Home Value Estimates: Understanding Their Purposes And Evaluating Their Results

Posted by Jasjeet Thind (https://www.zillow.com/data-science/author/jasjeett/)
🐦 (https://twitter.com/jasjeetthind) 🔗 (https://www.linkedin.com/in/jasjeetthind)
on May 24, 2017
in Machine Learning (https://www.zillow.com/data-science/category/analytic-applications/), Research
(https://www.zillow.com/data-science/category/research/)

Summary

- Redfin estimates are available on far fewer listed homes than Zillow (83% vs. 100%).
- Redfin estimates, prior to the selling agent suggesting a listing price, have a higher error rate than Zillow (9.1% vs. 7.8%).
- After the listing agent has communicated the price they intend to offer the house for sale, Redfin's error is lower than Zillow (2.7% vs. 4.1%).
- Redfin's estimate after a home has been listed for sale is not independent of the listing price itself, potentially imparting little information to a consumer beyond the listing price from which it is derived directly.

Introduction

The world of automated home valuations has definitely been heating up over the past couple of years. Several new entrants have launched products leading to a lot of discussion on everything from the theoretical—what an estimate should be striving to reflect—to the practical—how do various estimates compare in terms of accuracy and coverage.

With all of this recent activity in mind, we thought it might be helpful to examine some of these questions.

Let's start with the theoretical. Home value estimates can have a variety of purposes. The estimate might be intended to reflect the value of a property based on its rental potential (income capitalization approach) or based on the tangible costs of land and building materials (cost approach) or based on comparable sales.

For a comparable sales-based approach to home value, one must also decide which types of sales they would like to reflect with their estimate: all sales, distressed sales, non-distressed sales, intra-family sales, arms-length sales, and still other categories of sales.

Additionally, one must decide the degree to which their estimate reflects the view of the two parties in any asset transaction: the buyer and the seller. In asset transactions, the seller sets an offer or ask price whereas the buyer sets a bid price. In real estate, confusingly, the offer price is referred to as the "listing price" while the bid price is referred to as the "offer price." Typically, the listing price is higher than the offer price except in very hot markets where prices are being bid up because of too many buyers chasing too few sellers.

The Zillow Zestimate home valuation was designed to reflect the market value of a home based on comparable sales where those sales are full-value, non-distressed, arms-length transactions of real estate. Let's unpack this definition a bit.

A full-value sale means that the sale price reflects the entirety of the value being conveyed from seller to buyer and no other side-benefit is considered. For example, we don't want to consider the sale of a property for half its real value because the seller is getting something else of value from the buyer which is not reflected in the sale price. We also don't want the Zestimate to reflect the value of a home if sold by a county sheriff at a foreclosure auction nor if a parent sells a property to their children at a price below market value.

The Zestimate home valuation was also designed to be independent of any opinion from either the seller or buyer. Neither party is a neutral, unbiased observer in the transactions. Is the seller listing a home higher than market value in order to leave room for negotiation? Or are they trying to drum up more interest in the listing by pricing the home below market value? Is the buyer cherry picking comparable sales in order to justify a low-ball offer? We'd like to arrive at an estimate of value that is neutral to the opinions on either side of the transaction.

In the case of the Zestimate modeling framework, which contains numerous submodels, each estimating a home's value via different valuation approaches and data inputs, our goal of independence means that the listing price is not a factor in any of our valuation submodels. We will, however, use the listing price as a hint when selecting from amongst all available submodel estimates for the given home, but only when there is a substantial difference between the listing price and the submodel estimate that the system would have selected without reference to the listing price.

For example, consider a home with three submodel estimates: $40,000, $50,000 and $100,000. The modeling system may determine that the $50,000 estimate or some blend of the $40,000 and $50,000 estimates are the most accurate value for the Zestimate. However, if the home is then listed for sale at $95,000, the system will look for a submodel that could support that price, find the $100,000 estimate and present that as the most probable market value for the home. Importantly though, the estimate chosen must be arrived at independently from the list price. In other words, there must be a valuation approach or some set of comparable sales that produces that estimate, not merely the fact that there is a listing price in that range.

