14th Pacific Rim Real Estate Society Conference
Istana Hotel, Kuala Lumpur , January 2008

Automated Valuation Model Accuracy: Some Empirical Testing

Peter Rossini & Paul Kershaw
Centre for Regulation & Market Analysis
University of South Australia

Abstract

In recent years there has been a rapid move towards the adoption of Automated Valuation Models (AVMs) for basic residential valuations for mortgage security purposes. In Australia several commercial organisations now offer such residential valuations, but very little is known about the accuracy of these. While the accuracy of commercial models is almost impossible to measure, one problem is that few people understand the practical implications of the accuracy measures. Put simply most users of these products misunderstand the level of accuracy involved. This paper uses a data base of sales in Adelaide, South Australia to measure the accuracy of basic automated valuation models under a range of different model scenarios. Some very simple "models" based on the application of elementary descriptive statistics provide remarkably accurate estimates based on some evaluation criteria. Other models using more sophisticated hedonic price models produce better results but the small increase in accuracy may not justify the cost to develop and maintain these models. It is hoped that this paper will lead to a better understanding by users of these products of the accuracy of Automated Valuation Models.

Key words: Automated Valuation Models, Residential Valuation, Valuation Accuracy

Background

The paper results from a number of questions that have recently arisen in terms of the accuracy of Automated Valuation Models (AVMs). The first author of this paper has previously presented some of these issues (Rossini, 1999) but at the time AVMs were not commercially available in many jurisdictions and the paper reflected the expectations of accuracy at the time rather than actual performance. This paper updates this research now that AVMs are being used by consumers.

Early signs of the development of AVMs can be traced back to the 1960's and certainly their use has been well publicised for rating and taxation assessments, over many decades. The use of these outside of North America has been limited until the last decade and there is now evidence that they are being widely considered for security valuations on a much wider basis. The issues of AVM methodology, testing and accuracy has been addressed by the International Association of Assessing Officers (IAAO) over a long timeframe but this primarily deals with the issue of assessed values for rating and taxation where a standard set of accuracy tests (through ratio studies) has been established. However, there are very few clear benchmarks as to performance. Much of this will be due to different market situations. The degree of accuracy that you might expect from a model in North America where the assessments are used as the basis of significant property taxation[1] and markets operate opening and vigorously, will be different from locations where markets are thin and there is no significant local assessment authority.

In Australia AVMs have only recently become commercially available on a broad basis and Australia is considered to be an early stage market for these products (Downie et al. 2007). While some States such as South Australia and Western Australia have, for several decades, had significant data sets and used basic AVMs for taxation assessments, this has not been the case in the more populated Eastern States. Much of this is probably due to the differences in property taxation assessment methods. South Australia and Western Australia have used a capital value based assessment over this time while site values have been used over long periods in the Eastern States. Over the last twenty years a number of real estate data resellers have emerged or expanded to capitalise on the opportunity to sell data and produce extended products such as residential price indices and AVMs. The widespread use of the internet for real estate advertising has assisted them in collecting data[2] . These AVMs are now being used in Australia as a replacement for residential valuations in mortgage security situation. There is also evidence that some banks may be using them on an ongoing basis to check the security of residential loan portfolios.

---

[1] Where property taxation is used to derive a high proportion of government revenue the quality of assessments for rating and taxation becomes more significant and this has resulted in high quality data being collected and maintained with high quality assessments produced.

[2] Most of the major data resellers use a combination of secondary data from Government such as Land Title data and primary data. An association with an internet based real estate site affords them significant opportunities to augment the basic data at a low cost.

However very little is known about the accuracy of these models. In a recent discussion involving a consumer magazine that was trying to establish the validity of current AVMs, only one of the major sellers of AVMs was prepared to offer any suggestion of the level of accuracy and this was couched in very cautious terms. A suggestion that about 70% of values might be within + or – (plus or minus) 15% was given, but this varied on the location and amount of data. But this figure seemed very low and it was doubtful if it would be much better than an assessment based on a basic summary statistic (say the median value in the area).

This raised a number of questions.

1. Is there a well recognised methodology for measuring the accuracy of such models and is it independent of the AVM producer?
2. What are the test statistics that are used and how easy are these to interpret?
3. Are results of these tests readily available?
4. What are acceptable benchmarks given different circumstances?
5. In the absence of published accuracy results, could we develop some benchmarks for use in our local area by some empirical testing of basic models?

