Report for TEGoVA

The Accuracy of Automated Valuation Models (AVMs)

Professor George Andrew Matysiak

May, 2017

Summary

- There is ample general material and opinion on AVMs, covering such aspects as their increasing use, the types of models employed and their limitations.
- However, there is little hard impartial evidence on the accuracy of AVMs in the public domain. European vendors are reluctant to release details.
- AVM vendors emphasise that for AVMs to be effective, considerable volumes of up-to-date market data are necessary. This, in combination with a strict filtering of 'outliers', aims to ensure reliable estimates of market value.
- Emphasising the effective role and use of AVMs, *Hometrack's* view is: "Understanding the inputs and processes is key but it's the accuracy, quality and consistency of the outputs against a clear benchmark valuation that is most relevant to end users. Knowing when not to use an AVM is as important as having the confidence to use one."
- The vendors claim that there is continuous ongoing enhancement and development in AVM technology, including the use of sophisticated algorithms to calculate the value of properties. Extensive and detailed property databases are enabling this progress.
- The vendors themselves recognise that their AVM's will not provide accurate valuations in all situations, and caution the conditions under which their use is appropriate
- There is very little independent evaluation of the accuracy of AVMs, understandably, as the underlying data is not made available for analysis.
- The vendors argue that the 'accuracy' figures of the AVM need to be put into a wider perspective. Other than submitting information to rating agencies, AVM operators are unwilling to have their data/methodologies exposed to independent scrutiny.
- AVM accuracy results in the USA, for example from *HouseCanary* and *Zillow*, are made available on their Websites. These figures provide a point of reference.
- Based on an analysis of 666 US Counties, if +/- 10% is seen as an acceptable margin for error, *on average* some 70% of AVM valuations would fall within the +/- 10% bracket. This would likely be an upper limit for European AVMs.
- Depending on location, the distribution of valuations falling within the +/- 10 % bracket ranges from 20% to 92%.
- Despite high average accuracy levels, statistically-based valuations may be widely off the mark and need to be augmented by professional judgement.
- The margin for error will likely vary over different market conditions, types of property and countries.
- Debate regarding the role and accuracy of AVM valuations is an ongoing topic of discussion. For there to be a meaningful debate, AVM vendors need to make available access to their models for independent testing and verification of the models' output and accuracy.

Introduction

The study of real estate appraisal methods and the application of residential valuation and pricing approaches, including hedonic models, is a significant research area in academia. Recent developments have witnessed a substantial growth in residential Automated Valuation Model (AVM) providers, who offer their services routinely on a fee-based basis. This is also attracting academic attention. Although different underlying models are employed, fundamental to the approach are statistical and computing technicalities.

Despite traditional approaches being extensively employed in the valuation profession, over the last several years there has been a move towards automated valuation approaches as support in arriving at a valuation. AVMs are now widely used by real estate lenders, real estate professionals, Government, by the general public and are thus seen as complimentary to traditional valuations. The use of AVMs in assisting the processing of loan valuations is now established practice (Downie & Robson, 2008; CML, 2007). Indeed, an RICS (2013) information paper identifies the following areas where AVMs are used:

- Revaluation for credit decision in banks
- In-arrears assessment in banks
- Identification of fraudulent activity in banks
- Full valuation audits in banks
- Determining capital adequacy ratios in banks
- Mark-to-Market bank's portfolio of properties in banks
- Mass Appraisal for local taxes by government
- Estimating relocation compensation by government
- Cost/Benefit analysis for potential public expenditure
- Capital Tax planning for the individual

These computer-assisted quantitative methods have their advantages in that they are systematic and fast, thereby reducing reliance on labour input in providing an end-to-end valuation (Tretton,2007). By removing the human element, it is claimed by some advocates, that it also reduces inaccuracies due to reliance on human judgement. However, the overall attitude and degree of acceptance of such an automated approach to valuation varies.

When large numbers of properties are involved, manual valuation can be extremely time consuming. Consequently, automated valuation models have been employed to address this. Computer assisted mass valuation (CAMA) is now widely employed in different countries, particularly in assisting property tax assessment. The growth in AVMs has naturally evolved from the application of computer-based valuations in mass valuations.

The focus of this report is to examine the evidence regarding the accuracy of *commercially* available AVMs. Despite AVMs having been developed and refined over the years, they are still regarded as having shortcomings (Lipscomb, 2017) and their accuracy record in assessing market prices or values is called into question.

It should be acknowledged at the outset, that despite efforts to obtain information directly from commercial AVM providers, very little material was directly made available, particularly from European/UK vendors. Consequently, the results and conclusions in the report are largely based on secondary sources of information.

Investigating AVM accuracy

The following lines of research have been undertaken:

- Identifying existing academic work in the area
- Identifying publicly available information
- Identifying/contacting selected providers of AVMs in the UK and the US
- Identifying relevant reports/papers prepared by professional bodies
- Identifying trade press material and commentary on AVMs
- Identifying relevant Websites
- Sifting through 'popular' commentary (of which there is much) dotted around the web and various trade sources.

The findings reported in this document summarize the salient features which were identified as providing an insight into the question of AVM valuation accuracy. However, it should be stated at the outset, the information obtained on US AVMs far outweighed that for European AVMs, the comments on European AVMs being in respect of two UK AVM vendors.

Modelling residential property

This section provides a brief overview of the some of the common types of modelling techniques used to value properties, and is not a comprehensive synopsis, as this would deflect from the main focus which is to investigate the accuracy of AVMs.

There is a considerable and rich body of literature addressing the modelling of residential property prices and values. Computer aided valuation approaches encompass a variety of methods including: Computer Assisted Mass Appraisal (CAMA) (McCluskey et al,1997), multiple regression analysis, artificial neural networks (ANN) (Worzala et al,1995), fuzzy logic and, more recently, a rapidly evolving variety of machine learning and data mining oriented techniques (Zurada et al, 2011; Antipov & Pokryshevskaya, 2012).

Multiple regression analysis (MRA) has been extensively used and is the traditional method of choice. Typically, a multitude of property characteristics are taken into account in the modelling. However, the standard MRA approach has its limitations which are well recognised including: the inability to adequately deal with interactions among variables, non-linearity, heterogeneity, independence of errors and multicollinearity. Kilpatrick (2011) lists some of the issues associated with multiple regression analysis.

The standard MRA approach has been extended to accommodate, for example, spatial dependency type models. Alternative methods have been employed in attempting to address the issue of spatial (geographic proximity) reflected in the dependence of the error terms, namely spatial autocorrelation (Bourassa et al 2010). The application of spatial statistics has made significant contributions in modelling residential property prices (Bidanset & Lombard, 2014; Osland, L 2010).

