5-2-2022

Online Property Valuation, Price Discovery, and Market Efficiency in the Housing Market

Jeonghyun Chung
Georgia State University

Online Property Valuation, Price Discovery, and Market Efficiency in the Housing Market

BY

Jeonghyun Lee Chung

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree

Of

Doctor of Philosophy

In the Robinson College of Business

Of

Georgia State University

GEORGIA STATE UNIVERSITY
ROBINSON COLLEGE OF BUSINESS
2022

ACCEPTANCE

This dissertation was prepared under the direction of the *Jeonghyun L. Chung* Dissertation Committee. It has been approved and accepted by all members of that committee, and it has been accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Business Administration in the J. Mack Robinson College of Business of Georgia State University.

Richard Phillips, Dean

DISSERTATION COMMITTEE

Dr. Vincent Yao (Co-Chair)
Dr. Jonathan A. Wiley (Co-Chair)
Dr. Thao Le
Dr. Kristopher Gerardi (External – The Federal Reserve Bank of Atlanta)

ACKNOWLEDGEMENT

I would like to thank my committee members for their consistent guidance, advice, and support. I learned and developed so much - not only professionally, but also as a person. Thank you, Dr. Vincent Yao, for all of your care and support. To Dr. Jon Wiley, thank you for admitting me into the Ph.D. program and for your continued trust and encouragement. To Dr. Thao Le, thank you for all the ad hoc conversations, guidance, and friendship - one of the greatest gifts I received throughout my Ph.D. journey. Lastly, Dr. Kris Gerardi thank you for inspirational stories and advice. Only with the help from my committee members could I have completed this journey.

I dedicate this dissertation to my parents, Younbai Lee and Myungja Park, my husband, Robert Chung, and my lovely daughter, Anna Chung, for their love and encouragement. I am also thankful to all the faculty, staff, and colleagues at Georgia State University who supported me throughout this adventure. Without them, I could never finish my doctoral degree.

Lastly, to those in the PhD program, “I made it and you will make it soon”.

ABSTRACT

Online Property Valuation, Price Discovery, and Market Efficiency in the Housing Market

BY

Jeonghyun Lee Chung

March 28th, 2022

Committee Chair: *Dr. Vincent Yao and Dr. Jon Wiley*

Major Academic Unit: *Real Estate Department*

This dissertation studies the impact of publicly available property valuation on price discovery and sales outcomes in the housing markets using Zestimate, a popular property value estimate provided by the largest PropTech firm in the U.S. I find that sellers list their houses at inflated prices over the predicted values when their Zestimates are biased upwards. This initial overpricing leads to a subsequent 2 percent reduction in listing prices and delays sale by 3 days on average. However, such properties are eventually sold at a higher price than similar houses with undervalued Zestimates. On the other hand, sellers of properties with overvalued Zestimates are more likely to withdraw their homes from the market due to overpricing. My findings suggest that while easy public access to property value estimates should in theory improve market efficiency, inaccurate valuation can in fact have the opposite effect as market participants rely heavily on online property values in their decision-making.

Online Property Valuation, Price Discovery, and Market Efficiency in the Housing Market*

Jamie L. Chung†

Abstract

This paper studies the impact of publicly available property valuation on price discovery and sales outcomes in the housing markets using Zestimate, a popular property value estimate provided by the largest PropTech firm in the U.S. I find that sellers list their houses at inflated prices over the predicted values when their Zestimates are biased upwards. This initial overpricing leads to a subsequent 2 percent reduction in listing prices and delays sale by 3 days on average. However, such properties are eventually sold at a higher price than similar houses with undervalued Zestimates. On the other hand, sellers of properties with overvalued Zestimates are more likely to withdraw their homes from the market due to overpricing. My findings suggest that while easy public access to property value estimates should in theory improve market efficiency, inaccurate valuation can in fact have the opposite effect as market participants rely heavily on online property values in their decision-making.

Keywords: Asset Valuation, Housing Market, Mispricing, PropTech

JEL Classification: R00, O33, L85, L86

*I appreciate Georgia Multiple Listing Services(GMLS) and Zillow Group for providing real estate transaction data. I also thank Thao Le, Jon Wiley, Vincent Yao, the doctoral students at the department of real estate and finance, Georgia State University, and finance and economics faculty members at University of Nebraska - Omaha for helpful comments and suggestions.

†J. Mack Robinson College of Business, Georgia State University; Jeonghyun.chung@gmail.com.

1. Introduction

Unlike the stock market where market participants have easy, public access to audited reports on asset and company performance to evaluate a security's value, the real estate market is relatively opaque as firm generated and external analyst reports are not widely available. Publicly listed, algorithm derived online property value estimates therefore provide a potentially market moving intervention to the real estate market as these estimated values can be used as a previously unavailable, professionally derived reference price by inexperienced or unsophisticated market participants. This paper investigates the effects that this newly available pricing information has on the real estate market.

PropTech firms market their proprietary property value estimates as accurate to coerce the public that their platforms are useful and continued engagement is valuable in allowing buyers and sellers budget and plan for the anticipated transaction of a property. Among the digital real estate platforms providing an algorithm derived property value estimate, Zillow.com receives the most attention from market participants (**Figure 1**). Zillow's platform is a high probability first landing zone for buyers and sellers to go to when evaluating and researching the real estate market. As my research interests focus on how households utilize online property valuations, I use Zillow's property valuation metric called Zestimate as this number is free, easy to access and for many, there is no obvious superior alternative. By scraping property level Zestimates from Zillow's website, I investigate if Zestimates affect sellers' listing prices and if so, what impact it has on sales outcomes. This paper also explores whether Zestimates usage differs by market participant type and Zestimate volatility.

For those unfamiliar with a given property market and without access to financial models to determine a property's value, Zestimates potentially enable a more level playing field for all market participants. Previous literature on the topic of the impact analysts' forecasts and revisions have on security values indicates that such references have a material impact (Stickel, 1992; Park and Stice, 2000). Logically, as Zestimates can be perceived as an authoritative reference, a Zestimate could similarly affect property values. The effect of this online reference price relies on the accuracy of valuation as reference price can affect listing and selling price significantly. As Zillow provides a property's Zestimate immediately under a seller's asking prices, sellers can easily reference a property's Zestimate to inform at what price to list. If a seller sets a high listing price, the seller should

expect to wait longer for an acceptable offer than if the asset is priced lower. In a scenario where the publicly available reference price is undervalued, sellers may set listing prices lower than they had not seen a Zestimate, but may see properties sell faster than average. As many other market participants perceive the downward pricing behavior, other sellers are consequently more likely to lower their own asking prices, thus decreasing values across a market and vice versa. Both scenarios can therefore drive market inefficiencies. However, if publicly available property valuations are close to properties' predicted values, market participants may expect to buy and sell at a fair value with greater confidence, thus potentially increasing efficiency in the real estate market. Using property level Zestimate information and property transaction data, I find that when a Zestimate is too high, higher than a property's predicted value, sellers are more likely to set higher asking prices. This result implies that Zestimates anchor expectations, so prices correspondingly, move upward when Zestimates are high. Inaccurate estimates, therefore, can effectively create a self-fulfilling prophecy.

To measure how much mispricing affects sales outcomes due to flawed Zestimates, I decompose pricing misbehavior drivers into two categories. The first segment focuses on Zestimate's influence and the second segment focuses on the influence of other factors that cannot be captured in a Zestimate such as a seller's risk appetite, financing, or staging status has on sales outcomes. The preliminary results illustrate that pricing bias explained by Zestimate delays sales by 2.4 percent, which is 3 more days in addition to the delays caused by other pricing bias factors. The data indicates that sellers who overprice their listing due to heavy influence from the property's Zestimate will eventually apply a discount to the original listing price. However, these aggressively priced listings eventually close at higher prices, thus supporting the anchoring effect on listing prices (Northcraft and Neale, 1987; Genesove and Mayer, 2001; Bucchianeri and Minson, 2013). The magnitude of the impact is greater when the property price is biased due to other factors as a Zestimate can be publicly referenced.

This paper also provides a possible explanation for why some sellers withdraw properties. I find that Zestimates for withdrawn properties are estimated at 3.8 percent higher than that for similar sold properties. It indicates that these withdrawn properties likely had inflated Zestimates prior to listing. This hypothesis is further supported by evidence that sellers who withdraw their properties set asking prices higher than similar properties sold during the same time frame. Consequently, high

Zestimates could be a significant input into driving overpricing of assets. The findings of this paper imply that incorrect online property value estimates can make the housing market inefficient as they can contribute to increased transaction costs.

Do Zestimates affect market participants' listing price decisions differently depending on their level of access to information? To answer this question, I use institutional investors as a proxy for informed market participants as institutional investors often have a higher level of experience and access to more comprehensive, detailed resources and intelligence. Unsurprisingly, institutional sellers are less likely to rely on Zestimates. Institutional sellers' listing prices are 4.2 percent less related to properties' listed Zestimates while individual sellers typically price their homes more relying on the corresponding Zestimates. This result implies that unbiased Zestimates can possibly mitigate information asymmetries by providing a reference price to less informed sellers. However, preliminary results indicate that online property valuations do not fully overcome information asymmetries that exist between individual and institutional sellers in the housing market. Given the reach of platforms such as Zillow and the increasing reliance on their algorithm derived valuations on market actors, it is possible that if an algorithm driven model is incorrectly tuned, the scale of impact on housing market inefficiencies could be more significant than pundit's imagine.

This paper builds on recent work focusing on the accuracy of online property value estimates (e.g., Corcoran and Liu, 2014; Muralidharan, et al., 2018; Lee and Sasaki, 2018; Lu, 2019; Yu, 2020). By comparing selling price to Zestimate at sales, they conclude that Zestimate is unbiased appraised value. I contribute to the literature by evaluating the effects Zestimates may create at listings and sales outcomes. Especially, Zestimates at the time of a pre-listing are used in this study to accurately measure the effects Zestimate have on property listing prices. The timing of Zestimate is critical for my study as Zillow adjusts Zestimates immediately following a property's listing. This paper is also related to previous literature that investigate how mispricing affects real estate sales outcomes (e.g., Millaer and Sklarz, 1987; Yavas and Yang, 1995; Anglin et al., 2003; Haurin et al., 2010; Bucchianeri and Minson, 2013; Haurin et al., 2013; Anderson et al., 2014). My findings are consistent with previous literature that conclude a high-priced property takes longer to sell (Trippi, 1977; Anglin et al., 2003; Haurin et al., 2010) and is sold at a higher price than under-priced properties due to anchoring effects (Haurin et al., 2010; Bucchianeri and Minson, 2013). Previous literature attempts

to explain mispricing in the housing market by evaluating market participant types, the level at which they are informed, or proximity to a property. I contribute to this literature by providing evidence that incorrect online property value estimates can also lead to systematic overpricing of assets in housing markets.

The organization of this paper is as follows. Section 2 describes the conceptual background by reviewing past literature focused on online asset valuation in the housing market and mispricing in the real estate market. Section 3 addresses the data, methodology, and model used. Section 4 discusses the main empirical analysis and findings. The final section provides a succinct summary and touches upon the policy implications associated with the main findings.

2. Literature Review

As PropTech firms provide property value estimates using algorithm, the early studies focus on the accuracy of those valuations. Many scholars study the accuracy of Zillow's Zestimate by comparing sales prices to Zestimates at the time of sale. Using properties' Zestimates at the time of sale, past studies find that Zestimates can predict final transaction prices (Corcoran and Liu, 2014; Muralidharan, et al., 2018; Lee and Sasaki, 2018). Due to the difficulty associated with data acquisition, some papers use only a relatively small number of observations for their analysis. For example, Lee and Sasaki (2018) use 1,200 data points (averaging 4 properties from each US major county) and find that online property value estimates are strong predictors of final transaction prices. Lu (2019) and Yu (2020) use a larger number of samples to investigate if Zestimates exhibit racial biases that may exist in the market. They conclude that Zestimates' parameters do not include racial inputs, thus mitigating racial bias in Zillow's assessments. As my interests lay in whether Zestimates impact what price sellers' ultimately list at, using a properties' Zestimates at the time of closing is not appropriate as Zillow refreshes Zestimates as soon as properties are listed for sale and again when they are sold[1]. By scraping properties' entire Zestimate pricing change history, this paper uses the precise Zestimate information that sellers could actually see prior to listing their property for sale. Specifically, I use Zestimates one month prior to listing to investigate whether Zestimates affect sellers' listing prices

[1]Zestimate is estimated using observable variables available online, transaction events, and comparable properties' recent price information. https://zillow.zendesk.com/hc/en-us/articles/203512140-What-is-a-Zestimate-home-value-.

and sales outcomes.

The topic of mispricing in the housing market receives considerable attention from scholars (Millaer and Sklarz, 1987; Yavas and Yang, 1995; Anglin et al., 2003; Haurin et al, 2010; Bucchianeri and Minson, 2013; Haurin et al., 2013). The price at which to list a property is always challenging as the goals of selling a property quickly and at the highest possible price are seemingly inherently contradictory. Although the relationship between the degree of overpricing and days on market is positively related, the relationship between degree of overpricing and selling price is still unsettled. From a traditional finance perspective, an overvalued property should be expected to sell at a higher price as the initial listing price is used to anchoring price expectations upwards (Northcraft and Neale, 1987; Bucchianeri and Minson, 2013). Auction focused literature indicates that higher prices can also be derived by setting a price lower than an asset's intrinsic value (Bokhari and Geltner, 2011). This paper contributes to the existing literature by providing an additional factor for why mispricing occurs in today's housing markets. Algorithm driven online property valuation, which ideally provides true value of a property, can provide misinformation to market participants and derive mispricing in the housing market. Additionally, my finding supports the hypothesis that the anchoring effects are material. A property with an overvalued Zestimate does in fact sell at a premium relative to similar properties with lower Zestimates.

3. Data and Methodology

3.1. Zestimate Data

Zestimate is an algorithm driven property's market value estimated by Zillow.com. Zestimates are shown under the listing price for properties on market or under the recent sales price if it is off market, typically at the visual focus point of a website or application page (**Figure 2**). Given that Zestimates are shown immediately underneath listing prices (next to the recent sales price) for properties on market (off market), it is difficult for potential sellers and buyers to miss Zestimate information when searching property online. The historical Zestimate trend is also available on the same page helping market participants to understand the overall housing market and property specific price trend.

As Zillow is referenced by most home buyers and sellers, and since 2015 holds the greatest market share of any digital real estate platform in the US, I decide to use Zestimates for single family homes transacted in the Great Atlanta area from 2015 to 2019. First, I extract a list of addresses that meet the criteria of property type, sales period, and location from Georgia Multiple Listing Services (GMLS). I exclude new construction (properties less than 1 year old) from the sample as these properties typically do not have Zestimates. The initial sample includes 122,239 single family homes transacted from 2015 to 2019 in Atlanta-Sandy Springs-Roswell, GA MSA. Using the addresses of these transactions, I populate property characteristics (number of bed, number of bath, building size, built year, etc), transaction history (listing date, listing price, new listing prices, change dates, sales price, sales date, etc), along with current and historical Zestimates from the Zillow website. **Figure 2** illustrates an example of historical monthly Zestimate values that I scraped in this paper. 205 properties do not have Zesitmate information and these are eliminated from the final sample. I exclude additional 28,891 observations due to missing property characteristics, listing, and sales information. Key property characteristics such as number of bedrooms, bathrooms, building size and property age at sales, listing price, sales price, and Zestimate are winsorized at the 1 percent and 99 percent levels.

