

Using Tests To Find The Best Automated Valuation Model Vendor

You've narrowed your search down to a few AVMs. Now, how do you determine how they'll work for you?

BY DOUGLAS GORDON AND BUD SOTTILI

With the proliferation of the number of automated valuation model (AVM) providers, and the coverage and precision claims made by the vendors, how do you cut through the marketing hype to determine how these models will work for you?

The best way is to test the models yourself, using a sample derived from your own book of business. The primary objective of the test is to determine how the various models will perform for your company in a "live" setting.

Before deciding to conduct an AVM test, you should understand that the process requires considerable staff time to develop a good sample and to aggregate and analyze the results.

If this test is to be a reliable barometer of performance, you need to be very careful in designing the sample of properties provided and in choosing and analyzing the metrics that are used to measure model performance. Poorly designed or executed tests can result in considerable resources spent and misleading results produced.

We've developed some "best practices" in the design and analysis of AVM tests to help you choose which service to use.

However, if you don't have the time

or money to invest in conducting a test, survey some of the larger institutions that have already incurred the cost and effort to compare products for themselves. You could also ask your vendor for test statistics from previous test samples.

Make sure you are comparing apples to apples. It is absolutely impera-



SOTTILI



GORDON

tive you make sure each vendor has compiled the data in the same way. You can use the tools and templates in this article to analyze the data just as if you had conducted the test.

Test preparation

There are four steps to test preparation:

- Prepare a sample of your properties.
- Have the vendors value the properties in the sample.
- Analyze the result.
- Pick the winner.

A little planning can go a long way toward ensuring that you get the best results with minimum effort.

It's critical to understand the time line involved in the test. Best results are achieved when the sample is the freshest. There are two things you can do to minimize the timeline:

- Work with the selected vendors to understand how they process tests, the associated turnaround time and the required input and output data formats.

- Determine how long it will take you to assemble the sample set.

With this information, work with your staff and your vendors to set test dates. The schedule has to be set so all vendors can return the sample set on the same day. In most cases, a week should be more than sufficient to

Douglas Gordon is the director of model development for Freddie Mac's Strategic Information Services Unit. He has contributed to the development of Freddie Mac's automated underwriting, loan costing and property valuation models. Bud Sottili is the director of customer service for Freddie Mac's Strategic Information Services Unit, focusing on strategy development and marketing.

process the sample, and you might be able to reduce that further with good planning.

Developing a sample

An ideal test sample has a sufficient number of loans for which you have a good estimate of value to test against, the loans fairly represent your lending pattern and the loans have not been closed long enough to appear in public records. We'll discuss the importance of each of these characteristics.

Reliable estimate of value . It's difficult to draw reliable conclusions from samples of less than 1,000 loans. These loans must have a reliable estimate of value against which to test the model results.

There is broad consensus that the property's sale price in an arm's-length transaction is the best indicator of the value of the property. Purchase price reflects the price agreed upon by a willing buyer and seller in the open market and is the benchmark you should use.

Refinances do not have a "market tested" value and are not as good as purchase transactions for this type of test.

Loans should represent your lending pattern . Randomly selecting loans from your production should give you a sample that fairly represents your lending pattern.

After the selection is made, you should double-check that the loans are spread correctly over the locations where you lend, the mix of single-family homes to condos, etc. Sometimes the selection process can produce an unrepresentative sample, especially if the process is done manually.

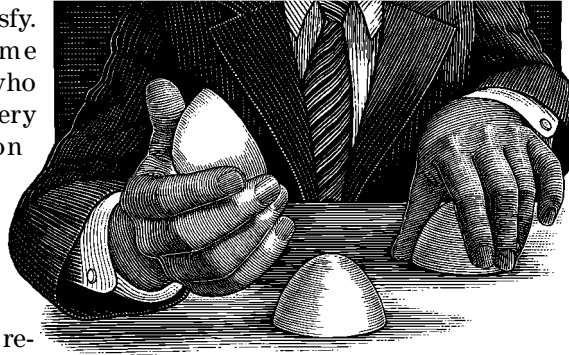
Loans should have closed very recently . Test data should include only very recent transactions, such as loans closed within the two months prior to the test. This is a key criteria which may be the most

difficult to satisfy. There are some AVM vendors who exclusively, or very heavily, rely on using public record data to generate their estimated values.

