Automated valuation models for commercial real estate in the Netherlands traditional regression versus machine learning techniques

Hilgers, B.A.J.

Award date:
2018

Automated Valuation Models for Commercial Real Estate in the Netherlands

Traditional Regression versus Machine Learning Techniques

Master Thesis

Student Info
B.A.J. (Bas) Hilgers MSc
0831665
Bashilgers@hotmail.com

Graduate Program
Eindhoven University of Technology
Architecture, Building and Planning
Real Estate Management and Development
August 2018

Supervisors
Eindhoven University of Technology
prof.dr. T.A. (Theo) Arentze
&
dr. I.V. (Ioulia) Ossokina

Cushman & Wakefield
drs. J. (Jacques) Boeve MRE FRICS RT RICS

Acknowledgement

I would first like to thank my main academic supervisor *prof.dr. Theo Arentze* – full Professor of Real Estate and Urban Systems and chair of the Real Estate Management and Development group in the department of the Built Environment of the Eindhoven University of Technology. His great knowledge about Hedonic modelling and academic research in general have provided me the necessary information to steer my research in the right direction and improved my results. Despite his busy schedule, he never failed to provide extensive feedback to my questions whenever I needed it. In addition, I want to extent this gratitude to *drs. Jacques Boeve MRE FRICS RT RICS* – international Partner at Cushman & Wakefield and responsible for the business line Valuation & Advisory in the Netherlands. His extensive knowledge about the valuation practice and vision for the future of the profession have helped me a lot in specifying my research with our monthly brainstorm sessions providing me with new inspiration. Third, I want to thank my second academic supervisor, *dr. Ioulia Ossokina* – Assistant Professor and Chair of Housing Research and Modelling, for her insightfull comments. Last, I want to thank my other colleques at Cushman & Wakefield for providing the additional knowledge needed to write this thesis.

Amsterdam, August 2018.

Abstract

Purpose – The purpose of this thesis is to investigate the potential of Automated Valuation Models (AVMs) for estimating the market value of individual Commercial Real Estate properties in the Netherlands.

Design/methodology/approach – With a unique complete dataset of 979 office property transaction from 2010 through 2018 obtained from Cushman & Wakefield that is enriched with information about building, location, lease and market factors, we study several methodologies that previous literature has shown to offer excellent explainability, reliability and predictability. These are the traditional Hedonic Price Model and the newer tree-based Machine Learning algorithms; Random Forest and (Extreme) Gradient Boosting. In addition, we introduce a new methodology by the name of Comparable Weighted Regression (CWR) that extends the Hedonic Price Model to allow for spatial- and temporal dependencies by weighting observations based on the degree of comparability to the subject property. Through a variety of error measures and cross-validation techniques we investigate which methodology provides not only the lowest Mean Absolute Percentage Error (MAPE), but also minimize the number of large errors as these are especially unwanted in practice.

Findings – The first hypothesis of this thesis addresses the importance of lease related factors in the prediction of the market value for Commercial Real Estate properties. We find through Leave-One Out Cross-Validation that the MAPE of the Baseline Hedonic regression model improves from 45.8 to 22.8 percent when lease factors are included. The second hypothesis investigates whether the prediction accuracy of the Hedonic improves when we incorporate spatial-temporal dependencies. We find that the MAPE decreases to 19.3 percent which is best among methodologies while the number of large errors are minimized. The third and last hypothesis studies whether a well-defined Hedonic regression model can outperform newer Machine Learning algorithms that have increased in popularity in recent years in both academia and practice. We find that the tuned (Extreme) Gradient Boosting outperforms the Random Forest algorithm with a MAPE of 21.6 percent, but which still performs worse than both traditional Hedonic and Comparable Weighted Regression.

Practical implications – As we find strong evidence that AVMs applied to the Commercial Real Estate sector benefit from including lease related factors into their model specification, such data should be gathered more extensively. Furthermore, data-driven Machine Learning techniques seem to have difficulties finding the underlying patterns in the data due to the relatively few transactions that take place in this sector. And as the estimates of this 'black-box' techniques are also more difficult to communicate and defend, traditional regression methodologies seem to fit the purpose of this thesis better than Machine Learning techniques. But with an optimal MAPE of 19.3 percent against the average of 10 percent of manual appraisals, the methodology and data still have a far way to go before practical application. AVMs that combine best from both worlds, such as the CWR, are likely to be the key to success.

Originality/value – The discussion whether Automated Valuations Models will disrupt the market or let it evolve is more relevant than ever. Surprisingly, literature that investigates the potential of such models for the Commercial Real Estate sector are practically non-existing. This thesis is a first study to compare traditional Hedonic Regression with Machine Learning techniques in this sector. In addition, we propose a new methodological framework, the CWR, that aims to counter some of the issues of traditional regression for the task at hand.

Acronym List

AI	Artificial Intelligence
ANN	Artificial Neural Network
API	Application Programming Interface
AVM	Automated Valuation Model
BAG	Cadaster in the Netherlands (Basis Administratie Gebouwen in Dutch)
CAMA	Computer Assisted Mass Appraisal
CART	Classification and Regression Tree
CBS	Central Bureau for Statistics (Centraal Bureau voor Statistiek in Dutch)
C&W	Cushman & Wakefield
CSM	Comparable Sales Model
CWR	Comparable Weighted Regression
GBM	Gradient Boosted Model
GIS	Geographic Information Systems
GIY	Gross Initial Yield
GWR	Geographically Weighted Regression
HTM	Hierarchic Trend Model
HPI	Hedonic Price Index
HPM	Hedonic Price Model
IAAO	International Association of Assessing Officers
kNN	k-Nearest Neighbours
OLS	Ordinary Least Squares
PDOK	Geoportal Public Maps (Publieke Dienstverlening Op de Kaart)
PropTech	Property + Technology
RF	Random Forest
SAR	Spatial Autoregressive Model
SER	Spatial Error
STAR	Spatial Temporal Autoregressive Model
TWR	Time Weighted Regression
VIF	Variance Inflation Factor
WLS	Weighted Least Squares (regression)
XGBoost	Extreme Gradient Boosted Model

Table of Contents

1. Introduction

In many applications, property valuation plays an important role: local governments need periodic valuations for tax purposes, main financial institutions must determine the collateral value behind a specific mortgage to price the risk of defaults, and (institutional) investors need property valuations to set reservation prices when acquiring or selling properties and to track the performance of their portfolio. Such valuations are still mainly performed manually, but with a growing amount of properties and smaller time intervals in which valuations are demanded, this task has become more challenging than ever. The recent increase in quality Commercial Real Estate data and new modelling techniques to handle it has opened new possibilities to the realm of Automated Valuation Models (AVM). This chapter introduces the topics covered in this thesis through the following sections:

1.1 Research Problem

1.1.1 Motivation

Many parties benefit from valuations being as accurate as possible. The main valuation method still used today is the direct comparison method that assesses the value of a property based on transaction prices of 'comparable' properties. However, since properties are never truly comparable – which especially is the case for Commercial Real Estate – prices need to be adjusted for differences in their characteristics such as unique building, location, lease and market factors. This process is often performed manually in practice which consumes a lot of time and manpower. AVMs have the potential to aid valuers in their process to meet new client expectations and could provide an advantage in terms of both costs reduction and increased prediction accuracy.

The advances made within the last decade in terms of computational power and econometrical modelling can help us greatly with the estimation of the Market Value of properties based on past transactions. In the owner-occupied housing market such AVMs are already widely in use. Examples are the Zillow's algorithm (Zestimate) that has revolutionized the residential valuation sector in the U.S., the Hierarchic Trend Model (HTM) used by Ortec Finance to value millions of houses in the Netherlands for taxation purposes (Francke, 2008) and companies like Geophy and PriceHubble that use the latest Machine Learning techniques in combination with 'Big' data to value Commercial Real Estate around the world. These examples show that AVMs indeed have the potential to outperform manual valuations in terms of prediction accuracy for only a fraction of the time and costs. Geophy (2017) even boldly states that within five years the market for valuations is entirely automated. Nevertheless, experience tells us that such models are highly dependent on the availability of quality data and so-called 'black-swan' events are especially difficult to capture within such models (David Geltner & Neufville, 2018). Additional research is needed to investigate the potential and limitations of AVMs applied to the Commercial Real Estate sector.

The gap in literature of AMVs applied to the Commercial Real Estate can be traced to the following non-exhaustive reasons:

- The number of Commercial properties and number of transactions are relatively low. Hence, studies can only be performed with small data samples.
- Commercial property characteristics are not centrally collected in contrast to housing where reliable and complete administrative data is typically available. In Commercial Real Estate, data collection depends on private companies who often aggregate their data from multiple sources (mostly broker and real estate news). Data is therefore often incomplete and more affected by entry errors. Also, data on the quality of the building such as state of maintenance and structure quality, among others, are rarely available.
- Commercial properties are more heterogeneous compared to housing. In combination with the lack of collected characteristics, econometric modelling is a challenge as such models tend to have a low fit and out-of-sample performance.

Recently, Cushman & Wakefield the Netherlands has taken the initiative to collect and store all data related to the Real Estate sector centrally in a personal Data Warehouse. With the acquisition of DTZ Zadelhoff in 2017 to form the largest Real Estate advisory firm in the Netherlands, the amount of Commercial Real Estate data is impressive. This research makes use of a first subset of this data to assess the accuracy of various AVM methodologies and discuss potential future implications of our results for the valuation profession.

We argue that recent advances in modelling techniques in the fields of Econometrics and Artificial Intelligence (AI), in combination with the ever-growing amount of quality data in both frequency and detail, provides a unique opportunity to investigate the potential of AVMs for the Commercial Real Estate sector. We focus on historical transaction prices of office properties throughout the Netherlands as this sector has most data available among Commercial Real Estate segments. Although potential determinants are derived from international literature and practice, the final model specification depends on the availability of the data.

1.1.2 Problem Description

One of the main challenges of this thesis is to model the spatial- and temporal dependencies within the model specification. The market value of a property is known to greatly depend on the value of a comparable property sold within the same local market. This might be the result of developers incorporating similar building technologies to meet market demands, building codes that result in homogeneous requirements, or real estate booms that lead to concentrated developments of comparable types at specific locations. Furthermore, investors and valuers likely consider these local market conditions when assessing the market and comparable buildings when valuing a property. We thus see that two identical properties could yield very different values if they are in different submarkets or are sold under different market conditions. Not controlling for these spatial and temporal effects within the model leads to both biased and inefficient estimates. Although proxies are often used, the uncaptured submarket information and discontinuities flow into the Hedonic residuals which has been one of the main reasons why previous literature show contradictory results.

1.2 Aim and Hypotheses

The *aim* of this thesis is to investigate the potential of AVMs applied to the Commercial Real Estate sector in the Netherlands. Based on newly collected transaction data from the real estate advisory firm Cushman & Wakefield, enriched with data from a wide variety of open source data such as BAG, CBS, PDOK and Google Maps, this thesis provides a first practical application of this 'Big' data. Both traditional Hedonic regression models as well as newer Machine Learning algorithms are applied to the same data in order to compare their potential for the Automated valuation of the heterogeneous goods that is Commercial Real Estate. The efficacy of the models is judged by their prediction accuracy and interpretability of the results. At the end of this thesis the best models are incorporated in an AVM application that provide valuers a first indication of the Market Value of individual office properties. A subsidiary goal is to create a useful reference material for new entrants to the AVM community, hence this thesis aims to describe the models and techniques in layman terms. To accomplish these goals three *hypotheses* are tested:

- **Hypothesis 1**: Lease related factors are important determinants of the Market Value, hence AVMs aimed at the Commercial Real Estate sector that include such factors provide significantly better prediction accuracy.
- **Hypothesis 2**: Incorporating methods to account for the dependencies over space and time as an extension of the traditional Hedonic regression framework reduces observable spatial and temporal patterns in the residuals and increase the predictive accuracy of the AVM.
- **Hypothesis 3**: A well-defined Hedonic regression model outperforms the most promising Machine Learning models (i.e. Random Forest Regression and Extreme Gradient Boosting) in terms of explainability, reliability and predictability with thin market data.

1.3 Research Design

The aims and objectives of this thesis are accomplished by employment of the following research methodologies.

a) A Critical Review of AVM Related Literature

The literature used in this thesis spans over three decades. It is comprised of published articles and other accessible sources that concern themselves with the theory and practice of AVMs. Particular emphasis is placed on explaining the general workings of the most common AVM methodologies, conclusions derived from previous literature and listing the advantages and limitations of applying these methodologies to the task of individual property valuations.

b) Acquire Data and Data Preparation

Data on actual office property transactions including a variety of determinants are gathered for the Dutch office market as a whole. The decision to focus on this rather broad market instead of for example a specific city or region has multiple reasons. First, the data is available at Cushman & Wakefield which provides a unique opportunity to conduct research at this scale. Second, the goal of the final model is to predict the value of a single property which can be situated anywhere in the Netherlands. Taking into account the full range of possibilities ensures that the findings can reasonably be generalised to any out-of-sample observation. Kempf (1999) furthermore show that although the pattern of demand is different across major office markets, empirical analysis does not detect any major, statistically significant differences in value influencing factors between regional markets. Hence, it might not be necessary to model each regional office property market separately. Last but perhaps most important, as Kok et al. (2017) also highlight, price effects from variables such as distance to the nearest station are similar throughout the Netherlands. By increasing the scale, we can increase the number of observations and in result increase the reliability of our estimates. The final dataset consists of 979 office transactions spread over 916 unique addresses from 2010 through 2018.

c) Establish Baseline Hedonic Regression Model

We first establish a Baseline Hedonic regression model that is used to demonstrate the significance of variables extracted from the international literature and practice. These variables can broadly be categorized in location, building, lease and market related factors (see *Figure 1-1*). Considerable effort has been made to develop an optimal model through extensive data preparation and feature engineering. This Baseline model is also used to evaluate the first hypothesis which is to investigate the importance of lease related factors to the prediction capabilities of an AVM model. Lastly, we use the final Baseline model specification obtained through both directional stepwise regression to compare the performance of other AVM methodologies against.

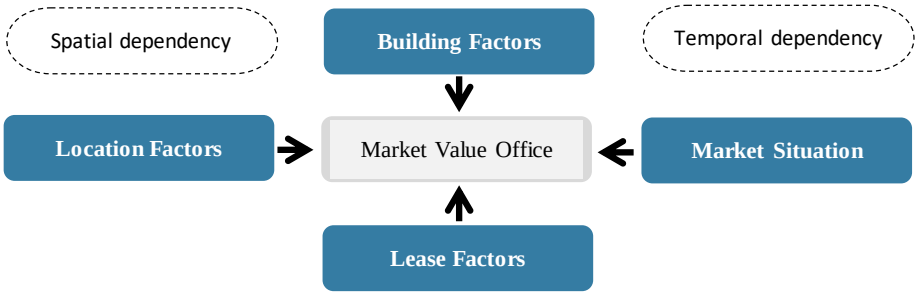


Figure 1-1 ■ Categories of Office Market Value Determinants

d) Examine Spatial and Temporal Model Extensions

The aim of this thesis is one of the classic goals of price statistics: the quantification of the true price of a given good with a certain quality. As Brachinger (2003) describes, "[…] the problem is that qualities change in time and the goods of today are no more the same as yesterday. So the goods actually available on the market are no more directly comparable with those which were available before. Therefore, for price comparison, prices have to be quality adjusted". In addition, as Tobler's (1970) first law of geography states: "Everything is related to everything else, but near things are more related than distant things". Heterogeneity across geographic stratums can thus also result in conventional OLS-based multiple regression analysis models' inability to accurately capture the variables' true effects. One of the main challenges of this thesis is to address both these spatial and temporal dependencies for the relatively thin market that is the office sector in order to optimize the prediction accuracy. Borst (2015) provides a practical overview of the most commonly applied spatiotemporal methods with the Geographical Weighted Regression (GWR) method showing the best performance in the Residential Real Estate sector. In this thesis we extend this framework to the Commercial Real Estate sector where dependencies are found in more dimensions than physical distance only.

e) Develop and Compare Tree-Based Machine Learning Models

Machine Learning techniques might provide more accurate estimates of the market than traditional linear regression models. These models may 'find' connections and patterns within data that might not be found with traditional methods (e.g. non-parametric relations). Although these algorithms are typically quite accurate, most are often described as 'black-boxes' since the complexities of their inner-mechanics render it nearly impossible to find direct relationships between input and output. We derive from the literature review that the Tree-based models offer the most potential as these models score high in both prediction accuracy and interpretability. Two extensions of a decision tree are applied in this thesis, namely Random Forest Regression and (Extreme) Gradient Boosting.

f) Analyse Results and Development of an AVM Tool

The results of all model formulations and methodologies presented in this research are compared and contrasted in accordance with the hypotheses in order to find an optimal model in terms of explainability, reliability and predictability. The final model is compared to the accuracy that valuers achieve in practice and implications are derived from these findings. The last step is to develop a tool that provide valuers of Cushman & Wakefield a first indication of the market value based on the results of this thesis. This tool will be further developed over the years.

1.4 Academic and Practical Relevance

1.4.1 Academic Relevance

Over the last decade, Automated Valuation Models have gained a lot of attention in academics with many studies showing the advantages and limitations of various modelling techniques. These studies are however all been aimed at the Residential market where data is numerous, while studies towards Commercial Real Estate have remained under-highlighted. This thesis therefore derives knowledge from international Residential literature and investigates whether these findings also hold for the Commercial Real Estate Sector in the Netherlands. In addition, a new methodological framework is proposed that assigns weights to observations based on its comparability.

1.4.2 Practical Relevance

Automated Valuation Models have the potential to aid valuers with the appraisal of Real Estate by providing additional information about relevant price effects and value determinants. Instead of looking at only a few comparable as is the case for manual appraisal, these data-driven models can use all relevant market data to estimate the (market) value of a property. If we can aid valuers with an AVM application that makes use of this information by only fraction, this would be highly beneficial for both the company and clients and our original goal would have been achieved.

1.5 Outline of the Thesis

This thesis is comprised of eight chapters. The organisation and relationships among chapters are shown in *Figure 1-2*. *Chapter 2* provides information about the setting where the AVM model is applied to and what the model needs to achieve. We first introduce the characteristics of the Dutch office market and discuss what the role of valuations in this whole. Furthermore, we provide a definition for the value used as dependent variable, inform about the current valuation accuracy in practice and discuss causes of potential prediction errors. *Chapter 3* provides a review of the most popular empirical modelling approaches used within the mass appraisal literature. After shortly describing each methodology, we review literature which apply these models and denote their strengths and weaknesses. The purpose of this chapter is to find methodologies that offer the most potential for the task at hand.

Chapter 4 further elaborates the methodologies that were found in Chapter 3 to offer the most potential. *Chapter 5* describes per phase how the data is prepared from raw towards the final cleaned dataset and *Chapter 6* covers the measures and methods applied to estimate and compare the performance of the different models. *Chapter 7* discusses the results of the various models that are evaluated against each other. Finally, *Chapter 8* summarises the study and draws conclusions regarding the applicability of the results. The thesis closes with a discussion about (potential) future implications of the findings and suggestions for further research.

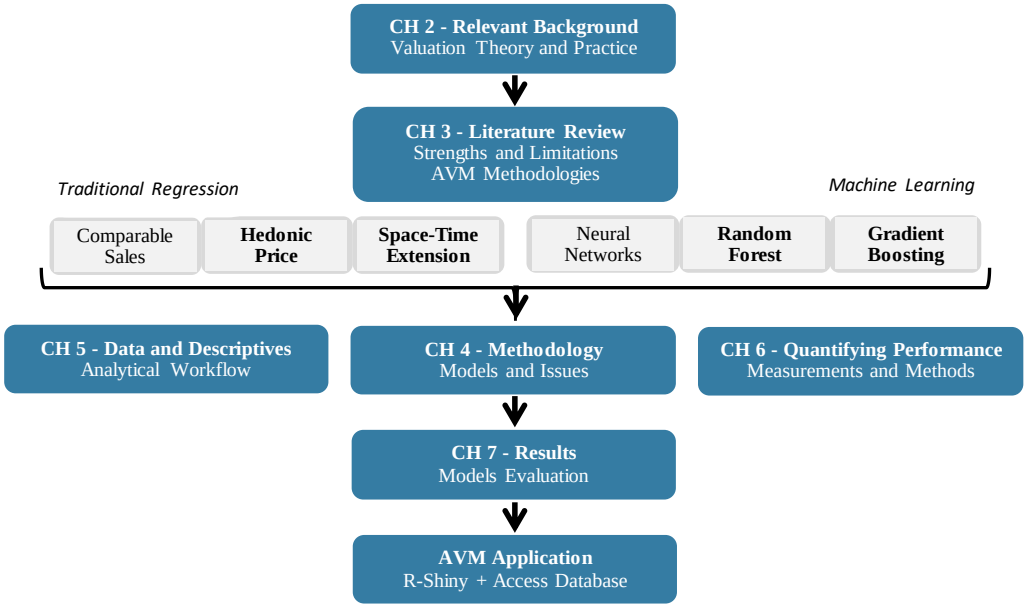


Figure 1-2 ■ Thesis Outline

2. Relevant Background

During the past decade a lot has changed in the way real estate is appraised. Reliable models seem to play an ever-increasing role in both determining and validating value estimates. Important arguments in support of new models are more objectivity, reproducibility, efficiency, cost reduction, among others. The employment of Automated Valuation Models (AVMs) can aid valuers in these changing landscapes as is the topic of this thesis. But before going into the details, this chapter discusses some relevant background to provide information about the setting where the AVM is applied to and what the model needs to achieve. The following sections are covered:

2.1 The Dutch Office Market

The Dutch real estate market has undergone some changes through the years. Deeper understanding of key developments can help improve any valuation model, hence this section briefly reviews related topics to the office market. In line with the well-established four-quadrant model of DiPasquale and Wheaton (1992), we discuss both the 'space market' with information about the current office market stock and the 'asset market' with information about (prime) rents, yields and recent market developments.

2.1.1 Office Market Stock

According to the 'Basisregistraties Adressen en Gebouwen' (BAG) - the government agency that manages building and address information in the Netherlands - the Dutch office market counts approximately 47.5 million square meters of office space. More than 15 thousand of these office properties have a Lettable Floor Area above the 500 square meters. Broadly, this market can be categorized by the 'top 4 (large) cities', the 'other Randstad locations' and the 'remainder of cities outside the Randstad region (IVBN, 2017). However, even within these seemingly similar locations, large differences can be found.

Amsterdam offers with a stock of around 6.9 million square meters the most office space in the Netherlands. This market mainly focusses on the financial sector and international trade with the Zuid-As as the best performing district. Prime rents here have risen to over 425 euro per square meter (Cushman & Wakefield, 2018). Den Haag is the second largest market in the Netherlands which houses the Dutch government within its city Centre. Prime rents equal 210 euro per square meter. Rotterdam is with a stock of 3,3 million square meters the third largest city in the Netherlands in terms of office stock. Their market is mainly aimed at international trade and insurance companies with a prime rent of 235 euro per square. Utrecht is with 2.5 million square meters the fourth largest city and focusses mainly on transport and the service sector. This can be traced back to its central position within the Netherlands. Prime rents equal 275 euro per square meter. The prime locations in these cities all have a Gross Initial Yields (GIY) of approximately 3.5 percent.

The Dutch Office market is also characterized by high level of vacancy throughout the country. This fact does however not acknowledge the significant differences that can be observed within fairly short distances. Medio 2018 the office vacancy counted approximately 5.5 million square meters or 11.6 percent of the total stock; a decrease of more than 3.7 percent compared to a year before (Cushman & Wakefield, 2018). More than half of the vacant offices are located in the Randstad. The largest amount of this vacancy is found in office districts that are built before 1995. It has long since been recognized that this vacancy is likely to be structural and we should therefore accept the hard fact that a new purpose has to be found for these buildings. In the last decade, of all the offices that were removed from the stock, around 60 percent was demolished and 40 percent transformed (Rijksdienst voor het Cultureel Erfgoed, 2013).

2.1.2 Office Market Developments

We observe a distinction between qualitatively high versus low real estate in the Dutch Office Market. When focusing on the top 4 cities and other prime office locations, we discover that in recent years these locations have done well with a high take-up, stable to rising rents, substantial lower vacancies and a competitive direct return and capital growth. On the other hand, at the periphery of large cities and border municipalities such as Capelle aan de IJssel, Hoofddorp and Nieuwegein, we see an

opposite trend happening. In these areas the vacancy rates have risen above the 30 percent, the average rent per square meter are below the national average of 133 euro per annum and are highly influenced by incentives that can rise up to 30 to 50 percent of the market rent, whereas the national average is around 10 to 30 percent (Cushman & Wakefield, 2017). Dynamis (2018) even expects that this dichotomy will only grow further in the near future. However, according to the Office Property Clock of JLL (2018), the Dutch market will soon move from a Rental Value Accelerating market to a Rental Value Growth Slowing market. This would mean that the recent developments in (top) rents and yields will soon slow down and move towards falling rental values as we see in Paris at the moment. Any AVM should investigate how these developments over time and space can best be modelled.

2.2 Value Definition

This section briefly discusses the role valuations have in the real estate market and which value definition is used throughout this thesis. These topics provide additional information about the purpose that the developed AVM has and what it aims to estimate.

2.2.1 The role of Valuation

Real property is defined as all the interest, benefits, rights and encumbrances inherent in the ownership of physical real estate (Pagourtzi et al., 2003). As the commercial property assets are known for their lack of continuous trading and transaction-based indices are rare, valuers perform a vital function in the market by acting as a surrogate for transaction prices. This most of the times comes in the form of a single Market Value estimate. Similar to the pricing in the bond or equity market, real estate appraisals are key to the interrelated process of acquisition, disposal and performance measurement. Nevertheless, within both academia and practice there is also some skepticism about the ability of valuers to fulfill this task in a reliable manner (e.g. Schekkerman, 2004).

