SmartZip AVMs

The industry’s best weighted valuation model of home prices
PART ONE

SUMMARY

PART TWO

METHODOLOGY

PART THREE

DATA AND DATA QUALITY

PART FOUR

MODELS

PART FIVE

VALIDATION

PART SIX

THE FUTURE OF AVM ACCURACY
SmartZip’s automated valuation model (AVM) estimates the fair market value of over 110 million homes and is trusted by leading real estate websites to generate off-market home values. SmartZip’s AVM uses multiple statistical models to estimate the fair market value of nearly every home in the US.

The purpose of this paper is to describe the SmartZip AVM including the methodology used, the data behind it, our stringent quality assurance process, and the full-scope validation of the AVM results.

**SMARTZIP AVM SNAPSHOT**

<table>
<thead>
<tr>
<th>110 M Homes, All 50 States</th>
</tr>
</thead>
<tbody>
<tr>
<td>~4% Median absolute percentage error nationwide</td>
</tr>
<tr>
<td>~11% Forecasted standard deviation nationwide</td>
</tr>
</tbody>
</table>
Our algorithms assess real estate data at various levels- from U.S. State, County, and City areas to Census tracts and blocks and neighborhood areas. Each algorithm is slightly different depending on the circumstances of available data and location, assessing variables related to publicly available property records: property attributes (such as beds, baths, building area, lot size, etc.), transaction histories, comparable sales, listing history, and monthly rents; as well as area demographics including income, school quality, crime, lifestyle, population growth, and job growth. More importantly, it relies on our neighborhood-level house price index (NHPI), which measures how much the same houses have sold for over time and in various market conditions.
SmartZip has current and historical data covering more than 110 million residential properties and over 200 million homeowners. We use public records data obtained from 3,000+ County Assessor’s offices and 2,600+ County Recorder’s offices across the U.S.

This data is cleaned, normalized, and paired with other data sets, such as listings data (sourced from MLS’s and real estate brokerages across the country), demographic data, other public records and metrics, consumer profile data (multi-sourced from up to 30 locations in some cases), and much more. The data is further normalized and validated against 3rd parties, (such as the USPS for address standardization and verification), and enriched through various geospatial tools for locality and context.
The SmartZip AVM is based on a multi-model approach, which means the algorithms are able to weigh the independent values of multiple variables that can lead to one final, predicted result. The independent variables we analyze include property, neighborhood and demographic attributes (type of property, bedrooms, building area, property age, school ratings, family income, etc.), which are then modeled against the SmartZip AVM results and market transactions over the last year. Because real estate values are based on hyperlocal factors, the level of regression will be selected from the highest geographical resolution possible that provides enough observations. For example, if there are more than 50 observations with valid AVM results or recent transactions in a Census Block, the regression will be carried out for that Block. If fewer observations are available, the regression will broaden its analysis to include the next available geographical level up, such as Block Group or Tract.

The various models are stochastically weighted to obtain the final SmartZip AVM result. A unique component of SmartZip’s AVM process is that the NHPI and AVM can adapt to sparse inputs and “borrow” data from properties that match the geographical area and property features of the home being evaluated. We also compare results against local listing prices and local sold prices, (that have been reported), in order to finalize our AVM values.
We follow a very rigorous process to ensure that our AVM results are of the highest quality possible — and we validate the results based on real, not hypothetical, market data. As any real estate expert knows, a home is only as valuable as what a buyer is willing to pay for it. For that reason, we measure the accuracy of our AVM results by benchmarking them against actual sales prices, not listing prices. From there, we give each AVM result a confidence score.

The majority of markets have a confidence score above 80. The markets with low AVM confidence are generally those without much historical data or where sales prices have not been disclosed.

The following evaluation metrics are used to determine the accuracy of our AVM results

1) Forecast Standard Deviation

Forecast standard deviation (FSD) quantifies the uncertainty in the AVM. It is calculated as the standard deviation of percentage error. The percentage error is calculated as:

\[
\frac{\text{Predicted home price} - \text{Actual sale price}}{\text{Actual sale price}}
\]

The lower the FSD, the closer the AVM is to the actual sale price. The SmartZip AVM has an FSD of 11%. Below is the spatial distribution of FSD across all US states. Most states have FSD below 15%; the states

![Forecasted Standard Deviation (FSD) for USA](image-url)
VALIDATION (CONTINUED)

(2) Hit Rate
Hit rate is a measure of the number of times (hits) the SmartZip AVM returned a valuation result for the requested property as a percentage of the total properties. The SmartZip AVM results have a hit rate of 90% nationwide.