Literature

The definition of an AVM has changed over time. In a significant new report on AVMs in an international perspective, an AVM is described as a

> *"…software systems that use one or more mathematical techniques to estimate the Market Value of a given property at a specified date. They automatically select the relevant market information and the appropriate method of aggregation by means of one or more proprietary algorithms that typically perform a statistical analysis of available property market information.*
>
> *They utilise databases of public and private property records, historical appraisal reports and multiple 'for sale' listing services that provide property characteristics and sales information. They offer the competitive advantage of speed and low cost. Their methodology aims to guarantee an objective valuation without appraiser subjectivity or bias." Downie, et al (2007) p 10*

They then go on to suggest that these may be based on a number of different methods such as indexing, TAV models, hedonic models, intelligent systems and econometric forecasts. This is probably a sensible modern definition of an AVM but it is considerably broader than that defined by the IAAO for a considerable period of time. The IAAO standards define an automated valuation model (AVM) as

> *"… a mathematically based computer software program that produces an estimate of market value based on market analysis of location, market conditions, and real estate characteristics from information that was previously and separately collected. The distinguishing feature of an AVM is that it is a market appraisal produced through mathematical modelling. Credibility of an AVM is dependent on the data used and the skills of the modeller producing the AVM" (IAAO, 2003 pp148)*

This concept of it being a market appraisal following the broad principles and data used in a typical valuation seem to be largely ignored by many modern AVM producers who are concerned with "results" rather than "process". This is probably the reason that professional organisations such as the Australian Property Institute (API) view their development with significant resistance and have moved to clarify where the boundaries are. While it is accepted that valuers need to be involved with the development of AVM rather than simply resist their introduction (Fortelney & Reed, 2005). It is also sensible that appropriate professional bodies clearly define the boundaries and the Australian Property Institute (API) have sought to do this (API, 2006) although their paper deals with professional valuers producing "desktop valuations" rather than non-valuers producing residential assessments though an AVM. Perhaps the valuation profession would be less concerned with a suggestion by Fitch (2007) in relation to AVMs in the UK when they suggest that AVMs will typically be an amalgamation of three modelling components namely comparable sales, repeat sales analysis and a hedonic price component all of which have strong ties to traditional valuation methods and concepts.

The emphasis for commercial providers of AVMs seems to be accurate results – regardless of the process. While this seems inherently reasonable (does a lending institution really care how you get the result as long is it's correct) advice from the baking and mortgage industry seems to suggest otherwise. There do not appear to be any significant benchmarks as to the required level of AVM accuracy. This is true even in the USA where such models have become widespread. In fact it appears that this sector is more concerned with process than having benchmark levels of accuracy (OCC ,2006, APRA, 2005, JP Morgan and Fujitsu, 2006) The Mortgage Brokers Association (2006) go on to state that there is a lack of standardisation in the AVM industry in the USA.

Accuracy Testing

There seems to be a number of emerging methods for testing AVMs and these have been grouped into three categories, but there is considerable intersection.

Commercial users – rating and taxation authorities

These are listed first due to the long history of these tests. The IAAO established a broad range of tests for assessments several decades ago. These are based on the ratio of assessed values to actual sale prices in ratio studies (IAAO, 2005). Ratio studies rely upon the use of the A/S ratio; the ratio of assessed value to the sale price (or independent valuation). The A/S ratios are then charted in various ways, described and inferential statistics used to determine the accuracy of the assessments. Typically the study uses a variety of parametric and non-parametric tests. These tests include:

1. Measures of assessment level; *mean*; *median*; weighted mean and geometric mean ratios.
2. Measures of Variability; Coefficient of Dispersion (*COD*); Coefficient of Variation (*COV*) and quartile ranges.
3. Measures of Reliability; confidence intervals and standard errors.
4. Vertical Inequities; Price Related Differential (*PRD*)
5. Hypothesis tests; Normality (for example Shapiro-Wilk test), two groups tests for equality (Mann-Whitney), three or more groups tests for equality (Kruskal-Wallis), sales chasing (Mann-Whitney).

While these provide a wide range of testing that considers many issues, there are no solid benchmarks. When reading material using these tests there are some suggested levels (see Rossini 2006a) but these are accepted as varying with circumstances.

Commercial users – banks and financial institutions

As AVMs are increasingly used for mortgage security purposes, a number of methodologies and tests have emerged and been published by financial ratings companies. Although not standardised, there seems to a level of similarity amongst several of these that provides a useful general method. The testing focuses on the difference between values assessed by AVMs compared to those assessed by surveyor (valuer) valuations (typically large numbers of surveyor estimates are gathered and compared to the AVM estimate of the same property on the same day) and the primary test statistic is the Forecast Standard Deviation (*FSD*). This is supported by Downie, et al. (2007) at an international level, Bahjat-Abbas (2006), Fitch (2007) and Standard and Poor's (2007) in the UK and Bradley & Nuetzel (2007) for the USA. The Forecast Standard Deviation is defined as the standard deviation of percentage forecast errors where

$$PercentageError = \frac{SurveyorValue - AVMvalue}{AVMvalue}$$

Fitch (2007) acknowledges the criticism that surveyor values my be biased compared with AVM estimates[3] 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. Given the established wide variations in valuations (see Skitmore et al (2007), Man, et al (2006), Newell, et al (2006) and Rossini (1999) for a Pacific Rim context) there must be some concerns at measuring the accuracy of an AVM against valuers opinions of value. On the basis of this test, an AVM that accurately predicted sale prices might be seen as producing serious errors if the valuer or surveyor produced under or overvaluations. Given that an error of plus or minus 10% is reasonably typical in manual valuations the test becomes somewhat clouded although useful for the lending sector. It is also obvious from the literature (Fitch, Bahjat-Abbas, Standard and Poor's op cit) that the lending institutions apply a variety of reductions to AVM estimates to reduce the downside risk of using these estimates. These vary for different classes of property and for different AVM providers.

