ASSESSING THE ACCURACY OF INDIVIDUAL PROPERTY VALUES ESTIMATED BY AUTOMATED VALUATION MODELS

GEORGE ANDREW MATYSIAK
MAY 2018
ABOUT THE AUTHOR

PROFESSOR GEORGE ANDREW MATYSIAK

George Matysiak is a financial economist and visiting Professor at Cracow University of Economics. He holds a visiting position at the School of Economics and Management (ISEG), Technical University of Lisbon. He has acted as a consultant to a variety of organisations on modelling, forecasting and investing in commercial real estate markets. Specialisms include: commercial real estate investment strategy; portfolio risk management; econometric modelling and forecasting commercial property markets; real estate pricing and risk premia; performance measurement; assessing real estate forecasting accuracy and Monte Carlo simulations of investment risk.

Prior to embarking on an academic career, he was as an Associate Director with the world’s largest real estate consultancy CBRE and previously in the research department of the insurance company Prudential Portfolio Managers (PruPim).

George Matysiak has delivered many real estate investment related training courses for companies including the Investment Property Databank, Government of Singapore Investment Corporation, the Investment Property Forum (IPF) and various executive training programmes. He has regularly presented at academic and professional/business conferences, including those organised by the National Bank of Poland.

Before settling in Poland, George Matysiak held the positions of Professor of Real Estate Investment at the University of Reading/Henley Business School and Senior Research Fellow at City University/CASS Business School, London. He is the author of many papers published in international journals focusing on real estate and economics and has twice been awarded the best paper published in the Journal of Property Investment & Finance. He is co-author of the popular text book Real Estate Investment: A Capital Market Approach.
ASSESSING THE ACCURACY OF INDIVIDUAL PROPERTY VALUES ESTIMATED BY AUTOMATED VALUATION MODELS

BY PROFESSOR GEORGE ANDREW MATYSIAK

SUMMARY

• This report follows an earlier report titled “The Accuracy of Automated Valuation Models (AVMs)” presented to TEGoVA in May 2017. It provided background information on AVMs and addressed their accuracy. In particular, the report found that European AVM vendors behave in a secretive manner and in the circumstances the report’s analysis of the accuracy of AVMs was based on information from the more transparent AVM vendors in the USA.

• The brief for this report has been to identify uncertainty surrounding AVM valuations and the criteria by which a valuer could judge the accuracy of an AVM estimate of value when being assisted by such AVM as a valuation tool.

• Currently, there are no ‘industry’ standard formats for assessing the accuracy of AVMs or the reporting of AVM outputs and attendant information. Following a review of relevant available sources, this report recommends clearly specified minimum reporting information from AVM vendors.

• A distinction needs to made between the reporting requirements for institutional lenders in connection with the valuation of property portfolios on the one hand and for valuers, lenders and borrowers in connection with the valuation of individual properties on the other. This report addresses the latter.

• The minimum required information recommendations are undemanding, in that the figures fall out of the AVM valuation and can therefore be routinely reported.

• In order to facilitate transparency and to assist valuers in assessing the accuracy of AVMs, it is recommended that AVM vendors provide in their reports the information for individual properties listed in Table 1 below.

• Table 2 in turn sets out some typical accuracy measures, together with definitions.

In seeking answers to these questions, one needs to look at valuation both as a process of reaching a value, and as the result of this process.
## TABLE 1: SUMMARY OF REQUIRED INFORMATION

<table>
<thead>
<tr>
<th></th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>The i) 50% and ii) 95% confidence intervals of the AVM valuation. That is, the estimated value ranges, containing the most likely lowest and highest property value</td>
</tr>
<tr>
<td>2.</td>
<td>A clear explanation, accompanied by a ‘legend’, of the ‘confidence score’ or ‘confidence level’</td>
</tr>
<tr>
<td>3.</td>
<td>Confirmation that comparables have been used in the AVM valuation. If not, what method was used in the AVM valuation?</td>
</tr>
<tr>
<td>4.</td>
<td>The standard deviation and the skewness of the comparable sales prices, or appraised values, used in the AVM valuation</td>
</tr>
<tr>
<td>5.</td>
<td>The AVM model’s overall accuracy, based on the comparable sales sample using:</td>
</tr>
<tr>
<td>i)</td>
<td>Mean Absolute Error</td>
</tr>
<tr>
<td>ii)</td>
<td>Median Absolute Error</td>
</tr>
<tr>
<td>iii)</td>
<td>‘Error Buckets’ for the percentage of valuations lying within +/- 5%, +/- 10% and +/- 20% of the Sales Price</td>
</tr>
<tr>
<td>6.</td>
<td>The number and the overall geographic distribution of the comparables used in the AVM valuation</td>
</tr>
<tr>
<td>7.</td>
<td>The range of comparables sales prices used in the AVM valuation</td>
</tr>
<tr>
<td>8.</td>
<td>Confirmation of the earliest and most recent sales dates of the comparables used in the AVM valuation</td>
</tr>
<tr>
<td>9.</td>
<td>If ‘adjusted’ comparable sales prices have been used, explanation of how they were adjusted</td>
</tr>
<tr>
<td>10.</td>
<td>Confirmation of the Benchmark used in arriving at the figures in 4. and 5. above, sales prices or valuations, in arriving at the overall accuracy figures</td>
</tr>
</tbody>
</table>

