AVMetrics
Flexible Solutions. Precision Results.
AUTOMATED VALUE METRICS

February 26, 2021

Federal Housing Finance Agency,
Office of Housing and Regulatory Policy,
400 7th Street SW, 9th Floor,
Washington D.C., 20219
(Transmitted via email to AppraisalRFI@fhfa.gov)

RE: FHFA Request for Information on Appraisal-related Policies, Practices, and Processes.

Attachments: MPT Examples and FSD Analysis Example

Dear FHFA Colleagues:

AVMetrics is pleased to see the FHFA's deep interest in appraisal-related policies, particularly where they involve Automated Valuation Models and Appraisal Waivers. AVMetrics' primary focus is the independent testing and measurement of AVMs for use in residential housing. Our sincere hope is that AVMs can be employed thoughtfully, carefully and judiciously in order that they can effectively support a safe, sound and stabile housing sector.

AVMetrics was founded in 2005 by Lee Kennedy after twenty years of appraisal analytics and risk management background. The company's primary mission is to improve the safe and sound use of Automated Valuation Models and other Alternative Valuation Products within the housing finance industry. Lee Kennedy continues to be an active member of the Collateral Risk Network (CRN), The Appraisal Foundation's Industry Advisory Council (IAC) and subcommittee member for AVM Quality Control Standards among other valuation industry activities.

The following pages include an executive summary and AVMetrics' responses to selected questions from FHFA's December 28, 2020 Request for Information on which we have subject matter expertise.

Sincerely,

Lee Kennedy,
CEO/Managing Director

2655 FIRST STREET, SUITE 250, SIMI VALLEY, CA, 93065
T (805) 421-5056 WWW.AVMETRICS.NET

Executive Summary

The lynchpin to many of the appraisal alternatives is an Automated Valuation Model – a subject which AVMetrics has studied assiduously for more than 15 years. We point out that even an excellent AVM can be improved by the use of a Model Preference Table®. MPTs enable better accuracy, fewer "no hits" and fewer overvaluations.

We also suggest an escalated focus on AVM testing, and we use our own research and citations of OCC Interagency Guidelines to emphasize the importance of testing to effectively use AVMs. We suggest that an "FSD Analysis" like the one we describe reduces risk by avoiding higher risk circumstances for using an AVM.

We suggest that the implementation of a universal MPT by the Enterprises will improve the collateral tools available and reduce the risk of manipulation by lenders. We also believe that a universal MPT can help redeploy appraisers to their highest and best use: the qualitative aspects of appraisal work. Our suggestion is that the GSEs endeavor to make the increased use of AVMs a benefit to appraisers, increasing their value-added and bringing them along in the transition.

Question A1.1:	Is there a need to provide new valuation solutions that address industry identified issues of appraiser capacity, turn-times, training, and rural and high-volume market coverage? What are those potential solutions? What are the risks of these policies and the challenges in implementing them?

We believe that there is a need for new valuation solutions. We can envision a solution that marries a highly qualified inspector evaluating property condition in conjunction with an AVM doing quantitative analysis and a trained appraiser reconciling the value to produce the best valuation for the lowest cost.

In the case of the GSEs, each has developed its own AVM to use in the appraisal waiver process. Because of the variability in performance that we observe in AVMs, we believe that relying on a single AVM is a drawback that creates increased risk in two ways.

First, using a single AVM leaves significant room for improvement. Having only one AVM reduces the percentage of times that that AVM will return a "qualified" value. This reduces how often a hybrid solution can be employed. Oftentimes at least one of the 25 other commercially available AVMs will have a better track record valuing property in any individual submarket. Incorporating multiple AVMs reduces the risk of an inaccurate valuation or a severe overvaluation. Second, at the point of use, there's significant risk of originators developing "target lock" on the AVM value estimate[1] without enough appreciation for the confidence in that value estimate.

We believe that good fundamentals for AVM usage include the following:

1) Establish minimum thresholds for AVM performance along multiple dimensions that include metrics for accuracy, precision and outliers[2]
2) Conduct frequent testing of multiple AVMs' performance using those metrics and comparing to the preestablished thresholds
3) Use a scoring system to create an overall grade for acceptability and to rank the AVMs based on performance and uncertainty metrics.

