Model-to-model validation: using validation models in real estate assessment

HERMANS, Luc, D. M.Sc., Ir. KATHMANN, Ruud M. and Ir. KUIJPER RO, Marco
Netherlands

Summary

This paper presents a framework to compare the results of an initial automated valuation model and the results of a validation automated valuation model. The framework is a reaction to developments in the automated valuation model market in the Netherlands. The framework consists of nine possible situations that can occur when the results are compared.

One of the most prominent advantages of the proposed framework is the fact that it allows ratio studies for objects without an observed sale price. Traditional ratio studies, and resulting model performance indicators, are all based on an observed sale price. Therefore, these indicators only give an indication of performance over a sample of the population. The framework proposed has the capacity to give an indication of how assessments compare to one another for the full population of properties.

Introduction

Trained data scientist, powerful computers and affordable statistical software are becoming ever more available. Due to these developments many industries are in a state of change. The real estate assessment market as well as the market for private sector valuation models is one of these industries. In this industry, one of these changes is the availability of a wide range of automated valuation models (AVMs). AVMs have been around for a few decades in the assessment industry, but because of new technologies and increased computer power these models are now not only more accessible but have also a much higher accuracy and reliability than for instance ten years ago.

Until now no consensus has been reached on what statistical approach results in the best AVM, but more important in the best tested AVM. This is partly due to the fact that the different statistical approaches address the impact of location differently.

In the Netherlands, AVM vendors have started to sell so called validation models. These validation models are automated valuation models which run alongside the already deployed AVM. This paper presents an approach on how to compare the results of such a validation model to the results of an initial AVM. First, a brief discussion on the observation of market value is presented. Secondly, traditional model performance indicators will be discussed briefly. Thereafter, the framework, containing nine possible situations will be elaborated. Finally, the paper will draw up concluding remarks and will present avenues of possible further research.

The terminology used in this paper is based on the use of the framework for real estate assessment purposes for real estate taxation. However, the framework can be applied between any two automated valuation models producing values and can therefore be used by any organization or individual estimating values for real estate.

2. Observation of market value

Over the course of history, market value has been defined in many ways. Currently one of the most prominent definitions is given by TEGoVA as part of the European Valuation Standards. This definition is in accordance with the definition of market value in the International Valuation Standards:

"Market value is the estimated amount for which a property should exchange on the date of valuation, in a transaction between a willing buyer and a willing seller, acting independently of each other after proper marketing, wherein the parties had each acted knowledgeably, prudently and without being under compulsion" (TEGOVA 2020)

The definition presented above indicates that a sale price is the preferred indicator for market value. However, this definition has some practical limitations. The most important limitation is the complexity of the definition. This complexity results in the fact that market value as such is unobservable (Clapp 1990). Almost no sale is made under the perfect conditions as assumed by the definition. Given that a property is the most valuable asset for most people there is almost always a form of compulsion. Furthermore, it is utopian thinking to assume that every sale is made on the basis of perfect knowledge of the market by both parties. Moreover, uneven negotiation skills can influence the sale price and deviate the observation from true market value. This explains why not every sale is a perfect approximation of the market value and you have to account for the noise around sales prices. Therefore, in the discussion on model performance indicators a lot of emphasis has been put on the operationalization of market value. The goal of this emphasis is to limit the influence of the error in the variable problem.

Traditionally, there are two approaches to measure market value. The first was introduced by Paglin and Fogarty (1972), they argue that market value is measured best through the use of sales price as an indicator. A decade later Kochin and Parks (1982) argued that using sales price as an indicator for market value introduces a bias towards regressive vertical inequity. To overcome this issue, they introduce the conceptualization of measuring market value through assessed value.

Ultimately, Clapp (1990) tries to end the debate. His argument is based on the notion that using the Paglin-Fogarty approach (using sales price) introduces a bias towards regressivity. Contrastingly, using the Kochin-Parks approach (using assessed value) introduces a bias towards progressivity. With this notion Clapp argues that neither one of the approaches is superior but both indicators contain information on the unobservable market value. Therefore, his argument is to use both sales price and assessed value as an indicator for market value.

A further shortcoming of the use of sale prices as the indicator for market value is the notion that only a limited number of properties within the population is sold each year. Traditional ratio studies, which will be more thoroughly explained in the next section, are all based on a known sale price. Therefore, the indicators only represent the model performance for a sample of the true population. The framework presented in this paper overcomes this problem by surpassing the need for an observed sale price.