Looking at the numbers

With that as background, it's perhaps not surprising that we read with great interest the recent SSRS report (https://www.redfin.com/blog/wp-content/uploads/sites/5/2017/02/Findings-Report.Final-Draft.Revised.2.2.17-FINAL.pdf) from February 2017 comparing the accuracy of Zillow and Redfin estimates of home value. We applaud Redfin for engaging SSRS to conduct an empirical evaluation of home value estimates provided by Zillow and Redfin. Prediction is an empirical business so ruthlessly evaluating one's own efforts and how they compare to others is a vital part of the scientific method that improves outcomes. At Zillow, such critical evaluation has long been a part of our culture as seen here (https://www.zillow.com/research/putting-accuracy-in-context-3255/), here (https://www.zillow.com/research/case-shiller-not-just-a-

home-price-index-533/), here (https://www.zillow.com/research/whats-the-real-discount-on-a-foreclosure-3229/), and here (https://www.zillow.com/research/zillow-stan-humphries-urban-institute-5504/).

Despite being independently conducted, the SSRS analysis did not address the issue of AVM accuracy in ways that are fairly conventional for such analyses. Typically, analyses of automated valuation models (AVMs) observe estimates on homes before they are listed for sale, compare those estimates to final sale prices, and compute the following two critical metrics:

1. Hit rate: The percentage of properties for which an estimate was available at all. This is important to consumers because it measures how often the estimate will be available to them in order to inform their decisions.
2. Accuracy: An error statistic such as median or mean absolute percent error which compares the estimate to the sale price.

In this case, however, the SSRS analysis did not compare the hit rate of the two estimates and, critically, the accuracy metric computed by SSRS was only computed on the estimate after the home had been listed for sale.

In this analysis, we set out to address these shortcomings by looking at homes which were listed for sale and subsequently sold and determining the hit rate and accuracy of the estimates, both before the home was listed for sale and after it was listed.

Data

We examined all homes in King County, Washington (where Seattle is located) that were first listed for sale on Zillow between December 23, 2016 and January 23, 2017. There were 1,453 homes listed for sale during this time of which 1,167 were single-family or condo homes. For each listing, we identified:

1. Whether the home could be found on Zillow and Redfin prior to the listing;
2. If found, whether a home value estimate was displayed prior to the listing; and
3. The estimated values before and after a home was listed for sale, the listing price, and the final sale price (if it sold).

Because we are interested in knowing the Redfin estimate before a home is listed for sale, we retrieved the Redfin webpages from a search engine web cache which keeps a copy of every internet webpage as of the last time that the search engine visited the page for search engine indexing purposes.

There were 582 homes for which a webpage prior to the initial listing date could be found on both websites and 359 of these homes had both Zillow and Redfin estimates and a subsequent sale.

There were 822 homes for which a webpage after the initial listing date could be found on both websites and 404 of these homes had both Zillow and Redfin estimates and a subsequent sale.

More information about how these data were obtained can be found in the Methodology section below.

Hit rate: Likelihood of a consumer finding a home value estimate

The hit rate was computed as the ratio of homes for which an estimate was provided on a webpage for a home, prior and subsequent to the listing, relative to the number of web pages found on the website (or in the web cache, in the case of Redfin), prior and subsequent to the listing.

Before a home was listed, 554 homes on Redfin had estimates out of 582 homes that were listed for sale and found in the web cache with a version of the Redfin page pre-dating the appearance of the listing, yielding a hit rate of 95%. For Zillow, 582 homes had estimates out of 582 homes that were listed for sale and were found on the Zillow website prior to the date of the listing, yielding a hit rate of 100%.

After a home was listed for sale, 684 homes on Redfin had estimates out of 822 homes that were listed for sale and found in the web cache, yielding a hit rate of 83%. For Zillow, 819 homes had estimates out of 822 homes that were listed for sale and found on the Zillow website, yielding a hit rate of almost 100%.

The significant decrease in coverage for Redfin once a home is listed for sale is presumably related to the fact that sellers can elect to remove the Redfin estimate from the listing, which can reduce the benefit to prospective buyers.

