices are determined based on the conditional expectation of δ_t at time t:

$$p_{i,t} = \mathbb{E}_{i,t}[\delta_t] \tag{23}$$

where $p_{i,t}$ is the transaction price for a property sold at time t. Agents do not observe the realisation of the fundamental δ_t and, as a result, estimate its value from available information. In particular, the informational structure is characterised by the presence of public and private signals. In every period, prices are determined after the observation of a private signal $s_{i,t}$ about the fundamental value:

$$s_{i,t} = \delta_t + \eta_{i,t} \text{ , } \eta_{i,t} \overset{iid}{\sim} \mathcal{N}(0, \sigma_\eta^2) \tag{24}$$

The noise terms are independent and identically distributed across individuals and time. If there are multiple transactions occurring in a given period, they are all formed based on a different private signal. There is also one publicly observable signal s_t arriving every k periods[24]:

$$s_t = \delta_t + u_t \ , \ u_t \overset{iid}{\sim} \mathcal{N}(0, \sigma_u^2) \tag{25}$$

The public signal noise is also identically and independently distributed across time. This signal represents any public information that agents might take advantage of to make inference about housing demand and prices[25]. Finally, agents also observe the full history of past transactions which they also use to extract the private signals that agents in previous periods have received. At every period t, therefore, the information sets of agents consist of the full history of past prices, the history of public signals and their own private signals. They thus form conditional expectations of δ_t and set prices accordingly, as follows:

$$p_{i,t} = \mathbb{E}[\delta_t | s_{i,t}, s^t, p^{t-1}] \tag{26}$$

where $s^t = \{s_0, s_{0+k}, s_{0+2k}, ...\}$, $p^{t-1} = \{p_0, p_1, ..., p_{t-1}\}$, denote the full history of public signals and transaction prices, respectively, that agents at time t observe, while $s_{i,t}$ is agent i's private signal. Given the model described in equations (22)-(26) above, we can trace the learning process of agents who act sequentially. Specifically, I solve for the posterior beliefs, and consequently prices, for both the Bayesian and the naïve case in order to compare the house price dynamics that these generate. Section A.2 in the Appendix provides detail on the procedure of forming posterior beliefs and the recursion that can be used to update beliefs given the new signals from a given period. It can be

[24]Here I assume that there is a single public signal arriving at $t = 0$ for simplicity of exposition, however, I vary the frequency of public signal arrival in the simulation exercise.

[25]As explained in Section A of the Appendix, the public signal might involve a housing price index published at regular frequencies. Alternatively, it can also be interpreted as representing local area characteristics or amenities visible to everyone.

shown that a Bayesian learner i in period t would form prices as follows:

$$\begin{aligned} p_{i,t} = \delta^i_{t|t} =& w_t s_{i,t} + (1-w_t)\widetilde{w}_{t-1}(a+\rho\bar{s}_{t-1}) + \\ & (1-w_t)(1-\widetilde{w}_{t-1})\widetilde{w}_{t-2}(a+\rho a+\rho^2\bar{s}_{t-2}) + ... + \\ & (1-w_t)(1-\widetilde{w}_{t-1})(1-\widetilde{w}_{t-2})...(1-\widetilde{w}_1)\widetilde{w}_0(a+\rho a+...+\rho^{t-1}a+\rho^t\bar{s}_0) + \\ & (1-w_t)(1-\widetilde{w}_{t-1})(1-\widetilde{w}_{t-2})...(1-\widetilde{w}_1)(1-\widetilde{w}_0)(a+\rho a+...+\rho^{t-1}a+\rho^t s^0) \end{aligned} \tag{27}$$

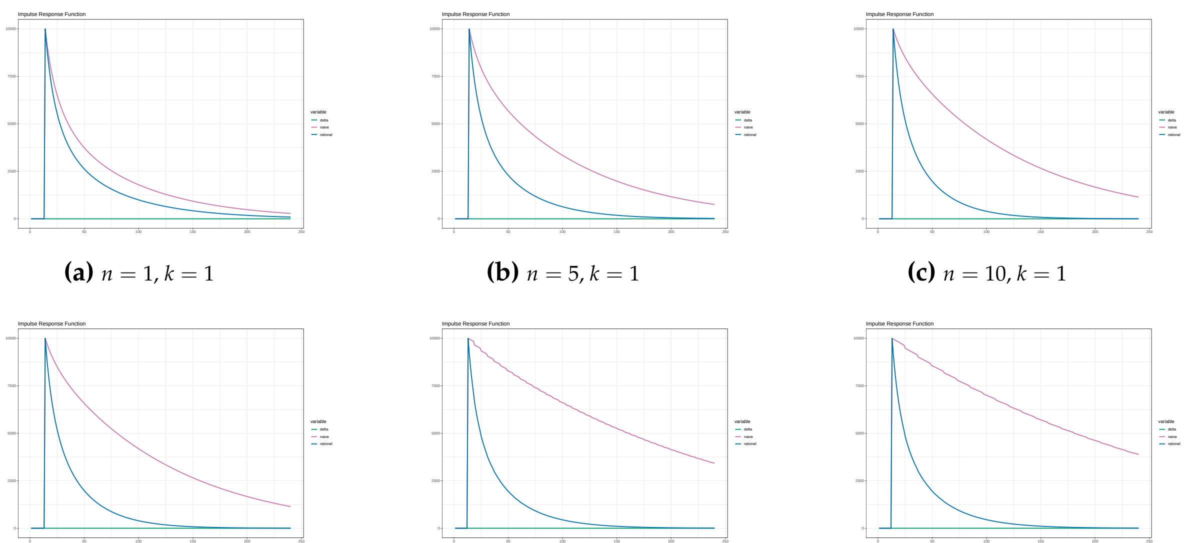
where $p_{i,t}$ is the price set by a Bayesian agent i in period t, $\delta^i_{t|t}$ is his conditional expectation of the fundamental value δ_t given all information available at time t, $s_{i,t}$ is agent i's own private signal, s^0 is the public signal arriving in the first period, $\bar{s}_{t-k}$, $\forall k \leq t$, is the precision-weighted average[26] of the private signals across all n agents in period $t-k$, and w_t and $\widetilde{w}_{t-k}$, $\forall k \leq t$ are weights that the agents in period t assign to all available signals. These are determined based on the signals' relative precisions with regard to the current state of the underlying as explained in Section A.2 of the Appendix. Similarly, we can show that naïve learners would form beliefs in a slightly different way:

$$\begin{aligned} \widetilde{p}_{i,t} = \widetilde{\delta}^i_{t|t} =& w_t s_{i,t} + (1-w_t)\widetilde{w}_{t-1}(a+\rho\bar{\widetilde{p}}_{t-1}) + \\ & (1-w_t)(1-\widetilde{w}_{t-1})\widetilde{w}_{t-2}(a+\rho a+\rho^2\bar{\widetilde{p}}_{t-2}) + ... + \\ & (1-w_t)(1-\widetilde{w}_{t-1})(1-\widetilde{w}_{t-2})...(1-\widetilde{w}_1)\widetilde{w}_0(a+\rho a+...+\rho^{t-1}a+\rho^t\bar{\widetilde{p}}_0) + \\ & (1-w_t)(1-\widetilde{w}_{t-1})(1-\widetilde{w}_{t-2})...(1-\widetilde{w}_1)(1-\widetilde{w}_0)(a+\rho a+...+\rho^{t-1}a+\rho^t s^0) \end{aligned} \tag{28}$$

$\widetilde{p}_{i,t}$ is the price set by a naïve agent i in period t, $\widetilde{\delta}^i_{t|t}$ is his conditional expectation of the fundamental value δ_t given all information available at time t, w_t and $\widetilde{w}_{t-k}$, $\forall k \leq t$ are the same weights as defined in the Bayesian case and $\bar{\widetilde{p}}_{t-k}$, $\forall k \leq t$ is the average price across all n agents in period $t-k$ weighted by the relative private signal precisions. Comparing

[26]As the precisions of all private signals are assumed to be the same, $\bar{s}_t$ is an equally-weighted average for all t.

equations (27) and (28), we can note that the difference between Bayesian and naïve agents is that naïve learners treat all past prices as independent signals, i.e., they fail to account for the fact that past agents have similarly set prices by looking at the actions of yet earlier agents. They, therefore, assign the same weights as the Bayesian agents but directly to the observed prices as opposed to the properly extracted signals. This would lead them to overweight stale news at the expense of more recent information since old news have already been accounted for in recent prices.



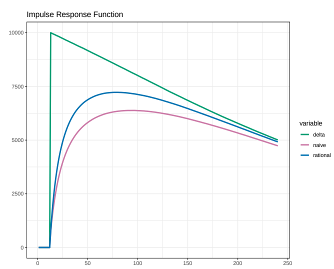
(a) $n = 1, k = 1$ **(b)** $n = 5, k = 1$ **(c)** $n = 10, k = 1$

(d) $n = 10, k = 1$ **(e)** $n = 10, k = 6$ **(f)** $n = 10, k = 12$

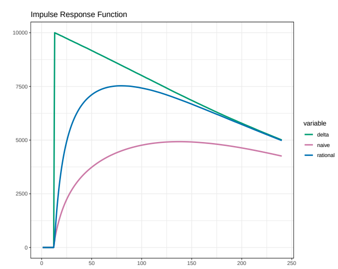
Figure 15 Impulse Response to a Shock to the Public Signal
The figure plots impulse responses of prices to a shock to the public signal in period 13. The response of naïve prices is depicted in pink, that of rational prices is in blue and the underlying state of demand is plotted in green. The various figures vary the number of prices in a given period (n) from 1 to 10 and the frequency at which the public signal arrives (k) from every month to every 12 months. The shock is standardized to correspond to a £10,000 increase in prices on impact.

To test the magnitude of the effect of naïve learning, I simulate a market with the above characteristics and compare the impact of various shocks on prices under the Bayesian and the naïve framework. For this, I first calibrate the model parameters, specifically the various signal precisions, using the empirical estimates from Section 5.2 above. The details of the calibration procedure are presented in Section A.1 in the Appendix.

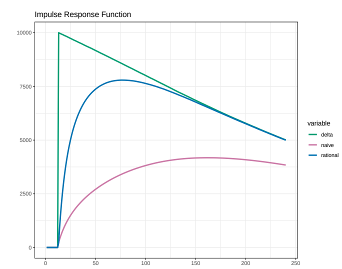
The parameters that govern the underlying process are estimated by running a monthly regression of log aggregate house prices on aggregate income. The predicted values of this regression are then used to fit an AR(1) process that yields estimates for a, ρ and σ_ε^2. Figure 15 depicts the impact of a shock to the public signal. I plot the response of naïve prices in pink, that of Bayesian prices in blue and the underlying in green. The shock is normalised to correspond to a 5% increase in prices which is about £10,000 for an average house priced at £200,000. In the first row figures, I hold the frequency of public signal arrival fixed by assuming that a new commonly observed signal arrives every period and I vary the number of agents: going from left to right, I plot impulse responses for the cases with one, five and ten transactions per month. Looking at the first figure which corresponds to the case with only one agent per period, we can see that naïve prices take longer to recover relative to the Bayesian case although the difference is not very large. This difference gets amplified, however, once we increase the number of agents per period. Figure 15b shows that, in the case with five transactions per month, after 20 years the effect on Bayesian prices has been eliminated while naïve prices are still about £700 larger which corresponds to 7% of the initial shock. The effect is even more striking in Figure 15c which assumes that there are ten agents per period: in this case, while rational prices have essentially converged to the truth in about 12 years, naïve prices are still higher by more than £2,500; even after 20 years more than 10% of the shock persists. Two opposing effects generate these results. First, note that the increase in the number of observations per period leads to Bayesian agents learning faster. Second, naïve agents are, on the contrary, harmed by the availability of more data since it takes them more time to converge as n increases. This implies that the increase in the amount of information about past prices is not necessarily beneficial in an environment where agents are prone to make wrong inference by double-counting commonly contained signals. Looking at the second row of Figure 15, I now keep the number of agents constant at ten per month and vary the frequency at which public signals are released, from every month in Figure 15d to every twelve months in Figure 15f. The results here are less surprising, namely, it takes both types of agents longer to converge when public news is more sparse, however, naïve learners are relatively more affected by this since they always overweight old news, therefore the shock, and even more so when new common signals arrive less frequently. This causes naïve prices to still be more than £3,500 above fundamentals, or 35% of the initial shock, even after 20 years.



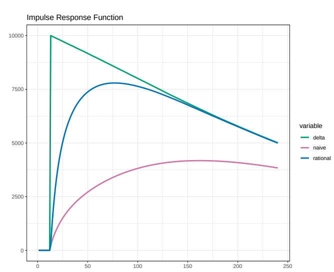
(a) $n = 1, k = 1$



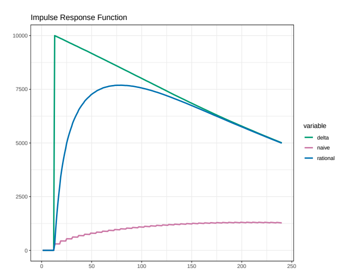
(b) $n = 5, k = 1$



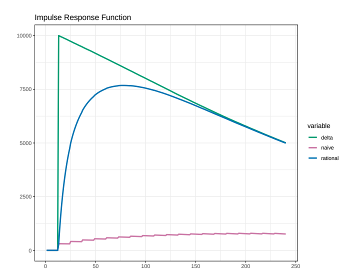
(c) $n = 10, k = 1$



(d) $n = 10, k = 1$



(e) $n = 10, k = 6$



(f) $n = 10, k = 12$

Figure 16 Impulse Response to a Shock to the Underlying
The figure plots impulse responses of prices to a shock to the underlying in period 13. The response of naïve prices is depicted in pink, that of rational prices is in blue and the underlying state of demand is plotted in green. The various figures vary the number of prices in a given period (n) from 1 to 10 and the frequency at which the public signal arrives (k) from every month to every 12 months. The shock is standardized to correspond to a £10,000 increase in the underlying on impact.

We can, so far, conclude that naïve learning leads agents to overreact to public signal shocks relative to Bayesian learners. This cannot be generalised, however, to any type of shock. In particular, Figure 16 displays impulse responses to a £10,000 shock to the underlying demand. The specifications across figures are the same as in Figure 15, namely I do the same comparative statics by varying the number of agents and the frequency at which public news gets released. We can observe a striking difference in the response of naïve and Bayesian learners relative to the previous example. Specifically, naïve agents underreact to the shock when this represents true changes in demand. This is intuitive as the real shock to delta gets suppressed by stale information coming from previous observations. Note that the effect is even more significant due to the high persistence in the fundamental as this implies that shocks to demand can take a very long time to recover from. As a result, in the worst case scenario of Figure 16f with ten agents and public signals arriving every twelve months, naïve prices are still about £4,000 away from the true fundamentals.

The above results suggest that pricing mistakes arising from naïve learning have an important economic impact which can be very long-lasting and cause changes in pricing patterns that are unrelated to true fundamentals. Moreover, the above findings shed light on some real-world dynamics of housing markets. In particular, they could potentially explain why real-estate prices in the UK were largely unaffected in the wake of the Brexit referendum vote even though this event pointed toward a significant drop in future housing demand. On the other hand, the naïve learning model would explain the effectiveness of stamp-duty tax holidays and similar policy measures as this type of public information is predicted to have a positive impact on housing market activity for an extended period of time even after their end.

7 Conclusions

In this paper, I have provided evidence on the learning behaviour of sellers in the market for residential housing. I have shown that valuation by comparables is a commonly used pricing method in the housing market which makes prices sensitive to the quality and quantity of past observations that gets released. Crucially, I have demonstrated that the failure to fully understand the structure of information flows leads prospective sellers to overweight signals coming from old data as this gets repeatedly embedded in subsequent observations. Finally, I have presented and simulated a model that shows how the excessive sensitivity to stale news can cause prices to exhibit large swings that might be unrelated to fundamentals.