After populating the data, I merge transaction information with historical Zestimate information to identify Zestimate values from time frames near the initial listing. **Figure 3** illustrates Zestimate trends around listing month, which justifies the decision to use pre-listing Zestimates in the analysis. Zestimates are steady before a property is listed, but values typically increase by 3 percent in the listing month[2]. This jump can be explained by the impact changing input parameters have on Zillow's pricing algorithm. Zillow's Zestimate incorporates property characteristics, transaction data such as listing prices, days on the market, and sales prices, transaction data of comparable properties, public records such as tax assessments and description, and market trends[3]. It implies that Zestimate is modeled to adjust its value as soon as a property is listed on market irrespective of how realistic the ask is. It suggests that Zestimates after listing month are not appropriate for my analysis to use as post-listing Zesitmates already account for seller's price preference. Furthermore, as potential sellers are more likely to utilize the pre-listing Zestimate number, I adopt Zestimates 1 month prior

[2] The average monthly home value change in the Atlanta, MSA for the same period of time is 0.57 percent with values ranging from 0.17 percent to 0.98 percent according to the Zillow Home Value Index (ZHVI).

[3] https://zillow.zendesk.com/hc/en-us/articles/203512140-What-is-a-Zestimate-home-value-.

to listing for the analysis[4].

Another benefit of the scraped data is that I can observe the entire listing price change history at property level[5]. The scraped data includes not only initial listing prices and dates, but also adjusted listing prices and price adjustment dates until sold or withdrawn. Although days on market is a solid measure of market efficiency, knowing when and how sellers change their listing prices provides more information to further our understanding of the housing market. Using historical changes to listing prices, I can test if sellers who price initial listings too far upwards due to high Zestimates are more likely to change listing prices and if so, in which direction changes are made.

3.2. Transaction Data

I primarily use the scraped data addressed above, but I also leverage residential transaction data such as GMLS and ZTRAX[6]. GMLS data provides in-depth information on real estate transactions such as property characteristics, listing price, listing date, sales price, and sales date. I use GMLS data to initially identify the sample and validate the quality of scraped data. ZTRAX is used to identify market participants' types. In addition to the details that GMLS data provides, ZTRAX provides information on whether a seller is an institutional investor[7]. I use institutional investors as proxies for informed market participants and exclude observations lacking values for seller types. Using property address and sales year and month, 85 percent of the sample is merged with ZTRAX data and 67 percent of sellers are successfully identified. In the proceeding section, I describe how I utilize this data.

3.3. Empirical Strategy

The main research question this paper attempts to answer is, "do online asset valuations affect sellers' pricing behavior?" To answer this question, I measure if overvalued Zestimates lead to overpricing. I use the difference between the log of properties' Zestimate and the log of predicted value as a proxy

[4]The results are robust when I use Zestimate_{-2m}, Zestimate_{-3m}, or the average $\text{Zestimate}_{[-3m,-1m]}$.

[5]Figure A1 displays an example of listing price change shown on Zillow website. GMLS provides original listing price, listing price at search, and listing price at contract. How many times and when to change the prices are missing.

[6]ZTRAX is residential property transaction data provided by Zillow.com

[7]This paper focuses on Zestimate at the pre-listing period. Thus, buyer's reaction to Zestimate is not aligned with this paper. The results on Zestimate effects at sales by buyer's types are available in the appendix.

for Zestimate's bias (Z-Bias) level. The difference between the log of properties' original listing price and the log of properties' predicted value is used as a proxy for seller's pricing bias level. To measure the Z-Bias and pricing bias, first I estimate predicted values of properties using a hedonic model with all transactions in the preceding 12 months. This method is commonly used when measuring the pricing bias in the real estate literature (e.g., Millaer and Sklarz, 1987; Yavas and Yang, 1995; Anglin et al., 2003; Haurin et al, 2010; Bucchianeri and Minson, 2013; Haurin et al., 2013; Ooi and Le, 2013) and in the finance literature (e.g., Basu and Markov, 2004)[8] . The predicted property value is estimated by property characteristics, sales time, and location information following the specification below:

$$ln(SalesPrice)_{i,t,z} = \alpha + \beta_i X_i + \gamma_{zyq} + \mu_{ym} + \epsilon_{i,t,z} \tag{1}$$

where $ln(SalesPrice)_{i,t,z}$ is the natural log of sales price for property i located in zip code z and transacted at time t. X_i is a matrix of the explanatory variables such as number of bedrooms, number of bathrooms, the log of building size in square feet, age, and age-squared. β_i is the vector of parameters, and $\epsilon_{i,t,z}$ is an error term. Local time-varying market trends are controlled for with the inclusion of zip code by year-quarter (γ_{zyq}) and year-month (μ_{ym}) fixed effects. Standard errors are clustered at zip code by year level. To estimate the predicted value of property i transacted at time t and in zip code z, I use 12 months of transaction data at the time of sales but lag the period by one month[9]. For example, properties sold from February 2018 to January 2019 are used to estimate the predicted value of properties sold in March 2019. The predicted value incorporates a lag of one month as Zestimate one month prior to listing is used in the analysis.

Predicted values are extrapolated market prices for properties. Using properties' predicted values, I test the hypothesis that overpriced listings are driven by overvalued Zestimates. To determine Zestimates' impact on pricing, I estimate the following specification:

$$PricingBias_{i,t,z} = \alpha + \beta_1 Z \cdot Bias_{i,t-1,z} + \mu_{ym} + \epsilon_{i,t,z} \tag{2}$$

[8]Finance literature measure analyst's forecast error as actual earnings minus forecast.

[9]I also estimate the predicted values one month prior to listing to match the timing of Zestimate used in this paper. The results are robust and available in the appendix.

where pricing bias of property i located in zip code z and listed at time t is measured by the the difference between the log of the actual listing price and the log of the expected listing price, the predicted value predicted in equation (1). Zestimate Bias (Z-Bias) is estimated by the difference between the log of Zestimates one month prior to listing and the log of properties' predicted value. The coefficient of Z-Bias, β_1, is the coefficient of interest that captures Z-Bias effects on mispricing. If β_1 is positive and significant, this behavior implies that a property is overpriced because a Zestimate is overestimated. If β_1 is insignificant, this indicates that market participants are likely not heavily reliant upon online pricing information. As Zestimates theoretically already account for observable information such as property characteristics and location, I only include year-month (μ_{ym}) fixed effect in the model[10].

After confirming that Zestimates affect sellers' listing decisions, I divide the pricing bias into two components. The first pricing bias factor is mispricing driven by misevaluated Zestimates. The second component is mispricing due to other factors not observed by Zestimate. The general mispricing examples are typically due to unobservable factors such as a seller's overconfidence and property-specific information not available to the public. Other possible explanations could include when sellers must receive a price higher than prior market performance due to scenarios such as where an asset holder is subject to outstanding pricing covenants or have debt obligations higher than a property's predicted value. These along with other reasons can influence at what price sellers list, and are not parameters that can be captured for incorporation into Zillow's Zestimate. Fitted values estimated from equation (2) capture the first segment while residuals indicate the second segment. The specifications of each component are described below.

$$\widehat{PricingBias}_{i,t,z} = \widehat{\alpha} + \beta_1 \widehat{Z \cdot Bias}_{i,t-1,z} \tag{3}$$

$$\widehat{\epsilon}_{i,t,z} = \widehat{PricingBias}_{i,t,z} - \widehat{Z \cdot Bias}_{i,t-1,z} \tag{4}$$

$\widehat{PricingBias}_{i,t,z}$ indicates the degree of mispricing where sellers set asking prices higher (lower) due to overestimated (underestimated) Zestimate, which is defined as $\widehat{PricingBias}_{Zestimate}$ in the further analysis. $\widehat{\epsilon}_{i,t,z}$ is the mispricing that cannot be explained by Zestimate, which is presented as

[10] I add zip code by year-quarter and year-month fixed effects, and cluster standard errors at the zip code by year level for robustness tests. The results are shown in **Table 2**.

$\widehat{PricingBias}_{Residuals}$ in the analysis. Using those two components, I estimate if mispricing driven by Zestimates affect sales outcomes via evaluating metrics such as days on market, sales price to listing price ratio, and final sales price. The specifications for each tests are:

$$ln(Time\ on\ Market)_{i,t,z} = \alpha + \beta_1 \widehat{PricingBias}_{i,t,z} + \beta_2 \widehat{\epsilon}_{i,t,z} + \lambda_{zyq} + \mu_{ym} + \varepsilon_{i,t,z} \quad (5)$$

$$ln(SalesPrice/predicted)_{i,t,z} = \alpha + \beta_1 \widehat{PricingBias}_{i,t,z} + \beta_2 \widehat{\epsilon}_{i,t,z} + \lambda_{zyq} + \mu_{ym} + \varepsilon_{i,t,z} \quad (6)$$

$$ln(SalesPrice/ListingPrice)_{i,t,z} = \alpha + \beta_1 \widehat{PricingBias}_{i,t,z} + \beta_2 \widehat{\epsilon}_{i,t,z} + \lambda_{zyq} + \mu_{ym} + \varepsilon_{i,t,z} \quad (7)$$

where i is property, t is sales time in year-month, and z is zip code. The log of days on market, the difference between the log of sales price and the log of properties' predicted value, and the difference between the log of properties' sales price and the log of properties' original listing price are used as dependent variables respectively. β_1 is the coefficient of interest which captures the mispricing caused by Z-Bias effects on sales outcomes while β_2 explains the impacts other mispricing factors not associated with Zestimates have on sales outcomes. Zip code by year-month fixed effect (λ_{zyq}) and year-month fixed effect (μ_{ym}) are included. Standard errors are clustered at zip code by year level.

3.4. Summary Statistics

Table 1 provides summary statistics for the final sample. The final sample includes 93,143 single family homes that on average, have 4 bedrooms and 3 bathrooms. The average listing price is \$301,741 which is greater than the average sales price of \$282,695. The average Zestimate one month prior to listing is \$283,786 which is 6 percent lower than listing price. Initial asking prices are 14 percent different from properties' predicted values. Zestimates are 7 percent off from the predicted value on average, which indicates that Zestimates are \$19,174 higher than predicted values on average. **Figure 4** shows that Z-Bias is well distributed. 45 percent of Zestimates are lower than predicted values while 55 percent of Zestimates are higher than predicted values. The median Z-bias is 3.6 percent.

I divide the pricing bias into two segments. $\widehat{PricingBias}_{Zestimate}$ is mispricing explained by Zestimates while $\widehat{PricingBias}_{Residuals}$ indicate mispricing due to other factors. Both values are

standardized to more fairly compare the impact these factors have on sales outcomes. Whether or not sellers change listing prices, and if so, to what extent and at what point are pricing adjustments made are two questions that must be evaluated to understand Zestimate's affect on property listings. The data describes that typically, properties stay on market for 93 days until sale and 30 percent of properties modify their listing price at least once prior to selling[11]. Those properties adjust asking prices 4 percent downwards from their initial listing price and sellers wait, on average, 58 days before modifying the listing price. **Table A1** summarizes the definitions of the variables used in the analysis.

4. Empirical Findings

4.1. Baseline Results

Do Zestimates affect the price at which sellers list at? To answer this question, I run the regression described in equation (2). **Table 2** Column (1) shares the Z-Bias effects on pricing bias when I only include year-month fixed effect. The result indicates that when Z-Bias increases by 1 percent, sellers are likely to list their property at a price 0.88 percent higher than its predicted value. It implies that misevaluated Zestimates can create mispricing in housing markets. Columns (2)-(4) illustrate the results for scenarios when Zestimates do not fully reflect observable information such as property characteristics and location. To control for those possibilities, I include different specifications. Column (2) includes zip code by year-month and year-month fixed effects to control local time-varying factors. Standard errors are clustered at zip code by year level. I additionally include property characteristics in Column (3) and Column (4). While Column (3) includes year-month and zip code fixed effects separately, Column (4) includes zip code by year-month and year-month fixed effects. Although the coefficients are smaller than the most relaxing scenario in Column (1), the results are still consistent that sellers are more likely to list their homes at a higher price when Zestimate is overestimated[12]. **Figure 5** displays the coefficient plots for the pricing bias associated

[11] I confirm that online listing price information changes are reasonable as 35 percent of properties in the MLS dataset experienced price modifications prior to sale, a figure similar to the 30 percent of properties in my data set.

[12] Additional test to control the timing of listing is conducted. Sellers who list their homes at the beginning of the month are more likely to use the previous month's Zestimate. Using the subsample of the homes that are listed on the first week of each month, I confirm that those sellers rely on Zestimates statistically and economically.

with Z-Bias. I divide the Z-Bias into 10 groups and test if a greater Z-Bias results in upwards mispricing at listing. The graph shows that Z-Bias and overpricing has an almost linear relationship.

This paper's primary assumption is that potential sellers check properties' Zestimates prior to listing. As I cannot observe if a seller referenced a property's Zestimate as input into determining a listing price, I use Google Trends' search outcomes of '*Zillow*' as a proxy for local household's dependence level of Zestimates[13]. Specifically, I measure if Z-Bias effects are greater in areas where households' search of the term Zillow is higher. The results in **Table 3** indicate that listing prices from sellers who reside in areas where households use Zillow more are more heavily influenced by Zestimate. A property whose Zestimate is inflated relative to its predicted value and located in an area where Zillow is in high use typically lists at a price 2.7 percent higher relative to a property whose Zestimate is less than its predicted value and in an area with less usage of Zillow's platform. Columns (2)-(6) run the same regression using the yearly subsample. The overall affect of Zestimate on listing prices becomes greater over time while the location effects correspondingly decrease over time. These results imply that Zestimates are used more frequently and have greater weight, making geographic usage variance trivial. Additionally, I divide Google search level into 10 groups to evaluate if the volume at which search queries are conducted impacts the extent of mispricing. I find that Z-Bias has an almost linear relationship with pricing bias even after controlling for households' interest of Zillow use. The figure is available in the appendix (Figure A2).

4.2. Sales Outcomes

To answer the questions, "do properties with overvalued Zestimates take longer to sell and how do overvalued Zestimates impact sales prices?", I divide mispricing into two categories. Using two components of the mispricing estimated in equation (3) and (4), I test if mispricing caused by Zestimates affect sales outcomes such as days on market, sales price, and asking price. **Table 4** represents the results from equation (5)-(7). Similar to outstanding literature, I find that increases in the degree of overpricing lead to commensurate increases in the number of days on market. One standard deviation increase of overpricing from other factors delays sales by 14 percent, leading to an additional 13 days on the market. When Zestimate related mispricing increases by one standard

[13] Google Trends provides quantitative interest level of relevant keywords during targeted time periods in targeted locations.

deviation, time on market increases by 2.4 percent (3 days) in addition to the delays associated with other mispricing. Column (2)-(3) showcase results from the subsample analysis. In Column (2), I use the sample where both mispricing explained by Zestimate and other mispricing factors are low and in Column (3), I use the sample where both values are high. When both mispricing levels are low, other mispricing factors do not affect days on market while Zestimate actually aids sellers in selling quicker when a Zestimate is undervalued. In the opposite scenario of Column (3), properties stay on market 6 days longer when Zestimate is overvalued than predicted value.