---

**Notes:**

Comparables are sold properties, similar to the property which is being AVM valued.

Recommendation 2: ‘Confidence score’ and ‘confidence level’ represent the AVM vendor’s assessment of how accurate, in their view, the AVM valuation is. It provides useful information on how much confidence the AVM vendor has in the resulting valuation figure. Each vendor will have their own definition, for example, it could be on a scale of 1-9 or some other scale. The basis of the calculations will differ for different AVM vendors. Consequently, Recommendation 2 asks for a clear explanation together with a legend showing the range of values.

Recommendation 5 iii) ‘Error Buckets’ is a term employed to describe the range in which the AVM valuations fall. For example, the Zillow ranges in Figure 3, show the percentage of AVM valuations falling within +/- 5%, +/- 10% and +/- 20% of the sales price. These ranges are referred to as “Error Buckets”.
**TABLE 2: DEFINITIONS AND MEASURES OF AVM ACCURACY**

<table>
<thead>
<tr>
<th>MEASURE</th>
<th>WHAT IS BEING MEASUREED</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absolute Error</td>
<td>The difference between the sales price and the original valuation, <em>ignoring</em> whether it’s a positive or negative amount.</td>
</tr>
<tr>
<td></td>
<td>The value of</td>
</tr>
<tr>
<td>Mean Absolute Error</td>
<td>The average of the Absolute Errors.</td>
</tr>
<tr>
<td></td>
<td>Absolute Error/N, where N = the number of valuations</td>
</tr>
<tr>
<td>Absolute Percentage Error</td>
<td>Absolute Error as a % Sale Price.</td>
</tr>
<tr>
<td></td>
<td>100* Absolute Error/Sale Price.</td>
</tr>
<tr>
<td>Median Absolute Error</td>
<td>The value which splits the errors, such that 50% are less than the middle value and 50% are greater than the middle value.</td>
</tr>
<tr>
<td></td>
<td>Order the Absolute Errors from the lowest value to highest value, then select the middle value, which is known as the median. (If there are an even number of values take the average of the two middle values.) The Absolute Errors are thus split into two groups containing an equal number of errors.</td>
</tr>
<tr>
<td>Percentage Error</td>
<td>100* (AVM Value – Sale Price)/Sale Price.</td>
</tr>
<tr>
<td>Mean Percentage Error</td>
<td>The average of the Percentage Errors.</td>
</tr>
<tr>
<td></td>
<td>Percentage Error/N, where N = the number of AVM valuations.</td>
</tr>
<tr>
<td>Standard Deviation of Percentage Errors</td>
<td>The spread of the Percentage Errors around the Mean Percentage Error.</td>
</tr>
<tr>
<td></td>
<td>Variance = [\sum_{1}^{N} \frac{(Percentage Error - Mean Percentage Error)^2}{N}]</td>
</tr>
<tr>
<td></td>
<td>Standard Deviation = (\sqrt{\text{Variance}})</td>
</tr>
<tr>
<td>Forecast Standard Deviation (FSD)</td>
<td>A measure of the spread surrounding the AVM valuation figure. This is the ‘uncertainty value’ which surrounds the AVM valuation, based on the model’s standard deviation error.</td>
</tr>
<tr>
<td>Confidence Interval</td>
<td>The range within which the sales price is likely to be.</td>
</tr>
<tr>
<td></td>
<td>For a 95% Confidence Interval the range is: (AVM Value-1.96*SD to AVM+1.96*SD), where SD = standard deviation error</td>
</tr>
<tr>
<td>Vendor’s bespoke ‘Confidence Score’ or ‘Confidence Level’</td>
<td>The vendor’s assessment of the how accurate the AVM valuation actually is. Vendors may/will have their own scale which needs to be clearly explained in the report.</td>
</tr>
<tr>
<td>‘Error Buckets’: % of AVM valuations that fall within +/-5%, +/-10% and +/- 20% of the sales price</td>
<td>The percentage of comparable valuations falling within the Error Buckets sales price ranges.</td>
</tr>
</tbody>
</table>
“For immovable property collateral, the collateral shall be valued by an independent valuer at or at less than the market value. An institution shall require the independent valuer to document the market value in a transparent and clear manner.”

