An Experiment within Valuation Practice

Towards Automating Commercial Real Estate Valuations?

In the new decade a lot is expected to change in the way real estate is valued. DNB (2019) recently published a report where an increasing number of stakeholders express concerns about the quality and independence of the current valuation practice. Advances made in terms of data quality, computational power and econometric modeling provide opportunities to improve estimations based on historical evidence. But experts also know that a lot of progress still has to be made before full automation can be achieved. Furthermore, with the decision of the ECB to ban fully automated valuations for real estate mortgages (Tweede Kamer, 2020), the need arises for hybrid approaches where man and machines work in conjunction, each capitalizing their own skills. This article investigates the implementation of data-driven methodologies in the current (commercial) residential valuation practices from a valuer's perspective and discusses findings from an experiment where model estimates are compared to manual valuations to analyze when and why the two might differ.

Bas Hilgers, Jurre Brantsma and Jacques Boeve

1 Introduction

In the absence of continuously traded, deep and securitized markets, real estate valuations perform a vital role in the property market by acting as a surrogate for transaction prices. The main method still used today to estimate valuation input parameters is a direct comparison approach where the value of a property is assessed based on transaction prices of comparable properties. Since properties are never truly comparable these results need to be adjusted for differences in their characteristics. This is often performed manually and is where the 'art' of the profession comes into play. However, this is also the part that is most sensitive to human error and (unintentional) subjectiveness.

There are thus strong arguments in support of new approaches with greater objectivity and reproducibility. In the Dutch owner-occupied residential sector, a large part of the valuations is appraised by so-called Automated Valuation Models (AVM). These statistical models can find patterns within large amounts of data to make predictions and/or provide inferences. Many studies (e.g. Horváth et al., 2016) even show that these models have the potential to outperform manual valuations. For commercial real estate, that is property intended to generate a profit either from capital gain or rental income, there is however yet little evidence on the applicability of such models. This can mainly be traced to the lack of quality data available and heterogeneity between properties which in general result in low fit and out-of-sample performance.

The main challenge the sector is currently facing is to develop and implement new methodologies that get the most out of the relatively little and noisy data available. Additional research is needed to better understand the opportunities, and perhaps more importantly, the limitations of such models as a single 'wrong' estimate could bear significant risk for the parties involved. This article therefore aims to investigate the feasibility of AVM for commercial residential valuations and how this might work in valuation practice. We cover the following research questions:

- How to develop an AVM for commercial real estate from a valuer's perspective?
- What data is needed for a market value estimate of commercial residential real estate?
- Which type of statistical models compare most to the thought process of the valuers?
- Do complex models provide enough gain to compensate for the loss of interpretability?
- Do we observe patterns in the results where manual reported values differ significantly from model estimates?

The outline of the article is as follows: First we introduce general valuation theory and discuss how AVM might fit in this framework for commercial residential real estate. Second, we cover data requirements of such models and describe the data used in this study. Third, we list methodologies applied and discuss the experimental setup of our research. Fourth, we report key findings and conclude with a discussion about the applicability of AVM in current valuation practice.