The methodology suggested in Fitch op cit is useful in that it suggests a method of AVM accuracy testing that is independent of the AVM creators. The different levels of reduction for each provider implies a variation in the level of accuracy across providers. They don't however suggest any standard benchmark levels or suggest if the AVM providers meet any specific level. The SFD used in these studies is suitable for comparing AVMs across different locations as well as for comparing models within the same location.

Academic papers

Academics have been producing technical papers on AVMs and general valuation modelling for many years and will typically use some test statistics to show levels of accuracy. It is not possible to consider all these papers but a small spectrum of recent papers in the Pacific Rim region suggest some broad methodologies.

The Mean Absolute Percentage Error (*MAPE*) is a widely used forecast accuracy test and appears in most academic papers (Ibrahim et al (2005), Lai & Fischer (2006), Rossini & Kershaw, (2005), Rossini (1997, 2000). It represents the average (absolute) error between the predicted value from the AVM and the actual sale price

---

[3] Typically the surveyor will have access to additional information such as a contract or asking price

of a property. It does not take into consideration the standard deviation. The MAPE is widely used and easy to understand by lay persons. As it is in percentage terms it makes a useful comparison across different models and across different data sets and locations.

The Root Mean Squared Error (*RMSE*) is also used in papers where different models are compared for the same data set or location (Lai & Fischer, 2006), Rossini 1997, 2000). The RMSE measures the error in dollar terms and is analogous to a standard error in a statistical model (the difference being only the denominator in most instances where the RMSE uses the number of observations and the SEE the degrees of freedom in the model). Being dollar based it is less useful for comparing AVMs in different locations where price levels vary.

Hit ranges are another typically used method for expressing the accuracy of an AVM (Lai & Fischer, 2006, Rossini, 1997). The percentage of assessments that fall within designated hit ranges is provided. Typically these hit ranges will be at 5%, 10%, 15% , 20% and 50%. This is a useful expression of the accuracy for several reasons. It can be compared to widely acceptable margins of error in valuation and it is easily understood by most people and often used by AVM providers.

These (and many more) academic papers often show that AVMs can be very accurate but typically these are the result of significant development for one-off situations. AVM providers are forced to produce more generalised models to cover large spatial areas and in a very timely manner, typically weekly. They must therefore use significantly generalised models. While academic studies may represent the possible accuracy levels it is unreasonable to expect that a typical commercial producer will meet these.

Methodology

The literature review in this paper has addressed four of the five issues mentioned in the background section. The remainder of the paper seeks to provide some guidance as to acceptable levels of accuracy for AVMs within Australia. The proposition is to establish a level of accuracy that is "unacceptably low" given typically available data and to make some suggestions of levels that might be consider acceptable at an empirical level. Academic papers that show high levels of accuracy from carefully devised models built with significant individual care and attention will represent the best possible outcomes but this does not give typical users a "feel" for what might be acceptable or useful. This paper starts at the other end and builds the most basic models as a point of reference. This is somewhat analogous to the approach taken in business forecasting where time series forecasts are typically compared to a naive model (what happened the last period will happen this period) and tested using the Theil's U statistic. A starting point for a typically users understanding of AVM accuracy would be the results from the most basic models.

To establish a minimum benchmark a series of models will be used to estimate values for properties that sold within a Local Government Area (LGA) in Adelaide, South Australia. A create-test methodology will be used which is typical of other studies using the same Adelaide based sales history data to compare a variety of models (Rossini and Kershaw, 2005, 2006, Rossini 1997, 2000). This means that data used to test the model is not used to create models; typical of a hold-out sample methodology used in most forecast testing.

Data

The research uses data from the Marion LGA, which is one of 19 LGA's in the greater Adelaide Metropolitan area. See Appendix 1 for the location of this LGA relative to the whole of Adelaide. The LGA has about 41,000 assessable properties of which approximately 36,000 are residential and 26,000 are detached houses and these are the focus of this research. The data set is based on detached residential properties that transacted in 2005 and 2006. In 2005 there are 1294 valid probable market transactions and these are used to produce a variety of models which are then used to predict values for the 1244 properties that had valid transactions in 2006. A simple SPAR index methodology is used to adjust values to 2006 levels since this method is found to be simple and efficient using the South Australian sale history data (Rossini and Kershaw 2006). While this is a relatively clumsy process compared to a typical AVM that would focus on using the most up-to-date data, it is simple and should produce results that can be easily replicated by others and form a suitable lower end benchmark. A commercial AVM should certainly produce something better!

Marion represents a heterogenous LGA with a mix in housing ages, styles and materials as well as variations in allotments sizes. Locational characteristics vary widely with some beach side suburbs, flat "plains" suburbs and elevated areas in the Adelaide foothills. Suburbs vary in size and homogeneity. The LGA was chosen for its variability's as this is likely to lead to less accurate but reasonable typical AVM accuracy.