The value is determined by a great number of characteristics associated with the subject property such as size, location, a range of quality attributes and future cashflows. Property valuation is concerned with the identification and analysis of these many characteristics, however difficulties arise in quantifying the influence of the sheer number of characteristics in order to come to a value estimate based on the latest information. This is even more so the case for Commercial Real Estate that is known for its heterogeneity. So even though a valuation is based on professional judgement, it remains partly a subjective opinion of value based on an assessment of influences that a valuer considers relevant to the value of the subject property at that time and is therefore as much an art as a science.

In addition, in many properties markets it is commonplace for the ownership of property to be separated from its use. That is, the price of exchange will be the same whether the purchaser has investment or occupation in mind while in reality the view of these two groups of bidders can be very different. An investor will view worth as the discounted value of the rental stream produced by the asset, whereas the owner-occupier will see the asset as a factor of production and assign to it a worth derived from the property's contribution to the profits of the business. Hence, any asset is likely to have several different values that is not always easy to distinguish. For any valuation it is thus important to ask the questions: value to whom, and for what? The next section answers these questions for the purpose of this thesis.

2.2.2 Market Value

Market Value is generally accepted in academics and practice as the 'same-for-all' value, that is the prime indication of the exchange price of an asset if it were to be sold in the open market. Given that the compelling reason for using Market Value is to ensure consistency in the valuation process, it is important that there is agreement in the details. Hence, the International Valuation Standards Comittee (2017) provided an international 'standard' definition that is as follows:

> *"Market Value is the estimated amount for which an asset should exchange on the date of valuation between a willing buyer and a willing seller in an arm's length transaction after proper marketing wherein the parties had each acted knowledgeably, prudently and without compulsion."*

In other words, the property is sold at the specified data in an open and competitive market where both buyer and seller act prudently and knowledgeably. The price is unaffected by undue stimulus with both parties acting in what they consider their best interest. Adequate market efforts are made and a reasonable time is allowed for exposure on the open market. Payments are made in cash or in comparable terms of financial arrangements without being affected by special considerations granted by anyone associated with the sale. It is important to realize that the above value definition is based on the past information. Valuers thus reflect the market, they do not make it.

We assume that Market Value can be derived from transaction prices. The idea is that it should be possible to estimate the Market Value based on historical transaction prices as this would be the price paid for the property on the open market. We realize that this rationale does not always hold as not all points in the above definition are satisfied; real estate properties are unique and the prices parties pay relate to the information that is available to seller and buyer and their negotiation skill. However, the more transactions of homogeneous products are available, the closer these transaction prices are likely to be to the actual Market Value (Francke, 2017). So, the best approximation of Market Value we can make with our AVM is to compare recent transactions of similar properties.

2.3 Manual Valuation Accuracy

Before we can derive conclusions from our model and discuss the potential of AVMs, it is important that we obtain some understanding about the accuracy of the current valuation practice. We first review literature that discuss differences between appraised and actual transacted values. Next, we look at the potential causes of these valuation errors.

2.3.1 Valuation Accuracy in Literature

The accuracy of valuations has been approached from various angles. Hager and Lord (1985) are one of the first found articles that investigate the differences between value and the estimates of appraisers. Although they compare the deviation between value estimates of less experienced valuers against those of experienced valuers, they highlight that accuracy can differ significantly over time, markets and even persons. The overall difference was approximately 13 percent. Schekkerman (2004) wrote a master thesis about the valuation accuracy aimed specifically at the Dutch market. He did compare appraised values with the actual transacted values and concludes that in over two-third of the office cases, the error was around 20 percent. Furthermore, 20 percent of the transactions were within a bandwidth of 5 percent of the last appraisal. This would suggest that a lot of gain could still be obtained by improving the valuation methods used currently.

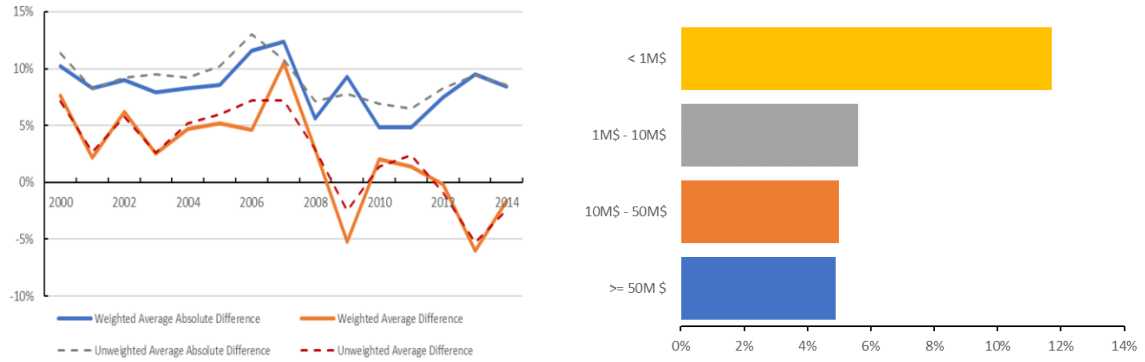


Figure 2-1 ■ Valuation Accuracy Offices in the Netherlands

Note: Left over time, right per value category. The valuation accuracy is expressed as the difference between the appraised value of the property adjusted for time to the sale price. Weighted values attach greater importance to more expensive properties. (*Adapted from MSCI, 2015)*

The MSCI (previously the ROZ-IPD) periodically publishes reports that compare last valuation estimates to the actual transaction prices. Most recently, MSCI (2015) analysed 12 office markets from 2005 through 2014. We see in *Figure 2-1* that the weighted average absolute difference in the Netherlands over the last decade was approximately 10 percent, the second lowest error among the countries across the world. Furthermore, we can see higher difference during the crisis of 2007 and lower in the aftermath. Most of the error in 2014 comes from properties that are worth less than 1 million dollars. A possible reason for this could be that these properties are often sold to smaller parties that have lower negotiation power and tend to pay a price above Market Value. Larger properties have an error of approximately 5 percent on average.

2.3.2 Causes Difference in Valuation and Sale Price

We've thus seen that the actual transaction price and the appraised value can significantly differ. Various causes can be uncovered for this. For example, van Gool and Have (2006) asked before the global economic crisis whether or not real estate appraisal methods were not causing bubbles in the market. They stated that increasing (structural) vacancy and incentives were insufficiently and inadequately incorporated in the valuations. In the years that followed this indeed seemed to be the case. Nowadays, the difference between the market rent and effective rent are well recognized in practice and have resulted in more stable value estimates. This shows that the market is continuously developing. Nevertheless, differences between appraised and transacted values continue to exist that are mainly rooted in 'appraiser behaviour'. McAllister et al. (2003) distinguish three causes:

- Historic appraisals influence current appraisals through an 'anchoring' bias. That is, appraisers often know previous estimates and are influenced by these past results.
- Most appraisal methodologies utilize historic transactions and thus introduce a delay in market change known as 'lagging' bias. Appraisers are thus in general slow to include non-transaction based information.
- Given historic transaction prices that does not necessarily represent Market Values, appraisers need to manually adjust information. This introduces additional 'noise' that cause errors in the estimates.

More data-driven models could offer a solution against appraiser-behaviour as they are less prone to the above described biases and have been one of the reasons of the recent interest in AVMs.

2.4 Chapter Summary

We've seen that the Dutch office market is a relatively small market with most of the stock situated within de Randstad region that covers the four largest cities in the Netherlands. However, even within this fairly short distance significant differences can be observed. Interestingly, in each of these cities a different sector is dominant. As Amsterdam is the main financial and international center, the highest rents and lowest yields can be found here that are still rising. On the other hand, in the periphery regions, rents are below the national average with high vacancy percentages and are likely to be structural. This dichotomy is expected to continue in the foreseeable future. Hence, we can conclude that any valuation model that aims to estimate values on a property level over multiple submarkets need to allow for such differences over space and time.

It is well-known that appraisals perform a vital function in the property market by acting as a surrogate for (expected) transaction prices. In order to provide a 'same-for-all' value for the assets, the standard definition of Market Value was adopted. In this thesis the assumption has been made that the historical transaction prices are Market Values and can thus be used to predict the Market Value of new data. However, it is important to realize that this might not always hold true. This has also been part of the reason for errors among manual appraisals together with other appraisal behavior biases and is part of the critique on current practice. More data driven models have the potential to offer more reliable estimates as they are less subjective to individual transactions, among other reasons. The aim of this thesis is then also to investigate whether the average 'error' rate of manual appraisals in the Netherlands which is around 10 percent can be outperformed by an AVM with the data at hand at Cushman & Wakefield. The next chapter reviews literature that provide information about the models that can be used for this task.

3. Automated Valuation Methods

Automated Valuation Models (AVMs) have been the object of study for many decades now. This chapter reviews some of the main methods that have been applied in literature and aims to unravel their advantages and limitations for the task at hand (see *Figure 3-1)*. First, some of the more traditional methodologies are covered, namely the Comparable Sales Method (CSM), Hedonic Price Model Regression (HPM) and the Spatial-Temporal Model Extension (ST). Next, the 'newer' Machine Learning algorithms are reviewed, namely Artificial Neural Networks (ANN), Decision Trees (CART), Random Forest (RF) and Gradient Boosting (GBM). Finally, these methods are evaluated based on potential prediction accuracy for individual properties valuation and the degree of interpretability and confidence of its results.

3.1 Traditional AVM Methods

The first known commercial application of AVMs was in North America at the start of the '90s (Matysiak, 2017). The method applied was basically an automated extension of a Comparable Sales Approach. During the last decennia, the Hedonic Price Theorem of Rosen (1974) gained momentum in both academia and practice and has grown to become the most established methodology for AVMs worldwide (European AVM Alliance, 2017). However, in recent years we observe a trend towards the application of Machine Learning algorithms that have proven to be hard to beat in terms of prediction accuracy[1]. This section reviews literature in which the more traditional models are applied, whereas the next section covers Machine Learning. As studies on Commercial Real Estate are relatively rare, insights are derived from the Residential real estate sector to fill the gap.

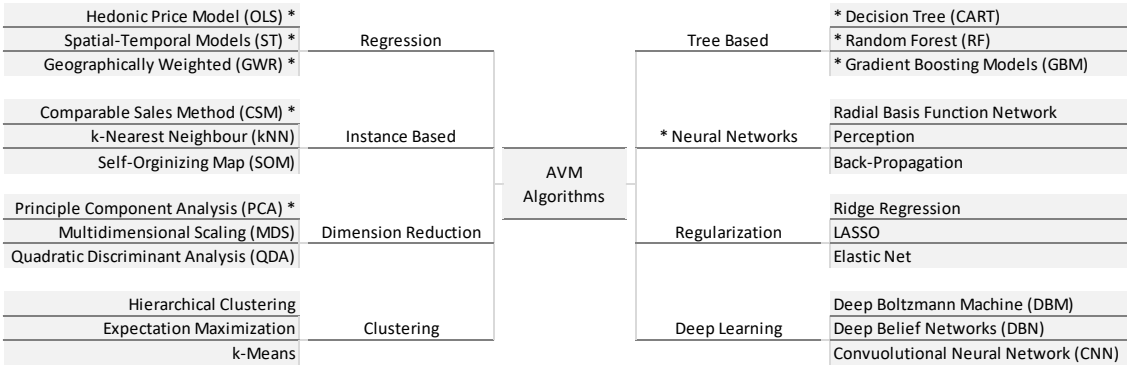


Figure 3-1 ■ Overview AVM algorithms

Note: On the left side we see the more traditional algorithms and on the right side the Machine Learning Algorithms. The (*) indicates whether the algorithm is covered in the literature review of this thesis. For more information about the various algorithms we recommend the overview of Brownlee[2].

3.1.1 Comparable Sales Method

The direct Sales Comparison Approach (CSM) is a method that produces an estimate of the Market Value for the subject property based on the transaction prices of similar properties. Lusht (2012, p. 83) sums the two fundamental assumptions of the CSM approach. First, that the comparable data should be a reliable indicator of the Market Value of the subject property and second that equal properties should transact for a similar price. Using statistical techniques these comparables can be selected in an automated manner. In some cases, adjustments must be made when properties are not one-on-one comparable. Here we assume that differences are corrected through a more qualitative process instead of regression. It is thus paramount that the right comparables are used to create a small homogeneous sample that need as little corrections as possible to derive an accurate value for the subject property.

3.1.1.1 Literature

Although the CSM is still often performed manually, intelligent tools have been developed that employ sophisticated algorithms to select the closest matching properties based on the individual characteristics and location of the property to be valued. The workings of such models have been covered by some relatively few academic literatures but are more applied in practice. Krause and Kummerow (2011) provide an example of an automated CSM that incorporates a statistical method to

[1]Hedonic Price Regression technically also fall under the category of Machine Learning. However, we assume that they are separated methodologies in this thesis.
[2] See https://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/

evaluate prediction errors. The nearest comparables are found by the algorithm and are automatically adjusted based on the differences between the properties. They find that the CSM outperforms Hedonic regressions in terms of prediction accuracy and also reduce spatial autocorrelation in the error terms. CoStar (2018) use various (Machine Learning) algorithms to select comparable properties for Commercial Real Estate[3]. They offer one of the largest databases for Commercial Real Estate and are involved in more than 85 percent of the transactions in the US. This software is then also used by many real estate companies in the US and often the base for commercial AVMs.

3.1.1.2 Advantages

AVMs based on CSM are easy to develop and typically return accurate estimates within a short amount time. The European AVM Alliance (2017, pp. 27-28) provides an overview of advantages:

- As the model selects the most appropriate set of comparables upon which to determine the value, high prediction accuracy can be achieved if the sample is homogeneous. In such case, the CSM method is likely to outperform other AVM methods as the CSM do not (or to a small degree) rely on pre-defined variables. Price effects that are thus not considered explicitly but are carried implicitly through the value information of its comparables. As Commercial Real Estate is relatively heterogeneous, the creation of a homogeneous sample can however be challenging.
- Unlike most statistical methodologies, the CSM do not require all features related to the transacted property to be available as input. Again, the model can derive these implicitly from its comparables. It can thus cope with missing data which is a common problem for real estate data.
- CSM can include a confidence interval based on its comparables which is a critical piece of information that is missing in e.g. Machine Learning algorithms. It indicates the extent to which the value estimate can be relied upon and thus provides an indication of risk.

3.1.1.3 Limitations

Although CSM can achieve relatively high accuracy even with incomplete data, there are several limitations to this method (European AVM Alliance, 2017, p. 28):

- Accuracy highly depends on the quality of the data available and is sensitive to outliers. If 'bad' comparables are included in a small sample they will skew the estimated value towards these observations. As Commercial Real Estate is a heterogeneous good, it can be challenging to evaluate which properties are truly comparable.
- Often not more than a handful of comparables are used to estimate the value. However, due to the infrequent trading of Commercial Real Estate it can be difficult to find comparable properties in the local market. If no reliable comparables can be found, no reliable value estimate can be generated.

3.1.2 Hedonic Price Regression Model

Hedonic Price Models (HPM) are widely applied to explain and simulate the pricing of heterogenic assets such as real estate. This valuation framework measures the contribution of systematic factors to the value of a property. By regressing the transaction prices against corresponding characteristics, one can equate supply and demand for quantitative and qualitative characteristics in a static framework. In other words, each property characteristic has a quantifiable influence on the property value and said contribution can be isolated. This approach is similar to the valuation process an appraiser adapts subconsciously. Rosen (1974) provided the theoretical framework of this methodology but the details of this economic theory are out of the scope of this thesis.

[3] In the Netherlands Momentum Technologies has developed a similar technology.

Table 3-1 ■ Overview Hedonic Price Literature

Authors	Sample size	Study area	Time span	**Building Factors**	Age	Building quality	Size	Height	Lot Area	Amenities in-house	Parking	Energy Label	**Lease Factors**	Rental Income	Lease Term	Vacancy levels	Incentives	Contract Features	**Location Factors**	District type	Accessibility	Distance centres	Amenities	Other attributes	**Other Factors**	Whole building	Buyer info	Seller info	Time on Market	Sale Conditions	**Control Variables**	Submarket dummies	Time dummies	Spatial Autoreg.	Temporal Autoreg.	Spatial-Temporal Autoreg.	**Total Variables**	(max) adj-R2
Dependent Variable - PRICE																																						
Hedonic																																						
Nappi-Choulet *et al.* (2007)	2,587	Paris	1991-05		X		X													X						X						X	X					0.89
Colwell & Munneke (2006)	477	Illinois	1995-97		X		X	X	X												X	X	X				X	X	X				X	X					0.84
Sivitanidou (1996)	539	Los Angeles	1987-92		X	X		X			X									X		X	X	X									X	X					0.59
Hough and Kratz (1983)	139	Chicago	1978		X	X	X	X		X	X																						X	X					0.72
Orr *et al.* (2003)	333	Schotland	1994-84		X	X							-	X															X					X					0.51
Chegut *et al.* (2013)	2,103	London	2000-09		X	X	X	X				X	-	X		X	X				X						X		X										0.61
Fuerst & McAllister (2011)	9,806	USA	1999-08		X		X	X	X			X	-	X							X		X										X	X					0.42
Dutch Market																																						
Brinkman (2014)	618	NLD	2011-14		X		X						X	X	X					X				X									X	X					0.67
Ziermans (2016)	386	Amsterdam	2002-12		X		X	X						X			X				X		X	-		X	X			X			X	X					0.63
Price Index																																						
Colwell *et al.* (1998)	427	Chicago	1986-93		X		X	X	X											X	X	X	X										X						0.84
Downs and Slade (1999)	935	Phoenix	1987-96		X	X	X	X	X													X					X			X			X	X					0.84
Hodgson *et al.* (2006)	1846	Phoenix	1997-04		X	X	X	X																									X	X					0.91
Munneke & Slade (2001)	890	Phoenix	1988-96		X		X	X	X													X					X			X			X	X					0.89
Spatial-Temporal																																						
Tu *et al.* (2004)	2,950	Signapore	1992-01				X	X																		X											X		0.80
Nappi-Choulet (2009)	2,587	Paris	1991-05		X		X													X						X						X	X		X	X		0.94
Chegut *et al.* (2015)	978	top markets	2007-13		X	X	X	X												X							X	X				X	X	X	X	X		0.57
Dependent Variable - RENTS																																						
Kempf (2016)	23,000	GER	1997-06		X		X	X					-	X		X	X															X	X					0.33
Koster *et al.* (2013)	4,792	NLD	1990-10		X	X	X	X					-				X				X			X					X			X	X					0.70
Dunse & Jones (1998)	477	Glasgow	1994-95		X	X	X	X		X	X		-							X												X						0.61
Glascock et al. (1990)	675	Baton Rouge	1985-88				X						-																			X	X					0.55
Mills (1992)	543	Chicago	1990		X		X			X	X		-							X			X									X	X					0.38
Oven & Pekdemir (2006)	52	Istanbul	2003		X		X	X		X	X		-	X		X	X			X	X	X	X	X															0.78
Mooradian & Yang (2000)	311	USA	1990				X						-		X		X															X	X					0.65
Ryan (2005)	520	San Diego	1986-95		X		X	X		X			-		X		X				X	X	X									X	X					0.52
Feurst (2007)	950	Manhattan	1999-04		X		X	X		X			-		X																		X					0.51
Slade (2000)	483	Phoenix	1991-96		X		X	X					-													X							X					0.32
Fisher and Webb (1996)	395	Chicago	1985-91		X	X							-	X																			X						0.44
Archer & Smith (2003)	647	Houston	2000		X		X	X					-		X							X											X	X					0.56
Bollinger *et al.* (1997)	907	Atlanta	1990-96		X	X	X	X		X	X		-								X	X	X	X		X								X					0.63
Dunse & Jones (2002)	539	Glasgow	1992-98		X		X			X	X		-									X											X	X	X				0.58

Note: This table provides an overview of the reviewed literature of traditional HPMs. The 'X' indicate whether the variable (category) was included in the study. '-' indicate that a variable was also used as dependent variable.

3.1.2.1 Commercial Real Estate Literature

The HPM is one of the most commonly used methodologies in the real estate pricing literature, especially in the Residential sector. Within the Commercial Real Estate sector however we only see a limited amount of applications, and the ones that do are confined to the U.S. and U.K. as the data in these markets is more developed. Nevertheless, improvement in information has encouraged development of this type of work in years to come. *Table 3-1* provides an overview of the variables used in noteworthy studies in Commercial Real Estate valuation.

The largest body of publication on office properties concern studies with rents as the dependent variable. Although this paper's aim is to analyse prices, rents provide a first series of articles that give interesting insight into value determining factors for Commercial Real Estate. Most notable for this thesis is the dissertation of Kempf (2015) about the construction of a Hedonic rent index for the German office market. As is the case for the Dutch market, he indicates that there is a gap in the literature for such models in Germany. Hence, through empirical analysis among real estate professionals he uncovers determining factors of rent. Probably the most interesting finding is that although the various markets have different economic forces, similar variables influence the rent levels. As such, he concludes that the whole market can be modelled by a single Hedonic price function.

Other notable studies of the application of HPM models that have rent as dependent variable but offer insight into potential determinants of price are as follows. Many aim to find significant dependent variables for rents in general (Dunse & Jones, 1998; Glascock et al., 1990; Mills, 1992). Factor models that group effects of variables have been applied to account for high dimensional data (Öven & Pekdemir, 2006). Hedonic modelling techniques are often used to investigate the impact of specific variables. For example, Gat (1998) and Hough and Kratz (1983) investigate the importance of design quality on rent, Mooradian and Yang (2000) the price effects of cancellation strategies of the lease and vacancy, Ryan (2005) the value of highway access and rail transit, and Koster et al. (2014) the premium on height within the Dutch market. The effects of different economic periods on the price determinants have also been a popular field of study (Fuerst, 2007; Slade, 2000), while many others use Hedonic analysis to create constant quality- price indices (Webb & Fisher, 1996; Wheaton & Torto, 1994) But most Hedonic studies look specifically at the influence of location on rent (e.g. Archer & Smith, 2003; Bollinger et al., 1998; Dunse & Jones, 2002).

As mentioned, few have attempted to uncover the price determinants of Commercial Real Estate and the ones that do mostly focus on Hedonic price index construction, not individual property valuation. The first application seems to be from Colwell et al. (1998) who adopt the index methodology of Fisher et al. (1994) to derive some useful insights about office value trends in the office market of Chicago during the '80s that contrasted the conventional wisdom that values had declined. Downs and Slade (1999) apply Hedonic price analysis to construct a transaction-based index for office assets and do this in full-disclosure of the data and model. Hodgson et al. (2006) achieve high explained variance through the use of an semiparametric approach which might be a first indication in this thesis that Machine Learning models might provide better estimates due to their non-parametric nature. Munneke and Slade (2001) allow the implicit prices of different quality characteristics to vary intertemporally, overcoming the potential bias imposed by holding implicit prices fixed and simply interpreting time dummy variables as in a conventional Hedonic approach.

Some papers investigate specifically the effect of 'being green' on the property price. For example, Fuerst and McAllister (2011) find that in the same submarkets, eco-certified buildings have both a rental and sale premium. Similarly, Chegut et al. (2014) find for London over the 2000 to 2009 period that the expanding supply of green buildings had a positive impact on rents and prices, but reduced rents and prices for other environmentally certified real estate.

Nappi-Choulet et al. (2007) apply the Hedonic method to analyse transaction prices of office properties in Paris. They highlight that the application of HPM is rare for Commercial Real Estate prices and that their article is the first transaction-based Hedonic price index in Paris. A 'constant quality' index allows to compare price trends for specific districts over time. They however neglect the cross-effects of spatial and temporal dependencies. Colwell and Munneke (2006) achieve high explanatory power by including the influence of bargaining strength as a determinant of price. Sivitanidou (1996) looks specifically at the influence of specified service employment centres in comparison to the property value per unit of land whereas Orr et al. (2003) use Hedonic regression to investigate the influence of time on the market on the transaction price. The literature that is recently gaining more attention considers both spatial and temporal effects in the Hedonic price framework and is covered in the next section.

Less than a handful of papers have made an attempt to apply a Hedonic price model to the Dutch office market. Similar to the hypothesis of this thesis, Brinkman (2014) investigates the extent to which AVM models can value office buildings in the Netherlands. He uses the appraised value as a proxy for the Market Value of the 618 properties in his dataset and seven independent variables. With multiple stepwise regression he explains 67 percent of the variation in his data; too low for practical application but still has a lot of room for improvement such as the inclusion of spatiotemporal effects. Ziermans (2016) on the other hand, looks at the Amsterdam Office Market and aims to uncover the determinants of incentives on rents. Although this study concerns a different dependent variable, the method applied and variables included proved some interesting insights about possible determinants of value. In particular, the involvement of a commercial advisor that counters some information asymmetry between buyer and seller is shown to have a significant effect on the value. Furthermore, through the use of panel data he accurately distinguishes trends over time.

3.1.2.2 Advantages

The HPM are estimated using multivariate (OLS) regression analysis and the advantages and limitations are in line with this well-known methodology. The main advantages include:

- The HPM include the ability to estimate value based on concrete choices on causation. These are translated into a mathematical model that is very versatile and if well specified, robust. It is thus possible to generate reliable estimates in thin markets (Francke, 2017), which is especially relevant for the Commercial Real Estate covered in this thesis. This is in contrast to most Machine Learning techniques.
- Effect of individual characteristics can be identified while controlling for other variables that might cause spurious relationships. The price effects of characteristics are isolated and the marginal contribution to the composite Market Value are easy to check and interpret. In addition, confidence intervals can be provided with each value estimate that are essential to distinguish the reliability of the value estimate.
- HPM are based on simple regression techniques which are easy to implement and understand. Furthermore, the HPM approach has been utilised extensively to investigate the relationship between prices through its characteristics, so literature provides priory knowledge.

3.1.2.3 Limitations

Although HPM is one of the most used methodologies in the real estate literature, the model has been criticised by many as a standard HPM suffers from various shortcomings.