2 Property Pricing

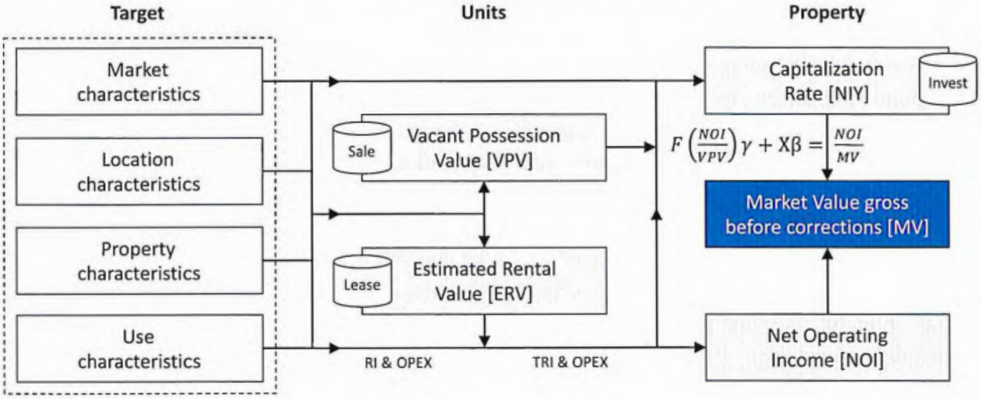
The valuation practice of commercial real estate is mostly concerned with estimating the market value which is defined as the most-likely price a property would transact for in an arm's length transaction at a given reference date. For this task, valuers often adopt a discounted cash flow method where all cash flows over the investment period are discounted back at a rate that reflects a market conform required rate of return. However, this approach has many moving parts that require subjective decisions which in turn can significantly influence the estimate. Therefore, valuers also often adopt a 'short-cut' capitalization method where the first year's Net Operating Income (NOI) is divided by a capitalization (cap) rate that is derived from realized comparable transactions. The final reported value is the deliberate outcome of the reconciliation of both with weights dependent on the certainty of assumptions underlying each method.

Automated Valuations

When discussing Automated Valuation Models (AVM) for commercial real estate, the general goal is to estimate this market value while improving on objectivity, reproducibility and efficiency through patterns in data. Depending on the purpose of the valuation, the interpretability of the result might be just as important as the accuracy of estimates. In this process, the estimation and communication of key market parameters that lead to the final market value could therefore provide essential insights to both valuer and client about the property's performance. There are many ways to model these, each with their advantages and disadvantages. For example, Geltner & De Neufville (2018) apply Monte Carlo simulations to generate Discounted Cash Flows scenarios which are less reliant on data availability but sensitive to chosen input parameter ranges, while Kok et al. (2020) utilize 'big' data and machine learning to optimize predictions but lose most interpretability in its complexity.

In this article, we focus on the estimation of the market parameters that are needed to automate a market capitalization approach (Figure 1). In the case of residential real estate, the Net Operating Income (NOI) is expressed as the net Theoretical Rental Income (TRI) and the Cap Rate is estimated via a set of characteristics of which the ratio between the Vacant Possession Value (VPV) and the NOI was found to be economically most significant. This approach as a base for AVM has the advantage that it aligns well with current valuation practice, including all the familiar steps, and that estimates can be derived from actual market evidence. The limitations are however that estimations could become unreliable when few comparables are available and that auxiliary models are required on multiple datasets to reach the final goal of a market value estimate. Nevertheless, cases where model results become unreliable can be identified and due to the alignment with valuation practice, valuers could add adjustments where necessary. Therefore, it might be more appropriate to define this as a Computer Assisted Modelling Approach (CAMA)

FIGURE 1 ▸ APPLIED AVM STEPS FOR COMMERCIAL RESIDENTIAL REAL ESTATE



or hybrid AVM rather than fully AVM. The remainder of the article focusses on the estimation of the key market parameters by which this automated capitalization approach can be constructed.

3 Data and Descriptive Statistics

Transactional Data

For an AVM to be able to generate estimates of the market parameters, relevant transactional data is required on realised prices including information on relevant characteristics. As shown in the previous section, for the valuation of commercial residential real estate we need estimates of the vacant possession value via sale transactions, market rent via lease transactions and cap rate via investment transactions. The sale and lease transactions in this study are obtained from the largest association of real estate agents and appraiser (NVM) and cover over 75 percent of all residential transactions in the Netherlands. The investment transactions are obtained from Cushman & Wakefield whose experts collect, enrich and validated transactions on a daily basis from a variety of sources. Table 1 provides an overview of variables included in the models.