If the sale or refinance is already in the public records, these models will return the public record value with very little deviation around it. This gives these models the appearance of amazing accuracy, at least on those loans. Sale transactions usually hit the public records within three months of the sale date. A time horizon of not more than two months back from the current date will eliminate most, if not all, of the contaminated records.

The older the loans in your test sample, the greater the number of contaminated records. This is why it's imperative to have all vendors return the test sample at the same time.

The preferred sampling strategy,



our experience, it's difficult to draw reliable conclusions from samples of much less than 1,000 loans. In the real world, it may be logistically impos-

sible to assemble a set of 1,000 "perfect" loans, so you'll need to apply some judgment.

It's probably better to test with less than 1,000 loans, rather than including large numbers of refinances or loans closed a long time ago. If you're adding refinances, make sure that you'll be able to identify them later so you can examine the results separately.

How to interpret the results

Once the test results are returned, you will need to aggregate and analyze them. This analysis can be done with standard spreadsheet tools and functions, unless your sample is very

Test Results Template

After you aggregate test results of AVM firms, a scorecard template can be useful for a side-by-side comparison of test results.

	Product	Countries Covered	Rank	Hit Rate	Rank	% Within +/- 10% Of Sale Price	Rank	% Within +/- 20% Of Sale Price	Standard Deviation Of Differences	Rank	% Of Values Within High-Low Range Of Model	Rank	Total Points
Company A													
Company B													
Company C													

SOURCE: Freddie Mac

and one that has been used by a number of institutions, is to include "final" but not yet "closed" transactions so they are not yet available in public records.

Sufficient number of loans . In

large, in which case advanced analytical tools such as SAS or SPSS may be required.

Since you will have so much raw data, it is important that you focus on the most critical results to assess

which AVM will provide the most value to your institution. Once the results are aggregated, a scorecard template can be useful for a side-by-side comparison of the test results.

Key results areas

County Coverage . This is the number of counties in your lending footprint where the model can be used. Although many AVMs have coverage in all 50 states, the actual number of counties covered in those states is a key differentiator.

Knowing the number of counties that are actually within your footprint will determine how often you will actually benefit by using that AVM. Ask for a list of county coverage before you even consider a test. Or, if your sample is large enough, you can derive this statistic from the test results.

Hit Rate . The hit rate is developed by dividing the number of properties where a value is returned by the total number of properties in the sample. Hit rate, like county coverage, is a key determinant of how often you will be able to take advantage of the AVM.

At this point, a high hit rate is generally between 60% to 80% when no prior sales data are provided. If your sample fairly represents your lending footprint, the aggregated hit rate will reflect your “real world” results. You may still want to review the hit rate in

the individual states and counties where you lend today and areas where you expect to expand your lending.

Accuracy . Accuracy can be measured in several ways. Consider several or all of the following accuracy measurements:

■ Percent within +/- 10% of sale price/ Percent within +/- 20 percent of sale price - This is the most basic accuracy statistic. Obviously, the smaller the range between valuation and actual sale price, the more accurate the model. This is an extremely useful comparison statistic for AVMs.

■ Standard deviation of differences - Subtract the AVM’s valuation from the actual sale price for each property valued, and develop the standard deviation of those differences. This is another indicator of the prediction ability of the AVM. A smaller standard deviation indicates AVM values that are closer to actual sales price.

■ Percentage of values within high-low range of model - Many models will return a range of values on properties. Accuracy can be measured by calculating the percentage of actual sales prices that fall within the predicted range and where they fell within the range.

Scoring the results

The test results template is an objective tool that you can use to score the results. The simplest methodology would

be to rank each of the AVMs’ performance across the key results area (1=best performance, 2=next best, etc.) and then to total the rankings. The lowest total is the winner.

Or, you can weight the rankings before totaling, depending upon the relative importance of each characteristic to you.

The final results from the above process produces a good starting point to determine which AVM vendor to use, but should not be used as the sole determinant. At a minimum, this process helps narrow your choices and gives you other things to consider:

■ What do you know about the vendor? Why would you expect the vendor to have expertise in this area? Do they have experience applying AVMs?

■ Are the data returned with the model values comprehensive, or would you still need to consult other data sources before closing the loan?

■ Does the vendor supply a confidence score with the result? Is it simple to understand and apply, or would you need to do to further analysis to determine the proper “cut points” for use?

Using a combination of an objective analytical approach with due diligence about the vendors will give your institution a good opportunity to successfully harness the power of AVM technology. **SME**