The sales data base uses data from the Lands Titles Office and the Office of the Valuer General and includes basic sales details as well as some property characteristics. Since the latter data is not available in all Australian jurisdictions the data is separated into two parts. A basic data set that is universally available and an enhanced set that is not. Results can then be compared against the data availability. In addition to this data the Capital Value assessed by the Office of the Valuer General is available with the sales data and this forms a very useful comparison. The data sets are summaries as follows.

Basic Data Set

This data set is readily available for many locations and is derived from secondary data that is easily obtainable at a low price. In this instance it is based on transaction data from the Lands Title Office including the address of the property, the date of sale and sale price. Such a data base could also be compiled from other sources at a lower level of accuracy and completeness, e.g. through advertisements of properties. These data are then enhanced by the addition of a geo code and land area typically provided form a digital cadastral data base. In this instance the geocode is based on a GPS coordinate derived from publicly available Google Map. This results in a data base for all properties that includes.

- Property Address – number, street name and suburb
- Sale Date
- Sale Price in dollars
- GeoCode in latitude and longitude.
- Land Area in hectares

Data sets including these variables are believed to be readily available in most jurisdictions.

Enhanced Data set

The enhanced data would not be available in many instances, is more expensive and difficult to collect.

- Building Area (this is an estimated equivalent area based on a weighted average of the area of all buildings where the area is based on building application or actual measurement)
- Rooms (number of main rooms in the dwelling based on building application or estimated)
- Year of Construction (based on building application or estimated)
- Condition Code (subject external observation)
- Building Style, Roof and Wall materials (based on building application or external observation)

Omitted Data

The dataset did not include some variables which research has shown produce superior results. Rossini (1998) used the same historical sales data including the enhanced data (but excluding the Geocode) to create AVMs across Adelaide using both multiple regression and Automated Neural Network (ANN) models. He added some additional data based on personal observations and found that factors such as view-outlook, street size, site features and interior condition-quality, led to significantly superior models but accepted that models based on the enhanced data used in this paper were adequate for rating and taxation purposes. It is accepted that far superior models could be created if these data were available.

Model Specifications

The models tested in this paper are separated by the data inputs. In each instance they are first estimated then indexed to 2006 values. The models are as follows

The SA Valuer Generals Capital Values for 2006.

These assessments (listed as CV2006) are provided with the sales data and form an appropriate point of comparison. They are based on a computer assisted manual valuation process where a manual valuation base is indexed using a simple regression based annual index following submarket analysis. The VG collects and maintains the additional data base that is used elsewhere in this paper and this together with further information is available to them in the process. The assessments result from a manual but computer assisted sales analysis process that should improve results over a fully automated process. These values are considered to be quite robust and have been widely used by local banks and lending institutions in South Australia as the basis for security valuations (at lower proportions of lending) for many years. These values are adjusted using a SPAR approach to bring them to current market levels to overcome both the market price increase and a systematic underestimate to meet statutory requirements.

Local Government Area (LGA), Suburb level and proximate median prices

These assessments use the basic data set and rely on pooling sales to find medians. These assessments rely simply on finding the median price of differing samples of properties.

MedLGA   properties are assessed at *median* of all properties in the *LGA*.

MedSub   properties are assessed at the *median* of all properties in the relevant *suburb*.

Med Geo3   properties are assessed at the *median* of the *three* physically nearest properties (based on Geo code)

Med Geo6   properties are assessed at the *median* of the *six* physically nearest properties

MedGeo12   properties are assessed at the **median** of the **twelve** physically nearest properties

MedGeo25   properties are assessed at the **median** of the **twenty five** physically nearest properties

Simple Hedonic Model using Multiple Regression

This assessment uses simple regression hedonic models that allows for variations in land area and holds location constant at about a neighbourhood level. It uses only the basic data set.

Hed30Sim   assessments based on individual hedonic models (separate model for each property) using the **thirty physically nearest** properties (based on Geocode); sale price Vs land

Hedonic Models using Multiple Regression

These assessments use multiple regression model/s where the location component is held constant at varying levels or where location is estimated using a simple land value response surface (LVRS). The enhanced data set is used with variables being specified or entered using a stepwise process. Each model is "automated" and will change with new data.

HedLGA   assessments based on a **single hedonic** model using all sales in the **LGA**;

           sale price Vs land and building area, condition code and year built.

HedSub   assessments based on **25 suburb** level **hedonic** models;

           sale price Vs land and building area, condition code and year built.

Hed30Com   assessments based on individual hedonic models (separate model for each property) using the **thirty physically nearest** properties (based on Geocode);

           sale price Vs land and building area, condition code and year built.

HedLVRS   assessments based on a **single hedonic** model using all sales in the **LGA**, sale price Vs land and building area, condition code and year built.

           sale price Vs **Land Value Response Surface** and **extended property characteristics** on a stepwise basis. This model uses a MRA based LVRS using a simple polynomial expansion of the geo code and follows the process used in Rossini and Kershaw (2006). This is a very "weak" surface and a far superior result would be expected using higher level approaches. The results for this staged regression model is shown in appendix 2 and 3.

