connected ANALYTICS

Best Practice Validation and Comparison for Automated Valuation Models (AVMs)

Supported by

CoreLogic®

October 2015
Dr Ross Gayler
Dr Debashree Sanyal
Roy Pugh
Siân King

Contents

Abstract

The residential mortgage market in Australia and New Zealand is currently estimated at more than $1.4 trillion and is growing at a faster rate than earnings. Constituting the lion's share of household debt nationally and posing a serious concentration risk for most of the retail banking system, it is of major importance to the stability of the Australia and New Zealand financial system to ensure that these loans are well managed and adequately secured.

The cornerstone of mortgage origination and portfolio management is the valuation estimate of the property the loan is secured against. A variety of techniques exist for generating valuations, but with ever growing competitive pressures necessitating high performance alternatives at low cost and high speed, Automated Valuation Models (AVMs) are becoming increasingly important to the lending industry, with as much as 30% of residential property valuations now being conducted using these tools. In addition, they have been officially recognised by APRA as being an appropriate valuation technique in appropriate circumstance, but only where such models are thoroughly analysed and lenders independently validate the reliability of the model outcomes.

Given the influence of these models on national finances, it is imperative that their outputs are as accurate as possible for the purposes for which they are intended, to ensure that lenders are in the best possible position to select the right tool for the right job. However, it is not an easy matter to reliably measure relative performance between AVMs. AVM providers build models using a variety of techniques, which are kept under lock and key as proprietary intellectual property, and also call on widely differing data sources as inputs. In addition, the performance statistics the AVM providers report for the models are not standardised across the industry. All this conspires to add to the difficulty in making comparative assessments between the models.

As such, it is at best difficult, and at worse misleading, to try and compare different AVM solutions, in particular with respect to their accuracy. The problem becomes exacerbated as the market for residential mortgages continually changes at different rates across the country.

Of the AVM providers, lenders and consumers affected by these models, only the lenders are currently in a position to address this lack of consistency. By acquiring suitable test data and constructing an independent, robust validation regime, the users of these models can (and in some cases do) evaluate the strengths and weaknesses of the models under various conditions related to their use.

Connected Analytics believes that the adoption of a consistent testing regime by the lending industry as a whole would not only promote internal understanding of model effectiveness, but would also lay the foundation for establishing industry wide standards, from which the collective focus would promote a culture of continuous improvement.

So, with the over-riding ambition of promoting industry best practice in AVM validation, this paper provides an overview of AVM usage in Australia and New Zealand; aims to increase transparency on the key issues any user of AVMs needs to understand when evaluating multiple AVMs; highlights the risks to any lender in making misinformed assumptions about an AVM that they then subsequently use to make business decisions; discusses the major challenges in establishing reliable validation tests for AVMs and concludes with recommendations for establishing an appropriate testing regime.

Australian & New Zealand Property Market

According to the market update provided by the Australian Prudential Regulation Authority (APRA) on 25/02/2015, mortgage lending provided by the Australian Authorised Deposit taking Institutions (ADIs) at the end of 2014 exceeded $1.278 trillion in residential term loans. While in New Zealand the Reserve Bank reported $0.199 trillion for the same period. With national population statistics of approximately 23.5 million and 4.5 million respectively, this equates to a $52,750 housing loan to every man, woman and child in the region (taking exchange rates at parity).

Yet despite the sizeable burden of servicing this debt, the willingness of these two populations to tolerate long-term financial pressure for the privilege of owning their own home(s) shows no sign of abating. Last year Australian mortgage debt grew by 9% (~$100b) while in New Zealand the figure is closer to 4.5%, with population growth rates of less than 2% and 1% respectively, the average individual commitment to home lending is on the increase.

Property Valuations

A key determinant of the amount banks are willing to lend against a property is the estimated valuation of the property, as this indicates the amount of security the lender believes it can fall back on if circumstances lead to the need for repossession.

This valuation estimate also becomes the denominator in the all-important loan-to-value ratio (LVR), a measure of the amount the lender will advance to the borrower as a proportion of the estimated value of the property they seek to purchase. The numerator in the LVR (i.e. the loan) is typically a function of the borrower's ability to service the loan, their risk characteristics and the valuation itself. Once set, the LVR then has important consequences for the pricing of the loan, the availability of other credit and whether or not the borrower will be required to take out Lenders Mortgage Insurance (LMI) to protect the banks against the increased default risk associated with higher levels of lending against value.

Property valuations then, are crucial to the loan transaction process, have significant impact on national debt levels and affect the day-to-day financial pressures borne by "average" individuals as they go about their business. Therefore, it is in the interest of lenders and borrowers alike that the valuations produced are as accurate as they reasonably can be. However, this is by no means easy to achieve. Property features are as individual as the people that occupy them, as are the economic, social and geographic landscapes the properties themselves inhabit.

Valuation Types

Historically, valuations were predominantly undertaken by qualified professional valuers who would visit properties of interest and make an assessment based on its observable features plus any specifics peculiar to its location. While these types of valuation still occur today, because of the costs involved and the time taken to process the valuation produced, they have been supplemented by new methods of estimation enabled by developing technologies and data analytics. The adoption of these alternative valuation tools has been driven by the competitive need to optimise processes and drive efficiency. A brief synopsis of the main types of property valuation available in Australia and New Zealand today is given below:

- Full Valuation – valuation performed by a qualified valuer who inspects the property and completes a short form or long form valuation. The content of this report is governed by the Australian Property Institute (API) and New Zealand Institute of Valuers (NZIV). Turnaround time is typically 3-5 days from the valuation being ordered
- Drive by / Kerbside - assessment performed by a qualified valuer who inspects the property from the street only. Turnaround time typically 1-3 days from the valuation being ordered
- Desktop - assessment performed by a qualified valuer who uses data and photos to value the property remotely. Turnaround time typically one day from the valuation being ordered
- Automated Valuation Model (AVM) – uses a statistical model that looks at a variety of data (property attributes, comparable property sales, subject property prior sales, geographic price performance etc.) to estimate the value of the property at a given point in time. The valuation can be performed real-time (on the spot) with a Forecast Standard Deviation (FSD) provided to measure the certainty in the valuation amount
- Existing valuations (often indexed up to reflect current conditions).

Which methods are employed, and in what composition, differs from lender to lender but roughly speaking, the higher the perceived risk (most commonly based on the absolute loan amount and/or the higher the LVR) the greater the need for accuracy and the higher the reliance on human judgement in the valuation process.

This approach is partly driven by the risk appetite of LMI providers, who set restrictions on the use of desktops and AVMs, in particular at higher LVR levels, and partly by the banks' internal preference for expert opinion over automated outputs at higher risk levels.

This could be taken to imply that both parties believe that expert driven evaluations provide the greatest accuracy available when needed.

Whether this is true is open to speculation. Human eyes are much better at evaluating non-standard information such as property features, upkeep or neighbourhood issues, than a mathematical model that generally relies on homogeneity of features, measured in sufficient volume, to achieve desirable levels of accuracy. On the other hand, models do not suffer as much from human frailties such as inconsistency of performance, variable standards in education and experience etc.

Resolution of this question is still someway off. There is growing data on the performance of statistical approaches to valuations by virtue of their data driven approach, however, less data is available to evaluate the quality of judgemental valuations, which still account for the majority of valuations across the lending industry.

Whether or not loan size and risk level are the appropriate criteria for selecting the approach to producing a valuation, as implied above, there are certainly circumstances where the trade off between expert opinions and statistical methods needs to be understood. In general, these circumstances relate to the volume and quality of data available for assessment. Where there is sufficient quantity and quality of appropriate data then AVMs perform at their most optimal. However when data is sparse, volatile or of questionable quality then it is prudent to fall back on traditional valuation methods.

While human judgement verses automated logic is not the central thrust of this paper, the determination of which method should be used and when, should ultimately be decided by a robust testing process, in line with the recommendations proposed in the concluding sections of this report.