McCluskey et al (2013) note that the relationship between property value and its explanatory attributes is highly complex and generally non-linear, which calls for more insightful approaches than the traditional MRA analysis. Other modelling approaches, for example artificial neural networks (ANNs), do not rely on any of the assumptions made by MRA and have been extensively explored (McCluskey et al 2012, 2013). The application of neural networks is one example attempting to capture complex interactions between the various characteristics considered to account for the value or price of a property.

There are many extensive reviews which summarise the variety of modelling methods employed in explaining residential property values. Refer to McCluskey et al (1997, 2013) for an overview of the various approaches.

Automated Valuation Models

AVMs are now widely employed in both the public and private sectors (Downie & Robson, 2008). TEGoVA provide the following, Definition 2.1, in their European Valuation Standards EVIP 6:

- ‘Automated Valuation Models (AVMs) can be defined as statistic-based computer programmes, which use property information (e.g. comparable sales and property characteristics etc.) to generate property-related values or suggested values.’

The International Association of Assessing Officers, IAAO (2003), describes an AVM as:

- ‘a mathematically based computer software programme that produces an estimate of market value based on analysis of location, market conditions, and real estate characteristics from information collected. The distinguishing feature of an AVM is that it produces a market valuation through mathematical modelling. The credibility of an AVM is dependent on the data used and the skills of the modeller producing the AVM.’

The RICS AVM Standards Working Group:

- ‘Automated Valuation Models use one or more mathematical techniques to provide an estimate of value of a specified property at a specified date, accompanied by a measure of confidence in the accuracy of the result, without human intervention post-initiation.’ (RICS 2013).

A key component in the RICS definition is the qualification *'…accompanied by a measure of confidence in the accuracy of the result…*'. The evidence regarding AVM accuracy is provided in the *Indicative accuracy figures* section of the report.

Finally, the European AVM Alliance, Glossary of Terms and Definitions:

- 'A system that provides an estimate of value of a specified property at a specified date, using mathematical modelling techniques in an automated manner.'

All four definitions of an AVM exclude any valuer involvement in arriving at an estimate of value.

A distinction should be made between the application of mass appraisal methods in valuation and AVMs. Grover (2016) provides a discussion on the defining characteristics of an AVM, and comments that, unlike mass appraisal methods, which are used to value entire populations of properties at a single point in time, an AVM valuation undertaken is in respect of a single property and is not tied to a specific date. Furthermore, the resulting AVM valuation is qualified by a measure of confidence in the accuracy of the result. Mass appraisal methods do not qualify the estimated values.

AVMs have their origins in North America, the first commercial application being in 1981, and began to be developed in the UK in the 1990s. Statistical and data mining methods are employed in estimating property values, which are calibrated on large databases of properties. Conventional AVMs originated by making use of statistical methods such as multiple regression analysis (MRA), with underlying hedonic-type models being the most extensively employed. However, many sophisticated AVMs have moved on and employ models based on machine learning and data mining techniques. Their quality will vary considerably, depending on such features as the available data, sample sizes together with the design and development of the model. This is persuasively demonstrated in the results obtained in various studies investigating residential property values and prices.

A flavour of the stages involved in the establishment of a commercial AVM capability in Germany is provided in Schultz et al (2014), discussing model development, the validation process and emphasising the importance of the removal of outliers.

No matter which quantitative or modelling approach is taken, an over-riding requirement in order to arrive at a robust AVM facility will be the need to establish a large and continuously updated database of property transactions. This is something which is strongly emphasized by all AVM vendors. The database will record a variety of individual property characteristics which are deemed to determine price or value. Consequently, given the growth in the size of available databases over the last few years, this has facilitated the development of alternative approaches, employing sophisticated machine learning and data mining applications by AVM vendors, in modelling, classifying and valuing properties.

As indicated, AVMs are computer-based applications, using a variety of statistical and algorithmic approaches in analysing the relationship between the price/value of a residential property and the property's underlying characteristics. The objective is to arrive at an estimate of the property's market value. The methods employed differ for different AVM vendors. Indeed, individual vendors will have several available models, making use of the most appropriate model in given circumstances. *However, all models, to a greater or lesser extent, will contain a degree of uncertainty surrounding the resulting property valuation.* A measure of this uncertainty is often provided by the AVM vendor.

An implicit assumption underlying an AVM model's prediction is that the property is in a marketable condition, with vacant possession and improved internally to normal standards. This is one limitation of AVMs, in that only a physical inspection can verify these assumptions (Robson & Downie, 2007).

Whilst there are a large number of AVM vendors, the inner workings of the models and details of their specification are not released, nor are 'accuracy' figures disclosed. Vendors do test their models regularly for accuracy, comparing individual property AVM valuations against achieved market prices, some claiming they have these figures independently assessed. However, these figures are not normally disclosed, and this non-disclosure puts a constraint on the analyses which can be objectively undertaken as regards an assessment of the reliability and accuracy of the models.

A report by Robson & Downie (2008) provides the results of a survey they undertook on AVMs. Needless-to-say, the AVM market will have developed since the survey was conducted, but it does provide an interesting broadly-based discussion of AVMs and attitudes towards them.

Robson & Downie discuss their findings on AVMs and the integration of AVMs within the valuation process, from an international perspective. There were 473 valuer responses, representing both lending and valuation organisations, described as senior professional members with 'much experience of mortgage valuations'. The results of the survey include the following findings:

- 71% of the valuers agreed that AVMs were inadequate for loan valuations as a result of no physical inspection.
- 87% of the valuers agreed that physical valuations were more accurate than AVMs, as a result of local knowledge.
- 90% of valuers agreed that the ability to evaluate comparables was a major advantage over AVMs.

It was also reported that 72% of the respondents expressed a desire 'to learn more' about AVMs. Consequently, AVM vendors should become more open and transparent, explaining the workings of their models, thereby enabling the models to be fully understood and scrutinized.

An overview of the global use of AVMs is provided in the Robson & Downie report. The use of AVMs was 'well established' in only three or four countries. It would be interesting to update the list of countries adopting the use of AVMs over the intervening period since 2008 as, no doubt, the numbers *may* have increased. However, given that the number of AVM providers across Europe is limited, it may be that the capacity for developing AVMs in some European countries is constrained due to factors such as: lack of market transparency, validating the reliability of transactions information, low transactions volumes in certain segments and a lack of comparables.

A recent discussion and general overview of the evolution of AVMs is provided in a compilation of articles by d'Amato and Kauko (2017). There are discussions about the application of AVMs and the problems encountered in using AVMs. It is noteworthy that the co-authors observe that, 'some institutions consider AVM assisted valuations more reliable than valuation in person.' This assertion needs to be more thoroughly clarified and supported by evidence.