(3) Median Absolute Percent Error (MAPE)
Median Absolute Percent Error (MAPE) is the median of the absolute difference between the AVM result for a home and its actual sale price expressed as a percentage of the sale price. Our AVM results have a MAPE of 4.0%, this means that half of our results are within ±4% of the sales price and half were outside of the ±4% of the sales price. Below is the MAPE distribution across USA states. The states with high MAPE typically represent non-disclosure areas where sale prices are not publicly available.
(4) Within 'X' % of the sale price

The final benchmarks show the number of homes with AVM results within 'X' % of the sale price. About 80% of SmartZip AVM results are within 10% of sales price and about 90% of SmartZip AVM results are within 20% of sales price.

<table>
<thead>
<tr>
<th>Year</th>
<th>-40%</th>
<th>-30%</th>
<th>-20%</th>
<th>-10%</th>
<th>0%</th>
<th>10%</th>
<th>20%</th>
<th>30%</th>
<th>40%</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012</td>
<td>2.09%</td>
<td>3.81%</td>
<td>7.31%</td>
<td>13.68%</td>
<td>40.93%</td>
<td>13.23%</td>
<td>8.65%</td>
<td>5.96%</td>
<td>4.32%</td>
</tr>
<tr>
<td>2013</td>
<td>1.21%</td>
<td>2.38%</td>
<td>5.30%</td>
<td>12.60%</td>
<td>50.51%</td>
<td>12.35%</td>
<td>7.39%</td>
<td>4.80%</td>
<td>3.47%</td>
</tr>
<tr>
<td>2014</td>
<td>0.87%</td>
<td>1.63%</td>
<td>3.40%</td>
<td>7.96%</td>
<td>59.27%</td>
<td>12.29%</td>
<td>6.92%</td>
<td>4.46%</td>
<td>3.20%</td>
</tr>
<tr>
<td>2015</td>
<td>0.89%</td>
<td>1.58%</td>
<td>3.43%</td>
<td>9.14%</td>
<td>59.50%</td>
<td>12.19%</td>
<td>6.44%</td>
<td>4.01%</td>
<td>2.82%</td>
</tr>
<tr>
<td>2016</td>
<td>0.71%</td>
<td>1.50%</td>
<td>3.61%</td>
<td>10.16%</td>
<td>60.80%</td>
<td>11.39%</td>
<td>5.88%</td>
<td>3.53%</td>
<td>2.42%</td>
</tr>
<tr>
<td>2017</td>
<td>0.43%</td>
<td>0.65%</td>
<td>1.71%</td>
<td>4.98%</td>
<td>74.91%</td>
<td>8.51%</td>
<td>4.42%</td>
<td>2.56%</td>
<td>1.84%</td>
</tr>
<tr>
<td>2018</td>
<td>0.33%</td>
<td>0.56%</td>
<td>1.35%</td>
<td>4.64%</td>
<td>78.52%</td>
<td>7.49%</td>
<td>3.47%</td>
<td>2.22%</td>
<td>1.40%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>within 10%</td>
<td>40.93%</td>
<td>50.51%</td>
<td>59.27%</td>
<td>59.50%</td>
<td>60.80%</td>
<td>74.91%</td>
<td>78.52%</td>
</tr>
<tr>
<td>within 20%</td>
<td>67.85%</td>
<td>75.46%</td>
<td>79.52%</td>
<td>80.83%</td>
<td>82.35%</td>
<td>88.40%</td>
<td>90.65%</td>
</tr>
</tbody>
</table>
Since 2010 SmartZip has tracked AVM benchmarks, a strong upward trend has emerged. This growing accuracy reflects not only our aggressive pursuit of new, high-quality data sets but also the very nature of data modeling, which is dependent on self-programming algorithms that backtest their own predictions and adjust accordingly.

Each and every day, SmartZip’s models backtest their predictions against real sold prices, and train themselves to better identify minute shifts in the market, or to a local buyer and seller demographics, across thousands of neighborhoods and millions of properties nationwide. Today, 90% of SmartZip AVM’s are within 20% of the final sales price — and our team is working tirelessly to close that final gap.