Principles for Calculating AVM Performance Metrics

by Hans R. Isakson, PhD, Mark D. Ecker, PhD, and Lee Kennedy

Abstract

An analysis of 5.3 million housing sales suggests that there are fundamental shortcomings with how automated valuation model (AVM) vendors currently calculate their AVM performance metrics, in particular the forecast standard deviation. The analysis demonstrates that the methodology used to calculate the values of performance metrics meaningfully impacts an AVM's credibility. This article proposes consistent methodologies to calculate AVM performance metrics that comply with well-established appraisal principles and allow a consistent evaluation and comparison of AVM performance. A case study's research AVM empirically illustrates that not following these principles yields overly optimistic AVM performance metric values.

Introduction

An automated valuation model (AVM)[1] is a computer software program that produces an estimate of the market value, called an *AVM valuation*, of a subject property given (1) the address of the subject property, and (2) property sales and characteristics data. AVM vendors blend many property transactions, acquired from public sources or data aggregators, with one or more valuation models, acquired from academic and professional publications or developed by their own analysts, into a product called an AVM, the details of which are a closely guarded trade secret. An AVM produces a valuation along with certain statistics, called *AVM performance metrics*, that assess the validity, accuracy, and precision of the AVM valuation. The focus of this article is AVM performance metrics.

Two recent events have made evaluation of overall AVM performance increasingly important. First, the Interagency Appraisal and Evaluation Guidelines require, among other things, that lending institutions independently assess the reliability of the AVMs they use.[2] Second, the Federal Deposit Insurance Corporation (FDIC), the Board of Governors of the Federal Reserve System, and the Office of the Comptroller of the Currency, have jointly increased the *de minimis* threshold, from $250,000 to $400,000, for residential real estate transactions that do not require an appraisal with a physical inspection of the property and neighborhood.[3] As a result, lenders will be allowed to make more residential mortgages secured by properties that are valued utilizing an AVM rather than a traditional appraisal.

Due to the proprietary intellectual property contained within an AVM, assessing AVM credibility, i.e., its validity, accuracy, and precision, is accomplished through an examination of the AVM's performance metrics.[4] Typically, users of

1. Throughout this work, the term "AVM" will be used to refer to commercial or professional-grade AVMs that value residential properties. That is, AVMs whose output is sold by AVM vendors to clients, in contrast to consumer-facing AVMs that typically provide output free of charge. See Valuation Analytics Workgroup, *The State of Automated Valuation Models in the Age of Big Data* (Mortgage Bankers Association (MBA), January 2019), 9–10; downloadable using story link at http://bit.ly/2HAhtvg.
2. *Interagency Appraisal and Evaluation Guidelines* (December 2, 2010), available at http://bit.ly/InteragencyGuidelines.
3. See Financial Institution Letter FIL-53-2019, "New Appraisal Threshold for Residential Real Estate Loans" (September 27, 2019), available at http://bit.ly/FIL-53-2019.
4. The terms "AVM performance metric(s)" and "performance metric(s)" are used interchangeably.

AVMs are dependent upon AVM vendors to provide reliable performance metrics, for example, the *forecast standard deviation* (FSD).[5] However, as Kane, Linné, and Johnson state, "Third-party verification is critical."[6] These third parties, including credit rating agencies (such as Fitch, Standard and Poor's, and Moody's) and independent AVM testing firms (such as AVMetrics), assess AVM reliability using performance metrics.

The purpose of this study is first to demonstrate that the calculation of performance metrics is not standardized across the AVM industry or AVM vendors, which can result in AVM vendors underreporting their FSDs. Second, five best-practice principles are recommended for AVMs, and a supporting statistical procedure is presented to implement these principles. The discussion explains how these steps would bring AVMs into better alignment with current appraisal practices. Moreover, if these principles are respected, then the values of the performance metrics associated with any model would be directly comparable to those of another model. The case study demonstrates that not following the valuation principles can result in an overly optimistic assessment of an AVM's performance. Consequently, it is recommended that AVM vendors adopt the valuation principles and that users of AVMs request conformity with these principles.

Review of the Literature

Most of the literature regarding AVM performance metrics appears in unpublished manuscripts,[7] self-published books,[8] industry websites,[9] or recent trade publications.[10] Exhibit 1 contains a list of common performance metrics, along with a glossary of abbreviations and definitions related to AVM performance metrics. For example, Gayler et al. recognize mean percentage sales error, mean absolute percentage sales error, FSD, and hit rate as important metrics for the evaluation of the performance of an AVM.[11] The Collateral Risk Management Consortium suggests using percentage sales errors, mean percentage sales error, and error buckets to assess AVMs.[12] CoreLogic recommends evaluating AVMs using the mean percentage sales error, median percentage sales error, FSD, and error buckets.[13] AVMetrics advocates that no more than 10% of AVM valuations should be more than 20% higher than their corresponding selling prices, suggesting a right tail 20% performance metric.[14] Kirchmeyer and Staas state that median absolute percentage (sales) errors (MAPEs) of less than 10% "are indicative of a strong AVM, while those ranging from 11% to 15% might also be acceptable for some lending programs."[15]

Error buckets, also called percent (predicted) error (PE) buckets, count the number of sales that are deemed accurate (i.e., the success rate of

5. The term "FSD" was originally coined by the Federal Home Loan Mortgage Corporation (Freddie Mac) for use with its Home Value Explorer AVM in the late 1990s to early 2000s. See Exhibit 1 for a list of common performance metrics.
6. M. Steven Kane, Mark R. Linné, and Jeffrey A. Johnson, *Practical Applications in Appraisal Valuation Modeling* (Appraisal Institute: Chicago, 2004), 171.
7. Peter Rossini and Paul Kershaw, "Automated Valuation Model Accuracy: Some Empirical Testing" (14th Pacific Rim Real Estate Society Conference, January 2008), available at http://bit.ly/38Ywpzr; and AVMetrics, *Automated Valuation Model (AVM) Tests* (2018).
8. James Kirchmeyer, "AVMs 101: A Guide to Automated Valuation Models" (Real Info, 2004); James Kirchmeyer and Peter Staas, "AVMs 201: A Practical Guide to the Implementation of Automated Valuation Models" (2008).
9. Veros, "Veros Confidence Score" (2019); Freddie Mac, "Confidence Levels" (2019), available at http://bit.ly/freddiemac_confidencescores.
10. International Association of Assessing Officers, *Standard on Automated Valuation Models (AVMs)* (July 2018), available at http://bit.ly/IAAO_AVM; MBA Valuation Analytics Workgroup, *State of Automated Valuation Models*.
11. Ross Gayler, Debashree Sanyal, Roy Pugh, and Siân King, *Best Practice Validation and Comparison for Automated Valuation Models (AVMs)* (CoreLogic, October 2015), available at http://bit.ly/2HYvbbe.
12. Collateral Risk Management Consortium, *The CRC Guide to Automated Valuation Model (AVM) Performance Testing* (2003), available at http://bit.ly/2Pq8XTU.
13. CoreLogic, *Automated Valuation Model Testing* (2011), available at http://bit.ly/AVMtesting.
14. AVMetrics, *Automated Valuation Model (AVM) Tests*, 25.
15. Kirchmeyer and Staas, "AVMs 201."

the AVM) at a given level of precision, typically ± 5%, 10%, 15%, and 20%.[16] In the study presented in this article, the notation PExx is used to refer to a specific error bucket, at a given ± (xx) percentage. For example, PE10 represents the ± 10% error bucket. Kirchmeyer originally suggested a success rate that at least 50% of AVM valuations should be within ± 10% of selling prices.[17] That is, the (percentage) success rate of an AVM at PE10 should be at least 50%. More recently, the Mortgage Bankers Association reported that "[a]lmost all counties in the United States experience [PE10] rates north of 70 percent,"[18] suggesting a success rate at PE10 of 70% or more.