Columns (4)-(6) provide evidence that the listing price in the housing market is used as an anchoring price, making selling prices higher than the predicted value when a property is overpriced. One standard deviation increase of pricing bias due to an overestimated Zestimate results in a sales price 2.9 percent higher than a property's predicted value. The effects are similar between low and high groups. The anchoring effects in the listing strategy can be also found when mispricing is due to other factors. However, sellers receive larger premiums when overpricing is driven by a high Zestimate than when overpricing is due to other factors. This behavior may be explained by buyers perceiving a high Zestimate as an authoritative signal of a property's worth. Although sellers receive a higher offer, they undergo listing price adjustments to sell (Column (7)-(9)). When sellers set a higher asking price, sellers discount their asking price by 4 percent when a listing price is overpriced due to reasons unobserved by Zestimates and by 1.5 percent when a property is overpriced due to a mispriced Zestimate. Zestimate plays a role only when both mispricing factors are overestimated. When both mispricing factors are high, one standard deviation increase of pricing bias from Zestimate makes final sales prices $4,104 lower than listing prices on average. Similar to real estate agents who are willing to wait longer to sell at a higher price (Rutherford, et al., 2005; Levitt and Syverson, 2008), sellers who overprice their properties when relying upon a high Zestimate also wait longer to sell, but eventually do receive a higher sales price.

Figure 6 illustrates the results graphically. I divide the overpricing level into 10 groups and plot the coefficients accordingly. Panel A shows the effects of two overpricing components on days on market. It is interesting to note that when Z-Bias is overestimated or underestimated, properties take longer to sell compared relative to when Zestimates are close to their intrinsic value. This U shaped behavior illustrates the impact Zillow can have when its Zestimates are publicly available to both

buyers and sellers. If a property's Zestimate value is significantly different than its list price, even if the Zestimate is unrealistically underpriced, market participants are more hesitant to purchase. However, the number of days until sales is linearly longer if a property is overpriced due to other reasons. Panel B indicates that sellers receive a higher sales price relative to the predicted value when they list their homes at a higher than predicted value. However, when a property price is underpriced due to an undervalued Zestimate, the sales price is typically lower than the predicted value. This result implies that buyers may also account for Zestimate information in their offers. Lastly, Panel C also provides evidence that listing prices are adjusted downward when a property is overpriced. Overall, the graphs illustrate that sellers who rely on overvalued Zestimate information typically sell slower and at a price lower than original asking price. However, the properties eventually sell at premium relative to properties with undervalued Zestimates.

4.3. Housing Market Efficiency

4.3.1. Listing Price Change

Sellers lower their asking prices to increase the probability of selling (Knight, 2002; De Wit, et al., 2013). Using properties' comprehensive listing price history, published online, I test whether overpriced properties are more likely to modify asking prices and if so, in which direction changes are made. **Table 5** Column (1) shows the results of the listing price change test when I use the entire sample. It indicates that overpriced properties are more likely to undergo pricing modification. The probability of a price change is 3.3 percent higher for properties with overestimated Zestimates relative to properties with fairly or underpriced Zestimates. In Columns (2)-(5), I use the subsample of properties that adjust listing prices at least once. Those properties decrease listing prices by $3,000 on average when Zestimate related mispricing increases by one standard deviation. The adjustment amount is similar between two mispricing segments. However, sellers who overprice homes due to Zestimate unrelated reasons are willing to wait longer to change prices downwards when compared to those who overprice homes due to Zestimate. One standard deviation increase of Zestimate related mispricing increases the number of days until change by 12 percent, indicating that sellers are willing to wait 7 additional days to make a price adjustment. Furthermore, overpriced sellers are less likely to raise their initial listing prices, but more likely to lower their asking prices. These findings are

consistent with the baseline results that mispricing delays a sale with price adjustments. Mispricing cases resulting from Zestimates typically adjust listing prices downward at greater intensity than other mispricing cases. As a Zestimate is mutually observable by sellers and buyers, sellers' discounted asking prices (perhaps moving closer to a Zestimate) may be easier and more obvious to perceive than when a property is mispriced for other reasons.

4.3.2. Sold vs. Withdrawn Properties

Evaluating why sellers pull a property from the market is an important area of study as a withdrawal, likely with little to no information provided as to why the withdrawal occurred, in the short term introduces ambiguity, likely to be perceived negatively, into the market. By comparing listing prices of withdrawn properties to properties that sold, this paper provides a potential answer to seller's withdrawal behavior. I identify withdrawn properties prior to sale from GMLS and scraped additional Zestimate information for these properties. I use the propensity score matching method to match the sample of withdrawn properties to sold properties controlling for property characteristics such as number of beds, building size, property age, and location. Zestimates one month prior to a listing event are also included in the matching process to provide a scenario where both sellers receive the same Zestimate signal immediately prior to listing. Panel A of **Table 6** shows the t-test results on control variables that similar properties are selected for both groups.

Panel B Columns (1)-(3) show the results when using the entire sample while Columns (4)-(6) display the results when using the matched sample. Column (1) provides an evidence that properties with high-valued Zestimates are more likely to withdraw. Specifically, 1 percent increase of Z-Bias results in 6 percent higher chance to withdraw. Column (2) and (3) indicate that sellers who withdraw before sale are more likely to misprice their homes upwards due to overestimated Zestimates relative to properties that successfully sold. Withdrawn properties initially listed at $905 higher than those that sold when Zestimate is estimated by 1 percent higher than its predicted value. In other words, sellers who withdraw their properties list their homes at $6,335 higher on average than those who successfully sold similar homes. The probability of withdrawal due to overestimated Zestimate is much higher when I use the matched samples. Sellers with inflated Zestimates prior to listing are more likely to overvalue their asset and consequently are more likely list at a price too

high, thus experiencing a 20 percent higher withdraw rate relative to similar properties in the same neighborhood with Zestimates closer to their predicted value. Other results are consistent even with the restrictive sample. These findings are consistent with the previous literature from Anglin, et al. (2003), who also find that withdrawn properties initially list at prices higher than sold properties. As this paper provides evidence that high Zestimates can cause sellers to misprice properties upwards, withdrawals would be expected to be more common for properties with Zestimates higher than the property's predicted value. **Figure 3** visualizes this results that Zestimates for withdrawn properties spike to a greater extent at listing then properties that sold. Zestimate spikes gradually disappear for withdrawn properties while Zestimates for sold properties are more stable. This behavior implies that Zestimates for withdrawn properties might suffer from greater upwards valuation errors than properties that sold.

4.3.3. Individuals vs. Institutional Seller

The ideal algorithm driven online asset valuation use case in the financial market is one that reduces information asymmetry among participants while improving operational efficiency. If the Zestimate is a truly fair property value, its cost-free, wide availability may help individual sellers close outstanding transaction performance gaps between them and institutional investors. However, we may see a difference in the extent of Zestimate reliance between individual and institutional sellers if institutional sellers hold information that Zestimate does not incorporate. If institutional sellers do not rely on Zestimates (or rely on to an extent less than individual sellers), this behavior may indicate that institutions rely more on proprietary information and processes such as institutional actors in other traditional security markets (Cehn and Jiang, 2006).

As institutional investors trade certain property types, I match the sample using a propensity score matching method. In addition to physical characteristics, the location and timing of sales are included in the matching process. Matched properties are located in the same zip code and are transacted within the 12 months period. Zestimate one month prior to listing month is also included in the model. The underlying strategy is that both individual sellers and institutional sellers view a Zestimate and perceive the valuation signal of the Zesitmate similarly. Panel A of **Table 7** exhibits matching outcomes. Homes sold by both individual and institutional sellers have similar

property characteristics. Pre-listing Zestimates for homes sold by individual sellers are $231,116 while Zestimates one month prior to listing for homes sold by institutional sellers are $235,748. As Zestimate does not consider seller's type in algorithm yet, similar Zestimates prior to listing imply that both individual and institutional sellers receive the similar pricing signal from Zestimates.

If Zestimates are accurate and used by sellers for pricing decision, individual sellers can take advantage of Zestimates by listing their homes at predicted values, which is possibly close to market value estimated by institutional sellers. Panel B of Table 7 shows how institutional sellers use Zestimate differently from individual sellers. Panel B Column (1) and (2) illustrates that institutional sellers are less likely to rely on publicly available information compared to individual sellers when they list residential properties for sale. While individual sellers list their homes at a price 70 percent related to Zestimates, institutional sellers list their homes at a price 55 percent related to Zestimate. It indicates that when a Zestimate is 1 percent overestimated than the predicted value, individual sellers list a home at 0.7 percent overpriced than the predicted value while institutional sellers list a property at 0.55 percent overpriced than the predicted value. Column (3) displays that Z-Bias effects on mispricing are 4.2 percent smaller when a property is listed by institutional sellers compared to when individual sellers list.

Next, I answer the question, "do individual sellers benefit from the availability of Zesitmates?". Panel C shows the mispricing effects on sales outcomes. My findings are contrary to the commonly held belief that informed investor performance is superior to the individual. When institutional sellers misprice upwards due to Zestimate reliance, their properties take 0.6 percent longer to sell compared to individuals' listed properties. However, the result is not statistically significant. This behavior could be explained by the first graph in **Figure 6** that even when the extent of overpricing resulting from high Zestimates is small, properties still take longer to sell. Institutional sellers receive slightly lower sales price than individuals if they overprice homes when referencing flawed Zestimate values. This result indicates that institutional sellers can fare worse than individual sellers when relying on Zestimates to set prices. This outcome is consistent with the findings that if a property is overpriced due to a high Zestimate, it takes longer to sell, discounting of the original asking price is required, and the property does eventually sell at a premium. However, if institutional sellers overprice a property due to reasons not captured by a Zestimate, they typically receive a higher

price than asking price, and eventually sell at higher prices than individual sellers. Though private information helps institutional sellers receive a higher price relative to individual sellers, this pricing effect is trivial when compared to parameters evaluated in the existing literature. As Zestimates provide a publicly available reference with virtually no access barriers, the performance gap between individual and institutional sellers may converge.

4.4. Zestimate Sensitivity

Market participants rely on Zestimates when they believe that Zestimates are helpful and accurate. As market conditions are inputs into Zillow's Zestimate algorithm, Zestimates correspondingly change as market conditions change or parameters are incorrectly input or weighted into the algorithm. For 42 percent of properties, Zestimates fluctuate within a 2-3 percent range in the 6 months prior to a listing while for the other 58 percent of properties in the sample, Zestimates can fluctuate as much as 35 percent downwards and as high as 130 percent upwards. Due to uncertainty, if properties' Zestimates are highly volatile, sellers may avoid referencing Zestimates when making a decision. In **Table 8**, I study if sellers' level of reliance on Zestimates in determining a listing price varies with properties' Zestimates' level of volatility. I measure monthly Zestimate growth for the time period starting from 6 months prior to a listing event to evaluate if Zestimates' impact on listing prices weaken when properties experience recent fluctuations in their Zestimates. When Zestimates are stable in the 6 months prior to listing, sellers are likely to list properties at a price 0.9 percent higher than properties' calculated predicted value when Z-Bias increases by 1 percent. However, when past Zestimates experience larger pricing shifts, the effects decrease, reducing Z-Bias' impact down from 0.9 percent to 0.57 percent. Lastly, I run a regression using a subsample of properties that experienced downward Zestimate trends in the 6 months prior to a listing and for properties that experienced upward Zestimate trends over the same time period respectively. The last two Columns illustrate that sellers rely more heavily on Zestimates when Zestimates exhibit an upward trend in the 6 months prior to listing. The Z-Bias effect on pricing bias is 0.69 percent when properties' recent Zestimates decline while the Z-Bias' effects on pricing bias increases to 0.81 percent when properties have upward Zestimate trends. These findings imply that sellers reference Zestimates if Zestimates provide what they discern as consistent and reliable valuations, and they rely on Zestimates to a

greater extent if a property experiences steady, upward valuations.

5. Conclusion

Like most industries, the real estate market is seeking to rapidly incorporate machine learning models into business practices to minimize inefficiencies and optimally price assets. Leveraging such capabilities, PropTech firms attempt to provide accurate property value estimates to the public. Using the largest PropTech firm's home value estimates data, Zillow's Zestimate, I investigate how this publicly available information affects sellers' listing price decisions and ultimately, sales outcomes. Contrary to the commonly held belief that open information helps markets more consistently arrive at optimal outcomes, my analysis appears to indicate that Zestimates can drive greater mispricing in the housing market. I find that sellers are more likely to list their asset at a price higher than similar properties when their Zestimates are overestimated. By decomposing mispricing caused by Zestimates from overall mispricing, I investigate the impact Zestimates have on sales outcomes. Sellers who overprice their properties due to high Zestimates experience longer listing periods and consequently adjust prices downward. However, properties priced high due to overestimated Zestimates do typically sell at a premium, which is consistent with previous literature on the effects anchoring can have on listing prices.

My findings imply that incorrect online property value estimates can make the housing market inefficient. In addition to delays of sales, mispricing due to inflated Zestimates is a possible explanation for sellers' withdrawal behavior. Sellers withdraw properties from the market as Zestimates prior to listing for withdrawn properties are typically overestimated to a larger extent than properties that sold. Those findings imply the increased transaction costs due to biased online property values in the housing market. I also find that Zestimates influence different market participant segments to varying extent depending on the participants' level of access to information. Individual sellers rely on Zestimates to a larger extent than institutional sellers when they make a listing price decision.

This paper sheds light on the economic impact algorithm derived valuations can have on markets. Incorrect parameter biases, that is the weights assigned to the various inputs in a data model can lead to improper valuations, and since these models can output values quickly and at global scale,

the inefficiencies propagated throughout a market can be significant. Inadvertently, this analysis indicates that there is still value in wisdom, that is human judgement derived through experience and a deep understanding of market fundamentals when evaluating properties' predicted value. Better informed, experienced sellers demonstrate that they are not as privy to online property values such as Zestimates.

References

[1] Anderson, R. I., Brastow, R. T., Turnbull, G. K., and Waller, B. D. Seller overpricing and listing contract length: The effects of endogenous listing contracts on housing markets. *The Journal of Real Estate Finance and Economics 49*, 3 (2014), 434–450.

[2] Anenberg, E. Information frictions and housing market dynamics. *International Economic Review 57*, 4 (2016), 1449–1479.

[3] Anglin, P. M., Rutherford, R., and Springer, T. M. The trade-off between the selling price of residential properties and time-on-the-market: The impact of price setting. *The Journal of Real Estate Finance and Economics 26*, 1 (2003), 95–111.

[4] Basu, S., and Markov, S. Loss function assumptions in rational expectations tests on financial analysts' earnings forecasts. *Journal of Accounting and Economics 38* (2004), 171–203.

[5] Bucchianeri, G. W., and Minson, J. A. A homeowner's dilemma: Anchoring in residential real estate transactions. *Journal of Economic Behavior & Organization 89* (2013), 76–92.

[6] Chen, Q., and Jiang, W. Analysts' weighting of private and public information. *The Review of financial studies 19*, 1 (2006), 319–355.

[7] Chinco, A., and Mayer, C. Misinformed speculators and mispricing in the housing market. *The Review of Financial Studies 29*, 2 (2016), 486–522.

[8] Clauretie, T. M., and Thistle, P. D. The effect of time-on-market and location on search costs and anchoring: the case of single-family properties. *The Journal of Real Estate Finance and Economics 35*, 2 (2007), 181–196.

[9] Corcoran, C., and Liu, F. Accuracy of zillow's home value estimates. *Real Estate Issues 39*, 1 (2014), 45–49.

[10] De Wit, E. R., and Van der Klaauw, B. Asymmetric information and list-price reductions in the housing market. *Regional Science and Urban Economics 43*, 3 (2013), 507–520.

[11] Genesove, D., and Mayer, C. Loss aversion and seller behavior: Evidence from the housing market. *The quarterly journal of economics 116*, 4 (2001), 1233–1260.

[12] Harding, J. P., Knight, J. R., and Sirmans, C. Estimating bargaining effects in hedonic models: Evidence from the housing market. *Real Estate Economics 31*, 4 (2003), 601–622.

[13] Haurin, D. R., Haurin, J. L., Nadauld, T., and Sanders, A. List prices, sale prices and marketing time: an application to us housing markets. *Real Estate Economics 38*, 4 (2010), 659–685.