Although I use the housing market as an ideal laboratory for analysing naïve inference and the effects thereof, the findings of this paper extend more generally to other markets where this pricing method is regularly employed, but also even more broadly to any social setting where economic agents use past observations to inform their decisions (e.g., leisure choices, political opinions, financial decisions). As long as the structure of the network through which information disseminates is not perfectly known to the individuals that form part of it, making inference by approximating past actions as being pure revelations of the private signals that previous actors have received would lead agents to place disproportionately high importance on early signals relative to more recent ones. Moreover, the findings above suggest that this behaviour can give rise to situations where agents overreact to noise and under-react to true changes in fundamentals. Crucially, the degree of (over)under-reaction is increasing with the availability of data on past actions and decreasing with the frequency at which agents receive new (public or private) signals about fundamentals. The results of this paper, therefore, give support to policies that facilitate access to reliable information about economic fundamentals. However, perhaps counterintuively, they predict that the improvement in the ability to observe the actions of other individuals might actually contribute to pricing mistakes.

References

Alevy, Jonathan E, Michael S Haigh, and John A List, “Information cascades: Evidence from a field experiment with financial market professionals,” *The Journal of Finance*, 2007, *62* (1), 151–180.

Anenberg, Elliot, “Information frictions and housing market dynamics,” *International Economic Review*, 2016, *57* (4), 1449–1479.

Bailey, Michael, Ruiqing Cao, Theresa Kuchler, and Johannes Stroebel, “The economic effects of social networks: Evidence from the housing market,” *Journal of Political Economy*, 2018, *126* (6), 2224–2276.

Baker, Malcolm and Richard Ruback, “Estimating industry multiples,” *Harvard University*, 1999.

Banerjee, Abhijit V., “A Simple Model of Herd Behavior,” *The Quarterly Journal of Economics*, 1992, *107* (3), 797–817.

Barberis, Nicholas, Robin Greenwood, Lawrence Jin, and Andrei Shleifer, “X-CAPM: An extrapolative capital asset pricing model,” *Journal of financial economics*, 2015, *115* (1), 1–24.

Bhojraj, Sanjeev and Charles MC Lee, “Who is my peer? A valuation-based approach to the selection of comparable firms,” *Journal of accounting research*, 2002, *40* (2), 407–439.

Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch, “A Theory of Fads, Fashion, Custom, and Cultural Change in Informational Cascades,” *Journal of Political Economy*, October 1992, *100* (5), 992–1026.

Bracke, Philippe and Silvana Tenreyro, “History dependence in the housing market,” 2020.

Brown, Jeffrey R, Zoran Ivković, Paul A Smith, and Scott Weisbenner, "Neighbors matter: Causal community effects and stock market participation," *The Journal of Finance*, 2008, *63* (3), 1509–1531.

Brunnermeier, Markus K and Christian Julliard, "Money illusion and housing frenzies," *The Review of Financial Studies*, 2008, *21* (1), 135–180.

Burnside, Craig, Martin Eichenbaum, and Sergio Rebelo, "Understanding Booms and Busts in Housing Markets," *Journal of Political Economy*, 2016, *124* (4), 1088–1147.

Carrillo, Paul E, "An empirical stationary equilibrium search model of the housing market," *International Economic Review*, 2012, *53* (1), 203–234.

Chen, Yongmin and Robert W Rosenthal, "Asking prices as commitment devices," *International Economic Review*, 1996, pp. 129–155.

Cho, Man, "House price dynamics: A survey of theoretical and empirical issues," *Journal of Housing Research*, 1996, pp. 145–172.

Dasgupta, Amil, Andrea Prat, and Michela Verardo, "The price impact of institutional herding," *The Review of Financial Studies*, 2011, *24* (3), 892–925.

DeAngelo, Linda Elizabeth, "Equity valuation and corporate control," *Accounting Review*, 1990, pp. 93–112.

DeMarzo, Peter M., Dimitri Vayanos, and Jeffrey Zwiebel, "Persuasion Bias, Social Influence, and Unidimensional Opinions," *The Quarterly Journal of Economics*, 2003, *118* (3), 909–968.

Eyster, Erik and Matthew Rabin, "Naïve Herding in Rich-Information Settings," *American Economic Journal: Microeconomics*, November 2010, *2* (4), 221–243.

_ and _ , "Extensive Imitation is Irrational and Harmful," *The Quarterly Journal of Economics*, 2014, *129* (4), 1861–1898.

Fracassi, Cesare, "Corporate finance policies and social networks," *Management Science*, 2017, *63* (8), 2420–2438.

Fuster, Andreas, David Laibson, and Brock Mendel, "Natural expectations and macroeconomic fluctuations," *Journal of Economic Perspectives*, 2010, *24* (4), 67–84.

Genesove, David and Christopher Mayer, "Loss aversion and seller behavior: Evidence from the housing market," *The quarterly journal of economics*, 2001, *116* (4), 1233–1260.

Giacoletti, Marco and Christopher A Parsons, "Peak-Bust Rental Spreads," *Available at SSRN 3482198*, 2019.

Glaeser, Edward L and Charles G Nathanson, "Housing bubbles," in "Handbook of regional and urban economics," Vol. 5, Elsevier, 2015, pp. 701–751.

_ and _ , "An extrapolative model of house price dynamics," *Journal of Financial Economics*, 2017, *126* (1), 147–170.

Grinblatt, Mark, Sheridan Titman, and Russ Wermers, "Momentum investment strategies, portfolio performance, and herding: A study of mutual fund behavior," *The American economic review*, 1995, pp. 1088–1105.

Guren, Adam M., "House Price Momentum and Strategic Complementarity," *Journal of Political Economy*, 2018, *126* (3), 1172–1218.

Han, Lu and William C. Strange, "The Microstructure of Housing Markets," in Gilles Duranton, J. V. Henderson, and William C. Strange, eds., *Handbook of Regional and Urban Economics*, Vol. 5 of *Handbook of Regional and Urban Economics*, Elsevier, 2015, chapter 0, pp. 813–886.

_ and William C Strange, "What is the role of the asking price for a house?," *Journal of Urban Economics*, 2016, *93*, 115–130.

Head, Allen, Huw Lloyd-Ellis, and Hongfei Sun, "Search, Liquidity, and the Dynamics of House Prices and Construction," *American Economic Review*, April 2014, *104* (4), 1172–1210.

Ho, Lee In, "On the Convergence of Informational Cascades," *Journal of Economic Theory*, December 1993, *61* (2), 395–411.

Hong, Harrison, Jeffrey D Kubik, and Jeremy C Stein, "Social interaction and stock-market participation," *The journal of finance*, 2004, *59* (1), 137–163.

_ , Jeffrey D. Kubik, and Jeremy C. Stein, "Thy Neighbor's Portfolio: Word-of-Mouth Effects in the Holdings and Trades of Money Managers," *Journal of Finance*, 2005, *60* (6), 2801–2824.

Ivković, Zoran and Scott Weisbenner, "Information diffusion effects in individual investors' common stock purchases: Covet thy neighbors' investment choices," *The Review of Financial Studies*, 2007, *20* (4), 1327–1357.

Kaplan, Steven N and Richard S Ruback, "The valuation of cash flow forecasts: An empirical analysis," *The journal of Finance*, 1995, *50* (4), 1059–1093.

Kim, Moonchul and Jay Ritter, "Valuing IPOs," *Journal of Financial Economics*, 1999, *53* (3), 409–437.

Kuchler, Theresa and Basit Zafar, "Personal Experiences and Expectations about Aggregate Outcomes," *Journal of Finance*, 2019, *74* (5), 2491–2542.

Lakonishok, Josef, Andrei Shleifer, and Robert W Vishny, "The impact of institutional trading on stock prices," *Journal of financial economics*, 1992, *32* (1), 23–43.

Liu, Jing, Doron Nissim, and Jacob Thomas, "Equity valuation using multiples," *Journal of Accounting Research*, 2002, *40* (1), 135–172.

McCrary, Justin, "Manipulation of the running variable in the regression discontinuity design: A density test," *Journal of econometrics*, 2008, *142* (2), 698–714.

Merlo, Antonio and Francois Ortalo-Magne, "Bargaining over residential real estate: evidence from England," *Journal of urban economics*, 2004, *56* (2), 192–216.

— , François Ortalo-Magné, and John Rust, "The home selling problem: Theory and evidence," *International Economic Review*, 2015, *56* (2), 457–484.

Murfin, Justin and Ryan Pratt, "Comparables Pricing," *Review of Financial Studies*, 2019, *32* (2), 688–737.

Ngai, L Rachel and Silvana Tenreyro, "Hot and cold seasons in the housing market," *American Economic Review*, 2014, *104* (12), 3991–4026.

Piazzesi, Monika and Martin Schneider, "Momentum traders in the housing market: Survey evidence and a search model," *American Economic Review*, 2009, *99* (2), 406–11.

Purnanandam, Amiyatosh K and Bhaskaran Swaminathan, "Are IPOs really underpriced?," *The Review of Financial Studies*, 2004, *17* (3), 811–848.

Smith, Lones and Peter Sorensen, "Pathological Outcomes of Observational Learning," *Econometrica*, March 2000, *68* (2), 371–398.

Wang, Guocheng and Yanyi Wang, "Herding, social network and volatility," *Economic Modelling*, 2018, *68*, 74–81.

Yavas, Abdullah and Shiawee Yang, "The strategic role of listing price in marketing real estate: theory and evidence," *Real Estate Economics*, 1995, *23* (3), 347–368.

Internet Appendix for "Learning from Past Prices - Evidence from the UK Housing Market"

A Structural Estimation

In this section I will sketch a stylised model that builds on the results from Section 2 in the main body of the paper in order to gain intuition regarding the way that covariances between prices and subsequent quotes is expected to change as the number of intermediate comparables grows under the Bayesian and naïve learning models. I will make the model more realistic compared to Section 2 by allowing the underlying state to change over time and for the presence of commonly observed signals. Subsequently, I will use the estimated coefficients from the empirical study of indirect effects obtained in Section 5.2 to provide some evidence regarding the magnitude of the impact of pricing mistakes on aggregate market dynamics.

For simplicity, let us assume that the log of house prices are determined by the fundamental δ_t which follows an AR(1) process with persistence parameter ρ and mean a:

$$\delta_t = a + \rho\delta_{t-1} + \varepsilon_t \,,\, \varepsilon_t \overset{iid}{\sim} \mathcal{N}(0,\, \sigma_\varepsilon^2) \tag{1}$$

This can be thought of as a reduced-form way of modelling the demand that sellers face. As a result, prospective sellers set listing prices based on their expectation of δ_t:

$$p_{i,t} = \mathbb{E}_{i,t}[\delta_t] \tag{2}$$

$$q_{i,t} = \mathbb{E}_{i,t}[\delta_t] \tag{3}$$

where $q_{i,t}$ is the log quote set by agent i at time t and $p_{i,t}$ is a transaction price for a property sold at time t. Agents do not observe the realisation of δ_t and will, therefore, try to estimate its value from available information. In particular, the informational structure is characterised by the presence of public and private signals. Each seller receives a private signal $s^q_{i,t}$ before choosing the quote:

$$s^q_{i,t} = \delta_t + \nu_{i,t} \,,\, \nu_{i,t} \overset{iid}{\sim} \mathcal{N}(0,\, \sigma_\nu^2) \tag{4}$$

The noise terms are independent and identically distributed across individuals and time. If there are multiple sellers in a given period, they all receive a different private signal. There is also one publicly observable signal s_t arriving every k periods:

$$s_t = \delta_t + u_t \, , \, u_t \overset{iid}{\sim} \mathcal{N}(0, \, \sigma_u^2) \tag{5}$$

In this case too, the noise is identically and independently distributed across time. The public signal represents any public information that sellers might use to make inference about housing demand and prices, for instance a housing price index published at regular frequencies. Alternatively, we can interpret it as representing local area characteristics or amenities visible to everyone. Finally, sellers also observe the full history of past transactions and listings which they also use to extract the private signals that agents in previous periods have received. The signal contained in past prices has the same form as the other two signals but possibly a different precision:

$$s_{i,t}^p = \delta_t + \eta_{i,t} \, , \, \eta_{i,t} \overset{iid}{\sim} \mathcal{N}(0, \, \sigma_\eta^2) \tag{6}$$

I make a distinction between the private signals embedded in prices and those from quotes to account for the fact that the final transaction price can be adjusted upon interaction with the buyer. In other words, the signal extracted from transaction prices contains additional information about demand and (idiosyncratic) buyer characteristics to the extent that buyers have some bargaining power. This approach of modelling private signals differently across quotes and prices can be seen as reduced-form way of capturing housing market features common in the housing literature without resorting to more complicated search models.[1] At every point in time, the information sets of agents consist of the full history of past prices, the history of public signals and their own private signals. They thus form conditional expectations of δ_t and set quotes accordingly, as follows:

$$p_{i,t} = \mathbb{E}[\delta_t | s_{i,t}^p, s^t, p^{t-1}, q^{t-1}] \tag{7}$$

[1]The specifics of the housing market microstructure and the bargaining process go beyond the scope of this paper. Here, I simply attempt to provide some evidence of the economic magnitude of the effect of naïve inference by sellers. Using a more evolved search model would render the interpretation of the results more difficult without changing the big picture. See Han and Strange (2015) for a survey of the literature on the microstructure of housing markets.

$$q_{i,t} = \mathbb{E}[\delta_t | s^q_{i,t}, s^t, p^{t-1}, q^{t-1}] \tag{8}$$

where $s^t = \{s_0, s_{0+k}, s_{0+2k}, ...\}, p^{t-1} = \{p_0, p_1, ..., p_{t-1}\}, q^{t-1} = \{q_0, q_1, ..., q_{t-1}\}$ denote the full history of public signals, transaction prices and listing prices, respectively, that agents at time t observe[2], while $s^p_{i,t}$ or $s^q_{i,t}$ is agent i's private signal.

A.1 Parameter Calibration

The goal of this exercise is to help us gain understanding about the way that the effect of a given price p on subsequent listings q would evolve as the number of interim comparables increases under the Bayesian learning framework. Specifically, I will use the setting of Figure 8 and Table 7 where I look at the effect of recent transaction prices on quotes posted in week two after the publication date. As in Table 7, there are four types of prospective sellers depending on what is in their information set:

- Seller 1 observes the newly published price data but has no comparable listing in the two-week period around the publishing date;

- Seller 2 observes the newly published price data and has at least one comparable listing in the week before and no comparable in the week after the publishing date;

- Seller 3 observes the newly published price data and has no comparable listing in the week before and at least one comparable in the week after the publishing date, and;

- Seller 4 observes the newly published price data and has at least one comparable listing both in the week before and the week after the publishing date.