An AVM's failure rate in a given error bucket is the complement of the AVM's success rate within that error bucket. The failure rate is a concept common in engineering, where it is defined as the frequency with which a component fails.[19] The failure rate concept is also found in other fields where the process fails to perform well, such as the percent of small business failures,[20] the percent of students failing a computer programing course,[21] hotel failures,[22] and commercial banks insolvencies.[23] In the study presented here, which focuses on sales where the AVM fails to accurately predict selling prices, the *failure rate* of an AVM in a particular error bucket (e.g., PE10) is defined as the frequency (percentage) with which an AVM *fails* to predict the value of a target property within the tolerance given by the error bucket (e.g., ± 10%).[24]

In addition, AVM vendors typically provide a confidence score, "which is often interpreted as meaning that the AVM estimate is within plus or minus 10% of the 'true' market value of the property with a high degree of confidence."[25] However, the definition and use of a confidence score are not standardized across AVM vendors.[26] For example, Veros describes its confidence score as a measure of accuracy between zero and 100 for which each decile generally corresponds to a 5% variance.[27] Realtors Property Resources uses an RVM confidence score of zero to five stars.[28] CoreLogic's PASS produces a confidence score between 60 and 100 that measures how well "sales data, property information, and comparable sales support the property valuation process."[29] Gordon states that a confidence score may or may not

16. Kirchmeyer, "AVMs 101"; R. Slump and A. Arora, "Property Valuations as a Part of the Rating Process, Fitch Ratings" (CRN Presentation, Las Vegas, NV, 2019); Veros, "Confidence Score: Exceeding Expectations with Meaningful Metrics," available at http://bit.ly/ConfidenceScore.

17. Kirchmeyer, "AVMs 101."

18. MBA Valuation Analytics Workgroup, *State of Automated Valuation Models*, 28.

19. Maxim Finkelstein, "Introduction," in *Failure Rate Modelling for Reliability and Risk*. (Springer Series in Reliability Engineering, 2008), 1–84.

20. John Watson, and Jim E. Everett, "Do Small Businesses Have High Failure Rates? Evidence from Australian Retailers," *Journal of Small Business Management* 34, no. 4 (1996): 45–62.

21. Jens Bennedsen and Michael E. Caspersen, "Failure Rates in Introductory Programming," *SIGCSE Bulletin* 39, no. 2 (2007): 32–36.

22. Paul Ingram and Joel A. C. Baum, "Chain Affiliation and the Failure of Manhattan Hotels, 1898–1980," *Administrative Science Quarterly* 42, no. 1 (March 1997): 68–102.

23. Adam B. Ashcraft, "Are Banks Really Special? New Evidence from the FDIC-Induced Failure of Healthy Banks," *American Economic Review* 95, no. 5 (December 2005): 1712–1730.

24. The failure rate of an AVM is not exclusively limited to the ± 10% error bucket. It can be used at any error bucket, for example, the failure rate ± 5%, 10%, 15%, or 20%. Also see Mark D. Ecker, Hans R. Isakson, and Lee Kennedy, "An Exposition of AVM Performance Metrics," *Journal of Real Estate Practice and Education* (forthcoming 2020); an earlier version of this research is available at http://bit.ly/2PCIemW.

25. James R. Follain and Barbara A. Follain, "AVMs Have Feelings, Too" (FI Consulting, September 4, 2007), available at http://bit.ly/2SFqJEK.

26. The confidence score of an AVM should not be confused with the confidence interval (level) of a statistical estimate. A confidence score most generally informs the level of the "AVM provider's confidence in the estimated values," which may or may not involve a confidence interval. See CoreLogic, "AVM FAQs" (2014), available at http://bit.ly/393dxzd.

27. Veros, "Veros Confidence Score."

28. See Realtors Property Resources, "What Is an AVM or RVM® Confidence Score?" (2018), available at http://bit.ly/RPRconfidencescore. The Realtors Property Resources property database is exclusive to the National Association of Realtors.

29. CoreLogic, "PASS® Automated Valuation Model" (2017), http://bit.ly/2PtPlst. Also see CoreLogic, "Forecast Standard Deviation and AVM Confidence Scores" (2017), http://bit.ly/3acguO9.

be related to the FSD and that "[s]uch a confusion of [confidence] scores and lack of connection to statistical performance in actual use forces lenders to guess at their risk management."[30]

For each individual target property being valued, AVM vendors may also report the target property's FSD.[31] Gayler et al. define an *FSD* as "the standard deviation of the percentage error, where the percentage error describes the relative difference between [AVM] valuation and price."[32] Freddie Mac qualifies the value of the FSD generated from its Home Value Explorer (HVE) AVM as high, medium, or low confidence. "High confidence" requires an FSD of 13 or less. "Medium confidence" arises from an FSD between 13 and 20, while "low confidence" occurs for valuations with an FSD greater than 20.[33]

Reporting of the FSD by AVM providers is ubiquitous; however, the FSD description is not standardized across the industry. For example, CoreLogic states that "[t]he FSD is a statistic that measures the likely range or dispersion an AVM estimate will fall within, based on the consistency of the information available to the AVM at the time of estimation."[34] Matysiak writes that the FSD is an "estimate of the amount of variation that can occur between the actual sales price and the forecast (the most probable market value) made by the AVM."[35] Gordon offers another definition, describing the FSD as "an AVM value's expected (forecasted) proportional standard deviation around actual subsequent sales price for the given property value estimate."[36]

The clearest mathematical definition of the FSD is that it is the standard deviation of the percentage sales errors for a collection of valuations.[37] However, the method of calculating an FSD for an *individual target property* is not consistent, meaning that it is not clear how an AVM provider is using the sampling distribution and/or parsing a data set to provide a unique FSD value for any one particular target property.

An AVM report typically contains a high/low range of value based on a ± 1 × FSD confidence interval around the AVM valuation.[38] This 1 × FSD interval is often interpreted by assuming that the underlying sales errors are normally distributed. Under normality, an AVM vendor has 68.26% confidence that the true market value of the target property lies within ± one FSD of the AVM valuation, or 95% confidence that the market value of the target property falls within ± 1.96 × FSD of the AVM valuation.[39] The assumption of normality allows the client to use the FSD-based confidence interval to test hypotheses regarding the market value of the target property. Therefore, it behooves the AVM vendor to test the distribution of percentage sales errors for normality before rendering any FSD-based inference, including a high/low value range.