Article 229(1), Capital Requirements Regulation Extracts from European Valuation Standards, European Valuation Guidance Note 11, “The Valuer’s Use of Statistical Tools”:

**1.5** As a general rule, the valuer should be aware that any analytical tool is only as reliable as:
- the data that is fed into it;
- the analytical model it uses.

**1.6** As huge volumes of data about real estate become more readily available, vast selection of ‘revolutionary’ products based on statistical analysis also becomes available on the market.

**1.7** AVMs are such statistical models, often a combination of complex models, aiming to produce a price estimate of a property as at a specific date.

**1.8** However, the valuer wishing to use an AVM in his work has to understand its role as a statistical tool in producing a valuation report that is in compliance with EVS, as by their nature AVMs are machine-based tools and do not have the advantage of viewing the subject property, understanding its context or appraising and discussing the outcomes.

...
1. INTRODUCTION

This report represents continuing work, following on from the report *The Accuracy of Automated Valuation Models (AVMs)*, presented to TEGoVA in May 2017. That report addressed the accuracy of AVMs, and found that European AVM vendors were opaque, the reported accuracy figures being based on those from two large US AVM vendors.

European law and European Valuation Standards allow valuers to use AVMs as one tool among others in reaching an estimation of value, but only insofar as the valuer is able to satisfy him/herself and the client of the relevance of the AVM report, its inputs and outputs. To enable this, it thus becomes essential that AVMs be more transparent and their accuracy verified. Can the claims made by AVM vendors be verified?

This report recommends *minimum reporting requirements* thereby enabling an assessment of AVM valuations. At the outset, a distinction needs to be made between two groups of AVM users: banks and valuers. Banks will require considerably more information, including details of the types of models employed and ‘Bulk’ accuracy test results. This report addresses the minimum information needed by valuers in order to gauge the usefulness and accuracy of the AVMs they propose to use as one of their tools.

The focus of the report is on identifying practical information to be provided in an AVM report and in particular, AVM accuracy figures. It will propose recommendations for the independent monitoring of AVM output, including an assessment of their reported accuracy. The required information recommendations made in the report comprise of what is seen as a minimum amount of detail which can be readily provided by the AVM vendor. Consequently, this information will enable valuers and other interested parties to assess AVM valuation accuracy. Several organisations, particularly in the US, have given considerable thought to AVM validation issues and the type of information which should be reported.

The recommendations which follow in the report draw on this material.

The report is structured as follows. Following the Introduction, Section 2 provides examples of accuracy figures for two major US AVM vendors and Section 3 briefly looks at data considerations. Section 4 looks at AVMs from a number of perspectives, including reporting and the minimum recommended information for AVM reports. Section 5 concludes.
2. THE DISTRIBUTION OF AVM ACCURACY FIGURES

In order to put AVM accuracy profiles into perspective, the distribution of various accuracy figures was obtained for two of the largest US AVM vendors namely, HouseCanary and Zillow. The figures are in respect of two data features, accuracy by size of property and accuracy by metropolitan location. US accuracy data has been used, as, unfortunately, European AVM vendors do not make their accuracy figures publicly available, and only provide the results of tests of accuracy to organisations such as banks making use of their AVM models.

This is of crucial importance for the individual buyer/borrower whose home is being valued.

The conclusion is that AVM valuations will contain ‘errors’. There will be a variety of sources contributing to the errors, mainly due to model and data issues, including:

- The AVM model is inadequate i.e. ‘model risk’ as the model is not well specified
- The market is not static and the AVM model is unable to capture the changed conditions
- The quality and extent of data coverage used in estimating the model
- The lack of physical inspection
- Outdated data when identifying comparables
- Pure ‘random’ error due to inexplicable factors – some effects are simply unanticipated

Assessing model risk and model validation is expected by regulators in the US to be undertaken by institutions using models. Indeed, amongst other conditions, federal banking agencies in the US require mortgage lenders that use AVMs to ‘have an understanding of how the model works’. Consequently, in assessing the validity of an AVM, lenders need to confirm the reasonableness of the model’s conceptual underpinnings and quantification methods.