Accuracy: How close are estimates to the final sale price

For each home with an estimate, the following values were retrieved:

a. Redfin estimate before listing
b. Redfin estimate after listing
c. Zillow Zestimate before listing
d. Zillow Zestimate after listing
e. For-sale price at time of listing
f. Final sale price

Pre-listing accuracy was determined by comparing (a) and (c) to (f). These metrics are shown in Table 1.

Post-listing accuracy was determined by comparing (b), (d) and (e) to (f). These metrics are shown in Table 2.

Table 1: Accuracy comparision (before listing)

	Zillow Zestimate	Redfin Estimate
Pct of homes with estimates	100%	95%
Within 5% of sale price	35%	29%
Within 10% of sale price	62%	53%
Within 20% of sale price	89%	80%
Median absolute percentage error	7.8%	9.1%

For estimate coverage, N=582 (number of webpages found on both sites prior to listing)

For accuracy metrics, N=359 (number of webpages found on both sites prior to listing, with Redfin and Zillow estimates and having sold)

When looking at estimate accuracy before a listing appeared (Table 1), the Zestimate achieved a lower error rate than the Redfin estimate. The median absolute percent error for the Zestimate was 7.8% compared to 9.1% for the Redfin estimate, with 89% of Zestimates within 20% of the final sale price versus only 80% of Redfin estimates. Sixty-two percent of Zestimates were within 10% of the final sale price versus only 53% of Redfin estimates.

Turning to accuracy after a home is listed for sale (Table 2), accuracy for both Zillow and Redfin improved substantially. Zillow's median absolute percent error improved to 4.1% while Redfin's error improved to 2.7%, more accurate than both the Zestimate and the initial listing price itself (3.6%).

In the case of the Zestimate, a chief reason for the gain in accuracy after a listing is active is that the home facts are updated with the new listing information whereas, before listing, the estimate is generally based on public record facts. Better facts result in a better estimate. As noted earlier, for a small number of listings for which the listing price is substantially different than the recommended estimate from the modeling system, the modeling system will attempt to find another submodel that supports the listing price. If no submodel supporting the listing price can be found, the prior recommended estimate is used. Otherwise, the submodel that independently supports the listing price is used and this can result in a more accurate prediction. The listing price itself, however, is not a factor in any valuation submodel.

Table 2: Accuracy comparision (after listing)

	Zillow Zestimate	Redfin Estimate	Listing Price
Pct of homes with estimates	100%	83%	100%
Within 5% of sale price	58%	75%	60%
Within 10% of sale price	84%	93%	83%
Within 20% of sale price	98%	99%	98%
Median absolute percentage error	4.1%	2.7%	3.6%

For estimate coverage, N=822 (number of webpages found on both sites after listing)

For accuracy metrics, N=404 (number of webpages found on both sites after listing, with Redfin and Zillow estimates and having sold)
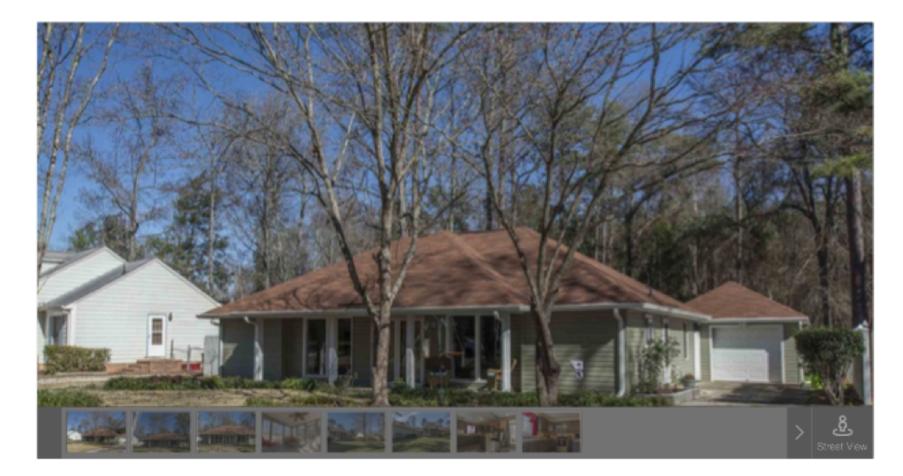
Redfin has pursued a more direct approach to using the listing price to inform its estimate, a decision which undoubtedly helps it achieve the 2.7% error rate observed here. While Redfin apparently uses some form of a comparable sales-based valuation approach when homes are

off-market (i.e., before they are listed for sale), their estimate is based directly on the listing price itself once a home is listed for sale.