To mitigate the drawbacks of using only a single AVM as the GSE's currently do, we recommend a Model Preference Table® of AVMs. An MPT is developed by testing every model extensively against valid benchmarks, scoring each model's valuation predictions and ranking the models based on the results. Exhibit 1 (attached) includes two examples of MPTs optimized for the target geography, property

[1] See, for example, the discussion of anchoring in "**Prevalence of GSE appraisal waivers: November 2020 originations"**
https://www.aei.org/research-products/report/prevalence-of-gse-appraisal-waivers-november-2020-originations/

[2] See Appendix B of Interagency Appraisal and Evaluation Guidelines, specifically under "Selecting an AVM(s)"
https://www.fdic.gov/regulations/laws/rules/5000-4800.html

type and price tier. We believe that a MPT minimizes the risk of unusable or inaccurate valuations or overvaluations.

To help the GSE's and other potential originators avoid the "target lock" issue, the valuation solution might present the results with risk metrics being more prominent than valuation.[3] The solution might also consider presenting different categories of risk, like letter grades. In this way, an "A" grade valuation would be relied on more heavily than a "B" grade valuation or a "C" grade. "A" grades might be eligible for higher LTVs, whereas "D" grades might require a different valuation type to lend beyond a very low LTV.

Question A1.2:	Are there opportunities for process improvements that allow non-traditional valuation services (inspection-only, desktop, exterior-only) to augment traditional appraisals? Please elaborate on the risks, challenges and benefits. Separately, are there opportunities to improve traditional appraisals to mitigate problems and concerns that have been observed to date?

It's obvious to most industry participants that there is significant risk in relying on inspection-only solutions when they depend on untested, poor AVMs that produce inaccurate valuations. This is why testing AVMs for accuracy, precision and error is necessary. We recommend using AVMs that are proven to meet predefined criteria through independent testing and are appropriate to the assignment.

Exhibit 2 (attached) shows an example of an acceptability analysis that an AVM user might perform. Our scoring methodology evaluates AVM valuations in batches based on their FSD. As expected, in this example lower FSD valuations perform better. A conservative AVM user might establish a cutoff of FSD <= 12 (the green area), which yields >75% usability with a weighted average MAE of 5.3% (Mean Absolute Error). A cutoff of 12 also yields a weighted average PPE10 of 87.9% (i.e., almost 90% of valuation predictions will be within +/-10% of the benchmark value), and a weighted average PPE>20% of 2.9% (i.e., less than 35 of valuation predictions will have an overvaluation potential of >20%).[4]

However, our research spanning more than a decade shows that even overall good-preforming models are less reliable in certain circumstances, so one of the less obvious risks that we would highlight is reliance on a "good" model that is poor in a specific geography, price level or property type. Models should be tested in each one of these subcategories in order to assess their reliability and risk profile.

[3] Our articles go into detail regarding the best ways to measure and represent AVM accuracy. Please see: Isakson, H., Ecker, M.D. and L. Kennedy (2020) "Principles for Calculating AVM Performance Metrics". The Appraisal Journal. Winter 2020, 88(1). pp 14-29. See Also: Ecker, M., Isakson, H. and L. Kennedy. (2021). An Exposition of AVM Performance Metrics. *Journal of Real Estate Practice and Education*. (Forthcoming). Currently found at: https://www.researchgate.net/publication/333670561_Exposition_of_AVM_Performance_Metrics_-_Single_File .

[4] It should be noted that not every AVM will perform similarly at each FSD level. Our internal testing shows that the same FSD score can indicate very different performance for different AVMs.

Identifying "reliable models" isn't straightforward. Performance varies over time as conditions change and models are tweaked. Performance also varies between locations, so a model that is extremely reliable overall may not be effective in a specific region. Furthermore, models that are effective overall may not be effective at all price levels, for example: low-priced entry-level homes or high-priced homes. Finally, very effective models will also produce estimates that they admit have lower confidence scores (and higher FSDs), and which should in all prudence be avoided, but without adequate testing and understanding may be inadvertently relied upon. Proper testing and controls can mitigate this problem.