- Specification of a Hedonic model has always been a critical issue. It all depends on data availability and assumptions arbitrarily chosen by the researcher (Varian, 2014). We for example

see many different proxies being used to capture locational effects with the choice of such independent variables affecting both the model fit and patterns reflected in the residuals. Simply including variables (and pairwise interactions) would be infeasible. Often, we need some kind of variable selection that is a simplification of reality.

- HPM assumes a predefined functional form that is often linear in real estate studies. However, many variables have a non-linear relationship relative to the value. Machine Learning models may therefore allow for more effective ways to model such complex relationships (Varian, 2014).
- In contrast to the CSM, HPM assumes that the value of a property is a function of its individual characteristics. All marginal price effects thus need to be modelled explicitly while no implicit value effects are included.
- Spatial autocorrelation of residuals violates the assumption that OLS of residuals must be uncorrelated and normally distributed with zero mean and constant variance. Such can impair the power of the traditional Hedonic mode. These effects are often neglected: two observations that are close in space or time might be correlated, and the omission of these correlation effects can lead to a bias in coefficient estimates and/or heteroskedastic issues (Nappi et al., 2009 p.2). If this is also the case for office properties, remedial steps are necessary if there are such discernible patterns in the residual errors of the model. Control variables might improve the model but have drawbacks such as discontinuities.
- Potential other problems relate to fundamental Hedonic regression assumptions such as the identification of supply and demand, disequilibrium, observations that need to be independent and identically distribution (i.i.d.) and errors that are the Best Linear Unbiased Estimator (BLUE).

3.1.3 Spatial-Temporal Regression Models

We've seen that a serious limitation of the HPM is that observations are likely to show interdependence over space and/or time and thus do not meet the key assumptions of Ordinary Least Squares (OLS). Consequently, predicted prices can become unreliable and results in the inability to accurately capture variables' true effects. Spatial-Temporal (ST) regression can be seen as an extension of the HPM that aims to model the interdependencies over space and times in its framework

The groundwork of Anselin (1988) distinguishes two kinds of spatial effects: Spatial autocorrelation and spatial heterogeneity. Briefly, the former refers to a functional relationship between observations, while the latter is connected to the lack of uniformity arising from space, potentially leading to spatial heteroscedasticity and spatially varying parameters. While spatial consideration in the form of dummy variables and distance coefficients can help improve models, they may fail to fully correct for spatial autocorrelation and introduce discontinuities. Spatial models aim to incorporate this geographical information into the model. A similar rationale holds for the combination with temporal interdependencies and has been the focus of many studies in recent years.

3.1.3.1 Literature

A number of model structures allow the model coefficients to be functions of space and time. In this section we briefly review the most relevant literature related to one of the following models: Spatial Autoregressive Models (SAR), Geographically Weighted Regression (GWR), Spatial-Temporal Autoregressive Models (STAR) and Hierarchic Trend Models (HTM).

Spatial Auto Regressive Models (SAR)

Spatial Autoregressive models, or simply Spatial Regression models, are a category of models that extent Multiple Regression Analysis (MRA) by incorporating spatial dependencies directly into its functional form. Spatial dependence parameters are estimated together with the regression coefficients based on the hypothesis of stationarity. These estimates are derived from its spatial neighbours through weight matrices with properties that can be considered neighbours to a subject property receiving a heigh weight and those who are not a low weight. In general, three SAR models can be distinguished:

- Spatial Lag Model (SLM) models the dependency in the dependent variable. It hypothesises that the price of the subject property is dependent on the prices of its neighbouring properties. The logic behind this model is that there are spillover effects in which the sale price of nearby property (in the same market) effects the value of said property more than those that are distant. Such correlations caused by unobserved characteristics are difficult to capture within traditional multiple regression.
- Spatial Error Model (SEM) models the spatial dependence in the error terms. It hypothesises that (part of) the error induced by a property is dependent on the error of nearby properties. Such spatial patterns are for example caused through omitted random factors within the model resulting in the spatial dependence in the error term.
- Spatial Durbin Model (SDM) takes into account both dependence of prices and errors of neighbouring properties. It thus basically combines a SLM and SEM into one model.

Geographically Weighted Regression (GWR)

Another approach that can cope with spatial autocorrelation is Geographically Weighted Regression, also named Local Regression Approach. Fotheringham et al. (2002) introduced the model and explained that the essential process of the GWR is to calibrate the spatially invariant version of the basic model at a number of points across space using a weighting scheme that places higher weights on data nearer to the calibration point. In other words, GWR uses simple Multiple Regression Analysis (MRA), but unlike MRA, GWR produces a different set of regression coefficients for observations by running a set of Weighted Least Squares Regressions (WLS) at different points in space with weights determined by a function of distance to its neighbours. Therefore, GWR is essentially the combination of many weighted MRAs that are performed in proximity of each subject property. The result is a set of coefficients that are a function of location. Many find that GWR outperform standard Hedonic price models and spatial expansions (in the Residential sector) in terms of explanatory power and predictive accuracy (e.g. Bitter et al., 2007; Borst, 2007; Huang et al., 2010). McCluskey et al. (2013) even states that this spatially weighted approach is the way forward in the developing mass appraisal.

Spatial-Temporal Auto Regressive Models (STAR)

While Spatial Autoregressive (SAR) models measure the co-movement between transaction prices of neighboring properties, Spatial-Temporal Autoregressive (STAR) models measure the co-movement of transaction prices with transactions that are close in space as well as time. Pace et al. (1998) proposed an original framework of an extended SAR model that includes temporal effects and find it to be a powerful predictor in a Residential real estate context. We observe a recent interest in this approach, but still relatively few of this literature is aimed at the Commercial Real Estate sector.

The methodology of both Tu et al. (2004) and Nappi-Choulet and Maury (2009) are derived from Pace et al. (1998). Tu et al. (2004) investigates spatial dependence in the office market of Singapore from 1992 through 2001 using a Bayesian spatial-temporal autoregressive model extension (B-STAR). The results indicate that by allowing for spatial dependence in the Hedonic methodology, the model is able to capture both the marginal effects of Hedonic properties as well as spatial dependence of the market. Furthermore, Bayesian estimation is able to correct the heteroskedasticity problem. As such the

structural prediction accuracy increases five to ten percent, therefore outperforming conventional Weighted Leased Squares with only a limited number of variables. The model is said to reduce the problems caused by infrequent trading of commercial properties. Nappi-Choulet and Maury (2009) on the other hand, look at the Paris property market from 1992 to 2005. Through the comparison of three models, one with endogenous temporal heterogeneity within a Bayesian setup, they add that size of spatial dependence strongly differ according to the position within the property cycle and that spatial drifts in the intercept should not be neglected. The findings thus suggest that spatial-temporal dependence is extremely relevant for transaction price indices.

Chegut et al. (2015) however, questions whether spatial dependence is an important factor in Hedonic models and price index construction. Geltner and Bokhari (2008) already noted that spatial dependence may not be a significant factor in Commercial Real Estate as segmentation across commercial property markets is very high. Hence, Chegut et al. (2015) compare a standard Hedonic price models and Spatial Temporal Autoregressive (STAR) models for six of the largest office markets globally. Results indicate that spatial dependence in these markets is statistically significant, but economically of limited importance. This is thus in contrast to the previous works of which a possible interpretation lies in the period studied, as the latter includes the Global Financial Crisis while the other studies included more stable periods.

Hierarchical Trend Models (HTM)
The hierarchical trend model (HTM) is a time-series model that could be categorized as a 'semi-parametric' model extension of the Hedonic Price Model. In this model the impact of space and time on transaction prices is modelled in an advanced and flexible manner. That is, some parameters vary over time while other parameters are constant. In the Netherlands, this model is widely used for the valuation of millions of houses for property tax purposes for over two decades without any significant problems (Francke, 2008). The price index produced by the HTM measures the price developments of a standardized house of constant quality over time. District and property specific trends are modelled as deviations of the common trend by random walks.

Francke and Vos (2004) first proposed the use of a structural time series model within a state space framework for estimating the evolution of the coefficients of the repeated sales model. This model is said to make better use of the available information by smoothing out some of the variability in the data and optimally extracting the signal from the noise. It is shown that, especially for small housing market segments, the HTM can produce price indices that are more accurate and up-to-date than traditional approaches. Francke et al. (2014) is the only application found to apply the HTM to the Commercial Real Estate sector. They provide a first attempt to construct a Commercial Real Estate price index for the Dutch market based on the database of 'Stichting Vastgoeddata (StiVAD)'. With only 40 observations for offices but positive results, the study is bound to be extended.

3.1.3.2 Advantages
Sufficient research has shown that spatial regression methods can improve traditional regression in the real estate sector. McCluskey et al. (2013) conclude that these models include the best of both interpretability and prediction accuracy and that the mass appraisal is likely to continue on this track. Additional advantages over HPM include:

- The existence of spatial autocorrelations that might be present in the data presents issues for traditional regression but are an opportunity for spatial regression models. If spatial patterns are observed, there are ways to benefit from this misspecification. That is, the intercorrelation between observations can help us improve our value estimates as we can find out which

properties are related to extract implicit value information. This is similar to the comparable sales method.

- With spatial-temporal regression models we can control for local markets conditions through a continuous function. This counters the limitations of traditional regression where arbitrarily chosen control variables for time and location are one of the main causes of different results in literature. With a spatial-temporal model we can apply a single model that allow for differences over space and time.

3.1.3.3 Limitations

The Spatial-Temporal Regression models like any other modelling has both advantages and limitations. Some that are relevant to the application of Commercial Real Estate are as follows.

- Many studies (e.g. Colwell et al., 1998) show that the cashflows that commercial properties can generate are among the main determinants of the value. This would suggest that comparable properties are not only a function of time and space, but additionally of the income generated, among other factors. Hence, the weight matrix of a Spatial Regression Models might be misspecified as the closest comparables might be more distant, while near and recent transacted properties might not be comparable at all.
- Price effects are difficult to extract as spatial predictions are not built into the model as coefficients as is the case with traditional Hedonic regression through for example dummy variables (McCluskey et al., 2013).
- Spatial models make various assumptions that do not always hold or require arbitrary choices from the researcher. Examples are that stationarity for temporal models and chosen time-invariant variables that might show differences over time (Francke, 2017).
- The models require a specialists' knowledge as these methods deal with advanced spatial- and time-series modelling. Furthermore, the methods require serious hardware. By comparison to a single or even a few models on sub-markets, GWR would for example have to run the model for every data point in the sample.

3.2 Machine Learning AVM Methods

In this section, popular Machine Learning methods are evaluated and compared based on their applicability for the automated valuation of Commercial Real Estate. These methods might provide a better prediction accuracy than the more traditional Hedonic regression as they can 'find' connections and non-linear patterns within the data. It is however well known that this often goes at the expense of the interpretability of the model. We therefore explore existing research on AVMs to evaluate the advantages and limitations of the most applied Machine Learning algorithms. The techniques covered in this paragraph are the Artificial Neural Network (ANN) and three Tree Based models; Classification and Regression Tree (CART), Random Forest (RF) and Gradient Boosting (GBM).

But before getting into these methods, it is important to point out a common misunderstanding. That is, in many literatures the term Machine Learning is used interchangeably with Artificial Intelligence (AI). While connected, both in fact represent different concepts. First coined in 1956 by John McCarthy, "AI involves machines that can perform tasks that are characteristic of human intelligence". Current AI is not yet intelligent. Rather it is a collection of techniques which cleverly apply maths to a specific domain, known as *narrow AI*. Examples are prediction algorithms for image and speech recognition tasks that are trained with huge amounts of (labelled) data.

Machine Learning on the other hand is simply a way of achieving AI. One could obtain AI without Machine Learning, but this would require to write an unbelievable amount of complex code. So instead, we can apply Machine Learning as a way to 'train' the algorithm such that it can learn 'how to' without being explicitly programmed. Various algorithms have been developed as we saw in *Figure 3-1* that learn in different ways. Broadly, these can be categorized into supervised learning, transferred learning, unsupervised learning and reinforced learning (Ng, 2017). When people however talk about Machine Learning, in general they refer to supervised Machine Learning which is most applied as is also the case in this thesis. Domingos (2015) summarizes the definition of (supervised) Machine Learning as follows:

> *"Every algorithm has an input and output: the data goes into the computer, the algorithm does what it will with it, and out comes the result. […] Machine Learning turns this around: in goes the data and the desired result and out comes the algorithm that turns one into the other."*

So, in order to apply Machine Learning to the task of property valuation, we thus require large amounts of data on both the input - the features of the properties – and the output – the value or historical transaction prices of properties in the market under investigation. The supervised Machine Learning algorithm can then, and only then, be processed to 'learn' from the data. The result is an algorithm that can be applied to estimate the value of properties based on the specific property features of new data.

The recent interest and growth in applications of Machine Learning has been ignited by the exponential growth of quality data and computational power. Although the focus has been on 'big data', maybe just as important is the generation of 'new data' (Anselin, 2017; Mullainathan, 2018). Think about satellite images, smart-cities, social media and GPS phone data. These data not only can help us create new levels of prediction accuracy, but also introduce whole new areas of research. This development in data has been accompanied by improvements in the underlying algorithms and techniques. However, relatively few of these have yet been applied in the academic literature related to real estate value estimates. The following sub-paragraphs discuss the algorithms that do.

Table 3-2 ■ Overview Reviewed ML Literature

	Market studied	Period studied	Sample size	HPM	ANN	CART	RF	GBM
Tay et al. (1992)	Singapore	1989	1,055	X	-	n/a	n/a	n/a
McCluskey et al. (2013)	Ireland	2002-04	2,694	X	-	n/a	n/a	n/a
Din et al. (2001)	Switzerland	1978-92	285	-	X	n/a	n/a	n/a
Peterson et al. (2009)	USA	1999-05	46,467	-	X	n/a	n/a	n/a
Lin et al. (2011)	USA	2009	33,342	-	X	n/a	n/a	n/a
Worzala et al. (1995)	USA	1993-94	288	X	X	n/a	n/a	n/a
Zuranda et al. (2011)	USA	2003-07	16,366	X	X	-	-	-
Kaoka (2002)	Finland	1993-97	small	n/a	X	X	n/a	n/a
Fan et al. (2006)	Singapore	1997-98	5,589	n/a	n/a	X	n/a	n/a
Onur et al. (2009)	Turkey	2007	1,049	n/a	n/a	X	n/a	n/a
Anitpov et al. (2012)	Russia	2010	2,848	n/a	n/a	-	X	n/a
Kok et al. (2017)	Netherlands	2011-16	5,018	n/a	n/a	n/a	X	X
Sangani et al. (2017)	USA	2016	90,275	n/a	n/a	n/a	n/a	X

Note: All the literature below is aimed at the automated valuation of the Residential sector as no Commercial Real Estate studies are found. The sample represents the total sample size. 'X' indicates whether a model outperforms the other models indicated by the '-', 'n/a' means that the model was not used in that particular research. Note that different specification can mean that more than one model performs best in a particular study.

3.2.1 Artificial Neural Network

Artificial Neural Network (ANN), or simply Neural Nets, are the algorithms where most news comes from. ANNs are an umbrella term for many different algorithms (see *Figure 3-1*). The development of ANNs has been inspired by attempts to replicate the way that we humans learn. The model consists of input and output layers, as well as one or more hidden layers that transform the input into something that the output layer can use. ANNs are excellent tools for finding patterns which are far too complex or numerous for a human programmer to extract and teach a machine to recognize. Briefly, the system assigns weights to connections between processing elements that are determined based on the patterns in a presented dataset. This way they are capable of adapting or 'mimicking' arbitrary and unknown functional forms with a arbitrarily specified degree of precision (Hastie et al., 2008)[4].

3.2.1.1 Real estate literature

ANNs have a relatively rich history in the property valuation literature. In recent years, ANN modelling techniques have become a serious alternative to and extension of more conventional property value modelling approaches. Many researchers have thus made the comparison between different AVM methods to investigate their potential. Abidoye and Chan (2017) critically reviews all relevant articles that adopt ANN for property valuation from 1991 through 2015 and highlight the strengths and weaknesses of the technique. The most relevant literature is reviewed below with in *Table 3-2* an overview of the samples and models used. We can immediately see that none of the ANN studies are aimed at the Dutch real estate sector and none to the Commercial Real Estate sector.

Tay and Ho (1992) were one of the first to investigate the potential of ANNs for the valuation of real property. They compare the performance of the ANN model with a traditional HPM for the valuation of Residential apartments in Singapore. The study finds that the ANN model outperforms the HPM in terms of an absolute error of 3.9 percent against 7.9 percent, respectively. Hence, they conclude that the ANN model is an easy-to-use, black-box alternative to the HPM. McCluskey et al. (2013) on the contrary do not find that the ANN performs better than the HPM. Even more so, they find that the GWR model provides far superior performance, which indicates that spatial dependencies are present within the Northern Ireland Residential market which the ANN model cannot cope with.

Din et al. (2001) also aimed to compare various real estate valuation models, but specifically look at the manner in which they respond to different environmental attributes scenarios. The Baseline model is a standard HPM that includes ordinal variables to measure environmental quality. They find that ANN models, which are non-linear per se, exhibit a similar general form of the price indices to the HPM. However, the price behaviours of the models' features notably differ depending on the input choice of the environmental variables. Variable selection thus still plays an important role in ANNs.

Most studies are aimed at the market in the USA. Peterson and Flanagan (2009) use a large sample size of more than 46 thousand observations and seven variables to find that the ANN performs better than the HPM. Lin and Mohan (2011) find similar results and add that the ANN is a more reliable and cost-effective method that the HPM for property valuation. Worzala et al. (1995) on the other hand find the HPM outperforms the ANN in most cases and thus does not support the previous findings. Similarly, Zurada et al. (2011) conclude that the ANN is not superior to other techniques such as HPM and Tree-based models and should be applied with caution. Literature thus provides contradictory results and application of this 'black-box' technique by the appraiser community should be done with caution.

[4] For a very practical example of the ANN to get more insight in the working of the model for the task of property valuation we recommend the article of Mora-Esperanza (2004).

3.2.1.2 Advantages

The application of advanced ANN algorithms for the task of property valuation are mentioned to have the following, non-exhaustive advantages:

- In line with most Machine Learning algorithms, the ANN model can handle nonlinear relationships that exits between property values and its attributes. It can also cope with time varying or uncertain attributes and thus be seen as a kind of 'model-free' regression that can learn from relationships not otherwise known (Abidoye & Chan, 2017).
- Where the HPM are characterized by a degree of subjectivity that often lead to different results, the ANN model does not require human (besides hyperparameter input). A priori theory is not needed to analyse the nonlinear or complex nature that might exist between the input and output. Success is largely due to its ability to discover complex structures that was not specified in advance (Mullainathan & Spiess, 2017). In addition, it can achieve high precision even when the dataset contains outliers and/or missing data.
- The very appeal of Machine Learning is that it can deal with high dimensional data (Anselin, 2017). The flexible forms allow us to fit varied structures of the data that can be 'big' and 'new' (or both). Traditional regression on the other hand would require some variable selection procedure that influence the results.
- Machine Learning algorithms are now easy to use. Convenient packages in R or Python can fit most Machine Learning algorithms with a few lines of code. This however also pose the risk that they are applied naively or their output is misinterpreted.

3.2.1.3 Limitations

Many papers however resonate issues with ANN for three obvious shortcomings:

- ANN models lack explanation facilities for their knowledge. The knowledge is buried in their structure and weight which makes it difficult to extract rules. ANNs are therefore often called a 'black-box' learning approach as relationships between input and output are difficult to interpret by humans. Recently however, various initiatives have been taken to improve their interpretability and is developing (Abidoye & Chan, 2017).
- Continuing on this fact, Machine Learning algorithms do not cope well with parameter estimations. Even when these algorithms produce regression coefficients, the estimates are rarely consistent due to correlations between variables. Thus important to take away is that at the time Machine Learning is only relevant in prediction tasks (Mullainathan & Spiess, 2017).
- ANN models require large amount of data and a long time to train due to the voluminous array of alternatives. The number of neurons used in the hidden layer of a model can aid the retrieval of an improved result, but unfortunately there is no consensus in the literature as regards the number of hidden neurons to be included in an ANN model. In addition, there is the danger of overfitting (Abidoye & Chan, 2017).

3.2.2 Tree-Based Models

Tree-Based Models are relatively straightforward statistical pattern recognition algorithms that are widely used in Data Science communities such as Kaggle with ensembles of decision Trees among the most often used algorithms in prediction modeling (Jeremy & Bowles, 2012). This section describes the few studies that apply either the Classification and Regression Tree (CART), the Random Forest (RF) or Gradient Boosted Trees (GBM) for real property valuation. These models all have a decision Tree at its core but differ in the way such Trees are combined to increase accuracy of the value estimates.

Briefly, the CART algorithm performs numerous tests to derive the best sequence for regressing and predicting the dependent variable based on rules for the independent variables. These tests identify the best 'splitters' which selects those variables and their interactions that are most important in determining the value estimate. The RF continues on this principle. However, it combines multiple decision Trees through a bagging approach that averages out noisy and unbiased data that reduces variance. The rationale behind the RF is that a combination of decorrelated Trees increases the prediction accuracy. Similarly, GBM combines multiple Trees that convert a set of weak learners into a single strong learner. However, the algorithm combines a boosting algorithm with Gradient descent to do so. In this iterative process, each successive Tree is built for the prediction residuals of the preceding Tree. *Chapter 4* provides more information about these methodologies.

3.2.2.1 Real estate literature

Over the past few decades, substantial technical literature on decision Trees has emerged. However, in the real estate research domain the number of studies that have attempted to take advantage of decision Tree techniques is only limited and to our knowledge even non-existing for Commercial Real Estate. As mentioned, *Table 3-2* provides an overview of studies covered in this review.

Fan et al. (2006) adopt a decision Tree approach to analyse the prices of Singapore's public housing market using 4.912 resale data. The article demonstrates the usefulness of this technique to identify important price attributes, relationships and finally to predict the values. Similarly, Onur and Hasan (2009) analyse empirically major factors that affect housing prices for the Residential market in Istanbul, Turkey. The built Trees show what homebuyers are more concerned about in terms of characteristics and could in a similar way be applied to Commercial Real Estate to identify what investors value.

Kauko (2002) wrote a dissertation about the value of location that explores and evaluates both the ANN and CART model as an alternative for traditional HPM. Of particular interest is how the different locational, environmental, and social factors impact housing market segments and house prices. He highlights that choosing the best model involves various trade-offs as each has their own strengths. The ANN represents a state-of-the-art technique that does not differ that much from HPM (with spatial extension) other than its non-linearity and computational effort. Trees on the other hand are more suitable for exploration of value effecting factors.

Zurada et al. (2011) provide one of the most extensive comparative study found in the AVM literature. Using a data sample of more than 16 thousand houses, three non-traditional regression-based methods and three Machine Learning methods and compared under various simulation scenarios. The result indicates that non-traditional regression-based methods perform better in most simulations, especially when the data is homogeneous, whereas Machine Learning methods perform better with heterogenous data. This indicates that for the heterogeneous Commercial Real Estate in this thesis, AI techniques might offer better performance. Antipov and Pokryshevskaya (2012) highlight that their research is a first attempt to apply the RF technique for the mass appraisal of real estate. They find that their RF model outperforms other models such as the HPM, kNN and ANN in Saint-Petersburg, Russia. In addition, the article proposes a CART-based technique that detects segments in which the model under- or overperforms.

Kok et al. (2017) are to our knowledge the only ones who investigate the potential of GBM for the automated valuation of multifamily assets. Using data from Geophy – a company that develops commercial AVMs - they state that their models outperform manual valuations in terms of average errors between the estimated and transacted prices. Interestingly, even so-called 'hyperparameters' such as crime rates or distance to music events are shown to have the potential to increase the predictive accuracy of AVMs and thus are all incorporated in their models.

3.2.2.2 Advantages

The often-mentioned advantage of the Tree-based algorithms is that they are more flexible than the often-criticized stringent assumptions of a standard Hedonic framework. (Mullainathan & Spiess, 2017; Varian, 2014; among others) cover the most important advantages and limitations.

- Tree based Machine Learning tends to work well where there are important non-linearities and interactions. The model is non-parametric thus does not require any assumption about underlying distribution of values of the predictor variables which makes the model easier to construct and explain. This important feature saves the developer time which would otherwise be spent determining whether variables have the right distribution, making transformation, among other things. Not to mention errors that are likely to be made during this process.
- Decision Tree algorithms are highly automated algorithms, even within the field of Machine Learning, and are easy to understand and apply. That is, little effort is required for data preparation to generate accurate value estimates. The algorithms cope with data that is missing variables, categorical and numeric, non-standardized, high dimensionality and can even perform feature selection for heterogeneous data. The model can thus, in contrast to the HPM, predict accurate values even when important variables are unknown (missing data).
- The algorithms supports a wide range of loss functions. Furthermore, the Tree's results can be visually represented as an interpretable tree-like structure and variable importance measures can also be derived. This provides a certain degree of interpretability in contrast to e.g. ANN.
- Even if Trees may not improve on predictive accuracy compared to e.g. linear models, it can still reveal some interesting insights about the data that are not apparent from a traditional of other Machine Learning techniques. For example, it might reveal a premium for older (monumental) buildings, whereas a linear model might assume a negative relationship.

3.2.2.3 Limitations

Although the Tree approach is an excellent tool to search patterns in the data, we also note that it suffers from several limitations.

- There is very little insight and control in what the model does (black-box approach). A single Tree can offer some insight about how predictors interact, but a forest of thousands of Trees cannot be easily interpreted. One can at best try different parameters or look at the variable importance that shows which variables cause the largest improvement in prediction accuracy.
- Learners can easily create overly complex Trees that do not generalize the model well. This is also known as 'overfitting' and is particularly threatening when the data are noisy. Careful tuning of the few parameters is thus key but makes the model slow to train (but fast to predict).
- Most decision Tree algorithms can only identify the single most significant splitter at a node. Even though other independent variables may produce a significant but relatively weaker effect on the value at the node, this influence cannot be analysed simultaneously within the built Tree framework. In other words, it is difficult for decision Tree algorithms to carry out a full consideration of the effect of independent variables. Since the algorithms are 'greedy' - meaning that it chooses the variable split that minimizes the error - Trees can have a lot of structural similarities and thus high correlation in their predictions.
- The regression model does not predict beyond the range of the training data. In addition, the RF algorithm is better at handling classification than regression problems. It cannot determine the precise continuous nature of the regression that is often required.
- Interestingly, Trees tend not to work all that well if the underlying relationship is actually linear.

