How Does Zestimate Affect Housing Market Outcomes Across Socio-economic Segments?

JOB MARKET PAPER

PRELIMINARY DRAFT: PLEASE DO NOT SHARE OR POST

Abstract

We study the impact of Zillow's Zestimate on housing market outcomes and how the impact differs across socio-economic segments. Zestimate is produced by a machine learning algorithm using large amounts of data and aims to be an unbiased prediction of a home's market value at any time. Zestimate can potentially help market participants in the housing market as identifying the value of a home is a non-trivial task. However, inaccurate Zestimate could also lead to incorrect belief about property values and hinder the selling process. Meanwhile, Zestimate tends to be significantly more accurate for rich neighborhoods than poor neighborhoods, raising concerns that the benefits of Zestimate accrue largely to the rich, widening socio-economic inequality. Using data on Zestimate and housing sales in the United States, we show that Zestimate benefits the housing market as on average it increases both buyer welfare and seller welfare. Moreover, Zestimate actually reduces socio-economic inequality, as our results reveal that both rich and poor neighborhoods benefit from Zestimate but the poor neighborhoods benefit more.

We build a structural model of a housing market where sellers and buyers face uncertainty about property values. Zestimate provides a signal of the property value. Our model captures two potentially countervailing effects of Zestimate. First, it reduces uncertainty in beliefs about property values. Second, it shifts the mean belief about property values towards the Zestimate. Since Zestimate predicts the property value with some error, the reported Zestimate could under or over estimate the property value. Hence, the mean belief about the property value may be shifted away from the true property value.

We estimate our model on a unique data set consisting of 3,724 properties listed in Pittsburgh between February and October 2019. The estimation results reveal that people in poor neighborhoods are more uncertain about property values initially before learning from Zestimate compared with those in the mid-range and rich neighborhoods. In a counterfactual analysis, we show that, on average, the introduction of Zestimate increases seller profit by 7.53%, buyer surplus by 4.42%, and total surplus by 6.16%. We also find that although Zestimate under-estimates the selling price of more than 40% of the properties in our sample, only 19.04% of the properties in the sample have lower seller profit with Zestimate than without. This suggests that having an undervalued Zestimate may still be better for the seller than not having a Zestimate due to the benefit of uncertainty reduction. In addition, Zestimate leads to the greatest total welfare increase in poor neighborhoods (7.09%) despite the fact that Zestimates are least accurate in these areas. One important reason that Zestimates are less accurate in poor neighborhoods is a lack of accurate data on home facts, which homeowners can voluntarily provide to Zillow to improve accuracy. In another counterfactual analysis, we increase Zestimate accuracy in poor neighborhoods to the same level as in other neighborhoods, and find that the positive impact of Zestimate on total surplus would further increase by 31.17%.

1. Introduction

Machine learning algorithms are now used in our daily life to facilitate important decision making. They impact our lives in myriad ways including access to credit, education, jobs and various other areas (Kleinberg et al. 2018, Agrawal et al. 2019, Lambrecht and Tucker 2019, Hansen et al. 2021, Calvano et al. 2020, Fu et al. 2021, Zhang et al. 2021). A major strength of machine learning algorithms is their capability of analyzing large amount of data quickly and identifying patterns that humans may not be able to identify easily. Therefore, they tend to makes better predictions than humans in many applications (Kleinberg et al. 2018, Fu et al. 2021). In recent years, technology advancement has led to better data infrastructure and easier access to computational resources, which provide foundations for estimating and analyzing complex models required by real-life problems. As a result, machine learning applications are becoming increasingly popular.

The goal of this paper is to study how a machine learning pricing algorithm affects the housing market. Housing is usually the key component of household wealth, the major collateral for bank lending and has the most significant long term impact on trends in wealth to income ratios (Piketty and Zucman 2014). According to the 2006 American Community Survey, 37% percent of owners with mortgages and 16% of owners without mortgages spend 30% or more of their income on housing costs.[1] As housing transactions typically involve large financial amounts and significant selling time, the related decisions are made with extra caution. Despite its importance, the housing market is often considered inefficient for several reasons. First, houses are heterogeneous assets. Each house is unique in its features (e.g. structure, floor plan, build quality), amenities, and locations. Second, buyers are heterogeneous in their valuations of a property. Thus, sellers are uncertain about how potential buyers value their properties. Third, there are non-trivial market frictions, such as search costs and various transaction costs, including agent fees, taxes, and moving costs. These factors make accurately pricing a property difficult, leading to a housing market that is far from efficient.

[1] https://www.census.gov/housing/census/publications/who-can-afford.pdf

The emergence of algorithms that predict market value of properties have the potential to improve housing market efficiency. Several online real-estate marketplaces, such as Zillow, Trulia and Redfin, have proprietary algorithms that estimate property values. These marketplaces display their estimates for free on their websites. Leveraging large amounts of data on millions of properties and massive computational power, these sophisticated algorithms are able to produce reasonable estimates of property values. Since these estimates that signal property values are easily available to buyers and sellers alike, they should work to reduce uncertainty in the housing market and, consequently, make it more efficient. However, uncertainty reduction is only part of the potential effects. As it is difficult to achieve 100% accuracy, these algorithms generate estimates of property values with errors. Such inaccurate estimates may shift beliefs about property values away from the true values, lending certainty to incorrect belief held by buyers and sellers that could impede the selling process. There has been multiple news articles and online discussions on the inaccuracy of these algorithms and how they cause issues for both buyers and sellers in the market.[2] In 2017, a homeowner even sued Zillow, claiming that the company's pricing algorithm repeatedly undervalued her house, which was "tortuously interfering with [the homeowner's] market value and marketability of her home".[3]

In this paper, we specifically focus on the impact of Zillow's pricing algorithm. Zillow is the most popular real estate website in the United States by number of visits, with approximately 36 million unique monthly visitors,[4] or 27.2% of the market share of visits.[5] It was also the first website to publish algorithm-generated home value estimates for properties nationwide. These estimates are called "Zestimates" and are available for more than 100 million U.S. homes. The

[2] For a few examples, see:
https://www.nytimes.com/2018/09/14/realestate/why-zillow-addicts-cant-look-away.html,
https://www.forbes.com/sites/johnwake/2019/06/30/new-zillow-zestimate-accuracy/
https://www.reddit.com/r/RealEstate/comments/af414e/how_accurate_is_the_zestimate_number_do_you_put_a/.

[3] https://www.courthousenews.com/wp-content/uploads/2017/04/Zillow.pdf

[4] https://www.statista.com/statistics/381468/most-popular-real-estate-websites-by-monthly-visits-usa/

[5] https://ipropertymanagement.com/research/zillow-statistics

Zestimate algorithm intends to provide unbiased signals of the expected selling price of properties. While this claim of being unbiased may be true at the population level across properties, the reported Zestimates tends to over- or undervalue properties at the individual level. Even though the Zestimate of a given property fluctuates from time to time, the value in one period is highly correlated with the value in previous periods. Thus, if a Zestimate undervalues (overvalues) a property in one period, it is likely to continue to undervalue (overvalue) that property in the future. According to Zillow, for on-market properties, 82.2% of Zestimates are within 5% of selling price, 95.5% of Zestimates are within 10% of selling price, and 98.8% of Zestimates are within 20% of selling price.[6] Both undervalued and overvalued Zestimate could be problematic, as an undervalued Zestimate could lead to lower belief in the property value and may result in a lower selling price, while an overvalued Zestimate could lead to incorrectly high expectations of the property value and result in inefficient search and a longer selling time. Thus, despite the fact that Zestimate can reduce uncertainty in the housing market, it is unclear how Zestimate affects buyer and seller welfare. This motivates our first objective, which is to examine how Zestimate affects the housing market in terms of market outcomes, including listing price, sales price, time on market, buyer welfare and seller welfare.

The second objective of this paper is to examine how the impact of Zestimate differs across neighborhoods. Housing is usually the largest asset of a household's portfolio, but it tends to account for a larger portion of family wealth for the poor than the rich. According to a study by Apartment List, [7] "those in the lowest [income] quartile make only 27% as much as the median household, but they still need to pay 79% of what the median household does each month to housing." The study also shows that the share of income spent on housing among homeowners steadily increases from around 10% to about 90% as we move from the highest income group to the lowest income group. Arguably housing matters more for the poor than for the rich. However,

[6] https://www.zillow.com/z/zestimate/

[7] https://www.apartmentlist.com/research/housing-markets-and-income-inequality

we notice that Zestimate is less accurate in poor neighborhoods than in other neighborhoods. Zillow measures Zestimate error by the percentage difference between the Zestimate and the selling price,[8] and we follow Zillow's approach to calculate Zestimate errors for each property in our sample. We divide the neighborhoods in our sample into three groups – poor neighborhoods, mid-range neighborhoods, and rich neighborhoods, and find that Zestimate error has the largest spread in poor neighborhoods, suggesting that Zestimate is less accurate in poor neighborhoods.

The Zestimate algorithm relies on finding comparable properties based on property features to generate an estimate of a property's market value. Zillow obtains the data on property features from public records and user-submitted information. If a property feature value is missing for a property on Zillow, then the Zestimate algorithm inevitably finds comparable properties that differ from the focal property on this feature. In this case, the comparable properties are less comparable to the focal property, and the Zestimate value generated based on these comparable properties would be less accurate. We found that on average fewer property features are available in poor neighborhoods, either because homeowners do not actively provide their property features on Zillow or because public records are less well-maintained, which could contribute to the lower Zestimate accuracy in these areas. Other possible causes include lower data accuracy and lower transaction volume in poor neighborhoods.

All other things being equal, less accurate Zestimate would lead to more harm or less benefit, because it provides a more erroneous signal and reduces uncertainty to a less extent. However, it is important to note that the impact of Zestimate also depends on the prior uncertainty about property values in absence of Zestimate, and it is unclear how poor neighborhoods and rich neighborhoods differ in the prior uncertainty. On the one hand, poor people tend to get less education and have less access to high-quality information or competent sales agents to facilitate their decisions. As a result, they may face greater uncertainty over property values. On the other hand, properties in

[8] Note that selling price is endogenous and may be affected by Zestimate, therefore this is a rough measures of Zestimate accuracy. We will employ other measure of Zestimate accuracy later in the paper.

poor neighborhoods are more standardized as opposed to properties in rich neighborhoods that usually have unique features. Therefore, the inherent uncertainty about property values may be lower.