3. Traditional model performance indicators

Model performance indicators are based on ratio studies. A ratio is a relative indicator between two observations. In the assessment industry, the most used ratio is the assessment to sales price ratio (ASR) (equation 1). These ratios form the basis for the subsequent calculations made to form model performance indicators. Traditionally, model performance for automated valuation models is being measured by three main concepts. Firstly, central tendency measures the appraisal level. This appraisal level gives an indication on how well the assessments align with the sale prices observed. Given that an assessor wants to be as close to the observed sale prices as possible with regards to the assessment level, the indicator for central tendency should be close to 1. There are several indicators for central tendency, however the median ratio is the preferred measure used in industry and literature. The choice for the median value is based on the notion that a median value is less prone to the influence of outliers as compared to a mean value.

Secondly, horizontal inequity indicates a situation where similar properties with the same market value are assessed differently. Horizontal inequity may be a result from many different underlying phenomena. Horizontal inequity in the real market (similar objects with the same market value are sold for different prices) can be caused by uneven negotiation skills or uneven knowledge about the real estate market. Horizontal inequity in assessment by the use of automated valuation models is mostly caused by inadequate data or incorrect parameters in the model. In essence, horizontal equity is measured by an indicator on variability between the assessment to sale price ratios within a similar group of properties. These groups can be thought of as value ranges, but more commonly they are calculated over different districts such as postal codes or within different groups of property types, such as row houses. There is a wide range of measures of variability. Presenting the advantages and disadvantages of each of these measures is beyond scope of this paper. However, the preferred measure in the assessment industry for property taxation purposes is the coefficient of dispersion (COD).

Thirdly, vertical inequity indicates a situation where properties with a different market value are assessed differently relative to their market value. Vertical inequity is regressive when lower value properties have a higher proportionate assessment value then higher value properties compared to their market value. Vertical inequity is progressive when higher value properties have a higher proportionate assessment value then lower value properties compared to their market value. By its nature vertical inequity is a more politically sensitive form of inequity. Regressive inequity can be interpreted as a tax rebate for wealthy taxpayers owning the high value properties. Due to this fact most of the academic debate has been focussed on identifying indicators for vertical inequity in real estate assessment.

The further academic debate on model performance indicators revolves around the functional form of the measuring equation. Early solutions start out with a linear functional form to measure vertical equity. Subsequent models implement non-linear approaches to finding vertical inequity. Currently the debate mostly focusses on the effect on the incorporation of location within the model performance indicators.

In the assessment industry for real estate taxation four main indicators are being used, based on the International Association of Assessing Officers (IAAO) Standard on Ratio Studies. These indicators are the median assessment to sales price ratio to indicate central tendency. The coefficient of dispersion (COD) to indicate horizontal inequity. For vertical equity two main measures are used. Firstly, the price related differential (PRD) and secondly, the price related bias (PRB).

The broader AVM industry often applies other measures such as the Mean Absolute Percentage Error (MAPE) and the Root Mean Squared Error (RMSE). The MAPE is an indicator of model performance that compares the forecast value (the estimated value) to the observed value (sale price). The MAPE presents the absolute difference between the forecast value and the observed value as a percentage deviating from the mean. The RMSE is also an indicator of model performance that uses the forecast value (the estimated value) and the observed value (sale price). However, the RMSE does not look at the absolute errors, but incorporates the squared errors. Often the MAPE is preferred over the RMSE for the sake of interpretability. Another measure for model performance is the hit rate, sometimes called an indicator range. This measure looks at the deviation between the estimated value to the observed value with a predefined bandwidth.

4. A reconceptualization of ratio studies to provide for model-to-model ratio studies

While each of the earlier mentioned measures for model performance has proven to be useful to assess the model performance of one AVM, the model performance indicators do not allow for a combinatory use of multiple models. The proposed framework in this paper facilitates the combination of two separate automated valuation models. A further advantage of the proposed framework is that the combinatory framework is not limited to properties with an observed sales price. Instead, each property in the population can be classified in one of the nine possible situations.

However, before this framework is possible, a new form of ratio should be calculated. The new ratio is the ratio between the estimated model value produced by the initial model compared to the estimated model value produced by the validation model. The equation of the “traditional” assessment to sale price ratio is given in equation 1. The equation of the “new” assessment to model value ratio is given in equation 2. The reconceptualization of the ratio can be defended by the arguments made in section 2. Though, before a model value can be taken as a decent indicator of market value, the performance of the validation model should be tested by the use of the traditional ratio studies and model performance indicators.

The calculation of these assessment to model value ratios allows for the calculation of traditional model performance indicators between the two separate automated valuation models.