It would be realistic to say that, as with any statistical or algorithmic model, AVMs depend on key assumptions which reflect a simplified depiction of how property transactions prices are determined in the market place. In reality, the assumptions will not be met.
No one should expect a model to be optimal under all circumstances.

One way to improve accuracy is to combine models, so-called ‘cascading’, which many AVM vendors do. The minimum required information recommended in this report does not include cascading details to be provided by AVM vendors.

What this tells us is that all AVM valuations will contain error and that the error will vary in different situations. For example, the error will vary over different market conditions, such as rising and falling markets, for different types of property, by location and by country. This has implications when addressing the question of accuracy:

- AVM accuracy needs to be evaluated and reported, at finer granular levels than currently is the practice and under different market conditions.

The practical questions are, how can the error be measured and how can this be reflected in the uncertainty surrounding the AVM valuation? The report’s required information recommendations address this.

In evaluating AVMs, three aspects in particular need to be addressed:

- The quality of the input data
- The AVM model(s) used to make the valuation prediction(s)
- The output of the model(s)
High quality data is of key importance in arriving at accurate AVM valuations. It is not the purpose of this report to address this aspect, other than to note it’s critical role and to make a few observations in passing:

- Current shortcomings of AVMs are much more information based than methodology based.
- Current condition of most properties is largely unknown from the typical property level datasets being used today (especially for properties which have not been listed or sold for a long time).

Examples:

- Does the home have a desirable open floor plan?
- Is there good natural light coming into the home?
- Is there good usable outdoor space on the lot?

Such data is available today, but only for a very small portion of the entire housing stock, and it is prohibitively expensive.

The view of a leading US AVM vendor is that as data becomes cheaper and more broadly available, the information it provides can only result in a decline in error rates. For example, high density 3D imaging using airborne laser scanning techniques (ALS) data together with image recognition is being used by leading US AVM vendors. It has been estimated that valuation error using such data can be reduced by up to 15%.

Based on correspondence with HouseCanary, the general view is that only marginal gains will result from new algorithms used by the final prediction model. The larger gains will come from having better quality data.

A caveat: developing new algorithms used for non-traditional data feature extraction and/or faster processing times will remain very important.

There is a lack of public information from AVM providers on data screening methods, for example distressed sales or special situation sales, and the resulting databases which are used in the underlying models. Other data issues include depth of data, the accuracy and completeness of the data, thin markets, timeliness and geographic coverage.

The European AVM Alliance (EAA) in their publication “Standards for Statistical Valuation Methods for Residential Properties in Europe” (2017) comment on data aspects as follows: “If data which is significant to estimate the value of a property is missing, or if the statistical process has shown the data to be inconsistent or unreliable, no statistical valuation should be provided or, alternatively, information on reduced expected accuracy should be produced.”

On data quality:

“Data preparation should entail thorough validation of the data. The extent and level of complexity of the validation procedure depends on the kind, origin and detail of the data. Statistical validation and cleansing routines should be used alongside screenings.
of samples to cross-check and improve the data quality.”

Historic transactions prices become less representative over time, particularly when residential markets are rising or falling rapidly. Consequently, AVM vendors should verify that timely comparable information is being used when reporting AVM valuations. One implication of this is addressed in the minimum required information (Section 4.3).

AVM vendors should provide a description of the sources of data and the subsequent data cleansing procedures addressing aspects such as missing data and outliers, thereby ensuring the accuracy of the data.

Actual sales should be valid transactions that reflect market value. Data should be consistent across the population of properties to be valued using the model. Examples include quality, physical condition, and effective age. Furthermore, the transactions data and characteristics should be representative of the underlying population or the subset of properties that may be subject to valuation using the AVM. A description of how missing data is handled should be provided.

The AVM vendor should make clear how comparable data is selected and the minimum number of such transactions required.

If a minimum number is unavailable, the steps involved in aggregating data in order to obtain a minimum should be outlined. This is a minimum requirement information recommendation (Section 4.3).

Where an AVM has been tested, an assurance should be given by the AVM provider that the test data are consistent, that is, have similar characteristics, with the data which were used in estimating the model(s). Furthermore, confirmation of how the test data were selected should be provided.

In the minimum required information recommendations (Section 4.3), other than referring to comparables data, no further detail on data aspects is required.
INTRODUCTION

It suffices to provide a brief background of the process of model validation. The following succinct explanation is extracted from a Connected Analytics publication, without comment.