The dependence of the Redfin estimate on the listing price is clearly visible in cases where the listing price is obviously erroneous. For example, a recent listing in South Carolina was inadvertently listed for sale at $129,900,000 instead of the correct listing price of $129,900 (making it, temporarily, the most expensive listing ever in Irmo, SC!). Figure 1 shows the webpage on Redfin for this listing (with an estimate of more than $130 million) while Figure 2 shows the equivalent page as it appeared on Zillow at the exact same time (with a Zestimate home valuation of around $120,000). The home eventually sold for $130,000.

Figure 1

REDFIN   1-877-973-3346   Buy ▾   Sell ▾   Real Estate Agents

209 Southampton Dr $129,900,000   3   2   1,387 Sq. Ft.
Irmo, SC 29063   Price   Beds   Baths $93655 / Sq. Ft.
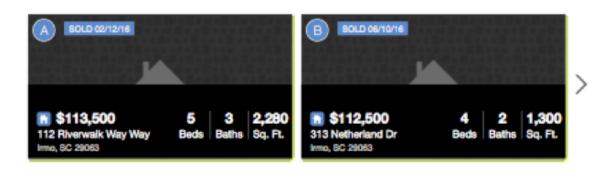Status: Active   Redfin Estimate: $130,960,956   On Redfin: 3 days

Redfin Estimate for 209 Southampton Dr

$130,960,956   Estimate based on 3 beds, 2 baths, 1,387 sq. ft.

+$1.1M over list price of $130M

Comparable Homes

A SOLD 02/12/16   B SOLD 06/10/16

$113,500   5   3   2,280 $112,500   4   2   1,300
112 Riverwalk Way Way   Beds Baths Sq. Ft.   313 Netherland Dr   Beds Baths Sq. Ft.
Irmo, SC 29063   Irmo, SC 29063

Figure 2

209 Southampton Dr,
Irmo, SC 29063

3 beds · 2 baths · 1,387 sqft Edit

Edit home facts for a more accurate Zestimate.

● OFF MARKET

Zestimate®: $119,852

Price this home

Rent Zestimate®: $1,090 /mo

This raises the interesting, almost philosophical question of whether accuracy gains are always preferred even at the cost of independence of the estimate from the opinion of one party in the transaction (in this case, the seller or their agent). The listing price is, of course, a very good predictor of the sale price since it is, literally, the price at which the seller will agree immediately to a sale. Any estimate based directly on the list price will also have this advantage. For example, if you set about estimating the price of a specific car by asking somebody who is about to sell the car the price they intend to ask for it and then make your estimate equal to the price they tell you, your "estimate" will likely be pretty close to the eventual sale price. Whether this "estimate" contains any additional value beyond the price the seller is asking is another question.

The South Carolina example shows the pitfalls of such an approach. In a situation where a seller is asking an unreasonable price for a property, the buyer would prefer to have an estimate that is independent of the seller's asking price, particularly since the seller is far from being a neutral observer in the transaction. The fact that most sellers don't ask an unreasonable price, thus making a prediction based on their asking price fairly accurate, obscures the fact that such a prediction is only accurate until it is not. It has no independence from the seller's own opinion.

Our intent when creating the Zestimate home valuation was to provide the consumer with a good starting point for the approximate value of a home, a starting point that was derived independently of any opinion from either the seller or buyer. When looking at an estimate that is derived directly from the listing price, the consumer is not getting an independent estimate or second opinion on price. Granted, a second opinion can sometimes be farther from the sale price than an estimate based entirely on the listing price, but the latter is guaranteed, by definition, to always be close to the list price even if the list price is not a reasonable price for the home. An independent estimate is not constrained in this way which makes it, arguably, more useful to a consumer.