The limitations of AVMs are well known and understood. For example: the inability to confirm or deny whether a property exists; the limited ability to address a property's condition; the limited ability to account for external influences; limited data coverage in some areas; limited ability to reflect any unique characteristics of a property, and so on. However, the crucial test is, can AVMs forecast 'accurately'?

Qualifying valuation estimates and measuring AVM accuracy

The uncertainty surrounding the prediction resulting from an AVM valuation is typically qualified by the AVM vendor. This can be achieved in a number of ways. For example, one measure is the so-called *Forecast Standard Deviation (FSD).* How is this calculated? First, an error is calculated, being equal *to the difference* between the estimated AVM value and the sales price of the sold property (or possibly, a valuer provided estimate of market price instead of a sales price). This error is then expressed as a percentage of the AVM value. The FSD is then calculated from all of the percentage errors across the AVM valued properties. Essentially, it summarises the distribution (the standard deviation) of the individual percentage errors, around the average value of all of the percentage errors. The FSD percentage is an *estimate* of the amount of variation that can occur between the actual sales price and the forecast (the most probable market value) made by the AVM; the lower the FSD, the smaller the error in predicting the resulting market value/sales price i.e. the closer will the AVM estimate be to the actual sales price.

If under-valuations and over-valuations are assumed to be equally likely, on average, approximately 2/3rds of AVM valuation errors will fall within +/- one FSD of the AVM estimate. For example, if the FSD is 10%, approximately 2/3rds of actual sale prices will fall within +/- 10% of the estimated AVM values. So, if a property was valued at €1 million, this could be interpreted that there is a 2/3rds chance that the actual sales price could be anywhere within the range €900,000 to €1,100,000.

The question is, what is the typical magnitude of the FSD figure in practice? In other words, how wide are these intervals? This is an empirical question which is addressed by looking at past AVM valuations and achieved sales prices in the market. As noted in the *Introduction*, many vendors do not report the overall accuracy figures of their AVM predictions. However, some information has been obtained from publicly available US based data which will be presented.

In commercial situations involving loans from banks and financial institutions for mortgage purposes, financial rating agencies have defined a measure of accuracy which has found common currency. The *benchmark* is the valuer's assessment, and accuracy is measured against the difference between the value estimated by the AVM compared to the valuer's assessment.

There are a number of ways in which accuracy can then be expressed. The most commonly quoted measure in the context of AVMs is the Forecast Standard Deviation, as discussed above, which is the standard deviation of the percentage forecast errors where:

$$\text{Percentage forecast error based on AVM} = \frac{\text{Surveyor Value} - \text{AVM Value}}{\text{AVM Value}}$$

It should be noted that the RICS (2013) cite an incorrect formula for *Percentage error based on automated value*, which they mistakenly attribute to *Fitch Ratings* (2007).

Using a measure relative to a valuer's opinion does not necessarily reflect the accuracy of the AVM relative to the *price achieved* in the market. One implication of this is that if an AVM accurately predicted the sale price, it could erroneously be judged as having produced an error, if the valuer had produced an under or overvaluation. Rossini & Kershaw (2008) comment on this as follows:

'Fitch (2007) acknowledges the criticism that surveyor values may be biased compared with AVM estimates however, since they represent the currently adopted standard for most lending institutions, it makes a better comparison from their point of view. In this regard, the testing by lending institutions does not measure the accuracy of the AVM compared to the market, simply the accuracy compared to the current alternative.' The current alternative being the valuer's assessment.

The reference to biased valuations arises because valuers will have access to additional information, such as a contract or the asking price.

Indicative accuracy figures

UK

Preamble

Having exchanged correspondence with *Hometrack, Rightmove* and the *European AVM Alliance (EAA*), the upshot is that they were all very guarded in terms of the details which they were willing to release. They took the view that looking at the accuracy figures in isolation can result in a misleading picture, and, as I was informed by Hometrack: '…there is a whole/use/application story that needs to sit alongside accuracy...', advising me to contact the EAA. Having contacted the EAA, their response was: ‘From our experience our experts in the EAA know communicating these complex issues via email is not a suitable way.” After a further exchange of emails, they eventually agreed to meet me in London – perhaps in May or June. I also contacted another EAA member, *Calcasa* who are based in the Netherlands, but they have not responded.

- My overall impression is that the UK AVM vendors are unwilling to engage in a discussion about their products or the valuation accuracy of their models.

The two leading UK AVM providers I have been in contact with, *Hometrack* and *Rightmove*, provided me with some very limited ‘taster’ information, being general overview material on their AVMs. Whilst broadly interesting, the information was far too inadequate to undertake any analyses, and so, reach any conclusions about the underlying models or their accuracy.

The information provided by *Hometrack* and *Rightmove* included the following:

- Both say they undertake monthly testing of their models for both accuracy and ‘consistency’. However, other than broadly based observations, they are unwilling to release the statistical results of these tests, let alone samples underlying the raw data, or background details of the AVM algorithms used in making the predictions.
- Rightmove say they are the largest AVM provider in the UK, claiming to have a database containing some 86 million property records. They say that they have stringent criteria, claiming to employ a thorough filtering process in selecting the properties used in their AVM models and in the forecasts, thereby arriving at robust valuations. Each month Rightmove recalibrate the models to ‘…ensure accuracy and consistency in the light of market liquidity and supply and demand which will ultimately impact on price movement in different areas.’ The final forecast is arrived at by combining the results of two different approaches. Each month 4,000 property valuation records across the UK are used as a ‘hold out’ sample, which Rightmove say they then test against the results of their AVM forecasts.
- As already noted, other than in broad terms, both Hometrack and Rightmove are very reluctant to release results of any accuracy tests which they undertake.
- Despite a pleasant exchange of emails and some broad observations on AVMs, Hometrack suggested I contact the European AVM Alliance, of which they are a founder member. This I did, having been given a contact at the EAA by Hometrack.

AVM Valuations versus physical valuations

As indicated, the very restricted amount of information provided does not permit any detailed analysis to be undertaken on Hometrack or Rightmove. Some figures provided by Rightmove hint at the types of numbers produced by their AVMs.