Evaluation Methods

The models are evaluated using the indicators suggested by the literature. The mean and median A/S ratio are calculated before the estimates are adjusted using the SPAR methodology. This provides a measure of how much adjustment has been required to bring them to market level. All other evaluators are calculated after the estimates are adjusted for market changes. The following evaluators are used.

General indicators	Commercial users comparison Statistic	IAAO Ratio Study Statistics
MAPE	FSD	Mean A/S (before indexing)
RMSE		Median A/S (before indexing)
Hit ranges		Normality test (Jacque-Bera)
Percentage within + or - 5%		COV
Percentage within + or - 10%		COD
Percentage within + or - 15%		PRD
Percentage within + or - 20%		
Percentage within + or - 50%		

In this paper all results are aggregated for the whole sample. In a future paper these will be stratified into smaller sub groups to establish the range of variability across different classes the manner shown by Fitch (2007). More homogenous areas will produce better estimates, less homogenous one significantly worse. As the study area has a representation of each it is suggested that these results will form a reasonable basis for a pooled benchmark.

Results

Table 1 shows the results for all the evaluators for all models. Highlighted cells indicate the "best" outcome for each evaluator for the given data set.

Table 1 - Evaluation results for all AVMs

Evaluator	CV2006	MedLGA	MedSub	MedGeo3	MedGeo6	MedGeo12	MedGeo25	Hed30Sim	HedLGA	HedSub	Hed30Com	HedLVRS
		Models using Basic Data							Models using Enhanced data			
MAPE	7.2%	15.9%	14.1%	13.7%	13.1%	12.9%	12.9%	13.0%	11.6%	9.2%	9.4%	9.7%
RMSE	$30,298	$79,563	$69,318	$ 64,383	$ 63,094	$ 63,799	$ 64,577	$ 67,094	$55,076	$48,167	$ 48,713	$ 45,475
Percentage within + or - 5%	43%	21%	24%	25%	25%	27%	27%	28%	27%	38%	40%	33%
Percentage within + or - 10%	72%	39%	44%	47%	48%	49%	48%	50%	51%	65%	65%	61%
Percentage within + or - 15%	90%	56%	61%	64%	65%	67%	65%	67%	70%	82%	80%	79%
Percentage within + or - 20%	98%	68%	74%	78%	79%	79%	79%	79%	83%	90%	89%	90%
Percentage within + or - 50%	100%	98%	99%	98%	99%	99%	99%	99%	100%	100%	100%	100%
FSD	9.1%	25.8%	21.2%	19.1%	18.9%	19.3%	19.4%	21.8%	17.3%	14.9%	14.2%	13.4%
Mean A/S (before indexing)	0.878	0.990	0.975	0.955	0.957	0.959	0.961	0.962	0.973	0.956	0.957	0.965
Median A/S (before indexing)	0.871	0.993	0.968	0.937	0.943	0.952	0.958	0.955	0.965	0.950	0.950	0.951
Normality test (JB)	16	5985	4027	4829	7134	7902	6435	67106	132329	173720	24178	62115
COV	9.176	20.423	18.228	18.208	17.258	16.943	16.857	17.346	14.763	12.546	12.979	12.593
COD	7.210	15.852	14.225	13.912	13.281	13.025	12.974	13.060	11.735	9.263	9.450	9.743
PRD	1.005	1.050	1.038	1.027	1.029	1.031	1.032	1.031	1.024	1.013	1.014	1.016

The Valuer General CV for 2006 provides an excellent point of comparison. The assessed values show a high degree of accuracy with a MAPE of only 7.2%; RMSE of $30, 298 and a Forecast Standard Deviation of only 9.1%. The hit ranges shows that 90% of estimates fall within +- 15%. The PRD indicates almost no level of regressivity and this is a very good result as some regressivity is expected in AVMs. The Jacque-Bera test for normality shows that the estimates are not normally distributed (a value below 6 is required with this data at a 95% level of confidence). This is an interesting finding because it must call into question the validity of the FSD which is difficult to interpret if the data is not normally distributed.

The models using the basic data give a clear idea of what can be expected with the most basic of models. The most basic estimate where every property is at the LGA median price, provides a degree of accuracy that would surprise some people. The model based on the median suburb value provides a very useful lower benchmark figure. Median suburb prices are often published and are extremely easily obtained. Estimating all properties at this level would be considered extremely naïve and most users would probably consider such an estimate to be of no value. However within this reasonably typical location, covering 25 suburbs, 61% of estimates would fall within +- 15% and 74% of estimates within +-20%. A MAPE of 14.1 percent does not sound ridiculously high.

The best estimates for the basic data can be made using models with a small number of observations geographically close to the estimated property. Simply finding the median value of the 12 closest properties yields results with a MAPE of 12.9% with 67% of estimates falling within +-15% and 79% of estimates being within +- 20%. Almost half of the estimates (49%) fall within the generally accepted range for valuation of +-10%. It is important to remember that these in **no way** comply to even the most basic requirements of a valuation method. Valuation practice would specifically prohibit the simple averaging of sale prices, and requires some adjustments for variations in properties. Arguably the hedonic models move towards this approach. The most basic hedonic model that allows for variations only in land size produces results very similar to that based on the median of the 12 closest properties.