Automated Valuation Models

Regardless of which method is deemed best overall, or best under certain circumstances, it is clear that the reliance on AVMs by Australian and New Zealand banks to provide property estimates is increasing. Informal evidence from two of the "big 4" banks in Australia indicates that more than 30% of all mortgage applications are valued solely using AVMs, while up to a further 20% of applications are valued using AVMs in conjunction with other methods.

But what is an AVM? Descriptions abound, typically referencing "mathematical", "statistical", "algorithmic" and other "logical" frameworks/tools which take as their inputs data from a variety of sources such as property features, market data, economic trends, price distributions etc. and may also overlay the valuations estimated by the model with known sales data if available.

The components that make up an AVM are borne of a variety of techniques and may include price indices, regression models and decision trees to name but a few. The picture is further complicated by lenders and AVM providers who sometimes combine various AVMs to produce a compound "hybrid" tool, called a cascade, which includes rule sets that determine under what conditions each individual AVM is called upon to contribute to a valuation estimate.

The unifying aspect of AVMs is that their various component features are developed analytically and are driven by data to generate their price estimates. However, this still leaves tremendous scope for the design of AVMs. In the end, the modelling methodology adopted by the provider may be influenced by the breadth and types of analysis data available at the time of model development, the skill set of the development analysts employed, implementation considerations and ever present time and cost constraints.

Whichever techniques are adopted, they are generally considered to be the intellectual property of the AVM provider and a source of competitive edge. As such, the specific detail behind an AVM design is rarely shared.

Despite the significant structural differences in AVMs and the data they process to generate a valuation, the majority of AVM providers typically offer the same type of outputs related to their models. These are:

- Valuation estimates or ranges: These are the actual predictions of the value of the property being considered. Most AVM providers and users take this to be an estimate of what would be the realised sales price on that day
- Accuracy statistics: Given the definition above, most take the absolute measure of accuracy as the difference between valuation and the subsequent sales price. This is sometimes transformed into a ratio, known as percentage error, by dividing the difference by sales price. This is scale independent and allows comparisons across properties of different values
- Hit rates: This is the average percentage of properties that the AVM claims it can produce a usable valuation for. Claims of approximately 75-80% hit rate seem to be fairly common although much depends on the definition of "usable"
- Confidence scores: These come in many forms but are typically of the form "X% of valuations will be within Y% of sales price"
- Forecast standard deviation (FSD): Formally, this is described as the standard deviation of the percentage error, where the percentage error describes the relative difference between valuation and price. Intuitively, FSDs tell us that if the error between valuation and price varies significantly, then the AVM does not perform consistently across the data being evaluated and further investigation is required to determine where the model is strong and where it is weak with respect to a sample of properties.

Note that, while providers typically report outputs similar to the above, there is no agreed industry standard as to how these metrics are defined and calculated. As such, generally speaking, they are not directly comparable across different AVMs.

However, with respect to the actual valuation estimate produced by a model, the majority of AVMs on the market today aim to predict the final sales price of a property under consideration. This is not as straightforward as it seems. Regardless of the quality and performance of an AVM, the sales data against which it is typically benchmarked contains random variance that can never be accurately modelled. For example, a given group of bidders competing for the same property at two separate auctions may end up with two very different sales prices, depending on the weather, the performance of the auctioneer, the time of year etc. Human sentiment is not consistent.

However, it would be surprising if there was not a strong correlation between valuation and sales price as the two are not independent. As mentioned, the valuation produced plays a role in how much a lender is prepared to advance to a borrower and at what price. The amount borrowed in turn dictates what the consumer can afford, which in turn affects property demand and so ultimately, price.

Overall, it can, and has been, demonstrated that well designed AVMs perform to an impressive level of accuracy if applied appropriately. No one should expect a model to be optimal under all circumstances. Rather, it is incumbent on the user of these models to ensure that they have the required tools at their disposal to evaluate their strengths and weaknesses relative to their intended purposes.

Uses of AVMs

So what are those purposes? As mentioned, AVMs already drive a significant proportion of the valuations used for originating mortgage loans. The ability to provide an on-the-spot valuation and approval in principle is an important device for locking in an applicant before they decide to shop elsewhere. This speed to market requirement is particularly important where an intermediary is involved, with the potential to introduce additional steps into the loan decision process (at the time of writing, approximately 50% of all residential mortgages in Australia and 30% in New Zealand are originated through a broker). To mitigate this risk, some lenders are making the AVMs they subscribe to available to intermediaries to enable rapid decision processes.

But loan origination is just one of several ways lenders are employing these versatile tools to get the most out of their technical advantages. As a consequence of their low cost relative to other forms of valuation, automated high-speed delivery and the objectivity of their outputs, AVMs are the obvious toolset for revaluing entire mortgage portfolios on a periodic basis. A few examples of where this has proved useful are:

- Using a dynamic LVR to define securitisation tranches
- Updating collateral management systems
- Using dynamic LVRs in product and customer behavioural scoring
- As an input into PD & LGD modelling and stress testing
- As an input into marketing propensity models.

With ongoing improvements in modelling methodology, data sources, computing power, low cost data storage and the growing variety of delivery mechanisms, it looks like AVMs are not only here to stay but will continue to grow in usage and influence.

Misuses of AVMs

So that is the good news. The bad news is that misusing AVMs, or at least using them without having a good understanding of the variation they can introduce, can be costly. Confidence levels of AVM accuracy considered acceptable across the AVM industry are typically expressed along the lines of: 70-80% of properties valued will fall within 10% of the realised sales values. However, even if 100% of valuations were within a 10% tolerance, when applied to a $100bn mortgage portfolio using AVMs between 30-50% of the time, this could imply that the loan security is underestimated by as much as $3-5bn due to expected errors in the model estimates. Note that all four of the major Australian banks have portfolios considerably in excess of $100bn.

The counter argument to this extreme view is that valuation errors at the individual property level might be expected to cancel out, at least to some degree, when rolled up to portfolio level. While this is probably true when applying appropriate, well-structured models during stable economic conditions, the implementation of a model that is unsuited to the portfolio (or segments thereof) or quickly changing house price trends can easily introduce systemic bias. In particular, in a falling market when risks are greatest, a model is more likely to lag the trend than lead it, resulting in over-pricing and the issues outlined above.

Of course, on-site valuations also come with an inherent error distribution, hence the earlier comment that valuations based on human judgement should be subject to a vigorous testing regime, just as we're advocating for AVMs.

In Australia currently, regulatory guidance for ADIs on valuation practices is provided by APRA's APG223. This regulation also governs the four largest New Zealand banks given they are owned by Australian banks.

"A prudent ADI contemplating the use of alternative valuation methods such as AVMs would subject proposals to thorough analysis and develop a risk management capability that includes ….ongoing monitoring of tools used, processes that capture evidence of action taken when values are deemed to be unreliable and periodical back-testing undertaken by the ADI to independently validate reliability of outcomes …" (see Appendix 1 for further relevant details from APG223).

This sets out what is expected of the lender in terms of valuation management. While the guidance is sound it is, generally speaking, non-specific and leaves considerable leeway for the user of AVMs to decide how it chooses to satisfy regulation.

So, the question occurs, to protect against these potential issues, should AVM providers or lenders apply a conservative adjustment to bias the models towards under-pricing? Unfortunately, this approach leads to few, if any, winners. Sellers want to sell for a profit and are likely to resist valuations that fall well short of their expectations and desires; borrowers would be restricted in what they were allowed to borrow and struggle to meet the sellers' demands; lenders would lend less and make less profit; and AVM providers would fall from popularity.

It seems that the only realistic way forward is for AVM providers to make the model outputs as accurate as possible and for users of AVMs to have clear testing processes in place to understand the variations in the estimate of value. This will inform the conditions under which the models should, or should not, be used.

Which AVM?