In summary, the impact of Zestimate on poor neighborhoods is ambiguous at a conceptual level. On the one hand, Zestimate may hurt poor neighborhoods as compared with other neighborhoods, because Zestimates are less accurate in poor neighborhoods. On the other hand, the prior uncertainty may differ in poor neighborhoods and other neighborhoods, and Zestimate could benefit the poor neighborhoods more if sellers and buyers in poor neighborhoods face greater uncertainty in absence of Zestimate. Resolution of this ambiguity demands empirical testing and motivates our second objective to identify how and to what extent Zestimate affects social inequality in terms of buyer welfare and seller welfare.

Recall that Zestimate has two potential effects: first, it reduces uncertainty about a property's value; second, it shift the belief about about a property's value. To tease out these two effects and evaluate the impact of different policy simulations on buyer and seller welfare, we build a structural model of the housing market.

We model the home selling process as a discrete-time, infinite horizon problem. A homeowner lists her property on the market and becomes a seller. As the seller lists the property, she sets a listing price, which is revealed to all the potential buyers, as well as a reservation price, which is the lowest offer value that the seller is willing to accept. In the subsequent periods while the property stays on the market, the seller can update the listing price and the reservation price at the beginning of each period. The potential buyers have heterogeneous valuations of the property because of their idiosyncratic tastes. In each period, one potential buyer decides whether to visit the property or not based on his current belief of the property's value. If the buyer decides to visit the property, then he incurs a visiting cost and realizes his own true valuation of the property after the visit. Next, the buyer and the seller engage in a bargaining process, in which the buyer reveals his valuation and the seller reveals her reservation price. If the buyer's valuation is higher than the seller's reservation price, then the property is sold. The final selling price is a bargaining outcome, which depends on the buyer's relative bargaining power, unless the bargaining outcome is higher than the listing price, in which case the property is sold at the listing price. If the buyer's valuation is lower than the seller's reservation price, then the buyer walks away while the seller stays on market until next period.

When a buyer walks away, the remaining buyers and seller go through an updating phase where they reassess their value of the property. To understand the nature of this updating, we need to set how information is generated and updated in the model. A property's worth is reflected by how potential buyers value it. While potential buyers have heterogeneous valuations of a property, the mean of true buyer valuations (i.e., valuations that would be revealed to them if they visited the property) can be viewed as a measure of the property value. With a slightly abuse of the term, we call the mean of true buyer valuations of a property the "true value" of the property. The buyers and the seller are uncertain about this true value of a property, and they form a belief about it based on observed characteristics and population level market trends. As a property stays on the market, the buyers and the seller gradually learn about the true value and update their beliefs. When a property is not sold after a buyer's visit, the seller and future potential buyers observe this signal and update their beliefs about the true value of the property. With these updated beliefs about true property values, buyers and sellers make decisions in subsequent rounds to maximize their expected utilities.

Now, we turn to the impact of Zestimate on this basic model. As mentioned above, Zestimate is an estimate of a property's market value, or the expected selling price. Note that the buyers and seller are uncertain about the expected selling price only because they are uncertain about the true value of the property, as they observe all the property features and know the entire selling process. Therefore, Zestimate effectively provides a signal of the true property value. Buyers and sellers update their beliefs about the true property value based on the Zestimate signal. By modeling Zestimate as a signal that updates beliefs about the true property value under a Bayesian learning framework, we capture the two effects of Zestimate: first, it reduces uncertainty in the belief about the true value; second, it shifts the mean belief about the true value.

We estimate our model using data on properties in Pittsburgh that were listed between February and October 2019. There are 3,724 properties in total, among which 1,080 properties are in poor neighborhoods, 1,681 properties are in mid-range neighborhoods, and 963 properties are in rich neighborhoods. For each property, we observe its Zestimate, Zestimate range, and Zestimate value history at roughly two-weeks interval. We also observe each property's on-market activity, including listing time, initial listing price, listing price updates, selling time, and final selling price. In addition, we observe a rich set of property features, including detailed location information, property structure, size, number of bedrooms, number of bathrooms, flooring, appliances and many other home characteristics.

The Key Findings

Our estimation results show that the average noise to information ratio[9] of Zestimate is 4.9303, which suggests that Zestimate has non-trivial effect on buyers' and sellers' beliefs about property values. In addition, the variance of the prior belief about true value in the absence of Zestimate is larger in poor neighborhoods than in other neighborhoods, which suggests that people in poor neighborhoods face greater uncertainty about property value in the absence of Zestimate, and therefore could potentially benefit more from an accurate signal of property value.

To examine the effect of Zestimate, in the first of our counterfactual analyses, we calculate the market outcomes and welfare when Zestimate is removed, and compare them with those in the case where Zestimate is included. We find that on average across all properties Zestimate leads to overall listing prices, higher selling prices, longer times on market, and more importantly, higher buyer welfare and higher seller welfare. Specifically, Zestimate leads to an average of 4.42% increase in buyer surplus, 7.53% increase in seller profit, and 6.16% increase in total surplus.

[9] The variance of the signal divided by the prior variance of the beliefs

Even though Zestimate underestimates the final sales price for about 40% of the properties, it only leads to a seller welfare loss for 19.04% of the properties. Moreover, while Zestimate is least accurate in poor neighborhoods, it leads to the greatest average total welfare increase in poor neighborhoods – 7.09% as compared to 5.67% and 5.96% in mid-range neighborhoods and rich neighborhoods, respectively. Specifically, the average seller profit increases in poor, mid-range, and rich neighborhoods are 8.58%, 6.89% and 7.47%, respectively, and the average buyer surplus increases in poor, mid-range, and rich neighborhoods are 4.75%, 4.36% and 4.15%, respectively.

Although our results show that the current implementation of Zestimate already leads to greater total welfare increase in poor neighborhoods (as compared to other neighborhoods), it is also clear that more accurate Zestimates could benefit poor neighborhoods more. Buyers and sellers in poor neighborhoods are not reaping the full benefits of the Zestimate algorithm. As discussed before, there are multiple reasons why Zestimate is less accurate in poor neighborhoods, and one important reason is poor input data quality. Zillow allows homeowners to provide or update home characteristics information to improve Zestimate accuracy,[10] yet we find on average fewer home characteristics are available for properties in poor neighborhoods. Therefore, homeowners could provide more information about their properties and benefit from a more accurate Zestimate. We undertake a second counterfactual analysis to evaluate how much poor neighborhoods are losing because of less accurate Zestimates. We increase Zestimate accuracy in poor neighborhoods up to the Zestimate accuracy in mid-range neighborhoods, which has the highest Zestimate accuracy among the three neighborhood groups. The results show that if, on average, Zestimates in poor neighborhoods were as accurate as Zestimates in mid-range neighborhoods, the average total surplus change would be 9.30%. Compared to the total welfare change of 7.09% with current Zestimate accuracy in poor neighborhoods, this suggests the potential positive impact of Zestimate on total surplus in poor neighborhoods could further increase by 31.17% with higher accuracy.

[10] https://www.zillow.com/sellerlanding/edityourhome/

Understanding the Mechanism

To understand the results, let us examine how Zestiamte impacts the housing market. Zestimate, as an estimate of the market value of the property, provides a signal of the true property value and influences the beliefs of both buyers and sellers. It has two effects on the beliefs about true property value: 1) it reduces the variance (uncertainty) of the belief about the true value; 2) it shifts the mean belief. We next discuss these effects one by one.

We first discuss the effect of variance (uncertainty) reduction on the seller's listing price. First note that the listing price is a commitment device that sellers use to attract buyers to visit the property, as a buyer knows that the final selling price is capped at the listing price. That is, by setting a listing price, the seller encourages buyer visits by passing some of the surplus to the buyer when the buyer-revealed valuation is significantly greater than the listing price. This implies that the lower the listing price, the greater will be the buyer's propensity to visit the property. Next, note that the uncertainty in buyer's valuations is positively related to the buyer's propensity to visit the property, which allows the buyer to learn about its true value and see if it is worth purchasing. The greater the uncertainty in buyer's valuations, the greater will be the option value of search for the buyer. Why? This is because as the uncertainty increases, then from the buyer's perspective, the probability that her revealed valuation upon visit is either very high or very low will also increase. The very high and the very low draws have an asymmetric impact on the buyer's expected benefit of visit. The buyer's surplus will be higher if she receives a very high draw upon visit as compared to a high draw, since in either case she will buy the property, and the magnitude of her surplus will be positively related to the magnitude of her draw. On the other hand, the buyer's surplus will be the same regardless of whether she gets a low or a very low draw upon visit, since she will not buy the property in either case. This asymmetric impact results in the expected benefit of visit to increase with the increase in uncertainty.

It follows from the above discussion that all else being the same, if the buyer's uncertainty decreases because of the information provided by Zestimate, then the buyer's propensity to visit the property will also decrease. Now to offset this decrease in visit probability, the seller will decrease the listing price in order to encourage more buyer visits. This implies that the greater the uncertainty reduction because of Zestimate, the lower will be the listing price.

We next discuss the impact of uncertainty reduction on the seller's reservation price – where the reservation price in each period is the lowest price that the seller will accept in that period. The seller's reservation price in a given period is directly related to the expected future profits that the seller would make if the property does not get sold - the greater the expected future profit of the property if it is does not get sold, the greater will be the reservation price in the current period. In what follows, we will explain why an increase in buyer's uncertainty will lead to a decrease in the expected future profits of the seller, and thereby a decrease in the reservation price.

If the property does not get sold in the present time period, it leads to a decrease in future buyers' valuations of the property. This is because if a property is not sold in the current period, it signals to the future buyers that the true value of the property belongs to the lower part of the distribution their beliefs. This decrease in valuations of future buyers will lead to lower future expected profits for the seller since the seller would have to decrease the listing price in the future. Note that the larger the variance of the buyers' beliefs about the true property value, the greater will be decrease in the valuations of future buyers if the property is not sold. Why? This is because high uncertainty in valuations implies that from the buyer's perspective, the true property value could come from a broad range of values. Thus if the property is not sold, then from the future buyers' perspective, there is a reasonable chance that the true value of property might be really low, which then leads to a large reduction in their valuations. This will lead to lower expected future profits, which then leads to a lower reservation price.