[14] Knight, J. R. Listing price, time on market, and ultimate selling price: Causes and effects of listing price changes. *Real Estate Economics 30*, 2 (2002), 213–237.

[15] Lambson, V. E., McQueen, G. R., and Slade, B. A. Do out-of-state buyers pay more for real estate? an examination of anchoring-induced bias and search costs. *Real Estate Economics 32*, 1 (2004), 85–126.

[16] Lee, Y. S., and Sasaki, Y. Information technology in the property market. *Information Economics and Policy 44* (2018), 1–7.

[17] Levitt, S. D., and Syverson, C. Market distortions when agents are better informed: The value of information in real estate transactions. *The Review of Economics and Statistics 90*, 4 (2008), 599–611.

[18] Lu, G. How machine learning mitigates racial bias in the us housing market. *Available at SSRN 3489519* (2019).

[19] Malik, N. Does machine learning amplify pricing errors in housing market?: Economics of ml feedback loops. *Economics of ML Feedback Loops (September 18, 2020)* (2020).

[20] Miller, N., and Sklarz, M. Pricing strategies and residential property selling prices. *Journal of Real Estate Research 2*, 1 (1987), 31–40.

[21] Muralidharan, S., Phiri, K., Sinha, S. K., and Kim, B. Analysis and prediction of real estate prices: A case of the boston housing market. *Issues in Information Systems 19*, 2 (2018), 109–118.

[22] Northcraft, G. B., and Neale, M. A. Experts, amateurs, and real estate: An anchoring-and-adjustment perspective on property pricing decisions. *Organizational behavior and human decision processes 39*, 1 (1987), 84–97.

[23] Ooi, J. T., and Le, T. T. The spillover effects of infill developments on local housing prices. *Regional Science and Urban Economics 43*, 6 (2013), 850–861.

[24] Park, C. W., and Stice, E. K. Analyst forecasting ability and the stock price reaction to forecast revisions. *Review of Accounting Studies 5*, 3 (2000), 259–272.

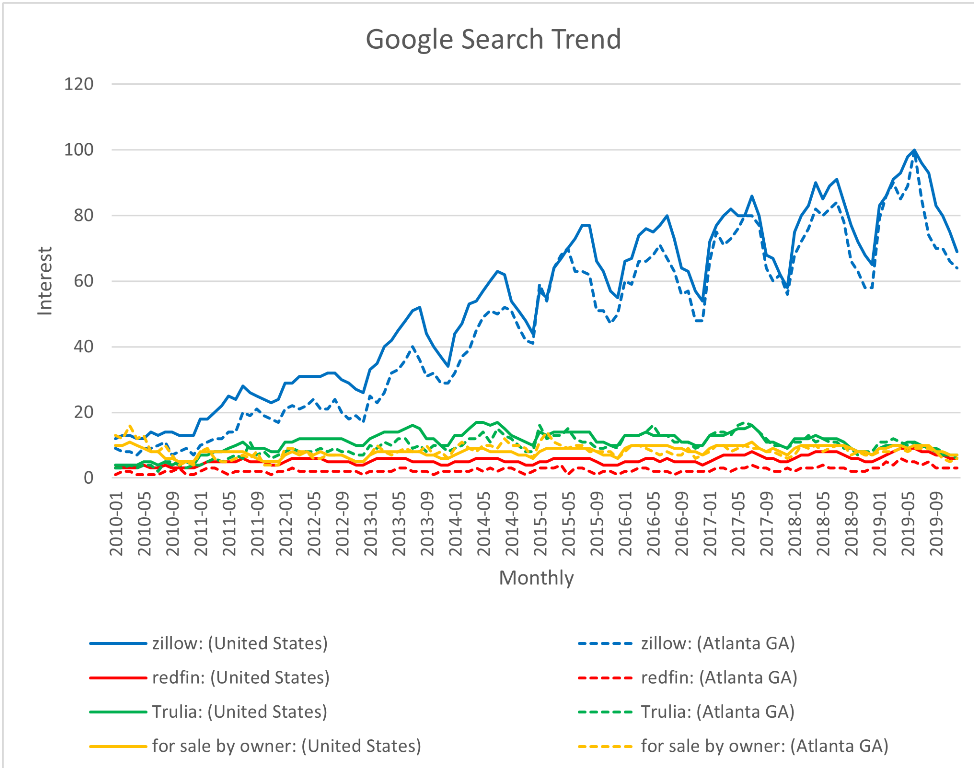
[25] Rutherford, R. C., Springer, T. M., and Yavas, A. Conflicts between principals and agents: evidence from residential brokerage. *Journal of financial Economics 76*, 3 (2005), 627–665.

[26] Stickel, S. E. Reputation and performance among security analysts. *The Journal of Finance 47*, 5 (1992), 1811–1836.

[27] Trippi, R. R. Estimating the relationship between price and time to sale for investment property. *Management Science 23*, 8 (1977), 838–842.

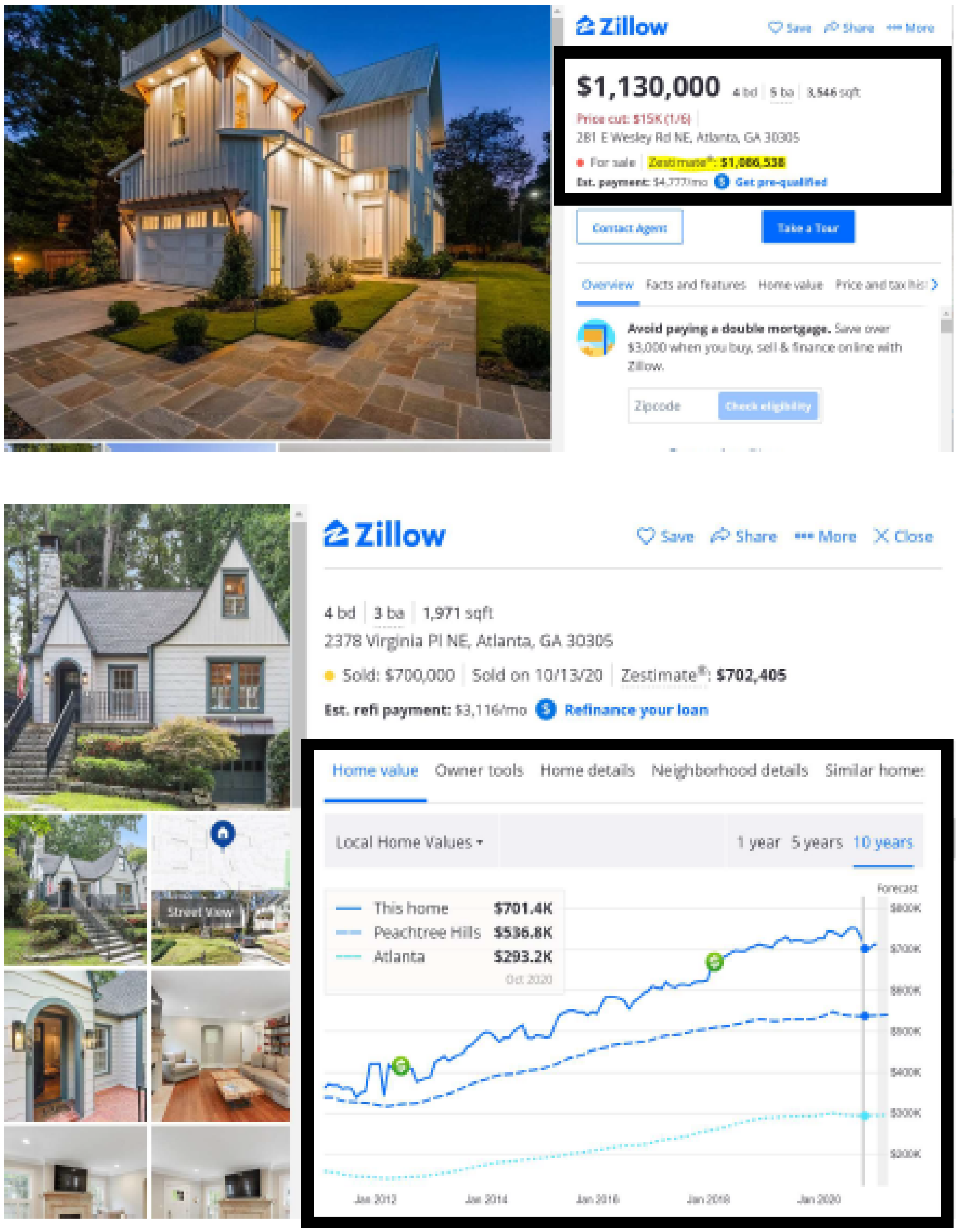
[28] Yavas, A., and Yang, S. The strategic role of listing price in marketing real estate: theory and evidence. *Real Estate Economics 23*, 3 (1995), 347–368.

Figure 1. Google Search Trend



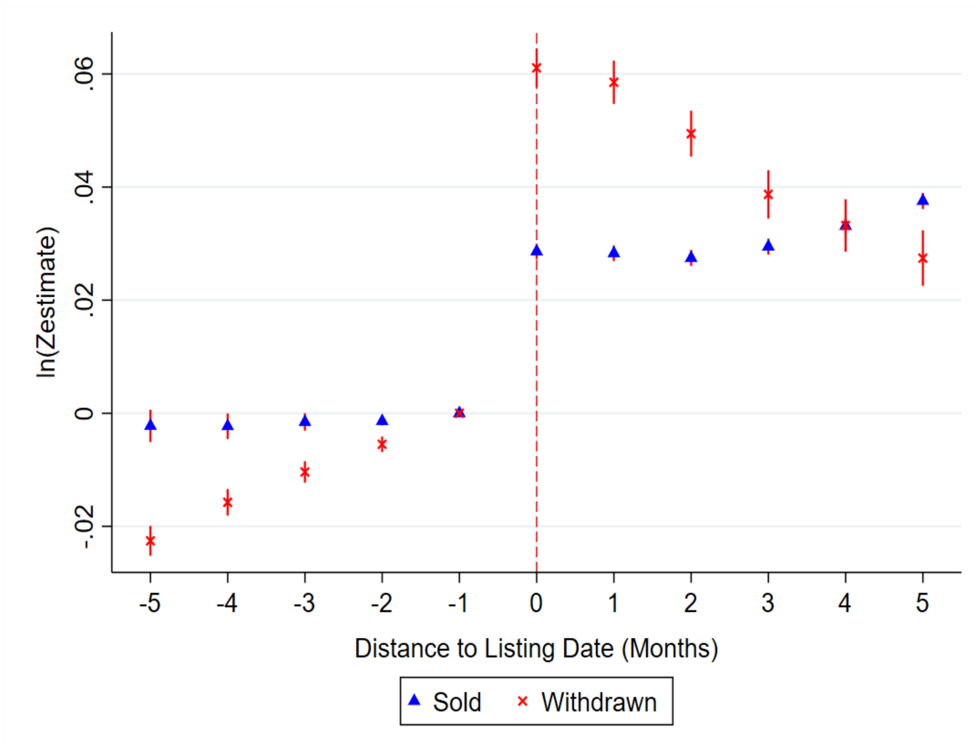
Notes: The figure shows that the general interest on Zillow increases continuously using Google search trend algorithm. Solid lines indicate the national level interest while dash lines show the interest of the metro Atlanta area.

Figure 2. Zestimate



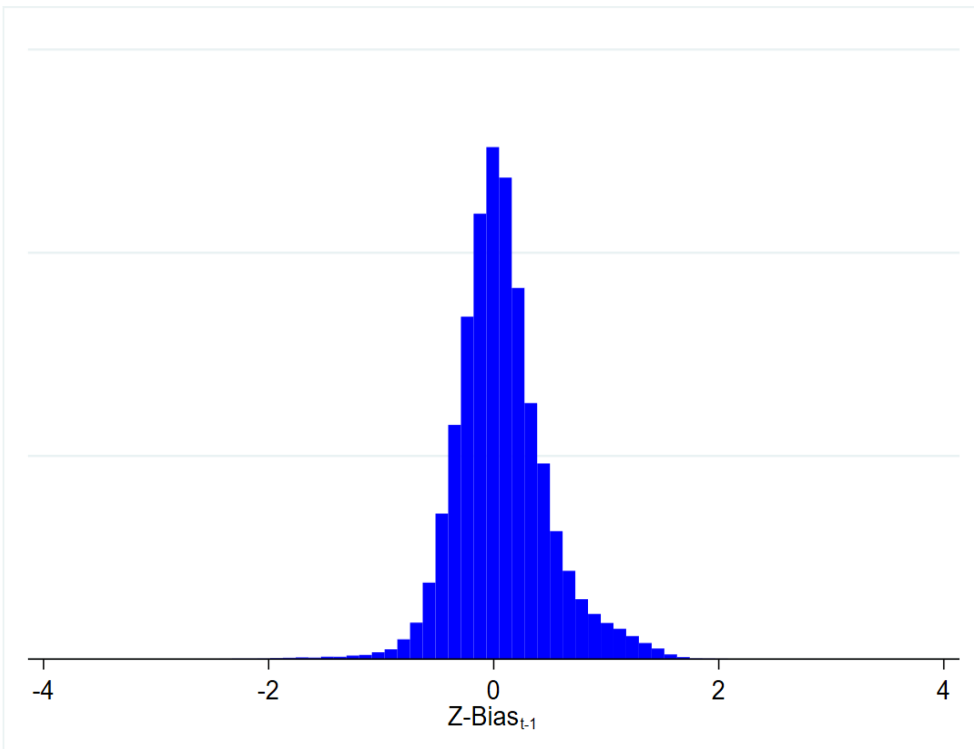
Notes: The figures show examples of Zestimates shown on Zillow's website. The current zestimates are shown underneath the listing price and the historical Zestimate trends are under 'Home Value' section.

Figure 3. Zestimate Changes



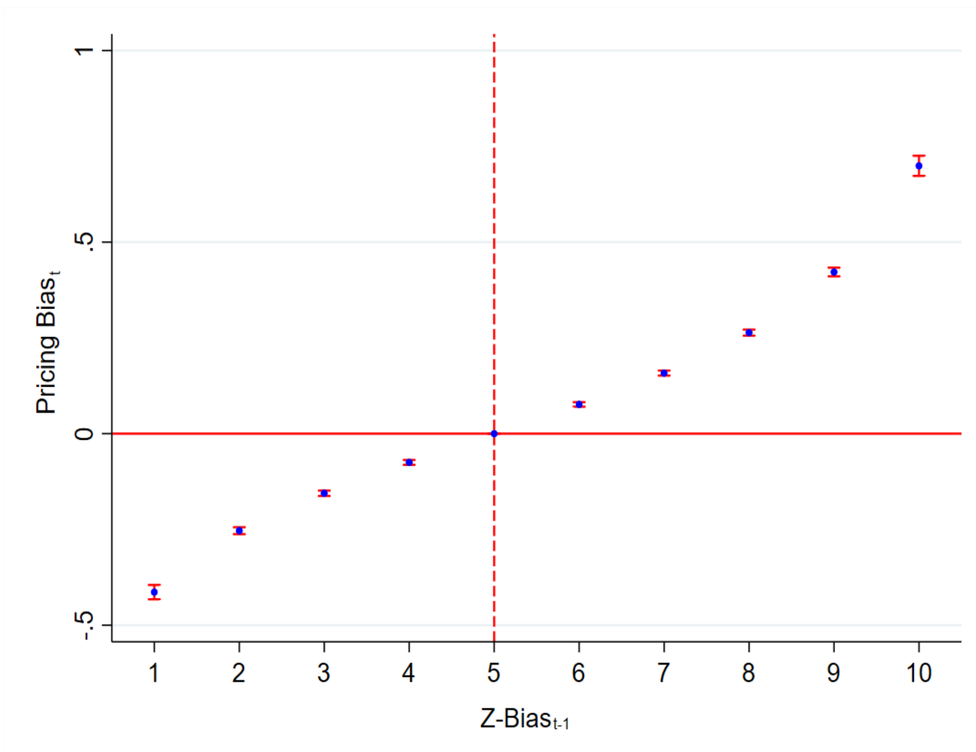
Notes: The figure shows Zestimate changes of sold properties (blue triangle) and withdrawn properties (red cross) around listing month respectively. Y-axis indicates Zestimates which are normalized at one month prior to listing month. X-axis represents the distance to listing month. 0 indicates the month when a property is listed. The red bars denote 95 percent confidence intervals.