As should be clear by now, the several providers of AVMs across Australia and New Zealand each provide models that differ in design and data usage, with the consequences that each model has its own particular strengths and weaknesses. To further confuse the distinctions between them, the performance metrics reported by the AVM providers (accuracy, confidence scores etc.) do not conform to any agreed definitional standards and hence require different interpretations. But it's not just a definitional issue. The tests undertaken by the AVM providers, which generate these statistics, do not conform to any industry standards either. In short, the AVMs on the market are not directly comparable based on industry supplied measures.

So where does this leave the user? The selection and continued use of a valuation model is not a trivial problem. We've already outlined the influence these models have and the risks of using them inappropriately. But lenders need to develop practical policy outlining which valuation method to apply in what circumstances and also, for those institutions that subscribe to more than one AVM, how to choose between them. In the current environment, with its notable lack of standardisation, the only sensible recourse open to users of multiple AVMs is to establish their own in-house testing regime.

The remainder of this paper explores the key considerations to bear in mind when constructing a testing regime for AVMs and concludes with Connected Analytics' recommendations for establishing better practice.

Model Validation

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.

> Given the estimated portfolio value under the influence of AVMs, it is curious that they are not subject to the same level of scrutiny by lenders as are other credit risk models. However, it is our contention that where possible, they should be.

As with most analytic models, when considering performance testing of AVMs they can be viewed as consisting of three standard components:

- Input data
- The predictive model (algorithm)
- The outputs of the model (valuation estimates).

Of these three components, the easiest and most important to test by the AVM users is the model outputs. These are the estimates that are used to make financial decisions and so they, above all else, need to be demonstrably accurate. Therefore, the focus of what follows is predominantly concerned with the key factors an AVM user needs to consider when establishing a testing regime of the model outputs, for the purposes of evaluating an AVM both in an absolute sense and relative to other AVMs.

However, it is acknowledged that not all AVM users have access to the resources they need to comprehensibly test the model outputs under all the conditions in which they are expected to perform, nor the resources to test the models as regularly as they should be to determine that they remain "current". In these circumstances, a solid understanding of the strengths and weaknesses of the input data and model construction take on extra importance, to allow the user to gauge whether a particular model is being (or will be) appropriately applied in an area of strength or not. Unfortunately though, information relating to data inputs and model methodology is not always readily available from the AVM provider for proprietary reasons. Therefore, while we address what we believe are the important considerations for these components below, to obtain the information needed to make an assessment of their qualities is likely to prove a headache for the user.

Errors or biases introduced at any one of these three stages can result in poorly performing valuations. Therefore, we consider each of the model components separately.

Input Data

In the terminology adopted by analysts, they talk of a model under development as being "trained" on a particular set of data. The term "trained" is very pertinent here because, like humans, a model can only "learn" from what it sees and, when implemented, will assume that all future input data contains the same trends and information value as the training data did.

So first and foremost, when considering the input data that an AVM sources, the primary question to ask is "is this input data consistent with the model development data"? If it isn't, the model cannot be expected to either recognise the input data or evaluate it correctly. Where this kind of issue shows up is if the model is applied to new geographies, new property types (with different arrangements of features) not present in the analysis sample, extreme price ranges not present in the analysis sample etc. This is a crucial point to be clarified. Stated another way it asks "is the profile of the properties to which the lender intends to apply the AVM consistent with the profile used to train the model"? If not and the user is not in a position to test AVM outputs, consider a different model or valuation method.

A related issue pertains to whether the AVM provider has changed the mix of data sources it subscribes to. This may introduce variability due to differing standards in data quality, alternative methods for recording property features and different approaches to how frequently data is updated. All of these variances can have a negative impact on model performance.

Another important point to consider is whether and how the model factors in sales data. Or, more specifically, how the AVM performance reporting is influenced by the inclusion of sales data.

This might be in the form of previous sales data for the property under consideration, which may or may not be adjusted to account for the effects of inflation, or adjusted to stay in line with price movements of "comparable" properties in the area. Or it may be appraisal data, produced before the final sale price is realised and used as additional intelligence to "inform" the model. The reason this is important is that overlays of supplementary information blur the reported statistics for the performance of an AVM. If a lender was to implement an AVM but was unable or unwilling to generate appraisals and factor them into the property valuation process, it is likely that the levels of valuation accuracy observed over time would be lower than those reported by the AVM provider.

As such, it is recommended that where possible, the AVM is tested purely on the lenders own sample, not the general population. Where this is not practical or possible then the lender should make sufficient enquiries with the AVM provider to ascertain that the AVM assumes a process similar to their own, and if not, to select a sample that best does so.

The above points can be summarised as "what are the criteria applied by the AVM provider for the inclusion or exclusion of data, when using a model to produce a valuation"? As indicated, these criteria can have a profound effect on AVM performance.

The Model

As mentioned earlier, there are numerous techniques available to design an AVM and typically, more than one technique is adopted in any given model. Each technique adopted will influence how the model behaves over time, for example:

- How sensitive is the model to variation? Do minor differences in property design produce significantly different valuation results or does the model dampen such variation?
- Does the model assign extra weighting to recent sales such that price valuations evolve quickly in a changing market?
- If different techniques are embodied into a model's design, how does it prioritise which technique is applied under which circumstances?
- Is the model easy to update? The more complex the model, the more effort and expense required of the AVM provider to keep the model up to date. In these circumstances a model may not recognise emerging trends.
- Does the model contain systemic bias? That is, does it tend to over-predict or under-predict on average?

As with input data, for those unable to undertake a rigorous testing of the AVM outputs, clarity with respect to model design may assist the lender to understand the most appropriate circumstances in which to apply a model. However, as commented earlier, this is typically not information the AVM provider is willing to share.

Note that while it might be reasonably argued that a standard approach to modelling methodology for AVMs would assist users to validate models, Connected Analytics does not believe it is a reasonable expectation that AVM providers adopt a common modelling practice, indeed competition and innovation in this space is what will see the performance of these models increase over time.

Firstly, from a competitive point of view, a standardised methodology would provide a disincentive to enter the AVM market. Secondly, if all models were designed along the same principles, then they would all embody similar strengths and weaknesses, with the likely consequence of a resulting range of property types, geographies etc. that would be inadequately measured.

Our view is that these are valuable tools to the lending market and it serves no one's purpose to commoditise them. The advantage of adopting different modelling approaches is that it increases the likelihood that there is an appropriate model for the job at hand. The challenge is to find it.

Model Outputs

As mentioned, the AVM outputs are most amenable to analysis as these are supplied to the user and are available for measurement against appropriate benchmarks. If more than one AVM is being tested for a particular purpose then they should be tested under exactly the same conditions for their results to be comparable, i.e. benchmarked using the same analysis sample against the same outcome prices. The important outputs to validate are:

- Accuracy of the AVM estimate. As discussed earlier, this is typically measured as the price difference between the model valuation estimate and the realised sales price (or functions of this difference)
- Forecast standard deviation or confidence scores. This tells us how the model accuracy varies across the segments under consideration (geography, price range, property type etc.). From this internal confidence scores can be constructed predicting the percentage of hits that are within a required tolerance
- Hit rate. In conjunction with accuracy and FSD statistics, enables the user to obtain a detailed picture of where AVMs need supplementing with other assessment methods.

Before testing takes place, the AVM user should consider the intended purposes of the AVMs and the levels of performance required for each purpose. For example:

- For originations the specific value of an individual property is of interest and so the accuracy of each transaction is the metric we are looking to optimise. The lender should be able to estimate the impact of valuation error on such quantities as LVR and estimate corresponding impacts on risk estimates for the mortgage applicants. Pricing the risk impact should then inform the level of valuation error tolerable. Note that these LVR sensitivities may vary depending on whether we are considering high or low risk applicants

- For account management, via behavioural scoring, valuation is typically factored into risk models via dynamic LVR variables. As for originations, specific individual risk may be important for credit limit increases, cross-sales etc. but in many models the LVR variable will be banded, which dampens the impact of individual error and so the accuracy tolerances may be acceptable at lower levels
- For revaluing entire portfolios, where the quantity of interest is the overall portfolio value (collateral management, capital allocation, etc.) then individual error becomes less important so long as the aggregated total is near the mark. Therefore the accuracy thresholds can be further reduced. Assuming the model does not suffer from serious systematic bias (to be checked for) then balanced coverage, i.e. hit-rates, becomes a more important consideration, with correspondingly higher thresholds established for this metric.