- Based on a snapshot of results for 2016, provided for 4 months by Rightmove, the monthly average errors were all negative, with an overall average valuation error across all properties being -1.5%, representing an average *undervaluation*. This figure is based on a comparison of Rightmove's AVM valuations against actual *physical valuations* of the properties collected by Rightmove.
- Whilst this average seems low, it is not strictly comparable with the US figures (being based on market prices and reported below in Table 1). where the average figure across 666 locations was a 6% error (see Table 1); the Zillow figures reflect the accuracy of comparing their AVM output with actual sales prices, not physical valuations.
- Rightmove take the view that their negative undervaluation error is preferable to a positive overvaluation error.
- There will be a distribution of errors, i.e. a standard error, of individual property valuation errors around the individual monthly averages which were provided by Rightmove. The distribution of individual errors around the average (the standard deviation) was not provided. However, Rightmove commented that their AVM provides 'consistency' in the reported valuation errors. This gives the impression that the values of the average monthly errors together with the *distribution* of these errors, across both smaller regions and different property types, were of similar orders of magnitude. Without more substantiated details this seems unlikely.
- My 'off-the-record' information suggests that there are UK regional variations in valuation accuracy.

As already indicated, it was not possible to undertake an analysis of the output of the AVM models. However, an indirect insight can be obtained. Credit rating agencies review the accuracy of AVM valuations as part of their rating process of residential mortgage-backed securities (RMBS), where residential properties have been valued by an AVM. Although the credit rating agencies do not release details of accuracy figures resulting from their investigations, the upshot of their analyses is reflected in the adjustments which they make to AVM determined values.

When looking at securitisation risk for residential mortgage-backed securities, Fitch Ratings (2012) looked at the underlying property values when assessing of the intrinsic risks in the mortgage loan. They subsequently make adjustments to the reported AVM valuations which reflect a number of factors, which they list:

- historical relative reliability of AVM values against surveyor values
- quantity and quality of data sources
- the quality of the model calibration framework
- frequency and quality of maintenance procedures for each AVM vendor
- lender's procedures around application of AVM values.

This is a rigorous list of checks for AVM originated valuations. As a consequence, Fitch Ratings apply a 'general adjustment' to all AVM valued properties where securitisation and covered bonds are involved. Over the years, there have been several updates to their treatment of automated valuations in securitisations and covered bonds. In the UK, for example, Fitch Ratings Criteria Addendum: UK (2016, Appendix 2)) states:

'General Adjustment: A 2.5% haircut will be applied to take into account the time lag between registration of a property and its eventual incorporation into the AVM database. This time lag is typically 1.5 months, up to a maximum of three months.'

In addition, there are other adjustments at pre-defined levels for low valued properties and for relative reliability levels, which vary, depending on the AVM provider. For example, a 3.5% valuation haircut is imposed for low valued properties in the case of Hometrack and 0% for Rightmove. These values are subject to regular review, and the adjustments may change from time-to-time in the course of a criteria review.

It is interesting that these 'haircuts' are not applied to ordinary physical valuations, although 'other adjustments *may* apply'. Clearly, Fitch Ratings make a distinction between AVM valued properties and physically valued properties in their assessment of the underlying valuation risk.

In summary, it is not possible to draw any insightful conclusions about the AVM accuracy for UK AVM vendors. The best one can do is to look to the accuracy figures in the US market and take a view about how the UK figures may stand relative to these.

US

AVM Valuations versus transactions prices

Having noted the sparse availability of UK/European AVM valuation data, there is US based information which can be drawn on. One of the largest vendors analysing the accuracy of their AVMs, for which data has been obtained, is the leading AVM provider in the US, *Zillow*. The accuracy results shown in this report are largely based on Zillow's figures. Some data was also obtained for another leading provider of AVMs, *HouseCanary*, and their accuracy figures are also reported.

The US AVM market is highly developed. At a recent Mortgage Brokers Association (MBA) forum in 2016, it was reported that nine out of 10 survey respondents were interested in technology which automates the mortgage loan process. Furthermore, it was claimed that "more than 85% of valuations were found to be accurate to the P10 standard (percentage of valuations within 10% of sales price)".

With high levels of perceived AVM accuracy, Fannie Mae and Freddie Mac, two of the largest funding sources of residential loans in the US, both employ AVMs as a crucial part of their valuation risk management systems. In November 2016, Fannie Mae expanded its use of AVMs by waiving the need for a physical valuation on certain refinancing loans, so-called Property Inspection Waivers.

In what follows, is a selection of information on valuation accuracy for several US AVM providers.

A small sample study using 500 sold properties in 2014 undertaken *by McEnearney Associates*, a Washington-area based real estate agent, provided some interesting figures. They looked at AVM valuations produced by Zillow, a leading US AVM vendor. The conclusion was that the estimated valuations were within +/- 5% 'half of the time.' In response to the study, Zillow commented: 'Our median error rate, nationwide, is currently 6.9 percent, which means half of all estimated values are within 6.9 percent of the final sale price.' Furthermore, they say '…which is why our local and national accuracy, is published prominently on our Web site and is updated every quarter.' It is important to note that Zillow's accuracy statistics are based on all closed sales in a market, not just a small sample of homes, such as 500, in a multiple listing service, which the McEnearney Associates study was based on.

Zillow also produced an analysis based on a much larger sample of 296,473 home sales across the USA, over a three-month period between March 1st 2012 and June 1st 2012. They compared sales prices with the initial prices at which the properties were listed, that is, the original asking price provided by a valuer.

So, how did the *Zestimate* accuracy (this is what Zillow call their AVM property price estimate) compare with the accuracy of the initial list price on a property, which the valuer's provided? It is expected that the list price will be fairly accurate, since it is derived by a real estate professional who will be familiar with both the property and the local market (although sometimes the seller may influence the price to be higher than an agent might prefer). Table 1 summarises the Zillow analysis:

Table 1: Accuracy comparison for 296,463 property sales

	Initial list price	Initial Zestimate	Final list price	Final Zestimate
Within 5% of sale price	48%	31%	56%	42%
Within 10% of sale price	71%	55%	80%	71%
Within 20% of sale price	90%	81%	94%	92%
Median absolute error	5.6%	9.2%	4.5%	6.3%

Source: Zillow

The accuracies for both the list price and the Zestimate in Table 1 show that the initial list price is within 5% of the final sale price for 48% of the properties, whereas the Zestimate achieved this accuracy for 31% of the properties. The initial list price is within 20% of the final sale price for 90% of the properties, whereas the Zestimate achieved this accuracy for 81% of the properties. The overall median absolute percentage error for the initial list price is 5.6% (50% of property valuations were within 5.6% of the sales price), whereas the median value for Zillow's estimate was 9.2%.

Table 1 also shows the accuracy for the *final list price* (the list price immediately prior to the sale, possibly after several price cuts) and the final Zestimate (Zillow's estimate of the sales price just prior to the sale). Both the list price and the Zestimate are much more accurate when measured closer to the time of sale, which is to be expected; the median absolute percent errors are 4.5% and 6.3% respectively, for the final list price and the final Zestimate.

