

MLS Data in AVMs: The Case For — and Against



Confidential

The recipient of this document agrees that at all times and notwithstanding any other agreement or understanding, it will hold in strict confidence and not disclose the contents of this document to any third party and will use this document for no purpose other than evaluating or pursuing a business relationship with CoreLogic. No material herein may be reproduced, in whole or in part, by any means without the expressed written consent of CoreLogic. Unauthorized distribution is strictly prohibited.

Table of Contents

MLS Data in AVMs: The Case For—and Against.....	1
The Case for Using MLS Data in AVMs	1
The Case Against Using MLS Data in AVMs	2
Finding the Solid Middle Ground	3
About CoreLogic.....	4

MLS Data in AVMs: The Case For—and Against

There's a great deal of buzz about incorporating MLS data into automated valuation models (AVMs). Some of our customers think the MLS offers fresh, comprehensive data; others view the use of MLS data as cheating. Whether it's a good idea or not depends entirely on which people you ask, their intended AVM uses, and their knowledge of AVM design. Using MLS data can provide benefits if models meet certain conditions and it can create false expectations if those conditions are not met. Understanding the challenge requires going back to the basic description of an AVM from the *Standard on Automated Valuation Models (AVMs)* issued by the International Association of Assessing Officers in September 2003:

“The distinguishing feature of an AVM is that it is an estimate of market value produced through mathematical modeling. Credibility of an AVM is dependent on the data used and the skills of the modeler producing the AVM.”

The last sentence of the definition contains the crux of the issue: “Credibility of an AVM is dependent on the data used and the skills of the modeler producing the AVM.” To that we might add that AVM integrity also rests on data validation, regular AVM blind testing, as-needed recalibration, and frequent data refresh rates—regardless of whether MLS data is used or not. The inclusion of MLS data (or any new data source) simply adds new layers of potential concern and more processes to ensure reliable results.

“MLS data” is a broad term for a wide variety of subject and comparable property data, including property characteristics, listing price, days on market and sales price. In AVM modeling, it is important to consider each of these data points independently. AVM providers must consider what data to include, how to test it, what quantities to use, and when to disregard it to create a consistent, accurate AVM. Knowing our client's intended AVM use is important too. Lenders and investors depend on AVMs for appraisal validation, fraud detection, portfolio analysis and secondary market due diligence, so the AVMs must be constructed to meet these needs.

The Case for Using MLS Data in AVMs

Any model benefits from a greater quantity of timely, accurate, relevant data. AVMs are no exception. AVM providers are constantly on the lookout for appropriate new data sources to increase the performance of our models. Traditionally, the industry has relied on AVMs that draw valuation data from county recorder's office sales records and county assessor's tax rolls, both widely accepted as reliable sources. Many commercial AVMs remain viable by relying solely on this traditional data. However, modelers must manage the systemic nature of county recorder and tax roll data including:

- ▶ There is always a lag time between the offer acceptance and sale closure and again between the closing and recording dates.
- ▶ No defined national standards for county records exist.
- ▶ In nondisclosure states, (Alaska, Idaho, Indiana, Kansas, Louisiana, Mississippi, Missouri, Montana, New Mexico, North Dakota, Texas, Utah and Wyoming) sales prices are not recorded.

By comparison, MLS data is current, available in all 50 states, and moving toward electronic standardization.

- ▶ The Real Estate Transaction Standard (RETS) that was mandated last year created a standardized data exchange system that operates in real time. MLS data is fresh. When an agent lists a property, changes a list price, or the sale closes escrow, updated information appears in the MLS as soon as the agent posts it.

- ▶ The MLS exists in every state, though sales prices are not necessarily reported in some nondisclosure states.
- ▶ MLS data commonly contains detailed information on property characteristics and condition.

In short, data from the MLS can help modelers more easily overcome some of the challenges inherent in county record data. We also find that the vast majority of MLS data is very accurate when subsequently compared against actual recorded data. Therefore, incorporating MLS data seems like a no-brainer. MLS data clearly meets the hurdle of being timely, accurate and relevant. So what's the fuss all about?

The Case Against Using MLS Data in AVMs

The main argument against using MLS data typically focuses on the list price of the subject property. This element, if given unfettered prominence in the AVM model, could skew test results and create a large gap between anticipated and actual model performance.

To understand how, it is important to know a little bit about AVM testing. Most AVM use is on properties that are not for sale — such as a second mortgage transaction or portfolio valuation — so it is important that AVM results be evaluated when the AVM does not have a list or sale price to guide it. However, a purchase price is really the best (and only) benchmark for a property's true value.

AVM testers exert considerable effort to establish a test set that evaluates the AVM against purchase transactions while ensuring that the AVM is “blind” to the purchase price. The most common blind test uses pending or recent property sales, and asks the AVM provider to also return the most recent sale information available in the vendor's database. Testing pending or recent property sales limits the extent to which the AVMs have recording data in their database. By requiring provision of the last known sale information, the few properties that *have* already been recorded can be removed from the test sample. This type of blind testing works great when recorded sale transactions are the AVM's primary data source. But what happens when MLS data — specifically list price — enters the equation? Is it still a “blind” test if the AVM knows the list price? If not, what does the test reveal?

Let's review an overly simplistic example of a bank testing an AVM to illustrate this point. Consider a hypothetical “AVM Z¹” that uses MLS data exclusively and the following formula:

1. If a list price is available in MLS for the subject property, then $AVM\ Z = \text{list price less the average list-to-sale discount in the zip code}$;
2. Otherwise, $AVM\ Z = \text{median sale price in the subject property's zip code}$.