Figure 4. Histogram of Z-Bias



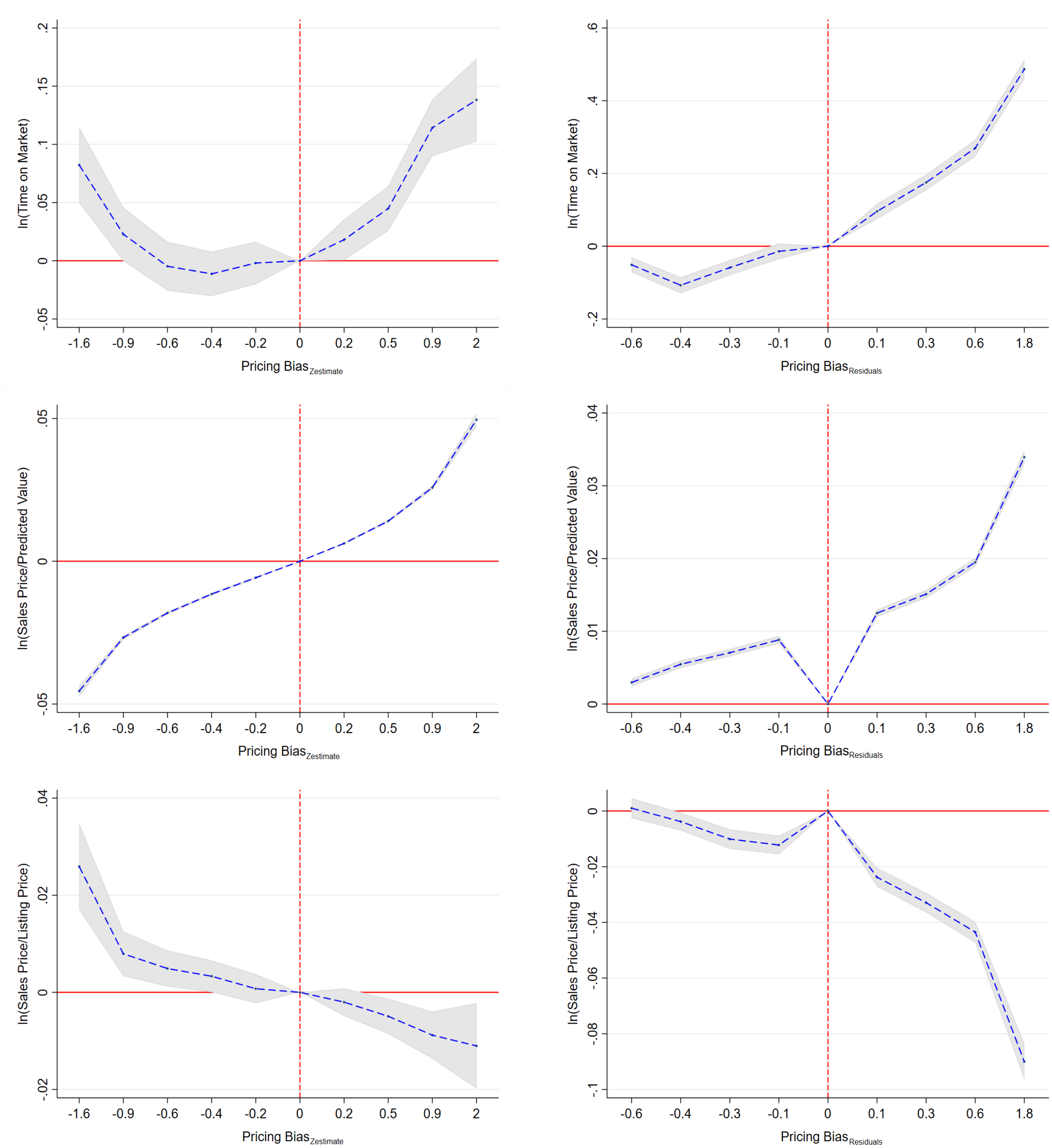
Notes: The figure shows the distribution of Z-Bias. The average Z-Bias is 0.068 while the median Z-Bias is 0.036. There are 49 bins and each bin's width is 0.11 bias level.

Figure 5. Coef. Plot for Pricing Bias



Notes: The figure shows the probability of pricing bias by Zestimate bias level. X-axis label indicates Zestimate bias group. 1 indicates the lowest Z-Bias group while 10 indicates the highest Z-Bias group. The red bars denote 95 percent confidence intervals.

Figure 6. Coef. Plots for Sales Outcomes



Notes: The graphs shows the impact of mispricing has on sales outcomes such as days on market, sales price compared to predicted value, and sales price to original listing price. Mispricing is decomposed to two factors: pricing bias associated with Zestimate and pricing bias from other factors. X-axis label indicates the mean value of each group. The shaded regions denote 95 percent confidence intervals.

Table 1 Summary Statistics

	(1) N	(2) mean	(3) sd	(4) min	(5) max
Num. of Bed	93,143	4	1	2	6
Num. of Bath	93,143	3	1	1	7
Bldgsize (SF)	93,143	2,574	1,197	924	7,007
Property Age	93,143	28	19	1	97
Sales Price ($)	93,143	282,695	188,251	31,500	1,150,000
Listing Price ($)	93,143	301,741	213,107	44,900	1,350,000
Zestimate$_{-1m}$($)	93,143	283,786	220,992	11,637	1,821,584
ln(Predicted Value ($))	93,143	12	0.45	11	13.71
Pricing Bias$_t$	93,143	0.14	0.41	-2.41	2.90
Z-Bias$_{t-1}$	93,143	0.07	0.41	-2.66	2.87
$\widehat{PricingBias}_{Zestimate}$	93,143	0	1	-13.23	19.33
$\widehat{PricingBias}_{Residuals}$	93,143	0	1	-6.59	6.74
ln(Listing Price/Zestimate)	93,143	0	1	-21.47	28.54
Time on Market (Days)	93,143	93	96	1	579
Num. of Change	93,143	1	2	0	9
Listing Price Change (%)	28,075	-0.04	0.05	-0.53	0.11
Days to Change	28,075	58	60	1	302

Notes: This table represents the summary statistics. Predicted value is estimated using 12 months of historical transactions prior to listing time, but lagged by one month. $PricingBias_t$ is mispricing level which is measured by the difference between the log of original listing price and the log of the predicted value. $Z \cdot Bias_{t-1}$ is Zestimate bias level that is calculated by the difference between the log of Zestimate one month prior to listing month and the log of the predicted value. $\widehat{PricingBias}_{Zestimate}$ indicates mispricing explained by Zestimate while $\widehat{PricingBias}_{Residuals}$ captures overpricing from other factors. $\widehat{PricingBias}_{Zestimate}$, $\widehat{PricingBias}_{Residuals}$, and ln(listing price/Zestimate) are standardized.

Table 2 Zestimate Effects on Pricing Bias

Dep. Variable: Pricing Bias$_t$	(1)	(2)	(3)	(4)
Z-Bias$_{t-1}$	0.878***	0.756***	0.748***	0.748***
	(0.0015)	(0.0128)	(0.0126)	(0.0127)
Property Characteristics	no	no	yes	yes
ZipCode × Year × Quarter FE	no	yes	no	yes
Year × Month FE	yes	yes	yes	yes
ZipCode FE	no	no	yes	no
Cluster	no	yes	yes	yes
Observations	93,143	93,078	93,115	93,078
R-squared	0.787	0.803	0.805	0.806

Notes: This table shows the baseline results for the Z-Bias effects on pricing bias. The dependent variable, $PricingBias_t$, is measured by the difference between the log of the listing price and the log of the predicted value. $Z \cdot Bias_{t-1}$ is calculated by the difference between the log of Zesitmate one month prior to listing and the log of the predicted value. Each column includes different controls and fixed effects. Column (1) includes year-month fixed effect only. Column (2) includes zip code by year-month and year-month fixed effects and standard errors are clustered at zip code by year level. Column (3) adds zip code fixed effects and property characteristics controls from Column (1). Column (4) additionally controls for property characteristics from Column (2). Standard errors are clustered at zip code by year level in Column (3) and (4). Standard errors are in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

Table 3 Zestimate Effects on Pricing Bias by Google Search level

Dep. Variable: Pricing Bias$_t$	(1) All	(2) 2015	(3) 2016	(4) 2017	(5) 2018	(6) 2019
Z-Bias$_{t-1}$	0.624***	0.498***	0.503***	0.651***	0.757***	0.808***
	(0.0373)	(0.0790)	(0.0576)	(0.0441)	(0.0560)	(0.0513)
Z-Bias$_{t-1}\times$ Gsearch	0.027***	0.039***	0.041***	0.026***	0.013*	0.010*
	(0.0041)	(0.0095)	(0.0069)	(0.0054)	(0.0071)	(0.0066)
Property Characteristics	yes	yes	yes	yes	yes	yes
City $\times$ Year FE	yes	yes	yes	yes	yes	yes
Year $\times$ Month FE	yes	yes	yes	yes	yes	yes
Observations	93,098	16,530	18,977	20,106	19,713	17,505
R-squared	0.801	0.790	0.803	0.824	0.829	0.809

Notes: This table shows the robustness test for the baseline result using Google search outcomes of *'Zillow'*. Z-Bias affects pricing bias even after controlling for the location variance by *'Zillow'* interest measured by Google trend. The dependent variable, $PricingBias_t$, is measured by the difference between the log of listing price and the log of the predicted value. The independent variable of interest is the interaction of Z-Bias and GSearch. $Z \cdot Bias_{t-1}$ is calculated by the difference between the log of Zesitmate one month prior to listing and the log of the predicted value. GSearch is Google's search score of *'Zillow'* from Google Trends. Google search interest is given at the city and year level. Property characteristics such as number of bedrooms, number of bathrooms, age, age squared, and the log of building size in square feet are included. Standard errors are clustered at city by year level. Standard errors are in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

Table 4 Pricing Bias Effects on Sales Outcomes

Dep. Variables:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	ln(Time on Market)			ln(Sales Price/Predicted Value)			ln(Sales Price/Listing Price)		
	All	Low-Low	High-High	All	Low-Low	High-High	All	Low-Low	High-High
$\widehat{PricingBias}_{Zestimate}$	0.024***	-0.075***	0.055***	0.029***	0.030***	0.029***	-0.015***	0.005	-0.014***
	(0.0087)	(0.0203)	(0.0150)	(0.0002)	(0.0003)	(0.0002)	(0.0024)	(0.0043)	(0.0021)
$\widehat{PricingBias}_{Residuals}$	0.141***	0.002	0.242***	0.012***	0.012***	0.010***	-0.042***	-0.044***	-0.074***
	(0.0052)	(0.0108)	(0.0119)	(0.0002)	(0.0005)	(0.0004)	(0.0023)	(0.0057)	(0.0045)
Property Characteristics	yes	yes	yes	yes	yes	yes	yes	yes	yes
ZipCode × Year × Quarter FE	yes	yes	yes	yes	yes	yes	yes	yes	yes
Year × Month FE	yes	yes	yes	yes	yes	yes	yes	yes	yes
Observations	93,077	27,553	27,151	93,077	27,553	27,151	93,077	27,553	27,151
R-squared	0.141	0.121	0.174	0.900	0.752	0.921	0.153	0.131	0.272

Notes: After diving pricing bias into two segments, pricing bias from Zestimate ($\widehat{PricingBias}_{Zestimate}$) and pricing bias from others ($\widehat{PricingBias}_{Residuals}$), I test how these two segments affect sales outcomes. Dependent variables are sales outcomes such as the log of days on market, the log of sales price to predicted value and the log of sales price to original listing price. Key independent variables are pricing bias from Zestimate,($\widehat{PricingBias}_{Zestimate}$), which is the estimated value from equation (2) and its residuals,($\widehat{PricingBias}_{Residuals}$), that indicates pricing bias from other factors. Both values are standardized. Column labeled 'All' shows the regression results from the entire sample. Using the median of pricing bias caused by Zestimate and residuals, I divide the sample into two groups: low and high. Low-Low includes the sample where both mispricing factors are low end while High-High includes the sample where both mispricing segments are high. Property characteristics such as number of bedrooms, number of bathrooms, age, age squared, and the log of building size in square feet are included. Standard errors are clustered at zip code × year level. Standard errors are in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

Table 5 Listing Price Change

Dep. Variables:	(1) I(Any Change)	(2) Price Change (%)	(3) ln(Days to Change)	(4) I(Positive Change)	(5) I(Negative Change
$\widehat{PricingBias}_{Zestimate}$	0.033***	-0.973***	0.123***	-0.052***	0.048***
	(0.0038)	(0.0936)	(0.0149)	(0.0047)	(0.0048)
$\widehat{PricingBias}_{Residuals}$	0.069***	-0.941***	0.207***	-0.013***	0.011***
	(0.0019)	(0.0703)	(0.0092)	(0.0028)	(0.0033)
Property Characteristics	yes	yes	yes	yes	yes
ZipCode × Year × Quarter FE	yes	yes	yes	yes	yes
Year × Month FE	yes	yes	yes	yes	yes
Observations	93,077	28,075	28,075	28,075	28,075
R-squared	0.084	0.160	0.146	0.143	0.140

Notes: This table shows how Zestimate affects sellers' listing price change decisions. Dependent variables are whether seller changes its listing price or not, listing price adjustment rate, days on market until price adjustment, indicator of positive change, and that of negative change, respectively. Key independent variables are pricing bias explained by Zestimate and pricing bias caused by other factors which are not associated with Zestimate. Both values are standardized. Property characteristics such as number of bedrooms, number of bathrooms, age, age squared, and the log of building size in square feet are included. Standard errors are clustered at zip code × year level. Standard errors are in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

Table 6 Sold vs. Withdrawn Properties

Panel A. Matching Outcomes

	Sold	Withdrawn	t-test
Bed	4.09	4.07	0.02 (1.31)
Bath	3.07	3.08	-0.01 (-0.51)
Property Age	22.31	22.14	0.17 (0.74)
ln(BldgSize)	7.88	7.88	0 (-0.89)
Zestimate ($)	247,707	250,196	-2,489 (-0.93)

Panel B. Z-Bias Effects on Pricing Bias

	(1)	(2)	(3)	(4)	(5)	(6)
	Entire Sample			Matched Sample		
Dep. Variables:	I(Withdraw)	Pricing Bias$_t$		I(Withdraw)	Pricing Bias$_t$	
Z-Bias$_{t-1}$	0.058***	0.843***	0.842***	0.223***	0.761***	0.759***
	(0.0429)	(0.0098)	(0.0098)	(0.3605)	(0.0640)	(0.0640)
I(Withdraw)		0.003***			0.002***	
		(0.0003)			(0.0003)	
Z-Bias$_{t-1}\times$ I(Withdraw)			0.003***			0.002***
			(0.0002)			(0.0003)
Property Characteristics	yes	yes	yes	yes	yes	yes
ZipCode $\times$ Year $\times$ Quarter FE	yes	yes	yes	yes	yes	yes
Year $\times$ Month FE	yes	yes	yes	yes	yes	yes
Pair FE	no	no	no	yes	yes	yes
Observations	118,910	118,910	118,910	18,530	18,530	18,530
R-squared	0.6070	0.8046	0.8046	0.6213	0.9033	0.9033