The overall conclusion is that the valuer estimates were more accurate than the Zillow estimates, over the period March 1st 2012 and June 1st 2012.

Corelogic, another large US vendor of AVMs, reported accuracy figures for the period 2007-2009. The number of properties analysed was not specified. Their results indicate that in excess of 80% of properties achieve a sales price within +/- 15% of the valuation. Some 72% of properties were accurately valued at the +/-10% level.

A recent analysis of 10 leading AVMs by the FNC Corporation in 2017 (now owned by Corelogic), for homes purchased in June and July of 2006 covering 48 US states, revealed that 50% of the valuations for each of their models were off *by at least* 20%. No details about the size of the sample were provided, i.e. the number of sales. Given only two month's data, one cannot infer if the 20% is a representative figure. However, it does show that orders of this magnitude where valuations are off by more than 20% half of the time, albeit over short periods, are possible.

Two large AVM venders, HouseCanary and Zillow, continually provide updated accuracy figures on their websites in some detail. Both appear to be transparent and open in disclosing the accuracy estimates. HouseCanary say that their figures are independently verified, whilst Zillow do not confirm whether or not their accuracy figures have been independently examined.

HouseCanary provide AVM forecasts across 50 US States. They claim to have the most accurate property valuations available online, the figures being validated by a third party every quarter. They report that for February 2017 sales, the median absolute percentage error was 4.8%, where half of the estimates were within +/- 4.8% and half outside the +/- 4.8% interval. Over the six months to February, the highest reported median figure was +/- 5.1%, which shows a consistency in the error profile. Table 2 summarises the figures across 50 States. It shows the maximum and minimum percentage of properties within a range of +/- 10% of the sales price, together with the maximum and minimum median figures across the 50 States. HouseCanary operate 8 individual AVMs. As can be seen, there are considerable differences in the distribution of accuracy across the 50 States.

Table 2: HouseCanary Valuation accuracy

Statistic	Within +/- 10%	Median
Min	39.9%	3.4%
Max	81.5%	15.0%

Source: HouseCanary Website

Zillow claim to be the largest AVM provider in the US. A detailed spreadsheet, which they make available, was downloaded from their website. This provides accuracy figures at the National, State and County level. The data is sizeable enough enabling descriptive statistics to be estimated, thereby facilitating a comparison of the distribution of accuracy figures across a wide range of locations. Table 3 provides summary statistics calculated from Zillow's raw figures, summarising some of the features at the individual County level, based on 666 Counties across the US.

Distribution of the accuracy across 666 individual locations (Counties) in the US:

Table 3: Percentage of valuations in Counties falling within the specified limits

Statistic	Within +/- 5%	Within +/- 10%	Within +/- 20%	Median
Average (across all locations)	48.7%	69.7%	84.7%	6.0%
Min (lowest % for a location)	9%	20%	37%	3%
First Quartile (25% of locations)	42%	63%	80%	4%
Third Quartile (75% of locations)	57%	79%	92%	7%
Max (highest % for a location)	76%	92%	100%	25%

Source: Estimated by the author based on raw data obtained from Zillow's website

Note: Min is the lowest percentage of valuations for a particular County falling within the specified +/- limits. Max is the highest percentage of valuations in a particular County falling within the specified limits.

The figures show the following:

- The median level of valuation error across 666 Counties in the US is 6.0%. Meaning, half of the errors nationwide were within 6% of the final selling price, and half had an error exceeding 6.0%.
- At the *individual County level,* the median ranged from 3% to 25%, which represents a wide range of variation across the different locations.
- On average, almost half of all valuations across all Counties were within +/- 5% of the sales price and half being in excess of +/- 5%. However, in one County only 9% of the valuations were within the 5% bracket. The highest recorded individual County accuracy figure was 76%.
- On average, the percentage of valuations across all Counties falling within +/- 10 % of the sales price is 70%. However, this can vary between 20% and 92%, depending on the County.

- On average, the percentage of valuations across all Counties falling within +/- 20% of the sales price is 85%. However, this can vary between 37% and 100%, depending on the County.

Distribution of Zillow's AVM valuation errors

The Appendix shows three histograms, displaying the distribution of the 666 Counties AVM accuracy errors: Figure A1: shows +/- 5% valuation accuracy rates; Figure A2: +/- 10% valuation accuracy rates and Figure A3: +/- 20% valuation accuracy rates. The histograms provide a detailed visual insight of the accuracy errors across the 666 US Counties. The Average accuracy figures from Table 3 are superimposed in order to provide a reference point. On balance, it appears that some 50% of the valuations are likely be outside the +/- 5% range of achieved sales price, which falls to 30% for the +/- 10% range and 15% for the +/- 20% range, as reported in Table 3.

As can be seen from 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. This has implications for valuer contribution.

Comparing average median figures, HouseCanary's figure was in the region 5% and the Zillow figure 6%. Other figures which can be compared are the percentage of properties with sales prices within +/- 10% of the AVM valuation. The minimum is 40% for HouseCanary and 20% for Zillow, the maximum figures being 81.5% and 92% respectively. However, we must be mindful that the HouseCanary figures are at the State level whilst Zillow's are at a smaller spatial area, namely County. Unfortunately, HouseCanary figures at the +/- 5% or +/- 20% levels are not available.

The Zillow figures, being more finely grained, show that the differences in accuracy figures across County locations within all +/- % bands can be significant, and delivers an important message. This shows that one cannot generalise the accuracy numbers nor apply the *same* accuracy number in all locations. Accuracy levels vary geographically.

If a +/- 10% accuracy figure is seen as being a realistic margin for error, and the Zillow figures are regarded as representative then, *on average*, the Zillow figures suggests that some 70% of AVM valuations would fall within the +/- 10% bracket.

Comment on US AVM accuracy figures

Given their number, the longer experience in developing and using AVM models, it may be assumed that the US models are leading-edge and more established compared to their European/UK counterparts. It may be argued the US models would define the upper limit for European/UK AVM accuracy levels which, if anything, would likely be lower. The US average of 70% of valuations falling within a bracket of 10% would seem an optimistic estimate for European/UK AVM valuations. Furthermore, given a 10% bracket, anything between 20% (the Zillow min) and 92%

(the Zillow max) is possible. If a 5% bracket was regarded as a tolerable margin for error, the levels of accuracy would fall considerably (see Table 3).