The primary goal of model validation is to determine whether the model under inspection is fit-for-purpose i.e. can the model be applied with confidence in the sphere of its intended uses and can the model be relied upon to perform in line with its stated performance metrics.

As a process, model validation is not new to banks. Internal risk models used for portfolio and capital management are routinely subjected to detailed inspection of:

- The soundness of the model methodology adopted
- Appropriate evidence of stakeholder and expert engagement
- The accuracy of the analytics undertaken
- The quality and relevance of data used in the analysis data set
- Performance of the model developed on independent data (i.e. a fresh data set not used for modelling)
- The quality of documentation produced
- System assurance testing to check the model is implemented as designed
- Evidence of appropriate oversight and governance from senior management.

4.1 AVM STANDARDS

Standard setting aimed at AVM vendors impacts on a variety of AVM features, which is to be welcomed. This report does not set out to define standards. In a broader setting, the earliest endeavours in establishing AVM standards can be traced back to 2003 (a brief overview is provided in d’Amamto & Kauko (2017, pp 3-21)).

Details are provided in Annex I

4.2 PROCEDURES FOR TESTING AVM OUTPUT ACCURACY

A vital question surrounding AVMs is their accuracy. A broad profile of accuracy figures was reported in Section 2, where it was seen that the figures can vary widely, depending on location and size of property. Whilst the figures represent aggregates, they suggest that more refined analyses need to be undertaken at lower levels of aggregation, termed ‘granular’ by the EAA; analyses of accuracy need to be undertaken for different compositions of data. For example, some of the questions regarding AVM accuracy include:

- Is the margin of error unchanging i.e. is it constant? It will most likely vary under different market conditions. How does the error vary between rising markets and falling markets?
- There are many circumstances which would need to be taken into consideration when assessing what would be an acceptable margin of error in valuing residential properties. Examples include:
- different market conditions/environments, such as rising/falling prices
up market versus down market (is the error asymmetric)
different size/value properties
quality of property
age of property
market liquidity e.g. the volume of transactions in thin markets
geographic location/different neighbourhoods

All of the above are likely to vary by country and within each country!

Particularly noteworthy is that the performance of AVMs will likely be very different when markets are falling, compared to relatively stable conditions or when markets are rising. So-called ‘valuation lag’ may arise in falling markets whereby valuations lag market conditions, leading to overestimates in value. The opposite situation arises when markets are rising, that is, AVM valuations may likely underestimate value. These situations have implications, particularly, for example, in the context of loan-to-value (LTV) cut-off ratios.

One consequence of these mis-valuations (i.e. biased valuations) is that the effective LTVs will be either higher or lower than intended by the lender, with the attendant impact on the lender’s loan portfolio risk. Furthermore, what is the margin of error for different portfolio loan books and is it smaller/larger for bigger portfolios?

Indeed, as far back as 2004, Fitch Ratings had issued an alert to vendors of US residential mortgage-backed securities (RMBS) that there was a significant delay/lag in translating market trends into the data used by AVMs (this decision was subsequently reversed in 2006).

In practice, AVM vendors as represented by the EAA, undertake accuracy tests for institutional clients and report back the results.

Testing for accuracy can be undertaken, using EAA’s descriptive labels, in two ways:

- Lender Tests (where the lender provides the data)
- Bulk Tests (based on the AVM vendor’s data)

Both are ‘blind’ tests using unseen data, which the AVM vendor has not used in estimating AVM valuations.

Lender tests are tests in which the subject properties (the sample) being tested are controlled by the user, typically a bank or mortgage lender. The user provides details of the sample properties to the AVM vendor, who then estimates the accuracy statistics. One issue here is that there may be an insufficient number of properties enabling a statistically robust conclusion to be drawn.

Bulk testing, on the other hand, involves testing properties extracted from the property database held by the AVM vendor. The vendor has to ensure that these will be strictly blind tests for the purpose of computing the result. Here the AVM user, the bank, has to take this on trust and is not in a position to verify the integrity of the procedure.
The results of these tests are not publicly disclosed, and consequently, it is not possible to obtain an insight into the accuracy of a vendor’s AVM valuations.

Nonetheless, for individual property AVM valuations, an indication of accuracy is necessary for transparency. Furthermore, this will provide valuers with essential information enabling them to take a view regarding the uncertainty surrounding an AVM valued property.