TABLE 1 ▸ OVERVIEW MARKET TRANSACTIONS AND VARIABLES PER MODEL

	Sale Transactions	Lease Transactions	Investment Transactions
Dependent variable	Price per m^2	Rent per m^2	Net Initial Yield
Market			
Transaction period	2015 – 2020	2015 – 2020	2015 – 2020
Submarket	Neighborhoods	Neighborhoods	Expert Delineated
Property			
Subtype	X	X	X
Size	X	X	X
Age	X	X	X
Energy label	X	X	X
Use			
TRI / VPV	-	-	X
Observations	1,261,595	221,690	4,211

TABLE 2 ▸ DESCRIPTIVE STATISTICS REPORTED MARKET PARAMETERS MANUAL VALUATIONS

	Vacant Possession Value (per m²)	Estimated Rental Value (per month)	Capitalization Rate (net TRI)
Observations	5,166	5,166	4,906
Mean	€ 2,910	€ 998	4.9 %
Std. Dev.	€ 1,031	€ 265	0.8 %
Min	€ 1,145	€ 290	2.1 %
Max	€ 8,171	€ 2,600	9.7 %

Valuation Data

The experimental setup of this article is to compare AVM results to reported values from manual valuations. So, besides transaction data to fit the models, we've collected a sample of 4,906 residential valuations executed in the year 2020 by the valuation department of Cushman & Wakefield, the Netherlands. The sample contains a wide variety of residential valuations spread over the whole Dutch market and made for different valuation purposes and types of clients including housing association and private sector. These reported values are manually calculated based on evidence from comparables and adjusted by experts to match the characteristics of the property to be valued. Table 2 provides descriptive statistics of the market parameter estimates in the valuation sample.

4 Methodologies and Experimental Setup

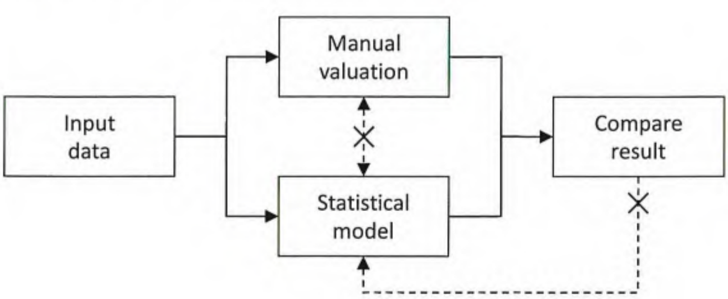
For the estimation of key commercial real estate market parameters, we investigate six different methodologies (referred to as models) of which results are compared to actual valuations. In the selection of models, we considered differences in complexity, alignment with valuation practice and findings from previous literature (e.g. Borst, 2007, McCluskey et al., 2013 & Zurada et al., 2011). The models applied are a parametric Hedonic Price Model (HPM) with time- and market fixed effects, a semi-parametric Geographically- and Temporal Weighted Regression Model (GTWR) and a non-parametric Random Forest (RF) and Gradient Boosting (XGB) model[1]. In addition, we apply a Simple Average Comparables Model (COMP) and a novice Corrected Comparables Model (CCM) which further extends the COMP approach by correcting for differences in characteristics by the coefficients of the HPM.

Rules Experiment

To make the experiment as objectively as possible we set some rules all models follow (Figure 2):

- The experiment covers the entire Dutch residential market. However, to fit the models, smaller samples are constructed based on characteristics of the property to be valued. That is, only transactions within the same (COROP) region are used and for the valuation of single-family properties, multi-family properties are

FIGURE 2 ▸ EXPERIMENTAL SETUP AND EVALUATION PROCESS



excluded and vice versa[2]. The Cap Rate model on the other hand is applied to the whole market due to data limitations and the fact that capital markets are more integration than the asset market.

- All models use the same data sample and transformations to fit the model. These data are also the same the valuer had at the time of valuation.
- No data is used that occurred after the valuation date as valuers also did not had this information and is in line with the definition of market value. New models are fitted at the beginning of each month.
- No information of the valuation is included in the models and no model information is used by the valuations.