3.2.3 Other Machine Learning Models

The models reviewed above are selected based on their potential for the task of automated valuation of Commercial Real Estate. It is however important to keep in mind that many more methods exist with each their own advantages and limitations. The choice is in general related to the amount and quality of the data available, among other reasons. Examples of algorithms not covered in this thesis are the self-organizing map, Fuzzy Logic, Analytical Hierarchy Process, Rough Set Theory, kNN, Monte Carlo Simulation, among others. In the bundle of articles of D'Amato and Kauko (2017), these methods and more are covered. An expected trend to come is the construction of hybrid models that combine the best of both traditional regression and Machine Learning. Ortec Finance (2017) provides a first look at such models. Mullainathan and Spiess (2017) highlight that an ensemble of methods consistently outperform individual, single models in terms of prediction accuracy. These topics are however out of the scope of this thesis.

3.3 Methods Evaluation

We've seen in this chapter that although many literatures are aimed at the price determination of real estate, surprisingly few are aimed at the Commercial Real Estate sector, with studies aimed specifically at individual property transactions or machine learning techniques to be practically non-existent. Nevertheless, due to the increasing amount and quality of the data in this sector this type of work is stimulated in years to come. Basically, we are at a point where the Residential sector was a few decennia ago. Hence, valuable lessons can be learned from this closely related field where data is already abundant. The broad literature discussed in this chapter provides some priori knowledge about the most promising methods for the automated valuation of individual Commercial Real Estate properties with in *Table 3-3* an overview of the final evaluation of the models covered.

Choosing the optimal model involves various trade-offs as every methodology has its own strengths and limitations. Among the more traditional models covered in the first section, the Hedonic Price Model (HPM) seems to offer most potential as a Baseline regression as this methodology is the most established in academic literature which provides priori information about significant determinants and potential model specifications. Nevertheless, we've also seen that this methodology suffers from some issues such as spatial-temporal dependencies that affect the value estimates of the model. The Geographically Weighted Regression was found to counter most of these limitations and provides one of the best prediction accuracies in studies applied to the Residential real estate sector. In line with the second hypothesis of this thesis we therefore extend the traditional hedonic framework to control for such dependencies. Note however that in the Commercial Real Estate sector these dependencies might be influenced by more factors than distance only.

Comparative studies that cover Machine Learning techniques often find mixed results, but in general we observe that the ANN outperform traditional regression and other Machine Learning algorithms. Nevertheless, this goes at the cost of the interpretability of the model which makes it unsuitable for the practical application in this thesis. Tree-based models are found to provide similar prediction accuracies but still provide some degree of interpretability of the results and are therefore applied in this thesis. Random Forest (RF) is one of the easiest to use algorithms but is also known to have difficulties with capturing the continues nature of the value estimates. Gradient Boosting (GBM) on the other hand is more difficult to apply due to its many hyperparameters, but can better cope with out-of-sample regressions. As only a few studies have made use of tree-based techniques for Automated Real Estate Valuation it makes the findings in this thesis all the more relevant.

The main difference between the traditional regression and the machine learning methods can be found within the assumptions of the structure between its characteristics and value. Hedonic regression assumes a linear specification which makes results easy to interpret and explained to non-statistic minded audience such as real estate investors or most appraisers. This statistical framework also makes it possible to test the assumptions made in a coherent manner and to formally compare competing models with each other. An often-mentioned objection against these parametric models is however that they are too rigid. Supposedly, too much structure is imposed on these models making them not 'flexible' enough to work with which causes lower prediction accuracy in certain scenarios. This is in sharp contrast to more data-driven RF and GBM models. The drawback of these models are that the results are much more difficult to interpret or to formally put to the test. Moreover, these data-driven methods demand a great many observations in order to calibrate the model. In the next chapter we dive deeper into the Methodologies of these two traditional and two machine learning methodologies.

Table 3-3 ■ Evaluation Traditional and Machine Learning Methods

	Considers individual property / location characteristics	Individual impact of factors observable (explainability)	Suitable for modelling and monitoring market trends	Can cope with non-linearity	Can provide individual confidence measure (reliability)	Can distinguish Spatial-Temporal Correlations
Traditional AVMs						
Comparable Sales (CSM)	Yes, explicitly and implicitly.	No	No	No	Yes	No
Hedonic Price (HPM)	Partly, only ones that have been quantified.	Yes	Yes, After conversion to HPI.	Partly, only when specifically modelled.	Yes	Partly, only through included coefficients.
Spatial-Temporal Extension (ST)	Yes, explicitly and implicitly.	Partly, more difficult to interpret	Yes, After conversion to HPI.	Partly, only when specifically modelled	Yes	Yes
Machine Learning AVMs						
Neural Network (ANN)	Yes, only ones that have been quantified.	No	Partly, only when specifically modelled.	Yes	No	No
Decision Tree (CART)	Yes, only in one Tree.	Yes, relative variable importance.	No, results unstable.	Yes	No	No
Random Forest (RF)	Yes, only ones that have been quantified.	Partly, relative variable importance.	Partly, better for classifications.	Yes	No	No
Gradient Boosting (GBM)	Yes, only ones that have been quantified.	Partly, relative variable importance.	Partly, only when specifically modelled.	Yes	No	No

Note: This table provides a summary of the evaluation of the most common AVM algorithms applied in the real estate literature for the purpose of appraising office properties. ST, RF and GBM seem to be optimal models to achieve the aim of this thesis. The methodologies applied in this thesis are shown in Bold.

4. Methodology

As we have seen in the previous chapter no single modelling technique is perfect with each having their own advantages and limitations. Based on the criteria of explainability, reliability and predictability, three models were chosen that offer most potential for generating automated valuations of individual Commercial Real Estate properties. These are the well-established Hedonic Price Model that is used as the Baseline estimate and the Tree-based Machine Learning algorithms: Random Forest and Gradient Boosting. In addition, we propose a new methodological framework under the name Comparable Weighted Regression that extends the Geographically Weighted Regression which was shown to provide optimal performance in the Residential valuation literature. Briefly, the new model allows for discontinuities in space and takes temporal effects into account in a rather straightforward manner, that is through the selection of close comparables. The chapter closes with an elaboration of software used to construct the AVM model for practical application.

4.1 Baseline Hedonic Regression Model

We first establish a reference model that serves as a starting point to judge the new Comparable Weighted Regression methodology and Tree-based Machine Learning models against. In the remainder of this thesis this model will be referred to as the 'Baseline model'. The remainder of this paragraph addresses two topics, namely the Baseline model specification and potential issues with this model.

4.1.1 Baseline Model Specification

The previous chapter already discussed the basic idea of Hedonic Theory with the prices of a good being explained by its underlying characteristics. As the actual prices are not known, Hedonic models shadow these prices by providing a quantification of the unobservable willingness one is expected to pay for the range characteristics. The simplest form of said relationship is linear where one unit growth in variable *x* (e.g. size) is expected to increase price by its estimated coefficient. The total price is then the sum of all such characteristics included in the model that are weighted by the price for each characteristic. Unexplained factors are included in the error term which are aimed to be minimized.

Although linear relationships are the easiest to interpret they are not always realistic. There are different forms possible of the functional relationships between the dependent and individual independent variables with a (semi)-logarithm functional form being the most common in real estate. Francke (2017) covers four reasons for this. First, the model specification should be in a multiplicative form as it is more natural to describe the differences in value between two equivalent properties in percentages rather than absolute value. Second, the value of a property is less than proportional to its characteristics. Consequently, this also means that not only the dependent variable might take different functional forms, but also the independent variables. In this thesis this is tested based on eyeballing multivariate plots and trial-and-error of models. Third, by using the logarithm of the transaction price the squared residuals are approximately minimized. Fourth is that the resulting residuals are closer to normality than the residuals of the transaction prices itself as the influence of outliers is less severe. Nevertheless, the deciding factor remains the increase of predictive accuracy.

The Baseline Hedonic regression model applied in this thesis is thus of the (semi)-logarithmic form:

$$ln(P) = \beta_0 + \beta_1 X_1 + \cdots + \beta_n X_n + \varepsilon \tag{4.1}$$

Where P in this vector notation is the price of the subject property and $X_1,\ldots,X_n$ are the set of characteristics (of which some are dummy variables). $\beta_1,\ldots,\beta_n$ represent the coefficients of each characteristic and ε the unexplained noise remaining. Note that X are often represented as a transformed function. The best model fit is calculated through Ordinary Least Squares (OLS). Stepwise regression (both ways) in combination with a trial-and-error process of variable selection and outlier analysis are applied to arrive at the final Baseline specification that provide the most robust Market Value predictions. Note that some of the independent variables are also in a transformed (logarithmic) form. Considerable effort has thus been made to construct an optimal Baseline model for comparison. The prediction accuracy is tested through various cross-validations.

To control for time, the model includes yearly time dummies that like any other dummy variable reflect price changes relative to the base category. To control for location, we include three locational variables: city category, centrality within the city, and district type. The coefficients of the variables are thus assumed to be static over time and space, which might not be a realistic assumption as we will later see. The *next chapter* provides more information about the variables.

4.1.2 Issues with Baseline Specification

There are several issues with the use of the Hedonic Price Theorem for the automated valuations of real assets. Besides the often-unrealistic stringent assumption mentioned, there are several issues that need special attention in the modelling phase. The list of issues addressed herein include:

a) Model (mis)specification : results are dependent on model specification.
b) Multicollinearity : lack of independence among the predictor variables.
c) Endogeneity : variable relate to both left and right side of the formula.
d) Heteroskedasticity : variability of error ε is unequal across variables.
e) Spatial-Temporal Dependencies : underlying process is not stationary over space-time.

Ad. a) Model (mis)specification

As we have seen in the literature review, a familiar problem with HPM is that economic theory does not provide priory information towards the specification of the model. It highly depends on both the availability of data and the manner in which it is processed by the researcher. These steps are often arbitrarily chosen and have been mentioned to be one of the main reasons why results in literature tend to deviate (e.g. Chegut et al., 2015, p. 29).

Misspecification of the model can lead to misleading interpretations of the true effects of the variables. This applies to both the size of the intercept and coefficients. They can be either too high or low which in turn leads to unreliable predictions. Furthermore, the standard errors of the coefficients are affected and the explained variance is lower. Omitted variables are among the main causes of model misspecification together with incorrect variable forms (Greene, 2011, pp. 97-98). When the right variables are added to the model specification, the explained variance begins to increase and the coefficients approximate those of the underlying model. Henceforth, we conclude that it is all the more important to specify in detail what variables are used and what transformations they have undergone in order to allow for reproducibility of the results[5].

Ad. b) Multicollinearity

Multicollinearity occurs if correlation exists among the independent variables in the model specification. For example, the size of a property is highly correlated with the total rental income that can be generated. Symptoms that can occur when estimating the coefficients in the presence of such multicollinearity are listed by (Greene, 2011, pp. 129-130):

- Small changes in data produce wide swings in the coefficient estimates.
- Coefficients may have high standard errors and low significance levels even though they are jointly significant and the R^2 of the regression is quite high.
- Coefficients may have the 'wrong' sign or implausible magnitudes.

There exist several methods for avoiding the pitfalls of data that exhibit multicollinearity. Among them is the use of step-wise regression, ridge regression, or alternative model formulations. In the empirical work of this thesis, care was taken to avoid excessive multicollinearity by choosing variables that are not highly correlated with one another. Remaining signs of multicollinearity were tested by the Variance Inflation Factor (VIF) with a rule of thumb of maximal ten.

[5] Castle et al. (2009) provide guidance how to choose the best regression equation.

Ad. e) Endogeneity

The first hypothesis of this thesis investigates whether the third dimension of lease related factors in addition to the first and second dimension of location and building factors, are an essential part to be included in the automated valuation of Commercial Real Estate. Few studies include lease factors in the price determination literature which might relate to the fact that such factors potentially introduce endogeneity to the Hedonic model specification. Endogeneity occurs when there is correlation between a variable *x* and other terms of the model; most commonly with both the dependent variable as well as the error term. In this case, problems arise as there is something that is related to the dependent variable that is also related to our independent variables hence causing biased estimates. For example, when we include rents as independent variable in our Hedonic price model we might introduce a term that correlates with both the price as well as with information remaining in the error term of the model specification. That is, if we assume that rents are set based on the transacted price and the independent variables which might not be the case in practice. As no test are available for endogeneity, theory and practical knowledge provide our guidelines.

Ad. c) Heteroskedasticity

According to the assumptions of the linear regression models, the error term is supposed to be independent and identically distributed (i.i.d.) and thus homoscedastic. In reality this assumption is often violated when fitting a Hedonic price function to real property sale transactions (Stevenson, 2004). This occurs as sub-populations often have different variabilities hence the errors are likely to be heteroskedastic. But what is the harm of this? It has been established that even if there is heteroskedasticity present in the errors, OLS provides unbiased estimates of the regression coefficients (Borst, 2007, p. 132). This is good news for the prediction, however, the problem arises with making statistical inferences about the coefficients. There is usually no way to know whether the variance estimates are too high or too low. Also, the familiar inference procedures based on the F-test and t-distribution are no longer appropriate.

There are a number of methods to deal with heteroscedasticity. The first is to add additional explanatory variables to the model to remove the condition. The second method is to use an estimation technique and variance estimators that are appropriate when the errors are independent and not identically distributed. In this thesis, we observe first signs of heteroskedasticity through eyeballing the residual plots. As Real Estate prices are well-known to show heteroskedasticity (Winson-Geideman et al., 2018), we use two Breusch-Pagan tests, one studentized and one non-studentized, to formally test this. In the presence of heteroskedasticity, we apply White's correction (White, 1980) to the standard errors to see whether the previous estimated variables show changes in significance and coefficients. If not the case, we have additional evidence for the validity of our model.

Ad. d) Spatial-Temporal Dependencies

Continuing with heteroskedasticity, correlations among the error term that are a violation of the key assumptions of OLS might also be caused specifically by spatial and/or temporal misspecification of the model. We've already seen that Spatial-Temporal autocorrelation is likely to be present in real estate as closer things tend to be more alike than those far apart (Borst, 2007). Interestingly, we observe a shift in the validation literature where researchers abandon the idea of trying to include all locational and temporal influences into the model specification as coefficients. Instead they seek to incorporate these dimensions more directly into the regression through new methodologies. Examples are the Geographically and Temporal Weighted Regressions (GWR/TWR) and Spatial-Temporal Auto Regression models (STAR). Our Baseline regression on the other hand still only includes traditional dummy controls for location and time with all the issues that come with them. In line with the first hypothesis of this thesis we therefore propose extension of the Baseline regression model that does consider these spatial-temporal effects by the model itself. This method is covered in the next section.

4.2 Comparable Weighted Regression Model

The previously described Hedonic price model is the most common statistical modelling technique used to determine the value of real property. A limitation of this approach however is that the true spatial and temporal effects are difficult to capture through proxy variables such as dummies for districts and years. Furthermore, the model unrealistically assumes that these effects are constant over space, that is, one model fits all. Geographically Weighted Regression (GWR), the statistical technique described by McCluskey et al. (2013) to have the most potential for AVMs aimed at Residential properties, counter these limitations by allowing the coefficients to vary over space. The GWR results in a set of local coefficient estimates derived more from near than distant neighbours.

This rationale seems plausible for Residential real estate where local market conditions largely determine the value of a house, think about recent sale prices in a neighbourhood with homogeneous properties that likely have similar values. For Commercial Real Estate however this might not be as straightforward for two reasons. First, Commercial Real Estate is much more heterogeneous than Residential real estate. Neighbouring properties might not be comparable at all due to unique characteristics or differences in the cashflows generated. Second, low transaction volumes in certain regions might result in the fact that the nearest neighbour(s) to a property are actually far away.

In practice, valuers then also not only select their comparables based on the physical distance between the properties, but also look at (other) location, building and lease characteristics in combination with local market conditions. It is not uncommon that for a comparable for an office building near Amsterdam central station an office building near Rotterdam central station is used. With the CWR approach we aim to allow these discontinuities while at the same time providing a way valuers can add value to the AVM by evaluating the weights of the comparable on which the results are dependent. The better the input of comparable properties is, the more the model can implicitly derive value information and the more likely the AVM prediction is closer to the actual Market Value.

4.2.1 Step 1: Determining Comparability Score

The first step of this two step procedure statistical model is to select the most appropriate set of comparables upon which to base the valuation of the subject property. We do however not use a cluster of the most comparable properties to derive the value from but construct a weight matrix with each field representing a score that relate to the degree of comparability between properties. Close comparables get a higher score than less comparable properties and thus form a distribution of weights that are used as input to the weighted (least squares) regression. Basically, this means that coefficients are derived more from close than distant comparables with the degree relative to the difference in weights. Nevertheless, we still use all observations in our dataset that provide us with valuable degrees of freedom and reduce effects of potential outliers.

The weighting scheme of the CWR is thus a function of an arbitrarily chosen selection of variables. In addition to the proximity to the subject property used exclusively in the GWR, the following variables are used to derive the comparability weights: transaction date to capture temporal effects, lettable floor area to capture the size of the building, year of construction as a proxy for building quality and type, (indexed) theoretical rental income per square meter to capture the cashflows generated, vacancy percentage and Weighted Average Lease Expiry to represent associated risk, and city category, centrality, and district type to capture the type of location, and the walkscore, distance to station and highway access and leefbaarometer proxies to capture the quality of the location. The importance of each variable relative to the other variables can be manually adjusted. The standard weights are based on a discussion session with three experienced valuers of which the results are shown in *Table 4-1*.

Table 4-1 ■ Comparable Sales Relative Weights

Weights	Variable	Formula	Weights	Variable	Formula
$W_1 = 3$	Distance	$DISTANCE_y$	$W_6 = 1$	Rental Income	$\lvert TRI_y - TRI_i \rvert$
$W_2 = 3$	Transaction Date	$\lvert YEAR_y - YEAR_i \rvert$	$W_7 = 1$	Lease Term	$\lvert WALE_y - WALE_i \rvert$
$W_3 = 2$	Lettable Floor Area	$\lvert LFA_y - LFA_i \rvert$	$W_8 = 2$	Walkscore	$\lvert WALK_y - WALK_i \rvert$
$W_4 = 2$	Construction Year	$\lvert BUILT_y - BUILT_i \rvert$	$W_9 = 1$	Leefbaarometer	$\lvert LBM_y - LBM_i \rvert$
$W_5 = 5$	City Category	$CITY_y = CITYi$	$W_{10} = 5$	Centrality	$CENTy = CENT_i$

Note: The total sum of weights is 25 with each variable contributing w/25 to the total comparability score. All variables used are standardized z-scores, except for City Category and Centrality. These two get a value of 0 if true or else 1.

The total weight that represents the degree of comparability between the observations and the subject property is then derived from the sum of the comparability of each variable. But before we can add these together we need to standardize the scale of each variable as the difference in e.g. rents is on a different scale than Year Built. For this task we use z-scores which measures the number of standard deviations from the mean or in our case the value of the subject property[6]. Further research would benefit from exploring the optimal distribution between variables and total scores. Machine Learning methods can aid to obtain these values but is out of the scope of this thesis.

4.2.2 Step 2: Weighted Least Squares Regression

The second step of the two-step procedure is to use the weighting scheme obtained in the first step to improve the estimates of the regression framework. The CWR is a relatively simple technique that extends the traditional regression framework by allowing local rather than global parameters to be estimated. These local parameters are not one dimensional as is the case with the GWR with space, but multidimensional where local basically means clusters of similar properties.

The vector with one weight per observation is derived from the comparability score among properties. That is, the function allows a continuous surface of the parameter values base on similarities among a range of property characteristics. In result, we expect there to be similar coefficients among clusters of a certain type of property that show high comparability. Note that traditional Hedonic regression can thus basically be seen as a special global case of the CWR in which the parameter surface is assumed to be constant with all observations receiving the same weights.

In the calibration of the CWR model it is assumed that observations that are more comparable to the subject property have more of an influence in the estimation of the models' coefficient than do data that are less comparable. In essence, the equation measures the relationships inherent in the model for a property that shows similar traits. In other words, an observation is weighted in accordance with its similarity to the subject such that the weighting of an observation is no longer constant in the calibration but varies based on the type of subject property. The way in which these weights vary dependent on the type of property distinguishes the CWR from traditional WLS and GWR.

Important to note is that the CWR not only produces localized parameters, but also localized versions of all regression diagnostics including the goodness-of-fit measures. These can be helpful to e.g. distinguish types of properties the model performs worst, among other things.

[6] In order to better communicate the degree of comparability between properties to the user, we transform the total summation of z-scores to a comparability score between 0 and 100.

4.3 Machine Learning Models

This section is devoted to the Machine Learning methodologies models applied in this thesis. Like predictions with traditional regression, we are interested in understanding the conditional distribution of some variable *y* given variable *x*. The main difference however, as we have seen, is that Machine Learning algorithms are able to find functions instead of being specified in advance. The more data is available, the better Machine Learning can find patterns in the data and the better the model can predict the Market Values of individual properties out-of-sample based on its features.

The selection of Machine Learning methods applied is based on the evaluation results of *Chapter 3*. Here we concluded that Tree-based algorithms offer the most potential to support appraisers in their valuation process as these models are easy to apply, offer some interpretability, and are shown to provide excellent performance out-of-sample. In particular, the 'bagged' extension of the Decision Tree; the Random Forest, and the 'Boosted' extension; the Gradient Boosted Tree offer much potential in theory. Hence, these methods are discussed in *Paragraph 4.3.2* and *Paragraph 4.4.3*, respectively. As both models have Decision Trees at its core, this method is first briefly explained in *Paragraph 4.3.1.* However, this method is not applied for valuation purposes for the aforementioned reasons. In line with the subsidiary goal of this thesis the methods are described more intuitively without going too deep into the mathematics such that this thesis can be used as guide with the actual AVM model.

4.3.1 Classification and Regression Tree (CART)

Although not used as a predictor of the Market Value in this thesis, the Decision Tree algorithm Classification and Regression Tree (CART) is at the base of both the Random Forest and the Gradient Boosted Trees. Hence, it is useful to know a little bit about the underlying rationale before we can move to the more advanced Tree algorithms.

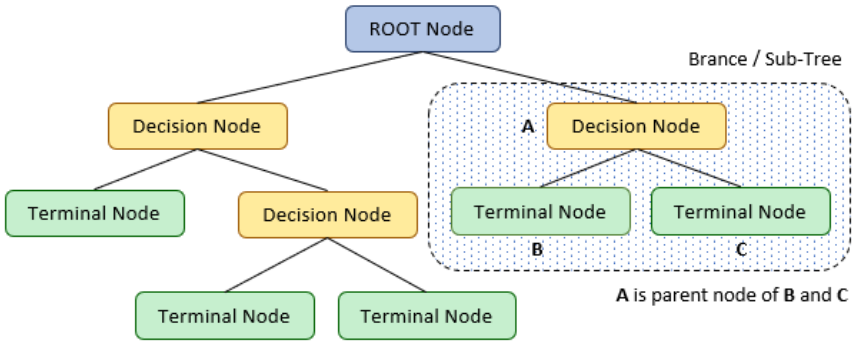


Figure 4-1 ■ Decision Tree Definition Scheme

The ultimate goal of CART is to construct (or "grow") a Tree that leads to a good out-of-sample prediction. This Tree can be represented by a flow-chart structure where each internal node denotes a test on attributes, each branch represents an outcome of the test and each leaf node or terminal node holds a class label (see *Figure 4-1*). The prediction function takes the form of a Tree that splits in two at every node. The top most node in a Tree are often called a root node and the predicted values of the model are obtained at the terminal nodes. For regression Trees, the estimated value is then simply the mean or average of all the instances of the training data that fall in that particular region. If new unseen property thus falls within a particular terminal node we can make a prediction through the mean value in this node. Basically, we could recreate this process as a linear function where each terminal node represents a product of dummy variables.

It is important to realize that this method introduces discontinuities in the estimates with the number of steps equalling the number of terminal nodes. When no stopping criterion is specified the process continues until the error of the training set is not reduced anymore with additional splits and/or the terminal nodes only have a few observations; every observation would practically be its own dummy. In result, the model would be a (near) perfect fit for the training data, while unseen (out-of-sample) data would contain large errors as the model does not generalize well. This is as we now know called *overfitting* and is a familiar threat of Machine Learning algorithms.

A solution to overfitting is to specify a splitting criterion that will stop the splitting process once it is breached, known as *regularization*. Instead of choosing the overall best Tree, we instead set a cost on complexity and choose among those that satisfy a certain criterion, e.g. maximum Tree depth. In general we see that the shallower the Tree the worse the in-sample fit, but the less likely it will overfit as the idiosyncratic noise of each observation is averaged out (Mullainathan & Spiess, 2017). The most popular approach to tackle overfitting is called *pruning* where a large Tree is first grown, but is later reduced by removing sections of the Tree that result in better explanatory power out-of-sample (Varian, 2014). With the right level of regularization, we thus aim to benefit from both the flexible form generated by Machine Learning without overfitting the model. Various techniques exist to find the optimal level of regularization – also known as tuning the algorithm – which are discussed with the Random Forest and Gradient Boosting methodologies.

The decision how the algorithm makes the strategic splits thus affects each Trees accuracy. But how do Trees decide where to split? Multiple functions can be used to decide how to split a node in two (or more) sub-nodes which are based on a predefined loss-function. The four most commonly used functions are the Information gain, Gigi-index approach, Chi-Squared and Reduction in Variance (Smith, 2017). The details of these algorithms goes beyond the scope of this thesis as they involve quite a lot mathematics. Most of the hard work is done automatically by basic statistical software and thus we don't need to concern ourselves with this. More importantly, it is more useful to grasp the idea of the basic concepts.