Thus, we see that the greater the uncertainty in the buyer's valuation, the lower will be the reservation price. In a similar vein, when the buyer has lower uncertainty in her beliefs about the true property value, the seller will have a higher reservation price. This implies that the greater the uncertainty reduction because of Zestimate, the higher will be the seller's reservation price.

We next discuss the impact of uncertainty reduction by Zestimate on the time on market. This follows directly from the impact of uncertainty reduction on reservation price. Since the actual draws of buyer valuations are not affected by the beliefs about the true value, a higher reservation price will imply that the property is less likely to be sold and will be on the market for a longer time. This implies that the greater the uncertainty reduction because of Zestimate, the greater will be the time on the market.

Finally, we discuss the impact of uncertainty reduction by Zestimate on the final selling price. This follows directly from the impact of uncertainty reduction on the listing price and the reservation price. Note that the final selling price will be a convex combination of the listing price and the reservation price, as determined by the relative bargaining power of the buyer and the seller. Since the decrease in uncertainty by Zestimate leads to an increase in the reservation price and a decrease in the listing price, it implies that the impact of uncertainty reduction on the selling price can go either way. In our case, we find that the uncertainty reduction has a non-significant effect on selling price.

The impact of the shift in the mean beliefs of the buyers' valuations (because of Zestimate) on the prices and the time on market is straightforward. If Zestimate increases the mean beliefs of the buyer's valuations, the seller would set a higher listing price and would have a higher reservation price, in anticipation of higher buyer valuations. Using the same logic as before, a higher reservation price will lead to longer time on market, and a higher reservation price and listing price will lead to higher selling price. On the other hand, if Zestimate decreases the mean belief of the buyer's valuations, then the seller would set a lower listing price and a lower reservation price, which will lead to shorter time on market and a lower selling price.

Next, we move to the discussion of the impact on welfare. Seller welfare is the difference between the selling price and the true value of the property, discounted by the time on market; buyer welfare is the difference between the valuation of the buyer who purchases the property, minus the total visiting costs of all buyers that have visited the property; and the total welfare is simply the sum of seller welfare and buyer welfare. Since the seller does not heavily discount future and the true value of the property is fixed, the total welfare is largely determined by the valuation of the buyer who purchases the property and the total visiting costs. If the property can be sold to a buyer that truly values the property without much more wasted visiting effort, then the total welfare would be high. In our case, the variance reduction effect and the mean shift effect when Zestimate increases the mean belief both lead to higher reservation price, which result in the property being sold to a buyer with higher valuation with minor increase in visiting cost, and therefore increase the total welfare. In contrast, the mean shift effect when Zestimate decreases the mean belief leads to lower reservation price and the property being sold to buyers with relatively low valuation, and therefore decrease the total welfare. In other words, the variance reduction effect increases total welfare for all the properties, and the mean shift effect increases total welfare for some properties while decreases total welfare for other properties. Therefore, Zestimate overall increases total welfare in the market, and even properties with under-valued Zestimate can still benefit from Zestimate because of the uncertainty reduction effect.

Seller welfare is largely determined by the final selling price. If a property is sold at a higher price within reasonable time, then the seller welfare would be higher. In our case, the variance reduction effect and the mean shift effect when Zestimate increases the mean belief would lead to higher selling prices, and therefore would increase seller welfare. The mean shift effect when Zestimate decreases the mean belief would lead to lower selling price, and therefore would decrease seller welfare. Combining these effects, Zestimate increases seller welfare on average across properties. Buyer welfare has two parts: the utility of the buyer who purchases the property, and the total visiting costs. As discussed before, the variance reduction effect and the mean shift effect when Zestimate increases the mean belief both result in the property being sold the a buyer with higher valuation with an increase in total visiting efforts. Our results suggest that the increase in the utility of the buyer who purchases the property is higher than the increase in the total visiting costs, and therefore buyer welfare increases. In contrast, the mean shift effect when Zestimate decrease the mean belief would lead to an decrease in buyer welfare. Combining these effects, Zestimate increases buyer welfare on average across properties.

Compared to people in other neighborhoods, people in poor neighborhoods face greater uncertainty about property values to begin with, probably because people in poor neighborhoods are generally less educated, and have less access to high quality information or agents to help them make decisions. Therefore, although Zestimate is less accurate in poor neighborhoods, it actually has larger impacts, and therefore leads to higher increases in both seller profit and buyer surplus in poor neighborhoods.

Related Literature

This paper is related to four streams of literature. The first is the literature on machine learning and human decision making. As machine learning algorithms are increasingly used to facilitate decision making, an important question to ask is whether algorithms can improve human decisions. Kleinberg et al. (2018) showed that a machine learning model can improve bail decisions, as the model can reduce crime rates with no change in jailing rates or reduce jailing rates with no change in crime rates, and these benefits can be achieved with reduced racial disparities. Similar results have been reported in the context of crowd lending(Fu et al. 2021) and resume screening (Cowgill 2018). This paper adds to the literature by showing that a popular pricing algorithm can benefit the housing market and increase both seller welfare and buyer welfare.

The second is the literature on the impact of machine learning algorithms on social inequality. Arising with the popularity of machine learning algorithms is the concern about the potential disparate impacts these algorithms may create. Lambrecht and Tucker (2019) conducted a field test on a popular advertising platform and found that job advertisements in science, technology, engineering, and math (STEM) fields are shown more often to men than to women, and the difference is largest for individuals in their prime career years. Obermeyer et al. (2019) found significant racial bias in a commercial prediction algorithm widely used in the healthcare industry, as black patients are considerably sicker than white patients in reality at the same level of predicted risk. Zhang et al. (2021) showed that Airbnb's smart-pricing algorithm reduced the revenue gap between white and black hosts conditional on the adoption, but black hosts were significantly less likely than white hosts to adopt the algorithm, so at the population level, the revenue gap increased after the introduction of the algorithm. This paper adds to the literature by showing that Zestimate may reduce socio-economic inequality as it leads to highest increase in total welfare in poor neighborhoods, despite of being least accurate in poor neighborhoods.

The third is the literature on housing market models and racial differentials in housing markets. Many models have been proposed to characterize housing market (Smith 1969, Wheaton 1990, Salant 1991, Merlo et al. 2015), and several papers have shown that blacks pay higher price than white for equivalent units in housing market (King and Mieszkowski 1973, Ihlanfeldt and Mayock 2009, Straszheim 1974), and similar racial discrimination has been found in online short term rental marketplace for both consumers (Edelman et al. 2017, Cui et al. 2020) and hosts (Zhang et al. 2021). Lu (2019) and Yu (2021) both show Zestimate might reduce racial disparities in housing market by providing less biased information. This paper adds to the literature by showing that Zestimate might reduce inequality because people in poor neighborhoods face the greatest uncertainty and therefore could benefit most from a pricing algorithm, yet they are still missing out some of the benefit due to the less accurate estimates.

The last is the literature on Bayesian learning. The Bayesian updating framework has been widely used in modeling consumer learning, and learning models have contributed greatly to our understanding of consumer behavior (Erdem and Keane 1996, Erdem 1998, Mehta et al. 2003, 2008, Erdem et al. 2008, Ching et al. 2013, Zhao et al. 2013, Huang et al. 2014) in marketing. In this paper, we apply the framework to model Zestimate as a signal that reduces uncertainty and shift mean belief. Different from most of the previous learning models where consumers learn towards true values from unbiased signals, in our case Zestimate signals are not unbiased at individual property level, and the under- and over- valuation of Zestimate can have systematic effects on housing market outcomes.

2. Model-free Evidence

This paper aims to examine the impact of Zillow's pricing algorithm. Zillow is the leading real estate website in the United States and the has been publishing algorithm generated home value estimate for individual properties, which is called "Zestimate", since 2006. Zestimate is currently available for 104 million properties in the United States, including both on-market properties and off-market properties. In this paper, we focus on on-market properties in Pittsburgh, Pennsylvania.

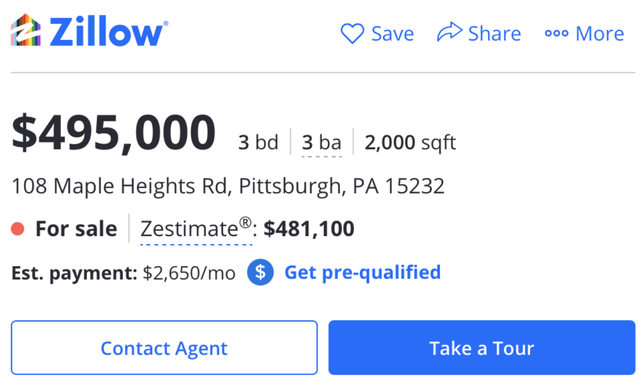

Figure 1 **Screenshot of the top of a property page on Zillow**

2.1. Data

Our sample consists of properties that were listed between February and October 2019 in Pittsburgh. There are 3,724 properties in total and they are spread across 140 different neighborhoods. Each property has an individual page on Zillow, where Zestimate is displayed on the top along with some other important information, including listing price, address, and size of the property, as shown in Figure 1. In addition, there is a section on the property that shows more detailed information about Zestimate, including Zestimate range and Zestimate history, as shown in Figure 2. While Zestimate is the estimated market value of a property, Zestimate range is the range in which the selling price is predicted to fall. According to Zillow, a wider Zestimate range "generally indicates a more uncertain Zestimate, which might be the result of unique home factors or less data avilable for the region or that particular home". Zillow also notes that "it's important to consider the size of the Estimated Sale Range (Zestimate Range) because it offers important context about the Zestimate's anticipated accuracy".[11] Zestimate history shows Zestimate values for a property in historical time points. It is important to note that these "historical Zestimates" are not necessarily the Zestimates that were displayed in the past. When Zillow updates the Zestimate algorithms, it may recalculate historical values using the the updated algorithms retroactively. In other words, these historical Zestimates are the estimated market values in the past based on the current Zestimate algorithm. While they are calculated based on the most updated algorithm, Zillow does not allow future information to influence a historical Zestimate, thus a sales in 2020 could not affect a historical Zestimate in 2019.