The object of interest for this analysis is the covariance of four types of quotes with the most recently published prices. For simplicity, I assume that the newly published comparable transaction price, denoted as p_0, has been determined based on agent 0's private

[2] As price data are published on a monthly basis, agents observe the history of past prices up until the previous month. Quotes can, however, be observed at a higher frequency on property websites. Here I assume that sellers can observe all past listings up to the previous week.

signal $s^p_{0,0}$ and a public signal[3] that englobes the full history of past information s^0:

$$p_0 = \mathbb{E}[\delta_0 | s^p_{0,0}, s^0] = w^p_0 \times s^p_{0,0} + (1 - w^p_0) \times s^0 \tag{9}$$

where the weight that agent 0 assigns to his private signal is proportional to the signal precision, i.e., $w^p_0 = \frac{\sigma_\eta^{-2}}{\sigma_\eta^{-2} + \sigma_w^{-2}}$. The variance of the posterior belief of agent 0 $P_{0|0}$ is given by:

$$P_{0|0} = (w^p_0)^2 \times \sigma^2_\eta + (1 - w^p_0)^2 \times \sigma^2_w \tag{10}$$

One period later, at $t = 1$, prospective sellers determine listing prices using available information. Type 1 sellers observe p_0, their own private signal $s^q_{1,1}$, a new public signal s_1[4] and they also directly observe the same public signal s^0 that has already been accounted for by agent 0. Rational sellers understand that p_0 already incorporates the original public signal s^0 and thus avoid double-counting it. They form their posterior belief about δ_1, and hence the quote, as follows:

$$\begin{aligned} q_{1,1} &= \mathbb{E}[\delta_1 | s^q_{1,1}, s_1, p_0, s^0] = \mathbb{E}[\delta_1 | s^q_{1,1}, s_1, p_0] \\ &= w^q_1 \times \left[\frac{\sigma_\nu^{-2}}{\sigma_\nu^{-2} + \sigma_u^{-2}} \times s^q_{1,1} + \frac{\sigma_u^{-2}}{\sigma_\nu^{-2} + \sigma_u^{-2}} \times s_1 \right] + (1 - w^q_1) \times (a + \rho \times p_0) \end{aligned} \tag{11}$$

where $w^q_1 = P_{1|0} \times \left(P_{1|0} + \left(\frac{\sigma_\nu^{-2}}{\sigma_\nu^{-2} + \sigma_u^{-2}} \right)^2 \times \sigma^2_\nu + \left(\frac{\sigma_u^{-2}}{\sigma_\nu^{-2} + \sigma_u^{-2}} \right)^2 \times \sigma^2_u \right)^{-1}$ and $P_{1|0} = \rho^2 \times P_{0|0} + \sigma^2_\varepsilon$ is the variance of agent 1's prior belief about δ_1 given the available information up to time 0. Note that under the Bayesian learning framework, agent 1 does not assign an explicit weight on the initial public signal s^0, rather he treats p_0 as a sufficient statistic for all information up to $t = 0$ being aware that it already embeds s^0. Naïve sellers, however, fail to recognise this, believing that p_0 is solely determined based on agent 0's private signal, i.e., they believe $p_0 = \widetilde{\mathbb{E}}[\delta_0 | s^p_{0,0}]$. As a result, they treat the newly observed price as independent from the public signal, leading them to assign an explicit weight to

[3]The public signal here can be interpreted as the prior belief of agent 0 based on his information set before receiving the private signal. For this reason, I allows this prior to have a different precision than the periodic public signals, as follows: $s^0 = \delta_0 + w_0$, $w_0 \overset{iid}{\sim} \mathcal{N}(0,\, \sigma^2_w)$.

[4]For this exercise, I assume that a new public signal arrives every period, i.e., k=1.

the public signal when forming beliefs about the state of housing demand:

$$\begin{aligned}\widetilde{q}_{1,1} &= \widetilde{\mathbb{E}}[\delta_1 | s^q_{1,1}, s_1, p_0, s^0] \\ &= w^q_1 \times \left[\frac{\sigma_\nu^{-2}}{\sigma_\nu^{-2} + \sigma_u^{-2}} \times s^q_{1,1} + \frac{\sigma_u^{-2}}{\sigma_\nu^{-2} + \sigma_u^{-2}} \times s_1 \right] + \\ &\quad (1 - w^q_1) \times [a + w^p_0 \times \rho p_0 + (1 - w^p_0) \times \rho s^0]\end{aligned} \tag{12}$$

where the weight w^q_1 is the same as for the Bayesian case. Notice that when forming his prior belief the naïve agents assign weights to p_0 and s^0 that a Bayesian learner would assign to the correctly extracted signals $s^p_{0,0}$ and s^0, respectively. The problem with naïve learners is that they believe that past actions are purely driven by private signals, $p_0 = s^p_{0,0}$, when instead p_0 has effectively been determined using all available information at $t = 0$, as in equation (9). This leads naïve agents to overweight the commonly observed public signal relative to Bayesian learners, as they account for it both directly (through its explicit weight in the prior belief) and indirectly (through its effect on p_0).

Type 2 sellers observe p_0 and another listing, denoted by $q_{0,1}$, posted before the price data publication date[5]. They also observe the original public signal s^0, the public signal from period 1, s_1, and a new private signal, $s^q_{2,1}$. The Bayesian learner would extract only the private signal embedded in $q_{0,1}$ and form beliefs as follows:

$$\begin{aligned}q_{2,1} &= \mathbb{E}[\delta_1 | s^q_{2,1}, q_{0,1}, s_1, p_0, s^0] = \mathbb{E}[\delta_1 | s^q_{2,1}, s^q_{0,1}, s_1, p_0] \\ &= w^q_2 \times \left[\frac{\sigma_\nu^{-2}}{2\sigma_\nu^{-2} + \sigma_u^{-2}} \times (s^q_{2,1} + s^q_{0,1}) + \frac{\sigma_u^{-2}}{2\sigma_\nu^{-2} + \sigma_u^{-2}} \times s_1 \right] + (1 - w^q_2) \times [a + \rho \times p_0]\end{aligned} \tag{13}$$

where $w^q_2 = P_{1|0} \times \left(P_{1|0} + 2 \times \left(\frac{\sigma_\nu^{-2}}{2\sigma_\nu^{-2} + \sigma_u^{-2}} \right)^2 \times \sigma^2_\nu + \left(\frac{\sigma_u^{-2}}{2\sigma_\nu^{-2} + \sigma_u^{-2}} \right)^2 \times \sigma^2_u \right)^{-1}$. Notice that both agent 2 and the interim agent forming quote $q_{0,1}$ observe s^0 and s_1. Since seller 2 knows their precisions, he can easily infer what the private signal embedded in $q_{0,1}$ is and avoid double counting information. His posterior belief about δ_1 is thus equal to a weighted-average of the prior belief (for which p_0 is again a sufficient statistic) and the

[5]Note that the interim agent does not observe p_0 and therefore cannot learn agent 0's private signal. He thus sets his quote $q_{0,1}$ based on the original public signal s^0, the new public signal from period 1, s_1, and his own private signal, $s^q_{0,1}$, as follows: $q_{0,1} = \mathbb{E}[\delta_1 | s^q_{0,1}, s_1, s^0] = w^q_0 \times \left[\frac{\sigma_\nu^{-2}}{\sigma_\nu^{-2} + \sigma_u^{-2}} \times s^q_{0,1} + \frac{\sigma_u^{-2}}{\sigma_\nu^{-2} + \sigma_u^{-2}} \times s_1 \right] + (1 - w^q_0) \times (a + \rho s^0)$, where $w^q_0 = (\rho^2 \times \sigma^2_w + \sigma^2_\varepsilon) \times \left(\rho^2 \times \sigma^2_w + \sigma^2_\varepsilon + \left(\frac{\sigma_\nu^{-2}}{\sigma_\nu^{-2} + \sigma_u^{-2}} \right)^2 \times \sigma^2_\nu + \left(\frac{\sigma_u^{-2}}{\sigma_\nu^{-2} + \sigma_u^{-2}} \right)^2 \times \sigma^2_u \right)^{-1}$.

average of the newly obtained signals in period 1, i.e., $s^q_{2,1}$, $s^q_{0,1}$ and s_1, weighted by their precisions. The naïve sellers of type 2 make the same mistake as the type 1 naïve sellers, i.e., they incorrectly believe that $p_0 = \widetilde{\mathbb{E}}[\delta_0|s^p_{0,0}]$ and $q_{0,1} = \widetilde{\mathbb{E}}[\delta_1|s^q_{0,1}]$. As a result, the effect of naïve inference is now two-fold as the initial public signal has been embedded both in p_0 and $q_{0,1}$. Since the private signals received by agent 2 and the interim agent who sets $q_{0,1}$ are equally-precise, agent 2 sets his quote as follows:

$$\begin{aligned}\widetilde{q}_{2,1} &= \widetilde{\mathbb{E}}[\delta_1|s^q_{2,1}, s_1, q_{0,1}, p_0, s^0] \\ &= w^q_2 \times \left[\frac{\sigma_\nu^{-2}}{2\sigma_\nu^{-2} + \sigma_u^{-2}} \times (s^q_{2,1} + q_{0,1}) + \frac{\sigma_u^{-2}}{2\sigma_\nu^{-2} + \sigma_u^{-2}} \times s_1\right] + \\ &\quad (1 - w^q_2) \times [a + w^p_0 \times \rho p_0 + (1 - w^p_0) \times \rho s^0]\end{aligned} \tag{14}$$

Similarly to seller 1, the naïve type 2 seller assigns weights that would be correct if past prices and quotes were truly equal to the private signals of the preceding agents. As this is not the case, however, seller 2 ends up overweighting the common signal through two indirect channels: its influence on p_0 and that on $q_{0,1}$.

Moving on to type 3 sellers, recall that the only difference with type 2 sellers is that they observe listing $q_{1,1}$, instead of $q_{0,1}$, which is set after the publication date and, in turn, directly observes p_0 as well. Rational agents would recognise this and extract the private signal from $q_{1,1}$ in order to avoid double counting the public signals s^0 and s_1 as well as the private information coming from p_0. Accordingly, it follows that the weights they would assign are the same as for the rational type 2 sellers:

$$\begin{aligned}q_{3,1} &= \mathbb{E}[\delta_1|s^q_{3,1}, q_{1,1}, s_1, p_0, s^0] = \mathbb{E}[\delta_1|s^q_{3,1}, s^q_{1,1}, s_1, p_0] \\ &= w^q_2 \times \left[\frac{\sigma_\nu^{-2}}{2\sigma_\nu^{-2} + \sigma_u^{-2}} \times (s^q_{3,1} + s^q_{1,1}) + \frac{\sigma_u^{-2}}{2\sigma_\nu^{-2} + \sigma_u^{-2}} \times s_1\right] + (1 - w^q_2) \times [a + \rho \times p_0]\end{aligned} \tag{15}$$

Similarly, naïve type 3 sellers will assign the same weights as naïve type 2 sellers under the beliefs that $p_0 = \widetilde{\mathbb{E}}[\delta_0|s^p_{0,0}]$ and $q_{1,1} = \widetilde{\mathbb{E}}[\delta_1|s^q_{1,1}]$. As a result, they set the quote as follows:

$$\begin{aligned}\widetilde{q}_{3,1} &= \widetilde{\mathbb{E}}[\delta_1 | s^q_{3,1}, s_1, \widetilde{q}_{1,1}, p_0, s^0] \\ &= w^q_2 \times \left[\frac{\sigma_\nu^{-2}}{2\sigma_\nu^{-2} + \sigma_u^{-2}} \times (s^q_{3,1} + \widetilde{q}_{1,1}) + \frac{\sigma_u^{-2}}{2\sigma_\nu^{-2} + \sigma_u^{-2}} \times s_1\right] + \\ &\quad (1 - w^q_2) \times [a + w^p_0 \times \rho p_0 + (1 - w^p_0) \times \rho s^0]\end{aligned} \tag{16}$$

An important distinction arises when comparing naïve agents of types 2 and 3. Namely, although they assign the exact same weights, the fact that $\widetilde{q}_{1,1}$, unlike $q_{0,1}$ is formed using information in p_0 could lead to different covariances with p_0 between type 2 and 3 sellers. This is because agent 3 has two different channels of influence from the private signal in p_0, one through the direct effect of p_0 on $\widetilde{q}_{3,1}$ and another due to the indirect effect of p_0 through $\widetilde{q}_{1,1}$. In addition, there are now four implicit weights on the public signal: the direct effect, the indirect effect through p_0 and the indirect effect through $\widetilde{q}_{1,1}$, which can further be decomposed into its direct effect on $\widetilde{q}_{1,1}$ and the indirect effect on $\widetilde{q}_{1,1}$ through p_0. Moreover, depending on the parameters, naïve agents of type 3 might either overweight or underweight the private signal contained in p_0 relative to Bayesian learners.

Finally, sellers of type 4 have the richest information set: they observe p_0 as well as both $q_{0,1}$ and $q_{2,1}$, in addition to the private and public signals, $s^q_{4,1}$, s_1 and s^0. As usual, rational type 4 sellers will extract and use only the private information from the intermediate quotes, leading to the following beliefs:

$$\begin{aligned}q_{4,1} &= \mathbb{E}[\delta_1 | s^q_{4,1}, q_{2,1}, q_{0,1}, s_1, p_0, s^0] = \mathbb{E}[\delta_1 | s^q_{4,1}, s^q_{2,1}, s^q_{0,1}, s_1, p_0] \\ &= w^q_4 \times \left[\frac{\sigma_\nu^{-2}}{3\sigma_\nu^{-2} + \sigma_u^{-2}} \times (s^q_{4,1} + s^q_{2,1} + s^q_{0,1}) + \frac{\sigma_u^{-2}}{3\sigma_\nu^{-2} + \sigma_u^{-2}} \times s_1\right] + \\ &\quad (1 - w^q_4) \times [a + \rho \times p_0]\end{aligned} \tag{17}$$

where $w^q_4 = P_{1|0} \times \left(P_{1|0} + 3 \times \left(\frac{\sigma_\nu^{-2}}{3\sigma_\nu^{-2}+\sigma_u^{-2}}\right)^2 \times \sigma_\nu^2 + \left(\frac{\sigma_u^{-2}}{3\sigma_\nu^{-2}+\sigma_u^{-2}}\right)^2 \times \sigma_u^2\right)^{-1}$. Naïve type 4 agents instead, believing that $p_0 = \widetilde{\mathbb{E}}[\delta_0 | s^p_{0,0}]$, $q_{0,1} = \widetilde{\mathbb{E}}[\delta_1 | s^q_{0,1}]$ and $q_{2,1} = \widetilde{\mathbb{E}}[\delta_1 | s^q_{2,1}]$, assign weights as follows:

$$\begin{aligned}\widetilde{q}_{4,1} &= \widetilde{\mathbb{E}}[\delta_1|s^q_{4,1}, \widetilde{q}_{2,1}, q_{0,1}, s_1, p_0, s^0] \\ &= w^q_4 \times \left[\frac{\sigma_\nu^{-2}}{3\sigma_\nu^{-2} + \sigma_u^{-2}} \times (s^q_{4,1} + \widetilde{q}_{2,1} + q_{0,1}) + \frac{\sigma_u^{-2}}{3\sigma_\nu^{-2} + \sigma_u^{-2}} \times s_1\right] + \\ &\quad (1 - w^q_4) \times [a + w^p_0 \times \rho p_0 + (1 - w^p_0) \times \rho s^0]\end{aligned} \tag{18}$$

The initial public signal s^0 influences type 4 sellers via six different channels, the private signal from p_0 affects $\widetilde{q}_{4,1}$ via two channels and the new public signal s_1 is also counted multiple times through its effect on $q_{0,1}$, $\widetilde{q}_{2,1}$ and the directly assigned weight. The results above show how agents overweight past news at the expense of more recent signals and this generates differences in the covariance patterns between quotes and prices under the naïve model relative to the Bayesian case. The equations also point to the fact that naïve agents tend to overweight commonly observed signals; on the other hand, it is not obvious whether they over- or under-weight the private signal coming from p_0 as this will depend on the relative signal precisions.