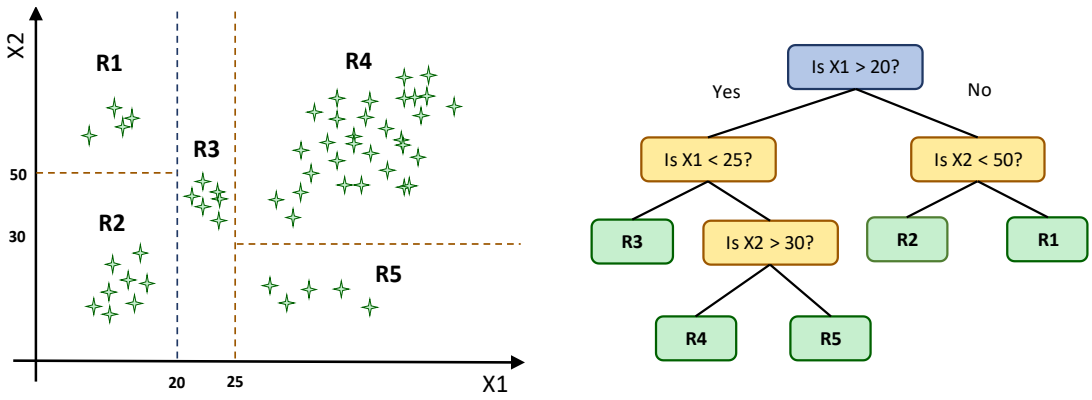


Figure 4-2 ■ Simplified Tree Splitting Process Example

Figure 4-2 shows a simplified example which can be done by hand, but with an underlying rational that is the same for even the most complicated loss-functions based algorithms. The creation of sub-nodes is based on increased homogeneity of resulting sub-nodes. In other words, can it group the data based on similar traits. Most Tree algorithms split the nodes on all available variables and then selects the splits which results in the most homogeneous sub-nodes (Smith, 2017). This process is finished when a stopping criterion has been reached or no splits are found that improve results.

We've thus seen that it can be quite easy to build, understand, interpret and visualize a CART model which could be used for all kinds of decision analysis. Nevertheless, the model also comes with many disadvantages as seen in Paragraph 3.2 and is unfit to be used for individual property valuations. Two techniques that improve on the limitations are called Bagging and Boosting (Hastie et al., 2008). Briefly, *Bagging* chooses some subset of the data at random with replacement and combines the results of the individual Trees to create a single predictive model. The Random Forest Model is an extension of this bagging method. *Boosting* on the other hand works in sequential manner were each Tree is fitted on a modified version of either the dataset or the model specification. Both are thus based on converting multiple weak learners (decision Trees) into one strong one. Interestingly, these models revolve around adding randomness to the data. This might seem paradoxical at first, but as we will see this randomness turns out to be a helpful way of dealing with the overfitting and increases the prediction accuracy significantly[7].

4.3.2 Random Forest

The Random Forest algorithm resolves some of the limitations of decision Trees by increasing the number of Trees built from the training data. Averaging the results of multiple Trees increases the model performance and creates finer grain predictions than a single Tree. However, instead of simply bagging the *N* number of Trees, each Tree is fitted on a bootstrap sample of the original training set and are constrained to a randomly chosen subset of variables such that sub-Trees are less correlated. It is a simple tweak but results in a significantly better predictor than simple bagging. The cause can be found in the fact that Trees are greedy algorithms, that is, the model always makes the choice that seems to be best at that moment and result in similar predictions. Varying the variables included thus allow for better estimation of the underlying functional form and importance of individual variables (Varian, 2014).

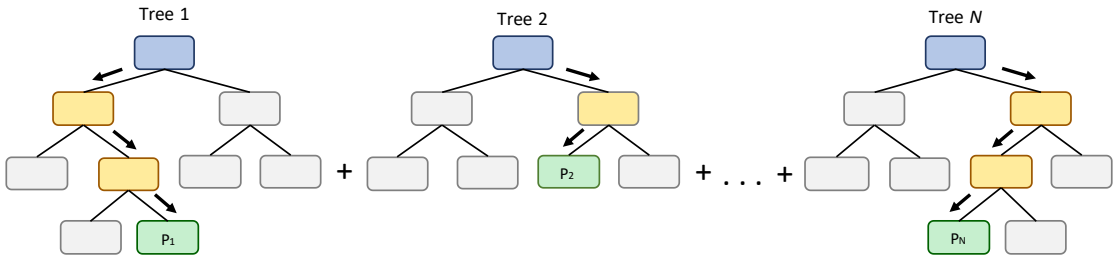


Figure 4-3 ■ Simplified Random Forest Scheme

In CART, when selecting a split point, the learning algorithm is allowed to look through all variables and all variable values in order to select the most optimal split-point. The Random Forest algorithm on the other hand changes this procedure in such a way that the learning algorithm is limited to a random sample of features to choose from. The number of features that can be searched at each split point must be specified as a parameter to the algorithm. For regression a good default is $m = p/3$ where m is the number of randomly selected features that can be searched at a split point and p is the total number of input variables (Breiman, 2001). Nevertheless, in this thesis the optimal hyperparameters are tuned through cross-validation techniques.

The following steps are taken by the Random Forest Algorithm to compute the value prediction (Hastie et al., 2008, p. 588). *Figure 4-3* provides a scheme of the steps described:

[7] For reproductive purposes it is paramount to choose a particular seed to replicate this randomness.

1. Assume number of observations in the training set is *N*. Then, the bootstrap sample of these *N* observations is taken at random but with replacement.
2. If there are *M* input features total, a number $m < M$ is specified such that at each node *m* variables are selected at random out of the *M*. The best split of these *m* is used to split each node. The value of *m* is held constant while we grow all the Trees in the forest.
3. Each Tree is grown to the largest extent possible and there is no pruning.
4. Predict new data by aggregation the predictions of the *n* Trees. That is, simply the average of each Trees prediction.

The regularization variables applied in this thesis to control the complexity of each individual Tree are the maximal Tree depth, the number of variables used in each Tree, the size of each bootstrap sample and the number of Trees generated[8].

4.4.3 Gradient Boosted Regression Tree

The Gradient Boosting algorithm is the second machine learning algorithm applied in this thesis. It has the potential to provide a more accurate estimate of the Market Value as the Random Forest has been often been criticised to provide better performance for classification problems than regression problems (e.g. Graczyk et al., 2010). Similarly to the Random Forest algorithm, Gradient Boosting converts a set of weak learners into a single strong learner. However, they differ in the way they create the weak learners during the iterative process and how results are added together. Gradient Boosting doesn't modify the sample distribution, to train on a newly sampled distribution, but instead the weak learner trains on the remaining errors (so-called pseudo-residuals) of the strong learner (see *Figure 4-3*). The algorithm namely involves three elements:

1. A loss function to be optimized
2. A weak learner to make predictions
3. An additive model to add weak learners to minimize the loss function

Ad. 1) Loss function: The goal of the model is to minimize the loss function. As we have seen, different loss functions exist but in general represents a quantitative value for the difference between the actual and predicted value. In regression the (root) mean squared error is the standard and thus also used in thesis. *Paragraph 6.1* provides more information about this measure and how it is calculated.

Ad. 2) Weak learners: Decision Trees are used as the weak learner in Gradient boosting. Specifically, regression Trees are used that output values for splits and whose output can be added together. This allow subsequent models' output to be added and "correct" the residuals in the prediction. We've already seen that Trees are constructed in a greedy manner choosing the best split points based on purity scores like Gini or to minimize the loss. Hence it is common to constrain the weak learners that can be tuned.

Ad. 3) Additive model: Trees are added one at a time, and existing Trees in the model are not changed. A Gradient descent procedure is used to minimize the loss when adding Trees. That is, after calculating the error or loss a new Tree is constructed based on Gradient decent that reduces the loss (i.e. follow the Gradient). The output for the new Tree is then added to the output of the existing sequence of Trees in an effort to improve the final output of the model. In other words, by

[8] The Caret package in R provides a built-in grid-search to find the optimal hyperparamaters for the Random Forest. The 'ranger' function is applied. See: *topepo.github.io/caret/available-models.html*

parameterizing the Tree and then modifying the parameters we move in the right direction that reduces the residual loss. The training stops once loss reaches an acceptable level or no longer improves.

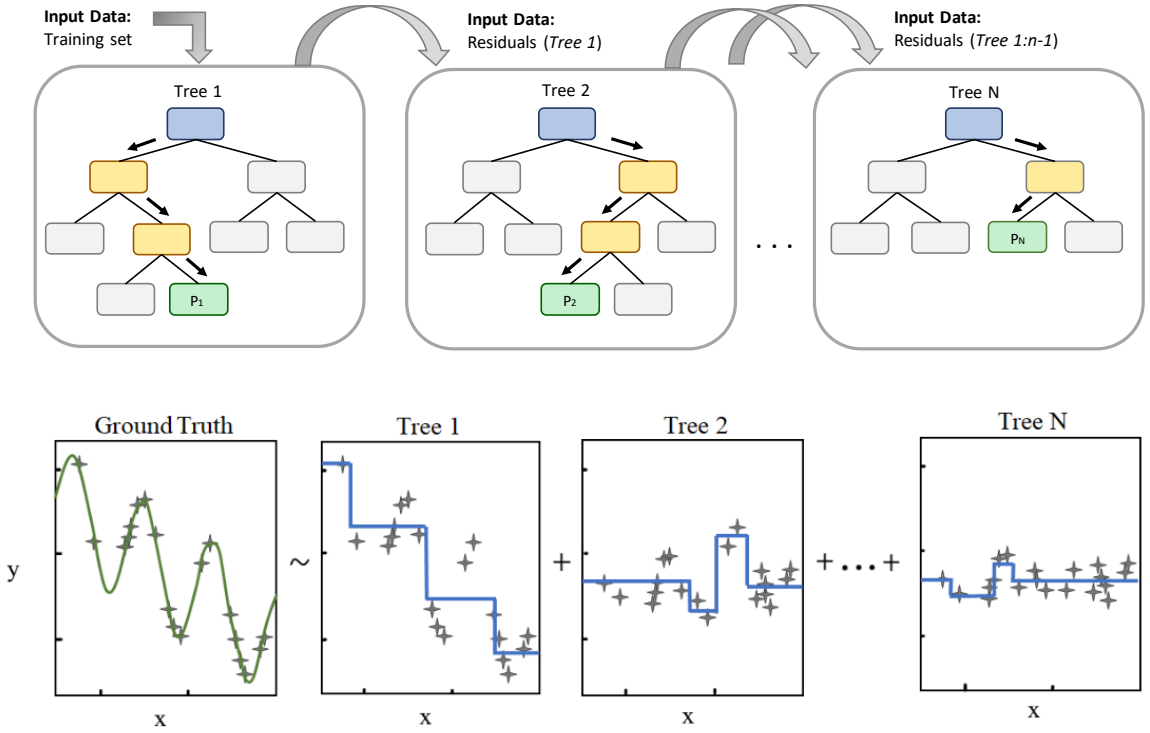


Figure 4-4 ■ Gradient Boosting Scheme and Example Plots

Extreme Gradient Boosting (XGBoost)

Gradient Boosted Trees are known for their high prediction accuracy. The algorithm has become increasingly popular in the Machine Learning community Kaggle and has offered many award-winning solutions in Data Science[9]. In particular the extension of the Gradient Boosted model XGBoost has risen in popularity over the last year. The main reason of its popularity is that it counters the main weakness of Gradient Boosted Trees, that Trees can only be modelled in sequence. XGBoost on the other hand allows the Trees to be built in a parallel fashion allowing to exploit multiple cores and increase the speed significantly. This speed is especially useful when tuning the many hyperparameters and large datasets. In addition, XGboost uses a more regularized model to control for overfitting, which is said to provide better performance out-of-sample (Chen & Guestrin, 2016).

Hyperparameter Tuning

As mentioned, any Machine Learning model is comprised of two types of parameters. First are the type of parameters that are learned through the Machine Learning algorithm. The second are the parameters that we can choose for regularization such as the maximal Tree depth or number of randomly selected variables. Optimizing these so-called 'hyperparameters' can offer a lot of room for improvement out-of-sample as we Gradient boosting is a greedy algorithm and overfits a training dataset quickly. It can therefore benefit from regularization methods that penalize various parts of the algorithm and generally improve the performance of the algorithm by reducing overfitting.

[9] See https://github.com/dmlc/xgboost/Tree/master/demo#machine-learning-challenge-winning-solutions

4.4 Application with R and R-Shiny

The software package R is used for the data preparation process, value predictions through means of traditional regression and Machine Learning, and development of the AVM application. R is a software package that is free, has a large user community for support, packages that work well together, and together with python is the most used for data analysis. In order to build a user-friendly application that can run the fairly complex models constructed in this thesis, we turn to the R-Shiny package in R-studio. What R-Shiny does is it creates an application through R-code that is shown as dynamic output in a webpage format. This package thus allows developers to create interactive web applications such that users can interact with the model as well as download and manipulate the data without needing knowledge of coding[10].

The main reason for choosing R-Shiny to construct a first version of the AVM model is that one can build these applications with R-code in a relatively short amount of time without any knowledge of HTML or JavaScript. Furthermore, no software needs to be installed to run the code as this can be done through an external server. Nevertheless, at the end of this project we also became familiar with the limitations of this software. Mainly that although the predefined settings of the application make it easy to develop the app, it limits the freedom given to developers to do anything beyond the standard. Furthermore, with increasing amounts of data, the R-language approaches its limits as it holds all objects in virtual memory. In the foreseeable future we probably continue with a combination of JavaScript as front-end and a combination of NodeJS and Python as back-end development. *Appendix D* provides a first look at the AVM application developed with R-Shiny.

[10] For learning data science, we recommend visit www.kaggle.com which has an active community that provides interesting cases with large datasets and where many members discuss their coding solutions.

5. Data and Descriptive Statistics

In most literature the data side of the research is skimmed over as the data and its preparation are relatively straightforward. However, with the ever-increasing amount of 'Big and New data' and new techniques to make this data useful we observe that the scientific community is spending more words on this part of the research than they did in the past. In this thesis, we extend this trend and put focus on details about the data munging steps as these are among the most crucial steps for the development of an AVM (see *Figure 5-1*). At the base of our analytical workflow are the data processing steps described by Winson-Geideman et al. (2018). Each section describes a step from gathering of the raw data to the final cleaned data ready for modelling. The R-code of this process is publicly available at *https://github.com/BasHilgers/Thesis_TUe_AVM*.

5.1 Data Gathering

In the first phase of our Analytical workflow we gather as much data as possible from a wide variety of sources and store it in a centralized database. The focus of this thesis is on individual transaction of office properties but more data than this is gathered which is not used in this research. Think about historical appraised values and transaction details. However, as the model evolves new ways can be found to make use of this data to further increase the accuracy. This thus means that not only ready to use structured data is collected but also unstructured data, such as valuation models and long text descriptions. This data can therefore best be classified as 'Big data' which Madden (2012) describes as: "too bit, too fast, or too hard for existing tools to process".

When dealing with AVMs or similar data-driven models it is important to realize that at start of development the data does not have to be perfect, just 'well-enough' with a model that makes good use of the data at hand while in its process improves by learning from its mistakes and collecting new data through its users. Related to this phase are companywide data collection plans with licenses, usage, security assessment, legal ramifications and procedures to be considered. These topics however are beyond the scope of this thesis[11]. In line with the 'Garbage-In, Garbage-out' principle it is paramount that together with the collected data as much as possible information about the reliability and sources is attached for evaluation purposes. Reliable data is one of the key steps that determines the final accuracy of the estimates and consequently the success of the model. Even more so since we are dealing with a relatively low number of observations which makes the model sensitive to outliers.

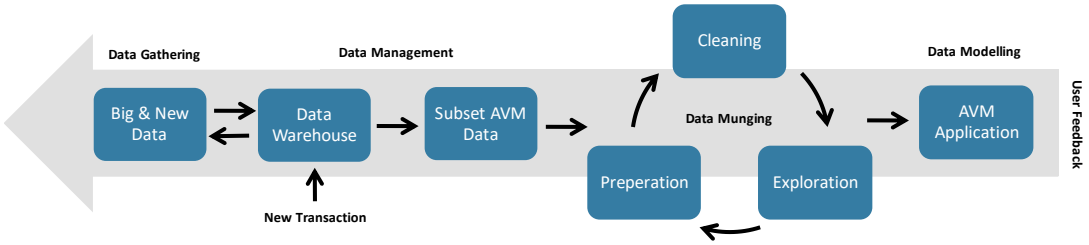


Figure 5-1 ■ Analytical Workflow of AVM Development

5.2 Data Management

The Data Management phase covers those operations and tasks that take the raw data and bring it together into a single location. The data are collected from the variety of sources that need to be related to each other. We therefore need to manage IDs, fieldnames, data structures, spatiotemporal formats, among other things. In general, we can distinguish two main activities: appending rows and arranging columns. The identifiers of the observations can be separated into the same categories as *Figure 1-1,* that is: Building, Location, Lease and Market including the transaction itself. The final database of this phase commonly goes by the name of a 'Data Warehouse' or 'Data marts' with questions such as software choice, IT architecture, storage capabilities, accessibility keys and pipelining codes for the continuous stream of new data to be considered. Again, these topics goes beyond the scope of this thesis[12]. In the final step of the Data Management, a subset is created with reliable and (mostly) complete observations that are used as input for the AVM analysis. An overview of these variables used for the AVM including their sources and percentage missing can be found in *Table 5-1*.

[11] See Winson-Geideman et al. (2018) for more information on these topic.
[12] For more info about data storage techniques and broader data mining, see Witten and Eibe (2005, p. 55).

Table 5-1 ■ Metadata about the AVM dataset

Variable	Source	Free / Confidential	Remark	Missing
Transaction Price	C&W	Confidential	-	0%
Transfer Date	C&W	Confidential	-	0%
Geocoordinates	BAG API	Free	Key on request	0%
Year of Construction	BAG API	Free	Key on request	0%
Year of Last Renovation	C&W	Confidential	-	NA
Lettable Floor Area	C&W / BAG API	Confidential / Free	Key on request	0%
Parking Spots	C&W	Confidential	-	NA
Building Height	PDOK 3D-hoogte	Free	Downloadable	0%
Energy Label	EP-online.nl	Free	Key on request	10%
Walkscore	Walkscore.com	Free	Key on request	0%
Leefbaarometer	Data.overheid.nl	Free	Downloadable	0%
Nearest Station	Google Maps API	Free	1,000 calls a day	0%
Nearest Highway	Personal GIS	Confidential	-	0%
Rental Income	C&W	Confidential	-	0%
Vacancy Percentage	C&W	Confidential	-	0%
Lease Term (WALE)	C&W	Confidential	-	0%
Rental Difference	C&W	Confidential	-	21%

Note: This table describes the Metadata (the data on the data) of the collected data from a variety of sources. Most data are obtained from Cushman&Wakefield (C&W) and is confidential. This data is enriched with various (semi)-public source data. The table also describes the initial percentage of missing observations per variable. NA is unclear, e.g. no renovation means not renovated or missing.

5.3 Data Preparation

The Data Preparation phase can be seen as the link between the managed data from our data warehouse and the initial analytical process. Briefly, it concerns the preparation of the variables (columns) and the observations (rows) that are used as input for the AVM. This step, like many, is an iterative process which is returned to both with the arrival of new data and throughout the analytical workflow. As we are working with incomplete Commercial Real Estate data, a lot of effort has been made to make the data complete. Some preselection excludes the use of several important value influencing factors such as groundlease and physical (quality) characteristics. In addition, if a single observation has limited information, these are labeled and removed from the dataset. The remaining variables are briefly elaborated with focus on their preparation[13].

5.3.1 Transaction Price and Other Critical Information

The AVM dataset has three mandatory variables that contain critical information about the dependent variable. These are the transacted price, the date of transfer and the geolocation. We've already seen that the dependent variable can best be taken as logarithm. The geolocation is included as spatial points based on longitude and latitude and the date of transfer is expressed in years[14]. Linked to this are the remaining features about the building, location, lease and market. Considerable effort has been made to fill all these incomplete yet reliable observations manually. The final AVM dataset consist of 979 office transactions between the years 2010 through 2018.

[13] For handling data preparation activities, we recommend the dplyr package: https://github.com/tidyverse/dplyr.
[14] Dates of transfer are collected as day-month-year format, but in this thesis only included as years.

5.3.2 Building Factors

Building factors significantly correlate to the price of a property (see *Figure 5-2*). Broadly speaking, we can make a distinction between size and quality. It is obvious that the floor size of the building is strongly correlated with the transacted price. For Commercial Real Estate however, instead of the total floor size of the building, we use the Lettable Floor Area as this is the space where revenues are generated. Other size related variables included are the number of parking spots and building height that might relate to a premium due to a certain building status. Missing values in building heights are filled by counting floors of images. The remaining variables do not contain missing observations.

We've seen that depreciation of quality is consistently significant across literature (Bokhari & Geltner, 2016). Unfortunately, the collection of Commercial Real Estate data in the Netherlands is still in its infancy with many missing observations [15]. A common alternative is to use the year of construction and/or renovation date as a proxy for the quality of the building. The rationale is that buildings that are built and/or renovated around the same period possess a similar depreciation rate and thus a similar quality. This of course does not always hold true, but in most cases provide reasonable estimates. We investigated three different forms of measurement and compared which provide most explanatory power to our model. These are dummies for similar building periods, the (effective) age variable that measures the number of years between the year (renovated) built and the year of transaction, and a spline function that creates a continuous non-linear function. In the end the dummy variables provided the highest prediction accuracy and is included in the final model specification. Energy labels are also considered as a building quality element that represents the 'greenness' of the building. Missing observations of this variable are median substituted based on building period groupings.

5.3.3 Location Factors

In terms of location related price affecting factors we include four indicators: type of location, amenities, liveability and accessibility. The *type of location* is measured through different variables in search of the one providing the highest predictive power. First, dummies of (combined) COROP regions are included as these are commonly used for analytical purposes. Second, we include indicators whether the city can be classified as large (Amsterdam, Rotterdam, Den Haag, Utrecht or Eindhoven), and/or whether the location is central within a city (within walking distance of an intercity station). Last, we distinguish different district types in line with C&W research reports (Office, Business, Mixed and Other). The level of *amenities* is determined by means of the Walkscore proxy. This tool gives a score per address based on the number and rating of the amenities within walking distance of 400 meters derived from different sources such as google maps and user input[16]. *Accessibility* is measured through distance to nearest train station (minutes walking) and highway access (minutes driving). Last, the *liveability* is measured through the 'Leefbaarometer' initiated by the Dutch government that provides a proxy for several quality measures of an area[17]. No missing observations are within the data.

5.3.4 Lease Factors

In line with the second hypothesis of this thesis, we investigate the importance of lease factors in the price determination of the value of Commercial Real Estate. Final variables included after a variable selection process are the Theoretical Rental Income (TRI) per square meter, percent of vacant LFA, weighted average lease expiry (WALE) excluding vacancy and whether the property was under or over-rented at time of sale. Broadly, these lease related factors can be categorized into cashflows generated

[15] See CoStar (2018) for an example how building factors are evaluated in practice.
[16] See https://www.walkscore.com/professional/research.php for more info.
[17] See https://www.leefbaarometer.nl/page/Help for more info.

with the property and the risk associated with these cashflows. As a proxy for *cashflow* generated (income) we use the TRI which was found to provide superior explanatory power over Rental Income and Estimated Rental Value. In addition, when correcting the total rents for rent that is paid additionally for parking we further increase our prediction accuracy. To capture the *risk* associated with the cashflow we include vacancy that proxies risk of no income generated, and a combination of WALE with Over- or Under rented to represent the risk associated with the certainty of income over time. No missing observations are present within the data as these are labeled and excluded from the AVM dataset.

5.4 Data Exploration

The next phase describes the Data Exploration that is often a process that is followed iterative with the previous Data Preparation and next Data Cleaning phase (see *Figure 5-1*). In this step we make sure we "get to know" the data prior to making key decisions on choosing analytical techniques such as data transformation and model specification. *Figure 5-2* below shows the correlation between continuous variables, *Figure 5-3* the distribution of transactions over time and *Figure 5-3* over space. *Table 5-2* provides an overview of the final list of variables and data formats. *Table 5-3* shows the descriptive statistics with visualizations in *Appendix A*.

Winson-Geideman et al. (2018) describe five considerations with Exploratory Data Analysis (EDA). The first and simplest is that it provides a verification of the previous data preparation steps. Second, possessing a better understanding of the basic "shape" or "form" of the data can help to redefine our data. Third, understanding patterns and distributions in the data form critical groundwork for outlier identification and/or missing data imputation. Once the basic shape of the data and the general relationship between fields are known, the selection of methods, test and procedures to handle outliers and missing data can be made more accurately and efficiently. A good idea during the EDA process is to not remove or clean any data observations, but rather that the EDA functions as an input to the final data cleaning exercise. This way different scenarios can easily be tested. Fourth, by identifying patterns and distributions in the data, EDA can assist in choosing modeling specifications and parameters. For example, spatial correlation can sign the use of more spatially-explicit methods such as Geographically Weighted Regressions. Finally, the EDA process can generate new insights on the current research questions or further research. By examining all dimensions of the data, often in combinations with other fields, relations that were previously undiscovered can be uncovered.