The US AVM accuracy figures need to be put into perspective. The figures reported in Tables1, 2 and 3, together with the distribution of the County accuracy profiles shown in histograms A1, A2 and A3, clearly show that there is a wide distribution of inaccurate AVM valuation figures across the US Counties. US AVM models should not be perceived as providing seemingly consistent and reliable valuations across the board; sometimes they do and sometimes they don't. They are no panacea.

Whist the use of AVMs is widespread in the US, this needs to be put into perspective. In an interview in 2016, the CEO of *Platinum Data Solutions*, a significant provider of valuation technologies to the mortgage banking industry, observed: "AVMs are going to get more and more mainstream, particularly as data and analytics get more sophisticated. AVMs won't take the place of an appraisal. There will always be a need for local knowledge and expertise, not to mention an on-site evaluation of the physical property."

Discussion

What is an acceptable margin for error? This is a long-established concept which is developed in case law and is not within the scope of this report. In this regard, perhaps there are guidelines which can be drawn upon from TEGoVA's membership experience?

One observation which can be made is that market conditions would likely raise or lower the margin. Indeed, although a different market from the residential market, in the commercial property markets there can be significant variability in valuation accuracy in different periods, including biased valuations when markets are moving relatively rapidly (Matysiak & Wang,1995) or when markets are 'thin'. An acceptable margin in commercial property appears to be in the region of +/- 10%. Indeed, figures reported by *MSCI* put the European country averages as varying between 9% and 12.6% over the eleven-year period 2005-2015, depending on country. However, recent UK case law may put the average in the +/- 15% bracket, depending on how 'specialised' the property is. The residential markets are more liquid with a greater volume of transactions than commercial real estate markets, and consequently, perhaps it would be expected to have lower margins for error in 'normal market' conditions.

The question then is, what is a 'normal' market and what would be an acceptable margin for error for residential properties in such a market? Indeed, there is a whole series of conditions which would need to be taken into consideration when looking to assess what would be an acceptable margin for error in valuing a residential property, including:

- Different market environments including rising/falling/volatile markets
- Different size/value properties
- Quality of property

- Age of property
- Market liquidity e.g. dependent on the volume of transactions
- Different neighbourhoods
- Geographical location
- Type of property

Many of these are likely to be country specific and so, the margin for error will vary in different countries.

Conclusion

Over the last 20 years, property valuation has evolved from traditional manual sales comparison methods and subjective valuer assessments based on comparables evidence, into mechanically oriented valuation models.

Advances in the availability of computer technology and data management systems have enabled the widespread development of AVMs. AVM vendors would emphasise that the model estimate provides an indicative valuation, an indication of the likely sales price the property would achieve in the open market. Furthermore, given the uncertainty surrounding the estimate, the extent of the uncertainty is also provided, which can be variously expressed.

Perhaps not unsurprisingly, there is little available published material on the accuracy of European AVMs. Indeed, there appears to be a reluctance to provide information or open-up methodologies more widely to independent scrutiny. In the circumstances, the accuracy of current AVM European-based services remains largely unverifiable.

Consequently, based on the limited amount of available data, it is not possible to reach unqualified conclusions about AVM valuation accuracy in European markets.

In comparison with European vendors, US AVM service providers are relatively much more open in making access to their accuracy results publicly available. Indeed, Zillow have recently introduced a facility for academics enabling access to their considerable database for research purposes.

The distribution of the accuracy figures of the US models, *across* both locations and *within* locations, appears to provide tolerable results which could be considered as acceptable levels of statistical confidence for AVM valuations. However, a purely derived statistical or data-mined valuation risks being widely off the mark, as reported in the various Tables in the report. Consequently, despite the high degree of accuracy reported by the US AVM vendors, there remains a requirement for professional judgement to augment model-based valuations, thereby arriving at a more broadly considered valuation estimate.

Expressed more resolutely, valuer involvement can be supported on purely statistical grounds in terms of Bayesian statistical reasoning. The AVM model estimates a value based on *prior* information, which is subsequently modified by any potentially new information not previously taken into account by the model, the *posterior* information, being the additional (local?) information the valuer will possess. Put prosaically, this may also be regarded as a check on the soundness of the model-based valuation.

AVM observers will always raise the question, and rightly so, how impartial any reported AVM accuracy figures are. Consequently, any reported AVM vendor figures together with attendant conclusions need to be critically evaluated. Property values are determined by a mix of qualities and conditions, a model only capturing the broad characteristics, leaving the detail out. Given that there will be differences in information/knowledge about a local market, which may not be widely disseminated, this makes for an imperfect market. Consequently, a thorough assessment of value requires not only experience where judgement is called on, but also knowledge of local market conditions where individual properties may be dissimilar in a variety of ways.

To put AVMs into perspective, Tretton (2007) takes the view that AVMs contribute to the process of arriving at a valuation, but ultimately the quality and accuracy are data and valuer led; there is no automated replacement for subjective professional judgement.

In order to address the questions raised in the *Discussion,* and provide a transparent basis for assessing the robustness and accuracy of AVMs, full details of the methodologies employed in arriving at values are required. However, reliance on analyses and results reported by AVM vendors themselves will always raise questions of how impartial the reported results really are. Consequently, any reported vendor conclusions need to be critically evaluated. In which case, consistent and transparent standards are required.

How accurate are valuations? In the present context, there are two separate questions which can be addressed regarding valuation accuracy profiles, namely:

- AVM valuations compared against professional physical valuations
- AVM valuations compared against achieved market prices

Most commentators would argue that physical valuations are most likely to be more accurate, given the idiosyncrasies of individual properties. This is an empirical matter requiring valuation and price data.

Regardless of how complex or elaborate an AVM might be, how much data was employed in estimating it or how many comparables were selected in arriving at the prediction, the bottom line test is, how ‘accurately’ did it predict property prices? The predicted estimate will be qualified by the forecast uncertainty, whether it be expressed by a range or some other measure of confidence, such as the FSD. If it does not predict accurately, the model is deficient. Consequently, prediction accuracy offers an objective way to validate the robustness of an AVM.

There needs to be more discussion about what is a fitting framework for assessing and evaluating AVMs. Several published papers have considered the issue of assessing and evaluating AVMs, and the type of considerations which should be taken into account in defining a suitable framework for this task.

As examples, in the US the issue of best practice in the AVM area has been addressed by bodies such as the CATC (2009). A CoreLogic (2015) paper provides an overview of AVM usage in Australia and New Zealand, providing interesting guidelines for best practice, validation and monitoring of AVMs. The report usefully identifies and discusses a number of aspects which need to be addressed when evaluating AVMs. One key message which emerges is that effective validation of AVMs is hampered by the lack of industry standardisation across virtually all aspects of the AVM process.