To more clearly see the differences in comovement patterns under the two models, I compute the covariances of the four sets of quotes with p_0 for both the Bayesian and naïve cases. It can be shown that the covariances in the Bayesian case will take the following forms:

$$Cov(q_{1,1}, p_0) = \rho Var(\delta_0) + (1 - w^q_1)(w^p_0)^2 \times \rho\sigma^2_\eta + (1 - w^q_1)(1 - w^p_0)^2 \times \rho\sigma^2_w \tag{19}$$

$$Cov(q_{2,1}, p_0) = \rho Var(\delta_0) + (1 - w^q_2)(w^p_0)^2 \times \rho\sigma^2_\eta + (1 - w^q_2)(1 - w^p_0)^2 \times \rho\sigma^2_w \tag{20}$$

$$Cov(q_{3,1}, p_0) = Cov(q_{2,1}, p_0) \tag{21}$$

$$Cov(q_{4,1}, p_0) = \rho Var(\delta_0) + (1 - w^q_4)(w^p_0)^2 \times \rho\sigma^2_\eta + (1 - w^q_4)(1 - w^p_0)^2 \times \rho\sigma^2_w \tag{22}$$

Note that the only difference in the covariance expressions above is in the weights $(1-w^q_i)$ that multiply both terms related to the respective inverse precisions of the private signal $s^p_{0,0}$ and the initial public signal s^0. As these weights are decreasing with i, it follows that Bayesian updating would imply that the covariances, and therefore the betas, should be monotonically decreasing with the increase in the number of intermediate comparables, regardless of the parameter values. In other words, as the information set of agents grows, they optimally assign a lower weight to each individual signal. Benchmarking the results of the empirical analysis from Section 5.2 against these predictions, we can reject the hypothesis that sellers in the housing market act in a Bayesian way. On the other hand, we can derive the same covariances for the naïve case and compare:

$$Cov(\widetilde{q}_{1,1}, p_0) = \rho Var(\delta_0) + (1 - w_1^q)(w_0^p)^3 \times \rho\sigma_\eta^2 + (1 - w_1^q)(1 - w_0^p)^2(1 + w_0^p) \times \rho\sigma_w^2 \quad (23)$$

$$\begin{aligned} Cov(\widetilde{q}_{2,1}, p_0) = \; & \rho Var(\delta_0) + (1 - w_2^q)(w_0^p)^3 \times \rho\sigma_\eta^2 + \\ & [w_2^q k_2(1 - w_0^q) + (1 - w_2^q)(1 - w_0^p)(1 + w_0^p)](1 - w_0^p) \times \rho\sigma_w^2 \end{aligned} \quad (24)$$

$$\begin{aligned} Cov(\widetilde{q}_{3,1}, p_0) = \; & \rho Var(\delta_0) + [w_2^q k_2(1 - w_1^q) + (1 - w_2^q)](w_0^p)^3 \times \rho\sigma_\eta^2 + \\ & [w_2^q k_2(1 - w_1^q) + (1 - w_2^q)](1 + w_0^p)(1 - w_0^p)^2 \times \rho\sigma_w^2 \end{aligned} \quad (25)$$

$$\begin{aligned} Cov(\widetilde{q}_{4,1}, p_0) = \; & \rho Var(\delta_0) + [w_4^q k_4(1 - w_2^q) + (1 - w_4^q)](w_0^p)^3 \times \rho\sigma_\eta^2 + \\ & [w_4^q k_4(1 - w_0^q)(1 + w_2^q k_2) + w_4^q k_4(1 - w_2^q)(1 - w_0^p)(1 + w_0^p) + \\ & (1 - w_4^q)(1 - w_0^p)(1 + w_0^p)](1 - w_0^p) \times \rho\sigma_w^2 \end{aligned} \quad (26)$$

where $k_2 = \frac{\sigma_\nu^{-2}}{2\sigma_\nu^{-2} + \sigma_u^{-2}}$ and $k_4 = \frac{\sigma_\nu^{-2}}{3\sigma_\nu^{-2} + \sigma_u^{-2}}$. The covariances will be larger in the naïve case due to overweighting of stale information embedded in p_0 relative to new signals coming from intermediate comparables. We can note that the original public signal will always be more heavily weighted in the covariances between naïve quotes and prices relative to the rational case. This is because, on top of assigning the optimal direct weight to it, naïve sellers also get indirectly influenced through its effect on previous prices/quotes. On the other hand, the private signal coming from p_0 might be both under- or over-weighted, depending on the relative precisions.

As the results in Section 5.2 cannot be reconciled with Bayesian updating, I postulate that sellers are subject to naïve learning and use these results in a calibration exercise. Specifically, I estimate the signal precisions described above, i.e., the precisions of the original and periodic public signal, σ_w^{-2} and σ_u^{-2}, and the two types of private signals, σ_η^{-2} and σ_ν^{-2}, using equations (23)-(26) and the results from Table 7. The parameters that govern the underlying process are estimated by running a monthly regression of log aggregate house prices on aggregate income. The predicted values of the above regression are then used to fit an AR(1) process that yields estimates for a, ρ and σ_ε^2. The calibrated parameters are the used in simulations in order to evaluate the aggregate impact of naïve inference on house prices in the long run.

A.2 Estimating the Magnitude of the Effect of Information Shocks under the Bayesian and Naïve Filters

Given the model described in equations (1)-(8) above, we can trace the learning process of sellers who act sequentially. Assuming there are n agents per period, the first set of agents sets prices using their own private signal and the public signal[6,7]. For simplicity, I here assume that there is a single public signal arriving at $t = 0$, however, I vary the frequency of public signal arrival in the simulations. I first describe the updating process for fully rational agents and subsequently specify how this differs from naïve updating. Let us denote the posterior belief of agents at time t by $\delta_{t|t}$, it then follows that:

$$p_0^i = \delta_{0|0}^i = \mathbb{E}[\delta_0 | s_0^i, s^0] = w_0 \times s_0^i + (1 - w_0) \times s^0 \tag{27}$$

where $w_0 = \frac{\sigma_\eta^{-2}}{\sigma_\eta^{-2} + \sigma_u^{-2}}$, as before. Denoting the variance of the posterior beliefs of sellers acting in period t by $P_{t|t}$, we have:

$$P_{0|0} = w_0^2 \times \sigma_\eta^2 + (1 - w_0)^2 \times \sigma_u^2 \tag{28}$$

The second set of agents in period $t = 1$ observe the same public signal s^0 and the n prices from the previous period, along with their own private signals. Unlike in the standard single-file example, their prior belief, therefore, is not simply equal to the posterior of any of the preceding agents, rather it is a function of the average private signal from the previous period. Denoting this prior belief by $\delta_{t|t-1}$, we obtain:

$$\delta_{1|0} = a + \widetilde{w}_0 \times \rho \bar{s}_0 + (1 - \widetilde{w}_0) \times \rho s^0 \tag{29}$$

where $\widetilde{w}_0 = \frac{n\sigma_\eta^{-2}}{\sigma_u^{-2} + n\sigma_\eta^{-2}}$ and $\bar{s}_0$ is the equally-weighted average of private signals at time 0. The variance of the prior belief of agents in period t is denoted as $P_{t|t-1}$. This can be

[6] I assume that same-period agents do not observe each other's actions and thus cannot use each other's signals to inform their decisions.

[7] For the simulation exercise, I focus solely on final transaction prices and, therefore, disregard the quote setting procedure. As a result, the only relevant type of private signal is the one embedded in final prices, i.e., $s_{i,t}^p$. For ease of exposition, I hereafter denote this signal simply by s_t^i.

computed recursively as follows:

$$P_{1|0} = \rho^2 \widetilde{P}_{0|0} + \sigma_\varepsilon^2 \tag{30}$$

where $\widetilde{P}_{0|0} = \widetilde{w}_0^2 \times \frac{1}{n}\sigma_\eta^2 + (1-\widetilde{w}_0)^2 \times \sigma_u^2$ in order to adjust for the fact that there are n private signals coming from period $t-1$. Each period agent i forms his posterior belief and the price by mixing the above prior and his private signal:

$$p_1^i = \delta_{1|1}^i = w_1 \times s_1^i + (1-w_1) \times \delta_{1|0} \tag{31}$$

where $w_1 = P_{1|0} \times (P_{1|0} + \sigma_\eta^2)^{-1}$ is the Kalman gain. From here onward, we can define the recursion through which agents form and update their beliefs in a sequential way. The prior beliefs are computed by adjusting for the number of observations from the previous period:

$$\delta_{t|t-1} = a + \widetilde{w}_{t-1} \times \rho \bar{s}_{t-1} + (1-\widetilde{w}_{t-1}) \times \rho \delta_{t-1|t-2} \tag{32}$$

where $\widetilde{w}_{t-1} = P_{t-1|t-2} \times (P_{t-1|t-2} + \frac{1}{n}\sigma_\eta^2)^{-1}$. The variance of this prior can be computed as follows:

$$P_{t|t-1} = \rho^2 P_{t-1|t-1} + \sigma_\varepsilon^2 = \rho^2 [P_{t-1|t-2} - P_{t-1|t-2}^2 (P_{t-1|t-2} + \frac{1}{n}\sigma_\eta^2)^{-1}] + \sigma_\varepsilon^2 \tag{33}$$

Finally, agent i forms his posterior belief updated for his private signal and sets the price accordingly:

$$p_t^i = \delta_{t|t}^i = w_t \times s_t^i + (1-w_t) \times \delta_{t|t-1} \tag{34}$$

where $w_t = P_{t|t-1}(P_{t|t-1} + \sigma_\eta^2)^{-1}$. Plugging in the expressions for the prior beliefs recursively, we can outline the way that prices depend on all past signals:

$$\begin{aligned} p_{i,t} = \delta_{t|t}^i = & w_t s_t^i + (1-w_t)\widetilde{w}_{t-1}(a + \rho \bar{s}_{t-1}) + \\ & (1-w_t)(1-\widetilde{w}_{t-1})\widetilde{w}_{t-2}(a + \rho a + \rho^2 \bar{s}_{t-2}) + ... + \\ & (1-w_t)(1-\widetilde{w}_{t-1})(1-\widetilde{w}_{t-2})...(1-\widetilde{w}_1)\widetilde{w}_0(a + \rho a + ... + \rho^{t-1} a + \rho^t \bar{s}_0) + \\ & (1-w_t)(1-\widetilde{w}_{t-1})(1-\widetilde{w}_{t-2})...(1-\widetilde{w}_1)(1-\widetilde{w}_0)(a + \rho a + ... + \rho^{t-1} a + \rho^t s^0) \end{aligned} \tag{35}$$

The difference between rational and naïve sellers is that naïve learners treat all past prices as independent signals by failing to account for the fact that past sellers have similarly set prices by looking at yet earlier prices. They, therefore, assign the same weights as the rational agents but directly to the observed prices as opposed to the signals extracted:

$$\begin{aligned}\widetilde{p}_{i,t} = \widetilde{\delta}^{i}_{t|t} =& w_t s^i_t + (1-w_t)\widetilde{w}_{t-1}(a+\rho\bar{\bar{p}}_{t-1}) + \\ & (1-w_t)(1-\widetilde{w}_{t-1})\widetilde{w}_{t-2}(a+\rho a+\rho^2\bar{\bar{p}}_{t-2}) + ... + \\ & (1-w_t)(1-\widetilde{w}_{t-1})(1-\widetilde{w}_{t-2})...(1-\widetilde{w}_1)\widetilde{w}_0(a+\rho a+...+\rho^{t-1}a+\rho^t\bar{p}_0) + \\ & (1-w_t)(1-\widetilde{w}_{t-1})(1-\widetilde{w}_{t-2})...(1-\widetilde{w}_1)(1-\widetilde{w}_0)(a+\rho a+...+\rho^{t-1}a+\rho^t s^0)\end{aligned} \tag{36}$$

This leads them to overweight old news at the expense of more recent information since these old news have already been accounted for by more recent sellers. To test the magnitude of the effect of naïve learning given the estimates obtained in the empirical analysis, I simulate a market with the above characteristics and compare the impact of various information shocks on prices in the rational and naïve settings.

B Additional Results

B.1 Tables

Table 1 Direct Effect of Transaction Prices on Quotes - Sample Refinements

The table presents the results of the following regression: $log(q_i) = \alpha + \beta_0 \times log(p_j) + \beta_1 \times log(p_j) \times Treated_i + \gamma Treated_i + Controls + \varepsilon_i$, where q_i is the listed price for property i, p_j is the transaction price for a comparable property j sold in the previous month and *Treated* is a dummy that turns on when the listing price has been set in the period following the price data publication date. Columns (1)-(2) present the results for the sample of listings that excludes those posted exactly on publishing dates; columns (3)-(4) restrict the sample to listings posted in the two weeks around the publishing date and; columns (5)-(6) limit the number of comparables to no more than 30 per listing. Controls for the time distance between the listing and the comparable transaction measured in days and its interaction with the log price are included in all columns. Fixed-effects included are: listing month-year dummies in all columns, and; transaction ID dummies in columns (2), (4) and (6). Standard errors double-clustered at the transaction and listing ID levels are reported in parentheses.

	No quotes on pub dates		Within 7 days of pub date		Less than 30 comps	
	(1)	(2)	(3)	(4)	(5)	(6)
Price × Treated	0.0041***	0.0031***	0.0056***	0.0038**	0.0035**	0.0033***
	(0.0016)	(0.0011)	(0.0021)	(0.0018)	(0.0015)	(0.0011)
Price	0.8400***		0.8457***		0.8387***	
	(0.0023)		(0.0037)		(0.0023)	
Treated	-0.0501***	-0.0351**	-0.0637**	-0.0406*	-0.0423**	-0.0373***
	(0.0193)	(0.0138)	(0.0254)	(0.0210)	(0.0183)	(0.0132)
Controls						
Price × Time distance	Yes	Yes	Yes	Yes	Yes	Yes
Fixed-Effects						
Month-year	Yes	Yes	Yes	Yes	Yes	Yes
Transaction ID	No	Yes	No	Yes	No	Yes
Observations	7,075,069	7,075,069	2,908,410	2,908,410	7,421,440	7,421,440
R^2	0.7052	0.8696	0.7058	0.8800	0.7056	0.8694
Within R^2	0.7033	0.0000	0.7036	0.0001	0.7038	0.0000

Two-way (Transaction ID & Listing ID) standard-errors in parentheses

*Signif Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 2 Direct Effect of Transaction Prices on Quotes - Sensitivity to Publication Days

The table presents the results of the following regression: $log(q_i) = \alpha + \beta_0 \times log(p_j) + \beta_1 \times log(p_j) \times Treated_i + \gamma Treated_i + Controls + \varepsilon_i$, where q_i is the listed price for property i, p_j is the transaction price for a comparable property j sold in the previous month and *Treated* is a dummy that turns on when the listing price has been set in the period following the price data publication date. In columns (1)-(2) dummies for the week day of the publishing date are also interacted with price p_j and *Treated*, i.e., the following specification is used: $log(q_i) = \alpha + \beta_0 \times log(p_j) + \beta_1 \times log(p_j) \times Treated_i + \sum_{d=1}^{5} \beta_d \times log(p_j) \times Treated_i \times Pub\ day\ d + \gamma_0 Treated_i + \sum_{d=1}^{5} \times \gamma_d \times Pub\ day\ d + Controls + \varepsilon_i$, where *Pub day d* is a dummy for the day of the week at which that publication date fell on. Columns (3)-(4) restrict the sample to cases where the publishing date occurred at the end of the month, while columns (5)-(6) to cases where it fell at the beginning of the next month. Controls for the time distance between the listing and the comparable transaction measured in days and its interaction with the log price are included in all columns. Fixed-effects included are: listing month-year dummies in all columns, and; transaction ID dummies in columns (2), (4) and (6). Standard errors double-clustered at the transaction and listing ID levels are reported in parentheses.