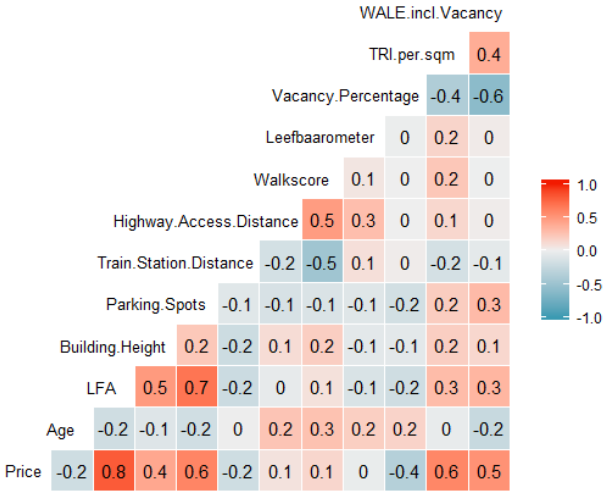


Figure 5-2 ■ Correlations (Pearson) between Variables

Table 5-2 ■ Variable Information

Variable	Type	Measured by	Transform	Sign
Transaction Price	RATIO	Net transaction price	Log	NA
Transfer Date	ORDINAL	Dummies with 8 levels	Dummy	NA
Geocoordinates	GEO	Longitude Latitude coordinates	Spatial	NA
Building				
Building Period	ORDINAL	Dummies with 8 levels	Dummy	(+)
Lettable Floor Area	RATIO	Size in sqm.	Log	+
Building Height	RATIO	Maximal height in meters	Log	+
Parking Spots	RATIO	Amount inside plus outside	Linear	+
Energy Label	ORDINAL	Dummies with 6 levels	Dummy	(+)
Location				
City Category	NOMINAL	Dummies large versus small	Dummy	(+)
Centrality	NOMINAL	Dummies central versus decentral	Dummy	(+)
C&W District Type	NOMINAL	Dummies with 4 levels	Dummy	NA
Walkscore	RATIO	Score from 0 to 100	Linear	+
Leefbaarometer	RATIO	Deviation from national average	Linear	+
Train Station Distance	RATIO	Minutes walking	Log	-
Highway Access Distance	RATIO	Minutes driving	Log	-
Lease				
Theoretical Rental Income	RATIO	Total in EUR per sqm	Log	+
Vacancy Percentage	RATIO	Amount as percentage of total LFA	Linear	-
Remaining Lease Term	RATIO	Weighted Average incl. vacancy	Linear	+
Rental Difference	ORDINAL	Dummies with 3 levels	Dummy	(+)

Note: This table provides an overview of the all variables used in the various models. Important to note is that not all variables are used in all models. The model specifications are defined in the methodology. The first column denotes the familiar variable names and the second column the data type. The third column provides additional information about the variable and the fourth column denotes the transformation applied to the variable. The last column shows the expected effect of the variable on the transaction price. The brackets indicate the expected effects of ordered dummy categories.

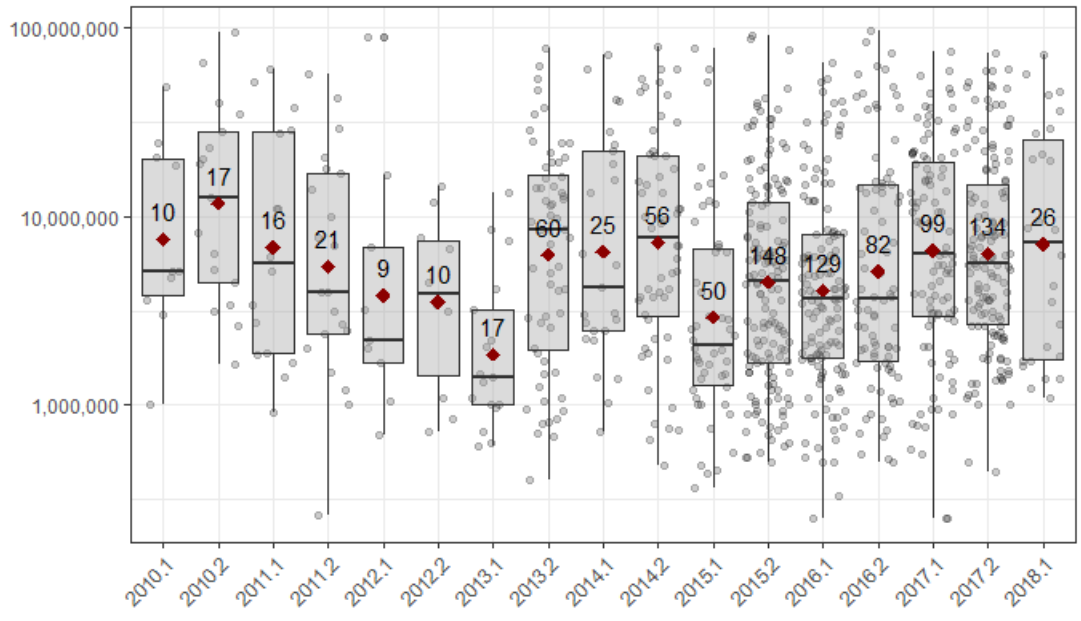


Figure 5-2 ■ Number of Transactions over Time

Table 5-3 ■ Descriptive Statistics

Variable	Mean	Std. Dev.	Min.	Median	Max.	Skewness	Kurtosis
Transaction Price (x100k)	118.4	163.8	2.5	48.0	974.8	2.3	5.7
Building							
Lettable Floor Area	7,134.9	6,616.1	400	4,811	47,322	2.0	5.0
Building Height	30.5	18.5	4.9	25.6	123.0	1.9	4.2
Parking Spots	94.2	119.2	0	60	1,050	2.7	11.1
Location							
Walkscore	73.8	19.7	18	77	100	-0.6	-0.5
Leefbaarometer	0.9	23.0	-73.0	1.9	79.0	0.1	0.5
Train Station Distance	22.7	19.8	0.4	17.0	143.8	1.9	4.8
Highway Access Distance	6.1	3.5	0.7	5.1	19.9	0.9	0.4
Lease							
TRI per sqm.	145.0	54.1	45.0	136.0	392.0	1.1	1.5
WALE incl. vacancy	3.6	3.7	0.0	2.5	21.5	1.5	2.6
Vacancy percentage	30.0	37.0	0.0	10.0	100.0	0.9	-0.7

Note: This table provides an overview of the summary statistics of the continuous (ratio) variables used in the full baseline specification.

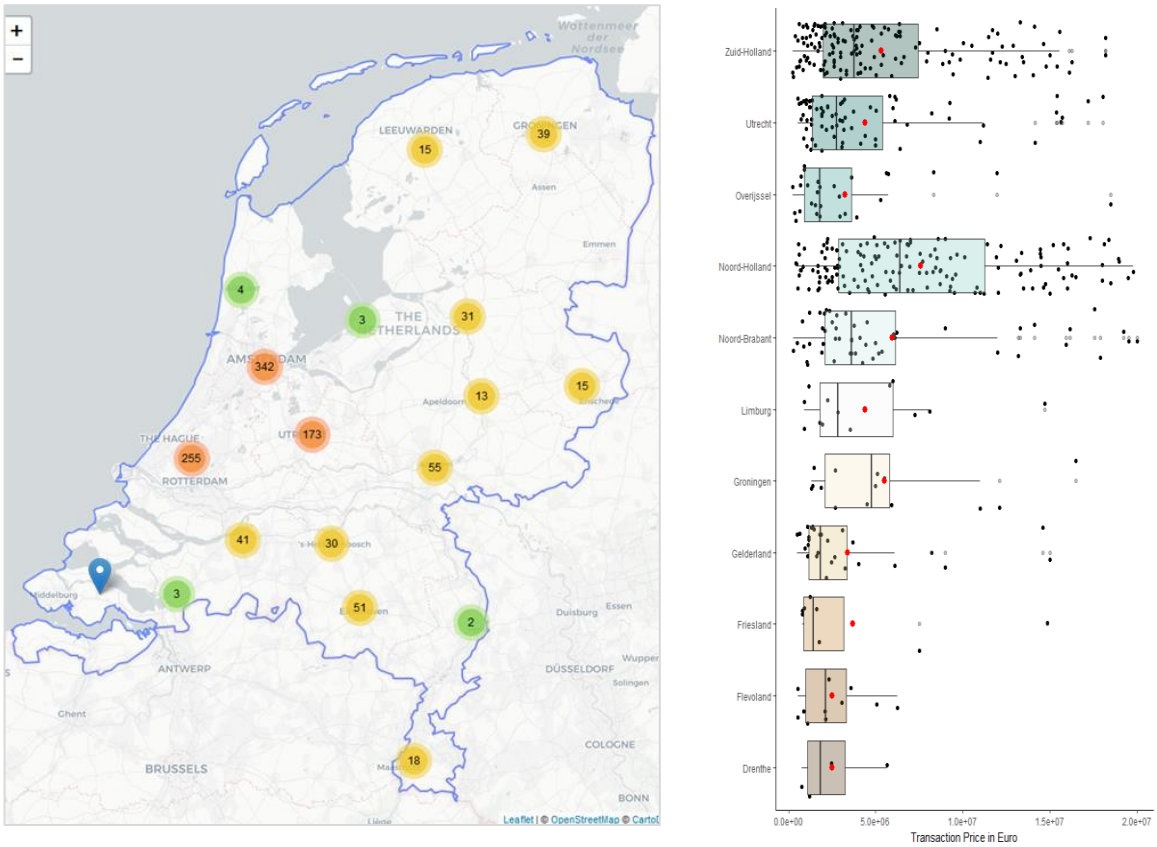


Figure 5-3 ■ Number of Transactions over Space

5.5 Data Cleaning

Up until now we have only identified and labeled outliers, errors and missing data. In this last Data Cleaning phase, we apply a range of criteria to treat the discordant values and prepare the final data for the modelling steps that follow. We developed a central strategy that handles the discordant values in an automated way such that the workflow can be repeated easily[18]. The result is a cleaned dataset that is ready for statistical testing and to make value predictions with. Important to realize however is that this phase is not done once but is an iterative process. Even after modelling, residuals can for example provide additional information about potential discordant values that needs to be handled. Documentation of the data provenance and rationale for removing any observation is a necessity.

But before getting into the actual cleaning steps, we discuss the crucial difference between data errors and outliers. *Data errors* are incorrect data that do not represent an actual, real world condition of the observation. These can occur for example through key-punch issues or measurement problems that skew the data. Within this thesis, with only around a thousand observations, data errors (and missing data) are checked and adjusted manually with the help of a customized dashboard in Access (see *Appendix D*). *Outliers* on the other hand are correct data values, but ones that are not representative of the phenomenon of study. They thus are technically valid observations, but they do not apply to a particular analysis at hand or are not representative of the underlying data generating process, findings and conclusions. Due to the heterogeneity of Commercial Real Estate and the small dataset, identifying and treating these outliers is difficult but all the more important. We therefore apply various techniques to label types of discordant values and apply sensitivity tests to investigate improvements in goodness-of-fit of our models. With repeated cross validations and transparent in the data provenance we aim to avoid overfitting the dataset during this process.

5.5.1 Univariate Discordant

The simplest form to spot discordant value is to look at one variable at a time. First, we checked the data manually for unlikely high or low values. The basic descriptive statistics and sorting columns provided a first indication whether the values are within expectations. Next, we used histograms to see if the distribution of the data in questions showed signs of discordance and labeled as informal outliers. Last, we applied more formal statistical measures on each isolated variable to identify outliers. As a rule of thumb, we identified univariate outliers if they surpass the cutoff point of three times standard deviation (see *Appendix A* for visualizations of each variable).

5.5.1 Multivariate Discordant

Single dimensions however fail to reveal discordant value in higher dimensions. For this reason, we apply some additional measures to identify these outliers. First, we use scatterplots and boxplots to informal identify outliers. We look for any values that are strongly different from the underlying trend. Different position of outliers can leverage the data in different ways and thus are labeled accordingly[19]. Next, we apply the so-called Mahalanobis' Distance to formally test outliers within multivariate situations. Much like the standard deviation this measure also looks at deviations but also includes covariances. Although this measure can handle more than two dimensions, in this thesis we limit ourselves to only two of which one is the dependent variable. *Appendix B* shows the outliers identified. Finally, the residuals of our models provide the last check on discordant values. We look at residuals per variables to informally identify discrepancies and use the formal cook SD to remove outliers.

[18] See *https://github.com/BasHilgers/Thesis_TUe_AVM* for more info about the code.
[19] For more information about leverage points, we refer to Winson-Geideman et al. (2018, pp. 90-91).

6. Quantifying Performance

An important step in the development of an AVM is to evaluate the prediction accuracy of the models' estimates. Without it models cannot objectively be compared to one another and the estimate also lacks confidence. For the AVMs applied to the Residential sector standards have been written by the IAAO (2013) with a new version available online that is not yet officially released. Standards for AVMs applied to the Commercial Real Estate sector are however not included but as uniform performance indicators are an absolute necessity in the market and with an increasing interest in AVM application these will most likely be introduced soon. For now, we rely on some of the most common approaches used in the scientific and data science community. In addition, the standard applied to the Residential sector are extended to the Commercial Real Estate sector to see whether these results hold significant meaning. This chapter aims to provides transparency in how prediction accuracies are measured and validated in this thesis.

6.1 Prediction Accuracy Measures

In order to communicate the performance of the prediction model we need some type of measurement that captures how accurate it can estimate values. In this research we have over a thousand observations on transacted prices *y* with relevant property features *x* through which we aim to compute a 'good' Market Value prediction given new values of *x*. 'Good' in the context of this thesis means it minimizes the difference between the actual transacted price and the value predicted by the model. In line with to financial decision-theory, one performance indicator might not tell the whole story as different measurements penalize different parts. It is therefore desirable to exploit multiple indicators that allow for objective comparison of the models. We thus apply, in addition to some qualitative measures, multiple quantitative accuracy measures to evaluate the AVM methodologies.

6.1.1 Root Mean Squared Error (RMSE)

In the traditional regression methodology, the RMSE is one of the most popular measures used as an indicator of prediction accuracy. This method differs from the Mean Absolute Error (MAE) in that it penalizes larger errors more than smaller ones by a squared term. For value predictions this feature comes in handy as large errors are especially undesirable. The Machine Learning models in this thesis are then also trained with this measure. *Equation 6-1* provides the formula with y_i as the observed Transaction Price for the i^{th} observation and $\hat{y}_i$ the Estimated Value with the model. *N* represents the total number of intances.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2} \tag{6-1}$$

6.1.2 Mean Absolute Percentage Error (MAPE)

Although the RMSE provides a good measurement to evaluate the performance of a model, especially when comparing models, the root term makes interpretation of the results difficult. To communicate the performance of the AVM model in a single number it is much more straightforward to use an absolute mean error measure (out-of-sample). However, as we have a broad range of values, we use the percentage difference of the error over the price of the property. This measure is known as MAPE which is formulated as follows:

$$MAPE = \left(\frac{1}{N}\sum_{i=1}^{N}\frac{|y_i - \hat{y}_i|}{y_i}\right) * 100\% \tag{6-2}$$

6.1.3 Coefficient of Dispersion (COD)

The Coefficient of Dispersion is the prediction accuracy measure standard in the Residential mass appraisal community (IAAO, 2013). It expresses the average deviation of the error ratio from the median as a percentage. A lower COD is thus desirable. It is however important to note that the value of this measure is highly dependent on the partitioning of the data and in result could show high variability which has been part of the critique on this measure (Borst, 2015). *Equation 6-3* provides the formulation where R_i denotes the ratio between the estimated value $\hat{y}_i$ and the transaction price y_i of the i^{th} observation and $\tilde{R}$ the median of these ratio.

$$COD = \left(\frac{1}{N}\sum_{i=1}^{N}\frac{|R_i - \tilde{R}|}{\tilde{R}}\right) * 100\% \tag{6-3}$$

6.2 Prediction Accuracy Evaluation

The ultimate goal of our predictive model is to generate optimal out-of-sample predictions for as many observations as possible whilst minimizing the amount of large errors. We have seen that it is often relatively easy to construct a predictor that works well in-sample but that fail miserably out-of-sample. This is a well-known problem in Machine Learning and has been given the definition of overfitting the data (see *Figure 6-1*). There are several ways to deal with this problem. Varian (2014) discusses three considerations:

1. Simpler models tend to work better for out-of-sample forecasts. In the case of Machine Learning algorithms various parameters exist that penalize models for excessive complexity. This is known as *regularization*.
2. It is conventional to divide the data into separate sets for training and testing. The training set is used to estimate a model and the test (holdout) set to evaluation how well the model performs on yet unseen data (see *Figure 6-2*).
3. The explicit numeric measure of model complexity for Machine Learning algorithms can be 'tuned' to produce optimal out-of-sample predictions. The standard way to test different values for these tuning parameters is (*k*-fold) cross-validation within the training set.

We can thus not only use cross-validation for tuning the hyperparameters but also to provide a reliable evaluation of our models' goodness-of-fit. That is, Cross Validation generate errors for unseen data that are similar to how the model would perform in real-world application. It is thus very suitable to find and communicate the overall models' performance and compare how well different models fit the data. In general, four main types of cross validation can be distinguished (Hastie et al., 2008, p. 241): The Holdout Method, *k*-fold Cross Validation, Leave one out Cross Validation (LOOCV) and the Bootstrap method.

In this thesis we use repeated *k*-fold Cross Validation to tune the hyperparameters of Machine Learning algorithms and to provide a quantification of the out-of-sample prediction capabilities. In addition, LOOCV is used to investigate how well the methodologies perform per property when maximizing the number of observations within the training set. Finally, a simulation is run similar to the LOOCV but with a test set of only the most recent data point. With the out-of-sample errors per estimate we can form a distribution per model that are used for analysis.

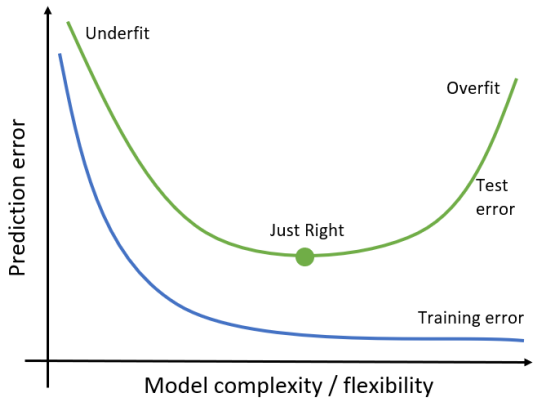


Figure 6-1 ■ Overfitting Example

Note: that as model complexity increases (e.g. increasing the maximum Tree depth), both the in-sample training error and the out-of-sample test error decrease. However, at a certain point we see that the test error starts to increase again. From this point the model starts to overfit the data. We aim with cross validation to find the 'just right' point.

7.2.1 *k*-fold Cross Validation

The most basic validation method is the holdout method. The disadvantage of this method is however that it requires a trade-off between the size of training versus test set. The way the data is split has effect on the error measure with different splits that can result in different values. A better approach is to use *k*-fold Cross Validation where we partition the data into k-folds and average the test results. *Figure 7-2* provides a scheme of the cross-validation technique used in this thesis. Notice that for Machine Learning an additional set is required to train the hyperparameters of the algorithm.

- *Training Set*: This set is used to build-up our prediction algorithm. By pairing the input with the expected output, the model tunes itself to the quirks of the training dataset.
- *Validation Set*: This set is used to compare the performance of the prediction algorithms that were created with training set using different hyperparameter settings. It tests different algorithm parameters to find the ones that provide the best performance out-of-sample.
- *Test Set*: At the last stage, we apply the optimal model from the iterative train-validation process to unseen (real-world) data. This way we can obtain an unbiased prediction error. Important is that after this measure is obtained the model cannot be tuned any further.

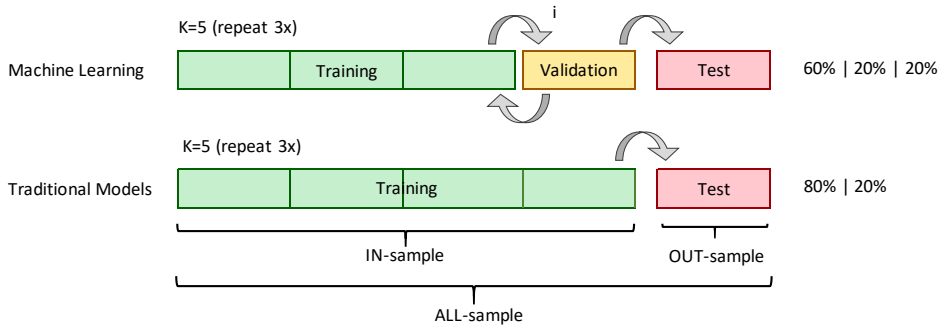


Figure 6-2 ■ (Repeated) K-fold Cross Validation Method Scheme

But how do we choose *k*? This is a trade-off between having a small *k* that matches a higher selection bias but lower variance in performance and a large *k* that has a smaller selection bias but higher variance in the performance. As our dataset is relatively small, we've chosen for a low *k* of 5 but in order to reduce the selection bias we repeat the method 3 times. The overall predicted error of the model is then the simple average of 3x5 error values.

7.2.2 Leave One Out Cross Validation (LOOCV)

LOOCV is an extension of the *k*-fold Cross Validation technique where the number of k-folds equal the number of observations *N*. This method takes a four-step procedure. First, we take one observation out of the training set. We then estimate a model with the remaining observations. Next, we predict the value of the observation that was taken out of the training set with the model. Last, we repeat this process for every single observation. The advantage of this method is that every point will at least be used once in the training and test set while the number of observations in the training set are maximized. The disadvantage is that it takes considerable more computation time.

As a final test of the model we perform a real-world simulation where the model is employed to estimate future data only. This test is basically an LOOCV that is trained with earlier data only compared to the new observation. After estimating this value, the observation is added to the dataset and the process is repeated. We perform this test on all transactions of 2018 in our dataset[20].

[20] We assume that the dummy of year 2018 equals 2017 when less than 30 observations are in this level.

7. Modelling Results

This chapter discusses the price modelling phase where we apply the different AVM methodologies to estimate the value of individual office properties in the Netherlands. We investigate whether a well-defined Hedonic Price model outperforms newer Machine Learning algorithms which have increased in popularity in both academia and practice in recent years. With just short of a thousand observations on actual transactions of commercial properties that show a lot heterogeneity, we might already expect that the strengths of Machine Learning algorithms cannot fully be exploited as these are often data-driven. We also propose an original Comparable Weighted Regression (CWR) method that is derived from the Geographically Weighted Regression (GWR) methodology, but instead of weighting distance only the CWR model weights comparability between observations on more than one dimension. We investigate whether we can reduce the issues with traditional Hedonic regression for this task and ultimately can improve the overall prediction accuracy. We finish with an evaluation of these models to find the one that provides lowest prediction errors while at the same time minimizing large errors which are especially undesirable.

7.1 Traditional Hedonic Regression

The first step in our price modelling phase is to establish a Baseline regression model. This model serves as a base for the comparison of this traditional Hedonic method against the newly proposed Comparable Weighted Regression method and Tree-based Machine Learning algorithms. Considerable effort has been made to develop an optimal model for this purpose. In this section we evaluate various model specifications on robustness of coefficients, significance and prediction accuracy in order to arrive to the final Baseline specification (see *Figure 7-1*). Both directional Stepwise regression aids us in this process. Furthermore, in line with the second hypothesis of this thesis, we investigate the importance of the inclusion of Lease related factors in AVMs applied to the Commercial Real Estate sector. *Appendix A* provides more information about the variables.

7.1.1 Building Factors

Lettable Floor Area

It is a known fact that the size of a property is significantly correlated with its value. For Commercial Real Estate this also holds true as this is the place where revenues of the business are generated. However, instead of the total Gross Floor Area (GFA) of a property we use the Lettable Floor Area (LFA) as it provides significantly higher explanatory power to the model. Since the price per unit LFA is not constant over space and time we need to include some type of control in the model to obtain unbiased estimates.

Model (1) in *Figure 7-1* provides a first estimate with the LFA as single predictor and some Spatial-Temporal control variables. As the relationship is curved with the dependent variable, we use the log transformed LFA. The resulting coefficient has the expected sign and remains relatively robust and highly significant over all model specifications. We observe with the adjusted R-squared that more than 80 percent of the variability of the data around its mean can be explained by this model with a MAPE of 53 percent. A (global) Moran's I test shows that significant spatial autocorrelation remains within the model, suggesting that we have yet to capture the spatial dependencies with the controls. It depicts a critical issue with the current methodology, that is that marginal effects of variables are not constant over space or time and how difficult it is to explicitly control for these and more effects.

Building Height

The height of a building is related to its representativeness and image. Fuerst (2007) shows that tenants are prepared to pay higher rents for higher floors in New York, and Koster et al. (2014) and van Assendelft (2017) find similar evidence in the Dutch office Market. We hypothesize that this premium is therefore also reflected in the price. Important to note is that building height is correlated with the LFA, however, as the correlation is less than 0.6 we assume that it is safe to use both in the model. The variable is log transformed.

Model (2-6) do not show significance at the conventional significance level of 0.05. Although the sign is as expected, the effect is relatively limited. The transformation to an indicator for Highrise buildings (6 floors or more) does not improve the results of the model. Reasons can be found in the fact that Highrise buildings are relatively low compared to other global markets. As a reference, the highest building in the Netherlands is 165 meter (Maastoren - Rotterdam), whereas the highest building in the United States is 541 meter (One World Trade Centre – New York). However, since the variable is significant with a significance level of 0.10 and the inclusion of the variable does seem to improve prediction accuracy of the model we include the building height in the Baseline model.