Notes: This table shows that Zestimate of withdrawn properties is more biased compared to properties that are sold during the sample period of time. Panel A shows the matching outcomes from 1:1 propensity score matching method with 0.01 caliper without replacement. Property characteristics such as number of bedrooms, number of bathrooms, age, age squared, and the log of building size in square feet are included. Zestimate one month prior to listing and the location of property is also controlled in the matching process. The last Column of Panel A shows the difference in values between two groups and statistical results are shown in parenthesis. Panel B represents the outcomes of Z-Bias effects on Pricing Bias. I(Withdraw) is a binary variable that indicates whether a property is withdrawn or not. Standard errors are clustered at zip code by year level. Standard errors are in parentheses; *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 7 Individual vs. Institutional Sellers

Panel A. Matching Outcomes

	Ind. Seller	Inst. Seller	t-test
Bed	3.68	3.69	-0.01 (-0.96)
Bath	2.8	2.8	0 (0.14)
Property Age	29.66	29.17	0.49 (1.25)
ln(BldgSize)	7.65	7.66	-0.01 (-1.28)
Zestimate($)	231,116	235,748	-4,632 (-1.27)

Panel B. Z-Bias Effects on Pricing Bias

Dep. Variable: Pricing Bias$_t$	(1) Ind. Seller	(2) Inst. Seller	(3) Both
Z-Bias$_{t-1}$	0.696***	0.553***	0.320***
	(0.0299)	(0.0308)	(0.0762)
Z-Bias$_{t-1}\times$ Inst. Seller			-0.042***
			(0.0148)
Property Characteristics	yes	yes	yes
Pair FE	no	no	yes
ZipCode × Year × Quarter FE	yes	yes	yes
Year × Month FE	yes	yes	yes
Observations	6,153	6,153	12,306
R-squared	0.818	0.721	0.894

Panel C. Mispricing Effects on Sales Outcomes

Dep. Variables:	(1) ln(Time on Market)	(2) ln(Sales Price/ Predicted Value)	(3) ln(Sales Price/ Listing Price)
$\widehat{PricingBias}_{Zestimate}$	-0.017	0.030***	-0.002
	(0.0218)	(0.0005)	(0.0050)
$\widehat{PricingBias}_{Zestimate}\times$ Inst. Seller	0.006	-0.001***	-0.012***
	(0.0197)	(0.0004)	(0.0043)
$\widehat{PricingBias}_{Residuals}$	0.124***	0.012***	-0.043***
	(0.0164)	(0.0005)	(0.0057)
$\widehat{PricingBias}_{Residuals}\times$ Inst. Seller	0.004	0.001***	0.016**
	(0.0199)	(0.0006)	(0.0068)
Property Characteristics	yes	yes	yes
Pair FE	yes	yes	yes
ZipCode × Year × Quarter FE	yes	yes	yes
Year × Month FE	yes	yes	yes
Observations	12,306	12,306	12,306
R-squared	0.181	0.792	0.245

Notes: This table shows how institutional sellers react to Zestimate differently compared to individual sellers. Panel A shows the t-test results from matching process. I use 1:1 propensity score matching with 0.01 caliper without replacement. Property characteristics and Zestimate one month prior to listing month is used in the matching process. The last Column of Panel A shows the difference in values between two groups and statistical results are shown in parenthesis. Panel B describes the results that institutional sellers utilize Zestimate information differently when compared to individual sellers. Panel C shows the impact mispricing has on sales outcomes such as days on market, the difference between sales price to predicted value, and the difference between original listing price to sales price. Inst.Seller indicates whether a seller is an institutional seller or not. Property characteristics such as number of bedrooms, number of bathrooms, age, age squared, and the log of building size in square feet are included. Standard errors are clustered at zip code by year level. Standard errors are in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

Table 8 Zestimate Volatility

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				$\mid \Delta(Zestimate_t) \mid$			
Dep. Variable: Pricing Bias$_t$	Within 3%	3% to 5%	5% to 10%	10% to 20%	Above 20%	Decreasing Trend	Increasing Trend
Z-Bias$_{t-1}$	0.895***	0.744***	0.670***	0.579***	0.571***	0.689***	0.808***
	(0.0102)	(0.0133)	(0.0193)	(0.0231)	(0.0266)	(0.0200)	(0.0102)
Property Characteristics	yes	yes	yes	yes	yes	yes	yes
ZipCode × Year × Quarter FE	yes	yes	yes	yes	yes	yes	yes
Year × Month FE	yes	yes	yes	yes	yes	yes	yes
Observations	38,896	37,145	19,977	10,090	4,719	27,609	65,398
R-squared	0.867	0.804	0.756	0.722	0.733	0.799	0.821

Notes: This table illustrates that Z-Bias' impact varies with Zestimates' volatility. $\mid \Delta(Zestimate_t) \mid$ indicates a monthly change of Zestimate values in the 6 months from listing event. For instance, 'Within 3%' indicates that Zestimates fluctuate stable staying within 3% range (± 3%) for the past 5 months of listing while '3% to 5%' indicates properties that the absolute values of monthly Zestimate changes range between 3% and 5%. 'Decreasing Trend' includes properties where Zestimate in the one month immediately prior to a listing is lower than Zestimates in the trailing 5 months prior to a listing. 'Increasing Trend' indicates properties where Zestimates in the month immediately prior to a listing are greater than Zestimates in the trailing 5 months. Mispricing bias is measured by the difference between the log of properties' initial asking price and their calculated predicted value while Z-Bias is measured as the difference between the log of a Zestimate in the one month immediately prior to a listing date and the log of properties' predicted value. Property characteristics such as the number of bedrooms, bathrooms, age, age squared, and the log of a building size in square feet are included. Standard errors are clustered at zip code by year level. Standard errors are in parentheses; *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

A. Appendix

A.1. Listing Price Change Data

The data used in this paper is from Zillow.com. In addition to properties' Zestimates, I also collect a log of modifications to property prices. Although the Georgia Multiple Listing Services (GMLS) provides original listing and final asking prices, the data does not indicate how many times a seller modified a listing price or the dates that adjustments occurred. **Figure A1** provides an example of listing price change information provided by Zillow's website.

A.2. Figure for Gsearch

For a robustness test, I use Google search analytics to examine if areas where households search on the keyword 'Zillow' heavily are influenced to a greater extent by properties' Zestimates than other areas. **Figure A2** presents the results graphically. The figure illustrates that Z-Bias and mispricing have a linear relationship even after accounting for Google search level. When Z-Bias is high and use of Google to search for 'Zillow' is high, asking prices are more likely to be listed above properties' predicted values.

A.3. Sold vs. Withdrawn Properties

Figure A3 illustrates the trends of Z-Bias of sold properties and withdrawn properties. Something interesting to note is that sold properties had stable Z-Bias trends prior to listing event while withdrawn properties typically saw sharp upward trends in Zestimates in the time period immediately before listing. Z-bias between sold and withdrawn properties is at a similar level 5 months prior to listing, but the gap diverges over time. The gap becomes largest in the month of listing indicating that listing prices for withdrawn properties are more likely to be higher than that sold. These findings imply that withdrawn properties had overly aggressive asking prices.

A.4. Accuracy of Zestimate

The accuracy of Zestimate is investigated by many scholars and they conclude that properties sell at prices very close to their Zestimates. Previous literature reference Zestimates in the month immediately prior to sale to determine their accuracy. To test whether sellers use Zestimates as an input to determine an asking price, I reference properties' Zestimates one month prior to listing. To test if Zestimates prior to listings are accurate, I calculate the mean squared errors of properties' sales price and Zestimate one month prior to listing. **Table A2** represents the mean squared errors of Zestimates and the predicted value estimated using the equation (1). The output indicates that the calculated predicted values that I estimate are closer to actual transaction prices relative to properties' Zestimates one month prior to listing. Overall, calculated predicted values are four times more accurate than Zestimates in predicting transaction prices. I also test the accuracy of Zestimates across different dimensions. Across all dimensions measured, the calculated predicted values used in the analysis are more accurate than Zillow's Zestimates. The gap in accuracy between the two values, that is gap of the mean squared errors between the predicted values and Zestimates, is smallest for properties less than 15 years old. When property values are greater than \$500,000, Zestimates are noisier, thus calculated predicted values are over 6 times more accurate than Zestimates.

A.5. Heterogeneity of Z-Bias

Table A3 displays which properties have more biased Zestimates. I run a regression on $Z \cdot Bias_{t-1}$ with property characteristics as explanatory variables. The result indicates that Zestimates are more likely to be mispriced upwards as a property's physical structure increases in size and age, thus implying that more expensive properties' Zestimates are more likely to be misvalued. It is consistent with the accuracy test shown in ${}^{T}ableA2^{t}hatZestimatesarenoisierswhenhomesaremoreexpensive.$

A.6. Alternative Specification

In the main analysis, I calculate properties' predicted value using transaction data from the trailing 12 month period, with a one month lag, prior to a property's sale. However, some may argue that predicted value has to be measured at the time of listing as my main interest is the effects of

Zestimates on sellers' listing behavior. Thus, I estimate the predicted value using the 12 months of transaction data from the time of listing month. For example, I use Zestimates in the time frame starting from August 2017 and to the end of July 2018 to calculate predicted values for properties *listed* in September 2018. This method allows me to measure Z-Bias by comparing the differences between Zestimates one month prior to a listing event and calculated predicted values one month prior to the listing month. **Table A4** illustrates the results from equation (2) using the calculated predicted values. The results are consistent with the baseline results, that is if a property's Zestimate is more than 1 percent higher than its predicted value, the asking price is likely to be 0.89 percent higher than the property's predicted value. The magnitude of Zestimates effects on listing prices is also similar to the main findings when using different specifications. The Z-Bias' impact on pricing bias ranges from 0.76 to 0.88 percent.

Table A5 illustrates pricing bias effects on sales outcomes when substituting calculated predicted values with the time a listing sits on market. Dependent variables are the log of days on market, the difference between the log of sales price and calculated predicted value, and the difference between the log of sales price and listing price. $\widehat{PricingBias}_{Zestimate}$ is the fitted value from equation (2), captured mispricing explained by Zestimates while $\widehat{PricingBias}_{Residuals}$ are the residuals from equation (2) that is mispricing caused by factors that are not associated with Zestimates. When the effects of mispricing due to inaccurate Zestimates become stronger, a property sits on the market longer. Listings that are mispriced by one standard deviation higher due to inaccurate Zestimates take 12.8 percent longer to sell (12 days). However, sellers can expect to receive an amount 3 percent higher price when asking prices are high by one standard deviation due to inaccurate Zestimates even when accounting for the average 1.6 percent discount from the initial list price.

A.7. Heterogeneity Test

I examine if households with greater internet access are more likely to be influenced by Zestimates. Using internet access level data from the federal communication and commission (FCC), I confirm that households living in areas with high levels of access to the internet are more likely to be influenced by Zestimates when determining at what price to list. **Table A6** indicates that sellers living in areas with low levels of internet access list properties at listing prices 63% correlated to Zestimate while

sellers living in areas with high levels of internet access list properties at prices 86% correlated to Zestimates.

A.8. Post-listing Zestimate Effects

I confirm that sellers overprice their homes with inflated Zestimates. After a property is listed on market, Zillow will immediately adjust a property's value with the revised estimate accounting for the price listed, the party listing (realtor or sale by owner), and market conditions. If Zestimate is revalued lower (higher) than listing price, buyers may perceive it as overpricing (underpricing) and are less (more) likely to put an offer, making sellers stay on market longer (shorter). To measure the extent to which the difference between the revised Zestimate and list price impacts the market, I use the difference between the log of listing prices and the log of Zestimates in the listing month. The measurement is useful as it is a direct comparison between listing price and Zestimate that market participants may actually use. **Table A7** displays the results after I standardize the difference between the log of listing price and the log of Zestimate. Column (1) illustrates properties take 2 more days to sell when listed for a price 1 standard deviation higher than a properties' Zestimates. Sellers do receive prices slightly higher than properties' predicted value when listed at a price higher than their Zestimate. However, sellers typically do have to lower their asking price by $15,000 on average to close a deal. These results indicate that market participants reference publicly available online valuations.

To test if the extent of the difference a listing price and a property's Zestimate plays a role on sales outcomes, I divide the difference between the asking price and Zestimate into 10 groups and calculate the effects on sales outcomes. **Figure A4** provides a graphical output of results. When listing price is lower than Zestimate, sellers sell their properties 15% faster than when listing price is similar to Zestimate. However, if listing prices are higher than Zestimates, sellers stay on market 3% to 15% longer. The magnitude becomes smaller for Groups 9 and 10, segmented pairings where asking prices are much higher than Zestimates. The second figure indicates that the delta between transaction prices properties' predicted values increase when listing prices are higher than their Zestimates. When listing prices are lower than properties' Zestimates, sales prices are 1% lower when asking prices are near the properties' Zestimate. The magnitude of the effect increases gradually as listing

prices rise upwards relative to their Zestimate. However, the effects are relatively small 1% range. The last graph illustrates that when listing prices are lower than Zestimates, properties sell at prices higher than asking prices. When asking prices are significantly higher than Zestimates, sellers are forced to discount significantly. These results imply that both parties rely on online property value estimates.

Next, I explore if sellers adjust listing prices if properties' Zestimates are lower than current asking prices, and if so, in what direction, and to what extent. **Table A8** illustrates that sellers listing properties at a price higher than their Zestimates in the month of their initial listing eventually adjust downwards. When properties' listing prices are one standard deviation more than their Zestimates, the probability that sellers adjust initial listing prices is 2.1% higher. Sellers adjust listing prices downward by 0.16% with one standard deviation increase on the difference between listing prices and a property's Zestimate. The more sellers overprice properties relative to their Zestimate, the faster sellers adjust their listing price. The probability of adjusting listing price downwards is 34% higher than that of upward price movements. The faster adjustment is likely due to the public availability of Zestimate, thus influencing potential buyers who will inundate a seller with bids or ignore the property completely. These results indicate that sellers continue to monitor Zestimates after listing.

A.9. Individuals vs. Institutional Buyers

This paper mainly investigates how online property value estimates, Zestimates, influence sellers' listing price decisions. However, as buyers have equal, relatively barrier less access to the same information, the paper also examines if and to what extent buyers rely on Zestimates when they purchase homes. I explore if buyers' property preferences change after accessing this information. Specifically, I examine if less informed buyers are more likely to purchase high-valued properties due to flawed Zestimates than more informed buyers. I leverage institutional buyers as a proxy for informed buyers as institutional buyers typically have access to more information and have more experience than individual buyers. Panel A of **Table A9** presents the propensity score matching outcomes between individual buyers and institutional buyers. I use a 1:1 propensity score matching method with a 0.01 caliper without replacement. Property characteristics, zip code, and transaction timing are controlled in the matching process. Unlike seller side analysis, I change the timing of Zestimate

to one month prior to sales to understand the buyer's reliance on Zestimate. The analysis indicates that similar properties are matched between two groups. Especially, both individual and institutional buyers receive similar signal from Zestimates before they purchase homes. Although number of bathrooms is statistically different between individual and institutional buyers, the difference of 0.04 bathrooms is economically trivial.