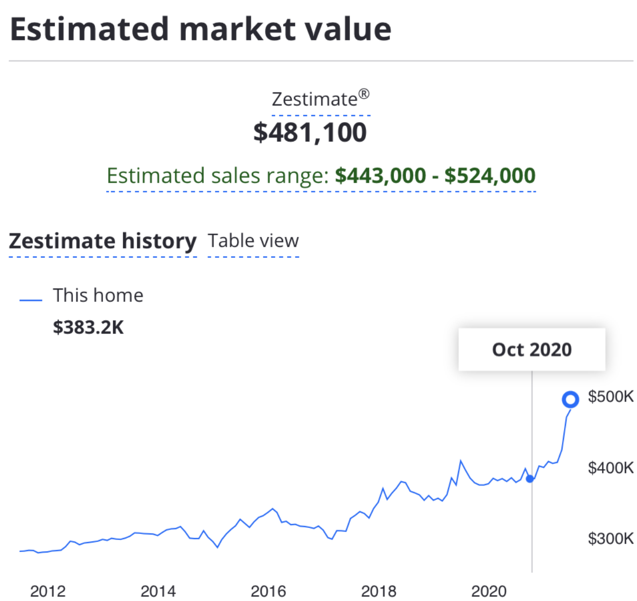

Figure 2 **Screenshot of the estimated market value section of a property page on Zillow**

For each property, we observe its Zestimate, Zestimate range and Zestimate history at roughly two-week intervals for the entire period when the property was on the market. We also observe its on-market activities, including listing time, initial listing price, listing price updates, selling time and final selling price. In addition, we observe a rich set of property features, including detailed

[11] https://www.zillow.com/z/zestimate/

location information, property structure, size, number of bedrooms, number of bathrooms, flooring, appliances, roof type, parking, basement and many other home characteristics. To examine the potential disparate impact of Zestimate across neighborhoods, we divide the 140 neighborhoods in our sample into three groups – poor neighborhoods, mid-range neighborhoods and rich neighborhoods – based on average property value in each neighborhood. Table 1 shows the summary statistics of our data.

2.2. Algorithm updates

Zestimate has been available for individual properties since 2006 when Zillow was launched. Over the years, Zillow has been working on improving the algorithm to generate more accurate Zestimate. According the published Zestimate accuracy statistics, the median error of Zestimate[12] has decreased from 6.9% in 2014 to 3.5% in 2019.[13] In 2018, Zillow launched a data science challenge called "Zillow Prize" on Kaggle to improve the Zestimate algorithm with over 1 million prize money.[14]. The winning ideas leverage neural networks and computer vision to incorporate image features and better identify patterns from real-time market data. These ideas eventually led to major Zestimate algorithm updates in 2019. During our observational period, there were two major algorithm updates, one in April 2019 and another in October 2019. These algorithm updates suddenly changed Zestimate values significantly for most of the properties without the anticipation from buyers or sellers. Moreover, these updates also changed historical Zestimate values, providing more accurate estimates of property market value in the past.

2.3. Reduced-form Analysis

In the section, we provide evidence of how Zestimate affects the observed market outcomes, such as listing price, selling price and time on market (TOM). The main challenge is that we do not observe the "true value" of a property, which could be a major confounding factor. For example,

[12] The error is computed by comparing the final selling price to the Zestimate that was published on or prior to the sale date.

[13] https://web.archive.org/web/*/https://www.zillow.com/zestimate/

[14] https://www.kaggle.com/c/zillow-prize-1

Table 1 Summary Statistics

Variable	Mean	Std. Dev	Min	Median	Max
Market Activities					
Listing Price	251323.57	145919.64	16000	210000	995000
selling price	235983.44	136551.06	16000	195000	965000
Time On Market when sold (in days)	47.14	59.91	1	23	300
Number of listing price update *	1.92	1.22	1	1	9
Ratio of selling price over last listing price	0.9674	0.0517	0.5948	0.9777	1
Ratio of selling price over initial listing price	0.9398	0.0789	0.4837	0.9600	1
Proportion of selling price = listing price	0.3541				
Proportion with listing price update	0.3549				
Zestimate Related					
Zestimate at listing time	242731.38	139783.09	20631	202704	996126
Ratio of Zestimate over selling price	1.0440	0.1270	0.4037	1.0190	2.0409
Ratio of Zestimate over initial listing price	0.9702	0.0687	0.3880	0.9785	1.6722
Key Home Features					
Number of bedrooms	3.07	0.97	0	3	8
Number of bathrooms	2.22	0.95	1	2	7
Floor size (in sqft)	1662.91	697.27	2	1498	12320
Basic Statistics					
Total number of properties	3724				
Number of neighborhoods	140				
Average selling price in poor neighborhoods	140551.39				
Average selling price in mid-range neighborhoods	215101.98				
Average selling price in rich neighborhoods	376210.21				

* among properties with price updates

if Zestimate is positively correlated with selling price, we do not know if Zestimate has positive impact on selling price or if it is only caused by a higher property value, which could be positively correlated with both Zestimate and selling price. Ideally, to evaluate the impact of Zestimate, we should focus on the extend to which Zestimate is deviated from the “true value” and examine the impact of such “Zestimate errors”. Zillow measures Zestimate errors using the difference between Zestimate and selling price. While this measure looks reasonable on the surface, it is problematic because selling price is endogenous and may be affected by Zestimate.

We overcome this challenge by leveraging the algorithm updates. As mentioned before, the algorithm updates changed not only the current Zestimate, but also the historical Zestimate values. That is, upon updating the algorithm, Zillow also recalculated the estimates of property values in the past, using the data available at that time and the updated (more accurate) algorithm. Therefore, for any time point before an algorithm update, we observe two difference Zestimate values: the original Zestimate that was shown at the time, and the updated Zestimate that became available with the launch of the new algorithm. The updated Zestimate is a more accurate estimate of market value at that time, and it would not affect any event that happened before the algorithm update since it became available only after the algorithm update. Therefore, we use updated Zestimate available after the algorithm update in October 2019 as a proxy for the true value of a property. For each property, we calculate the Zestimate error at the time when it was listed as follows

$$\text{ZestError} = \frac{\text{Original Zestimate} - \text{Updated Zestimate}}{\text{Updated Zestimate}}. \tag{1}$$

Table 2 Reduced-form Regression Results

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	ln(ListPrice)	ln(SoldPrice)	TOM	PriceUpdate	SoldAtListPrice	ln(ListPrice)	ln(SoldPrice)	ln(SoldPrice)
ZestError	0.326***	0.252***	0.860***	1.343**	-1.612**	0.530***	0.376***	0.427***
	(0.0161)	(0.0276)	(0.311)	(0.597)	(0.635)	(0.0265)	(0.0465)	(0.0613)
ZestRange						0.104***	0.0240	-0.109***
						(0.0145)	(0.0253)	(0.0359)
ZestError*ZestRange						-0.761***	-0.516***	-0.780***
						(0.0984)	(0.172)	(0.242)
UpdatedZestimate	0.982***	0.987***	0.713***	0.966***	-1.174*	0.984***	0.986***	1.005***
	(0.00495)	(0.00849)	(0.101)	(0.185)	(0.622)	(0.00490)	(0.00858)	(0.0104)
Constant	0.228***	0.211*	-3.905***	-10.87***	3.165***	0.197***	0.217*	-0.0411
	(0.0641)	(0.110)	(1.284)	(2.448)	(1.001)	(0.0635)	(0.111)	(0.148)
Observations	3,715	3,715	3,715	3,684	3,724	3,715	3,715	2,089
R-squared	0.988	0.966				0.989	0.966	0.972
home facts	YES	YES	YES	YES	YES	YES	YES	YES
neighborhood	YES	YES	YES	YES	YES	YES	YES	YES

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Next, we examine how Zestimate error affects the market outcomes. Table 2 shows the regression results. In all the regressions, we control for the updated Zestimate (as a proxy for property true value), a rich set of home facts (e.g. number of bedrooms, number of bathrooms, parking, flooring and lot size) and neighborhood fixed effects. Columns 1 - 3 in Table 2 show that as Zestimate error increases (which implies higher Zestimate conditional on property value), the initial listing price price is higher, the final selling price is higher, and the property stays on the market for longer time. We also examine the effect of Zestimate error on the probability of adjusting listing price while the property is on the market and on the probability of being sold at the listing price. Since more than 98% of the listing price adjustment in our sample is downward, the price adjustment is a signal of the initial listing price being too high. In contrast, a property being sold at the listing price suggest that eventually the buyer find the listing price reasonable or even undervalued, and therefore is willing to purchase the property at the listing price without further bargaining. Columns 4 - 5 show that as Zestimate error increases, the property is more likely to be over-priced, and in contrast, as Zestimate error decreases, the property is more likely to be under-priced.

Last, we examine the effect of Zestimate uncertainty. As mentioned before, Zestimate range indicates the uncertainty of Zestimate and provides information about the anticipated accuracy of Zestimate. Zestimate range includes a low estimated value and a high estimated value, and we calculate the following value as a measure of Zestimate uncertainty:

$$\text{ZestRange} = \frac{\text{High Estimated Value} - \text{Low Estimated Value}}{\text{Zestimate}}. \tag{2}$$

Column 6 - 7 show the impact of Zestimate uncertainty. First, Zestimate range moderates the effect of Zestimate error. As Zestimate range (uncertainty) increases, the effects of Zestimate error on listing price and selling price become weaker. Second, higher Zestimate range leads to higher listing price, but does not have significant impact on selling price. However, a significant fraction of the properties were sold at listing price in our sample, and if we exclude these properties and focus on properties sold below listing price, then higher Zestimate range leads to lower selling price among these properties.

The regression analysis provides evidence that Zestimate affects the initial listing price, the final selling price and the time on market. In addition, the effects of Zestimate are moderated by Zestimate uncertainty, and Zestimate uncertainty itself leads to higher listing price and lower selling price when a property is sold below listing price. Based on these evidences, we build a structural model of the housing market, which we present in the next section.

3. Model

3.1. Setup

We model the home-selling process as a discrete-time, infinite horizon problem, where each time period spans 2 weeks. When a homeowner decides to sell her property, she lists her property on the market and becomes a seller. As the seller lists the property, she sets a listing price, which is revealed to all the potential buyers, as well as a reservation price, which is the lowest offer value that the seller is willing to accept. In the subsequent periods while the property stays on the market, the seller can update the listing price and the reservation price at the beginning of each period. We denote the listing price of property i at time t as p^s_{it}, and the reservation price as r^s_{it}.