	Pub day of week		Pub day at end of month		Pub day at beginning of month	
	(1)	(2)	(3)	(4)	(5)	(6)
Price × Treated	0.0057***	0.0034***	0.0040**	0.0033***	0.0107*	0.0057
	(0.0019)	(0.0011)	(0.0016)	(0.0011)	(0.0063)	(0.0045)
Price × Treated × Monday	-0.0019	0.0007*				
	(0.0034)	(0.0004)				
Price × Treated × Tuesday	-0.0070**	−0.0000				
	(0.0030)	(0.0003)				
Price × Treated × Wednesday	0.0040	-0.0004				
	(0.0030)	(0.0003)				
Price × Treated × Thursday	-0.0048*	0.0000				
	(0.0029)	(0.0003)				
Price	0.8402***		0.8411***		0.8225***	
	(0.0023)		(0.0023)		(0.0097)	
Treated	-0.0669***	-0.0383***	-0.0483**	-0.0365***	-0.1131	-0.0537
	(0.0233)	(0.0134)	(0.0190)	(0.0138)	(0.0789)	(0.0576)
Controls						
Price × Time distance	Yes	Yes	Yes	Yes	Yes	Yes
Fixed-Effects						
Month-year	Yes	Yes	Yes	Yes	Yes	Yes
Transaction ID	No	Yes	No	Yes	No	Yes
Observations	7,466,950	7,466,950	6,959,170	6,959,170	507,780	507,780
R^2	0.7050	0.8689	0.7072	0.8701	0.6735	0.8525
Within R^2	0.7032	0.0000	0.7055	0.0000	0.6697	0.0002

Two-way (Transaction ID & Listing ID) standard-errors in parentheses

*Signif Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 3

Direct Effect of Transaction Prices on Quotes - Existing Houses and Per Price Range

The table presents the results of the following regression: $log(q_i) = \alpha + \beta_0 \times log(p_j) + \beta_1 \times log(p_j) \times Treated_i + \gamma Treated_i + Controls + \varepsilon_i$, where q_i is the listed price for property i, p_j is the transaction price for a comparable property j sold in the previous month and *Treated* is a dummy that turns on when the listing price has been set in the period following the price data publication date. Columns (1)-(2) present the results for the sample of listings that excludes newly-built properties; columns (3)-(4) restrict the sample to quotes that are below median, and; columns (5)-(6) to quotes above median. Controls for the time distance between the listing and the comparable transaction measured in days and its interaction with the log price are included in all columns. Fixed-effects included are: listing month-year dummies in all columns, and; transaction ID dummies in columns (2), (4) and (6). Standard errors double-clustered at the transaction and listing ID levels are reported in parentheses.

	Existing houses only		Price below median		Price above median	
	(1)	(2)	(3)	(4)	(5)	(6)
Price × Treated	0.0045***	0.0029***	0.0046*	0.0040**	0.0081**	0.0028
	(0.0015)	(0.0011)	(0.0025)	(0.0017)	(0.0032)	(0.0020)
Price	0.8431***		0.5467***		0.5934***	
	(0.0023)		(0.0034)		(0.0048)	
Treated	-0.0545***	-0.0333**	-0.0547*	-0.0468**	-0.1000**	-0.0306
	(0.0186)	(0.0134)	(0.0294)	(0.0206)	(0.0407)	(0.0248)
Controls						
Price × Time distance (days)	Yes	Yes	Yes	Yes	Yes	Yes
Fixed-Effects						
Month-year	Yes	Yes	Yes	Yes	Yes	Yes
Transaction ID	No	Yes	No	Yes	No	Yes
Observations	7,225,115	7,225,115	3,919,114	3,919,114	3,089,563	3,089,563
R^2	0.7093	0.8719	0.3981	0.7039	0.4363	0.7309
Within R^2	0.7076	0.0000	0.3958	0.0000	0.4327	0.0000

Two-way (Transaction ID & Listing ID) standard-errors in parentheses

*Signif Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 4 Direct Effect of Transaction Prices on Quotes - Controlling for Listing Agency

The table presents the results of the following regression: $log(q_i) = \alpha + \beta_0 \times log(p_j) + \beta_1 \times log(p_j) \times Treated_i + \gamma Treated_i + Controls + \varepsilon_i$, where q_i is the listed price for property i, p_j is the transaction price for a comparable property j sold in the previous month and *Treated* is a dummy that turns on when the listing price has been set in the period following the price data publication date. Controls for the time distance between the listing and the comparable transaction measured in days and its interaction with the log price are included in all columns but (3). Column (3) instead includes time distance (measured in weeks) dummies and their interaction with log price. Fixed-effects included are: real-estate agent dummies in all columns; listing month-year dummies in columns (2)-(4), and; transaction ID dummies in column (4). Standard errors double-clustered at the transaction and listing ID levels are reported in parentheses.

	(1)	(2)	(3)	(4)
Price × Treated	0.0037***	0.0042***	0.0039***	0.0023**
	(0.0013)	(0.0013)	(0.0013)	(0.0010)
Price	0.5581***	0.5455***	0.5418***	
	(0.0019)	(0.0019)	(0.0026)	
Treated	-0.0444***	-0.0524***	-0.0481***	-0.0240**
	(0.0154)	(0.0154)	(0.0153)	(0.0115)
Controls				
Price x Time distance	Yes	Yes	No	Yes
Price x Time distance dummies	No	No	Yes	No
Fixed-Effects				
Agency ID	Yes	Yes	Yes	Yes
Month-year	No	Yes	Yes	Yes
Transaction ID	No	No	No	Yes
Observations	7,443,824	7,443,824	7,443,824	7,443,824
R^2	0.7899	0.7925	0.7925	0.8970
Within R^2	0.3690	0.3540	0.3540	0.0000

Two-way (Transaction ID & Listing ID) standard-errors in parentheses

*Signif Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 5 Direct Effect of Transaction Prices on Quotes - Within Listing Price Updates
The table presents the results of the following regression: $log(q_i) = \alpha + \beta_0 \times log(p_j) + \beta_1 \times log(p_j) \times Treated_i + \gamma Treated_i + Controls + \varepsilon_i$, where q_i is the listed price for property i, p_j is the transaction price for a comparable property j sold in the previous month and Treated is a dummy that turns on when the listing price has been set/updated in the period following the most recent price publication date. Only listings that have at least one treated and one untreated quote are included. Controls for the time distance between the listing and the comparable transaction measured in days and its interaction with the log price are included in all columns but (5). Column (5) instead includes time distance (measured in weeks) dummies and their interaction with log price. Listing ID fixed effects are included in specifications (3)-(7). Additional fixed-effects include: listing month-year dummies in all columns but (2); transaction ID dummies in columns (2), (6) and (7), and; order of quote update dummies in column (7). Standard errors double-clustered at the transaction and listing ID levels are reported in parentheses.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Price × Treated	0.0028**	0.0037***	0.0013***	0.0011***	0.0011***	0.0033***	0.0021***
	(0.0014)	(0.0012)	(0.0003)	(0.0002)	(0.0002)	(0.0003)	(0.0002)
Price	0.8353***		-0.0024***	0.0003	0.0003		
	(0.0026)		(0.0003)	(0.0003)	(0.0004)		
Treated	-0.0362**	-0.0466***	-0.0173***	-0.0104***	-0.0106***	-0.0400***	-0.0221***
	(0.0165)	(0.0139)	(0.0034)	(0.0030)	(0.0030)	(0.0037)	(0.0030)
Controls							
Price × Time distance	Yes	Yes	Yes	Yes	No	Yes	Yes
Price × Time distance dummies	No	No	No	No	Yes	No	No
Fixed-Effects							
Listing ID	No	No	Yes	Yes	Yes	Yes	Yes
Month-year	Yes	Yes	No	Yes	Yes	Yes	Yes
Transaction ID	No	Yes	No	No	No	Yes	Yes
Order of price update	No	No	No	No	No	No	Yes
Observations	3,817,123	3,817,123	3,817,123	3,817,123	3,817,123	3,817,123	3,817,123
R^2	0.7136	0.9203	0.9962	0.9971	0.9971	0.9989	0.9991
Within R^2	0.7061	0.0001	0.0006	0.0008	0.0009	0.0131	0.0011

Two-way (Transaction ID & Listing ID) standard-errors in parentheses

*Signif Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 6 Direct Effect of Transaction Prices on First Quotes - Before vs After March 2012

The table presents the results of the following regression: $log(q_i) = \alpha + \beta_0 \times log(p_j) + \beta_1 \times log(p_j) \times \textit{Post March 2012}_i + \beta_2 \times log(p_j) \times \textit{Treated}_i + \beta_3 \times log(p_j) \times \textit{Treated}_i \times \textit{Post March 2012}_i + \gamma_1 \textit{Post March 2012}_i + \gamma_2 \textit{Treated}_i + \gamma_3 \times \textit{Treated}_i \times \textit{Post March 2012}_i + \textit{Controls} + \varepsilon_i$, where q_i is the listed price for property i, p_j is the transaction price for a comparable property j sold in the previous month, Post March 2012 is a dummy that equals one for listings published starting from March 2012 and Treated is a dummy that turns on when the listing price has been set/updated in the period following the most recent price publication date. Only the initial quotes of listings are included. Controls for the time distance between the listing and the comparable transaction measured in days and its interaction with the log price are included in all columns but (3). Column (3) instead includes time distance (measured in weeks) dummies and their interaction with log price. Fixed-effects included are: listing month-year dummies in all columns but (1) and; transaction ID dummies in column (4). Standard errors double-clustered at the transaction and listing ID levels are reported in parentheses.

	(1)	(2)	(3)	(4)
Price × Treated × Post March 2012	0.0094***	0.0090***	0.0089***	0.0033
	(0.0027)	(0.0027)	(0.0027)	(0.0024)
Price × Treated	-0.0060**	-0.0056**	-0.0060**	0.0001
	(0.0025)	(0.0025)	(0.0025)	(0.0021)
Price × Post March 2012	0.0460***	0.0461***	0.0461***	
	(0.0023)	(0.0023)	(0.0023)	
Price	0.7917***	0.7917***	0.7901***	
	(0.0027)	(0.0027)	(0.0033)	
Treated	0.0667**	0.0625**	0.0687**	-0.0071
	(0.0302)	(0.0301)	(0.0301)	(0.0248)
Post March 2012	-0.5650***	-0.5862***	-0.5865***	
	(0.0270)	(0.0278)	(0.0278)	
Treated × Post March 2012	-0.1083***	-0.1033***	-0.1030***	-0.0314
	(0.0321)	(0.0320)	(0.0320)	(0.0282)
Controls				
Price × Time distance	Yes	Yes	No	Yes
Price × Time distance dummies	No	No	Yes	No
Fixed-Effects				
Month-year	No	Yes	Yes	Yes
Transaction ID	No	No	No	Yes
Observations	10,585,043	10,585,043	10,585,043	10,585,043
R^2	0.6773	0.6782	0.6782	0.8561
Within R^2	–	0.6756	0.6756	0.0000

Two-way (Transaction ID & Listing ID) standard-errors in parentheses

*Signif Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 7 Effect of Transaction Prices on First Quotes Around Placebo Publishing Dates

The table presents the results of the following regression: $log(q_i) = \alpha + \beta_0 \times log(p_j) + \beta_1 \times log(p_j) \times Treated_i + \gamma Treated_i + Controls + \varepsilon_i$, where q_i is the listed price for property i, p_j is the transaction price for a comparable property j sold in the previous month and Treated is a dummy that turns on when the listing price has been set/updated in the week before (first four columns) or one week after (last four columns) the closest price publication date. Only the initial quotes of listings are included. Controls for the time distance between the listing and the comparable transaction measured in days and its interaction with the log price are included in all columns but (3) and (7). Columns (3) and (7) instead include time distance (measured in weeks) dummies and their interaction with log price. Additional fixed-effects include: listing month-year dummies in all columns but (1) and (5) and; transaction ID dummies in columns (4) and (8). Standard errors double-clustered at the transaction and listing ID levels are reported in parentheses.

	7 days before				7 days after			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Price × Treated	0.0021	0.0019	0.0012	0.0012	0.0007	0.0004	-0.0000	-0.0013
	(0.0021)	(0.0021)	(0.0021)	(0.0018)	(0.0021)	(0.0021)	(0.0021)	(0.0018)
Price	0.8430***	0.8428***	0.8410***		0.8406***	0.8409***	0.8411***	
	(0.0031)	(0.0031)	(0.0035)		(0.0042)	(0.0042)	(0.0040)	
Treated	-0.0279	-0.0266	-0.0169	-0.0110	-0.0127	-0.0076	-0.0014	0.0139
	(0.0255)	(0.0254)	(0.0255)	(0.0212)	(0.0254)	(0.0253)	(0.0254)	(0.0214)
Controls								
Price × Time distance	Yes	Yes	No	Yes	Yes	Yes	No	Yes
Price × Time distance dummies	No	No	Yes	No	No	No	Yes	No
Fixed-Effects								
Month-year	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Transaction ID	No	No	No	Yes	No	No	No	Yes
Observations	2,831,845	2,831,845	2,831,845	2,831,845	2,900,284	2,900,284	2,900,284	2,900,284
R^2	0.7021	0.7029	0.7029	0.8785	0.7068	0.7077	0.7077	0.8817
Within R^2	–	0.7006	0.7005	0.0000	–	0.7055	0.7055	0.0000

Two-way (Transaction ID & Listing ID) standard-errors in parentheses

*Signif Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 8 Indirect Price Effects Through Intermediate Listings - Full Sample

The table presents the results of the following regression: $log(q_i) = \alpha + \beta_0 \times log(p_j) + \beta_1 \times log(p_j) \times$ *Comps in w-1* $+ \beta_2 \times log(p_j) \times$ *Comps in w+1* $+ \beta_3 \times log(p_j) \times$ *Comps in all weeks* $+ \gamma_1 \times$ *Comps in w-1* $+ \gamma_2 \times$ *Comps in w+1* $+ \gamma_3 \times$ *Comps in all weeks* $+$ *Controls* $+ \varepsilon_i$, where q_i is the listed price for property i, p_j is the transaction price for a comparable property j sold in the previous month, Comps in w-1 is a dummy that turns on if the listing-transaction pair has at least one other comparable match in the week before the price publication date but none in the week after, Comps in w+1 is a dummy for pairs that have at least one match in the week after but none in the week before, and Comps in all weeks is a dummy for listing-transaction pairs with at least one match in each week. The table also reports the p-values of linear hypothesis tests of the difference in the price coefficients. Controls for the time distance between the listing and the comparable transaction measured in days and its interaction with the log price are included in all columns but (1), while controls for the number of comparables in the current week and each of the two previous weeks are included in columns (3)-(4). Listing month-year fixed effects are included in column (4). Standard errors double-clustered at the transaction and listing ID levels are reported in parentheses.

	(1)	(2)	(3)	(4)
Price	0.8242***	0.8231***	0.8231***	0.8226***
	(0.0020)	(0.0044)	(0.0044)	(0.0043)
Price × Comps in week -1$^{(1)}$	0.0089***	0.0089***	0.0089***	0.0083***
	(0.0032)	(0.0032)	(0.0032)	(0.0032)
Price × Comps in week +1$^{(2)}$	0.0176***	0.0176***	0.0177***	0.0169***
	(0.0029)	(0.0029)	(0.0029)	(0.0029)
Price × Comps in all weeks$^{(3)}$	0.0239***	0.0240***	0.0238***	0.0232***
	(0.0025)	(0.0025)	(0.0025)	(0.0025)
Comps in week -1	-0.1140***	-0.1138***	-0.1151***	-0.1071***
	(0.0392)	(0.0392)	(0.0392)	(0.0391)
Comps in week +1	-0.2130***	-0.2130***	-0.2150***	-0.2067***
	(0.0356)	(0.0356)	(0.0356)	(0.0356)
Comps in all weeks	-0.2854***	-0.2858***	-0.2890***	-0.2837***
	(0.0303)	(0.0303)	(0.0303)	(0.0303)
(2)−(1)	0.0087***	0.0087***	0.0088***	0.0086***
p-value	(0.0082)	(0.0083)	(0.0084)	(0.0091)
(3)−(2)	0.0063**	0.0064**	0.0061**	0.0063**
p-value	(0.0154)	(0.0147)	(0.0179)	(0.0158)
(3)−(1)	0.0150***	0.0151***	0.0149***	0.0149***
p-value	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Controls				
Price x Time distance	No	Yes	Yes	Yes
Nb. of comps per week	No	No	Yes	Yes
Fixed-Effects				
Month-year	No	No	No	Yes
Observations	2,932,258	2,932,258	2,932,258	2,932,258
R^2	0.7067	0.7068	0.7068	0.7077
Within R^2	–	–	–	0.7050

Two-way (Transaction ID & Listing ID) standard-errors in parentheses
*Signif Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 9 Indirect Price Effects Through Intermediate Listings - Before March 2012

The table presents the results of the following regression: $log(q_i) = \alpha + \beta_0 \times log(p_j) + \beta_1 \times log(p_j) \times$ *Comps in w-1* $+ \beta_2 \times log(p_j) \times$ *Comps in w+1* $+ \beta_3 \times log(p_j) \times$ *Comps in all weeks* $+ \gamma_1 \times$ *Comps in w-1* $+ \gamma_2 \times$ *Comps in w+1* $+ \gamma_3 \times$ *Comps in all weeks* $+$ *Controls* $+ \varepsilon_i$, where q_i is the listed price for property i, p_j is the transaction price for a comparable property j sold in the previous month, Comps in w-1 is a dummy that turns on if the listing-transaction pair has at least one other comparable match in the week before the price publication date but none in the week after, Comps in w+1 is a dummy for pairs that have at least one match in the week after but none in the week before, and Comps in all weeks is a dummy for listing-transaction pairs with at least one match in each week. The regressions are estimated using data from the sample before March 2012. The table also reports the p-values of linear hypothesis tests of the difference in the price coefficients. Controls for the time distance between the listing and the comparable transaction measured in days and its interaction with the log price are included in all columns but (1), while controls for the number of comparables in the current week and each of the two previous weeks are included in columns (3)-(4). Listing month-year fixed effects are included in column (4). Standard errors double-clustered at the transaction and listing ID levels are reported in parentheses.