Once a validation model passes the traditional ratio studies the assessment to model value ratios can be used to calculate the traditional measures for model performance between the initial automated valuation model and the validation automated valuation model. In this comparison the focus should be put on the measures for central tendency (the median assessment to model value ratio) and the measure used for the indication of variability (in this case the coefficient of dispersion as calculated by the use of the assessment to model value ratio). The steps described above are visualized in image 1. Please note that the populations differ. The first and second step (“traditional” ratio studies for both the initial and validation model) are limited to the properties with an observed sale price. The last step (the “traditional” ratio studies between outcomes of the initial AVM and the outcomes of the validation AVM can be calculated over the full population of properties.

$$ASR = \frac{Model\ Value}{Sales\ price} \tag{1}$$

$$AMR = \frac{Initial\ Model\ Value}{Validation\ Model\ value} \tag{2}$$

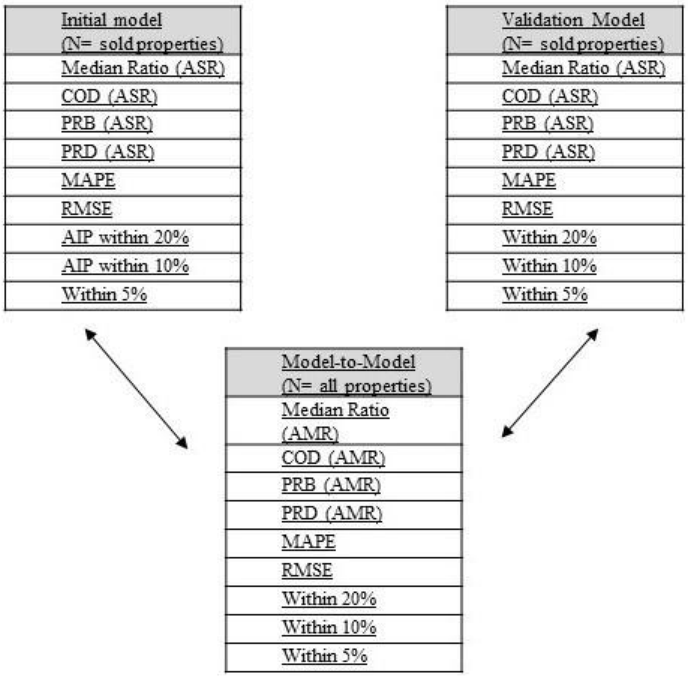

Image 1. Extension of "traditional" ratio studies.

5. Model-to-model validation framework

The availability of three indicators for market value (the observed sales price, the initial model value, and the validation model value) allows for a comparison of ratios at the object-level. The availability for comparable indicators per property allows for a more detailed look into the model performance with respect to the accuracy of the market value. Furthermore, spatial analysis can be undertaken to identify the correspondence between the indicators at the lowest (geographic) level - being that of the property itself. This paper identifies nine possible scenarios however, it is important to note that the simulations are exclusive, an object can only be categorized in one situation. Further, some categories can also be sub-categories, an approach to single out specific situations. The proposed situations are presented in Table 1. First the situation without the availability of an observed sales price is presented, with the subsequent situations with an available observed sales price determined. It is important to note that the situations only give an indication of the difference between both models and that the real market value might differ significantly.

Situation	Description	Initial model value	Validation model value	Observed sale price	AMR	ASR
1.	the validation model result is higher than the initial model value result and there is no observed sale price	200.000	240.000	X	0,83	X
2.	the validation model result is lower than the initial model value result and there is no observed sale price	200.000	160.000	X	1,25	X
3.	the validation model result is equal the initial model value result and there is no observed sale price	200.000	200.000	X	1	X
4.	the validation model result is almost equal the initial model value result and there is no observed sale price	200.000	195.000	X	1,03	X
5.	the validation model result is higher than the initial model value result and the observed sale price is exactly equal to the initial model value	200.000	240.000	200.000	0,83	1
6.	The validation model result is equal to the observed sale price and the initial model value is significantly lower than the other two indicators.	200.000	220.000	220.000	0,91	0,91
7.	The validation model result is equal to the observed sale price and the initial model value is significantly higher than the other two indicators.	200.000	180.000	180.000	1,11	1,11

8.	The validation model result and the initial model value are both close to the observed sale price.	200.000	195.000	205.000	1,03	0,98
9.	The validation model result, the initial model result and the observed sale price are all exactly equal	200.000	200.000	200.000	1	1

Table 1. Nine situations of the proposed framework.