In each period t, a potential buyer arrives and views the listing. We refer to this potential buyer as "the buyer" hereafter. From the listing, the buyer observes a set of home characteristics as well as the listing price p^s_{it}, and decides whether to visit the property or not. Visiting incurs a search cost c_{it}, and buyers have heterogeneous search costs. We denote the visiting cost of the buyer in period t for property i as c_{it}, and

$$c_{it} \propto \mathcal{LN}(c_0, \sigma_0^2) \tag{3}$$

If the buyer chooses to visit the property, then his own valuation of the property, denoted as v_{it}, is revealed to him after him touring the property. We assume that the true distribution of buyer valuations for property i is

$$v_{it} \sim \mathcal{LN}(\lambda_i, \sigma_v^2) \tag{4}$$

Once the buyer realizes his own valuation of the property v_{it}, he engages in a bargaining process with the seller. We follow Chen and Rosenthal (1996) and assume the following bargaining rule based on the Nash bargaining solution (Nash 1950, Draganska et al. 2010, Zhang and Chung 2020): If the buyer's valuation v_{it} is greater than the seller's reservation price r^s_{it}, then a transaction happens and the selling price will be given as

$$p_i = \begin{cases} (1-\theta)r^s_{it} + \theta v_{it}, \quad \text{if } r^s_{it} \leq v_{it} \leq \frac{p^s_{it}-(1-\theta)r^s_{it}}{\theta}; \\ p^s_{it}, \quad \text{if } v_{it} > \frac{p^s_{it}-(1-\theta)r^s_{it}}{\theta}. \end{cases} \tag{5}$$

where θ is the bargaining power parameter. That is, the selling price is such that a fixed fraction θ of the total surplus goes to the seller and the remaining fraction $(1-\theta)$ goes to the buyer, unless this price would be higher than the listing price p_s, in which case the selling price equals to the listing price. If the buyer's valuation is lower than the seller's reservation price, then no transaction happens and the seller stays on the market until the next period.

A key component of our model is buyer and seller's uncertainty about property values. The expected value of buyer value's natural logarithm, λ_i, can be viewed as a measure of the value of the property i. The value of a property is determined by a rich set of property features (such as floor plans, build quality and location) and how buyers in market value these features. Buyers and sellers can observe all the property features, while the researchers only observe some of the features. We denote the part of property value driven by features that researchers (as well as buyers and sellers) observe as $\gamma\mathbf{X_i}$, where $\mathbf{X_i}$ is a vector of the observed features and γ is a vector of parameters. We further denote the part of property value driven by features that researchers do not observe (but buyers and sellers observe) as u_i.

Even though buyers and sellers observe all the features, they may be uncertain about the current market condition or how the market value the features, thus they may be uncertain about property value. We capture such uncertainty in another component of property value, denoted as α_i. Formally, λ_i can be written as

$$\lambda_i = \alpha_i + \gamma\mathbf{X_i} + u_i, \quad u_i \sim \mathcal{N}(0, \sigma_u), \tag{6}$$

where α_i captures the part of property value that is unobservable to buyers and sellers (as well as the researchers). While buyers and sellers do not know the value of α_i for each individual properties, they know the true distribution of α_i across all properties, which we denote as $\alpha_i \sim \mathcal{N}(\alpha_0, \sigma_0^2)$. Without additional information, they use the true distribution of α_i as the initial prior belief about α_i for all the properties.

Since α_i is a component of property value, the uncertainty about α_i is equivalent to the uncertainty about λ_i. For the ease of understanding, we will focus on the uncertainty about λ_i in the rest of the paper. We denote the belief about λ_i in period t as $\hat{\lambda}_{it} \sim \mathcal{N}(\mu_{it}, \sigma_{it}^2)$, and we have

$$\mu_{i1} = \alpha_0 + \beta \mathbf{X_i} + u_i, \tag{7}$$

$$\sigma_{i1}^2 = \sigma_0^2. \tag{8}$$

While a property stays on the market, the seller and the buyers update their belief about α_i based on signals from two sources: Zestimate and realized buyer valuations. Zestimate is an algorithm generated estimate of a property' market value, or the expected selling price. Apparently, the expected selling price is determined by the selling process and property values (λ_i). Since buyers and sellers are only uncertain about property values, the value of Zestimate is to provide a signal about property values. In other words, with the knowledge of the entire selling process, sellers and buyers are able to infer a signal of property value implied by a Zestimate. We denote Zestimate for property i at time t as Z_{it}, and the implied signal of property value as ζ_{it}. Since Zillow intends to provide Zestimate as an accurate estimate and and does not have incentive to systematically over- or under-valued a property, we assume that ζ_{it} is an unbiased signal about property value λ_i:

$$\zeta_{it} \sim \mathcal{LN}(\lambda_i, (\sigma_{it}^z)^2) \tag{9}$$

The mapping from Z_{it} to ζ_{it} is convoluted, and we approximate it by the following mapping:

$$\ln(\zeta_{it}) = a_z + b_z \ln(Z_{it}), \tag{10}$$

where a_z and b_z are two parameters to be estimated.

The standard deviation of Zestimate signal, σ^z_{it}, is not directly observed, but Zillow shows "Zestimate Range" along with Zestimate for each property. While Zestimate is the estimated market value, Zestimate range describes "the range in which a sale price is predicted to fall, including low and high estimated values".[15] We denote low estimated value and the high estimated values as Z^L_{it} and Z^H_{it}, respectively. As Zestimate range is a measure of Zestimate uncertainty, we use it to approximate σ_{zit} and assume the following:

$$\sigma^z_{it} = a_s + b_s \frac{Z^H_{it} - Z^L_{it}}{Z_{it}}. \tag{11}$$

As the seller initially lists a property on the market, both the seller and the buyers update their belief about λ_i with the Zestimate signal at the begining of the first period. Although Zestimate fluctuates from time to time, Zestimate in subsequent periods is highly correlated with Zestimate in previous periods and hardly provides any additional information. Thus, the seller and the buyers do not learn from Zestimate and update their belief about λ_i in the subsequent periods. The exception is when Zillow updates the algorithm and Zestimate changes significantly. In this case, the update Zestimate can be viewed as a new signal, and the seller and the buyers will learn from this new signal and update their belief at the beginning of the first period after the algorithm update. We denote the posterior belief about λ_i in period t after learning from Zestimate (if any) as $\hat{\lambda}^p_{it} \sim \mathcal{N}(\mu^p_{it}, (\sigma^p_{it})^2)$. Under the Bayesian updating framework, we have

$$\mu^p_{it} = \begin{cases} \mu_{it} + \frac{\sigma^2_{it}}{\sigma^2_{it} + (\sigma^z_{it})^2}(\ln(\zeta_{it}) - \mu_{it}), & \text{if } t = 1 \text{ or Zestimate algorithm updated;} \\ \mu_{it}, & \text{otherwise.} \end{cases} \tag{12}$$

$$(\sigma^p_{it})^2 = \begin{cases} (\frac{1}{\sigma^2_{it}} + \frac{1}{(\sigma^z_{zit})^2})^{-1} & \text{if } t = 1 \text{ or Zestimate algorithm updated;} \\ \sigma^2_{it}, & \text{otherwise.} \end{cases} \tag{13}$$

[15] https://www.zillow.com/z/zestimate/

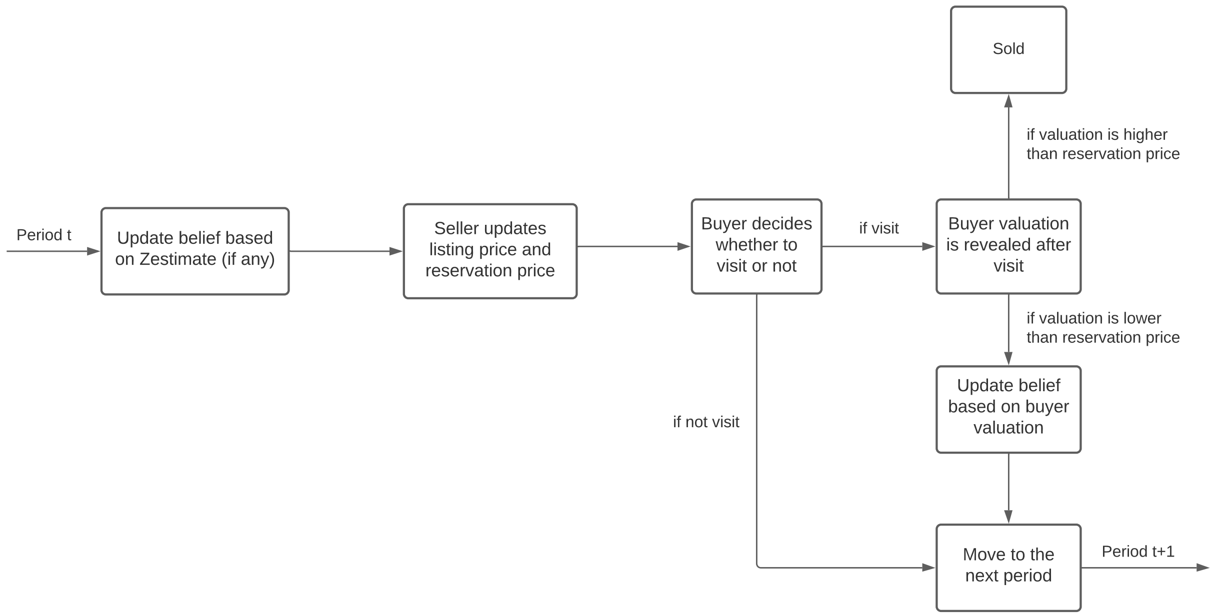

Figure 3 The Sequence of Events During a Period

Another source of signals is the realized buyer valuations. As mentioned before, if a buyer visits a property but realizes a valuation v_{it} that is lower than the seller's reservation price, then buyer will walk away and the seller will stay on the market until the next period. As the buyer engages in a bargaining after realizing his own valuation of the property, his valuation v_{it} is revealed to both the seller and the future buyers (through real estate agents) during the process. Thus, if a buyer visits the property and no transaction happens, then the seller and the future buyers update their belief about λ_i with the information of the buyer valuation v_{it} under the Bayesian updating framework, and the posterior belief becomes the prior belief about λ_i in time period $t+1$. If no buyer visits the property, then no additional information becomes available, and in the next period the new potential buyer and the seller hold the same prior belief about v_i as in this period. Formally,

$$\hat{\lambda}_i^{t+1} \sim \mathcal{N}(\mu_{i(t+1)}, \sigma^2_{i(t+1)}) \tag{14}$$

where

$$\mu_{i(t+1)} = \begin{cases} \mu^p_{it} + \frac{(\sigma^p_{it})^2}{(\sigma^p_{it})^2+\sigma^2_v}(\ln(v_{it}) - \mu^p_{it}), & \text{if a buyer visits in period } t; \\ \mu^p_{it}, & \text{otherwise.} \end{cases} \tag{15}$$

$$\sigma^2_{i(t+1)} = \begin{cases} (\frac{1}{(\sigma^p_{it})^2} + \frac{1}{\sigma^2_v})^{-1} & \text{if a buyer visits in period } t; \\ (\sigma^p_{it})^2, & \text{otherwise.} \end{cases} \tag{16}$$

Figure 3 shows the sequence of events in our model. In each period, buyers and sellers first update their belief based on Zestimate signals if this is the initial listing period or if Zillow updates the Zestimate algorithm. Based on the current belief, the seller chooses the listing price and the reservation price. Observing the listing price and the property characteristics, the buyer decides whether to visit the property or not. If the buyer does not visit the property, then the seller moves to the next period. Otherwise, the buyer's valuation is revealed after his visit. If the realized buyer valuation is higher than the seller's reservation price, then the property is sold; otherwise, the buyer walks away, and the seller updates her belief based on the buyer valuation and move to the next period.