Conclusion

Here, we've tried to extend Redfin's analysis to make it more similar to rigorous approaches typically found when comparing automated valuation models of which the Zillow Zestimate and Redfin estimate are two examples.

In doing so, our analysis finds that Redfin estimates can be found on far fewer listed homes than Zillow Zestimates (83% vs. 100%) and, prior to the selling agent suggesting a listing price, have a higher error rate (9.1%) than Zillow (7.8%). After the listing agent has communicated the price they intend to offer the house for sale, Redfin's error (2.7%) is lower than Zillow (4.1%), although this is not unexpected since the Redfin estimate is based directly on the listing price itself at this point whereas the Zillow Zestimate home valuation is not. Their approach guarantees an accurate estimate if the offer price of the listing agent reflects a fair market value but, in our opinion, imparts little information to a consumer beyond the listing price from which it is derived directly.

Methodology

For this analysis, we are interested in obtaining estimates for homes before and after they have been listed for sale and we are interested in computing the error rate of those estimates compared to a final sale price. This is a tricky endeavor given that it is unknown in advance which of the 135 million homes in the country will be listed for sale. In the case of Zillow, we display historical estimates of value so it is relatively easy to identify homes that have recently sold, observe their information when they were listed for sale, and observe home value estimates both before and after the date of the listing.

Unfortunately, Redfin does not display such historical estimates which means that, when we observe a sale transaction today, we are not able to recover the home value estimates before the sale occurred or the home was listed for sale.

Our approach to solving this problem proceeded as follows. First, we monitored recent listings in the geography of interest (King County, Washington). When a home was listed for sale, we retrieved a copy of the home's page as it existed on Redfin prior to the listing date via the Google web cache. Google's web cache keeps a copy of every internet webpage as of the last time that Google visited it for search engine indexing purposes. Estimates on webpages recovered from the cache were not used unless the corresponding webpage was within a few days prior to the initial listing date. From these cached pages prior to a home being listed for sale, the Redfin estimate was captured. A second observation of the Redfin estimate was recovered from the Google cache in the same fashion a few days later after the cached version had been updated with the version of the Redfin page which showed the home to be listed for sale currently.

For each of the two observations of the Redfin estimates, before and after a home became listed for sale, the value of the Zillow Zestimate home valuation was retrieved from our internal databases where the date of the Zestimate matched the date of the corresponding Redfin observations. While the corresponding Zillow observation came from our internal databases because of ease of access, we also confirmed that the Zestimate value contained in our database for a given day matched the value found on the Google cached version of the Zillow webpage on that same day.

As noted previously, the hit rate was computed as the ratio of homes for which an estimate was provided on a webpage for a home, prior and subsequent to the listing, relative to the number of web pages found on the website (or in the web cache, in the case of Redfin), prior and subsequent to the listing.

In terms of accuracy, the following data points were captured for each home:

a. Redfin estimate before listing
b. Redfin estimate after listing
c. Zillow Zestimate before listing
d. Zillow Zestimate after listing
e. For-sale price at time of listing
f. Final sale price

The Redfin median absolute percent error (MAPE) for estimates *before* listing is computed as the median of the set of absolute values of [(a) – (f)] / (f) per listing.

The Zillow MAPE for estimates *before* listing is computed as the median of the set of absolute values of [(c) – (f)] / (f) per listing.

The Redfin MAPE for estimates *after* listing is computed as the median of the set of absolute values of [(b) – (f)] / (f) per listing.

The Zillow MAPE for estimates *after* listing is computed as the median of the set of absolute values of [(d) – (f)] / (f) per listing.

The MAPE for the listing price itself is computed as the median of the set of absolute values of [(e) – (f)] / (f) per listing.

Zillow and Zestimate are registered trademarks of Zillow, Inc.

SHARE

Want to work at Zillow?

View Openings (http://www.zillow.com/jobs/openings/)

Follow Us   Facebook (https://www.facebook.com/Zillow)   Twitter (https://twitter.com/zillow)
Google+ (https://plus.google.com/+Zillow)   LinkedIn (https://www.linkedin.com/company/zillow)