In a July 2020 article on the effectiveness of confidence scores[5], we showed that high-performing, commercial-grade AVMs can still produce substantial over-valuations. 3% of the valuations provided by one of the best AVMs had very low confidence scores, and almost one-third those valuations exceeded the benchmark value by 20% or more. To the credit of the model developer, they identify the low confidence of the estimate, but the risks are that the GSE's and other potential users may not have an option to use another AVM or may not realize how meaningful that low confidence score (or high FSD) is.

That same study showed that the lowest confidence valuations by one model can often be valued much more reliably by other models. That is demonstrated by analyzing the 129,000 lowest-confidence valuations from the highest-performing model. You might think that if the best model in the study could not effectively value those properties, it was probably because they were not good prospects for accurate AVM valuations. In many cases that is undoubtedly true. However, the second-best model overall actually produced 38,000 high-confidence valuations that tested quite well on those same properties (about 29% were effectively estimated by the second model). In other words, even the best model will not perform well in some places, and other models can fill in those gaps effectively. In our opinion, there is risk that the GSE's and potential other originators who rely on only one model without understanding its strengths and weaknesses will make riskier lending decisions.

Question A1.3:	Do appraisal waivers have a place in Enterprise appraisal policy and process, and if so, for what segment of loans? What are the current risks to Enterprise safety and soundness in how appraisal waivers are offered? Would caps or other limits on their usage be appropriate?

We believe that appraisal waivers clearly have a place in Enterprise appraisal policy, and the current policies of limiting appraisal waivers based on loan purpose, LTV and other indicators of risk are good mitigations.

Question A1.4:	Would utilizing alternative inspection workforces, such as insurance adjusters, real estate agents, and appraisal trainees assist with addressing appraiser capacity concerns? Are there risks of using third-party non-appraisers? If yes, How?

[5] https://www.avmetrics.net/2020/07/02/for-avms-confidence-isnt-overrated/

Please see our answer to Question C1.1.

Question A1.5:	Is there a need for additional policies and controls to balance potential risks with efficiency benefit from appraisal modernization? If yes, please provide your recommendations.

We believe that there are significant risks from increased reliance on models that are less effective, either overall or in specific circumstances. Those risks can be mitigated with better information on the accuracy, precision and error of the models being relied upon. Regular, comprehensive testing in real-world conditions against valid benchmarks is essential to understand the strengths and weaknesses of any model.

In order to ensure that adequate testing is done regularly, we recommend that a control be implemented to create transparency around how the GSE's or other originators are performing their testing. This could be done in a variety of ways. One method might require the GSE or lending institution to indicate their last AVM testing date on each appraisal waiver.

Rating agencies are already providing another form of controls on appropriate AVM usage. In August of 2020, Moodys downgraded J.P. Morgan Chase's Mortgage Acquisition Corp as an aggregator of prime jumbo mortgages largely because of their valuation cascade practices and the lack of performance data on those practices.[6]

Regardless of how it's done, the goal would be to create a mechanism that would increase commitment to appropriate testing. The GSE's could provide a leadership role by demonstrating how they would like lending institutions to demonstrate their independent AVM testing as required by OCC 2010-42 and 2011-12.

Question B2.2:	How can the Enterprises improve their collateral tools currently available to lenders?

In our almost two decades of experience, our testing has found that no one model can be the most accurate in every circumstance. This has led to the conclusion that the best automated collateral tool recognizes the strengths of the many commercially available models at a very granular level. At this time, there are at least 25 commercially available residential models. Their relative strengths and weaknesses can be identified through comparative analysis. The best automated tool would use that information to deploy those commercially available models in an optimized manner.

[6] See https://www.moodys.com/research/Moodys-decreases-JP-Morgan-Mortgage-Acquisition-Corps-assessment-to-Average--PR_430194

We believe that the Enterprises can best improve their collateral tools that they make available to lenders by providing them with a Model Preference Table® customized to each specific valuation need.

Question B2.4:	How can lenders manipulate automated underwriting systems when seeking an appraisal wavier? For example, lenders changing the loan amount, submitting data changes multiple times, or submitting to both Enterprises and delivering to the one who offers the waiver? How do the Enterprises minimize this manipulation?