Using the matched sample, I test if institutional buyers utilize Zestimate differently in the housing market. Panel B represents how institutional buyers rely on Zestimates when purchasing properties. Z-Bias is measured as the difference between the log of Zesitmate one month prior to sales and the log of predicted value while pricing bias is measured by the difference between the log of final asking price and the log of predicted value. The final asking price measured by the listing price at sales is used to estimate Zestimate's influence on buyer's purchase decision. Buyers are likely to evaluate the final asking prices rather than original listing prices. Less informed buyers may choose a property that is overpriced as sales when Zestimate is overestimated. However, I find that there is no statistically significant difference of property choice associated with Zestimate between individual buyers and institutional buyers. Individual buyers do not particularly select homes that have overvalued Zestimates. Panel C show mispricing effects on sales outcomes. Mispricing associated with Zestimate lowers asking price, but sellers eventually receive premium. However, I am unable to find a difference in sales outcomes between individual buyers and institutional buyers when I account for mispricing caused by Zestimates. Sales outcomes are similar when I account for mispricing associated with seller's private information. The results imply that Zestimates may be used as an anchor price by individual sellers and a source of price discovery that is perceived of similar accuracy to an in-depth valuation exercise executed by an institutional investor.

A.10. Local vs. Non-local Buyers

I identify local buyers that reside in the same state as where a property is located as another proxy for informed buyers[14]. Non-local buyers that travel from other states to purchase a property typically pay a premium relative to local buyers due to higher search costs and greater reliance on price anchors (e.g., Harding et al., 2003; Lambson, et al, 2004; Clauretie and Thistle, 2007; Chinco and

[14] I am unable to identify non-local sellers due to data limitation.

Mayer, 2016). Non-local buyers may depend on Zestimates more than local buyers to overcome asymmetric information. Non-local buyers can take advantage of Zestimates if Zestimates provides unbiased expected property values mitigating asymmetric information local buyers may have. To test if non-local buyers rely on Zestimates differently than local buyers,I first match the sample using a 1:1 propensity score matching method with a 0.01 caliper without replacement. I use property characteristics, Zestimates one month prior to sales, and zip codes in the matching process. I also limit the transaction time period to 12 months to control time-varying trends. Panel A of **Table A10** describes how property characteristics are statistically indistinguishable between local and non-local buyers. Especially, all parties see the same or very similar Zestimates value prior to executing a transaction.

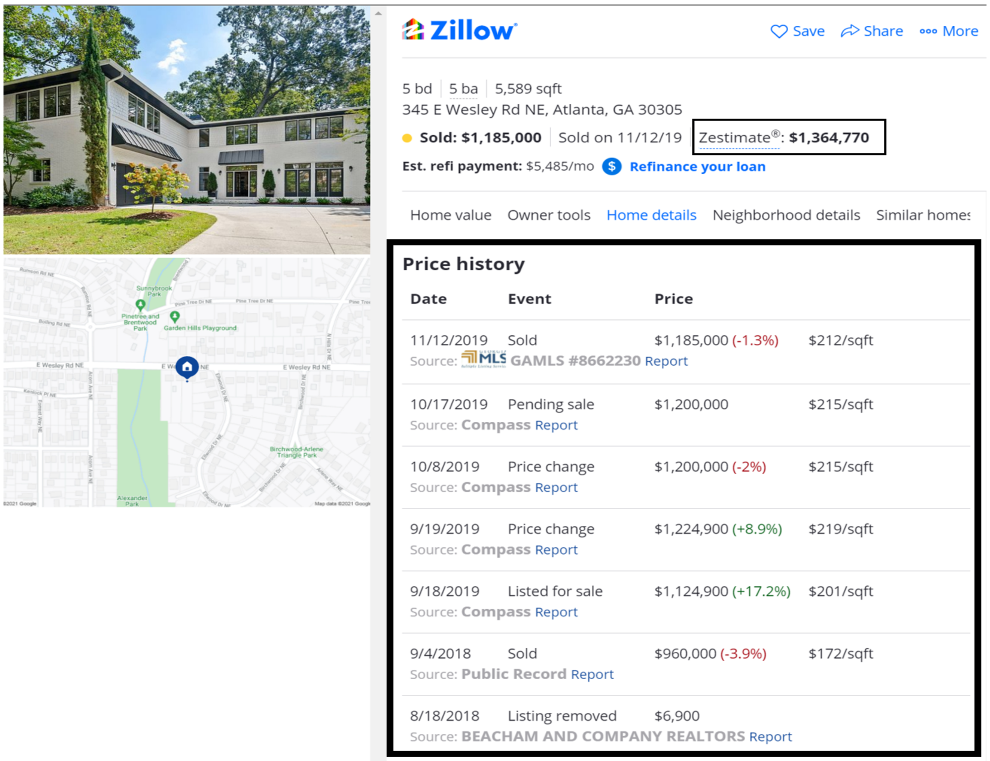
Panel B shows the results Z-Bias has on pricing bias. Z-Bias is measured by the difference between the log of a Zestimate and the log of a property's predicted value. I use Zestimates one month prior to transaction execution to investigate how buyers use online property valuations. A property's pricing bias is captured by the difference between the log of asking price at sale and the log of the property's predicted value. Column (1) shows the subsample analysis on Z-Bias effects on pricing bias using the sample of local buyers only. In column (2), I use a subsample of non-local buyers only while I use both samples in column (3). The results indicate that non-local buyers rely more heavily on Zestimate than local buyers. Non-local buyers are more likely to purchase homes that are overpriced due to a high Zesimate. The coefficient of the interaction term between Z-Bias and non-local buyer indicates that when a property's Zestimate is 1% higher than its predicted value, homes purchased by non-local buyers are listed at a price 10% higher than similar homes purchased by local buyers.

Do non-local buyers' higher reliance on Zestimates cause them to transact properties at price levels similar to local buyers? Panel C shows the results of mispricing effects on sales outcomes between local and non-local buyers. Column (1) shows mispricing effects on properties' days on market. On average, non-local buyers take 6 days longer to purchase a property than local buyers when Z-Bias is high. The delays could be due to higher search costs and/or overpricing driven by high Zestimates. Column (2) represents the result when non-local buyers pay premiums when they rely on a property's Zestimate. When a property is overpriced due to highly estimated Zestimate,

non-local buyers do pay premiums relative to local buyers. However, the pricing premium is tiny. The results on discount rate in Column (3) shows that non-local buyers are willing to pay 1.3% more than asking price when a Zestimate is overestimated by one standard deviation. Although non-local buyer's sales outcomes are different from local buyers when a property is overpriced due to its high Zestimate, non-local buyers do not pay premiums for mispricing resulting from other factors that are not captured in Zestimates. These findings imply that Zestimates are used as a reference price to non-local buyers, but this one parameter does not fully overcome the information asymmetries associated with geographic distance.

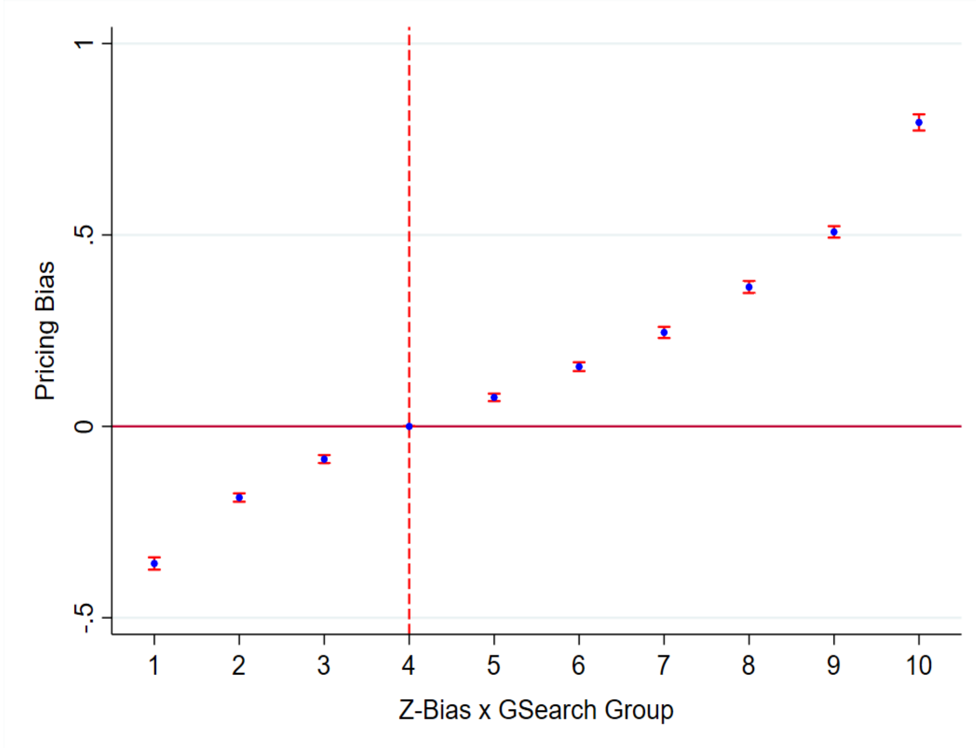
B. Figures

Figure A1. Price Change History



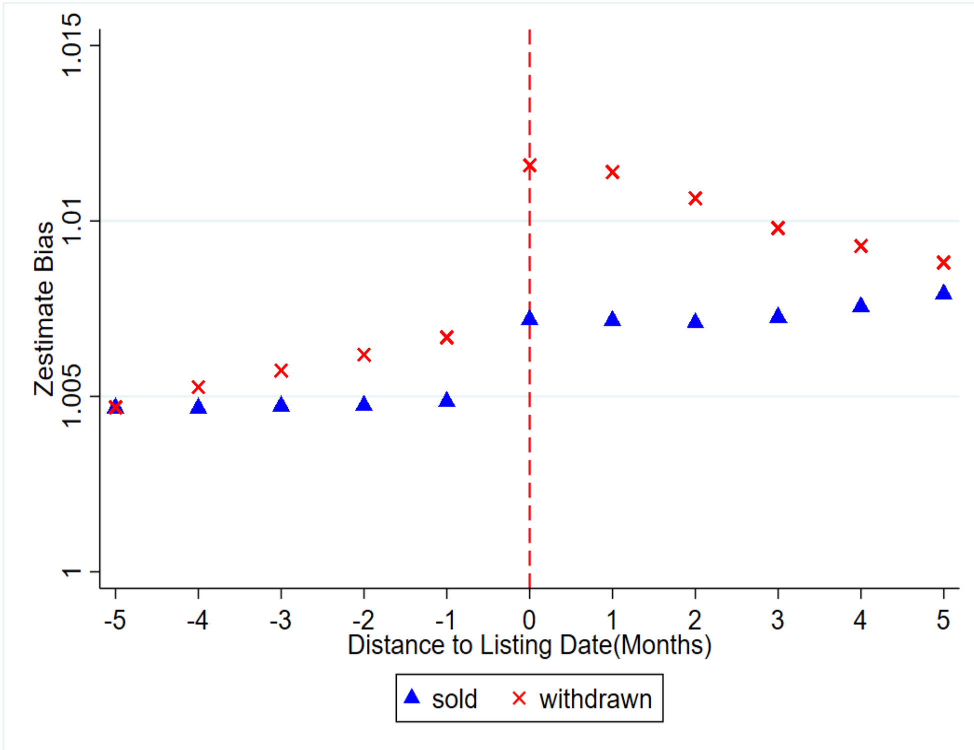
Notes: The figure shows an example of listing price change history on zillow website.

Figure A2. Coef. Plot for Pricing Bias by Z-Bias and Google Search



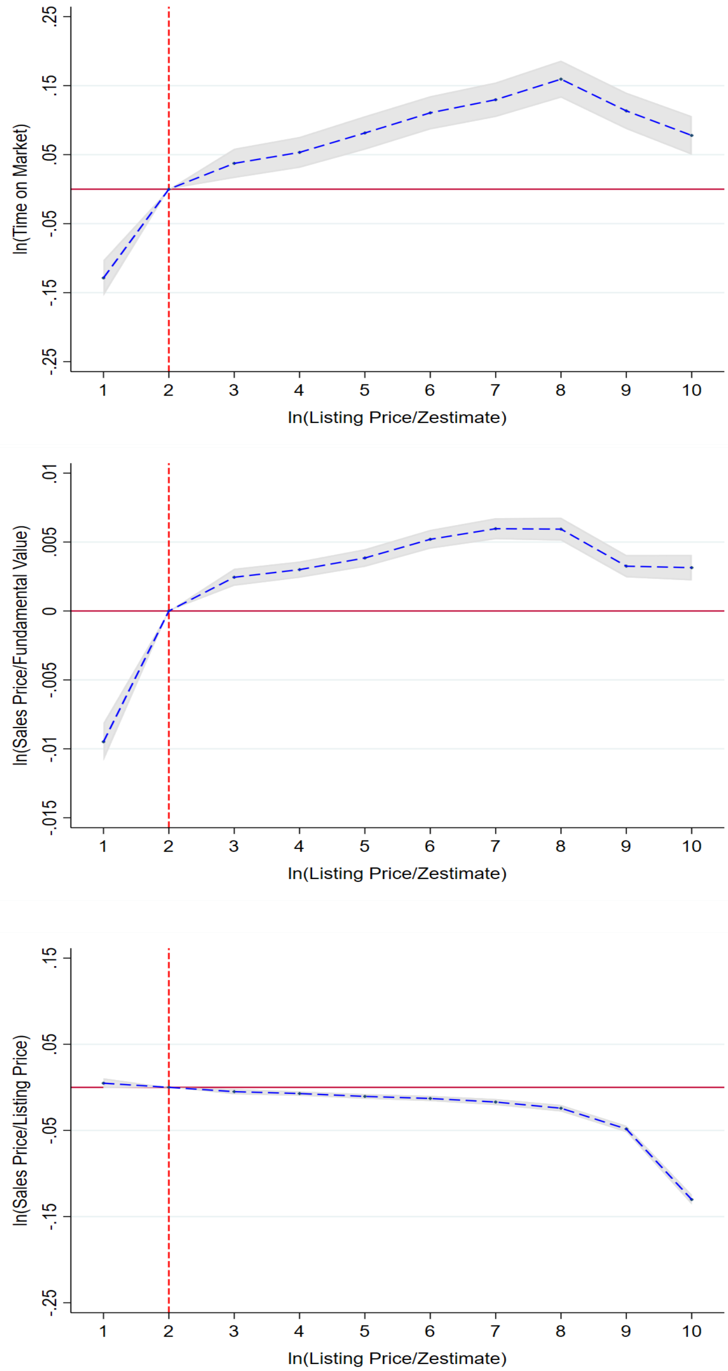
Notes: The figure shows the coefficient of pricing bias associated with Zestimate bias and Google Search trend of 'Zillow'. X-axis label indicates the group. 1 is the lowest while 10 is the highest group where Z-Bias is high and Google Search level is high. Confidence interval is at 95 percent.

Figure A3. Sold vs. Withdrawn



Notes: The figure shows Z-bias changes of sold properties and withdrawn properties when it is on market respectively.

Figure A4. Coef. Plots of Post-listing Zestimate Effects on Sales Outcomes



Notes: The graphs shows the impact of the difference between listing price and Zestimate has on sales outcomes such as days on market, sales price compared to predicted value, and sales price to original listing price. Zestimate is used at the time of listing month. X-axis label indicates groups of the difference between listing price to Zestimate. 1 indicates the lowest listing price compared to Zestimate group while 10 indicates the highest listing price compared to Zestimate group. The red dash line indicates the group where listing prices are closest to Zestimates. The shaded regions denote 95% confidence intervals.