As with all things related to AVMs, there are no industry aligned benchmarks that can be used as guides on where to set respective thresholds. In any case, the lender is better off estimating values that are relevant to their own portfolios and risk appetite.

Validation Test Data

Once the required performance thresholds have been established for lending or management purposes as outlined above, the next step is to construct an analysis data set to undertake the validation tests. Even though there are many things to consider, the main thrust of the analysis is to segment the data based on key discriminators of sales price (geography, property types etc.) that are relevant to AVM users lending practices, and then to calculate the statistics outlined above to see how well the valuations perform on each segment, relative to the desired thresholds.

Therefore the first step in constructing the validation data set is to define the segments to be tested. Any segment that exhibits different price dynamics should be isolated for separate testing, where data is available and sufficiently robust, to enable accurate and reliable analysis

Potential segments include, but are not necessarily limited to:

- Units v houses v other property types
- Low and high value properties
- Large and small properties
- Different geographies
- Refinances v new purchases.

The lender should seek to understand the basis for including or excluding data from an AVM development sample and should expect the supplier of the AVM to provide this readily. Model performance on data defined similarly to the inclusions/exclusions is likely to be volatile and should be tested separately to avoid muddying overall results.

After the segments are defined, the following considerations need to be addressed to the construction of the validation data set:

- The AVM value: There are two ways to obtain this value. If the lender has previously received stored and dated AVM values as part of their business operations, they can simply retrieve these. However, in instances whereby the lender is using AVMs for the first time, or expanding their use and/or they do not have sufficient numbers of observations for robust analysis, they will have to obtain a retrospective file from an AVM provider.

Great care needs to be exercised in understanding the AVM provider's process when recreating the retrospective AVM estimate.

The retrospective should be an accurate reflection of the AVM that would be available in a live environment which means excluding the use of any data which occurred after the AVM date AND any date which would not have been available to the AVM at the time (i.e. occurred but the AVM provider was not yet aware of it and/or had not yet loaded and made it available within its systems)

Example: Assuming we wanted an AVM estimate of number 64 High St. on the 1st October. If number 66 sold on the 15th October, we obviously could not use the information because it was not available at the time of the AVM calculation. However, assume number 62 sold on the 8th Sept.

Whether that information can be used or not would depend on whether it would be reasonable for the AVM model to have access to this information, which means it would have had to have been loaded AND made available, prior to the AVM being run on the 1st October.

One of the ways certainty in retrospective AVMs can be achieved is for AVM providers to supply a file that is held and tested at a later stage with subsequent sales. This method has been used by auditors in New Zealand to give certainty that AVM providers are not using data they do not have access to at the time of valuation.

- Target variable: Despite the fact we know that sales price incorporates variance, this is the value that most organisations aim to test against, as it is the final sales price that ultimately drives all the financial transactions. If insufficient sales data is available from within the organisation performing the testing, then data may be obtained from an external data provider (who will typically also be one of the AVM suppliers)

  Note: If limited data availability necessitates the use of appraisal estimates then it may be possible to use as a substitute. However, due to the nuances around this data, it should not be "mixed in" with sales data (i.e. they should be analysed separately) and any results would need to be carefully interpreted

- Volume: Each segment to be tested should contain a sufficient number of records to deliver statistically significant and robust results for the metrics calculated on that sample. A constraint here is the potential for the AVM provider to limit the number of properties allowed for evaluation

- Random sampling: If possible, the properties used in the validation data set should be randomly drawn from the available data sources to avoid the unintentional introduction of bias. If conducted properly, the aim of this exercise is to ensure that sample mix of properties to be tested is representative of the population that the AVM user typically lends to

- Independence: Ideally, the validation sample would be sourced independently by the lender to avoid any potential for the AVM supplier providing data and/or services (such as address matching) that are over and above what would be available in the live environment. However, when the lender does not have sufficient data to undertake the analysis they should at least be very specific with regard to the sample construction and ask the provider to detail all steps, including the output results, in the process so that they can validate these against on-line processes.

One area whereby an AVM provider could "enhance" their results would be in data matching. If on first pass they obtained a low match rate(say 70%) then it could be possible that they would try additional matching processes, including manually reviewing the files to ensure a maximum match rate (say 90%).

Although this might appear to be a genuinely valuable additional step, in reality this could provide the lender with a false appreciation of the solution, building expectations they would achieve 90% when in reality only 70% would be realised in an automated environment without the capability of the manual process.

- History: It's important to understand how the models perform over time and in changing conditions. For instance, how does the model perform under various market conditions? A model that over-predicts in one direction may under-predict in the other. Likewise, does the model show signs of instability as the mortgage portfolio evolves over time due to a range of dynamic drivers such as change in the customer profile, origination policy amendments, directional change in lending strategy etc.?

  Care needs to be exercised when analysing historic data. If the history goes so far back that the model has been significantly modified during that period, then some of the results generated will reference a model that is no longer active and not be relevant to today's model(s). Also, be mindful of any changes the AVM has undertaken with respect to data suppliers, data formats, data cleansing techniques etc.

> AVMs may perform very differently during rising markets compared to falling markets. In particular falling markets may result in valuation lags that lead to underestimates in collateral requirements.

- Point of scoring: If using historic samples to compare the performance of two competing AVMs on a given segment, it is important to ensure that the time between production of the AVM estimates and the sales price used for benchmarking is consistent for both models and/or the methods used to index the values are consistent.

Analysis

With segments defined and a well-designed validation data sample constructed, the analysis can begin. As with any analysis, it is important to bear in mind that any results generated are only relevant to the context in which they were produced. If conditions change such as data supply, economic drivers or policy updates then results of the analysis may be invalidated, or at the very least subject to distortion. This necessitates that the testing itself needs to incorporate sensitivity analysis to identify volatile parameters that should be closely monitored after the analysis results are implemented. If the sensitive parameters move beyond acceptable trigger ranges, then this should be taken as an indicator that the analysis may need to be repeated or updated.

In addition to context, careful thought needs to be given to the interpretation of any metrics generated from the analysis, as there are subtle differences in what they measure. Small data samples are easily influenced by outliers that can seriously distort an average based on using the mean. In these circumstances it may make more sense to rely on a median or modal value, or alternatively identify the outliers and exclude them from the analysis.

AVM Metrics

Prior to undertaking any analysis, it is important to document how the AVM performance is going to be evaluated, which metrics are to be calculated, whether they will be given equal importance or will they will be weighted in accordance with their relative value to the lender.

Hit Rates

Often inappropriately considered the simplest of the AVM statistics to calculate and interpret, as it simply represents the proportion of successful valuations returned on a sample of properties, great care should be taken in understanding what constitutes a "hit" across different AVM suppliers.

To start with, is it important that the property itself be considered a hit?

Some AVM providers, if unable to locate the specific property may still return an AVM which would be based on the average AVM in the building, street or suburb

Although it might not seem appropriate to provide an AVM without locating the building, it is not such a straightforward issue. If for instance the building that was located contained 20+ properties all with previous sales prices of within 5% of each other, then the AVM returned would likely be of high accuracy.

However, some lenders use AVMs not only as a valuation estimate but also as pseudo fraud check. Applying the rationale that if there is an AVM against the property then it is most likely to exist, would not be a valid assumption in the process above.

Also, simply producing an AVM is not necessarily sufficient. For instance, if no data was available specifically about the property then a potential AVM estimate could be the region average. Not very accurate, but should this at least be considered a hit?

If the AVM provider already knows what level of minimum quality they require from an AVM then they may wish to apply this upfront and only consider an AVM hit when the property has been located and the AVM accuracy is above this threshold.

As such, if hit rates vary significantly across multiple AVM providers then it would be prudent to understand why, so that such differences can be appropriately considered within the evaluation. Sometimes having a lower hit rate could be a sign of more accurate matching and so should be a positive aspect of the AVM provider and not something considered undesirable.