We agree that there is an opportunity for manipulation by lenders by means of exploiting two characteristics of the Enterprises' appraisal waiver systems. First, the fact that the two GSEs can be applied to separately allows them to be played against each other. Second, the fact that the two GSEs have different internal AVMs could lead to an adverse selection process, with lenders more likely to submit loans to the GSE with the AVM that has produced an overvaluation.

We recommendation a solution in which both GSE's use a common platform for the acceptance of waivers and that that common platform uses a universal Model Preference Table®. Not only would there be less chance of adverse selection, but the platform would always be using the most competent model for a property.

Question B2.6:	Is there any data or evidence you could share regarding the performance of alternative appraisal solutions versus traditional appraisals?

AVMetrics has done these studies in the past and would be happy to collaborate with the GSEs on testing the performance of alternative appraisal solutions against traditional appraisals. AVMetrics could contribute its robust testing capability and leverage the appraisal and alternative appraisal data available to the GSEs. That combination would undoubtedly reveal new insights about when and where each methodology has advantages.

Question C1.1:	What do you envision the impact of appraisal process improvements as described in this RFI to be on the appraisal industry? What impact, if any, has increasing use by the Enterprises of alternative appraisal solutions had on the availability and/or quality of traditional appraisals?

As the demand for valuation services continues to grow, the appraisal profession continues to shrink. This disparity can already be felt in rural and underserved areas with delays and increasing fees. There have been a number of requests to the ASC for appraisal waivers[7]. The only way to fill that gap will be to

[7] See Federal Register, Appraisal Subcommittee, Notice of Received Request for Temporary Waiver, dated May 30, 2019, available at: https://www.federalregister.gov/documents/2020/08/13/2020-17660/appraisal-subcommittee-order-extending-commercial-real-estate-transaction-temporary-waiver-relief

incorporate new ways of effectively increasing appraiser productivity such as the modernization being discussed in this RFI process. We believe that such modernization can increase the capacity of the valuation industry overall. However, there are risks, and depending on how the modernization is handled, appraiser retirements could limit any increase to overall industry capacity (please see our answer to Question C1.2).

Our belief is that appraisers are a valuable and limited resource, and they should be employed at their highest and best use. Trying to be a "manual AVM" is not their highest and best use. "Cookie cutter" or non-complex appraisals are not their highest and best use. Their expertise should be focused on the qualitative aspects of the valuation process such as condition, market and locational influences, not the quantitative (facts) such as bed and bath counts. Although AVM's are getting much better – and new tools that specifically model this type of qualitative data are being tested[8] – they still do not match the ability of a trained appraiser at capturing and analyzing the qualitative aspects of a property.

The natural solution is merging the robust data processing capabilities of an AVM with the qualitative assessment skills of appraisers. This would logically take the form of an AVM augmented by qualitative assessments of physical property condition, market and location influences, etc. We believe that modernizations along these lines will make appraisals more accessible by redeploying appraiser resources to their highest and best use, i.e., solving the valuation issues surrounding complex properties.

Question C1.2:	What would be the impact of appraisal policy and process improvements to the mid or late career appraiser? Do you believe late career appraisers would delay retirement if they could focus on specific valuation services like desktop appraisals? Or alternatively, would late career appraisers cease operations due to technology adoption challenges?

See also Federal Register, Appraisal Subcommittee, Order extending Commercial Estate Transaction Temporary Waiver Relief, dated August 13, 2020, available at: https://www.federalregister.gov/documents/2019/05/30/2019-11282/appraisal-subcommittee-notice-of-received-request-for-a-temporary-waiver

See also Appraisal Subcommittee, Frequently Asked Questions (FAQs): Temporary Waiver Authority, dated December 6, 2019, available at: https://www.asc.gov/Documents/OtherCorrespondence/2019.12.06%20-%20Temporary%20Waiver%20FAQs%20.pdf

[8] Two examples include: NatureQuant: https://www.naturequant.com/naturescore/ AVMetrics has conducted preliminary analysis on the NatureScore, and it has shown a positive correlation to the traditional appraisal process of valuing external influences on a property.

Cape Analytics: https://capeanalytics.com/real-estate/ They have developed model digestible property condition data, and they are currently being used for conditional property adjustments enhancing an AVM.

The appraisal industry has been through many transitions, and in each one, experienced appraisers have chosen to retire instead of struggle with the change. Our hope is that proactive management of the transition can minimize the losses to the industry from premature retirements.