Table 7-1 ■ Traditional Hedonic Regression Results

Variable	(1)	(2)	(3)	(4)	(5)	(6)
Lettable Floor Area *(Log)*	1.1422***	1.2022***	1.2018***	1.1062***	0.9623***	0.9610***
Building Height *(Log)*		0.0851*	0.0080	0.0043	0.0334	0.0362*
Parking Spots *(Log)*		0.0001	0.0003	0.0005***	0.0002***	0.0002**
Energy Label: Below C		-0.2440***	-0.2464***	-0.1847***	-0.0128	-0.0138
Energy Label: C		-0.1754***	-0.1733***	-0.1363**	-0.0552*	-0.0538*
Energy Label: B		-0.0019	-0.0228	0.0181	0.0301	0.0303
Energy Label: Above A		0.4090**	0.2840*	0.3292**	0.2321***	0.2259***
Year Built: before 1906		0.3649**	0.1175	-0.0307	0.3216***	0.3214
Year Built: 1906-1945		-0.0929	-0.3093*	-0.4386***	0.0920	0.0975
Year Built: 1946-1970		-0.5465***	-0.6926***	-0.7317***	-0.1031	-0.0998
Year Built: 1971-1990		-0.7618***	-0.8159***	-0.7229***	-0.1050*	-0.1114*
Year Built: 1991-2000		-0.7752***	-0.7709***	-0.6800***	-0.1185**	-0.1192**
Year Built: 2001-2010		-0.6124***	-0.5487***	-0.5101***	-0.1188**	-0.1231**
Walkscore			0.0093***	0.0067***	0.0033***	0.0034***
Leefbaarometer Score			0.4257***	0.6075***	0.2287***	0.2407***
Train Station Distance *(Log)*			0.0400	0.0207	0.0042	
Highway Distance *(Log)*			0.0364	-0.0095	0.0064	
TRI per sqm. *(Log)*					1.0465***	1.0458***
Vacancy Percentage					-0.3211***	-0.3243***
WALE incl. Vacancy *(Log)*					0.2264***	0.2266***
Rental Difference: Under					-0.1066***	-0.1074***
Rental Difference: Over					-0.1240***	-0.1246***
District Type: Business	-0.2948***			-0.1132	-0.0291	
District Type: Mixed	-0.2740***			-0.1350**	-0.0366	
District Type: Other	0.0258			0.0238	0.0245	
City category: Large	0.5865***			0.5590*	0.2517***	0.2453***
Centrality: Central	0.1513			0.0512	-0.0018	
Transfer Year 2010	0.6198***	0.5022***	0.4143***	0.4622***	0.0526	0.0508
Transfer Year 2011	0.3077***	0.4236***	0.3490***	0.2300**	-0.0479	-0.0450
Transfer Year 2012	0.1033	-0.1548	-0.1320	-0.0437	-0.1150	-0.1149
Transfer Year 2013	-0.1559**	-0.1951**	-0.2282***	-0.2811***	-0.3941***	-0.3903***
Transfer Year 2014	-0.1198	-0.1532*	-0.2117***	-0.2412***	-0.3827***	-0.3794***
Transfer Year 2015	-0.2719***	-0.2389***	-0.2625***	-0.2773***	-0.2447***	-0.2443***
Transfer Year 2016	-0.0623	-0.1124*	-0.0959*	-0.1022**	-0.1546***	-0.1525***
Transfer Year 2018	0.2905**	0.1990	0.2259*	0.2876***	0.1565***	0.1540***
Intercept	5.5071***	5.7967***	5.1845***	5.9739***	1.7816***	1.8161***
R^2	0.80	0.79	0.82	0.85	0.96	0.94
MAPE OLS *(Out-of-Sample)*	52.3%	51.7%	50.0%	43.8%	21.8%	21.9%
MAPE GLS *(Out-of-Sample)*	52.1%	51.5%	49.9%	43.7%	21.7%	21.8%
LOOCV	53.4%	53.2%	51.8%	45.5%	22.8%	22.6%
Simulation 2018	51.2%	50.9%	46.1%	38.1%	20.8%	19.5%

Note: This table provides an overview of the stepwise regression results. The dependent variable is the Log Transaction Price (net) with 1091 observations and *, **, *** denotes significance at the 0.10, 0.05 and 0.01 level, respectively. Model (1) only includes the most significant variable (LFA) and some spatial-temporal control variables. Model (2) adds more building factors. Model (3) captures location through value influencing factors and Model (4) combines these with spatial controls. Model (5) also includes lease factors and Model (6) denotes the result of the (both directional) Stepwise regression. GLS is weighted by 1/SD(resid)^2 with an 80/20 sample split.

Parking Spots

Parking facilities are one of the most important amenities for office properties as their makes it possible for employees and customers to reach the office by car. Hence, we hypothesize the number of parking spots is a function of the price that results in a premium. Although the effect is only limited, the variable is significant through most model specifications. It is however important to note that the inclusion of this variable as is remains controversial. Not only is the variable highly correlated (0.7) with the LFA, but *Appendix A* also shows an unusual amount of zero parking spots which was likely filled in when the number of spots were unknown. Furthermore, when no or limited parking facilities are present on-site these may be arranged somewhere off-site. Further research is encouraged.

Energy Label
The awareness and recognition of environmentally friendly offices have increased in recent years. Many literature has also proven that green buildings can add value (Fuerst & McAllister, 2011). For investors there may be higher net operating income due to increased demand from occupiers, lower void rates, lower costs of ownership and an element of protection from future regulatory changes. In turn, this is likely to be reflected in the value of the property. We assume that this 'greenness' can be captured through the use of energy labels as a proxy within our model.

Model (2-6) do not find consistent significance results through the model specifications for the various categories. The buildings with an energy label-above-A get a high premium compared to label-A which is as expected for the above described reasons. Strangely enough, properties with label-B have a positive price effect compared to label-A in *Model (4-6)* but are also insignificant. As the remaining categories have the expected signs and the prediction accuracy increases we leave the variable as is in the model. Important to note is that per 1 January 2023 all office buildings must have at least label-C. This is expected to decrease the demand and thus the price in our Hedonic price model for below C properties in the near future and deserves further research.

Year Built or Last Renovated
In Hedonic price studies aimed at the Residential sector, the quality and maintenance of both the interior and exterior of the building provide significant explanatory power to the model. Unfortunately, for the Commercial Real Estate market in the Netherlands this information is not available. Many studies therefore advocate the building age as a proxy for quality. The intuitive reasoning is that office buildings become physically, economically and functionally obsolete (Bokhari & Geltner, 2016). In other words, the building age proxies the depreciation rate of the building. In this thesis we find that dummy variables for building periods provide superior explanatory power over the use of effective age as a linear or spline function.

Most building period categories are significant in *Model (2-4)* which is a result of the preliminary feature engineering process. With the inclusion of lease factors in *Model (5-6)* this significance however partly disappears signing potential multicollinearity with these variables. In addition, the years of renovation did not add value to the model which may be caused by the large amount of missing data and large deviations in quality can still exist within a category. A rating system for the office building properties like CoStar's Building Rating System (2018) deserves some thought. For now, as the coefficients are as expected, the variable is included as is in the Baseline model.

7.1.2 Location Factors

Walkscore
Walkscore is a tool that provides a score of a location based on the amount and rating of amenities within walking distance (400 meters). Each type of destination is given equal weight and the points for each category are summed and normalized to produce a score from 0 to 100[21]. Although not (yet) validated within the Netherlands, various studies make use of this measure and find significance (e.g. Kok & Jennen, 2012; Pivo & Fisher, 2011). Clustering of services in certain areas is often found to increase efficiency of labour, all other things being equal, and increases in value can be afforded to pay. *Model (3-6)* show highly significant estimates with the expected sign across all specifications and is thus included in the Baseline model as a log-linear function. Important to note is that the scores are observed only in the present hence the score of past transaction could have changed over time.

[21] See https://www.walkscore.com/professional/research.php

Leefbaarometer
Leefbaarometer is used as a proxy for the livability of the location based on the five quality subscores (Housing, Residents, Facilities, Safety and Physical environment) and is included as the sum of the means from the national average. Important to note is that potential multicollinearity can occur in combination with the Walkscore as these also capture facilities. Nevertheless, as both variables remain highly significant across all model specifications and have the expected sign and decrease the overall MAPE of the model, we assume it is beneficial to include both measures within the Baseline model.

Accessibility
The accessibility of the location is measured through the minutes driving (without traffic) to the nearest highway access and the minutes walking towards the nearest train station. Whether this train station provide intercity trains is already included in the centrality control measure. Literature provides contradictory results with some finding statistical significance (e.g. Debrezion et al., 2006) while others do not. As the effect of these value influencing factors decrease more with distance the relationship is assumed to be a negative logarithmic function of price.

Model (3-5) find that both the duration to station and to the nearest highway access are not significant and do not have the expected signs over all model specifications. These variables are thus dropped from the final Baseline model specification. Furthermore, the effect of other public transportation modes is excluded from the analysis as it is assumed that every office property is accessible by some kind of transportation mode. The less accessible properties would likely also be further from the station and thus are partly included in the duration to the nearest station.

Control: City Category
The first control variable of location is the city category. We make a distinction between Large and Small cities where the large city category includes Amsterdam, Schiphol, Rotterdam, Den Haag, Utrecht and Eindhoven while the small include the remaining cities and towns. This control is significant through all model specifications.

Control: Centrality
The second control variable of location is the centrality within the city. A property assumed to be located central when it is in walking distance from an intercity train station. In line with the Walkscore proxy this distance is set to 400 meters. The control is not significant in all model specifications with some unexpected signs and is therefore excluded from the final Baseline model.

Control: District Type
The last control variable of location is the type of district. A distinction has been made between Office, Business, Mixed and Other district type. Office and the miscellaneous district types seem to have a premium over Mixed and Business districts. However, as these are not significant they are dropped in the final Baseline model.

7.1.3 Lease Factors

In the Residential Sector, building and location are often seen as the two dimensions that determine the value of a property. Commercial Real Estate theory however suggests that the underlying future cashflows that a property generates have a significant impact on the value of the property and thus introduce a third dimension. In line with the second hypothesis of this thesis, we investigate the influence of lease related factors on the ability of the model to predict transaction. We should keep in mind that the inclusion of lease factors potentially poses multicollinearity and endogeneity problems. Results should therefore be interpreted with caution.

Theoretical Rental Income
We hypothesize that the higher the rental income per square meter, the higher the incoming cashflows and thus improve the value of a property. We use Theoretical Rental Income (TRI) as this measure includes rent that is generated would all space have been let and thus is independent of vacancy. Together with the Lettable Floor Area, the TRI provides most explanatory power and remains highly positive significant over all model specifications. This variable likely also captures some missing information about building and location factors as these are often priced in the rents.

Vacancy Percentage
If there is vacancy at the time of sale there is the risk that less income will be generated with the property and can affect the value negatively. This however does not hold for investors who buy the property for own use; the so-called owner-users. As this information is available, we assume no vacancy when this is the case with a lease term of five years. The results show that the remaining vacancy is highly significant with a negative linear relationship with price.

Remaining Lease Term
Although few Hedonic studies are found that include lease terms, in practice it is well established that the remaining term of contracts is highly correlated with the value of the property. We observe that in general lower rent is offered on longer leases as the risk of vacancy after a tenant leaves are lower and consequently results in lower transaction costs such as agent fees. A longer (weighted) average lease term thus provides less risk and thus investors are prepared to pay a higher price. On the other hand, if the property is underrented, one could receive higher rents on the open market and might decrease the value. The price effect is thus dependent on local market conditions.

Our model shows highly significant positive coefficients which are as expected as we controlled for under or over rented properties. Note that we use the WALE including vacancy opposed to excluding vacancy as the former provides significantly higher explanatory power. This is as expected as WALE excluding vacancy does not give a fair representation of the risk when large parts of the building are vacant. We find new evidence of the significant importance of the remaining lease term on the value of an office property in the Netherlands.

Rental Difference
If the rent is below the market, the property could receive higher value on the open market. Not all observations had information about the Market Rent per square meter, so we applied median substituted based on the five nearest neighbors. Results of this variable are highly significant but do not have the expected sign. Reasons could be the missing data, input errors or amount of difference between the Rental Income and Market Rent. As the variable is however highly significant and increase prediction accuracy we keep the it in the final model specification.

7.1.4 Model Diagnostics

We run several model diagnostics to check the validity of the final Baseline *Model (6)*, denote issues remaining and to test where and how the model could potentially be improved.

Residual Analysis
First and foremost, we check the residuals of the model to see if there are no obvious patterns remaining. If for example an underlying trend is visible the model failed to capture an important piece of information. Both the residuals against the fitted values and against individual explanatory variables are at random hence are assumed to be valid. *Appendix C* shows the residual plots of the model.

Cook's Distance
Second, we check the residuals for outliers based on cook's distance. Influential points that are more than four times the mean are labelled as these observations might distort the outcome and accuracy of the regression. These influential data points are worth checking for validity but in this thesis are all dropped from the analysis (see *Appendix C*).

Multicollinearity
Third, we test for multicollinearity through the use of the variance inflation factor (VIF). Variables that exceed a threshold of 10 are assumed to be multicollinear. As all pairs are below the threshold we assume that no multicollinearity is present within the final model specification.

Heteroskedasticity
Fourth, we test for heteroskedasticity. Real Estate prices are known for its heteroskedasticity hence we apply two Breusch-Pagan tests; one studentized and one non-studentized. Both tests show evidence of heteroskedasticity within the model. We thus apply White's correction to the standard errors to see if any of the previously selected variable show significant change in both significance and coefficients. This does not seem to be the case hence we are a little more convinced about the validity of our model[22].

Spatial Autocorrelation
Last, we test the model for spatial autocorrelation. A global Moran's I test finds significant spatial autocorrelation within the model. We use Lagrange Multiplier to determine whether the spatial dependence is in the dependent variable (spatial lag – SAR) or in the model errors (spatial error – SER). Both seem to be the case. We can thus conclude that significant spatial autocorrelation remains within the model despite the various control variables introduced. This is as expected since we did not control for the correlations within each control as this would leave to few observations per category (e.g. office district dummies). A model that can include the spatial and temporal dependencies as a continuous function might therefore be preferable as AVM model for individual property valuations.

7.2 Comparable Weighted Regression

We found four issues with the Baseline Hedonic regression model that pose a threat to the predictive capabilities of the AVM model. First, OLS regression estimates do not provide reliable standard errors due to the presence of heteroskedasticity. Second, the model contains significant levels of spatial dependencies both in the dependent variable and residuals. Third, the time dummy variables provide a very broad approximation with its yearly intervals on a national level. It is however common knowledge that differences are present at smaller intervals and can vary over space (or even per variable). Last, the stringent parametric assumptions do not provide optimal fits to the data.

We therefore propose a newly developed Comparable Weighted Regression (CWR) model that tackles the first three issues. Similar to the output of an GWR model, the CWR provides 'local' coefficients that can vary instead of being static as is assumed in a global (OLS) model. Note however that in contrast to the GWR it is difficult to plot the 'local' dynamic coefficients on a map as we are dealing with more dimensions than only space. A combination with non-parametric Machine Learning techniques could also tackle the fourth issue but goes beyond the scope of this thesis. *Table 7-2* shows the results of the Comparable Weighted Regression.

[22] See *https://github.com/BasHilgers/Thesis_TUe_AVM* for the code and outcomes of the diagnostic tests.

Table 7-2 ■ Comparable Weighted Regression Results

Variable	Baseline (6)	Min	Mean	Median	Max	SD
Lettable Floor Area *(Log)*	0.9610	0.9527	0.9676	0.9677	0.9833	0.0051
Building Height *(Log)*	0.0362	0.0158	0.0378	0.0374	0.0607	0.0104
Parking Spots *(Log)*	0.0002	0.0001	0.0002	0.0002	0.0004	0.0001
Energy Label: Below C	-0.0138	-0.0358	-0.0094	-0.0098	0.0148	0.0111
Energy Label: C	-0.0538	-0.0664	-0.0505	-0.0506	-0.0346	0.0043
Energy Label: B	0.0303	0.0133	0.0338	0.0333	0.0520	0.0074
Energy Label: Above A	0.2259	0.2063	0.2578	0.2601	0.2848	0.0156
Year Built: before 1906	0.3214	0.2698	0.3428	0.3397	0.4157	0.0275
Year Built: 1906-1945	0.0975	0.0411	0.0962	0.0932	0.1649	0.0211
Year Built: 1946-1970	-0.0998	-0.1328	-0.0965	-0.0967	-0.0598	0.0122
Year Built: 1971-1990	-0.1114	-0.1578	-0.1176	-0.1188	-0.0865	0.0127
Year Built: 1991-2000	-0.1192	-0.1599	-0.1237	-0.1251	-0.0925	0.0113
Year Built: 2001-2010	-0.1231	-0.1604	-0.1289	-0.1294	-0.0928	0.0118
Walkscore	0.0034	0.0027	0.0033	0.0033	0.0038	0.0002
Leefbaarometer Score	0.2407	0.1740	0.2388	0.2426	0.2930	0.0284
TRI per sqm. *(Log)*	1.0458	0.9775	1.0369	1.0377	1.1005	0.0277
Vacancy Percentage	-0.3243	-0.3701	-0.3244	-0.3247	-0.2768	0.0193
WALE incl. Vacancy *(Log)*	0.2266	0.2024	0.2284	0.2288	0.248	0.0098
Rental Difference: Under	-0.1074	-0.1550	-0.1277	-0.1279	-0.1041	0.0128
Rental Difference: Over	-0.1246	-0.1207	-0.1016	-0.1017	-0.0768	0.0080
City category: Large	0.2453	0.2198	0.2494	0.2495	0.2751	0.0097
Transfer Year 2010	0.0508	0.0289	0.0610	0.0569	0.1689	0.0344
Transfer Year 2011	-0.0450	-0.0748	-0.0424	-0.0426	-0.0005	0.0138
Transfer Year 2012	-0.1149	-0.1674	-0.1231	-0.1220	-0.0763	0.0129
Transfer Year 2013	-0.3903	-0.4294	-0.3888	-0.3908	-0.3360	0.0180
Transfer Year 2014	-0.3794	-0.4031	-0.3819	-0.3824	-0.3550	0.0082
Transfer Year 2015	-0.2443	-0.2652	-0.2441	-0.2442	-0.2135	0.0071
Transfer Year 2016	-0.1525	-0.1661	-0.1510	-0.1512	-0.1357	0.0056
Transfer Year 2018	0.1540	0.1083	0.1524	0.1521	0.2053	0.0143
Intercept	1.8161	1.4418	1.8012	1.7941	2.1407	0.1670
R^2	0.94	0.94	0.95	0.95	0.96	0.01
LOOCV	22.6%	-	21.9%	-	-	-
Simulation 2018	19.5%	-	19.3%	-	-	-

Note: This table shows ranges of the (local) coefficients based on clusters of comparable buildings that are used as weights. The Baseline is the model where the weight are the same for every observations and is thus identical to Model (6) specification of Table 7-1. LOOCV and Simulation 2018 are again represented as MAPEs.

We derive the following conclusions from the results. The LFA remains relatively stable over clusters of similar properties which might come as a surprise as theory suggest that certain type of buildings have a higher price per square meter. Most likely, these effects are captured in the rent or other variables included in the model. Building height on the other hand show a broader range with micro analysis showing that the larger coefficients in general come from large city, high-rise buildings. The global function thus seems to underestimate the price effect of building height for these types of buildings. Parking spots remain relatively stable which is also unexpected for similar reasons as the LFA. Energy Labels on the other hand are not stable with the coefficient of Energy label C even switching signs. Furthermore, Energy label-B seem to consistently have a premium over Energy label-A. It can thus be questioned whether this variable should be included in the final model specification. The Walkscore shows least variations in the coefficients, which might be a result of the inclusion of the Leefbaarometer that correlates with the Walkscore. Lastly, the Lease factors show no surprises.

Overall, the CWR method, although slow in terms of computation time, shows positive signs of improvement over the traditional Hedonic regression methodology. It thus indeed seems to tackle some of the issues present in standard OLS estimates. In addition, relative weights of variables are yet to be optimized and there is thus still room for improvement. The better comparables are found, the more accurate the results.

7.3 Machine Learning Methods

In the third step of the price modelling phase we compare the results of the traditional Hedonic Baseline regression and Comparable Weighted Regression with two Tree-based Machine Learning algorithms. The application of such algorithms are increasingly popular in both literature and practice, especially in the Residential sector, with many to find superior performance accuracy over traditional regression (e.g. Antipov & Pokryshevskaya, 2012; Zurada et al., 2011). In line with the third and last hypothesis of this thesis, we investigate whether these findings also hold for the Commercial Real Estate sector where transactions are much rarer and data is more heterogeneous.

In order to guide our findings, we first apply the Random Forest and Extreme Gradient Boosting (XGBoost) Algorithms with standard parameter settings. Next, we tune the hyperparameters of these models through repeated *k*-fold cross validation based on the RMSE to adjust the model to the traits of the data[23]. The tuned models naturally improve the standard model but including both in our results lets us analyse the extent. Similarly, we include both In-sample and Out-of-Sample accuracy results to indicate signs of overfitting. We finish the section with an analysis of the variable importance.

Table 7-3 ■ Prediction Accuracy AVM Models

Model	Out-of-Sample	In-Sample	LOOCV	Sim 2018	COD
Baseline Regression	21.8 %	21.9 %	22.8 %	20.8 %	-18.1 %
Stepwise Regression	**21.7 %**	21.9 %	22.6 %	19.5 %	-16.6 %
Baseline Comparable Weighted	-	-	22.4 %	21.1 %	-25.8 %
Stepwise Comparable Weighted	-	-	**22.2 %**	**19.3 %**	-22.1 %
Random Forest	27.0 %	13.7 %	29.6 %	20.9 %	-95.0 %
Tuned Random Forest	27.7 %	11.4 %	28.4 %	23.3 %	-49.0 %
Gradient Boosting	27.9 %	**0.2 %**	27.3 %	21.3 %	-119.3 %
Tuned Gradient Boosting	27.7 %	21.8 %	26.5 %	21.6 %	**-15.3 %**

Note: The Baseline model equals model (5) and the Stepwise model (6) in Table 7-1. The Comparable Weighted Regression methods make use of the same model specifications. The Machine Learning models are run with both standard and tuned hyperparameter settings. The sample is split into 80% In-sample and 20% Out-of-Sample. COD defines the Coefficient of Dispersion parameter, LOOCV the Leave-One Out Cross Validation and Sim 2018 a real-world simulation with arrival of new recent data only. Bold represents the best result among models. Except for the COD, all percentages denote MAPEs.

7.3.1 Model Comparison

We've already derived in the previous section that the Comparable Weighed Regression outperforms traditional Hedonic Regression in terms of prediction accuracy. *Table 7-3* confirm these results for different cross validations that are applied to obtain true, unbiased and reliable error measures of the models. The Coefficient of Dispersion (COD) is however higher which indicate that the errors are more dispersed around the median. This is generally an unwanted feature and should be weighed against the increase in prediction accuracy when choosing an optimal model. *Section 7.4* analysis this further. Note also that the simulation of 2018 indicates improvement of the model over time.

Surprisingly, the tuned Random Forest Model actually seems to perform worse than the standard hyperparameter settings in terms of MAPE. When we however look at the out-of-sample RMSE in our code we see that that this error measure has indeed decreased which is apparently not seen in the MAPE. As the RMSE penalizes outliers more it is likely that these improved while the mean error remained similar. This is also reflected in the COD that decreased.

[23] See code at *https://github.com/BasHilgers/Thesis_TUe_AVM* for the values of the hyperparameter.

The XGBoost algorithm improves noticeable over the standard Gradient Boosting in terms of prediction speed. This comes in handy with large datasets and cross validations with many folds. Within this time the method can fit the data nearly perfect when no cost is put on complexity as we can observe from the in-sample error. Out-of-sample the model however has large errors with a high COD as the function does not generalize well. This depicts the strong tendency of Machine Learning to overfit the data. When the many hyperparameters of this algorithm are optimally tuned with repeated *k*-fold cross validation, prediction accuracy improves, and the COD reduces to lowest among models. The results are however still in favour of traditional Hedonic regression.

All-in-all we can conclude that Machine Learning Methods do not provide better prediction accuracy over traditional regression methods for the Commercial Real Estate data at hand. Potentially the moderate sample size together with a limited number of covariates make it difficult for the algorithms to find the underlying patterns in the data. The Random Forest performs worst of all models and is in line with previous literature that notes that this algorithm performs better at classification tasks. The Gradient Boosted model shows better speed, flexibility, and the lowest COD, but still does not manage to outperform the Comparable Weighed Regression in terms of prediction accuracy.

7.3.2 Variable Importance

A nice feature of Tree-based Machine Learning algorithms is that it comes with relative variable importance and thus brings back some interpretability to this black-box method (see *Figure 7-1*). With it we can interpret the results to check for no surprises in the important features, allow for additional feature selection and guide the direction of feature engineering.

As Hastie et al. (2008, p. 593) explain: "At each split in each Tree, the improvement in the split-criterion is the importance measure attributed to the splitting variable, and is accumulated over all the Trees in the forest separately for each variable". So in other words, as we train each Tree we do a permutation experiment were we scramble the values of a variable and check whether the accuracy of the Tree changes. If it does by a lot, the variable is very important and vice versa. Note however that in this thesis the relative importance is not captured by the mean decrease in accuracy but rather the decrease in node impurity (through the Gini coefficient). This method works in a similar way for both the Random Forest and the Gradient-Boosted Algorithm.

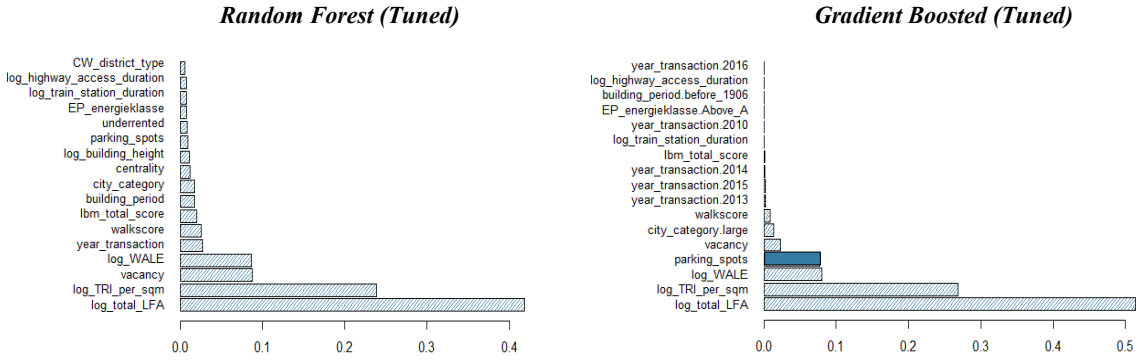


Figure 7-1 ■ Variable Importance

Note: The left plot represents the Random Forest with tuned hyperparamaeters and the left the Gradient Boosted model. Both derive it values from the Gini splitting index. The Gradient Boosted algorithm is estimated with a sparse matrix of features and the figure is thus truncated to show only the most important categories.