Typically reasons for being unable to return an AVM include:

- Insufficient address data being supplied (missing house number or postcode for instance)
- Property not yet built (brand new estates may not have full information registered yet)
- Property has been redeveloped (for instance #1 The Street is demolished to make way for two town houses, 1A & 1B The Street)
- Non-standard or over-size houses that occupy multiple plots.

In such instances, the results should be consistent across each AVM provider.

However, as there is no definitive address database in Australia or New Zealand, each AVM provider will search against their own proprietary data sources. As such, other reasons for differences in AVM supply across different providers would include:

- Property is considered to be commercial, not residential
- Property is considered to be land only
- Key data pertaining to the property (number of bedrooms, bathrooms, land size etc.) is not available
- The accuracy of the AVM (due to a variety of reasons

including lack of comparable sales) does not meet the minimum standard

- Differences in the providers address matching routines (exact v fuzzy logic etc..
- Address not found.

All but the last reason are likely to cause issues for any AVM provider. If the last cause can be verified it may indicate the need for greater reliance on an alternative AVM, for that sample of properties. The best way to check this is to compare the hit-rate results for several AVMs on the same properties.

Also to note is the distribution of hits for properties in a given segment, particularly by geography, as it may be possible that an AVM provider has greater access to data in a region than their competitors. Such a result, assuming accuracy was equal, may lead to a conclusion that different AVMs providers should be used in different regions to maximise the overall country "hit" rate.

Note that some providers exclude AVM outputs with low confidence scores or high FSDs, deeming these records unreliable. In these circumstances, these records will not be included in hit-rates, indicating a lower rate than if left in. The consequence is that those attempting to control the quality of their outputs through the use if FSDs can appear to be performing worse than an AVM provider that leaves them in.

In all cases, the lender should work with the provider to investigate lower than expected hit rates or any apparent bias identified as a result of model validation.

Accuracy

The difference between sales price and valuation is the standard metric measured to determine accuracy. However, on its own it has its limitations. It is not sensible to compare a $100,000 gap between a valuation of $1,000,000 and a sales price $900,000, with a $5,000 gap between a $50,000 valuation and a $45,000 sales price, if the objective is to declare the smaller gap more accurate. They both represent a drop of 10% from valuation to sales price.

True, most people would rather save $100,000 than $5,000 but if the purpose of the test was to determine how effective an AVM was across a portfolio, the conclusion should be that it is equally effective everywhere, i.e. it is always over-pricing by 10%.

So, should accuracy be measured at an absolute level or a relative one? The answer is – it depends on what you want to know. Understanding changes in relative accuracy across the range of portfolio sales prices is useful for setting policy regarding the involvement of other valuation methods. On the other hand, knowing the true size of the error is important for granting actual loans.

At an individual property level, AVM estimates are more likely to accurately reflect sales price if:

- There is sufficient volume of "similar" properties in the given segment
- The model is current and was trained on a relevant sample
- The market is not moving rapidly.

To the extent that accuracy at the individual level requires volume, there is a clear relationship between accuracy and hit rates (see previous section). The more hits there are, the more chance of training a model on similar properties.

At a portfolio level, sufficient accuracy may be observed if only the last two conditions hold. However it is likely under conditions of low homogeneity of property features that the individual valuations will exhibit high variance (see FSDs below).

With respect to accuracy, an important distinction needs to be drawn between random error and systemic error.

Random error is a consequence of the fact that all models are approximations. This type of error is generally regarded as unpredictable, within the constraints of model design and the expectation is that taken over sufficient volume the errors will typically cancel out. Not much help with respect to individual transactions but a source of comfort when conducting portfolio level valuations.

Systemic error, on the other hand, is more concerning as it is likely that an AVM is either routinely over-predicting or under-predicting. This is both troublesome at the individual and the portfolio level and is typically caused by structural problems in the model and/or bias that has been introduced to the input data sources. This is normally detected through monitoring at the portfolio level and is corrected for by recalibrating the model.

A more difficult influence on accuracy of AVM valuations to detect, which has been alluded to a couple of times in the text, is retrospectively adding data that was not available at the point in time the valuation would have been performed in a live environment. This occurs when additional data is introduced to bolster the accuracy of the AVM valuations. For example, Valuer General sales data typically come in from the State entities within three months of the valuation. If comparable data is used retrospectively that was not available to the AVM provider at the date of valuation, it may artificially inflate the accuracy of the model for mortgage origination purposes.

In essence, this data overwrites or modifies the AVM outputs but to all intents and purposes is indistinguishable from model generated outputs. The main concern here is that misrepresentation of retrospective data availability could result in valuation strategy expectations not being met, i.e. it would result in a difference between production results and testing results, thereby putting at risk the success of such a project.

Of course, if the additional data is available to the user and can be factored into their lending processes then all well and good. But even here the AVM and supplementary data should be evaluated separately, so that the true contribution of each component to the valuation performance is transparent.

The resolution to this is not straightforward. In theory, it would be possible to construct a validation sample of recent sales from the user's portfolio data which were previously unknown to the AVM provider, in the hope that they had no supporting data already on hand and little time to get any new data. In practice this is fraught with difficulty. With today's high-speed transmission of data and the need to allow a sufficient timeframe between AVM valuation and the realised sales price, the timing of this operation would have to be so nimble as to be impractical.

More realistically, the user should insist that all post-AMV valuation data be backed out of the exercise and ask the provider to demonstrate that this has been done. In the long run, it is hoped that the industry moves away from retro-fitting validations.

Forecast Standard Deviations

Calculating the variance of accuracy estimates across a sample is a vital aid to understanding whether an AVM performs consistently at its average performance level or whether it performs better under some circumstances than others. Defined as the standard deviation of percentage error, FSD is the metric that provides this information on variance of performance.

Of the metrics calculated to evaluate AVMs, FSDs are perhaps the most informative. If the standard deviation component is calculated according to basic statistical theory, this metric opens up the AVM to the rigour of full statistical analysis.

Simply interpreted, a high value of the FSD tells us that the performance of the model varies considerably across the segment, while a small value indicates greater consistency of accuracy. Put statistically, one standard deviation should contain approximately 68% of all percentage errors, while two standard deviations will contain approximately 95% of all percentage errors. Therefore if a standard deviation is small, then two standard deviations is not going to be particularly big, and we can surmise that 95% of all percentage errors are smaller than this number. I.e., accuracy error is not a problem under these circumstances.

With statistically derived standard deviations it is an easy step to construct confidence intervals, i.e. make statements about the proportion of percentage errors falling within a specified range. By constructing confidence scores based on these confidence intervals, FSDs provide a means of linking the confidence scores directly to the accuracy of the validation estimates.

Unfortunately, not all providers take this route. For example, some providers base their confidence scores on the number of properties of a given type, in a particular suburb. These confidence scores cannot be linked directly to the error in valuation accuracy and so are less useful for making inter-portfolio comparisons.

Given the valuable nature of FSDs, like accuracy estimates, they should be validated at segment level to inform lending policy within that segment.

It is possible to reverse engineer FSDs to gain insight into the structure of the AVM. By grouping FSDs into bands (say ten), and ranking them from high FSDs to low FSDs, it is feasible to profile each FSD group to determine which properties, geographies etc. are most stable with respect to the valuations produced, verses those that are least stable. This in turn allows the tester to form a view as to which type of data was predominantly used to train the model and whether it is relevant to the lenders needs. If this type of analysis were to be relied upon, then it would also be necessary to validate FSDs by level, not just by segment.

Connected Analytics AVM Validation/ Comparison Recommendations

The following section presents the practices Connected Analytics believes should be adopted as a baseline for establishing a robust AVM validation regime in the Australian and New Zealand lending markets. In analytics, as in most things, there are many ways to skin a cat. Therefore the recommendations given are neither definitive nor exhaustive. However, Connected Analytics believe they form a sound foundation for effective AVM management and a basis upon which to start building an industry wide set of validation standards that would be to the benefit of all.