There are a number of additional studies using performance metrics. Such studies include one by Clapp and O'Connor, who report the mean absolute percentage sales error and its 25th, 50th, and 75th percentiles to evaluate seven different

30. Douglas Gordon, "Metrics Matter" (Thomson Corp. and National Mortgage News, 2005), available at Freddie Mac http://bit.ly/2vkFsMf.

31. CoreLogic, "AVM FAQs."

32. Gayler et al., *Best Practice Validation and Comparison for AVMs*.

33. Freddie Mac, "Confidence Levels" (2020), available at http://bit.ly/freddiemac_confidencescores.

34. CoreLogic, "Forecast Standard Deviation and AVM Confidence Scores" (2017), available at http://bit.ly/3acguO9.

35. George A. Matysiak, *The Accuracy of Automated Valuation Models (AVMs)* (Brussels, Belgium: The European Group of Valuers' Association, May 2017), 7, available at http://bit.ly/2PTnQyp.

36. Gordon, "Metrics Matter," 1.

37. Gayler et al., *Best Practice Validation and Comparison for AVMs*, 5. The *FSD* definition by Gayler et al.—"the standard deviation of the percentage sales errors"—is used for all FSD calculations performed by the authors in this work.

38. CoreLogic, "Forecast Standard Deviation and AVM Confidence Scores."

39. See Appendix A, "Statistical Tables," in Neil A. Weiss, *Introductory Statistics*, 10th ed. (Pearson, 2016), A2–3.

Exhibit 1 Glossary of Common AVM Performance Metrics

AVM Performance Metric	Abbreviation	Definition (Source)
Coefficient of Dispersion	COD	The average percentage deviation of the AVM's valuation-to-sales price ratios from the median AVM valuation-to-sales price ratio (Pokryshevskaya and Antipov, 2011)
Coefficient of Variation	COV	The standard deviation divided by the mean AVM valuation-to-price ratio (Kane, Linné, and Johnson, 2004)
Confidence Score	None	A value that indicates the AVM vendor's confidence in its AVM valuation (CoreLogic, 2014)
Failure Rate	None	The complement of the PE bucket (Ecker, Isakson, and Kennedy, 2020)
Forecast Standard Deviation	FSD	The standard deviation of a set of percentage sales errors (Gayler et al., 2015)
Hit Rate	None	The percentage of properties for which an AVM returns a value (MBA, 2019)
Mean Percentage Sales Error	MPE	The mean of a set of percentage sales errors (CoreLogic, 2011)
Median Absolute Percentage Sales Error	MAPE	The median of a set of absolute percentage sales errors (Kirchmeyer and Staas, 2008)
Median Percentage Sales Error	None	The median of a set of percentage sales errors (CoreLogic, 2011)
Percentage Predicted Error Bucket	PE%	The percentage of AVM valuations within a specified ± percentage of selling prices (Kirchmeyer, 2004; CoreLogic, 2011)
Percentage Sales Error	None	The AVM valuation minus its selling price, for a target property, which is then divided by the selling price (CRC, 2003)
Price-Related Difference	PRD	The mean valuation-to-selling price ratio divided by the weighted (by selling prices) mean ratio (IAAO, 2018)
Right Tail 20%	None	The percentage of AVM valuations more than 20% higher than their corresponding selling prices (AVMetrics, 2018)

valuation models.[40] Pokryshevskaya and Antipov use mean average percentage sales error and the coefficient of dispersion (COD), which is the average percentage deviation of the median sales error.[41] Rossini and Kershaw describe several performance metrics for which they also suggest performance thresholds.[42] They report the mean absolute percentage sales error, FSD, and COD, in addition to three error buckets: PE10, PE15, and PE20. Kane, Linné, and Johnson suggest using the COD, together with the coefficient of variation (COV), both of which assess horizontal equity.[43] Lastly, the IAAO advocates an additional metric, namely, the price-related difference (PRD), which measures the vertical equity of the property tax system.[44]

AVM Valuation Performance Metrics Example

To illustrate the calculations of these performance metrics, a research AVM was constructed, labeled the test valuation model (TVM). The TVM is a regression model containing fifteen housing characteristics employed as independent variables.[45] The purpose of this analysis is to illustrate that performance metrics are sensitive to their calculation methodology and to additionally show how these metrics change when applying several best-practice principles.

The following analysis demonstrates how different statistical methodologies, using the same valuation model and the same data set, result in different performance metric values. To start, a base case of performance metrics is calculated using *internally* fitted (regression) valuations. The data set employed in this empirical demonstration consists of 53 housing sales in 2012, located in a submarket of Cedar Falls, Iowa. The properties' locations are denoted with individual circles in Exhibit 2. For the analysis, the first house to sell in 2013 was arbitrarily chosen as the target property, which is indicated by the filled box (■) in Exhibit 2. The TVM uses the 53 sales to produce a valuation of $159,427 as of January 1, 2013, for the target property. The TVM performance metrics associated with the target property are calculated from the 53 comparable property sales based on the internally generated predicted values from the regression; the performance metrics are reported in Exhibit 3.

The TVM performance metrics in Exhibit 3 are indicative of an acceptable AVM. The TVM performs well with regard to the Rossini and Kershaw thresholds, and the Kirchmeyer error buckets, and the TVM's failure rate (at ± 10%) is 39.6%. With an FSD of 13.4, the TVM earns a "medium confidence" score based on Freddie Mac's thresholds (and it is only 0.5 away from attaining a "high confidence" score). Only 4 of the 53 comparable sales (7.5%) have TVM values more than 20% larger than their respective selling prices.[46] Lastly, the TVM has a PRD value of 1.0156, which suggests that it is slightly overvaluing inexpensive houses, more so than it undervalues expensive houses, but the model performs reasonably well for houses close to the median and mean selling prices.

40. John M. Clapp and Patrick M. O'Connor, "Automated Valuation Models of Time and Space: Best Practice," *Journal of Property Tax Assessment & Administration* 5, no. 2 (2008): 57–67.

41. Elena B. Pokryshevskaya and Evgeny A. Antipov, "Applying a CART-Based Approach for the Diagnostics of Mass Appraisal Models," *Economics Bulletin* 31, no. 3 (2011): 1–8; and IAAO, *Standard on Ratio Studies* (Kansas City, MO: International Association of Assessing Officers, April 2013), available at http://bit.ly/2PtgUaT.