	(1)	(2)	(3)	(4)
Price	0.7656***	0.7510***	0.7504***	0.7548***
	(0.0037)	(0.0085)	(0.0085)	(0.0085)
Price $\times$ Comps in week -1$^{(1)}$	0.0257***	0.0257***	0.0258***	0.0229***
	(0.0064)	(0.0064)	(0.0064)	(0.0064)
Price $\times$ Comps in week +1$^{(2)}$	0.0303***	0.0303***	0.0306***	0.0267***
	(0.0056)	(0.0056)	(0.0056)	(0.0056)
Price $\times$ Comps in all weeks$^{(3)}$	0.0216***	0.0215***	0.0180***	0.0135***
	(0.0047)	(0.0047)	(0.0047)	(0.0047)
Comps in week -1	-0.3179***	-0.3178***	-0.3240***	-0.2814***
	(0.0762)	(0.0762)	(0.0762)	(0.0759)
Comps in week +1	-0.3700***	-0.3696***	-0.3813***	-0.3282***
	(0.0673)	(0.0673)	(0.0673)	(0.0671)
Comps in all weeks	-0.2494***	-0.2488***	-0.2345***	-0.1697***
	(0.0565)	(0.0565)	(0.0565)	(0.0565)
(2)−(1)	0.0046	0.0046	0.0048	0.0038
p-value	(0.4899)	(0.4922)	(0.4782)	(0.5784)
(3)−(2)	-0.0087*	-0.0088*	-0.0126**	-0.0132**
p-value	(0.0901)	(0.0893)	(0.0146)	(0.0102)
(3)−(1)	-0.0041	-0.0042	-0.0078	-0.0094
p-value	(0.4910)	(0.4859)	(0.1883)	(0.1107)
Controls				
Price x Time distance	No	Yes	Yes	Yes
Nb. of comps per week	No	No	Yes	Yes
Fixed-Effects				
Month-year	No	No	No	Yes
Observations	1,355,042	1,355,042	1,355,042	1,355,042
R^2	0.6014	0.6014	0.6024	0.6044
Within R^2	–	–	–	0.6018

Two-way (Transaction ID & Listing ID) standard-errors in parentheses

*Signif Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 10 Chain Effects of Prices on Future Listings per Order of Match

The table presents the results of the following regression: $log(q_i) = \alpha + \sum_{k=1}^{10} \beta_k^{pre} \times log(p_j) \times \textit{Comp Order k Pre} + \sum_{k=1}^{10} \gamma_k^{pre} \textit{Comp Order k Pre} + \sum_{k=1}^{10} \beta_k^{post} \times log(p_j) \times \textit{Comp Order k Post} + \sum_{k=1}^{10} \gamma_k^{post} \textit{Comp Order k Post} + \textit{Controls} + \varepsilon_i$, where q_i is the listed price for property i, p_j is the transaction price for a comparable property j sold in the previous month and Comp Order k Pre (Post) is a dummy that turns on when quote i is the k-th sequential match to transaction j in the period before (after) the price data publication date. The sample includes listings in the one-month period surrounding the publication date that have a comparable transaction which has at least one treated and one untreated match. Controls for the time distance between the listing and the comparable transaction measured in days and its interaction with the log price are included in all columns but (1). Column (3) also includes listing month-year fixed effects. Standard errors double-clustered at the transaction and listing ID levels are reported in parentheses.

	(1)	(2)	(3)
Price	0.8388***	0.8404***	0.8401***
	(0.0011)	(0.0020)	(0.0020)
Price × 2nd Untreated	0.0010	0.0011	0.0011
	(0.0016)	(0.0016)	(0.0016)
Price × 3rd Untreated	0.0043**	0.0045**	0.0044**
	(0.0020)	(0.0020)	(0.0020)
Price × 4th Untreated	0.0054**	0.0057**	0.0056**
	(0.0023)	(0.0023)	(0.0023)
Price × 5th Untreated	0.0060**	0.0063**	0.0062**
	(0.0027)	(0.0027)	(0.0027)
Price × 6th Untreated	0.0039	0.0042	0.0041
	(0.0033)	(0.0033)	(0.0033)
Price × 7th Untreated	0.0068*	0.0071*	0.0070*
	(0.0039)	(0.0039)	(0.0039)
Price × 8th Untreated	0.0021	0.0024	0.0025
	(0.0044)	(0.0044)	(0.0044)
Price × 9th Untreated	0.0036	0.0040	0.0041
	(0.0055)	(0.0055)	(0.0055)
Price × 10th or more Untreated	0.0000	0.0004	0.0003
	(0.0038)	(0.0038)	(0.0038)
Price × 1st Treated	0.0015	0.0023	0.0025
	(0.0014)	(0.0017)	(0.0017)
Price × 2nd Treated	0.0039**	0.0048**	0.0050***
	(0.0016)	(0.0019)	(0.0019)
Price × 3rd Treated	0.0054***	0.0064***	0.0065***

Continued on next page

Table 10 – *Continued from previous page*

	(1)	(2)	(3)
	(0.0019)	(0.0022)	(0.0022)
Price × 4th Treated	0.0106***	0.0116***	0.0116***
	(0.0022)	(0.0025)	(0.0025)
Price × 5th Treated	0.0082***	0.0093***	0.0092***
	(0.0026)	(0.0029)	(0.0029)
Price × 6th Treated	0.0098***	0.0109***	0.0107***
	(0.0031)	(0.0033)	(0.0033)
Price × 7th Treated	0.0117***	0.0129***	0.0126***
	(0.0037)	(0.0039)	(0.0039)
Price × 8th Treated	0.0177***	0.0188***	0.0186***
	(0.0044)	(0.0046)	(0.0046)
Price × 9th Treated	0.0099*	0.0110**	0.0108**
	(0.0051)	(0.0052)	(0.0052)
Price × 10th or more Treated	0.0101***	0.0113***	0.0112***
	(0.0036)	(0.0039)	(0.0039)
2nd Untreated	-0.0115	-0.0131	-0.0122
	(0.0198)	(0.0198)	(0.0198)
3rd Untreated	-0.0508**	-0.0535**	-0.0524**
	(0.0236)	(0.0237)	(0.0237)
4th Untreated	-0.0635**	-0.0669**	-0.0661**
	(0.0280)	(0.0282)	(0.0281)
5th Untreated	-0.0686**	-0.0726**	-0.0720**
	(0.0328)	(0.0330)	(0.0330)
6th Untreated	-0.0406	-0.0449	-0.0437
	(0.0395)	(0.0397)	(0.0396)
7th Untreated	-0.0749	-0.0797*	-0.0777*
	(0.0469)	(0.0471)	(0.0470)
8th Untreated	-0.0158	-0.0207	-0.0210
	(0.0529)	(0.0532)	(0.0531)
9th Untreated	-0.0269	-0.0319	-0.0328
	(0.0659)	(0.0661)	(0.0661)
10th or more Untreated	0.0194	0.0140	0.0174
	(0.0457)	(0.0462)	(0.0461)
1st Treated	-0.0165	-0.0288	-0.0305

Continued on next page

Table 10 – *Continued from previous page*

	(1)	(2)	(3)
	(0.0174)	(0.0200)	(0.0200)
2nd Treated	-0.0442**	-0.0582**	-0.0590***
	(0.0199)	(0.0229)	(0.0228)
3rd Treated	-0.0599**	-0.0751***	-0.0747***
	(0.0233)	(0.0264)	(0.0263)
4th Treated	-0.1213***	-0.1372***	-0.1360***
	(0.0269)	(0.0300)	(0.0299)
5th Treated	-0.0894***	-0.1059***	-0.1035***
	(0.0318)	(0.0347)	(0.0346)
6th Treated	-0.1101***	-0.1269***	-0.1229***
	(0.0377)	(0.0403)	(0.0403)
7th Treated	-0.1278***	-0.1451***	-0.1399***
	(0.0449)	(0.0472)	(0.0471)
8th Treated	-0.1997***	-0.2172***	-0.2118***
	(0.0529)	(0.0551)	(0.0551)
9th Treated	-0.1079*	-0.1256**	-0.1199*
	(0.0612)	(0.0632)	(0.0631)
10th or more Treated	-0.0996**	-0.1177**	-0.1125**
	(0.0438)	(0.0470)	(0.0469)
Controls			
Price x Time distance	No	Yes	Yes
Fixed-Effects			
Month-year	No	No	Yes
Observations	11,292,009	11,292,009	11,292,009
R^2	0.7081	0.7081	0.7089
Within R^2	–	–	0.7063

Two-way (Transaction ID & Listing ID) standard-errors in parentheses
*Signif Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 11 Chain Effects of Prices on Future Listings per Order of Match - Linear Effects

The table presents the results of the following regression: $log(q_i^T) = \alpha + \beta_0^T \times log(p_j) + \beta_1^T \times log(p_j) \times$ *Nb. prior comps*$^T + \gamma_1 \times$ *Nb. prior comps*$^T +$ *Controls*$^T + \varepsilon_i^T$, where q_i^T is the listed price for property i, p_j is the transaction price for a comparable property j sold in the previous month and Nb.prior comps is the number of previous listings that have been matched to the same transaction. The subscript T is an indicator for whether the regressions use the sample of treated listings, i.e., listings posted in the period after the price data publication date ($T = 1$), or the set of untreated ones ($T = 0$). Controls for the time distance between the listing and the comparable transaction measured in days and its interaction with the log price are included in all columns but (1) and (4) and listing month-year fixed effects are included in columns (3) and (6). Standard errors double-clustered at the transaction and listing ID levels are reported in parentheses.

	Untreated			Treated		
	(1)	(2)	(3)	(4)	(5)	(6)
Prior × Nb. prior comps	0.0001	0.0001	0.0001	0.0007***	0.0008***	0.0007***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Price	0.8409***	0.8414***	0.8408***	0.8420***	0.8435***	0.8436***
	(0.0011)	(0.0022)	(0.0022)	(0.0011)	(0.0030)	(0.0030)
Nb. prior comps	0.0005	0.0003	0.0008	-0.0072**	-0.0076**	-0.0067**
	(0.0034)	(0.0035)	(0.0034)	(0.0032)	(0.0033)	(0.0033)
Controls						
Price x Time distance	No	Yes	Yes	No	Yes	Yes
Fixed-Effects						
Month-year	No	No	Yes	No	No	Yes
Observations	5,577,243	5,577,243	5,577,243	5,714,766	5,714,766	5,714,766
R^2	0.7068	0.7068	0.7076	0.7094	0.7094	0.7103
Within R^2	–	–	0.7048	–	–	0.7076

Two-way (Transaction ID & Listing ID) standard-errors in parentheses

*Signif Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 12

Chain Effects of Prices on Future Listings per Order of Match - Before March 2012

The table presents the results of the following regression: $log(q_i) = \alpha + \sum_{k=1}^{10} \beta_k^{pre} \times log(p_j) \times \textit{Comp Order k Pre} + \sum_{k=1}^{10} \gamma_k^{pre} \textit{Comp Order k Pre} + \sum_{k=1}^{10} \beta_k^{post} \times log(p_j) \times \textit{Comp Order k Post} + \sum_{k=1}^{10} \gamma_k^{post} \textit{Comp Order k Post} + \textit{Controls} + \varepsilon_i$, where q_i is the listed price for property i, p_j is the transaction price for a comparable property j sold in the previous month and Comp Order k Pre (Post) is a dummy that turns on when quote i is the k-th sequential match to transaction j in the period before (after) the price data publication date. The regressions use only data from the period before March 2012. The sample includes listings in the one-month period surrounding the publication date that have a comparable transaction which has at least one treated and one untreated match. Controls for the time distance between the listing and the comparable transaction measured in days and its interaction with the log price are included in all columns but (1). Column (3) also includes listing month-year fixed effects. Standard errors double-clustered at the transaction and listing ID levels are reported in parentheses.

	(1)	(2)	(3)
Price	0.7875***	0.7777***	0.7786***
	(0.0023)	(0.0042)	(0.0042)
Price × 2nd Untreated	0.0057*	0.0052	0.0047
	(0.0034)	(0.0034)	(0.0034)
Price × 3rd Untreated	0.0094**	0.0086**	0.0077**
	(0.0038)	(0.0038)	(0.0038)
Price × 4th Untreated	0.0076*	0.0066	0.0054
	(0.0045)	(0.0045)	(0.0045)
Price × 5th Untreated	0.0102**	0.0090*	0.0076
	(0.0051)	(0.0051)	(0.0051)
Price × 6th Untreated	0.0073	0.0060	0.0045
	(0.0057)	(0.0057)	(0.0057)
Price × 7th Untreated	0.0047	0.0033	0.0018
	(0.0064)	(0.0064)	(0.0064)
Price × 8th Untreated	0.0021	0.0007	-0.0009
	(0.0075)	(0.0075)	(0.0075)
Price × 9th Untreated	-0.0011	-0.0027	-0.0044
	(0.0085)	(0.0086)	(0.0086)
Price × 10th or more Untreated	-0.0175***	-0.0193***	-0.0213***
	(0.0058)	(0.0059)	(0.0059)
Price × 1st Treated	0.0014	-0.0034	-0.0036
	(0.0030)	(0.0034)	(0.0034)
Price × 2nd Treated	0.0028	-0.0024	-0.0030
	(0.0033)	(0.0038)	(0.0038)

Continued on next page

Table 12 – *Continued from previous page*

	(1)	(2)	(3)
Price × 3rd Treated	-0.0002	-0.0058	-0.0068
	(0.0038)	(0.0043)	(0.0043)
Price × 4th Treated	0.0071*	0.0013	0.0001
	(0.0042)	(0.0047)	(0.0047)
Price × 5th Treated	0.0070	0.0011	-0.0001
	(0.0048)	(0.0053)	(0.0053)
Price × 6th Treated	-0.0031	-0.0092	-0.0105*
	(0.0054)	(0.0059)	(0.0059)
Price × 7th Treated	0.0070	0.0008	-0.0005
	(0.0061)	(0.0065)	(0.0065)
Price × 8th Treated	0.0036	-0.0027	-0.0040
	(0.0069)	(0.0073)	(0.0073)
Price × 9th Treated	-0.0051	-0.0114	-0.0127
	(0.0077)	(0.0081)	(0.0081)
Price × 10th Treated	-0.0183***	-0.0249***	-0.0263***
	(0.0056)	(0.0062)	(0.0062)
2nd Untreated	-0.0666*	-0.0606	-0.0543
	(0.0397)	(0.0398)	(0.0397)
3rd Untreated	-0.1110**	-0.1009**	-0.0895**
	(0.0453)	(0.0454)	(0.0454)
4th Untreated	-0.0867	-0.0739	-0.0584
	(0.0533)	(0.0536)	(0.0536)
5th Untreated	-0.1160*	-0.1011*	-0.0824
	(0.0600)	(0.0604)	(0.0602)
6th Untreated	-0.0808	-0.0646	-0.0452
	(0.0677)	(0.0681)	(0.0680)
7th Untreated	-0.0485	-0.0311	-0.0109
	(0.0760)	(0.0764)	(0.0763)
8th Untreated	-0.0128	0.0052	0.0255
	(0.0887)	(0.0892)	(0.0892)
9th Untreated	0.0310	0.0495	0.0725
	(0.1015)	(0.1020)	(0.1016)
10th or more Untreated	0.2482***	0.2696***	0.2952***
	(0.0692)	(0.0702)	(0.0701)