Quantifying Performance

While there do not yet exist uniform criteria for the evaluation of AVM, the IAAO (2013) advocate a Median Average Percentage Error (MdAPE) and a Coefficient of Dispersion (COD) measure as standards. Nevertheless, since the distribution of absolute errors are right skewed, the median is more optimistic and less sensitive to outliers than the mean. In practice however the variation of errors is just, if not more, important than the average errors as valuers are responsible for the reported value and large outliers can have serious consequences. Hence, we instead apply a Mean Average Percentage Error (MAPE) and standard deviation of errors to quantify the differences between approaches (Steurer & Hill, 2020). Important to highlight is that in our setup the performance of results does not refer to out-of-sample accuracy of models on realised transactions, but rather to what extent the results align with the reported values by expert valuers.

5 Results

The research questions of this article are answered by comparing automated model predictions of market parameters to what the valuers have reported at the end of the manual valuation process (Figure 2). In particular: Do more complex models provide better performance and if so, is it worth the cost of complexity? Does the number of relevant transactions significantly influence model performance? And do we observe other relevant patterns within the results where performance differs significantly?

Model Complexity

Table 3 shows the overall percentage differences between model results and reported values from manual valuations and the overall variance of the differences. We can observe that more complex models (RF, XGB) perform better than simpler ones (COMPS, CCM, HPM, GTWR). This is as expected as these types of models are trained to optimize predictions and consider relationships in higher dimensions (Mullainathan & Spiess, 2017). Nevertheless, the results also show that the simpler models do not perform much worse, given the same data, while maintaining a better interpretability. That is, we can analyse how a prediction was constructed, either directly through comparables or through marginal

TABLE 3 ▸ OVERALL MAPE PER MODEL PER MARKET PARAMETER WITH STANDARD DEVIATION BETWEEN BRACKETS

	Vacant Possession Value	Estimated Rental Value	Capitalization Rate
COMPS	9.7% (8.6%)	9.6% (9.2%)	10.5% (6.7%)
CCM	9.9% (8.3%)	10.4% (8.9%)	11.2% (6.2%)
HPM	9.6% (8.2%)	10.8% (9.8%)	10.9% (6.8%)
GTWR	10.8% (8.9%)	11.7% (9.9%)	11.0% (6.4%)
RF	9.0% (8.3%)	**8.8%** (9.0%)	9.4% **(5.9%)**
XGB	**8.5% (7.1%)**	9.7% **(8.5%)**	**8.8%** (6.1%)

price effects of characteristics. Depending on the purpose of the valuation, the risk a single 'wrong' prediction might bring or when seeking a synergistic relationship between 'man and machine', this interpretability might be more valuable than the percentage points increase in overall prediction accuracy.

We furthermore can observe that in contrast to the findings of McCluskey et al. (2013), we do not find the GTWR to outperform other models. Although these spatial-temporal econometric models solve some limitations parametric models have, such as constant parameters and MAUP effects (Helbich, 2013), many (e.g. LeSage, 2004 & Borst, 2017) argue that these types of models are great for analytical purposes but not optimal for predictions. Nevertheless, their semi-parametric nature do provide compelling features for real estate value estimations and deserve further investigation.

Transaction Density

Table 4 shows the same model performances but split by transaction density based on the amount of transactions available within proximity. We find significant evidence that in higher transaction density areas the performance of the models is better than in areas with low transaction density. The transaction noise in these regions is most likely lower due to the information availability. Nevertheless, for lease and investment transactions we do not observe clear patters, which may relate to the fact that these models tend to have more difficulties to fit the data regardless of transaction density. Interestingly, we can also observe that complex models perform better than simpler models when transaction density is high, but worse in areas where transaction are scarce. In these markets, models that make stronger assumptions seem to align better with the process of a valuer and is also in line with the bias-variance trade-off theorem on model complexity (Bishop, 2018). Most likely, these models overfit subsets of the sample which is a risk when utilizing these types of models in production.

Unobserved Heterogeneity

Besides the density of transactions available another key factor for accurate estimations is the relevance of the transactions for the property to be valued. For example, if a valuer wants to determine the value of a house with a large garden one requires enough similar property transactions in-sample. When moving to commercial real estate these differences in characteristics (heterogeneity) become more abundant of which many are unobservable (Francke & Minne, 2019). In our study we find significant differences between the performance of housing association and private sector valuations which is likely the result of differences in building quality not observable in our datasets. The properties of housing association are in general of lower quality than the data in our datasets hence, all models tend to overestimate these observations as shown in Table 5. Models that implicitly include this unobserved heterogeneity perform better and might hold more potential in current practice where (commercial) real estate data is still lacking in many ways.