42. Rossini and Kershaw, "Automated Valuation Model Accuracy."

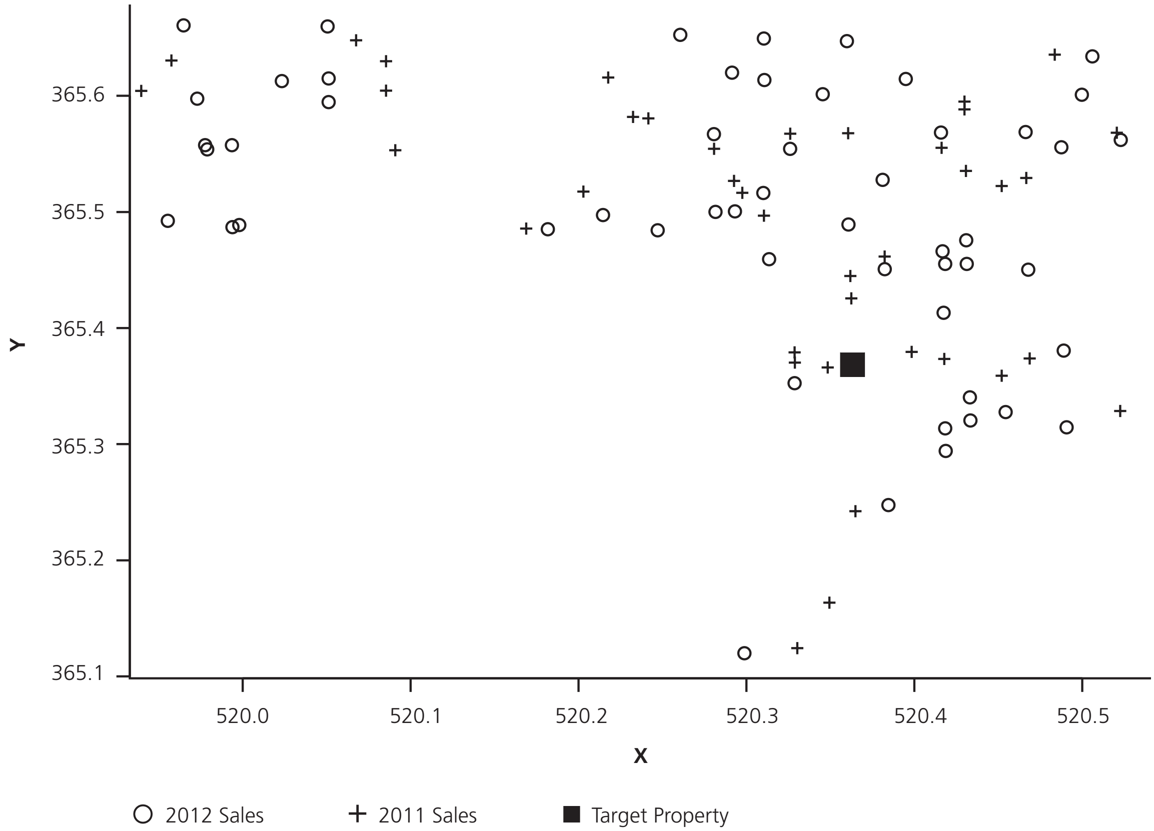
43. Kane, Linné, and Johnson, *Practical Applications in Appraisal Valuation Modeling*. Horizontal equity is the notion that people in the same circumstances should be treated the same or that similar properties should have similar tax assessed values. Following the IAAO's *Standard on Automated Valuation Models (AVMs)*, both the COD and COV assess horizontal equity as they measure spread of AVM valuation to selling price ratios about the center (mean ratio for COV and median ratio for COD).

44. IAAO, *Standard on Automated Valuation Models (AVMs)*. Vertical equity means that higher-valued properties have higher assessed values than lower-valued properties. For this analysis, following IAAO's *Standard on Automated Valuation Models (AVMs)*, vertical equity is assessed through the PRD statistic by comparing AVM value to selling price ratios for the most and least expensive houses.

45. See Chapter 8 in Kane, Linné, and Johnson, *Practical Applications in Appraisal Valuation Modeling*, and Ecker, Isakson, and Kennedy, "Exposition of AVM Performance Metrics," for additional details about the model.

46. This is within AVMetrics' right tail 10% suggested threshold; see AVMetrics, *Automated Valuation Model (AVM) Tests*.

Exhibit 2 Location of House Sales in Cedar Falls, Iowa



Note: The X-Y coordinates are State Plane Coordinates, Iowa North, NAD 1983, where each unit represents 10,000 feet.

The TVM regression-based predicted values for each of the 53 comparable sales, which produce the performance metrics seen in Exhibit 3, provide internal measures of model performance. That is, all 53 housing sales that produced the target property's valuation are re-used to determine each comparable sale's valuation. As such, the resulting metrics in Exhibit 3 tend to be overly optimistic, compared to the prediction of a new, external-from-the-model observation.[47] Although the internal calculations yield favorable performance metric values, an AVM should use an *external*, cross-validation methodology to judge how well the model predicts market values for housing sales that were *not* used to construct the model. A leave-one-out (LOO) cross-validation procedure is recommended. The LOO procedure removes each sale in the original data set, one at a time, and generates a valuation for that left-out property from the remaining $n - 1$ sales.[48] This process is repeated until each property in the original data set has been valued. Consequentially, under a LOO validation methodology, each house will never be used in the model to (indirectly) value itself.

47. "A result of this model development process is that the error mean square MSE will tend to understate the inherent variability in making future predictions from the selected model." John Neter, Michael H. Kutner, Christopher J. Nachtsheim, and William Wasserman, *Applied Linear Statistical Models*, 4th ed. (McGraw-Hill/Irwin, 1996), 435.

48. Another common validation technique is a *k*-fold analysis, which is a mass appraisal technique that cycles through valuing 100(*k*/*n*) percent of the data. The leave-one-out (LOO) process is a *k*-fold analysis that values one observation at a time.

Exhibit 3 TVM Performance Metrics

AVM Metric	Value
Mean Sales Error	$1,014
Mean Percentage Sales Error (MPE)	0.84%
Median Sales Error	$–1,832
Median Percentage Sales Error	–1.51%
Mean Absolute Sales Error	$13,788
Median Absolute Sales Error	$10,611
Mean Absolute Percentage (Sales) Error	9.8%
Median Absolute Percentage (Sales) Error (MAPE)	6.5%
FSD	13.4
Percentage of Estimates within ± 10% (PE10)	32/53 for 60.4%
Failure Rate at ± 10%	21/53 for 39.6%
Percentage of Estimates within ± 15% (PE15)	40/53 for 75.5%
Percentage of Estimates within ± 20% (PE20)	47/53 for 88.7%
Percentage of Estimates more than 20% (Right Tail 20%)	4/53 for 7.5%
Coefficient of Variation (COV) of TVM/Sale Price	0.13317 or 13.3
Coefficient of Dispersion (COD) of TVM/Sale Price	9.86
Regression *R*-Squared (Coefficient of Determination)	0.7272
Adjusted *R*-Squared	0.6165
PRD of TVM/Sale Price	1.0156
Mean Selling Price of 53 Sales	$143,767
Median Selling Price of 53 Sales	$130,000
Mean TVM Valuation for 53 Sales	$143,062
Median TVM Valuation for 53 Sales	$138,650