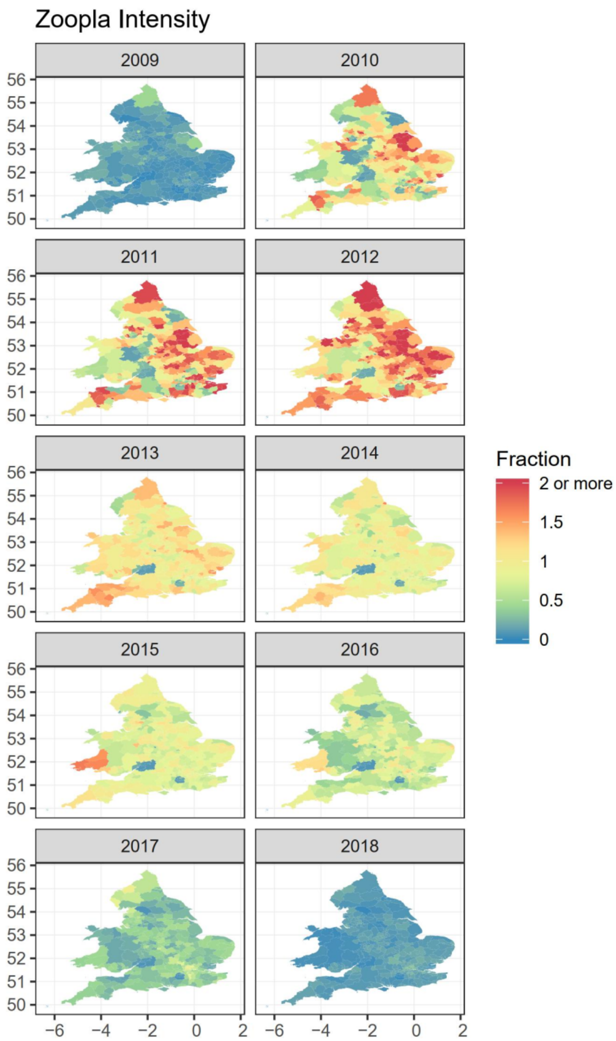
Continued on next page

Table 12 – *Continued from previous page*

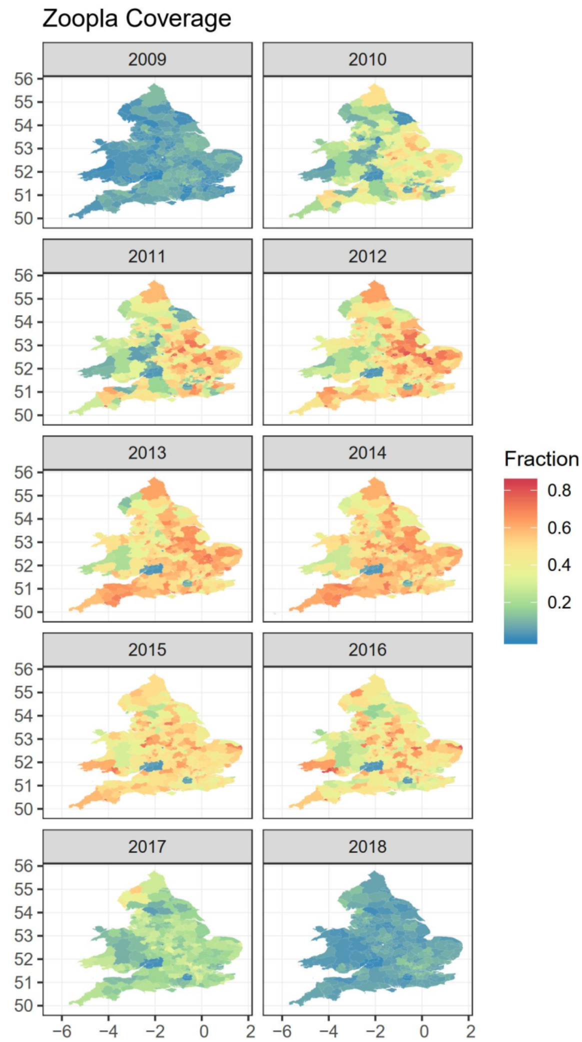
	(1)	(2)	(3)
1st Treated	-0.0201	0.0375	0.0404
	(0.0355)	(0.0404)	(0.0403)
2nd Treated	-0.0336	0.0302	0.0385
	(0.0395)	(0.0452)	(0.0451)
3rd Treated	0.0038	0.0718	0.0854*
	(0.0452)	(0.0512)	(0.0512)
4th Treated	-0.0815	-0.0108	0.0050
	(0.0501)	(0.0560)	(0.0559)
5th Treated	-0.0762	-0.0033	0.0131
	(0.0567)	(0.0630)	(0.0630)
6th Treated	0.0479	0.1223*	0.1405**
	(0.0644)	(0.0698)	(0.0698)
7th Treated	-0.0713	0.0041	0.0228
	(0.0722)	(0.0773)	(0.0773)
8th Treated	-0.0251	0.0513	0.0702
	(0.0824)	(0.0873)	(0.0871)
9th Treated	0.0810	0.1579	0.1764*
	(0.0918)	(0.0966)	(0.0967)
10th or more Treated	0.2571***	0.3363***	0.3567***
	(0.0674)	(0.0748)	(0.0746)
Controls			
Price x Time distance	No	Yes	Yes
Fixed-Effects			
Month-year	No	No	Yes
Observations	4,772,899	4,772,899	4,772,899
R^2	0.60688	0.60689	0.60811
Within R^2	–	–	0.60669

Two-way (Transaction ID & Listing ID) standard-errors in parentheses
*Signif Codes: ***: 0.01, **: 0.05, *: 0.1*

B.2 Plots



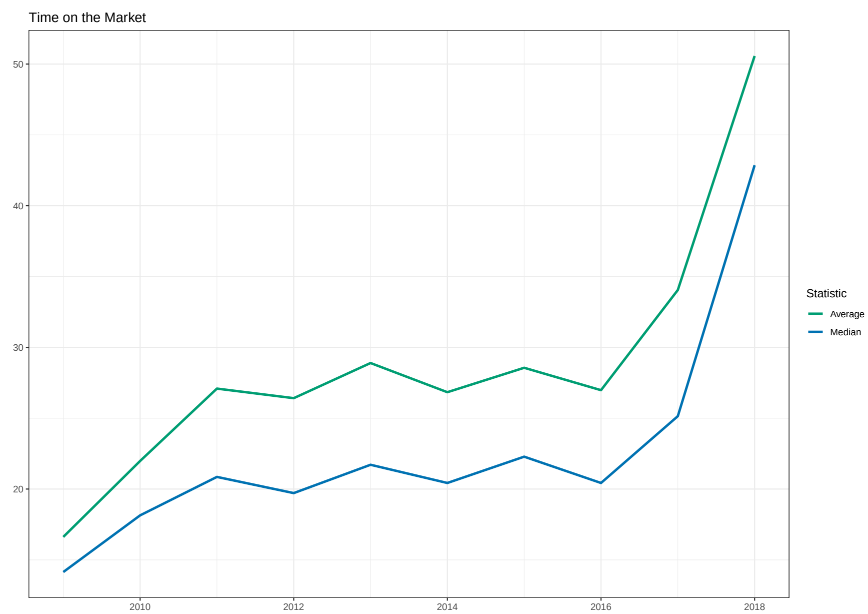
(a) Listings-to-Transactions Ratio



(b)
Fraction of Transactions Listed on Zoopla

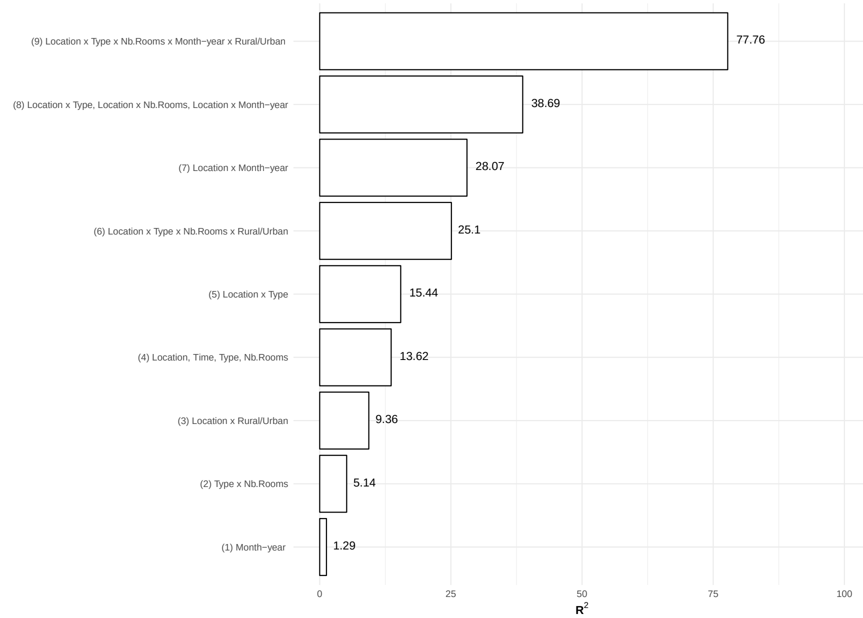
Figure 1 : Geographic Coverage
The figure plots heat maps of the relative geographic coverage of the Zoopla listing data between 2009-2018 by year across England and Wales. Figure 1a displays the total number of listings as a fraction of transactions shifted by six months (average TOM), while Figure 1b the total number of transactions that were matched to their respective listings in the Zoopla data.

(a) Price Discount

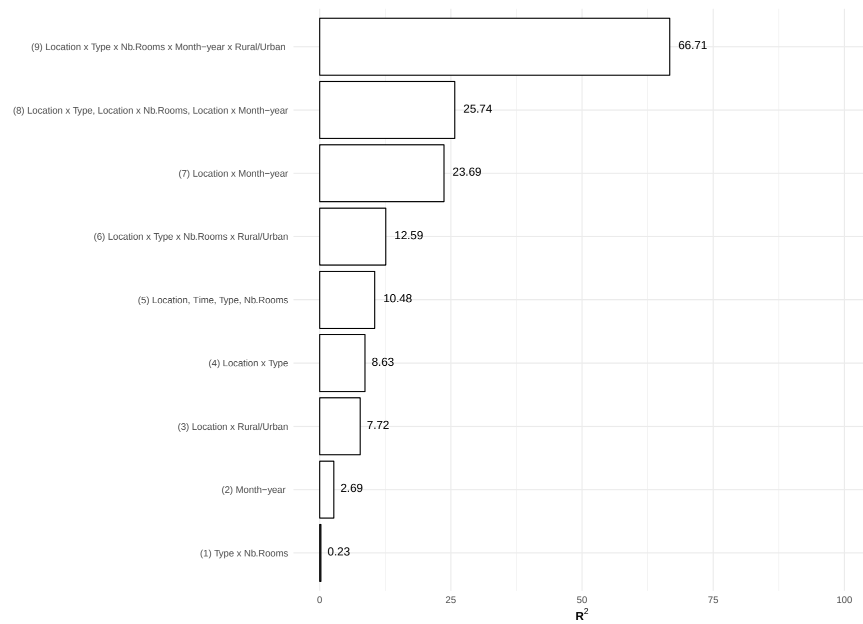


(b) Time on the Market

Figure 2 Time-series of Price Discount and Time on the Market
The figure displays the time-series of price discount and time on the market (TOM) for the set of property listings that were matched to their respective ex-post transactions in the sample from 2009 to 2018. Figure 2a plots the time-series of the percentage difference between the first listed price and the final transaction price, while Figure 2b shows the time-series of time on the market measured as the number of weeks since the property was first listed. The green lines show the time-series of the average values and the blue lines represent the median values.



(a) Price Difference in pounds (£)



(b) Percentage Price Difference

Figure 3 : Fraction of Variation in Price Discount Explained by Fixed Effects
The figure displays the percentage of the variation in difference between listing and transaction prices that is explained by observable characteristics, measured as the R-squared from a regression of price differences on various fixed effects, for the set of property listings that were matched to their respective ex-post transactions in the sample from 2009 to 2018. Figure 3a shows the variation explained in the absolute price difference and Figure 3b in the percentage price discount. Fixed effects included are: month-year of the listing or transaction; property type (detached, semi-detached, terraced house or a flat); number of rooms in the property, where properties with between 6 and 10 rooms are placed in one bucket and properties with more than 10 rooms in another; location, measured as the address outcode, and; a rural/urban area indicator from the 2011 Census classification of Output Areas.

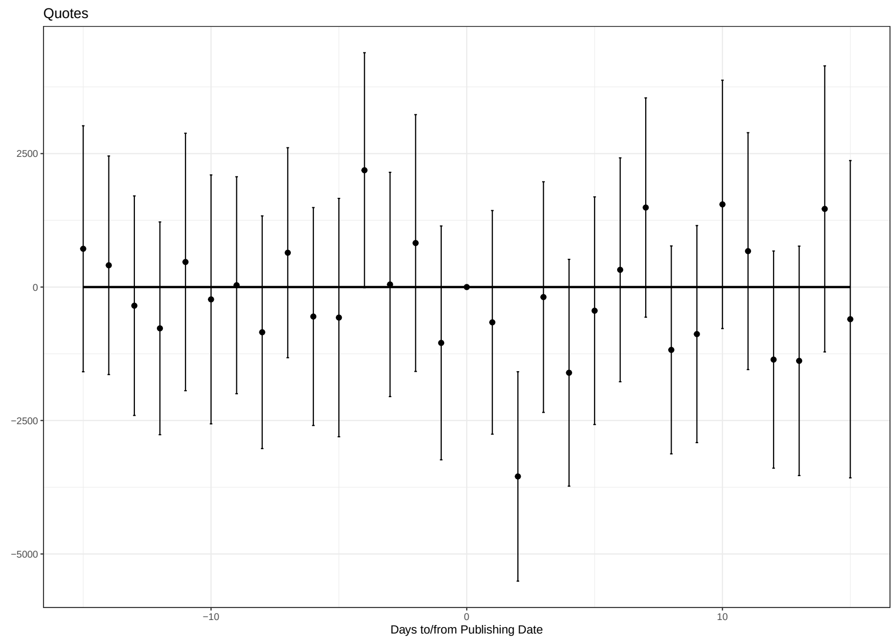
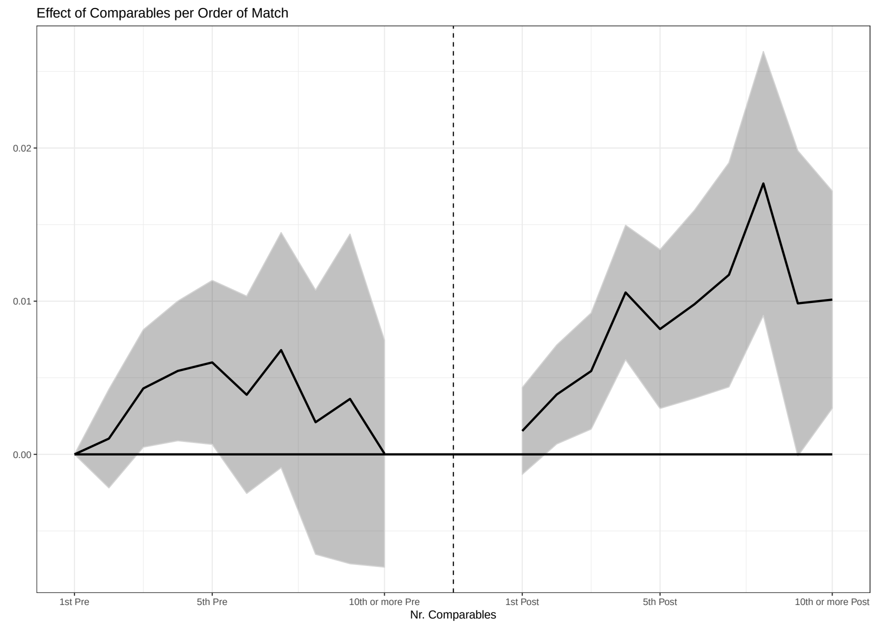
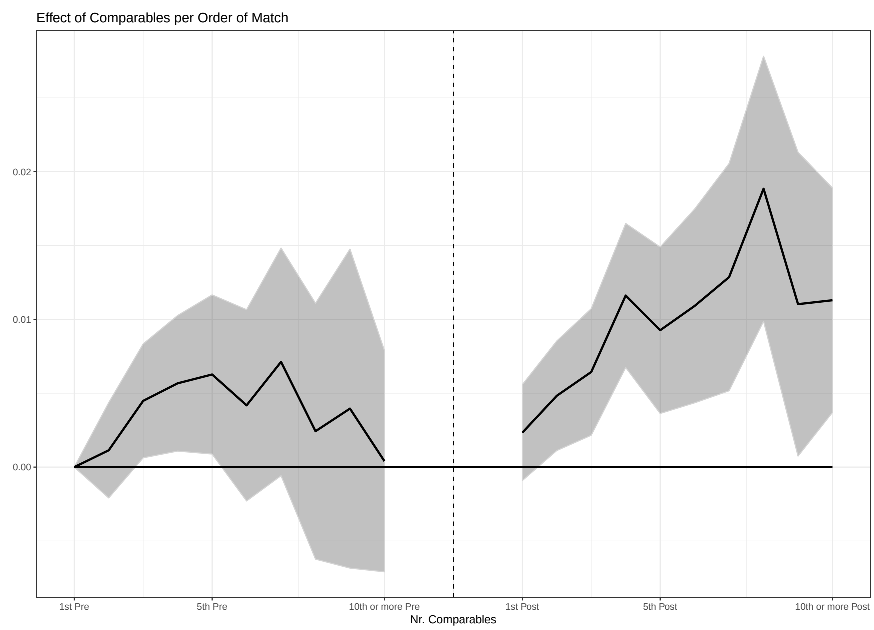