Our bank is underwriting first mortgage applications for properties that will sell in the next month. We know the anticipated purchase price for all associated homes. We conduct a test by comparing AVM Z values against these anticipated sale prices. AVM Z returns their AVM value (likely the discounted list price in MLS) and the last known sale price (the most recent recorded sale *prior* to the current transaction.) We are thrilled to see that 95 percent of AVM Z's values are within 5 percent of our benchmark property value. AVM Z's self-reported last known sale for each of these properties is sufficiently old for us to believe that AVM Z is returning blind results.

With our blind performance study in hand, we establish policies for AVM use on second mortgage lending. We start obtaining AVM Z values on second mortgage applications. The properties are not listed for sale, so AVM Z returns the median home price in the zip code as the value for each property. We make our lending decisions based on the false belief that each AVM Z value has a 95 percent probability of being within 5 percent of the property value. How accurate are the actual AVM Z values in the second mortgage files? We don't really know. And we don't realize that we don't know. The unpredicted deviation between expected and actual accuracy could be very costly to our bank.

We see how AVM testing can be artificially optimistic if list price data is used inappropriately. If we can accurately reflect the expected AVM performance, does that solve the problem? Unfortunately, no. There's still a concern with the potential for undue influence. MLS listing data is created by real estate brokers and home sellers who have a financial interest

¹ CoreLogic® has no reason to believe that any commercial AVM is actually this basic, but we kept the example simple for the sake illustration.

in selling a property for the highest possible price. This doesn't mean that MLS data is inaccurate, just that we need to appreciate the goals of the primary MLS users if we are going to apply the data for a different purpose. Listing prices often reflect seller optimism more than true market values.

Beyond systemic optimism, AVM providers also must recognize fraudulent attempts to misprice a home. An AVM that regurgitates a list price could be manipulated simply by listing a property at a desired price point. This would be troublesome for the MLS as well as the AVM provider and the client. And it would call into question the AVM's appropriateness for use in fraud detection, appraisal review and loan funding.

Finding the Solid Middle Ground

With the pros and cons on the table, we can look at the ways AVM users and providers can take advantage of MLS data while helping clients manage risk. We believe that using MLS data in AVMs is appropriate, as long as it is managed carefully. In fact, CoreLogic has been meticulously managing certain supply-side data elements within our models for several years.

- ▶ **Data Sources** – AVM users should require AVM providers to identify whether MLS data is used in the model; whether the source is primary or secondary; and describe controls to ensure that all data is confirmed. Of course, the presence and terms of individual contracts will be confidential, but AVM providers should still be able to provide basic information.
- ▶ **Data Validation** – A solid AVM will include an estimation function when examining MLS data that determines whether the MLS data seems consistent with other data sources. If an inconsistency between MLS and public records data is too great, the MLS data should be eliminated, the confidence score reduced, or the AVM should not return an estimate.
- ▶ **Model performance** – AVM results must be tested against actual sale prices—before sale data makes it into the AVM database—and secondary transactions, such as refinance appraisals. A good AVM can accurately and consistently value properties regardless of whether the property has recently been listed for sale. At CoreLogic, we perform nightly, automated blind testing of the entire market using exhaustive data sets that provide each transaction with its own audit trail. We also routinely test our model against lender pipeline data on second mortgage and refinance transactions. AVM users should ask to see proof of blind testing and test results at specified intervals.
- ▶ **Component Data Elements** – A good AVM will consider each data element and make a conscientious decision about how to validate it and where in the model to use it. Property characteristic data from MLS can be effectively used on the subject and surrounding properties to help the model select better comps. Other fields, such as the subject property list price, are better suited for secondary model components, such as improving confidence score calculations or causing the AVM to suppress a valuation.

Should MLS data be used in an AVM? We think the answer is yes, as long as appropriate controls are in place. Reputable AVM providers want AVMs that perform predictably well across all uses. They go to great lengths to ensure that all data, including MLS data, is appropriately tested and incorporated into the model.

About CoreLogic

CoreLogic is a leading provider of consumer, financial and property information, analytics and services to business and government. The company combines public, contributory and proprietary data to develop predictive decision analytics and provide business services that bring dynamic insight and transparency to the markets it serves.

CoreLogic has built the largest and most comprehensive U.S. real estate, mortgage application, fraud, and loan performance databases and is a recognized leading provider of mortgage and automotive credit reporting, property tax, valuation, flood determination, and geospatial analytics and services.

More than one million users rely on CoreLogic to assess risk, support underwriting, investment and marketing decisions, prevent fraud, and improve business performance in their daily operations. Formerly the information solutions group of The First American Corporation, CoreLogic began trading under the ticker CLGX on the NYSE on June 2, 2010. The company, headquartered in Santa Ana, Calif., has more than 10,000 employees globally with 2009 revenues of \$2 billion. For more information visit www.corelogic.com

CoreLogic
4 First American Way
Santa Ana, CA 92707

FOR MORE INFORMATION PLEASE CALL 1-866-774-3282

© 2010 CoreLogic

CORELOGIC is a registered trademark of CoreLogic
Proprietary and confidential. This material may not be reproduced in any form without expressed written permission.

17-MLSAVM-1010-00



CoreLogic®

corelogic.com