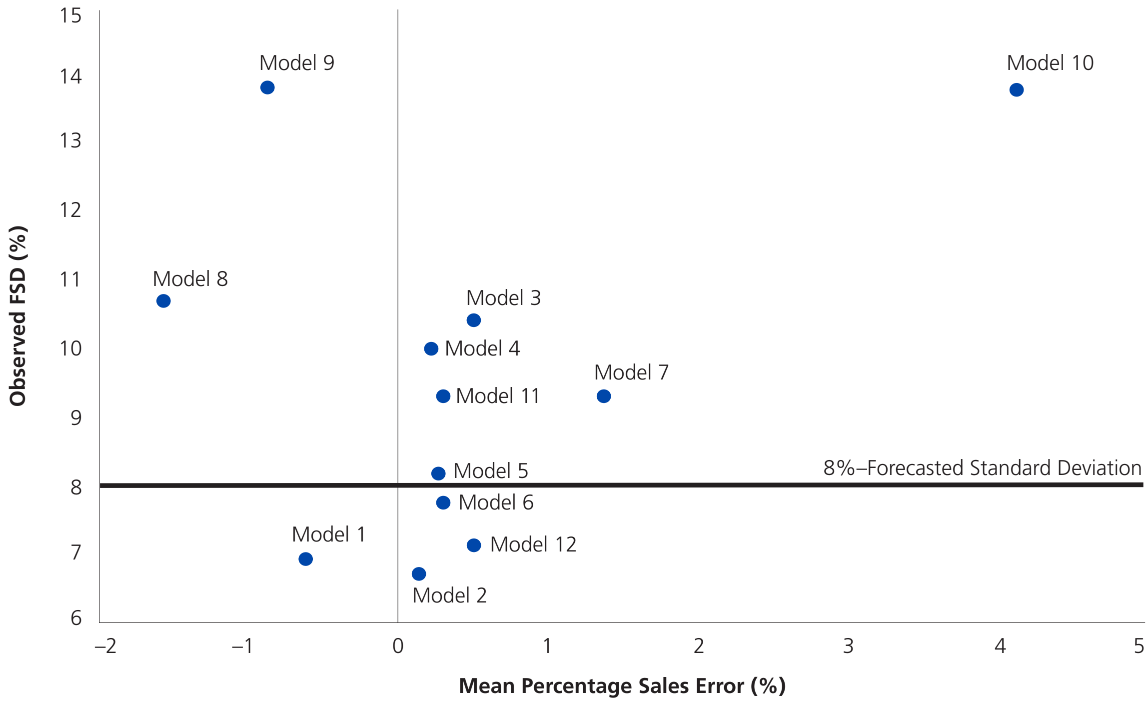
Lack of Standardization of AVM Performance Metrics

Next, an analysis was constructed to empirically illustrate that current methodologies used by AVM vendors to (self-) report performance metrics—in particular, the FSD—for an individual target property are not consistent. Due to the inconsistency, it is not clear how an AVM parses a data set of housing sales to provide a unique value (for example, the FSD) for an individual target property.[49]

An analysis that looks at the accuracy and precision of AVMs by aggregating housing sales to a common, vendor-reported FSD is called an "AVM-by-FSD" analysis. This analysis allows AVM testers to corroborate a vendor-reported FSD with an observed FSD. Exhibit 4 shows the performance of fourteen AVMs, whose identities are removed for confidentiality, involving 683,802 properties, where the FSD for each of these properties had a vendor-reported value of 8 (the horizontal line in Exhibit 4). In Exhibit 4,

49. Some third-party companies independently test commercial AVMs. For example, AVMetrics provides participating AVM vendors with sales of hundreds of thousands of test properties on a biweekly basis. Vendors are allowed 48 hours from receipt of the test properties to return AVM valuations, along with an FSD, for each test property. Each test property provided has passed quality checks to ensure that its sale price is representative of market value (i.e., an arm's-length transaction) and is unknown to the vendor. AVMetrics then compares each returned AVM valuation to its corresponding selling prices to calculate the AVM's performance metrics, including the observed FSD, which is calculated using all housing sales with a common vendor-reported FSD. AVMetrics uses a proprietary analysis to uncover whether or not the property's selling price is truly unknown to the AVM that is valuing it; this includes, but is not limited to, having the vendor report the last known selling price and date for each test property. See AVMetrics, *Automated Valuation Model (AVM) Tests*, 3.

Exhibit 4 AVM Performance for 683,802 Target Properties Valued by Fourteen* AVMs, Each with a Self-Reported FSD of 8



* Two AVMs produce nearly identical results for their mean percentage sales errors and the observed FSDs (Models 11 and 13; Models 7 and 14), and as a result, only twelve AVMs are presented in this exhibit.

Models 5 and 6 are the AVMs with observed FSDs (8.1 and 7.7, respectively) closest to their self-reported value of 8.

As seen in Exhibit 4, a majority of the fourteen AVMs analyzed underreport their FSDs, with eight models having their observed FSDs above 9 (when it should be 8). This suggests that the vendor-reported FSD of 8 is an overly optimistic assessment of AVM precision for the majority of AVMs examined. Most notable in Exhibit 4 are Models 8, 9, and 10, which are the AVMs with the highest observed FSDs (10.7, 13.9, and 13.8, respectively) and deviate by more than 60% from the vendor-reported FSD value of 8.

The underreporting of the FSD is not unique to those test properties with a vendor-reported FSD of 8. Exhibit 5 shows the performance for most of the 327 total AVM/FSD combinations, each with at least 100 hits, which results in over 5.3 million housing sales. In Exhibit 5, 138,118 properties, each with a vendor-reported FSD of 12, had a calculated or observed FSD of 20.4 (averaged over fifteen AVMs). That is, the vendor-reported FSD of 12 is underreported by 8.4 or by 70.3% on average.

Inspecting Exhibit 5, AVMs are consistently underreporting their FSDs, with the best performing having vendor-reported FSDs of 7 and 8. Overall, for the 327 total AVM/FSD combinations, 77.1% (252 of the 327) have underreported their FSDs. Moreover, the average FSD difference (observed FSD minus vendor-reported FSD) for all 327 is 4.3 or, on average, a 54.9% underreporting percentage computed using all 5.3 million sales. In addition, only 25 AVM/FSD combinations (7.6% of the 327) have their observed FSD within ± 10% of the vendor-reported FSD, which suggests a lack of consistency regarding the FSD definition and/or how AVM providers calculate an FSD for each target property. These findings are not a trivial matter, as overly optimistic vendor-reported FSDs make AVMs appear substantially more reliable than they actually are.

AVM providers should adopt and employ uniform methodologies when calculating perfor-

Exhibit 5 AVM Performance by Vendor-Reported FSDs for 5.3 Million Target Properties Valued by as Many as Fifteen AVMs

Vendor-Reported FSD	Number of AVMs	Hits	Observed FSD	FSD Difference: Observed FSD minus Vendor-Reported FSD	FSD Difference (%)
1	5	29,076	9.0	8.0	802.0
2	7	69,671	7.1	5.1	375.0
3	9	406,350	7.6	4.6	152.6
4	12	620,462	7.4	3.4	83.8
5	14	870,006	7.1	2.1	42.7
6	15	485,109	8.0	2.0	33.4
7	14	715,546	8.3	1.3	19.0
8	14	683,802	9.5	1.5	18.5
9	14	305,621	12.5	3.5	38.9
10	15	230,958	15.5	5.5	54.7
11	15	155,956	19.7	8.7	78.8
12	15	138,118	20.4	8.4	70.3
13	15	120,975	22.1	9.1	70.0
14	15	84,143	21.8	7.8	55.4
15	15	74,297	23.7	8.7	58.1
20	12	79,238	27.1	7.1	35.3
25	5	7,475	28.3	3.3	13.4
30	4	3,980	32.0	2.0	6.8
40	1	530	24.9	–15.1	–37.8
Total	327	5,344,833	19.6	4.3	54.9

mance metrics, in particular the FSD and percentage sales errors, to help mitigate the issues mentioned. Moreover, lenders who use FSDs should demand that vendors align their statistical calculation methodologies to comply with well-established appraisal principles, such as those set forth in the next section.