TABLE 4 ▸ MAPE PER MODEL PER MARKET PARAMETER SPLIT BY TRANSACTION DENSITY

	Vacant Possession Value		Estimated Rental Value		Capitalization Rate	
Transaction Density	High	Low	High	Low	High	Low
COMPS	**9.2%**	14.3%	**9.8%**	13.9%	7.1%	6.6%
CCM	**10.9%**	15.1%	10.4%	**9.9%**	**8.5%**	**7.9%**
HPM	**9.1%**	13.0%	10.8%	**6.3%**	**7.5%**	7.6%
GTWR	**10.0%**	10.6%	**9.6%**	11.6%	**7.1%**	7.3%
RF	**7.3%**	12.1%	8.8%	**6.2%**	5.9%	**6.0%**
XGB	**8.9%**	16.3%	9,3%	**6.1%**	**4.4%**	6.8%

TABLE 5 ▸ MAPE PER MODEL PER MARKET PARAMETER SPLIT BY SOCIAL AND PRIVATE PROPERTIES

	Vacant Possession Value		Estimated Rental Value		Capitalization Rate	
Segment	Social	Private	Social	Private	Social	Private
COMPS	10.0%	**9.6%**	13.5%	**8.3%**	14.4%	**7.0%**
CCM	10.8%	**9.8%**	14.2%	**9.1%**	16.1%	**7.8%**
HPM	11.3%	**9.0%**	15.4%	**9.2%**	15.2%	**7.0%**
GTWR	14.4%	**9.6%**	18.1%	**9.8%**	15.4%	**7.1%**
RF	9.5%	**8.9%**	13.8%	**7.2%**	13.7%	**5.5%**
XGB	9.4%	**8.5%**	13.6%	**8.3%**	13.1%	**4.9%**

6 Conclusion and Discussion

The aim of this article was to investigate the potential of Automated Valuation Models (AVMs) for the estimation of market parameters of commercial residential real estate. We've set up an experiment where we compared the outcomes of 4,906 residential valuations made by expert valuers of Cushman & Wakefield in 2020 to the estimates of different (types of) 'data-driven' methodologies. We covered what information would at least be needed to automate the valuation process and investigated significant patterns that arose between the two.

The evaluation of the experiment shows interesting results to our research questions. First, we observe that more complex models, given the same data, perform better overall than simpler ones. Nevertheless, most interpretability is lost in the process and depending on the purpose of the valuation the gain in accuracy might not be worth the costs. Second, we observe that all models perform better in markets where many relevant transactions are available, but also that simpler models provide more robust estimations than complex ones. Third, we observe that while the methods applied are important, quality and availability of relevant market data are paramount. When data is not representative for the property to be valued excessive residuals emerge in all models. Finally, in contrast to the findings of DNB (2019) we do not observe a systematic over-valuation in the results but rather an under-valuation compared to model estimates. Most likely, valuers are more conservative towards previous reported values (smoothing and lagging) than the current market (McAllister et al., 2003). Further research would benefit from investigating whether the findings hold for more extensive model specifications and different property segments.

So, we have seen that with accurate data even the simplest models can provide insightful results while the opposite does not hold true. We strongly believe that it is the synergistic relationship between man and machine that holds the future of the valuation profession. But much research and transparency in practical application still need to be established before these new data-driven approaches become the new valuation standard.

About the Authors

Bas Hilgers MSc is a PhD-candidate at the Vrije Universiteit Amsterdam on the topic of commercial real estate valuations. He also works at the Valuation & Advisory department of Cushman & Wakefield where he investigates the implementation of data-driven methodologies in valuation practice.

Jurre Brantsma MSRE MRICS RT is partner at Cushman & Wakefield and co-heading the residential valuation team. He mainly focuses on large institutional clients that invest in residential properties and is involved in increasing innovations within the Valuation & Advisory department of Cushman & Wakefield.