C. Tables

Table A1 Variable Definitions

Variables	Definition
Predicted Value	Estimated market values using a henonic model with the past 12 months historical transaction but one month lagged
Pricing Bias	The difference between the log of original listing price and the log of predicted value
Z-Bias	The difference between the log of Zestimate one month prior to event and the log of predicted value
$\widehat{PricingBias}_{Zestimate}$	Mispricing explained by Z-Bias
$\widehat{PricingBias}_{Residuals}$	Mispricing associated with other factors that are not captured in Zestimate (i.e. Seller's financing tightness, risk appetite, staging status, etc)
Inst. Seller	Sellers who are not individual sellers (i.e. PropTech firms, REITs, etc)
Inst. Buyer	Buyers who are not individual buyers (i.e. PropTech firms, REITs, etc)
Non-Local Buyer	Buyers who reside out of states at the time of purchase

Notes: This table summarizes the definitions of key variables used in the analysis.

Table A2 Zestimate vs. Predicted Value

All	**obs**	**Squared Error**	**MSE**
Zestimate	108,049	20,000	0.19
Predicted Value	108,049	5,705	0.05
Bedrooms $\leq$ 3	**obs**	**Squared Error**	**MSE**
Zestimate	44,595	9,880	0.22
Predicted Value	44,595	2,814	0.06
3 < Bedrooms	**obs**	**Squared Error**	**MSE**
Zestimate	63,454	10,120	0.16
Predicted Value	63,454	2,891	0.05
0 < Property Age $\leq$ 15	**obs**	**Squared Error**	**MSE**
Zestimate	30,669	3,434	0.11
Predicted Value	30,669	2,088	0.07
16 < Property Age $\leq$ 30	**obs**	**Squared Error**	**MSE**
Zestimate	40,513	4,170	0.10
Predicted Value	40,513	1,016	0.03
30 < Property Age	**obs**	**Squared Error**	**MSE**
Zestimate	36,867	12,396	0.34
Predicted Value	36,867	2,602	0.07
Sales Price $\leq$ 25,000	**obs**	**Squared Error**	**MSE**
Zestimate	60,620	8,317	0.14
Predicted Value	60,620	3,775	0.06
250,000 < Sales Price $\leq$ 500,000	**obs**	**Squared Error**	**MSE**
Zestimate	36,254	6,219	0.17
Predicted Value	36,254	1,007	0.03
500,000 $\leq$ Sales Price	**obs**	**Squared Error**	**MSE**
Zestimate	11,175	5,464	0.49
Predicted Value	11,175	924	0.08

Notes: This table shows the mean squared errors of Zestimate one month prior to listing month and the predicted value that I estimate using the hedonic model. Errors are estimated by the difference between the log of each values and the log of actual selling price. Squared error represents the sum of the squared errors and MSE indicates the mean squared error which is squared errors divided by the number of observations. This table suggests that the predicted values are statistically better estimated than Zestimates.

Table A3 Heterogeneity of Z-Bias

	(1)
Dep. Variable: Z-Bias$_{t-1}$	
Num. of Bed	0.0052*
	(0.003)
Num. of Bath	-0.0407***
	(0.002)
ln(BldgSize)	0.0222***
	(0.007)
Property Age	0.0024***
	(0.000)
ZipCode × Year × Quarter FE	yes
Year × Month FE	yes
Observations	90,308
R-squared	0.673

Notes: This table shows the heterogeneity of Z-Bias based on sales price and property characteristics. Standard errors are clustered at zip code × year level. Standard errors are in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

Table A4 Zestimate effects on Pricing Bias

Dep. Variable: PricingBias_t	(1)	(2)	(3)	(4)
Z-Bias_{t-1}	0.8868***	0.7771***	0.7644***	0.7675***
	(0.002)	(0.014)	(0.013)	(0.014)
Property Characteristics	no	no	yes	yes
ZipCode × Year × Quarter FE	no	yes	no	yes
Year × Month FE	yes	yes	yes	yes
ZipCode FE	no	no	yes	no
Cluster	no	yes	yes	yes
Observations	89,403	89,332	89,377	89,333
R-squared	0.795	0.811	0.810	0.813

Notes: This table shows the baseline results for the Z-Bias effects on pricing bias. Instead of using the past 12 months transaction from the time of sales, I use the 12 months transaction from the time of listing with one month lagged to estimate the predicted value. The dependent variable, $PricingBias_t$, is measured by the difference between the log of the listing price and the log of the predicted value. $Z \cdot Bias_{t-1}$ is calculated by the difference between the log of Zesitmate one month prior to listing and the log of the predicted value. Each column includes different controls and fixed effects. Column (1) includes year-month fixed effect only. Column (2) includes zip code by year-month and year-month fixed effects and standard errors are clustered at zip code by year level. Column (3) adds zip code fixed effects and property characteristics controls from Column (1). Column (4) additionally controls for property characteristics from Column (2). Standard errors are clustered at zip code by year level in Column (3) and (4). Standard errors are in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

Table A5 Pricing Bias effects on Sales Outcomes

Dep. Variables:	(1) ln(Time on Market)	(2) ln(Sales Price/ Predicted Value)	(3) ln(Sales Price/ Listing Price)
$\widehat{PricingBias}_{Zestimate}$	0.128***	0.029***	-0.016***
	(0.0085)	(0.0002)	(0.0022)
$\widehat{PricingBias}_{Residuals}$	0.158***	0.012***	-0.041***
	(0.0051)	(0.0002)	(0.0022)
Property Characteristics	yes	yes	yes
ZipCode × Year × Quarter FE	yes	yes	yes
Year × Month FE	yes	yes	yes
Observations	89,333	89,333	89,333
R-squared	0.160	0.905	0.154

Notes: This table shows the results when I use the predicted value that is measured by the time of listing instead of sales. Instead of using the past 12 months transaction from the time of sales, I use the 12 months transaction from the time of listing with one month lagged to estimate the predicted value. After diving pricing bias into two segments, pricing bias from Zestimate ($\widehat{PricingBias}_{Zestimate}$) and pricing bias from others ($\widehat{PricingBias}_{Residuals}$), I test how these two segments affect sales outcomes. Dependent variables are sales outcomes such as the log of days on market, the log of sales price to predicted value and the log of sales price to original listing price. Key independent variables are pricing bias from Zestimate,($\widehat{PricingBias}_{Zestimate}$), which is the estimated value from equation (2) and its residuals,($\widehat{PricingBias}_{Residuals}$), that indicates pricing bias from other factors. Both values are standardized. Property characteristics such as number of bedrooms, number of bathrooms, age, age squared, and the log of building size in square feet are included. Standard errors are clustered at zip code × year level. Standard errors are in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

Table A6 Z-Bias Effects by Internet access level

Dep. Variable: Pricing Bias_t	(1) Low Internet	(2) Mid Internet	(3) High Internet
Z-Bias_{t-1}	0.633***	0.797***	0.856***
	(0.0217)	(0.0159)	(0.0089)
Property Characteristics	yes	yes	yes
ZipCode × Year × Quarter FE	yes	yes	yes
Year × Month FE	yes	yes	yes
Observations	28,354	22,746	42,015
R-squared	0.567	0.861	0.828

Notes: This table shows that the impact of Zestimate bias on pricing bias by internet access level. Property characteristics such as number of bedrooms, number of bathrooms, age, age squared, and the log of building size in square feet are included. Standard errors are clustered at zip code× Year level. Standard errors are in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

Table A7 Post-listing Zestimate Effects on Sales Outcomes

Dep. Variables:	(1) ln(Time on Market)	(2) ln(Sales Price/ Predicted Value)	(3) ln(Sales Price/ Listing Price)
ln(ListingPrice/Zestimate)	0.022***	0.003***	-0.055***
	(0.0039)	(0.0002)	(0.0030)
Property Characteristics	yes	yes	yes
ZipCode × Year × Quarter FE	yes	yes	yes
Year × Month FE	yes	yes	yes
Observations	93,078	93,078	93,078
R-squared	0.103	0.207	0.595

Notes: This table shows that how Zestimate affects sales outcomes when properties are priced higher than Zestimate. Original listing price and Zestimate on listing month are used to measure the difference between the log of listing price and the log of Zestimate. The value of ln(Listing Price/Zestimate) is standardized. Property characteristics such as number of bedrooms, number of bathrooms, age, age squared, and the log of building size in square feet are included. Standard errors are clustered at zip code by year level. Standard errors are in parentheses; *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A8 Post-listing Zestimate Effects on Listing Price Change

Dep. Variables:	(1) I(Any Change)	(2) Price Chage (%)	(3) ln(Days to Change)	(4) I(Positive Change)	(5) I(Negative Change)
ln(ListingPrice/Zestimate)	0.021***	-0.156***	-0.248**	-0.365***	0.339***
	(0.0014)	(0.0142)	(0.1044)	(0.0438)	(0.0418)
Property Characteristics	yes	yes	yes	yes	yes
ZipCode × Year × Quarter FE	yes	yes	yes	yes	yes
Year × Month FE	yes	yes	yes	yes	yes
Observations	93,064	28,058	28,058	28,058	28,058
R-squared	0.082	0.186	0.122	0.143	0.140

Notes: This table shows that how Zestimate affects listing price change behaviros when properties are priced higher than Zestimate. Original listing price and Zestimate on listing month are used to measure the difference between the log of listing price and the log of Zestimate. The value of ln(Listing Price/Zestimate) is standardized. Property characteristics such as number of bedrooms, number of bathrooms, age, age squared, and the log of building size in square feet are included. Standard errors are clustered at zip code by year level. Standard errors are in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

Table A9 Individual vs. Institutional Buyers

Panel A. Matching Outcomes

	Ind. Buyer	Inst. Buyer	t-test
Bed	3.59	3.59	0 (-0.04)
Bath	2.63	2.59	0.04 (2.06**)
Property Age	31.32	31.8	-0.48 (-1.51)
ln(BldgSize)	7.57	7.58	-0.01 (-0.41)
Zestimate($)	207,405	208,805	-1,400 (-0.43)

Panel B. Z-Bias Effects on Pricing Bias

Dep. Variable: Pricing Bias$_t$	(1) Ind. Buyer	(2) Inst. Buyer	(3) Both
Z-Bias$_{t-1}$	0.711***	0.726***	0.692***
	(0.0235)	(0.0195)	(0.0193)
Z-Bias$_{t-1}\times$ Inst. Buyer			0.050
			(0.0146)
Property Characteristics	yes	yes	yes
Pair FE	no	no	yes
ZipCode × Year × Quarter FE	yes	yes	yes
Year × Month FE	yes	yes	yes
Observations	7,049	7,049	14,098
R-squared	0.871	0.818	0.836

Panel C. Mispricing Effects on Sales Outcomes

Dep. Variables:	(1) ln(Time on Market)	(2) ln(Sales Price /Predicted Value)	(3) ln(Sales Price /Listing Price)
$\widehat{PricingBias}_{Zestimate}$	-0.006	0.030***	-0.009*
	(0.0221)	(0.0005)	(0.0052)
$\widehat{PricingBias}_{Zestimate}\times$ Inst. Buyer	-0.010	0.001	0.005
	(0.0194)	(0.0004)	(0.0044)
$\widehat{PricingBias}_{Residuals}$	0.113***	0.012***	-0.044***
	(0.0158)	(0.0006)	(0.0071)
$\widehat{PricingBias}_{Residuals}\times$ Inst. Buyer	0.024	-0.000	-0.001
	(0.0175)	(0.0008)	(0.0089)
Property Characteristics	yes	yes	yes
Pair FE	yes	yes	yes
ZipCode × Year × Quarter FE	yes	yes	yes
Year × Month FE	yes	yes	yes
Observations	14,098	14,098	14,098
R-squared	0.170	0.882	0.227

Notes: This table shows that how Zestimate is utilized differently by buyer's type. Insitutional buyers are used as a proxy for informed buyers. Panel A shows the t-test results from matching process. Columns called t-test show the difference in values between two groups and statistical results are shown in parenthesis. I use 1:1 propensity score matching with 0.01 caliper without replacement. Zestimate one month prior to sales is used in the matching process. Inst.Buyer indicates whether a buyer is an institutional buyer or not. Panel B describes the results of Z-Bias effects on pricing bias while Panel C shows the results of mispricing effects on sales outcomes. Mispricing bias is measured by the difference between the log of final asking price and predicted value. Z-Bias is measured by the difference between the log of Zestimate one month prior to sales date and the log of predicted value. Property characteristics such as number of bedrooms, number of bathrooms, age, age squared, and the log of building size in square feet are included. Standard errors are clustered at zip code by year level. Standard errors are in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

Table A10 Local vs. Non-Local Buyers

Panel A. Matching Outcomes

	Local Buyer	Non-Local Buyer	t-test
Bed	3.68	3.69	-0.01 (-0.41)
Bath	2.75	2.74	0.01 (0.56)
Property Age	26.76	26.74	0.02 (0.07)
ln(BldgSize)	7.64	7.65	-0.01 (-1.09)
Zestimate($)	218,381	219,366	-985 (-0.28)

Panel B. Z-Bias Effects on Pricing Bias

Dep. Variable: Pricing Bias$_t$	(1) Local Buyer	(2) Non-Local Buyer	(3) Both
Z-Bias$_{t-1}$	0.688***	0.780***	0.686***
	(0.0276)	(0.0462)	(0.0250)
Z-Bias$_{t-1}\times$ Non-Local Buyer			0.099***
			(0.0283)
Property Characteristics	yes	yes	yes
Pair FE	no	no	yes
ZipCode $\times$ Year $\times$ Quarter FE	yes	yes	yes
Year $\times$ Month FE	yes	yes	yes
Observations	4,497	4,497	8,994
R-squared	0.789	0.780	0.763

Panel C. Mispricing Effects on Sales Outcomes

Dep. Variables:	(1) ln(Time on Market)	(2) ln(Sales Price /Predicted Value)	(3) ln(Sales Price /Listing Price)
$\widehat{PricingBias}_{Zestimate}$	-0.031	0.029***	-0.012**
	(0.0234)	(0.0005)	(0.0056)
$\widehat{PricingBias}_{Zestimate}\times$ Non-Local Buyer	0.061***	0.001**	0.013**
	(0.0230)	(0.0005)	(0.0057)
$\widehat{PricingBias}_{Residuals}$	0.148***	0.012***	-0.042***
	(0.0150)	(0.0008)	(0.0095)
$\widehat{PricingBias}_{Residuals}\times$ Non-Local Buyer	-0.049***	-0.001	-0.008
	(0.0184)	(0.0011)	(0.0128)
Property Characteristics	yes	yes	yes
Pair FE	yes	yes	yes
ZipCode $\times$ Year $\times$ Quarter FE	yes	yes	yes
Year $\times$ Month FE	yes	yes	yes
Observations	8,994	8,994	8,994
R-squared	0.207	0.902	0.279

Notes: This table shows that how Zestimate is utilized differently by buyer's geographic location. Out of state buyers at closing are used as a proxy for non-local buyers who are less informed compared to buyers from the same state. Panel A shows the t-test results from matching process. Columns called t-test show the difference in values between two groups and statistical results are shown in parenthesis. I use 1:1 propensity score matching with 0.01 caliper without replacement. Zestimate one month prior to sales is used in the matching process. Non-local buyer indicates whether a buyer is from the same state or not. Panel B describes the results of Z-Bias effects on pricing bias while Panel C shows the results of mispricing effects on sales outcomes. Mispricing bias is measured by the difference between the log of final asking price and Predicted value while Z-Bias is measured by the difference between the log of Zestimate one month prior to sales date and the log of predicted value. Property characteristics such as number of bedrooms, number of bathrooms, age, age squared, and the log of building size in square feet are included. Standard errors are clustered at zip code by year level. Standard errors are in parentheses; *** p<0.01, ** p<0.05, * p<0.1.