Valuation Principles for AVMs

It is suggested that AVMs incorporate four valuation principles that appraisers already observe when performing traditional appraisals. This would be a logical step for AVM vendors since both an appraiser and an AVM have the common purpose of valuing exactly one target property. Moreover, these principles, if implemented, would provide greater uniformity in assessing the reliability of AVMs when (1) building a valuation model that uses comparable sales to estimate the target property's market value, and (2) constructing a data set to compute performance metrics, which may be different from the original data set of comparable sales. The following discusses each valuation principle recommended for implementation and then illustrates the effect of the application of each principle by recalculating the performance metrics shown in Exhibit 3.

The first valuation principle that should be incorporated into the AVM-building and validation process is the principle that **comparable properties should be very close substitutes for the target property**. Appraisers understand the concept of comparable properties, but it is difficult to build an AVM that can select a set of comparable properties as well as a well-trained appraiser. For all practical purposes, comparable properties are essentially equivalent to each other (and the target property), as these comparable properties and the target property compete for the same set of buyers.[50] Much appraisal and academic literature addresses the issue of substitutability among properties using submarkets. For example, Schnare and Struyk, Palm, and Watkins use predefined submarkets, such as ZIP codes, census block groups, school districts, etc. from which comparable properties are selected.[51] Bourassa et al., Goodman and Thibodeau, Tu, and Isakson and Ecker[52] use statistical analyses to allow the housing sales data itself to identify (possibly non-contiguous) submarkets containing similar properties.

The second valuation principle that should be incorporated into the AVM-building and validation process is that **a property should never be used to value itself**. Obviously, the sale of a target property should never be included in the original data set that values itself. Again, appraisers would almost never violate this principle. However, if the target property has been recently flipped (bought and resold within a few months), then it is possible for the sale of the target property to be included in the original data set used by an AVM. The more common violation of this principle occurs indirectly, when the valuation model uses the coefficients that it calculated from all properties, including the property being valued, to estimate that property's market value.

Comparable sales should (1) never include the target property, even indirectly, (2) be selected from the same submarket as the target property, and (3) be sold as close in time to the valuation date as possible, but never after it. Therefore, the third valuation principle for AVMs is that **the data set should contain no sales that post-date the valuation date of the target property**. That is, sales that occur after the valuation date should not be used to value the target property.[53] Doing so produces what Thanos, Dube, and Legros call "arrow of time" violations.[54]

To avoid arrow of time violations, the data set should not include either the entire post-dated sale itself or any individual variable that is post-dated. Mixing and matching pre- and post-dated variables for an individual house is especially egregious, because housing characteristics can change over time, for example, because of remodeling. A common violation of the arrow of time principle occurs when the assessed value, in a tax assessed value AVM, does not comport to the date of sale for a comparable house and/or the valuation date for the target property.

The fourth valuation principle that should be incorporated into AVMs is another time-related principle, which provides that **sales of comparable properties should span similar market conditions**. That is, the comparable sales chosen should have sold during market conditions similar to those that

50. In applying the sales comparison approach, "The goal is to find a set of comparable sales ...as similar as possible to the subject property to ensure they reflect the actions of similar buyers." Appraisal Institute, *The Appraisal of Real Estate*, 14th ed. (Chicago: Appraisal Institute, 2013), 381. Comparable sales "are located in the same area and are very similar in size, condition and features" as the target property. Brendon DeSimone, "What Are Comps? Understanding a Key Real Estate Tool" (Zillow, July 13, 2015), available at http://bit.ly/ZillowComps.

51. Ann B. Schnare and Raymond J. Struyk, "Segmentation in Urban Housing Markets," *Journal of Urban Economics* 3, no. 2 (1976): 146–166; Risa Palm, "Spatial Segmentation of the Urban Housing Market," *Economic Geography* 54, no. 3 (1978): 210–221; and Craig A. Watkins, "The Definition and Identification of Housing Submarkets," *Environment and Planning* 33, no. 12 (2001): 2235–2253.

52. Steven C. Bourassa, Foort Hamelink, Martin Hoesli, and Bryan D. MacGregor, "Defining Housing Submarkets," *Journal of Housing Economics* 8, no. 2 (1999): 160–183; Allen C. Goodman and Thomas G. Thibodeau, "Housing Market Segmentation and Hedonic Prediction Accuracy," *Journal of Housing Economics* 12, no. 3 (2003): 181–201; Yong Tu, "Segmentation, Adjustment and Disequilibrium," in *Housing Economics and Public Policy*, ed. Tony O'Sullivan and Kenneth Gibb (Oxford, UK: Blackwell Science, 2003); and Hans R. Isakson and Mark D. Ecker, "The Influence of Leaking Underground Storage Tanks on Nearby House Prices," *Journal of Economic Insight* 44, no. 1 (2018): 45–67.

53. The calibration and validation of retrospective AVMs can easily include sales that post-date the valuation date. In addition, using the internal residuals from a regression analysis as sales errors, as demonstrated in Exhibit 3, would also include some predicted values based upon post-dated sales.

54. Sotirios Thanos, Jean Dubé, and Diègo Legros, "Putting Time into Space: The Temporal Coherence of Spatial Applications in the Housing Market," *Regional Science and Urban Economics* 58 (May 2016): 78–88.

existed as of the target property's valuation date. Of course, all comparable sales must occur on or before the valuation date, but the question of how old the comparable sales can be requires knowledge of the local market conditions.[55]

As a general rule, it is easier to control for changing market conditions by selecting comparable properties closer in time to the valuation date than to build time-related trends into a valuation model. It is difficult, at best, to decide the appropriate statistical methodology to account for changing market conditions. It is simpler to choose comparable sales that occur closer in time to the target property's valuation date (e.g., that hold market conditions constant), than it is to decide, mathematically, how the model should reflect changing market conditions when the sales span changing conditions. If the valuation model within the AVM does not account for changing market conditions—or fails to use comparable sales that reflect similar market conditions—then the AVM will produce non-credible (biased and/or imprecise) valuations.

Improved AVM Performance Methodology Example

Although the internal metrics presented in Exhibit 3 indicate that the TVM is an acceptable valuation model, the calculation of these metrics violates the second recommended valuation principle, namely that a property should never be used to value itself. Specifically, the metrics in Exhibit 3 are an internal measure of model performance, because in the test the TVM-predicted values for each of the 53 houses were created using the regression coefficients already derived from these 53 houses. As a result, it would be better to use the LOO methodology to provide the TVM's performance metrics. In this way, the original data set of comparable sales would "do double duty," and could be used to value the target property and also provide, using a LOO strategy, the values of the performance metrics.