3.2. Buyers' Problem

In this section, we characterize buyer decisions in the model. In each period, a buyer arrives and update his belief about λ_i with information of previous buyer valuations and Zestimate (if any). In the search stage, the buyer decides whether to visit the property or not. The buyer in this stage does not know the value of his own valuation, and there are two source of uncertainty for him about his valuation: the uncertainty about the λ_i and the uncertainty about his idiosyncratic taste. Since the buyer knows the distribution of idiosyncratic taste and has a belief about λ_i, the buyer's belief about his own valuation before visit is

$$\hat{v}_{it} \sim \mathcal{LN}(\mu^p_{it}, (\sigma^p_{it})^2 + \sigma^2_v) \tag{17}$$

The realized buyer valuation is the utility of owning the property for the buyer. From Equation 5, we know that the buyer will purchase the property at the listing price if

$$\hat{v}_{it} > \frac{p^s_{it} - (1-\theta)r^s_{it}}{\theta} \equiv \bar{v}_{it}, \tag{18}$$

and he will purchase the property at the weighted average of the seller's reservation price and his own valuation if his own valuation is higher than the seller's reservation price but lower than $\bar{v}_{it}$. If his valuation is lower than the seller's reservation price, the buyer will walk away and get zero utility. Therefore, the utility of visiting property i for the buyer in period t is:

$$\begin{aligned} u_{it} = &\int_{r_{it}^{s}}^{\bar{v}_{it}} (1-\theta)\cdot(\hat{v}_{it} - r_{it}^{s}) f(\hat{v}_{it})\, \mathrm{d}\hat{v}_{it} \\ &+ \int_{\bar{v}_{it}}^{\infty} (\hat{v}_{it} - p_{it}^{s}) f(\hat{v}_{it})\, \mathrm{d}\hat{v}_{it} \\ &- c_{it}, \end{aligned} \tag{19}$$

where c_{it} is the visiting cost.

The buyer will choose to visit the property if and only if the utility of visiting is positive. We denote buyer's visiting decision as $A_{it} \in \{0,1\}$, and denote the probability of buyer visit as q. Then

$$q = \Pr(A_{it} = 1) = \Pr(u_{it} > 0) = \Pr(c_t < \bar{u}_{it}), \tag{20}$$

where

$$\bar{u}_{it} = \int_{r_{it}^{s}}^{\bar{v}_{it}} (1-\theta)\cdot(\hat{v}_{it} - r_{it}^{s}) f(\hat{v}_{it})\, \mathrm{d}\hat{v}_{it} + \int_{\bar{v}_{it}}^{\infty} (\hat{v}_{it} - p_{it}^{s}) f(\hat{v}_{it})\, \mathrm{d}\hat{v}_{it}. \tag{21}$$

If the buyer chooses visit, then he will observe his own evaluation, v_{it}, which is a draw from the true valuation distribution $\sim \mathcal{LN}(\lambda_i, \sigma_v^2)$, after the visit. The buyer will purchase the property if and only if v_{it} is higher than the seller's reservation price, and the transaction price is given by Equation 5. Otherwise, the buyer walks away and no transaction happens.

3.3. Seller's Problem

We now move to the seller's decisions. In each period, after update her belief about λ_i with information of previous buyer valuations and Zestimate (if any), the seller decides on the listing price and the reservation price to maximize her ex ante profit. Apparently, the seller's profit crucially depends by buyers' valuations. Similar to a buyer in the search stage, the seller faces two source of uncertainty about a buyer's valuation: the uncertainty about λ_i and the uncertainty about the buyer's idiosyncratic taste. With the belief about λ_i and the knowledge about the distribution of idiosyncratic taste, the seller's belief about the distribution of buyer valuations is again

$$\hat{v}_{it} \sim \mathcal{LN}(\mu^p_{it}, (\sigma^p_{it})^2 + \sigma^2_v) \tag{22}$$

Let p^s_{it}, r^s_{it}, and $\pi_i(\mu_{it}, \sigma_{it})$ denote the listing price, the reservation price, and the seller's optimal ex ante profit function when the seller's belief about λ_i is $\mathcal{N}(\mu_{it}, \sigma^2_{it})$, respectively. There are four possible outcomes in each period. First, no buyer visits and no transaction happens, thus the seller stays on the market until the next period with the same belief about λ_i. Second, a buyer visits and realizes a valuation that is higher than $\bar{v}_{it} = \frac{p^s-(1-\theta)r^s}{\theta}$, thus the property is sold at the listing price. Third, a buyer visits and realizes a valuation that is lower than $\bar{v}$ but higher than the seller's reservation price, thus the property is sold at the weighted average of the buyer's valuation and seller's reservation price. Last, a buyer visits and realizes a valuation that is lower than the seller's reservation price, thus the property is not sold and the seller stays on the market until the next period. In this case, the seller will update her belief about λ_i with the buyer valuation. Let $\mu_{i(t+1)}(\hat{v}_{it})$ and $\sigma^2_{i(t+1)}$ denote the mean and the variance of posterior belief (which is the prior belief in the next period) when the buyer valuation is $\hat{v}_{it}$, we have

$$\mu_{i(t+1)}(\hat{v}_{it}) = \mu_{it} + \frac{\sigma^2_{it}}{\sigma^2_{it} + \sigma^2_v}(\hat{v}_{it} - \mu_{it}), \tag{23}$$

$$\sigma^2_{i(t+1)} = (\frac{1}{\sigma^2_{it}} + \frac{1}{\sigma^2_v})^{-1}. \tag{24}$$

Note that the seller chooses listing price and reservation price after observing Zestimate and updating the belief about λ_i, thus the impact of Zestimate is reflected in the state variables μ_{it} and σ_{it}. Moreover, the seller does not consider the impact of Zestimate in future periods, because she cannot anticipate Zestimate algorithm updates, and therefore does not expect any additional

Zestimate signal. Therefore, the seller's ex ante profit function is:

$$\begin{aligned}\pi_i(\mu_{it},\sigma_{it}) = \max_{p^s_{it},r^s_{it}}\{&(1-q)\cdot\beta\cdot\pi_i(\mu_{it},\sigma_{it})\\ &+q\int_{\bar{v}_{it}}^{\infty} p^s_{it}\cdot f(\hat{v}_{it})\,\mathrm{d}\hat{v}_{it}\\ &+q\int_{r^s_{it}}^{\bar{v}_{it}}[\theta\hat{v}_{it}+(1-\theta)r^s_{it}]f(\hat{v}_{it})\,\mathrm{d}\hat{v}_{it}\\ &+q\cdot\beta\int_{\infty}^{r^s_{it}}\pi_i(\mu_{i(t+1)}(\hat{v}_{it}),\sigma_{i(t+1)})f(\hat{v}_{it})\,\mathrm{d}\hat{v}_{it}\},\end{aligned} \tag{25}$$

where q is the probability of buyer visit derived in Equation 20 and it is a function of listing price and reservation price. Rearrange the profit function, we have

$$\begin{aligned}\pi_i(\mu_{it},\sigma_{it}) = \max_{p^s_{it},r^s_{it}}\{\frac{q}{1-\beta(1-q)}\{&\int_{\bar{v}_{it}}^{\infty} p^s_{it}\cdot f(\hat{v}_{it})\,\mathrm{d}\hat{v}_{it}\\ &+\int_{r^s_{it}}^{\bar{v}_{it}}[\theta\hat{v}_{it}+(1-\theta)r^s_{it}]f(\hat{v}_{it})\,\mathrm{d}\hat{v}_{it}\\ &+\beta\int_{\infty}^{r^s_{it}}\pi_i(\mu_{i(t+1)}(\hat{v}_{it}),\sigma_{i(t+1)})f(\hat{v}_{it})\,\mathrm{d}\hat{v}_{it}\}\}.\end{aligned} \tag{26}$$

The seller's reservation price is the lowest price that she is willing to accept. Thus, the optimal reservation price should be the value at which she is indifferent between selling the property and moving into the next period. Thus, the optimal reservation price solves the following equation:

$$r^s_{it} = \mathbb{E}(\pi_i(\mu_{i(t+1)}(\hat{v}_{it}),\sigma_{i(t+1)})|\hat{v}_{it}<r^s_{it}) = \frac{\beta\int_{\infty}^{r^s_{it}}\pi_i(\mu_{i(t+1)}(\hat{v}_{it}),\sigma_{i(t+1)})f(\hat{v}_{it})\,\mathrm{d}\hat{v}_{it}}{\Phi(\frac{\ln(r^s_{it})-\mu_{it}}{\bar{\sigma}_{it}})}. \tag{27}$$

By First Order Condition, we know that that the optimal listing price satisfies

$$\frac{\partial\pi_i(\mu_{it},\sigma_{it})}{\partial p^s_{it}} = 0. \tag{28}$$

Equations 27 and 28 give the optimal solutions for listing price and reservation price. After deciding on the optimal listing price and the optimal reservation price, the seller waits for buyer visit and sell the property if a buyer values the property more than the reservation price. If no buyer visits or no buyer has valuation more than the reservation price, the seller will stay on the market until the property eventually gets sold. We denote the period in which the property is sold as T_i.