- **Consequently, a recommended minimum requirement is that standardised results are reported in AVM individual property valuation reports (Section 4.3).**

The accuracy of the tests is evaluated relative to a benchmark value, which may be either a valuation of the sample properties by a professional valuer or confirmed achieved market sales prices.

It is preferable to have a sales price benchmark reference against which to assess accuracy, as in the U.S., and not a valuation-based benchmark, which European vendors tend to use.

- **It is the opinion of the author that actual sales prices should be used as the relevant benchmark.**

Each AVM valuation should be accompanied by measures of uncertainty, including a range of values which the eventual sales price is likely to be within, that is, the confidence interval. Various other statistical measures should also be made available (Section 4.3).

### 4.3 MEASURING, REPORTING AVM PERFORMANCE AND ASSOCIATED ACCURACY

This section lists the **minimum required information**, thereby facilitating more transparency in AVM vendors’ reports to valuers/individuals. A number of aspects need to be taken explicitly into account in AVM reports. On reporting, the European AVM Alliance (2017) comment:

"Reports should contain information on the origin of the property data (i.e. Sales Prices, Surveyor Valuations etc.) as well as information on non-property specific data that are used to calculate the value. This includes the granularity of those data. Value estimates produced by Comparables based AVMs should be reported in conjunction with confidence information (i.e. Confidence Levels or Forecast Standard Deviation). Information on the number of comparables selected to calculate the value should also be stated."

The following 10 pieces of information are recommended for inclusion in AVM reports. The list contains the individual property’s risk metrics together with risk measures of the associated comparable information. The comparables risk information provides a perspective on how accurate the AVM model is when tasked with valuing the comparables data, where sales prices are known. The model’s value estimates can then be compared with the known sales prices.

1. The i) 50% and ii) 95% confidence intervals of the AVM valuation
2. A clear explanation, accompanied by a ‘legend’, of the ‘confidence score’ or ‘confidence level’
3. Confirmation that comparables have been used in the AVM valuation. If not, what method was used in the AVM valuation?
4. The distribution of the standard deviations of the comparables used in the AVM valuation, if they are used
5. The AVM model’s overall accuracy, based on the comparable sales sample using:
   i. Mean Absolute Error
   ii. Median Absolute Error
   iii. ‘Error Buckets’ for the percentage of valuations lying within +/- 5%, +/- 10% and +/- 20% of the Sales Price
6. The number and the overall geographic distribution of the comparables used in the AVM valuation
7. The range of comparables sales prices used in the AVM valuation
8. Confirmation of the earliest and most recent sales dates of the comparables used in the AVM valuation
9. If ‘adjusted’ comparable sales prices have been used, explanation of how they were adjusted
10. Confirmation of the Benchmark used in arriving at the figures in 4. and 5. above, sales prices or valuations in arriving at the overall accuracy figures.

4.4 ADVICE ON THE USE OF AVMS

Much has been written on this aspect, particularly in the USA by the Appraisal Foundation and the International Association of Assessing Officers (IAAO).

THE GUIDANCE OF THE APPRAISAL FOUNDATION (U.S.)

The Appraisal Foundation produced its latest Uniform Standards of Professional Appraisal Practice, effective from January 2018 (2018-2019, edition). Advisory Opinion 18 (AO-18) is concerned with the use of AVMs, where advisory opinions are issued to address how an appraiser may use an AVM and to illustrate the applicability of appraisal standards in specific situations and offer advice for the resolution of appraisal issues and problems.’ Specifically,

- What steps should an appraiser take when using an AVM as a tool in the development of an appraisal or appraisal review concerning an individual property?

The AO lists five conditions under which AVMs can be used, namely:

- Does the appraiser have a basic understanding of how the AVM works?
- Can the appraiser use the AVM properly?
- Are the AVM and the data it uses appropriate given the intended use of assignment results?
- Is the AVM output credible?
- Is the AVM output sufficiently reliable for use in the assignment?

THE GUIDANCE OF THE IAAO

When using an AVM, it is noted that an appraiser should have a basic understanding of how the AVM analyses data to determine whether the AVM measures and reflects market activity for the subject property. Furthermore, it will be sufficient to describe the AVM’s overall process and verify that the AVM consistently produces results that accurately reflect prevailing market behaviour for the subject property. There is no requirement to understand or explain the underlying AVM’s algorithm or the statistical or mathematical formulae.

