Comparison of Estimates of Residential Property Values

Prepared for: Redfin, a residential real estate company that provides web-based real estate database and brokerage services

Prepared by: Aniruddha Banerjee, SVP, Advanced Analytics, SSRS
Yvonne Shands, Research Director, SSRS

Rev. January 27, 2017
**Introduction and Summary**

Home prices are predicted by three sources: Redfin, Zillow, and Homes.com. All three companies provide estimated values of listed homes, i.e., their predicted sale prices. After properties are actually sold, and their actual sale prices are known, it becomes possible to determine how accurately each company predicted those sale prices. For this purpose, some widely used metrics for judging predictive accuracy may be employed.

SSRS was asked by Redfin to independently assess, using appropriate and commonly-used metrics, the relative prediction performance of Redfin, Zillow, and Homes.com. For this purpose, SSRS used a database of 25,000 Redfin-listed properties (assembled as described below), of which 6,338 had actually sold by November 30, 2016 (the end of the data collection period). For the Redfin and Zillow comparison, SSRS computed the metrics to assess relative prediction performance only for the properties (1) that were recorded as sold by both companies and (2) for which estimated values were available from both companies. SSRS followed the same procedure for the comparison between Redfin and Homes.com.

Based on this analysis of independently assembled data, SSRS concludes that the metrics employed demonstrate that Redfin performed better in predicting actual sale prices than either Zillow or Homes.com.

**Data**

Redfin provided SSRS a datafile overnight every Sunday through Thursday of the data collection period (October 19, 2016 to November 30, 2016). This file included all residential properties nationwide that Redfin identified as currently in “pending sale” status for the 194 census-defined U.S. metropolitan areas where Redfin has listing and estimate data. Also included were the estimated values of those properties, as calculated by Redfin at the time the data were sent.

SSRS randomly selected between 500 and 1,000 properties from these daily feeds to be accumulated in its data sample. This incremental random selection process stopped when 25,000 properties had been accumulated in the SSRS sample. Of these, 24,789 properties across 7,531 zip codes and 189 census-defined metropolitan areas had unique addresses and were used for the analysis.

Within 36 hours of being selected into the SSRS sample, each selected property was searched by street address on [www.zillow.com](http://www.zillow.com) (for Zillow-listed properties) and [www.homes.com](http://www.homes.com) (for Homes.com-listed properties) to determine if the selected property could be found on either site. If found, the listing for that property was fully reviewed by an SSRS data entry clerk to identify whether an estimated value was available on that site for that property. If an estimated value was available, it was entered by the clerk into a CfMC data entry system. This process was repeated for each property listed on each of the two sites.

---

1 These properties included single-family, condos, co-ops, townhouses, etc.
SSRS transferred a daily file of selected properties back to Redfin so that Redfin could monitor the curing process for each pending sale. Once a minimum of 5,000 of the 25,000 properties had cured (i.e., gone to final sale), Redfin notified SSRS that it would make available final sale price data as listed on the MLS. Redfin sent the file of final MLS sale prices for all selected properties to SSRS on 12/2/16.

This final sale price was considered to be the actual value of the home with which to compare the estimated value (i.e., predicted sale price) provided by Redfin, Zillow, or Homes.com. Using the metrics described below, SSRS tested for the relative prediction performance of all three companies.

Methodology
Two sets of metrics were used for the purpose:

- Percentile distribution of the absolute value of the prediction error rate (ABSPER), using the SSRS-assembled data on sold properties. The percentile distribution comprised the 50\textsuperscript{th} (median), 75\textsuperscript{th}, 90\textsuperscript{th}, and 95\textsuperscript{th} percentiles.
- Percentage of predicted sale prices (PPSP) that fell within particular ranges of the actual sale price, where the ranges selected were ±3\%, ±5\%, ±10\%, and ±20\%.

ABSPER was computed for all three companies by taking the absolute value of the ratio \(\frac{\text{predicted sale price} - \text{actual sale price}}{\text{actual sale price}}\).\textsuperscript{2} To enable a head-to-head comparison of ABSPERs between Redfin and Zillow (or between Redfin and Homes.com), SSRS selected only the sold properties that both had valued.

PPSP was computed for all three companies as the percentage of predicted sale prices within particular ranges of the actual sale price. For example, with a ±3\% range, SSRS computed the percentage of predictions that were within plus or minus 3\% of the actual sale price. Again, for head-to-head comparisons of PPSPs, SSRS selected only the sold properties that both Redfin and Zillow (or Redfin and Homes.com) had valued.