4. Estimation

The parameters in our model include the following: the mean and variance of log visiting cost; the mean and variance of prior belief about α_i; the coefficients of home characteristics; the variance of log buyer valuations; the relative bargaining power; the intercept and the slop in the linear transformation of Zestimate to Zestimate signal; the intercept and the slop in the linear transformation of Zestimate range to Zestimate standard deviation; and the seller's discounting factor. Note that relative bargaining parameter θ cannot be identified in our data. Since we directly observe neither buyer valuation nor seller's reservation price, we cannot distinguish the case where the seller's bargaining power is high from the case where the buyer's valuation is high. To evaluate the relative bargaining power, we refer to the "buyer-seller index" on Zillow, which ranges from 0 to 10 and indicates the extend to which a market is a seller's market. The average buyer-seller index in Pittsburgh during our observational period is 8.7, therefore we fix the bargaining parameter θ at 0.87. We also fix the discounting factor β at 0.99. In order to examine the potential difference in the uncertainty across the three neighborhood groups, we allow the three neighborhoods to have different prior belief about α_i. That is, we assume that buyers and sellers use the distribution of α_i in their neighborhood groups as the prior belief about α_i. We use σ_0^P, σ_0^M and σ_0^R to denote the standard deviation of prior belief in poor neighborhoods, mid-range neighborhoods and rich neighborhoods, respectively. Similarly, α_0^P, α_0^M and α_0^R denote the mean of prior belief in the three neighborhood groups. Thus, the set of parameters we need to estimate is

$$\Theta = \{c_0, \sigma_c, \sigma_0^P, \sigma_0^M, \sigma_0^R, \alpha_0^P, \alpha_0^M, \alpha_0^R, \sigma_v, a_z, b_z, a_s, b_s, \beta, \gamma\}. \tag{29}$$

4.1. The Likelihood Function

We estimate the model using Maximum Likelihood Estimation. In this section, we derive the likelihood function. For each property, we observe its listing price in each period (p_{it}^s), the period in which it was sold (T_i), the final selling price p_i, the Zestimate (Z_{it}) and the Zestimate range (Z_{it}^H, Z_{it}^L) in each period. We first solve the seller's problem by Equation 27 and 28, and with the observed listing price and Zestimate we infer the seller's belief about λ_i in each period (μ_{it}, σ_{it}).

Let $S_{it} \in \{0,1\}$ denotes whether there is a new Zestimate signal in period t, then the likelihood of observing Zestimate Z_{it} when Zestimate range is $[Z_{it}^H, Z_{it}^L]$ conditional on α_i is

$$l_{it}^z(\Theta|\alpha_i) = [\frac{1}{a_s + b_s \frac{Z_{it}^H - Z_{it}^L}{Z_{it}}} \phi(\frac{a_z + b_z \ln(Z_{it})) - \lambda_i}{a_s + b_s \frac{Z_{it}^H - Z_{it}^L}{Z_{it}}})]^{S_{it}}. \tag{30}$$

Note that with μ_{it} and σ_{it}, we can infer buyer visits and buyer valuations while a property stays on the market. Specifically, if $\mu_{it} \neq \mu_{i(t+1)}$, then it implies that a buyer visits the property in period t, and the buyer valuation is

$$v_{it} = \exp\{(\mu_{i(t+1)} - \mu_{it}) \frac{\sigma_{it}^2 + \sigma_v^2}{\sigma_{it}^2} + \mu_{it}\}. \tag{31}$$

The probability that a buyer does not visit the property is

$$l_{it}^n(\Theta) = 1 - q_{it}. \tag{32}$$

where q_{it} is given by Equation 20. And the likelihood of a buyer visiting the property with realized valuation of v_{it} conditional on α_i is

$$l_{it}^w(\Theta|\alpha_i) = q_{it} \cdot \frac{1}{\sigma_v} \phi(\frac{\ln(v_{it}) - \lambda_i}{\sigma_v}). \tag{33}$$

If a property is sold below the listing price, we know that a buyer visits and his valuation is

$$v_{iT_i} = \frac{p_i - (1-\theta) r_{iT_i}^s}{\theta}, \tag{34}$$

thus the likelihood of being sold below the listing price is

$$l_i^p(\Theta|\alpha_i) = q_{it} \cdot \frac{1}{\sigma_v} \phi(\frac{\ln(v_{iT_i}) - \lambda_i}{\sigma_v}). \tag{35}$$

If a property is sold at the listing price, we know that a buyer visits and his valuation satisfies

$$v_{iT_i} \geq \frac{p_{iT_i}^s - (1-\theta) r_{iT_i}^s}{\theta} \equiv \bar{v}_{iT_i}, \tag{36}$$

thus the likelihood of being sold at the listing price is

$$l_i^s(\Theta|\alpha_i) = q_{it} \cdot \Phi(\frac{\ln(\bar{v}_{iT_i}) - \lambda_i}{\sigma_v}). \tag{37}$$

Using the previous definitions, we construct the likelihood contribution of an observation sold below the listing price as

$$l_i^{p_i<p_i^s}(\Theta) = \int [\prod_{i=1}^{T_i-1} (l_{it}^n(\Theta))^{(1-A_{it})} \cdot (l_{it}^w(\Theta|\alpha_i))^{A_{it}} \cdot (l_{it}^z(\Theta|\alpha_i))^{S_{it}}] \cdot l_i^p(\Theta|\alpha_i) \cdot (l_{iT_i}^z(\Theta|\alpha_i))^{S_{iT_i}} f(\alpha_i)\, \mathrm{d}\alpha_i, \tag{38}$$

and likelihood contribution of an observation sold at the listing price is

$$l_i^{p_i=p_i^s}(\Theta) = \int [\prod_{i=1}^{T_i-1} (l_{it}^n(\Theta))^{(1-A_{it})} \cdot (l_{it}^w(\Theta|\alpha_i))^{A_{it}} \cdot (l_{it}^z(\Theta|\alpha_i))^{S_{it}}] \cdot l_i^s(\Theta|\alpha_i) \cdot (l_{iT_i}^z(\Theta|\alpha_i))^{S_{iT_i}} f(\alpha_i)\, \mathrm{d}\alpha_i. \tag{39}$$

Therefore the likelihood of observing the sample is

$$L(\Theta) = \prod_i (l_i^{p_i<p_i^s}(\Theta))^{\mathbb{1}(p_i<p_i^s)} \cdot (l_i^{p_i=p_i^s}(\Theta))^{\mathbb{1}(p_i=p_i^s)}, \tag{40}$$

and the MLE parameter estimates are the ones that maximize the log-likelihood of observing the sample.

4.2. The Estimation Results

Table 3 shows the result of the estimated parameters. The average normalized visiting cost is 0.0603, and the variance is 0.0117, suggesting that buyers incur non-trivial cost in visiting a property. The estimates of the variance of prior belief suggest that buyers and sellers in the poor neighborhoods face the greatest uncertainty about property values, followed by buyers and sellers in rich neighborhoods, and buyers and sellers in mid-range neighborhoods face the least uncertainty. This may be because properties in mid-range neighborhoods are more comparable to each other and that the turnover rate is higher in mid-range neighborhoods. The standard deviation of log buyer valuations is 0.1913, suggesting significant heterogeneity in buyer valuations. The estimates of a_z and b_z suggest that the Zestimate signal about mean buyer valuation is a fraction of the raw Zestimate value. This is intuitive as the market value (expected selling price) tends to higher than the mean buyer valuation. The estimates of a_s and b_s imply that the average standard deviation of Zestimate signal is 0.2323, which is significant higher than the standard deviation of the prior belief.

Table 3 Parameter Estimates

Parameter	Description	Estimate	Standard Errors
c_0	The mean of log visiting cost	-3.5893	0.0211
σ_c	The standard deviation of log visiting cost	1.2078	0.0087
σ_0^P	The standard deviation of prior belief in poor neighborhoods	0.1090	0.0016
σ_0^M	The standard deviation of prior belief in mid-range neighborhoods	0.1018	0.0012
σ_0^R	The standard deviation of prior belief in mid-range neighborhoods	0.1022	0.0017
α_0^P	The mean of prior belief in poor neighborhoods	7.8243	0.5004
α_0^M	The mean of prior belief in mid-range neighborhoods	8.3667	0.4993
α_0^R	The mean of prior belief in rich neighborhoods	8.5967	0.4227
σ_v	The standard deviation of log buyer valuations	0.1912	0.0003
a_z	The intercept in the linear transformation of Zestimate	-0.1196	0.0012
b_z	The slope in the linear transformation of Zestimate	1.0006	0.0005
a_s	The intercept in the linear transformation of Zestimate range	0.0436	0.0011
b_s	The slope in the linear transformation of Zestimate range	1.2584	0.0035

5. Counterfactual Analysis

Our structural model along with the estimated parameters allow us to examine market outcomes under different conditions. In this section, we conduct counterfactual analysis to analyze the impact of Zestimate and Zestimate accuracy. We first calculate a posterior distribution of α_i for each individual property, and use the mean of the posterior distribution as the value of α_i in simulations. For each condition, we simulate 500 paths and calculate the average initial listing price, the average final selling price, the average TOM, the average buyer welfare, the average seller welfare and the average total welfare for each property in our sample.

5.1. The (Disparate) Impact of Zestimate

In the first counterfactual analysis, we compare the scenario where Zestimate is present to the scenario where Zestimate is removed. When Zestimate is removed, buyers and sellers make decisions based on their prior belief about property values, and only update their belief on buyer valuations. In this case, the uncertainty in belief is higher and the mean belief is not shifted by an over-valued or under-valued Zestimate signal. Table 4 shows the the basic statistics of the initial listing price, the final selling price and the TOM under the two conditions. We can see that on average Zestimate leads to higher listing price, higher selling price and longer TOM.

Table 4 Market Outcomes with and without Zestimate

	Listing Price		selling price		TOM(# of periods)	
	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
without Zestimate	246940.21	143298.64	214117.08	124775.17	4.6672	1.0889
with Zestimate	242436.78	140347.61	215940.75	125511.93	5.0701	1.1799

We also calculate the expected buyer welfare and the expected seller welfare for each property under the two conditions. The seller welfare is calculated as difference between the selling price and mean of buyer valuations, discounted by the time it takes to sell:

$$\text{SellerWelfare}_i = \beta^{T_i - 1} \cdot (p_i - \exp(\lambda_i + \frac{\sigma_v^2}{2})) \tag{41}$$

The buyer welfare has two parts: the utility of purchasing, which is the difference between the final buyer valuation and the selling price, and the total visiting of all buyers who visit the property:

$$\text{BuyerWelfare}_i = (v_{iT_i} - p_i) - \sum_{t=1}^{T_i} c_{it} A_{it} \tag{42}$$

The total welfare is simply the sum of seller welfare and buyer welfare.