Figure 4 : Variation In Listing Prices Around Publishing Dates - Before March 2012
The figure plots the results from a regression of listing prices on dummies for the signed number of days between the listing date and the price data publication date for the sample before March 2012. The regression is specified as follows: $q_i = \alpha + \sum_{\Delta=-15}^{15} \gamma_\Delta + FE + \varepsilon_i$, where the fixed-effects correspond to the characteristics the matching is based on, i.e., location, property type, number of rooms and month-year, and Δ is a dummy for the signed difference in days between the date on which a listing is posted and the publication date. The baseline coefficient is the one for listings posted exactly on the publication date. The vertical lines represent the 95% confidence bounds of the point estimates for the average listing prices.



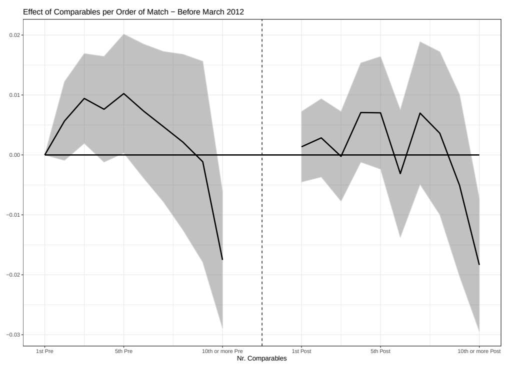
(a) Baseline regression



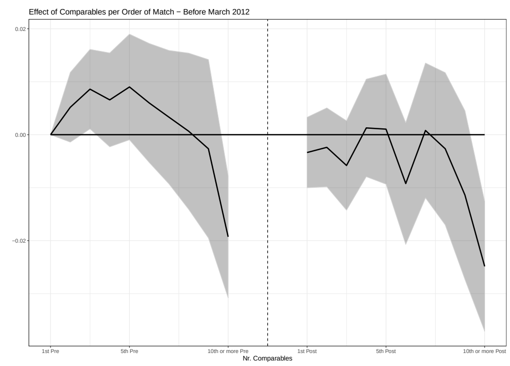
(b) Time distance controls

Figure 5 : Coefficients on Comparable Transaction Prices by Order of Match

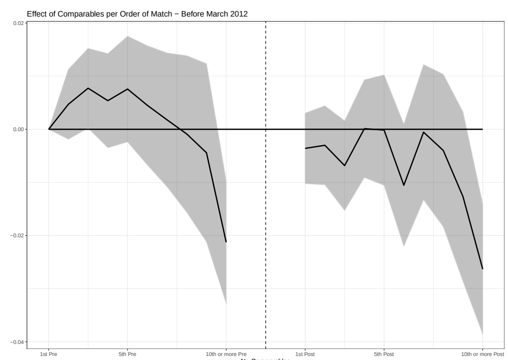
The figure plots the price coefficients from the following regression along with their 95% confidence bounds: $log(q_i) = \alpha + \sum_{k=1}^{10} \beta_k^{pre} \times log(p_j) \times \textit{Comp Order k Pre} + \sum_{k=1}^{10} \gamma_k^{pre} \textit{Comp Order k Pre} + \sum_{k=1}^{10} \beta_k^{post} \times log(p_j) \times \textit{Comp Order k Post} + \sum_{k=1}^{10} \gamma_k^{post} \textit{Comp Order k Post} + \textit{Controls} + \varepsilon_i$, where q_i is the listed price for property i, p_j is the transaction price for a comparable property j sold in the previous month and Comp Order k Pre (Post) is a dummy that turns on when quote i is the k-th sequential match to transaction j in the period before (after) the price data publication date. The sample includes listings in the one-month period surrounding the publication date that have a comparable transaction which has at least one treated and one untreated match. Figure 5a is the baseline regression with no controls and Figure 5b includes controls for the time distance between the listing and the comparable transaction measured in days and its interaction with the log price.



(a) Baseline regression



(b) Time distance controls



(c) Time distance controls and month-year fixed effects

Figure 6 :

Coefficients on Comparable Transaction Prices by Order of Match - Before March 2012

The figure plots the price coefficients from the following regression along with their 95% confidence bounds: $log(q_i) = \alpha + \sum_{k=1}^{10} \beta_k^{pre} \times log(p_j) \times \textit{Comp Order k Pre} + \sum_{k=1}^{10} \gamma_k^{pre} \textit{Comp Order k Pre} + \sum_{k=1}^{10} \beta_k^{post} \times log(p_j) \times \textit{Comp Order k Post} + \sum_{k=1}^{10} \gamma_k^{post} \textit{Comp Order k Post} + \textit{Controls} + \varepsilon_i$, where q_i is the listed price for property i, p_j is the transaction price for a comparable property j sold in the previous month and Comp Order k Pre (Post) is a dummy that turns on when quote i is the k-th sequential match to transaction j in the period before (after) the price data publication date. The sample includes only data before March 2012 of listings in the one-month period surrounding the publication date that have a comparable transaction which has at least one treated and one untreated match. Figure 6a is the baseline regression with no controls, Figure 6b includes controls for the time distance between the listing and the comparable transaction measured in days and its interaction with the log price and Figure 6c includes month-year fixed effects in addition to time distance controls.

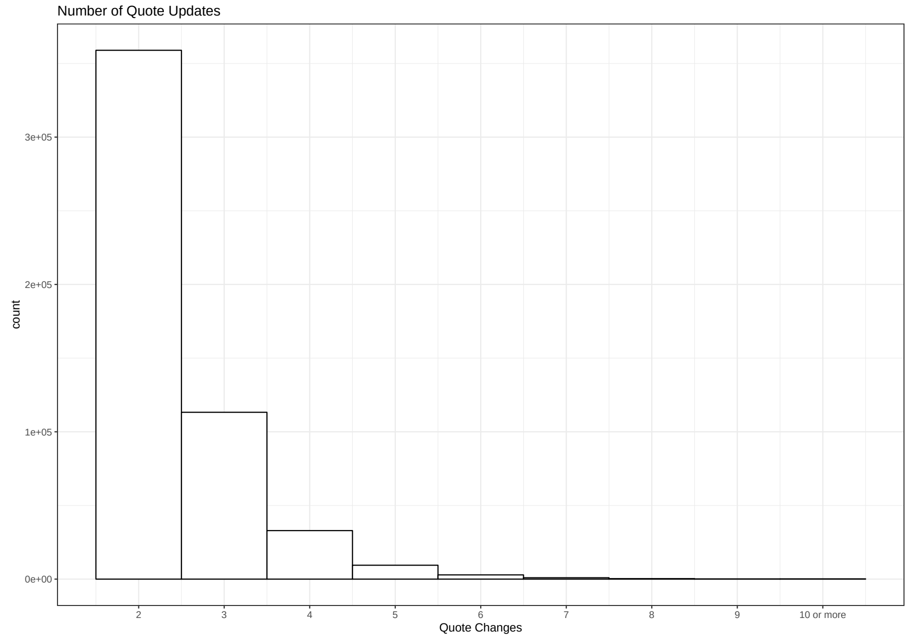


Figure 7 Number of Quote Updates per Listing
The figure plots the histogram of the total number of price changes per listing. The sample includes listings posted after March 2012 that have at least one comparable match.

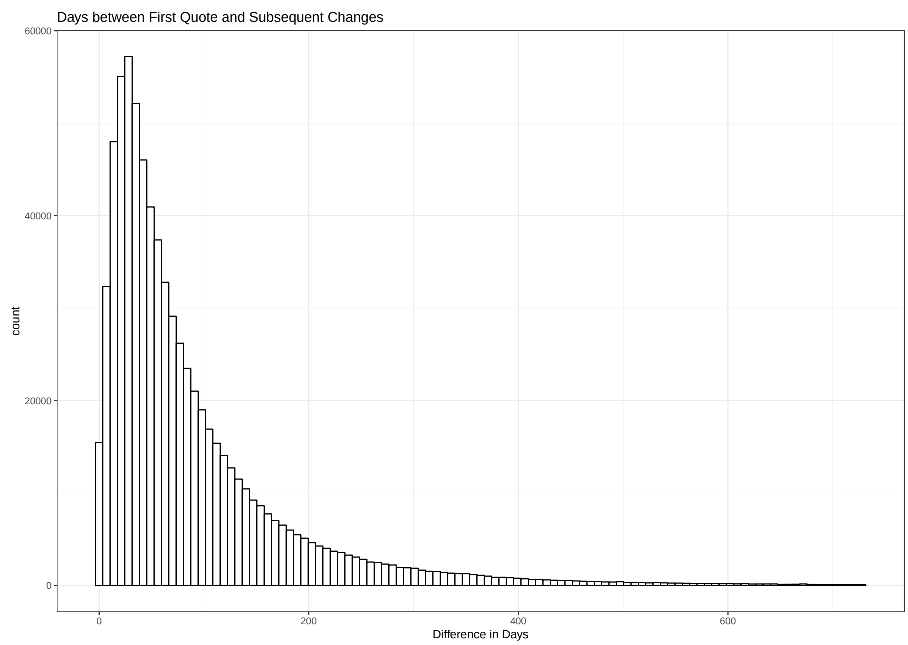
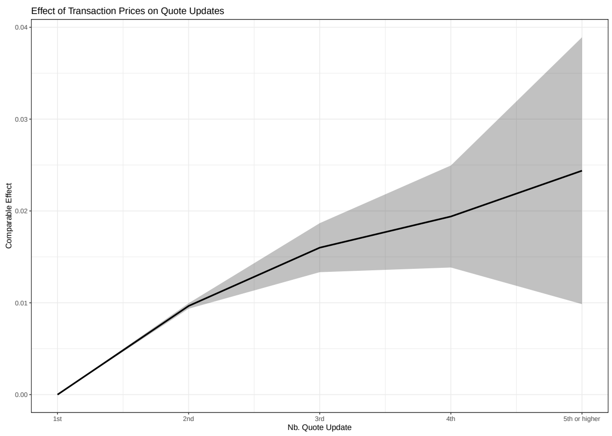
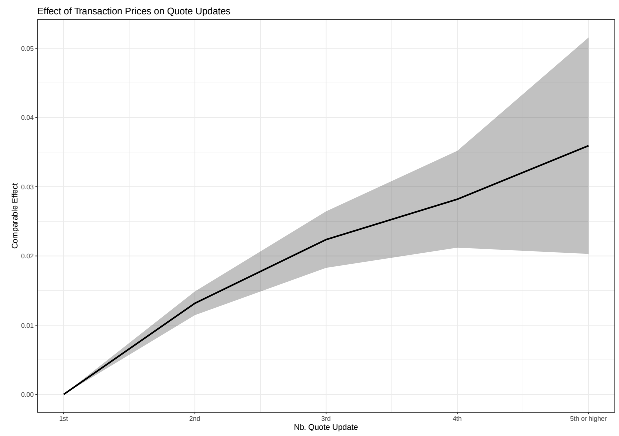


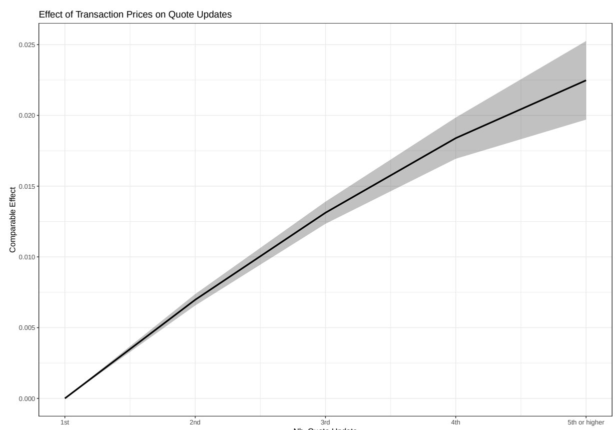
Figure 8 Distance in Days between Initial Listing Date and Subsequent Quote Changes
The figure plots a histogram of the difference in days between quote changes and the initial date of the listing. The sample includes listings posted after March 2012 that have at least one comparable match.



(a) Baseline Regression



(b) Time distance controls



(c) Time distance
controls, Month-year and Listing ID Fixed Effects

Figure 9 : Effect of Comparable Transaction Prices on Listing Price Updates - Coefficients
The figure displays the price coefficients from the following regression along with their 95% confidence bounds: $log(q_i^n) = \alpha + \beta_1 \times log(p_j) + \sum_{n=2}^{5} \beta_n \times log(p_j) \times \textit{Update Number } n + \sum_{n=2}^{5} \gamma_n \textit{Update Number } n + \textit{Controls} + \varepsilon_i^n$, where q_i^n is the n-th listed price update for property i, p_j is the transaction price for a comparable property j sold in the month before property i was initially listed and Update Number n is a dummy that turns on when quote i is the n-th change to the listing price for property i. The sample includes listings in the post March 2012 period that have at least one price change and a comparable transaction that has been published just before the listing has been first posted. Figure 9a is the baseline regression with no controls, Figure 9b includes controls for the time distance between the listing and the comparable transaction measured in days and its interaction with log price and Figure 9c includes month-year and listing ID fixed effects in addition to time distance controls.

References

Han, Lu and William C. Strange, "The Microstructure of Housing Markets," in Gilles Duranton, J. V. Henderson, and William C. Strange, eds., *Handbook of Regional and Urban Economics*, Vol. 5 of *Handbook of Regional and Urban Economics*, Elsevier, 2015, chapter 0, pp. 813–886.