AVM Validation Principles

Validation findings should be actionable

Determining accuracy rates, hit rates etc. is one thing but doing something about them is another. Once a model's strengths and weaknesses have been profiled, the validation findings should be accompanied by an action plan to include:

- When and how the model(s) should be deployed (e.g. structured feedback to affected risk management and lending functions, appropriate amendments to lending policy etc.)
- The design of future monitoring and test activities
- When one model should be used in preference to another
- Interaction and cooperation with the AVM provider(s) to continuously improve AVM performance in line with the lenders needs

Validation findings should be informative

Several commenters aired the opinion that, generally speaking, AVMs were poorly understood by the lending industry and expressed the view that many users of AVMs would benefit from ongoing education with respect to AVM management. To that end, lenders should not only seek to understand validation findings in the context of their own portfolio management but also to understand the advantages/disadvantages arising from their use of particular AVMs, relative to their peers.

Validation findings should be incorporated into credit risk governance

It would appear that the growth in usage and influence of AVMs within lending institutions has evolved gradually and, consequently, to a certain extent unnoticed, such that they are not subject to the rigorous governance standards applied to mainstream credit risk models. Connected Analytics is of the view that these models are very important to modern credit risk management and will continue to be so. As such, they should be subject to the same high standards of review and control as are other models.

Validation findings should be based on sound statistical analysis and should be capable of independent reproduction

Perhaps the single biggest factor affecting a lenders ability to interpret and properly understand the AVMs is the lack of standardisation across the industry. While individual lenders can currently do little to immediately influence the industry position, an important first step is to ensure that their own internal validation practices are based on sound analytic principles that enable the lender to draw unbiased, defendable inferences from their test results.

Validation findings should be relevant

It has been emphasised several times throughout this paper that models need to be tested relative to the purpose(s) for which they are intended for use. Performance results pertaining to originations are not necessarily valid for portfolio management requirements. Similarly, nation-wide averaged performance results stand a good chance of being unacceptably inaccurate when applied to specific geographic segments of interest. It follows that the validation thresholds that a performing model is expected to meet should be tailored to the specific uses of the model.

Validation results should be current

Portfolios change, properties change, economies change, models degrade etc. Given the potential size of the financial impacts of systematically under-performing models, there is little excuse for not establishing regular monitoring and/ or re-testing of AVMs in use.

AVM Validation Priorities

The priorities recommended below have been distilled from the above commentary - informed by stakeholders and literature - and by and large are given in the order that we believe they should be considered.

1. **Establish Validation Purpose(s)**

   For example, testing an AVM's performance with respect to loan origination. Before any analysis is undertaken the validation team, in conjunction with relevant stakeholders, should list and rank the most important requirements for the AVM based on how it is going to be used within the organisation. When comparing multiple models they should define upfront the criteria and parameters for assessment of the models, and whether they are looking for one overall "winner" or whether they are looking to optimise overall AVM performance by using multiple models in conjunction with each other

2. **Define Validation Thresholds**

   In line with priority 1 and based on internally estimated financial impacts resulting from error rates of the metrics being measured, the lender should have pre-defined the validation thresholds considered acceptable for AVM performance. These thresholds should form the basis of triggers to escalate performance issues along the governance chain, should the required standards not be achieved.

3. **Define Valuation Segmentation**

   The purpose of segmentation is to obtain detailed validation results with respect to the lending activities deemed most important to the lender. These are typically based on perceived differences in house price dynamics such as geography, price range, property type etc. The lender should, where data volume and quality allow, design the validation tests in line with pre-identified segments of interest.

4. **Define Test Components**

   This is to identify the components of the AVM that need to be tested separately to evaluate their overall influence on model performance, such as

- If the valuation estimate for a property is based on indexing previous sales data for that property, then results such as this should be tested separately from model generated results
- If the validation sample has some instances of sales data being used as the target variable and others where appraisal data is used as the target, then these should be tested separately.

Connected Analytics strongly recommends that falling markets be tested separately from rising markets. AVMs may perform very differently under these conditions. In particular, as mentioned earlier, falling markets may result in valuation lags that lead to underestimates in collateral requirements.

If data volume allows the test components identified should be validated against the segments defined in priority 3.

5. **Design Test Data**

   Priorities 1-4 define the validation plan. Before the plan can be executed, a validation data set needs to be constructed. The composition and quality of the validation sample is crucial to the efficacy of the test statistics used to judge the AVM performance. Therefore, all the points below are strongly recommended by Connected Analytics.

- The sample of properties used to test the valuation estimates of one or more AVMs should have sufficient breadth and depth such that there are statistically significant volumes of properties to test the segments and components outlined in priorities (3) and (4) above

  Where this is not possible, consider the following for each segment:

- Use influence statistics to identify and account for outliers
- Use medians instead of means to calculate test statistics
- Consider combining segments where performance might be expected to be similar or priority is low
- Employ a valuation strategy that supports the AVM with the use of other valuation techniques
- Remove any properties from the analysis sample that are not representative of properties that will be valued using an AVM going forward, such as policy exclusions. Incorporating them into the validation statistics may give a misleading view on how the AVM will perform operationally
- Ideally, the data sample should be drawn from the lender's internal database, with the appropriate AVM(s) outputs appended at the time of valuation. These cases can then be directly compared to the sales price achieved on the day. Unfortunately, given the breadth and depth of sample required to test all relevant segments, few institutions have this level of data available and will need to send a sample of properties to an AVM provider for retrospective scoring. Where this is the case:
- Aim to construct the validation sample such that the

properties sent for matching are outside the AVM providers prior experience, i.e. so that they are unlikely to have additional information on record to "assist" the AVM make its valuation

- Insist on understanding the AVM provider's processes for capturing, storing and using data within a retro-spective process
- Remove any data from the testing that would not have been available at the time of valuation to ensure a true reflection of how the model is performing during mortgage origination
- Ensure that the retrospective valuations are generated as at the time of sale and also that the valuation attached is generated by an AVM that is still in use and unmodified since the valuation. There is no point validating a model that is no longer in use. To this point, ensure that the test sample consists of a sufficient volume of recent transactions
- Seek to understand the reason why the AVM provider excluded any records as part of their matching exercise
- Connected Analytics advocates using the property sales price as the target variable to benchmark performance testing of an AVM. As discussed, it is not perfect, but at the end of the day it is what we try to predict at the point of lending. Appraisals or other price estimates should only be used as the target variable where sales price is not available and should be tested separately
- If more than one AVM is to be validated, they should be tested on the same sample for results to be directly comparable.

6. Calculate Test Statistics

With priorities 1-5 in place we can now test the performance of an AVM on our validation data segments. As discussed earlier, we are primarily concerned with validating the model outputs and the statistics that we support as most important to judge AVM performance are accuracy, forecast standard deviation and hit-rate. With these defined, most other metrics of interest can be derived from them.

Note that the AVM provider will supply accuracy rates, FSDs etc. We advocate the lender calculate their own metrics in line with the definitions below and where performance appears significantly different from that reported by the AVM provider, seek to understand why.

As mentioned, the lender should define the acceptable thresholds of the test metrics based on the estimated financial impacts of under-performance. This will help interpret the validation results when multiple AVMs are being tested. The lender needs to weigh up the performance of an AVM on priority segments, not just at an overall level. In this way, the preferred AVM may not actually have the best levels of accuracy overall, but instead performs strongly on those segments that are most important to a lenders activities.

Accuracy

We believe that the measures of accuracy should be defined in the following two ways:

- Absolute Error = sale price – AVM valuation
- Percentage Error = (sale price – AVM valuation) / (sale price)

Which metric is employed in a given test depends on the purpose of the test. For originations we want the AVM valuation to be as close as possible to the realised selling price, so Absolute Error would be the metric of interest. When comparing the relative performance of an AVM across segments or across time Percentage Error makes more sense as it is scale independent and unaffected by such factors as inflation.