The independent validation and standards of validation of European AVMs needs to be promoted more vigorously, otherwise the role of AVMs will continue to be misunderstood and contested. Until such time as AVM vendors make their models more widely available to open scrutiny, their claims of robust and accurate models cannot be regarded as impartial.

Two recommendations follow on from this report:

- It would be valuable to undertake a survey of how extensively AVMs are used, together with users' attitudes to and experience of AVMs. The survey would address a series of structured questions, including the potential/lack of potential for AVMs in particular markets or countries.

- TEGoVA should consider the value of creating an AVM working group, with a remit of defining industry standards of best practice, including testing and reporting standards, together with monitoring the development of AVMs. Furthermore, TEGoVA may take the initiative and facilitate an educational role in the area of AVMs.

This report has provided the key practical points which have emerged as a result of the research undertaken on AVMs. An important proviso needs to be made regarding the information contained in the report. *The information reported here is based on what has been provided by the AVM vendors, directly or indirectly, and has not been independently verified.* This qualification should be kept in mind.

References

Evgeny A. Antipov and Elena B. Pokryshevskaya (2012), Mass appraisal of residential apartments: An application of Random forest for valuation and a CART-based approach for model diagnostics, *Expert Systems with Applications,* Vol.39, pp. 1772-1778.

Steven C. Bourassa, Eva Cantoni, and Martin Hoesli (2010), Predicting House Prices with Spatial Dependence: A Comparison of Alternative Methods, *Journal of Real Estate Research,* Vol.32, No.2, pp. 139-160.

Paul E. Bidanset and John R. Lombard (2014) Evaluating Spatial Model Accuracy in Mass Real Estate Appraisal: A Comparison of Geographically Weighted Regression and the Spatial Lag Model*, Cityscape: A Journal of Policy Development and Research*, Vol.16, No.3, pp. 169-182.

CATC (2009), Proposed Interagency Appraisal and Evaluation Guidelines - Docket No. OP-1338.

CML (2007) *Automated Valuation set to grow sharply.* CML News and Views, No.10 June 2007, Council of Mortgage Lenders.

CoreLogic (2015), "Best Practice Validation and Comparison for Automated Valuation Models (AVMs).

Downie, M.L. and Robson, G. (2008), Automated Valuation Models: An International Perspective, Council of Mortgage Lenders, London.

Fitch (2007) *Criteria for automated valuation models in the UK.* 2nd June 2006.

Fitch (2016) *Criteria Addendum: UK. 3rd February 2017.*

Grover, R. (2016) "Mass valuations", *Journal of Property Valuation & Investment*, Vol. 34, No. 2, pp. 191-204.

I.A.A.O. (2003), Standard on Automated Valuation Models, International Association of Assessing Officers, Kansas City, MO.

Kauko, T. and d'Amato, M. (2008), "Introduction: suitability issues in mass appraisal methodology", in Kauko, T. and d'Amato, M. (Eds), Mass Appraisal Methods: An International Perspective for Property Valuers, Wiley-Blackwell, Chichester, pp. 1-24.

Kauko, T. and d'Amato, M. (2017)," Advances in Automated Valuation Modelling", Springer.

Kilpatrick, J. (2011), "Expert systems and mass appraisal", *Journal of Property Investment & Finance*, Vol. 29, Nos 4/5, pp. 529-550.

Lenk, M.M., Worzala, E.M. and Silva, A. (1997), "High-tech valuation: should artificial neural networks bypass the human valuer?", *Journal of Property Valuation & Investment*, Vol.15, No.1, pp. 8-26.

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

Matysiak, G.A. and Wang, P. (1995), “Commercial property market prices and valuations: analysing the correspondence”, *Journal of Property Research*, Vol.12, No.3, pp.181-202.

William McCluskey, William Deddis, Adam Mannis, Dillon McBurney and Richard Borst. (1997), “Interactive application of computer assisted mass appraisal and geographic information systems”, *Journal of Property Valuation & Investment*, Vol.15, No.5, pp. 448-465.

McCluskey, W.J. and Borst, R.A. (1997), “An evaluation of MRA, comparable sales analysis, and ANNs for the mass appraisal of residential properties in Northern Ireland”, *Assessment Journal*, Vol.4, No.1, pp. 47-55.

McGreal, S., Adair, A., McBurney, D. and Patterson, D. (1998), “Neural networks: the prediction of residential values”, *Journal of Property Valuation & Investment*, Vol.16, No.1, pp. 57-70.

McCluskey, W.J., M. McCord, P.T. Davis, M. Haran, and D. McIlhatton (2013), “Prediction Accuracy in Mass Appraisal: A Comparison of Modern Approaches”, *Journal of Property Research,* 30, No.4, pp. 239–265.

Osland, L. (2010), “An Application of Spatial Econometrics in Relation to Hedonic House Price Modelling”, *Journal of Real Estate Research*, Vol 32, No.3, pp. 289-320.

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

Rossini, P. and Kershaw, P. (2008) “Automated Valuation Model Accuracy: Some Empirical testing, *Fourteenth Annual Pacific-Rim Real Estate Society Conference*, Kuala Lumpur, Malaysia, January, 2008.

Schulz, R., Wersing, M. and Werwatz, A. (2014), “Automated valuation modelling: a specification exercise”, *Journal of Property Research*, Vol.31, No.2, pp. 131-153.

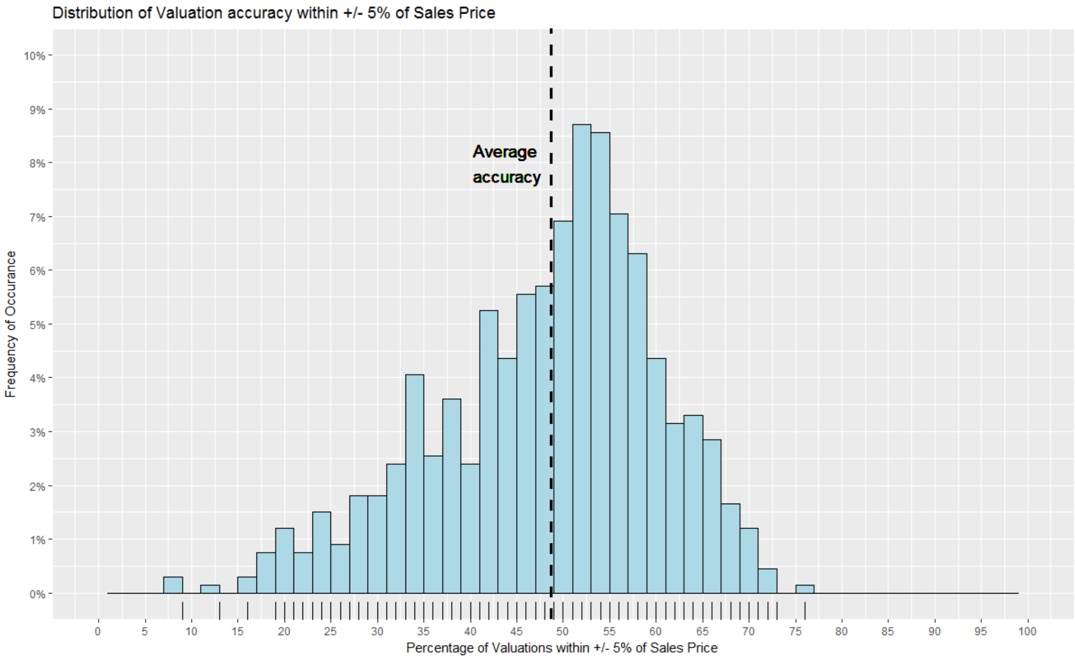
Tretton, David. (2007), “Where is the world of property valuation for taxation purposes going?”, *Journal of Property Investment & Finance*, Vol.25, No.5, pp. 482-514.