The first scenario (1) presents the situation where the validation model result is higher than the initial model value result and there is no observed sale price. which occurs when there is no available observed sales price and the assessment to model value ratio (AMR) crosses the lower value of an arbitrary bandwidth. In this paper the bandwidth for undervaluation is set at an AMR of less than 0.85. Inversely, the second situation (2) relates to the situation where the validation model result is lower than the initial model value result and there is no observed sale price, which occurs when there is no available observed sales price and the assessment to model value ratio crosses the upper value of the arbitrary bandwidth. In this paper this upper value is taken to be 1.15., Both the upper and lower boundary of the arbitrary bandwidth in this paper are drawn from the 15% performance indicator for the COD as proposed by the IAAO. A third possible situation (3) pertains to situation where the validation model result is equal the initial model value result and there is no observed sale price which transpires when there is no available observed sales price, however the result of the validation model is exactly equal to the value of the initial model, therefore the AMR is equal to 1. The last possible situation (4) without the availability of an observed sales price is the situation where the validation model result is almost equal the initial model value result and there is no observed sale price. This occurs when the initial model value and the validation model value differ, but this deviation is within the boundaries of the arbitrary bandwidth.

Several other viable scenarios can be determined which account for an available observed sales price and therefore an available assessment to sales price ratio. The first possible situation (5) is the situation where the validation model result is higher than the initial model value result and the observed sale price is exactly equal to the initial model value which arises when the validation model value crosses the upper or lower value of the arbitrary bandwidth and the assessment to sales price ratio is exactly equal to 1. This situation indicates a suspicion of sales price chasing. The second possible situation (6) would be the situation where the validation model result is equal to the observed sale price and the initial model value is significantly lower than the other two indicators. This arises when both the AMR and the ASR cross the lower value of the arbitrary bandwidth. The third possible situation (7) represents the situation where the validation model result is equal to the observed sale price and the initial model value is significantly higher than the other two indicators. which tends to occur when both the AMR and the ASR cross the upper value of the arbitrary bandwidth. Sold Compromise represents the next situation (8) is the situation where the validation model result and the initial model value are both close to the observed sale price. Which arises when both the ASR and the AMR are within the boundaries of the arbitrary bandwidth. The last possible situation (9) is the situation where the validation model result, the initial model result and the observed sale price are all exactly equal which occurs when the observed sales price, the initial model value and the validation model value are exactly equal.

Once the properties are sorted in the possible situations analysis can be done on the outcome. A frequency table can identify the relative occurrence of the different situations. This might give an indication on whether or not most properties are listed in an “acceptable” situation. Furthermore, the properties can be mapped using a GIS. This allows to find pockets of situations and might give an indication of local model performance. In any way, the classification into the different situations allows for the concentrated use of available resources.

6. Conclusions and further research

The presented framework allows for the comparison between outcomes of an initial AVM with the outcomes of a validation AVM. By sorting the outcomes based on the presented assessment to sales price ratio and the assessment to model value ratio assessors can focus their time and budget on the properties that fall within one of the undesirable situations. One of the most important advantages of the framework is that it allows for a comparison of indicators for the full population of properties whereas the “traditional” ratio studies only allow for model performance indicator calculations of properties with an observed sale price. If this framework is repeated results can be checked for multiple models. However, by doing this we always have to define a base model (the initial model).

A more fundamental advantage of using this framework is that we can use two separate AVMs without having to explain both models thoroughly. For instance, the initial model can be based on a general linear model with the advantages of explainability. Whereas the used validation model can be based on Artificial Intelligence, with all drawbacks in explainability. By subsequently applying the framework we can still define how the AI-based validation model output compares to the output of the explainable general linear initial model.

However, the framework has some limitations. The most prominent limitation is the fact that the framework can only be used for comparing two automated valuation models. With current developments in mind, it can be hypothesized that assessing offices will deploy ever more automated valuation models. Therefore, research should focus on finding a way to combine multiple automated valuation models where none of the deployed automated valuation models has to be defined as the baseline (initial) model. Further research should focus on developing a way to interpret the results of multiple models. Additionally, the framework is now focused on the Dutch situation. Further research should focus on applying the framework in a different country context.

7. Literature

Clapp, John M. 1990. "A New Test for Equitable Real Estate Tax Assessment." *The Journal of Real Estate Finance and Economics* 3, no. 3: 233-249.

Kochin, Levis A. and Richard W. Parks. 1982. "Vertical Equity in Real Estate Assessment: A Fair Appraisal." *Economic Inquiry* 20, no. 4 (October): 511-532.

Paglin, Morton and Michael Fogarty. 1972. "Equity and the Property Tax: A New Conceptual Focus." *National Tax Journal* 25, no. 4 (December): 557-565.

TEGoVA. 2020. *European Valuation Standards,* ninth edition

8. Contact information

e-mail: l.hermans@waarderingskamer.nl

9. Author information

Luc Hermans – is GIS-specialist at the Netherlands Council for Real Estate Assessment

Ruud Kathmann – is member of the management team at the Netherlands Council for Real Estate Assessment

Marco Kuijper – is strategy developer at the Netherlands Council for Real Estate Assessment