Table 5 Average Welfare Change by Zestimate

	Population	Poor	Mid-Range	Rich
Total Surplus	6.16%	7.09%	5.67%	5.96%
Seller Profit	7.53%	8.58%	6.89%	7.47%
Buyer Surplus	4.42%	4.75%	4.36%	4.15%

Table 6 shows the the average welfare change caused by Zestimate. First, we can see that Zestimate on average increases both seller welfare and buyer welfare in the population, suggesting that overall Zestimate benefits the market. Note that in our sample Zestimate under-estimates selling price for more than 40% of the properties, yet only 19.04% of the properties have a lower seller welfare with Zestimate than without, suggesting that having an undervalued Zestimate may still be better for the seller than not having a Zestimate because of the benefit of uncertainty reduction. Moreover, although Zestimate is least accurate in poor neighborhoods, we find that Zestimate leads to the greatest total welfare increase and seller welfare increase in poor neighborhoods, suggesting that it actually benefits poor neighborhoods most.

5.2. Improving Zestimate Accuracy in Poor Neighborhoods

The previous counterfactual analysis shows that currently Zestimate leads to greater welfare increase in poor neighborhoods and therefore helps to bridge the gap in housing inequality. However, it is also clear that more accurate Zestimate could benefit more and poor neighborhoods are losing some of the benefits because of the less accurate Zestimate. There are generally fewer features available on Zillow for properties in poor neighborhoods, and even the available features may be outdated or erroneous due to less well-maintained public records. Luckily, Zillow allows homeowners to provide or update home characteristics information to improve Zestimate accuracy.[16] Therefore, homeowners could provide more information about their properties and benefit from a more accurate Zestimate.

Table 6 Average Surplus Change in Poor Neighborhoods with More Accurate Zestimate

Zesetimate	Current	Improved
Total surplus	7.09%	9.30%
Seller profit	8.58%	10.12%
Buyer surplus	4.75 %	7.45%

[16] https://www.zillow.com/sellerlanding/edityourhome/

To evaluate how much poor neighborhoods are losing because of the less accurate Zestimate, in the second counterfactual analysis, we increase Zestimate accuracy in poor neighborhoods to be the same as that in rich neighborhoods. As Zestimate becomes more accurate, it leads to more uncertainty reduction and it is less deviated from true property value. Table 6 shows the average surplus change in poor neighborhoods as we increase Zestimate accuracy. If on average Zestimate in poor neighborhoods were as accurate as that in rich neighborhoods, then the average total surplus change would be 9.30%. Compared to the total surplus change of 7.09% with the current Zestimate accuracy in poor neighborhoods, this suggests that the positive impact of Zestimate on total surplus in poor neighborhoods could further increase by 31.17% with higher accuracy.

6. Conclusion

In this paper, we study the impact of Zillow's Zestimate on the housing market. Zestimate is an estimate of a home's market value generated by a machine learning algorithm leveraging large amount of data and computational power. As it provides an additional signal about property value, Zestimate has the potential to benefit the market by reducing the uncertainty in the market. However, as achieving perfect prediction is extremely difficult if not impossible, Zestimate is more or less erroneous, and it tends to be an over-estimate for some of the properties and under-estimate for other properties. Both under-valued and over-valued Zsetimate could be problematic, as under-valued Zestimate could lead to lower belief property value and may results in lower selling price, while over-valued Zestimate could lead to incorrectly high expectation about property value and result in inefficient search. Thus, our first objective is the examine how Zestimate affect housing market in terms of market outcomes and welfare. Meanwhile, we notice that Zestimate tends to be significantly more accurate for rich neighborhoods compared to poor neighborhoods, which raises concerns that Zestimate may widen the socio-economic inequality. Thus, our second objective is to identify how and to what extent Zestimate affects the social inequality in the housing market.

We build a structural model of housing market where sellers and buyers face uncertainty about property values and Zestimate provides an unbiased signal of the property value. Our model captures two potentially countervailing effects of Zestimate: First, it reduces the uncertainty in the belief about property value; second, it shifts the mean belief about property values towards its value. The estimation results reveal that people in poor neighborhoods face greater uncertainty in their prior belief caompred to those in the mid-range and rich neighborhoods. In a counterfactual analysis, we show that the introduction of Zestimate increase both seller profit and buyer surplus, and having an undervalued Zestimate could still be better for the seller than not having an Zestimate. In addition, Zestimate leads to the greatest total surplus increase in poor neighborhoods despite being the least accurate. While poor neighborhoods are benefiting most from Zestimate, they are still missing out some of the benefits because of the less accurate Zestimate. In another counterfactual analysis, we show that poor neighborhoods are losing significant amount of positive impact of Zestimate due to lower accuracy.

There are several limitations of the paper. First, we do not model the case where multiple buyers come and engage in a bidding war, which usually leads to a selling price higher than the listing price. While such cases are rare in our data, it is common in a hot market. Zestimate may affect the housing market differently and future research may investigate the dynamics in this case. Second, due to the computational constrains, we look at only one local market. Future work may study the impact of Zestimate across multiple markets, which may reveal interesting insights on the impact of Zestimate under different conditions.

References

Agrawal A, Gans JS, Goldfarb A (2019) Artificial intelligence: the ambiguous labor market impact of automating prediction. *Journal of Economic Perspectives* 33(2):31–50.

Calvano E, Calzolari G, Denicolo V, Pastorello S (2020) Artificial intelligence, algorithmic pricing, and collusion. *American Economic Review* 110(10):3267–97.

Chen Y, Rosenthal RW (1996) On the use of ceiling-price commitments by monopolists. *The RAND Journal of Economics* 207–220.

Ching AT, Erdem T, Keane MP (2013) Learning models: An assessment of progress, challenges, and new developments. *Marketing Science* 32(6):913–938.

Cowgill B (2018) Bias and productivity in humans and algorithms: Theory and evidence from resume screening. *Columbia Business School, Columbia University* 29.

Cui R, Li J, Zhang DJ (2020) Reducing discrimination with reviews in the sharing economy: Evidence from field experiments on airbnb. *Management Science* 66(3):1071–1094.

Draganska M, Klapper D, Villas-Boas SB (2010) A larger slice or a larger pie? an empirical investigation of bargaining power in the distribution channel. *Marketing Science* 29(1):57–74.

Edelman B, Luca M, Svirsky D (2017) Racial discrimination in the sharing economy: Evidence from a field experiment. *American Economic Journal: Applied Economics* 9(2):1–22.

Erdem T (1998) An empirical analysis of umbrella branding. *Journal of Marketing Research* 35(3):339–351.

Erdem T, Keane MP (1996) Decision-making under uncertainty: Capturing dynamic brand choice processes in turbulent consumer goods markets. *Marketing science* 15(1):1–20.

Erdem T, Keane MP, Sun B (2008) A dynamic model of brand choice when price and advertising signal product quality. *Marketing Science* 27(6):1111–1125.

Fu R, Huang Y, Singh PV (2021) Crowds, lending, machine, and bias. *Information Systems Research* 32(1):72–92.

Hansen KT, Misra K, Pai MM (2021) Frontiers: Algorithmic collusion: Supra-competitive prices via independent algorithms. *Marketing Science* 40(1):1–12.

Huang Y, Vir Singh P, Srinivasan K (2014) Crowdsourcing new product ideas under consumer learning. *Management science* 60(9):2138–2159.

Ihlanfeldt K, Mayock T (2009) Price discrimination in the housing market. *Journal of Urban Economics* 66(2):125–140.

King AT, Mieszkowski P (1973) Racial discrimination, segregation, and the price of housing. *Journal of Political Economy* 81(3):590–606.

Kleinberg J, Lakkaraju H, Leskovec J, Ludwig J, Mullainathan S (2018) Human decisions and machine predictions. *The quarterly journal of economics* 133(1):237–293.

Lambrecht A, Tucker C (2019) Algorithmic bias? an empirical study of apparent gender-based discrimination in the display of stem career ads. *Management Science* 65(7):2966–2981.

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

Mehta N, Chen X, Narasimhan O (2008) Informing, transforming, and persuading: Disentangling the multiple effects of advertising on brand choice decisions. *Marketing Science* 27(3):334–355.

Mehta N, Rajiv S, Srinivasan K (2003) Price uncertainty and consumer search: A structural model of consideration set formation. *Marketing science* 22(1):58–84.

Merlo A, Ortalo-Magné F, Rust J (2015) The home selling problem: Theory and evidence. *International Economic Review* 56(2):457–484.

Nash JF (1950) The bargaining problem. *Econometrica* 18(2):155–162.

Obermeyer Z, Powers B, Vogeli C, Mullainathan S (2019) Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 366(6464):447–453.

Piketty T, Zucman G (2014) Capital is back: Wealth-income ratios in rich countries 1700–2010. *The Quarterly Journal of Economics* 129(3):1255–1310.

Salant SW (1991) For sale by owner: When to use a broker and how to price the house. *The Journal of Real Estate Finance and Economics* 4(2):157–173.

Smith LB (1969) A model of the canadian housing and mortgage markets. *Journal of Political Economy* 77(5):795–816.

Straszheim MR (1974) Housing market discrimination and black housing consumption. *The Quarterly Journal of Economics* 19–43.

Wheaton WC (1990) Vacancy, search, and prices in a housing market matching model. *Journal of political Economy* 98(6):1270–1292.

Yu S (2021) Algorithmic outputs as information source: The effects of zestimates on home prices and racial bias in the housing market. *Available at SSRN 3584896* .

Zhang L, Chung DJ (2020) Price bargaining and competition in online platforms: An empirical analysis of the daily deal market. *Marketing Science* 39(4):687–706.

Zhang S, Mehta N, Singh PV, Srinivasan K (2021) Can an ai algorithm mitigate racial economic inequality? an analysis in the context of airbnb. *Marketing Science* .

Zhao Y, Yang S, Narayan V, Zhao Y (2013) Modeling consumer learning from online product reviews. *Marketing Science* 32(1):153–169.