In a regression, the *predicted residual sum of squares* (PRESS) statistic has traditionally been used to identify unique individual observations (outliers) and to determine a set of statistically significantly independent variables that contribute to explaining the dependent variable.[56] The PRESS statistic implements a LOO methodology that systematically pulls each of the n sales, one at a time, from the original data set and uses the remaining $n - 1$ sales to estimate the market value of that removed property. This process is repeated by cycling through each of the n sales, one at a time, employing the same AVM that was originally used to value the target property.[57] A PRESS value can be used to calculate (with the property's selling price) the PRESS sales error. In fact, any performance metric seen in Exhibit 3 can be calculated using the PRESS value. For example, the PRESS-based FSD for the target property evaluated using the TVM is 19.6, an increase from the internal prediction-based FSD of 13.4. Such a large increase in the FSD indicates that the TVM does not predict new observations nearly as well as it explains the house prices for the existing 53 comparable sales and, as a result, the initial FSD of 13.4 is providing an overly optimistic assessment of the model's predictive performance.

Although the PRESS methodology abides by the second valuation principle—a property should never be used to value itself—it violates principle three—the data set should contain no sales that post-date the valuation date of the target property. In fact, all but the most recent comparable sale will have at least one sale post-dating it. In particular, to value the oldest of the 53 sales in the Cedar Falls data set, the PRESS procedure

55. For discussions of comparative market analyses, see Simon Stevenson, "Modeling Housing Market Fundamentals: Empirical Evidence of Extreme Market Conditions," *Real Estate Economics* 36, no. 1 (2008): 1–29; Robert Novy-Marx, "Hot and Cold Markets," *Real Estate Economics* 37, no. 1 (2009): 1–22; James VanderHoff, "Adjustable and Fixed Rate Mortgage Termination, Option Values and Local Market Conditions: An Empirical Analysis," *Real Estate Economics* 24, no. 3. (1996): 379–406; and David Dale-Johnson and Stanley W. Hamilton, "Housing Market Conditions, Listing Choice and MLS Market Share," *Real Estate Economics* 26, no. 2 (1998): 275–307.

56. Douglas C. Montgomery, Elizabeth A. Peck, and G. Geoffrey Vining, *Introduction to Linear Regression Analysis* (Hoboken, NJ: John Wiley & Sons, Inc., 2012).

57. Technically, rerunning of the AVM regression n times is not required, as the PRESS residual can be calculated using the original regression that valued the target property. See Montgomery, Peck, and Vining, *Introduction to Linear Regression Analysis*, 598–600.

uses the remaining 52 sales, which each post-date this oldest sale. As a result, none of the PRESS-based metrics are presented other than the FSD of 19.6, discussed above. The PRESS methodology should be modified to abide by the four valuation principles discussed in the previous section. This modified LOO procedure is referred to here as a "Generalized PRESS" (GenPRESS) methodology. As long as the AVM uses a set of comparable sales to produce its valuation, regardless of the valuation model being a regression or not, then the GenPRESS procedure can provide the values of the performance metrics.

To enforce the no-post-dated-sales principle of the GenPRESS methodology, additional housing sales are needed to value the oldest comparable properties. For example, if an AVM uses n comparable sales to value the target property, then the oldest of these n sales cannot be valued using the remaining $n - 1$ post-dated sales; new comparable sales would need to be selected. These newly chosen comparable properties must be selected in concordance with the time and submarket criteria established when picking the original comparable sales used to value the target property. Then, the AVM is run, using these newly gathered sales, exactly as it was when valuing the target property, to value each of the original comparable sales.

To calculate the TVM's GenPRESS regression-based predicted value for the oldest of the 53 properties in the Cedar Falls data set, property sales in 2011 were gathered. As seen in Exhibit 2, 40 property sales, denoted by the plus (+) symbol, occurred in the same submarket as the target property in 2011, which is within one year of the selling date for the oldest sale in the original data set (January 1, 2012). A market analysis (House Price Index) in Rosburg et al. indicates that relatively stable market conditions existed in 2011 and 2012 in Cedar Falls.[58] As a result, the TVM was run with these 40 property sales in 2011 to value this oldest comparable property. When using a twelve-month moving window, the number of property sales valuing each of the 53 original comparable sales in the original data set ranged from 40 to 57.

A GenPRESS predicted value was calculated for each of the original 53 comparable properties and was used to (re-)calculate the values of the performance metrics reported in Exhibit 6. Any substantial difference between the respective statistics in Exhibits 3 and 6 provides an assessment of "the applicability of the model to data beyond those on which the model is based."[59]

All the GenPRESS-based performance metrics in Exhibit 6 show a poorly performing AVM compared to the corresponding metrics in Exhibit 3. In particular, the FSD has risen from 13.4 to 24.7, while the failure rate (at ± 10%) has increased from 39.6% to 66.0%. The difference in the respective values of the performance metrics in these two tables is entirely attributed to employing the GenPRESS methodology in which no post-dated sales were used to value each comparable sale. Exhibit 6 reveals that the results seen in Exhibit 3 make the AVM appear to be more accurate, precise, and reliable. As a result, the methodology used to calculate the values of performance metrics meaningfully impacts an AVM's credibility.

Therefore, a fifth principle is applicable when resampling sales in a cross-validation analysis to calculate performance metrics. Specifically, **valuations produced by an AVM in a cross-validation analysis should use the same methodology originally used to value the target property**. That is, the performance metrics associated with a target property (for example, the FSD) should be calculated using the data set that was originally used to value the target property using a leave-one-out methodology.[60] The recommendation to use the GenPRESS methodology is more than just a method to compute the AVM's

58. See Exhibit 3 in Alicia Rosburg, Hans Isakson, Mark D. Ecker, and Tim Strauss, "Beyond Standardized Test Scores: The Impact of a Public School Closure on House Prices," *Journal of Housing Research* 26, no. 2 (2017): 119–135.

59. Neter et al., *Applied Linear Statistical Models*, 435.

60. Other philosophical choices exist for cross-validation, such as creating a holdout data set, for example, by setting aside n of the 53 comparable sales in the Cedar Falls data set. The remaining $53 - n$ sales are then used to value the target property, while the n withheld sales are used to calculate the values of the performance metrics. See Kane, Linné, and Johnson, *Practical Applications in Appraisal Valuation Modeling*, 171. We advocate including all comparable property sales in the original data set to value the target property and then using the LOO GenPRESS methodology to calculate the values of the performance metrics, primarily because the GenPRESS avoids sacrificing any comparable sales needed to value the target property.

Exhibit 6 GenPRESS Performance Metrics for the TVM

AVM Metric	Value
Mean Sales Error	$806
Mean Percentage Sales Error (MPE)	3.5%
Median Sales Error	$–5,245
Median Percentage Sales Error	–3.1%
Mean Absolute Sales Error	$24,374
Median Absolute Sales Error	$17,764
Mean Absolute Percentage (Sales) Error	18.2%
Median Absolute Percentage (Sales) Error (MAPE)	13.9%
FSD	24.7
Percentage of Estimates within ± 10% (PE10)	18/53 for 34.0%
Failure Rate at ± 10%	35/53 for 66.0%
Percentage of Estimates within ± 15% (PE15)	32/53 for 60.4%
Percentage of Estimates within ± 20% (PE20)	36/53 for 83.7%
Percentage of Estimates more than 20% (Right Tail 20%)	10/53 for 18.9%
Coefficient of Variation (COV) of TVM/Price	0.23909 or 23.9
Coefficient of Dispersion (COD) of TVM/Price	18.51
PRD of TVM/Price	1.0290
Average Regression *R*-Squared (Coefficient of Determination) for the 53 Regressions	0.7524
Average Adjusted *R*-Squared for 53 Regressions	0.6463
Average Number of Sales for 53 Regressions	50
Mean Selling Price of 53 Sales	$143,767
Median Selling Price of 53 Sales	$130,000
Mean TVM GenPRESS Valuation for 53 Sales	$144,573
Median TVM GenPRESS Valuation for 53 Sales	$139,216

performance metrics following this fifth principle. It also provides (1) guidance on how to resample a sales data set to calculate performance metrics, and (2) a straightforward and consistent method to calculate a unique FSD value for each target property.