Markets and innovation operate by providing the incentive to blaze new paths to the goal of producing more from less, i.e., increasing productivity. Unfortunately, many appraisers already experienced the recent changes as an ever-increasing pressure to extract more from them for less remuneration. We recommend two approaches to retaining appraisal talent in the industry.

First, these innovations – appraisal waivers, AVMs, hybrids, etc. – should be implemented in such a way as to focus appraisers on their highest and best use (for which they should be paid commensurate with their value-added). In other words, if some of the productivity gained by implementing new policies and processes can accrue to the appraisers, then they will experience these changes as increases in productivity and earning potential, and they will be more supportive.

Second, a change management approach to the transition that uses effective communication can help reduce frustration for appraisers. If appraisers are left to figure out on their own how to avoid "creative destruction," they will likely have a more negative opinion of modernization. However, if they are actively led through industry communications to understand the advantages of outsourcing lower value-added activities and being allowed to focus on higher value-added activities, they may have greater buy-in.

Exhibit 1: Model Preference Table examples. Here are two small snapshots of a large table that optimizes the AVM choice based on 3 criteria: geography (in the first case, Los Angeles County, CA and in the second, Fairfax County, VA), property type (SFR, Condo, TRH and PUD) and price tier (in this case, just three). Even though these examples are "blinded" to remove the actual AVM names, a couple inferences can be made based on the results. First, in some sub-categories, there may be no models that performed well enough to be appropriate for use. Second, very often the best model in one category will not be first (or even in the top five) in an adjacent category.

It is interesting to note that in the nine categories shown in Los Angeles (three property types x three price tiers), four models appear six times (models 6, 7, 17 and 21), but no model appears in every category. In the Fairfax, VA table's nine categories, no model appears more than five times (models 4, 6, and 21). These results demonstrate that no single model is best across all the subcategories, and in fact, no single model is even capable of serving all of these niches.

1 Los Angeles County, CA example

BLINDED VERSION							
		Bucketing	**1st**	**2nd**	**3rd**	**4th**	**5th**
		National; All Prop Types; All Price Tiers	Model 17	Model 21	Model 7	Model 20	Model 6
LA County		LA County, All PropTypes; All PriceTiers	Model 17	Model 7	Model 21	Model 2	Model 20
LA County	PropTypes	**LA County; Condo; All Price Tiers**	Model 17	Model 21	Model 7	Model 2	Model 6
LA County	PropTypes	**LA County; PUD; All Price Tiers**	Model 7	Model 17	Model 21	Model 2	Model 6
LA County	PropTypes	**LA County; SFR; All Price Tiers**	Model 17	Model 7	Model 21	Model 2	Model 20
SFR	Price Tiers	**LA County; SFR; $50K - $250K**	Model 21	Model 7	Model 6	Model 20	Model 25
SFR	Price Tiers	**LA County; SFR; $250K - $750K**	Model 17	Model 21	Model 7	Model 20	Model 6
SFR	Price Tiers	**LA County; SFR; $750K+**	Model 17	Model 7	Model 21	Model 2	Model 9
PUD	Price Tiers	**LA County; PUD; $250K - $750K**	**No Model Scoring below $250K**				
PUD	Price Tiers	**LA County; PUD; $250K - $750K**	Model 2	Model 4	Model 12	Model 6	Model 20
PUD	Price Tiers	**LA County; PUD; $750K+**	Model 7	Model 17	Model 21	Model 2	Model 25
Condo	Price Tiers	**LA County; Condo; $50K - $250K**	Model 17	Model 21	Model 6	Model 5	Model 9
Condo	Price Tiers	**LA County; Condo; $250K - $750K**	Model 17	Model 21	Model 7	Model 2	Model 6
Condo	Price Tiers	**LA County; Condo; $750K+**	Model 17	Model 7	Model 2	Model 6	Model 4

Exhibit 1: (MPT Example continued)