(IAAO, 2003, p. 30)
Given the evolution of AVMs over the last two decades, there remains an absence of regulatory standards. The use of more complex methods will likely pose additional problems in auditing AVMs due to the increased competence requirements imposed upon auditors and the reliance on more experimental testing methods to support AVM analysis. What does seem apparent, however, is that rigorous standards for the independent scrutiny of AVMs are needed.

From a practical perspective, some form of independent certification that such standards had been met would provide confidence in AVM valuations. This is for the future. However, a more undemanding and attainable requirement is for AVM vendors to provide transparent information requirements in their AVM reports.

The report has identified a minimum set of information which will help to open the transparency door. Indeed, it may be that the AVM vendors’ industry will recognise that providing more transparency in their reports along the lines suggested would facilitate a wider and more supportive acceptance of their product.

The current buzzword in data mining/machine learning investment circles is ‘quantamental’, which captures the essence of combining the output from an algorithm and fundamental analysis. Consequently, it may be that the partnership of AVM output and valuer analysis will result in synergies.

5. CONCLUSION
HOUSECANARY

HouseCanary kindly provided figures for the percentage of properties, by size i.e. value band, whose values fell within +/- 2 per cent of the eventual sales price. This is a very demanding margin for error. Figure 1 shows the distribution of the figures.

FIGURE 1

It is seen that highest levels of accuracy are around 40% for properties within the US $200k – 600K price range. For low valued properties, less than US$ 100k, some 17% of properties were valued with the +/- 2 per cent range. Interestingly, accuracy reaches a maximum of 42% for properties within the US$400-500k bracket and steadily declines to a figure of 21.5 per cent for properties with values in excess of US$ one million.
Figure 2 summarises the distribution of accuracy figures by size of property for different levels of accuracy. The figures show the percentage of properties valued within +/- 5% of the sales price and within +/- 20% of the sales price, across 12 price bands.

The figures may be summarised as follows:

- What stands out for the lower valued properties, less than US$ 50K: less than 10% of the valuations were within +/- 5% of the achieved sales prices, rising to 27% within +/- 20% of the sales price.
- For property values in the range US$ 100K – US$ 1 Million, the figure shows that in excess of 85% of values are within +/- 20% of the sales price, with 80% of properties valued in excess of US$ 1 Million falling within the +/- 20% band.
- If 10% is seen as an acceptable margin of error, for properties with values in excess of US$ 100k, some 75% of the valuations fall within +/- 10% of the sales price, and hence, one-quarter of AVM generated appraisals will have errors greater than +/- 10%.

The median absolute percentage error across all properties (674 metropolitan statistical areas, MSAs, as at June 2017) is 5.6%, i.e. half of the errors nationwide were within 5.6% of the final selling price, and half exceed 5.6%.
**Zillow**

Zillow claim to be the largest provider of AVMs in the US. The following histogram, Figure 3, compares the distribution of Zillow’s AVM accuracy rates across 864 US Counties. Different location, counties, will have different valuation errors and the histograms show the distribution of AVM valuation accuracy within each of the individual 864 Counties for different levels of accuracy.

**FIGURE 3:**

COMPARISON OF DISTRIBUTIONS FOR DIFFERENT ACCURACY LEVELS

Source: Zillow and author’s summary
The figures may be summarised as follows:

- The median level of valuation error across the US is 4.3% (as at June 2017), i.e. half of the errors nationwide were within 4.3% of the final selling price, and half had an error exceeding 4.3%.

- Top histogram: shows +/- 5% valuation accuracy rates; middle histogram: +/- 10% valuation accuracy rates and bottom histogram: +/- 20% valuation accuracy rates. The histograms provide a detailed visual insight across 864 US Counties.

- The average accuracy across all counties is superimposed in order to provide a reference point. On balance, it appears that some 50% of the valuations are likely to be outside the +/- 5% range of achieved sales price, which falls to almost 30% for the +/- 10% range and 17% for the +/- 20% range.

- As shown in the histograms, given the skewed nature of the distributions, even at the wider range of +/- 20%, there exist a significant proportion of valuations in many locations which lie outside the specified ranges of accuracy.

The key features of Zillow’s figures are:

- At the individual County level, the median ranged from some 2% to 32%, which represents a wide range of variation in accuracy across the different locations.

  » Almost half (47%) of all valuations across all Counties were within +/- 5% of the sales price, half being in excess of +/- 5%.

- One quarter of the Counties (215) had less than 35% of the valuations falling with +/- 5% of the sales price.

- However, 10% of the Counties had less than 25% valuations within the 5% bracket. The lowest recorder accuracy was 3.5% and the highest recorded 80%.