All comparisons of Redfin and Zillow predictions were conducted on 5,661 properties (the number of sold properties in common that both had valued). Similarly, all comparisons of Redfin and Homes.com predictions were conducted on 5,074 properties (again, the number of sold properties in common that both had valued).

These two metrics relate to predictive accuracy as follows. First, lower ABSPERs mean greater predictive accuracy. Thus, in any head-to-head comparison of the predictive accuracy of two companies, the one with lower ABSPERs at the 50\textsuperscript{th}, 75\textsuperscript{th}, 90\textsuperscript{th}, and 95\textsuperscript{th} percentiles is unambiguously more accurate.

Second, higher PPSPs mean greater predictive accuracy. Again, in any head-to-head comparison of the predictive accuracy of two companies, the one with the higher PPSPs within the ±3\%, ±5\%, ±10\%, and ±20\% ranges (around the sale price) is unambiguously more accurate.

\textsuperscript{2} No distinction was thus made between positive and negative prediction errors, i.e., estimated values that, respectively, exceeded or fell below the actual sale price.
Findings and Conclusion

SSRS’ computations of these metrics are shown in the tables below. Table 1 pertains to the percentile distribution of ABSPERs, and Table 2 pertains to PPSPs by selected ranges. Both tables show comparisons between Redfin and Zillow (and also between Redfin and Homes.com) for the sold properties they had in common.

Table 1. Percentile Distribution of ABSPERs, Comparing Redfin to Zillow and Homes.com

<table>
<thead>
<tr>
<th>ABSPER Percentile</th>
<th>Redfin in Common</th>
<th>Zillow</th>
<th>Redfin in Common</th>
<th>Homes.com</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>2.06%</td>
<td>5.95%</td>
<td>2.05%</td>
<td>10.26%</td>
</tr>
<tr>
<td>75</td>
<td>4.21%</td>
<td>13.44%</td>
<td>4.19%</td>
<td>21.00%</td>
</tr>
<tr>
<td>90</td>
<td>8.92%</td>
<td>35.66%</td>
<td>8.83%</td>
<td>44.03%</td>
</tr>
<tr>
<td>95</td>
<td>14.60%</td>
<td>68.11%</td>
<td>14.60%</td>
<td>80.00%</td>
</tr>
<tr>
<td>N</td>
<td>5,661</td>
<td>5,661</td>
<td>5,074</td>
<td>5,074</td>
</tr>
</tbody>
</table>

Table 2. PPSPs within Selected Ranges, Comparing Redfin to Zillow and Homes.com

<table>
<thead>
<tr>
<th>PPSP Range</th>
<th>Redfin in Common</th>
<th>Zillow</th>
<th>Redfin in Common</th>
<th>Homes.com</th>
</tr>
</thead>
<tbody>
<tr>
<td>3%</td>
<td>63.66%</td>
<td>29.31%</td>
<td>64.03%</td>
<td>16.44%</td>
</tr>
<tr>
<td>5%</td>
<td>79.69%</td>
<td>44.11%</td>
<td>79.58%</td>
<td>26.63%</td>
</tr>
<tr>
<td>10%</td>
<td>91.26%</td>
<td>67.27%</td>
<td>91.47%</td>
<td>49.13%</td>
</tr>
<tr>
<td>20%</td>
<td>96.59%</td>
<td>83.09%</td>
<td>96.96%</td>
<td>73.45%</td>
</tr>
<tr>
<td>N</td>
<td>5,661</td>
<td>5,661</td>
<td>5,074</td>
<td>5,074</td>
</tr>
</tbody>
</table>

The noteworthy finding from Tables 1 and 2 is that Redfin’s estimated values (or, predictions of the actual sale price) are unambiguously superior (in terms of predictive accuracy) than those of either Zillow or Homes.com, for properties that each listed in common with Redfin.

Tables 1 and 2 show that Redfin’s ABSPERs are consistently lower than those of either Zillow or Homes.com at all percentile levels. They are, in fact, substantially lower at the higher percentile levels, implying a much narrower distribution of ABSPERs around the median for Redfin than for either Zillow or Homes.com. This signifies less spread in prediction error rates and greater proximity overall of estimated values to actual sale prices.
Tables 1 and 2 also show that Redfin’s PPSPs are consistently higher within designated ranges around actual sale prices, when compared to those of Zillow or Homes.com. In fact, Redfin’s PPSPs are substantially higher at the wider ranges, implying that Redfin’s estimated values cluster around actual sale prices at higher rates than the estimated values of either Zillow or Homes.com.

Based on these findings, SSRS concludes that Redfin’s performance in predicting actual sale prices (prior to sale of properties that subsequently sold) was superior to those of either Zillow or Homes.com.