All of the hedonic models using the enhanced data provide improvements to the estimates. This is expected and provides clear evidence that improved AVM performance is largely data driven. Even the basic LGA level model with four variables provides a significant improvement in all the models using the basic data set. This model can be improved by placing a very simple LVRS and this model shows that even a relatively simple MRA model that includes a handful of property characteristics and a simple LVRS to pick up major location variation can produces quite meaningful estimates.

Outside of the VG capital values this model produces the best estimates when evaluated using the RMSE and FSD. The best estimates in terms of MAPE and hits is produced by estimating each property using a 4 variable MRA model specific to that suburb. In this model location is held constant within each suburb but this is sufficient to produce very meaningful results. This model shows overall pooled results with a MAPE of 9.2% with 82% of estimates within +-15% and 90% of estimates within +- 20%. The IAAO statistics would suggest that the outcome is probably suitable for rating and taxation assessments.

None of the estimates produce percentage errors that are normally distributed using the Jacque-Bera test. The presumption that the errors are normally distributed is not supported in these models and this means that using the FSD to imply confidence levels is erroneous.

None of the models produce estimates that outperform or even approach those of the VG's capital values. Given that these values are readily available and are so such accuracy it is not surprising that they are used as good evidence to support home mortgages at low lending levels. Significantly better models would be needed to outperform these estimates that follow a simple indexing to equate them to market levels.

Importantly, none of the models used in this research would be considered to be complex. Those using only the basic data are simplistic. Those using the enhanced data are simple to estimate and could be estimated quite well using software such as Excel. All are within the learning curriculum of a modern undergraduate property student.

Discussion

The results for this simple research help establish some simple benchmarks that might be expected from an AVM. The basic models suggest an ***absolute minimum benchmark***. They imply limits where no significant allowance has been made of most major value determining characterises. These benchmarks based on a large pooled set of estimates would probably be: MAPE : 13%; 50% of estimates within +- 10%; 65% of estimates within +- 15%, 80% of estimates within +- 20%; FSD less than 19%; COV of A/S ratios less than 17 and COD less than 13. AVMs that produce such accuracy levels would have to be argued as being of no real value to users.

The best of the models using the enhanced data is still easy to estimate but may provide a level of accuracy that is below individual manual valuations but which do make adjustments in value for major property and location characteristics. The benchmarks would seem to set ***reasonable level of acceptance*** for estimates using this data. Reasonable bench marks might be: MAPE 10%, 65% of estimates within +- 10%; 80% of estimates within +- 15%, 90% of estimates within +- 20%; FSD less than 15%; COV of A/S ratios less than 13 and COD less than 10.

The results suggest that other indicators may be more suitable. A simple test based on ANOVA or the coefficient of Determination would be sensible. Regression modellers would recognise that the reason that most of these indicators are reasonable even in a model based on the average (or median) is that the major component of price is a constant and that regression statistics such as the coefficient of determination measure the explained portion of the variation in price. Hence a model estimating all properties at the mean price would yield a coefficient to determination of zero as none of the variations in price (about he mean) has been explained. A simple way to show this is to plot the actual prices against the estimates. The resulting correlation coefficient is also a useful indicator of accuracy. In practice many better statistical tests are available but these are probably rejected due to interpretation difficulties. The current crop of evaluators is also intuitively attractive because they are easy to interpret and also make the estimates appear to be precise. The downside of this is that a lay person will gain a false impression of the accuracy. Hence the need for proper benchmarks.

Conclusions

This paper highlights some of the dangers in using basic evaluators in determining the accuracy of AVMs. It also suggests two levels of benchmarks that are appropriate. The fist being a minimum level that suggests that an AVM provides no better accuracy than basic submarket averages. The second being levels that might be considered to be a reasonable level of acceptance. These benchmarks are.

Absolute minimum benchmark	*Reasonable level of acceptance*
MAPE : 13%	MAPE 10%
50% of estimates within +- 10%	65% of estimates within +- 10%
65% of estimates within +- 15%	80% of estimates within +- 15%
80% of estimates within +- 20%	90% of estimates within +- 20%
FSD less than 19%	FSD less than 15%
COV of A/S ratios less than 17	COV of A/S ratios less than 13
COD less than 13	COD less than 10

Bibliography

IAAO (2005) "Ratio Study Revision – Draft 12", http://www.iaao.org/_tmp/Standard_on_Ratio_Studies_Draft_12.pdf

IAAO (2003). "Standard on Automated Valuation Models (AVM's)." Assessment Journal(Fall 2003): 109-154.