At a minimum, Connected Analytics recommends that the data segments and higher levels of aggregation are accuracy tested for the following:

i. Current Average Error: Is the AVM error within acceptable thresholds based on a recent sample of data?

ii. Longitudinal Average Error: Is the AVM error deterio-rating over time on some segments?

iii. Current Bias: Does the AVM consistently over/under predict on some segments? The financial impacts of bias are not likely to be symmetric and so their validation thresholds should be set accordingly. That is, over-prediction carries more risk of credit loss than under-prediction.

iv. Longitudinal Bias: Does the AVM over/under predict due to a trending market?

v. Sensitivity Analysis: Do minor differences in the definition of segments lead to large differences in accuracy?

vi. Once the accuracy statistics are calculated the segments should be ranked from high to low accuracy and their defining features analysed. In this way, the systematic strengths and weaknesses of an AVM can be understood with respect to the segment definitions and factored into future lending decisions.

Forecast Standard Deviation (FSD)

This is the standard deviation of the Percentage Error. We recommend that Percentage Error be defined as described in the last section and the standard deviation be calculated according to standard statistical sampling theory.

The advantage of this approach is that it links AVM volatility directly to sales price and is amenable to standard statistical analysis. Confidence intervals based on measures such as the number of properties in a suburb are not directly related to price and so cannot offer accurate estimates of price volatility.

We recommend that tests i-vi outlined above for accuracy be repeated for FSDs. The purpose here, of course, is to understand what drives volatility.

The results have to be interpreted hand-in-hand with the accuracy statistics. The average error rate might be acceptable for a given segment, but if it is accompanied by a high FSD then we know (and can calculate how much) that a given percentage of valuations will be outside of our accuracy tolerance.

Hit Rate

This is the proportion of successful valuations returned on a sample of properties.

For hit rate it is important to understand why hits are or are not achieved. It is recommended that lenders ascertain from AVM providers exactly what processes they apply to locate properties. In particular, if an individual property is not matched, does it get excluded as not being a hit or does it get an averaged valuation based on a higher aggregate level, e.g. apartment block, street, suburb.

Also, the lender should enquire as to what criteria the AVM provider applies to determine whether a hit stays in the returned file or gets excluded. For example, high FSDs or missing property information, such as the number of bedrooms.

7. Engagement & Governance

Once the strengths and weaknesses of a given AVM have been determined they should be distributed to appropriate stakeholders and reviewed critically by senior management. As mentioned at the start of this paper, the use of AVMs appears to be growing “under the radar”. They are simply too influential to allow this to continue.

8. Monitoring & Regular Testing

Model validation is typically used to provide an on-the-spot assessment of current AVM performance on the lenders portfolio. However, these are not set-and-forget exercises as model performance can deteriorate over time. One AVM provider we spoke to estimated that approximately 30% of models fall by the wayside over a seven year period. In addition, the conditions under which the AVM is being applied may change (economy, policy…). Therefore, it is incumbent on the lender to keep an eye on the relationship between sale price and AVM valuation and to make sure that it does not deviate too far from the latest validation results.

Connected Analytics recommends that regular monitoring, at least monthly, should be established to track the ongoing performance of an AVM on the lender’s portfolio. This is to ensure that the use of an AVM in recent transactions continues to perform in line with the results obtained from the validation exercise.

However, a limitation of monitoring is that it only observes the recent transactions undertaken by the lender so will be blind to wider pricing activity outside of this experience. Therefore it is recommended that the monitoring be supported by a program of comprehensive validation exercises of the type described in steps 1-7, to ensure that lenders understand the impacts of wider market trends.

Ideally re-validations would occur at least every six months. However, if the monitoring demonstrated that expected levels of performance were not being met before six months was up then the next scheduled validation exercise should be bought forward.

In line with in-house monitoring, the lender should also seek information from the AVM provider about planned changes to model structure or data inputs. Any significant changes should act as a signal to prepare a thorough validation exercise.

If a lender decides to use an AVM for purposes that have not been incorporated into previous validation exercises, then it is unreasonable to expect performance levels observed during the validations to automatically carry over to the new activities. In these circumstances Connected Analytics would advocate enhanced human oversight.

Conclusion

It is hoped that this paper managed to emphasise three key messages:

1. AVMs are powerful tools. Their influence is significant and is on the increase

2. The effective management of AVMs is complex and requires:

a. Clarity of policy concerning appropriate use of models and their interaction with other valuation approaches

b. Appropriate governance and oversight

c. Robust validation and monitoring processes to accurately and continually assess performance

3. Currently, effective validation is hampered by the lack of industry standardisation across virtually all aspects of the AVM lifecycle.

Feedback is welcomed on our attempt to address point (2c) with the long term goal in mind of fixing point (3).

Appendix 1 – Australian Valuation Regulations (relevant sections of APG223)

Connected Analytics comment: This regulation is also relevant to New Zealand market as the 4 largest NZ banks are reviewed by APRA given their Australian ownership.

Security valuation

The valuation of underlying collateral can be undertaken in a range of ways. A full on-site valuation is good practice, although APRA acknowledges the benefit of tailoring an ADI's valuation policy to its circumstances. APRA expects that an ADI will document when and what type of valuation method is appropriate.

Valuation methods

61. Techniques such as desk-top assessments, kerb-side valuations, automated valuation models (AVMs) and reviews of contracts of sale are all acceptable valuation assessments, in the appropriate context. As the risk associated with collateral increases, or the coverage of a given loan by collateral decreases, the need for specialist valuation also increases.

A prudent ADI contemplating the use of alternative valuation methods such as AVMs would subject proposals to thorough analysis and develop a risk management capability that includes:

(a) a hierarchy of acceptable methods of determining value that is appropriate to the level of risk;

(b) analysis of the strengths and weaknesses of the relevant approaches/models being considered, including an understanding of the methodologies used, sources of data employed and how the service provider may be able to assist in re-engineering the ADI's processes, and details of back-testing or auditing arrangements undertaken by the service provider;

(c) clarity of the output to be provided and how it would be integrated with the ADI's processes;

(d) ongoing monitoring of tools used, processes that capture evidence of action taken when values are deemed to be unreliable and periodical back-testing undertaken by the ADI to independently validate reliability of outcomes; and

(e) appropriate training for staff on operational requirements.

Where an ADI relies on a panel of approved valuation professionals, sound credit risk management practice would provide for the panel to be periodically reviewed by senior risk management staff of the ADI. Valuer selection would be conducted by the ADI's risk management area, rather than sales staff, and the involvement of ADI sales or product staff in panel management would be minimal.

An ADI is required to have regard to the principles governing security valuation practices, including valuation of security in the form of property, as detailed in Attachment B of Prudential Standard APS 220 Credit Quality (APS 220). Sound risk management practices would include valuation reports prepared with professional skill and diligence, valuers selected on the basis of appropriate professional qualifications and maintaining comprehensive valuation documentation for the term of the loan. The scope and extent of a valuation report would be commensurate with the property value and inherent risks.

65. Attempts by an ADI or third-party lending staff to pressure valuers to over-value properties are an indicator of poor practice and improper behaviour. A robust risk management framework would capture such instances, prompting appropriate investigation and swift and appropriate action where necessary.

66. The valuation management process itself may be effectively outsourced to third parties. However, it is good practice for any override of valuation requirements to be limited to senior risk management staff of the ADI.

67. Good practice would be to ensure that claims against collateral are legally enforceable and could be realised in a reasonable period of time if necessary. This would include the ADI confirming that the:

(a) borrower has, or will have when the loan is extended, a clear title to the property;

(b) characteristics of the property are as they have been represented; and

(c) property serving as collateral is appropriately insured at the time of origination and is maintained under the contractual terms of the mortgage.

68. An ADI would typically seek to ensure that property serving as collateral could be readily linked to related residential mortgage lending facilities. Underinvestment in such collateral tracking capability could leave an ADI open to greater operational risks and losses.