Worzala, E., Lenk, M. and Silva, A. (1995), “An exploration of neural networks and its application to real estate valuation”, *The Journal of Real Estate Research*, Vol.10, No.2, pp .185-201.

Zurada, J., Levitan, A. and Guan, J (2011) “A comparison of regression and artificial intelligence methods in a mass appraisal context”, *Journal of Real Estate Research*, Vol.33, No.3, pp. 349-388.

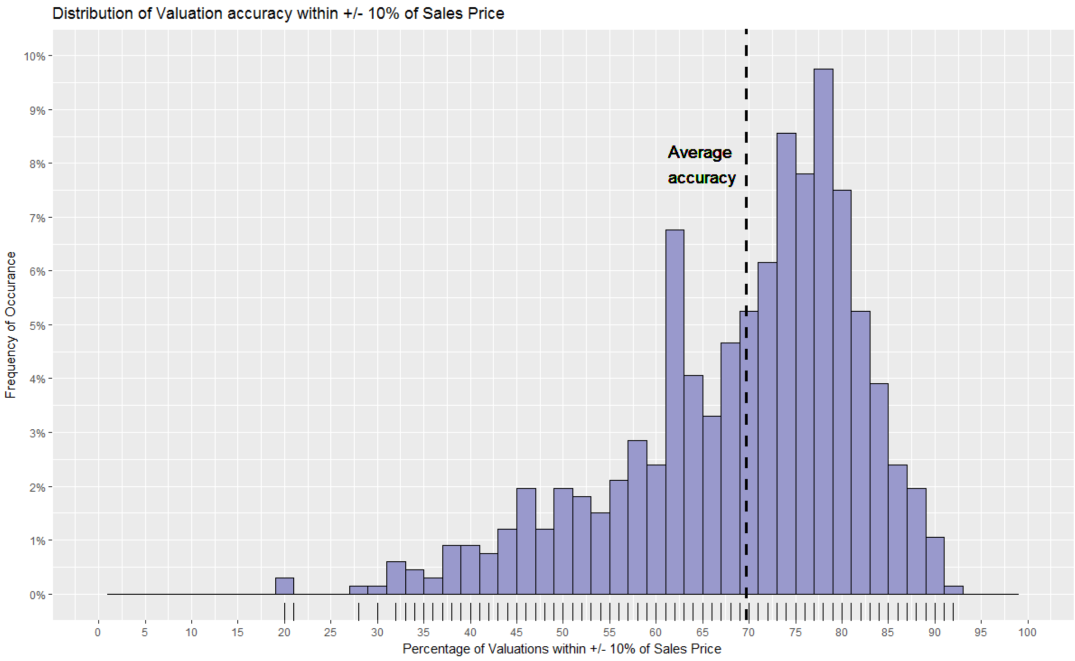
Appendix: Distribution of Zillow's valuation accuracy figures

Figure A1: 5% Accuracy rates



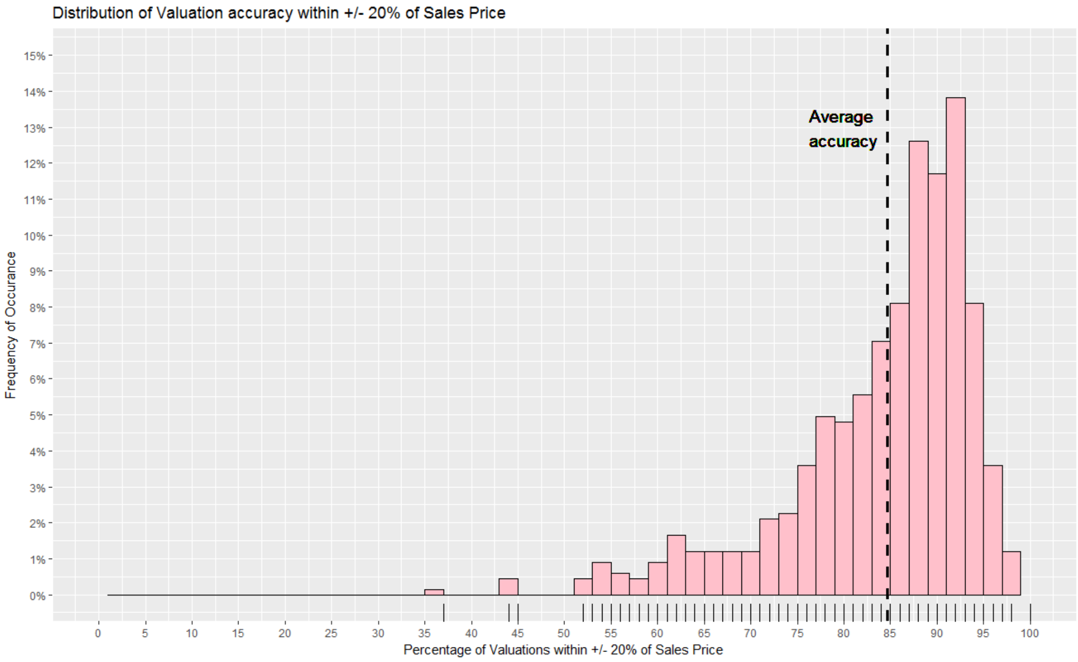
Source: Zillow and author's summary

Figure A2: 10% Accuracy rates



Source: Zillow and author's summary

Figure A3: 20% Accuracy rates



Source: Zillow and author's summary