API (2006) "Draft Exposure of the Residential Desktop Operating Framework", December 2006. The Australian Property Institute

APRA (2005) "Residential Valuation Practices by ADIs and LMIs". The Australian Prudential Regulation Authority. http://www.apra.gov.au/ADI/upload/Survey-Results-Residential-valuation-practices-by-ADIs-and-LMIs.pdf

Bahjat-Abbas (2006) "Automated Valuation: a Ratings Agency Perspective". http://www.rics.org/NR/rdonlyres/1BBE4F9B-32DC-4118-85CA-22AFF0F996AE/0/NadiaBahjatAbbas.pdf

Bradley, M. & Nuetzel, P.(2007) "On-Target or Off-Base? How to Ensure AVM Accuracy, Secondary Marketing Executive October 2007 .pp32,33,38

Downie, M. and Robson, G. (2007) "Automated Valuation Models: an international perspective", Research Report form The Council of Mortgage Lenders, CML London. http://www.cml.org.uk/cml/filegrab/1AutomatedValuationModelsHB.pdf?ref=5550

Fitch (2007) "Criteria for automated valuation models in the UK". 22 May 2007 www.fitchratings.com

Fortelney, A. & Reed, R. (2005) "The Increasing Use of Automated Valuation Models in the Australian Mortgage Market". Australian Property Journal, May 2005.

Ibrahim, M., Cheng and Eng. K (2005) "Automated valuation model: an application to the public housing resale market in Singapore ", Property Management, Vol 23 No, 5, 2005, pp357-373.

JP Morgan and Fujitsu (2006) "Australian Mortgage Industry Report Volume 4: Property Valuation Processes Reviewed - putting the meat back in the sausage". http://www.fujitsu.com/downloads/AU/mortgage_reports/vol_4.pdf

Lai, P. & Fischer, D., (2006), "Artificial Neural networks and Computer Assisted Mass Appraisal", 12th Annual Conference of the Pacific Rim Real Estate Society, January 22 to 25, 2006

Man, KF. And Ng, A. (2006) "An empirical analysis of Valuation accuracy and variation – Hong Kong evidence. The 12th Annual Conference of the Pacific Rim Real Estate Society, January 22 to 25, 2006

Mortgage Brokers Association (2006) Issue paper, Automated Valuation Models

Newell, G. and MacFarlane, J. (2006) "The Accuracy of Property forecasting in Australia", The 12th Annual Conference of the Pacific Rim Real Estate Society, January 22 to 25, 2006

OCC (2006) OCC Bulletin 2000-16. US Comptroller of the Currency, Administrator of National Banks http://www.ffiec.gov/ffiecinfobase/resources/retail/occ-bl2000-16_risk_model_validation.pdf

Rossini, P. and Kershaw, P., (2005) "Using a Hybrid Automated Valuation Model to Estimate Capital and Site Values", Eleventh Annual Pacific-Rim Real Estate Society Conference, Melbourne Australia, January 23rd - 27th January, 2005

Rossini, P. and Kershaw, P., (2006a) "Can a Hybrid Automated Valuation Model Outperform Individually Assessed Capital and Site Values", The 12th Annual Conference of the Pacific Rim Real Estate Society, January 22 to 25, 2006

Rossini, P. and Kershaw, P., (2006b) "Developing a weekly residential price index using a sales price appraisal ratio", The 12th Annual Conference of the Pacific Rim Real Estate Society, January 22 to 25, 2006

Rossini, P.A., (1997), "Artificial Neural Networks versus Multiple Regression in the Valuation of Residential Property", Australian Land Economics Review, November 1997 Vol 3 No 1

Rossini, P.A., (1998), "Improving the Results of Artificial Neural Network Models for Residential Valuation", 4th Pacific Rim Real Estate Society Conference, Perth 1998

Rossini, P.A., (1999), "Accuracy Issues for Automated and Artificial Intelligent Residential Valuation Systems", International Real Estate Society Conference, Kuala Lumpur, 26-30 January, 1999

Rossini, P.A.,(2000), "Using Artificial Intelligence for Real Estate Forecasting", Pacific-Rim Real Estate Society Conference, Sydney, 24-27 January, 2000

Skitmore, R., Irons, J and Armitage, L (2007) "Valuation Accuracy and Variation: A Meta Analysis", The 13th Annual Conference of the Pacific Rim Real Estate Society, January 21 to 24, 2007

Standard and Poor's (2004) "Guidelines for the use of automated valuation models for U. K. RMBS transactions". 20 February 2004. www.standardandpoors.com

Contact

Peter Rossini

Email peter.rossini@unisa.edu.au

University of South Australia, School of Commerce, City West Campus

GPO Box 2471, Adelaide, SA 5001, Australia

Appendix 1 – Location of study Area, Marion Local Government Area within Adelaide Metropolitan Area

Gulf St Vincent

CBD

Marion LGA

(c) Peter Rossini
Uni of South Australia

Appendix 2 – Land Value Response Surface – Regression Equation

Model Summary(b)