69. To access mortgage risk-weight treatments of less than 100 per cent under Prudential Standard APS 112 Capital Adequacy: Standardised Approach to Credit Risk (APS 112), an ADI is required to ensure loans are secured by residential property (either as a single property or in a group where loans are secured by more than one property).

The lower risk-weight does not apply in circumstances involving property used for mixed purposes, i.e. where the property also accommodates a component of non-residential use. In addition, if a borrower's income/business is to invest or speculate in or develop multiple residential properties, the concessional risk-weight for residential loans is not applicable.

An ADI would, as a matter of good practice, develop a policy on when a borrower (or connected group of borrowers) providing collateral in the form of mortgages over multiple residential properties is more akin to commercial lending than residential lending. This is particularly the case where one borrower holds multiple housing stock in the same title/ deposited plan. In addition, where a developer or commercial borrower chooses to hold a number of residential properties longer term, as opposed to selling them, the risks are more likely to be of a commercial rather than residential nature. In such cases, APRA would not expect the provisions in APS 112 applying to loans secured by residential mortgages to be applied. Instead, the exposures would be treated as commercial real estate and risk-weighted at 100 per cent (refer to paragraph 22 of Attachment A to APS 112). For ADIs accredited to use the internal ratings-based (IRB) approach to credit risk, such exposures would be treated as 'retail IRB, 'corporate IRB' or 'income producing real estate (IPRE)' under Prudential Standard APS 113 Capital Adequacy: Internal Ratings-based Approach to Credit Risk (APS 113), as appropriate. APRA's guidance on identifying IPRE exposures was outlined in a letter to ADIs in October 2009.6

Risk-weights for capital adequacy purposes determined by reference to APS 112 are based on the LVR calculated at the point of origination. Reliance on a valuation other than the valuation at origination would generally require a subsequent formal revaluation by an independent accredited valuer. In particular, it would be imprudent for the ADI to use indexed-based valuation methods that calculate capital adequacy on a dynamic LVR basis. This is because an index does not necessarily mean that all properties in a particular area have exhibited the same increase or decrease as reflected in the index. That said, APRA supports efforts by an ADI to better understand its portfolio on a dynamic LVR basis for internal management purposes, such as overall portfolio risk assessment.

Loan-to-valuation ratios

72. Although mortgage lending risk cannot be fully mitigated through conservative LVRs, prudent LVR limits help to

minimise the risk that the property serving as collateral will be insufficient to cover any repayment shortfall. Consequently, prudent LVR limits serve as an important element of portfolio risk management. APRA emphasises, however, that loan origination policies would not be expected to be solely reliant on LVR as a risk- mitigating mechanism.

73. A prudent ADI would monitor exposures by LVR bands over time. Significant increases in high LVR lending would typically be a trigger for the Board and senior management to review risk targets and internal controls over high LVR lending. APRA

has no formal definition for high LVR lending, but experience shows that LVRs above 90 per cent (including capitalised LMI premium or other fees) clearly expose an ADI to a higher risk of loss.

74. Lending at low LVRs does not remove the need for an ADI to adhere to sound credit practice or consumer lending obligations. A prudent lender would seek to ensure that a residential mortgage loan has reasonable expectations of being repaid without recourse to the underlying collateral. An overall sound assessment would be based on the borrower's repayment capacity at the time of loan origination rather than an overriding presumption that the value of collateral will appreciate.

- ADIs typically require a borrower to provide an initial deposit primarily drawn from the borrower's own funds. Imposing a minimum 'genuine savings' requirement as part of this initial deposit is considered an important means of reducing default risk. A prudent ADI would have limited appetite for taking into account non-genuine savings, such as gifts from a family member. In such cases, it would be prudent for an ADI to take all reasonable steps to determine whether non- genuine savings are to be repaid by the borrower and, if so, to incorporate these repayments in the serviceability assessment
- A prudent ADI would exhibit greater caution when relying on collateral values in periods of rapid growth in property prices. It may be appropriate for an ADI to strengthen its LVR constraints or re- assess its risk appetite in markets exhibiting rapid price appreciation
- Sound credit practice would include recalculating LVRs at the time of any top-up loan and other formal loan increases during the life of the loan. Any subsequent refinancing, including any second mortgage, charge or lien, would also typically result in the calculation of a new LVR at the point of refinancing. Such calculations would be based on appropriate and contemporary property valuations

  Further, particular caution would need to be exercised in relation to any draw-down on the equity in the property, especially if the draw- down would increase the current LVR above the level originally agreed. Finally, any significant increase in loan exposure would normally be subject to a full assessment of the borrower's repayment capacity
- In the case of valuation of off-the-plan sales, developer valuations might not represent a sustainable resale price. Consequently, in such circumstances, a prudent ADI would make appropriate reductions in the off-the- plan valuations in determining LVRs or seek independent professional valuations. Similarly, developer discounts would not be treated as part of the borrower's deposit for LVR calculation purposes: such discounts reduce the sale price, but do not increase the borrower's deposit.

79. Where an ADI's risk appetite allows for higher LVR lending, good practice would provide that the additional risk in this lending would be mitigated by measures such as stronger serviceability- adjusted loan pricing and additionally, in the case of IRB banks, higher expected loss provisions and capital. APRA does not consider the sole use of coverage of loans by LMI as a sufficient control to mitigate high LVR risk.

Appendix 2 – References

- "Standard on Automated Valuation Models (AVMs)"- International Association of Assessing Officers (Sept. 2003)
- "Automated Valuation Model Accuracy: Some Empirical Testing"- Peter Rossini & Paul Kershaw (Jan 2008)-Paper presented at 14th Pacific Rim Real Estate Society Conference Istana Hotel, Kuala Lumpur
- "Australian House Prices: A Comparison of Hedonic and Repeat-sales Measures"- James Hansen (2006)- RBA research paper
- "Anatomy of Risk Management Practices in the Mortgage Industry: Lessons for the Future"- Clifford V. Rossi - University of Maryland-May 2010
- "The Role of Conditional Logic in AVM Cascade Creation"- CoreLogic White paper - Susan Allen, Nov 2009
- "Innovation in AVM Testing: When what you see is really what you get"- CoreLogic White paper Sept 2009
- Forecast Standard Deviation & AVM Confidence Scores - CoreLogic 2010
- "The AVM Cascade Industry"- Andrew Caplin, Roy Lowrance - New York University - Oct 2010
- "Valuation Opinion-sometimes less is more" –Susan Allen, Joni Pierce CoreLogic White paper , Oct 2010
- CoreLogic Risk Summit Brochure , July 2014
- "Automated Valuation Models: an international perspective" - Mary Lou Downie and Gill Robson, Northumbria University, Oct 2007.

References: (Sourced from RP data):

- "Best practice in Automated Valuation Model (AVM) validation and monitoring"- Internal discussion paper –RP data- Nigel Butler , Jan 2014
- "RP Data AVM Reference"- RP Data - Joe Cuffaro - April 2012
- "Automated Valuation Models: Are You Buying What You Think You Are Buying?" -Riskmark White paper
- "Automated-valuation-model-testing" - CoreLogic document -July 2011
- "AVM Cascades Methodology and Validation-Model and Data Description"- First American CoreLogic - July 2008
- "Best Practices in AVM testing practices"- First American Real Estate Solutions- Robert Walker - Oct 2005
- "Automated Valuation Model - Performance Metrics"- First American Real Estate Solutions - Christopher L. Cagan
- "Bias and Precision in the Measurement of Residential Real Estate Performance" - Rismark Working Paper- Danika Wright, 2006
- "Monthly AVM report" - RP Data - February 2015
- "Best Practices In Automated Valuation Model Validation"- Collateral Assessment & Technologies Committee (CATC)– Jan 2009
- "Comparing the Accuracy of Short Forms, EVRs and AVMs in Relation to Sales"- RP Data - Joe Cuffaro, Oct 2012
- "Comparing the quality of property valuation methodologies"- RP Data white paper - Joe Cuffaro (RP Data), Dr Alun Pope (Rismark International) , Nigel Butler - March 2012.