The GenPRESS methodology also assesses the AVM's accuracy for a data set that was not used to create the model, especially when including additional comparable properties when valuing the earliest sales in the original data set. The GenPRESS-based performance metrics evaluate the quality of the AVM's prediction of new or external properties because each target property is left out of the training data set, in contrast to the usual internally based predicted value.

Discussion and Conclusions

AVM vendors should adopt the five best-practice valuation principles recommended in this article, coupled with the GenPRESS methodology to calculate performance metrics and allow a more realistic assessment of AVM performance. The example demonstrates that, for the Cedar Falls data set, applying these five valuation principles together with the GenPRESS methodology produces a degradation in the values of performance metrics, as seen in the comparison of the results in Exhibit 3 to those in Exhibit 6. In other words, not following these principles provides an overly optimistic evaluation of the AVM's performance.

The research AVM presented demonstrates that the values of performance metrics highly depend upon their calculation methodologies. For example, it was shown that three different statistical methodologies produce three different FSD results: the target property's FSD changes from 13.4, using regular regression predicted values, to 19.6, using traditional PRESS predicted values (that allow the use of post-dated sales), and finally to 24.7, using GenPRESS predicted values (that do *not* allow the use of post-dated sales). In particular, the latter two FSD values (19.6 and 24.7) are each calculated using a LOO methodology. The increase from 19.6 to 24.7 reveals the substantial effect of eliminating post-dated sales. Moreover, the values of nearly all performance metrics substantially deteriorate when enforcing the two principles that no property should be used in the model to value itself, even indirectly, and no sales should post-date the property being valued.

The research also has shown that the vendor-reported FSDs for commercial AVMs are not nearly as credible as currently being reported. The analysis of 5.3 million housing sales indicates that 85.0% of AVMs with a vendor-reported FSD of 15 or below are overly optimistic in their reported precision (by 5.3 or 83.3%, on average). Overall, the observation that AVM vendors are inconsistent in their calculation of confidence scores, along with their underreporting of FSDs, should be a concern for the currently under-regulated AVM industry. Standardization of the calculation of performance metrics should be employed by AVM vendors. Adopting quality control standards could lessen the impact of future regulatory mandates of the Consumer Financial Protection Board (CFPB).

Unfortunately, exactly how AVM vendors calculate performance metrics for any one target property is part of AVMs' proprietary intellectual property. The analysis of 327 AVM/FSD combinations using 5.3 million sales in this research suggests that AVM vendors are underreporting their FSDs. These overly optimistic vendor-reported FSDs suggest that implementation of the five valuation principles, together with the GenPRESS methodology, could serve as a means to consistently evaluate and compare AVM performance. It is fully expected that similar results would be seen for commercial AVMs, as presented for the TVM for the Cedar Falls data set, because Neter et al. have already established that internally based metrics produce overly optimistic results.[61]

What remains unclear is by how much the values of the performance metrics would change, especially the FSD, if five valuation principles together with the GenPRESS methodology were adopted by AVM vendors. Only through scrutiny of the internal workings of AVMs, by the AVM vendors themselves, can the efficacy of any set of best-practice principles be determined. In sum, AVM vendors are encouraged to adopt a set of principles, such as those detailed in this work, that comply with well-established appraisal practices and allow AVM clients to trust the credibility and comparability of AVMs, measured through their performance metrics.

About the Authors

Hans R. Isakson, PhD, is a professor emeritus of economics at the University of Northern Iowa. He has taught undergraduate and graduate level courses in real estate appraisal at the University of Georgia, Washington State University, University of Texas at Arlington, and University of Northern Iowa. He has authored and coauthored over forty articles and has presented dozens of papers on real estate valuation before various academic and professional associations. He was the corecipient of the Appraisal Institute's Arthur A. May Award in 1979 for his article "The Impact of Market Experience upon Appraisers' Energy Awareness." He has testified as an expert witness regarding the reliability of AVMs in over a dozen residential mortgage-backed securities (RMBS) lawsuits. **Contact: hans.isakson@uni.edu**

Mark D. Ecker, PhD, is a professor of statistics at the University of Northern Iowa. He has authored or coauthored over three dozen articles and provided statistical consulting services to expert witnesses in numerous high-profile court

61. There may be AVM vendors currently abiding by a set of valuation principles like those mentioned, and as a result, little to no improvement in AVM credibility would be seen.

cases involving AVMs. His current research includes developing new statistical techniques to measure the value impact on house prices of proximity to a hog lot, nuclear power plant, or high-quality school. **Contact: Mark.Ecker@uni.edu**

Lee Kennedy founded AVMetrics in 2005, which focuses on AVM model validation, testing, and collateral consulting. Prior to founding AVMetrics, he led the Alternative Valuation Group for a large lender. As a recognized expert in the testing and use of AVMs, and an authority on current AVM regulations and guidelines, Kennedy serves as a frequent speaker and panelist in the industry. He has published many articles on AVM-related topics and is the editor of *AVMNews* as well as a frequent expert witness in cases on AVM usage. Kennedy is a certified residential property appraiser in California. **Contact: Lee.Kennedy@avmetrics.net**

Disclosure

The authors have been retained as expert witnesses to assess AVMs in dozens of high-profile cases. Any opinions or points of view expressed in this study represent a consensus of the authors. Any products and manufacturers discussed in this article are presented for informational purposes only and do not constitute product approval or endorsement by the authors.

Additional Reading

Suggested by the Authors

Chagani, Ershad. "The Potential of Machine Learning Real Estate Valuation Models." *Cornell Real Estate Review* (blog), March 28, 2018. http://bit.ly/3cnqwOg.

CoreLogic. "About Automated Valuation Models (AVMs)." 2017. http://bit.ly/2uJyvnL.

Falessi, Davide, Likhita Narayana, Jennifer Fong Thai, and Burak Turhan. "Preserving Order of Data When Validating Defect Prediction Model." 2018. http://bit.ly/2TcMqfn.

Additional Resources

Suggested by the Y. T. and Louise Lee Lum Library

Appraisal Institute—Online Store

http://bit.ly/AI_Online_Store

- **Education**
  - *Real Estate Finance, Statistics, and Valuation Modeling*
- **Publications**
  - *A Guide to Appraisal Valuation Modeling*
  - *An Introduction to Statistics for Appraisers*
  - *Practical Applications in Valuation Modeling*

Freddie Mac

- **Forecast Standard Deviation**
  http://www.freddiemac.com/hve/fsd.html
- **Home Value Explorer AVM**
  http://bit.ly/freddiemac_HVE