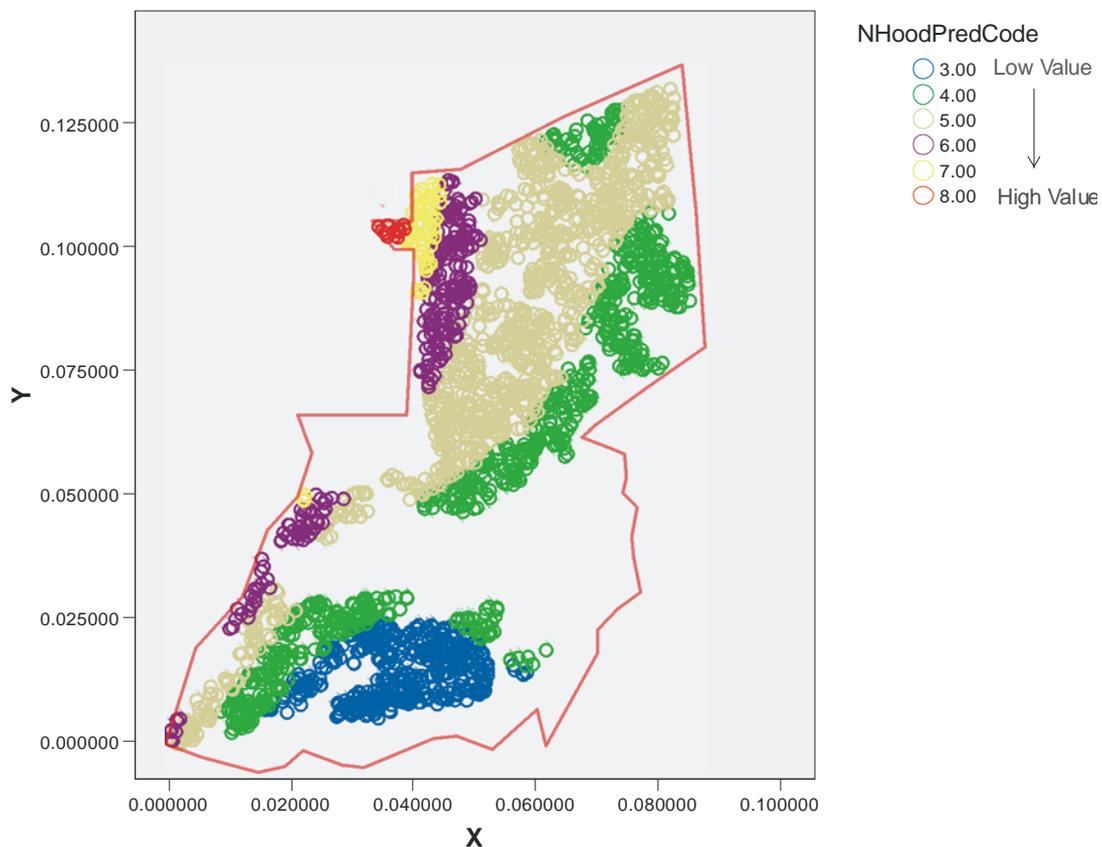| Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.472	.223	.220	.88154941

ANOVA(b)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	561.634	10	56.163	72.270	.000
	Residual	1958.366	2520	.777		
	Total	2520.000	2530			

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta	B	Std. Error
1	(Constant)	.579	.132		4.387	.000
	XCord	-116.285	12.202	-2.288	-9.530	.000
	YCord	41.959	11.421	1.626	3.674	.000
	XSqd	2261.895	342.015	4.198	6.613	.000
	YSqd	638.986	138.852	3.140	4.602	.000
	XbyY	-947.921	359.885	-2.753	-2.634	.008
	YSqdX	-12899.393	3543.197	-4.465	-3.641	.000
	XSqdYCub	521512.896	184719.579	1.584	2.823	.005
	XCubY	135518.533	69619.452	2.344	1.947	.052
	XQuart	-188398.827	61253.167	-2.326	-3.076	.002
	YQuart	-11017.310	5018.209	-.765	-2.195	.028

a Dependent Variable: Standardized Residual

Appendix 3 – Hedonic model including LVRS "LOCATION" variable

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
	.818	.670	.668	$42,474.51

ANOVA(j)

Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	4606390907269.8	9	511821211918.8	283.7	.000
Residual	2269538286252.1	1258	1804084488.2		
Total	6875929193521.9	1267			

Coefficients(a)

Model	Unstandardized Coefficients		Standardized Coefficients			Collinearity Statistics	
	B	Std. Error	Beta	t	Sig	Tolerance	VIF
(Constant)	70176.49	7436.89		9.43	.000		
Build Area (sqm)	1125.06	28.52	.687	39.44	.000	.86	1.15
Land Area (hectares)	238821.99	57658.95	.068	4.14	.000	.96	1.04
LOCATION (	64053.15	2546.48	.421	25.15	.000	.93	1.06
Wall (stone) Dummy	12765.16	5276.59	.040	2.41	.016	.96	1.03
Style (Archit) Dummy	83064.75	15718.58	.089	5.28	.000	.91	1.08
Style (SAHT) Dummy	-23293.07	5038.35	-.114	-4.62	.000	.43	2.31
Style (Convent) Dummy	-12253.40	4079.21	-.074	-3.00	.003	.43	2.31
Style (Spanish) Dummy	-30014.33	11655.37	-.044	-2.57	.010	.89	1.11
Style (Ranch) Dummy	-58282.14	24857.70	-.038	-2.34	.019	.97	1.02