2 Fairfax, VA county example

BLINDED VERSION							
		Bucketing	**1st**	**2nd**	**3rd**	**4th**	**5th**
		National; All Prop Types; All Price Tiers	Model 17	Model 21	Model 7	Model 20	Model 6
Fairfax County		Fairfax County, All PropTypes; All PriceTi	Model 17	Model 21	Model 7	Model 6	Model 4
Fairfax Co	PropTypes	**Fairfax County; Condo; All Price Tiers**	Model 17	Model 21	Model 4	Model 6	Model 3
Fairfax Co	PropTypes	**Fairfax County; PUD; All Price Tiers**	**No Model Scoring for PUD**				
Fairfax Co	PropTypes	**Fairfax County; SFR; All Price Tiers**	Model 17	Model 21	Model 7	Model 20	Model 2
Fairfax Co	PropTypes	**Fairfax County; TRH; All Price Tiers**	Model 21	Model 7	Model 6	Model 4	Model 3
SFR	Price Tiers	**Fairfax County; SFR; $50K - $250K**	**No Model Scoring for SFR under $250k**				
SFR	Price Tiers	**Fairfax County; SFR; $250K - $750K**	Model 17	Model 21	Model 20	Model 7	Model 6
SFR	Price Tiers	**Fairfax County; SFR; $750K+**	Model 7	Model 2	Model 4	Model 6	Model 20
TRH	Price Tiers	**Fairfax County; TRH; $250K - $750K**	Model 7	Model 16	Model 21	Model 4	Model 23
TRH	Price Tiers	**Fairfax County; TRH; $250K - $750K**	Model 21	Model 7	Model 4	Model 6	Model 3
TRH	Price Tiers	**Fairfax County; TRH; $750K+**	Model 20	Model 23			
Condo	Price Tiers	**Fairfax County; Condo; $50K - $250K**	Model 17	Model 21	Model 4	Model 6	Model 3
Condo	Price Tiers	**Fairfax County; Condo; $250K - $750K**	Model 17	Model 10	Model 4	Model 12	Model 6
Condo	Price Tiers	**Fairfax County; Condo; $750K+**	**No Model Scoring for Condo over $750k**				

Exhibit 2: FSD Analysis of an AVM showing possible acceptability thresholds. Below, a single AVM in a single geography is analyzed by grouping all usable valuations by the AVM-provided FSD. In this example, “Model P”, provided 1,125 usable valuations in South Carolina with an FSD of 7. This analysis shows how they scored on three metrics covering accuracy, precision and outliers, and it provides a composite score that can be used for ranking. This composite score and the cutoffs are examples of what an AVM user might employ. However, the specific metrics and scoring should be appropriate to the application of the AVM user.

3 FSD Analysis

Vendor	Model	State	FSD	Records	Hits	MAE	PPE10	PPE20+	Composite Score	Point score break	% of total Population	
c	P	SC	7	1,125	1,125	3.3%	96.6%	1.2%	10.621		76.4%	In "Green"
c	P	SC	8	996	996	3.6%	94.2%	0.9%	10.475			
C	P	SC	9	411	411	5.2%	87.6%	3.2%	10.001			
C	P	SC	10	547	547	7.1%	80.1%	4.8%	9.468	1.00		
C	P	SC	11	439	439	8.6%	74.5%	5.5%	9.069			
C	P	SC	12	317	317	9.8%	69.1%	8.5%	8.676	2.00		
C	P	SC	13	232	232	12.7%	61.6%	10.8%	8.085		87.75%	Green and Yellow
C	P	SC	14	197	197	13.1%	51.8%	10.2%	7.523	3.00		
C	P	SC	15	140	140	13.7%	50.0%	6.4%	7.431			
C	P	SC	16	111	111	17.5%	39.6%	13.5%	6.581			
C	P	SC	17	111	111	15.0%	45.9%	12.6%	7.076			
C	P	SC	18	109	109	18.1%	34.9%	11.0%	6.312			
C	P	SC	19	63	63	20.4%	31.7%	19.0%	5.933			
C	P	SC	20	52	52	19.3%	44.2%	19.2%	6.677			
C	P	SC	21	53	53	21.9%	35.8%	18.9%	6.078			
C	P	SC	22	51	51	22.4%	33.3%	17.6%	5.925			
C	P	SC	23	29	29	22.7%	34.5%	27.6%	5.873			
C	P	SC	24	15	15	26.9%	26.7%	33.3%	5.153			
C	P	SC	25	21	21	19.6%	33.3%	19.0%	6.066			