Jacques Boeve Msc MRE FRICS RT was until March 2021 head of Valuation & Advisory and international partner at Cushman & Wakefield Netherlands. Nowadays he works as an independent real estate advisor. In addition, he regularly gives lectures at universities and courses to real estate specialists on future-proof valuation and the practical application of automated valuation modeling.

For more information about the applied models, additional discussion or other questions, please contact: bas.hilgers@cushwake.com

FOOTNOTES

1 Considerable effort has been made to optimize out-of-sample performance of each model given the data available. Extensive details about the methodologies and model specification are however out of the scope of this article.

2 Further research could benefit from investigating whether this specification choice results in loss of valuable information.

3 For commercial real estate, not only the distance and time might be important for the weighting matrix, but also physical characteristics (Hilgers, 2018).

REFERENCES

- Bishop, C. M. (2006). *Pattern recognition and machine learning* (pp. 147-152). Springer.
- Borst, R. (2007). *Discovering and applying location influence patterns in the mass valuation of domestic real property*. (Doctoral dissertation, University of Ulster).
- Borst, R. (2015). *Improving mass appraisal valuation models using spatio-temporal methods*. International Property Tax Institute.
- De Nederlandsche Bank. (2019). De kwaliteit en onafhankelijkheid van woningtaxaties. *Occasionele Studies*. 117(1).
- Francke, M. & van de Minne, A. (2020). Modeling unobserved heterogeneity in hedonic price models. *Real Estate Economics*.
- Geltner, D. M., & De Neufville, R. (2018). *Flexibility and real estate valuation under uncertainty: A practical guide for developers*. John Wiley & Sons.
- Hager, D. P., & Lord, D. J. (1985). The property market, property valuations and property performance measurement. *Journal of the Institute of Actuaries*, 112(1), 19-60.
- Helbich, M., Brunauer, W., Vaz. E. & Nijkamp, P. (2013). Spatial heterogeneity in hedonic house price models: The case of Austria. *Tinbergen Institute Discussion Paper*.
- Hilgers, B. A. J. (2018). *Automated valuation models for commercial real estate in the Netherlands: Traditional regression versus machine learning techniques*. (Master's thesis, Eindhoven University of Technology).
- Horváth, A., Imre B. & Sápi, Z. (2016). The international practice of statistical property valuation Methods and the possibilities of introducing automated valuation models in Hungary. *Financial and Economic Review*. 15(4), 45-64.
- International Association of Assessing Officers. (2013). *Standard on mass appraisal of real property*.
- Kok, N., Koponen, E. L., & Martínez-Barbosa, C. A. (2017). Big data in real estate? From manual appraisal to automated valuation. *The Journal of Portfolio Management*, 43(6), 202-211.

- LeSage J. P. (2004). A family of geographically weighted regression models. In: Anselin L., Florax R.J.G.M. & Rey S. J. (Ed.) Advances in spatial econometrics. *Advances in spatial science*. Springer.
- McCluskey, W. J., McCord, M., Davis, P. T., Haran, M., & McIlhatton, D. (2013). Prediction accuracy in mass appraisal: A comparison of modern approaches. *Journal of Property Research, 30*(4), 239-265.
- McAllister, P., Baum, A., Crosby, N., Gallimore, P., & Gray, A. (2003). Appraiser behaviour and appraisal smoothing: Some qualitative and quantitative evidence. *Journal of Property Research, 20*(3), 261-280.
- Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives, 31*(2), 87-106.
- Steurer, M. & Hill, R. (2020). Metrics for measuring the performance of machine learning prediction models: An application to the housing market. *Unpublished work*.
- Tweede Kamer der Staten-Generaal. (2020). Aanhangsel van de handelingen: Huizenbezitter honderden euro's extra kwijt aan taxatie.
- Zurada, J., Levitan, A., & Guan, J. (2011). A comparison of regression and artificial intelligence methods in a mass appraisal context. *Journal of Real Estate Research, 33*(3), 349-387.