- The average percentage of valuations across all Counties falling within +/- 10% of the sales price is almost 70%. However, this can vary between 20% and 98%, depending on the County.

- On average, the percentage of valuations across all Counties falling within +/-20% of the sales price is 83%. However, this can vary between 32% and 100%, depending on the County.
There have been a variety of publications establishing AVM standards, especially by US organisations, notably:

- **The Appraisal Standards Board** in its Uniform Standards of Professional Appraisal Practice (USPAP), particularly Advisory Opinion 18 (AO-18)
- **IAAO Standard on Automated Valuation Models (AVMs)**, recent unpublished update Exposure Draft, September 2017
- **Collateral Risk Management Consortium**, CRC Guide to Automated Valuation Model (AVM) Performance Testing
- **Collateral Assessment & Technologies Committee (CATC)**, Best Practices in Automated Valuation Model (AVM) Validation

Several AVM providers and companies analysing AVMs have produced reports discussing, and recommending, approaches towards validating AVMs. Of particular note, **CoreLogic** in the US have produced several noteworthy Whitepapers on the subject, including Automated Valuation Testing (2011) and, Connected Analytics, in Australia, Best Practice Validation and Comparison for Automated Valuation Models (AVMs) (2015).

The following observations draw on of some the salient features arising in these publications.

The Collateral Risk Management Consortium (2003) issued their CRC Guide to Automated Valuation Model (AVM) Performance Testing (2003). The guide represents an extensive review of AVM performance testing practices. The CRC noted that lending institutions struggle with unique AVM databases, proprietary technology and systems and very different quality control measures, recognising that there was no consistent way to evaluate or rank AVMs in relation to any industry standards. The guide recommended a number of tests for evaluating residential AVMs. Many of the suggested metrics have found currency in current discussions on AVMs.

- Hit rate
- Comparison of AVM valuation with sales price
- AVM percentage error
- AVM bias
- Number of outlier outcomes
- Comparison of AVM percentage error to vendor confidence level

In 2003, the International Association of Assessing Officers (IAAO) in the US published a detailed report providing standards for the use of AVMs. The IAAO standard went into considerable detail regarding many aspects of AVM development covering such matters as data capture, model specification, calibration and the delivery of outputs.

The European AVM Alliance (ibid) set out broad guidelines and standards for statistical valuation methods. The standards outline the types of statistical valuation, reviewing their suitability for different uses. Furthermore, they discuss how the different parties such as lenders, investors and regulators should evaluate the resulting performance of each approach.
The declared objective in setting out standards by the EAA is described as follows:

“These Standards for Statistical Valuation Methods for Residential Property in Europe are intended to provide for the first time a coherent set of information and descriptions aimed at increasing the understanding, transparency and clarity on the wide array of existing Statistical Valuation Methods. The document focuses on principles, definitions and minimum requirements for Statistical Valuation Methods applicable across European jurisdictions.”

It should be remarked that the EAA are not very prescriptive in setting out their ‘standards’. Indeed, the recommendations fall well short of what is set out in the US documents, being broadly framed in general terms. It remains to be seen if any standards are implemented in practice.
REFERENCES


Collateral Assessment & Technologies Committee (CATC 2009), Best Practices in Automated Valuation Model (AVM) Validation. 2nd ed.


CoreLogic White paper (Sept 2009), Innovation in AVM Testing: When what you see is really what you get.

CoreLogic (2010), Forecast Standard Deviation & AVM Confidence Scores

CoreLogic (July 2011), Automated-valuation-model-testing

EMF/EAA (May 2016), Joint paper on the use of automated valuation models in Europe

European AVM Alliance (March 2016), Response of the EAA to the BCBS second consultative document on ‘Revisions to the Standardised Approach for credit risk’.


CoreLogic (December 2010), Statement on new interagency appraisal and evaluation guidelines.

Connected Analytics (2015), Best practice validation and comparison for automated valuation Models (AVMs).

TEGoVA (2017), The Accuracy of Automated Valuation Models (AVMs).

Lipscomb, C.A. (March 2017), Valuation: The Next Generation of AVMs, Fair & Equitable, IAAO.


IAAO (2003), Standard on Automated Valuation Models (AVMs), International Association of Assessing Officers, Chicago IAAO.

IAAO (September 2017), Exposure Draft Standard on Automated Valuation Models (AVMs), Chicago IAAO.

RICS (2013), Automated Valuation Models (AVMs), The Royal Institution of Chartered Surveyors, London