As expected, the Lettable Floor Area (LFA) is the most important variable in both algorithms. We also find additional evidence for the importance of Lease related factors in the valuation of Commercial Real Estate. The Theoretical Rental Income (TRI) per square meter is the second most important variable and the remaining Lease Term (WALE) and Vacancy percentage are also high on the list. The Rental Difference is the only Lease factor of which the importance is limited. Again, this could be due to the incomplete and subjective information in this variable and interpretation of this result should be done with caution.

Surprisingly, the number of Parking Spots is found to be an important predictor in the Gradient Boosted model whereas the Random Forest and traditional regression do not find this result. This seems to be an example of the case where a Machine Learning algorithm derive its results from correlations rather than causation. Since the number of Parking Spots is highly correlated with the LFA (>0.7) these might be used as substitutes for each other in the splitting criteria of the Trees. This does however not mean in reality that adding parking spots cause high increase in price. The remainder of the variables are of relative little importance, but when combined determine a significant part of the estimate. With practically no limit to the number of features that can be included, this characteristic where no manual variable selection has to take place represents one of the strengths of these algorithms. Further research in this area could benefit the AVM model. No additional feature engineering steps were taken after the additional information from this section.

7.4 Model Evaluation

In conclusion, the Comparable Weighted Regression provides superior performance over traditional Hedonic Regression and the (tuned) Machine Learning models: Random Forest and (Extreme) Gradient Boosting. That is, the CWR has a MAPE obtained through LOOCV of 22.2 percent while the best Hedonic and the Machine Learning methods have a MAPE of 22.6 and 26.5 percent, respectively (see *Table 7-3*). Compared to manual appraisal which has a mean percentage error of approximately 10 percent, the model and data still have long way to go before practical application.

In addition to an optimal (mean) performance accuracy, we aim to minimize the amount of large errors as these are especially unwanted. *Table 7-4* provides an overview of the percent of instances that are measured within a certain error percentage. Again, the CWR find the better performance in all cases. The results show that approximately 30 percent of the instances can be estimated within a 10 percent error. The larger errors should be checked manually for false information in further research as these may skew the results and thus can still increase the performance of all models.

Table 7-4 ■ Error Distribution per AVM model – LOOCV

Percentage Difference	Baseline	Stepwise	CWR	RF	GBM
< 1%	3%	3%	**4%**	3%	2%
< 2%	5%	6%	**8%**	6%	5%
< 3%	9%	9%	**10%**	8%	8%
< 5%	15%	15%	**16%**	14%	14%
< 10%	28%	29%	**30%**	25%	26%
< 25%	61%	62%	**63%**	54%	58%
< 50%	90%	90%	**90%**	85%	86%
< 75%	**100%**	**100%**	**100%**	95%	96%
< 100%	**100%**	**100%**	**100%**	97%	99%

Note: Percentage Difference measured as the difference between the predicted value of a property and the actual transfer price (net). Percentages per model denote cumulative numbers. Sample size of 979 estimated with LOOCV. Bold represents optimal results among the five models.

8. Conclusion and Discussion

The primary motive for undertaking this thesis was to study the potential of Automated Valuation Model for the valuation of Commercial Real Estate properties in the Netherlands. We therefore investigated different methods and discussed various considerations related to this task. In this chapter we summarize our results and provide a discussion about the implications of our findings. We finish this thesis by highlighting some of the limitations remaining in the model and provide suggestions for additional research.

- 7.1 Conclusion
- 7.2 Discussion
- 7.3 Further Research

8.1 Conclusion

The aim of this thesis was to investigate the potential of Automated Valuation Models (AVMs) for the estimation of the Market Value of individual Commercial Real Estate properties in the Netherlands. With a final cleaned dataset of 979 office property transactions obtained from Cushman & Wakefield over the period 2010 through 2018, we studied both well-established traditional Hedonic regression methods and newer Machine Learning algorithms. Furthermore, we proposed a new original method, named Comparable Weighted Regression, that extents the traditional Hedonic regression in a way that gives higher weights to observations that are more comparable to the subject property. Similar to the Geographically Weighted Regression, this method counters some of the rigid assumptions of traditional regression. But instead of deriving weights from distance only, it looks at higher dimensions to allow for discontinuities over space which is very similar to the process that a valuer adopts.

This thesis is one of the first studies to investigate the potential of AVM applied to the Commercial Real Estate sector and the first one to compare both more traditional regression and newer Machine Learning methods. With the goal of developing a predictive model that generates optimal (out-of-sample) predictions for as many observations as possible while minimizing the amount of large errors, we provide insight whether these models have the potential to outperform manual appraisals in this sector. Based on an extensive literature review of potential methods that can be used for this task, we selected four methodologies that offer the most potential for practical application. These are traditional Hedonic regression, both with and without a spatial-temporal model extension, and the tree-based machine learning algorithms Random Forest and (Extreme) Gradient Boosting. In order to provide a fair comparison of these methods, we use the same variables in each model. These are categorized by the following categories[24]:

Building
- Lettable Floor Area
- Building height
- Parking spots
- Energy label
- Year built (or last renovated)

Lease
- (Theoretical) Rental Income per square meter
- Vacancy percentage
- Weighted Average Lease Expiry (including vacancy)
- Rental difference: Over- or under-rented

Location
- Leefbaarometer score (overall quality of the area)
- Walkscore (amenities within walking distance)
- *Distance to nearest highway access (accessibility by car)
- *Distance to nearest train station (accessibility by public transport)

Market
- City category: Large versus small cities
- *Centrality: Central versus decentral position in the City
- *District type: Office, business, mixed or other
- Year of transaction

[24] The * indicates whether a variable was excluded from the final Baseline Regression specification through both way stepwise regression.

The first hypothesis addressed in this thesis investigates the importance of lease related factors in the prediction of the Market Value for Commercial Real Estate properties. It is a well-known fact in the valuation practice that the value of Commercial Real Estate strongly depends on the income that is generated through a property. Surprisingly few academic studies and models applied in practice however include such variables as this information is often confidential or incomplete. With the unique dataset obtained from Cushman & Wakefield this thesis finds that these features can greatly improve the accuracy of the model. All lease related factors remain highly significant through all model specifications and improve the Mean Absolute Percentage Error (MAPE) obtained with Leave-One Out Cross Validation (LOOCV) of the Baseline Hedonic model from 45.8 to 22.8 percent. Note however that including these factors potentially introduce endogeneity to the model and a large part of the improvement is likely due to omitted variable bias. These results should thus be interpreted with caution. Nonetheless, this thesis provides strong evidence that AVMs applied to the Commercial Real Estate sector could benefit from including lease related factors into their model specification.

The second hypothesis in this thesis investigated whether the traditional Hedonic framework can be improved by allowing for effects of spatial- and temporal dependencies. We first reviewed the most popular models that allow for these effects and concluded that the Geographically Weighted Regression was often found to provide superior performance in studies applied to the Residential Real Estate sector. Issues that however remained were that this method does not control for temporal dependencies and that comparable properties that are distant are not considered by the model. We therefore introduced a new Comparable Weighted Regression model that solved these issues in a rather straightforward way. Weights are based on a comparability score that includes both spatial and temporal similarities, among other factors. This thesis provides first evidence that such weighted regression can improve the prediction accuracy with an MAPE that decreased to 19.3 percent in the best model specification and the number of large errors that was lowest among the four methodologies. Note that the current results still offer a lot of room for improvement as the better comparables are found, the higher the prediction accuracy will be.

The third and last hypothesis addressed in this thesis studied whether a well-defined Hedonic price model can outperform the newer Machine Learning algorithms that have increased in popularity in recent years in both academia and practice. The Random Forest and (Extreme) Gradient Boosting algorithms were found to provide most potential among algorithms as these methodologies are in line with the process that a valuer adapts, are known to provide excellent performance in both speed and prediction accuracy, and the interpretability of the trees' results is highest among 'black-box' techniques. Our results indicate that with the data at hand, Machine Learning algorithms have difficulties finding the true underlying patterns. The tuned Gradient Boosting outperforms the Random Forest algorithm with a MAPE of 21.6 percent, but which is still worse than the prediction accuracy of both the traditional Hedonic and Comparable Weighted Regression models. With more observations and the benefit of such algorithms to deal with high-dimensional (fat) data that include missing variables, this accuracy can probably still improve significantly. Nevertheless, as the number of transactions in the Commercial Real Estate sector remains relatively low, such data might be difficult to obtain.

8.2 Discussion

The valuation profession is likely to face a period of significant change in the upcoming years. These changes mainly relate to different client expectations and technological developments in 'Big' Data, Blockchain, Artificial Intelligence and probably above all the widespread application of Automated Valuation Models (AVMs). Whether these changes are a threat or an opportunity for valuers is still unclear. This section briefly discusses what changes AVMs will bring, what form it can take and how valuers can position themselves in this whole.

8.2.1 The Future of the Valuation Profession

Valuations have a rich history and play a crucial role in many real-estate related decisions. With the paradigm shift initiated by digital transformation, the sector however seems to be at a crossroad. One way is the road where the digitization will disrupt the current practice (Schumpeter's theorem), whereas the other leads to the evolution of the sector. Geophy (2018) clearly visions the former to be correct and boldly states that within five years the market for valuations is almost entirely automated. What however is left out of this statement is that the traditional valuation practice is also moving forward. Through the years the sector has become increasingly more sophisticated with noticeable improvements in the use of data and consistency and transparency of the results (RICS, 2017). One thing that will however not change is the goal to provide clients with an unbiased estimation of the (market) value. The main question is thus whether these developments could help the valuer deliver more accurate and efficient valuations, or alternatively completely or partially replace the role of the valuer. The added value for the client will ultimately be the key.

8.2.2 The Role of Artificial Intelligence

Artificial Intelligence applied to the Automated Valuation of real estate has the potential to provide excellent predictions of the (market) value of a property that with the right amount of relevant data can be more accurate than manual appraisals. In addition, parametric approaches can include confidence intervals that depict the certainty of a value estimate falling within a certain range which can be helpful as a risk measure. It is thus clear that AVMs are here to stay and as models become more sophisticated, their accuracy and usability will only increase. But whether such models are applied with Machine Learning or rather with more traditional regression techniques is still under discussion.

We already highlighted that Machine Learning algorithms have the advantage that they can find patterns in high dimensional data which leads to prediction results that are difficult to beat by traditional regression approaches. Currently, we then also see that AVMs that are applied in practice incorporate some form of Machine Learning in their estimation process. But with the models almost exclusively applied to the Residential real estate sector where data is more homogeneous and abundant this is not a surprise as this is where such models perform best. Accurate estimates can then be generated in only a fraction of the time and costs with little value that humans can add. However, if such models fail to find appropriate data these can become dangerously unreliable. In general, it holds that the more complex the algorithm, the better the estimation results, but also the more difficult it is to unravel where this 'black-box' derived the value from. This makes it nearly impossible to defend individual estimates should there be a legal requirement to do so. In such cases, the interpretability of traditional regression models that allow for statistical testing might be preferred over the increase in prediction accuracy of Machine Learning. As Commercial Real Estate falls within the category where quality data is more limited, interpretability of the results area high on the list. It might even be the task of the valuer to choose between algorithms on a case to case basis. It is no doubt that AVMs will increasingly become an indispensable tool for valuers. But before that, more research is needed.

8.3 Further Research

The research field of Automated Valuation of real estate is broad. In this thesis we only scratched the surface of possible techniques and data that can be exploited to generate an accurate prediction of the (market) value of Commercial Real Estate. With the expected increase in (publicly) available data related to these kind of properties from companies that specialize in this type work such as Real Capital Analytics and Vastgoeddata, in combination with initiatives from large market players as Cushman & Wakefield to structure their data assets, this work is likely to boom in upcoming years. So to conclude this thesis, we have some recommendations for further research based on our experience.

- A natural next step is to extent the current study that focused on the Office sector to other Commercial Real Estate sectors such as Retail, Logistics, Commercial Residential or even more exclusive sectors such Hotels, Healthcare or Datacenters. These sectors do have some overlap with the office market sector so results and methods of this thesis might be still useful. However, as each sector have their own distinct variables that influence the value, additional research is encouraged.
- We have shown that the original Comparable Weighted Regression framework of this thesis offers the potential to outperform traditional Hedonic regression and Machine Learning algorithms in the valuation of Commercial Real Estate. Further research is needed to test whether these results still hold with an increase in data that are less heterogeneous and more local, for example in the residential real estate sector. In line with this, additional academic research may benefit from investigating how an algorithm can best find the closest comparable in an automated way based on its characteristics.
- The data could be enriched with a wide set of variables. Commercial Real Estate data for AVMs is a typical example of 'fat' data, which opposed to 'tall' means it has a lot of predictors relative to the number of observations (transactions). Various studies find that 'hyperlocal' metrics such as proximity to music events, green space and local crime can affect the value of real estate. Even unconventional 'new' data which was not useful for research in the past such as image and language information might help to create better value predictions. Further research might concern itself with how the high-order interactions of this 'Big' data can best be incorporated into a prediction model for Commercial Real Estate. Hybrid models that combine the strengths of Machine Learning with the interpretability and statistical validation of traditional regression might be the key. A good start would be to apply the LASSO technique.
- It can be questioned whether the actual transacted prices of Commercial Real Estate are representative of the market value of that property. Often the conditions of a transaction are not in line with the definition of market value and including such transactions in the estimation model may skew the results. Examples are forced sales, trophy buildings or even a McDonalds that might have value for this company with its unique design but are much less valuable for others. In other words, the price for which a property transacts is often the highest bid that can differ significantly from what the rest of the market is willing to pay for it. Additional research is needed to investigate these effects on the accuracy of the valuation model.
- This thesis excludes periods of economic distress. Future research studies should investigate the performance of AVMs applied to the Commercial Real Estate sector over a full market cycle and study how a model can adjust to values that divert from the fundamentals when the market is in distress. It might be interesting to see how risk related factors such as the risk of a tenant leaving might affect the value in such times.

Bibliography

Abidoye, R. B., & Chan, A. P. C. (2017). Artificial neural network in property valuation: application framework and research trend. *Property Management, 35*(5), 554-571.

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*: Springer.

Anselin, L. (2017). Spatial Data Science. Retrieved 1 March, 2018, from https://www.youtube.com/watch?v=lawWM6jQYEE&feature=youtu.be

Antipov, E. A., & Pokryshevskaya, E. B. (2012). Mass appraisal of residential apartments: An application of Random forest for valuation and a CART-based approach for model diagnostics. *Expert Systems with Applications, 39*(2), 1772-1778.

Archer, W. R., & Smith, M. T. (2003). Explaining Location Patterns of Suburban Offices. *Real Estate Economics, 31*(2), 139-164.

Bitter, C., Mulligan, G. F., & Dall'erba, S. (2007). Incorporating spatial variation in housing attribute prices: a comparison of geographically weighted regression and the spatial expansion method. *Journal of Geographical Systems, 9*(1), 7-27.

Bokhari, S., & Geltner, D. (2016). Characteristics of Depreciation in Commercial and Multifamily Property: An Investment Perspective. *Real Estate Economics, 0*(0).

Bollinger, C. R., Ihlanfeldt, K. R., & Bowes, D. R. (1998). Spatial Variation in Office Rents within the Atlanta Region. *Urban Studies, 35*(7), 1097-1118.

Borst, R. (2007). *Discovering and Applying Location Influence Patterns in the Mass Valuation of Domestic Real Property.* (Doctor of Technology), University of Ulster.

Borst, R. (2015). *Improving Mass Appraisal Valuation Models Using Spatio-Temporal Methods*. Toronto: International Property Tax Institute.

Brachinger, H. W. (2003). Statistical Theory of Hedonic Price Indices: Department of Quantitative Economics, University of Freiburg/Fribourg Switzerland.

Breiman, L. (2001). Random Forests. *Machine Learning, 45*(1), 5-32.

Brinkman, J. A. A. (2014). *Het modelmatig waarderen van kantoorobjecten : een onderzoek naar de mogelijkheden om de waarde van kantoorobjecten modelmatig vast te stellen.* (MSRE), Amsterdam School of Real Estate, Amsterdam.

Castle, J., Qin, X., & Reed, W. (2009). *How To Pick The Best Regression Equation: A Review And Comparison Of Model Selection Algorithms*.

Chegut, A., Eichholtz, P., & Kok, N. (2014). Supply, Demand and the Value of Green Buildings. *Urban Studies, 51*(1), 22-43.

Chegut, A. M., Eichholtz, P. M. A., & Rodrigues, P. J. M. (2015). Spatial Dependence in International Office Markets. *The Journal of Real Estate Finance and Economics, 51*(2), 317-350.

Chen, T., & Guestrin, C. (2016). *XGBoost: A Scalable Tree Boosting System*. Paper presented at the the 22nd ACM SIGKDD International Conference, San Francisco.

Colwell, P. F., & Munneke, H. J. (2006). Bargaining Strength and Property Class in Office Markets. *The Journal of Real Estate Finance and Economics, 33*(3), 197-213.

Colwell, P. F., Munneke, H. J., & Trefzger, J. W. (1998). Chicago's Office Market: Price Indices, Location and Time. *Real Estate Economics, 26*(1), 83-106.

CoStar. (2018). CoStar Building Rating System: COMPS [Press release]

Cushman & Wakefield. (2018). Nederland compleet (medio 2018). Amsterdam.

D'Amato, M., & Kauko, T. (2017). *Advances in Automated Valuation Modeling: AVM After the Non-Agency Mortgage Crisis* (Vol. 86).

David Geltner, & Neufville, R. d. (2018). *Flexibility and Real Estate Valuation under Uncertainty: A Practical Guide for Developers*. Oxford: Wiley-Blackwel.

Debrezion, G., Pels, E., & Rietveld, P. (2006). The Impact of Rail Transport on Real Estate Prices: An Empirical Analysis of the Dutch Housing Market. *Tinbergen Institute Discussion Paper, No. TI 06-031/3*.

Din, A., Hoesli, M., & Bender, A. (2001). Environmental Variables and Real Estate Prices. *Urban Studies, 38*(11), 1989-2000.

DiPasquale, D., & Wheaton, W. C. (1992). The Markets for Real Estate Assets and Space: A Conceptual Framework. *Real Estate Economics, 20*(2), 181-198.

Domingos, P. (2015). *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World*. New York: Basic Books.

Downs, D., & Slade, B. (1999). *Characteristics of a Full-Disclosure, Transaction-Based Index of Commercial Real Estate* (Vol. 5).

Dunse, N., & Jones, C. (1998). A hedonic price model of office rents. *Journal of Property Valuation and Investment, 16*(3), 297-312.

Dunse, N., & Jones, C. (2002). The existence of office submarkets in cities. *Journal of Property Research, 19*(2), 159-182.

Dynamis. (2018). Sprekende Cijfers Kantorenmarkt.

European AVM Alliance. (2017). Standards for Statistical Valuation Methods for Residential Properties in Europe. London.

Fan, G.-Z., Ong, S. E., & Koh, H. C. (2006). Determinants of House Price: A Decision Tree Approach. *Urban Studies, 43*(12), 2301-2315.

Fisher, J. D., Geltner, D. M., & Webb, R. B. (1994). Value indices of commercial real estate: A comparison of index construction methods. *The Journal of Real Estate Finance and Economics, 9*(2), 137-164.

Fotheringham, S., Brunsdon, C., & Charlton, M. (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. West Sussex: Wiley.

Francke, M. (2017). *Advanced Valuation Models* (Vol. Reader): University of Amsterdam.

Francke, M., van Gool, P., & van de Minne, A. (2014). Prijsindex voor Commercieel Vastgoed: Gebasseerd op Beleggingstransacties. *Real Estate Research Quarterly, 13*(4).

Francke, M., & Vos, G. (2004). The Hierarchical Trend Model for Property Valuation and Local Price Indices. *The Journal of Real Estate Finance and Economics, 28*(2), 179-208.

Francke, M. K. (2008). *The hierarchical trend model.* Chichester: Wiley-Blackwell.

Fuerst, F. (2007). *Office Rent Determinants: A Hedonic Panel Analysis.* Real Estate & Planning Working Papers rep-wp2008-12. Henley Business School, Reading University.

Fuerst, F., & McAllister, P. (2011). Green Noise or Green Value? Measuring the Effects of Environmental Certification on Office Values. *Real Estate Economics, 39*(1), 45-69.

Gat, D. (1998). Urban Focal Points and Design Quality Influence Rents: The Case of the Tel Aviv Office Market. *The Journal of Real Estate Research, 16*(2), 229-247.

Geltner, D. M., & Bokhari, S. (2008). A technical note on index methodology enhancement by two-stage regression estimation. *Technical Report.*

Glascock, J. L., Jahanian, S., & Sirmans, C. F. (1990). An Analysis of Office Market Rents: Some Empirical Evidence. *Real Estate Economics, 18*(1), 105-119.

Graczyk, M., Lasota, T., Trawiński, B., & Trawiński, K. (2010). Comparison of Bagging, Boosting and Stacking Ensembles Applied to Real Estate Appraisal (pp. 340-350). Berlin, Heidelberg: Springer Berlin Heidelberg.

Greene, W. H. (2011). *Econometric Analysis* (I. Edition Ed.). Harlow: Pearson Education Limited.

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.

Hastie, T., Tibshirani, R., & Friedman, J. (2008). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (S. Edition Ed.): Springer.

Hodgson, D. J., Slade, B. A., & Vorkink, K. P. (2006). Constructing Commercial Indices: A Semiparametric Adaptive Estimator Approach. *The Journal of Real Estate Finance and Economics, 32*(2), 151-168.

Hough, D. E., & Kratz, C. G. (1983). Can "good" architecture meet the market test? *Journal of Urban Economics, 14*(1), 40-54.

Huang, B., Wu, B., & Barry, M. (2010). Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. *International Journal of Geographical Information Science, 24*(3), 383-401.

IAAO. (2013). *Standards on Mass Appraisal of Real Property*. Kansas City: International Association of Assessing Officers.

International Valuation Standards Comittee. (2017). Market Value Definition.

IVBN. (2017). Investment in Office Property in the Netherlands: a European Perspective. Den Haag.

Jeremy, H., & Bowles, M. (2012). *The Two Most Important Algorithms in Predictive Modeling Today*. Paper presented at the Strata Conference.

JLL. (2018). Office Property Clock. Amsterdam.

Kauko, T. (2002). *Modelling the Locational Determinants of House Prices: Neural Network and Value Tree Approaches* (doctoral), Utrecht University, Utrecht.

Kempf, S. (2015). *Development of Hedonic Office Rent Indice: Examples for German Metropolitan Areas*. Wiesbaden: Springer Gabler.

Kok, N., & Jennen, M. (2012). The impact of energy labels and accessibility on office rents. *Energy Policy, 46*, 489-497.

Kok, N., Koponen, E.-L., & Martínez-Barbosa, C. A. (2017). Big Data in Real Estate? From Manual Appraisal to Automated Valuation. *43*, 202-211.

Koster, H. R. A., van Ommeren, J., & Rietveld, P. (2014). Is the sky the limit? High-rise buildings and office rents. *Journal of Economic Geography, 14*(1), 125-153.

Krause, A., & Kummerow, M. (2011). An Iterative Approach to Minimizing Valuation Errors Using an Automated Comparable Sales Model. *Journal of Property Tax Assessment & Administration, 39*(2), 39-52.

Lin, C. C., & Mohan, S. B. (2011). Effectiveness comparison of the residential property mass appraisal methodologies in the USA. *International Journal of Housing Markets and Analysis, 4*(3), 224-243.

Lusht, K. (2012). *Real Estate Valuation: Principles and Applications*. State College: KML publishing.

Madden, S. (2012). From databases to big data. *IEEE Internet Computing, 4*(6).

Matysiak, A. (2017). The Accuracy of Automated Valuation Models (AVMs): TEGoVA.

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.

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.

Mills, E. S. (1992). Office Rent Determinants in the Chicago Area. *Real Estate Economics, 20*(2), 273-287.

Mooradian, R. M., & Yang, S. X. (2000). Cancellation Strategies in Commercial Real Estate Leasing. *Real Estate Economics, 28*(1), 65-88.

Mora-Esperanza, G. (2004). Artificial Intelligence Applied to Real Estate Valuation. *Catastro, April 1*, 255-265.

MSCI. (2015). Real Estate Index Analysis: Valuation and Sale Price Comparison Report: MSCI.

Mullainathan, S. (2018). Smarter Algorithms, Better Policy. from https://www.youtube.com/watch?v=cuGWl3t_1MI

Mullainathan, S., & Spiess, J. (2017). Machine Learning: An Applied Econometric Approach. *Journal of Economic Perspectives, 31*(2), 87-106.

Munneke, H. J., & Slade, B. A. (2001). A Metropolitan Transaction-Based Commercial Price Index: A Time-Varying Parameter Approach. *Real Estate Economics, 29*(1), 55-84.

Nappi-Choulet, I., Maleyre, I., & Maury, T. P. (2007). A Hedonic Model of Office Prices in Paris and its Immediate Suburbs. *Journal of Property Research, 24*(3), 241-263.

Nappi-Choulet, I., & Maury, T. P. (2009). A Spatiotemporal Autoregressive Price Index for the Paris Office Property Market. *Real Estate Economics, 37*(2), 305-340.

Ng, A. (2017). The State of Artificial Intelligence. Retrieved 1st January, 2018, from https://www.youtube.com/watch?v=NKpuX_yzdYs

Onur, Ö., & Hasan, Ş. (2009). Housing price determinants in Istanbul, Turkey: An application of the classification and regression tree model. *International Journal of Housing Markets and Analysis, 2*(2), 167-178.

Orr, A. M., Dunse, N., & Martin, D. (2003). Time on the market and commercial property prices. *Journal of Property Investment & Finance, 21*(6), 473-494.

Ortec Finance. (2017). A hybrid approach to house valuation.

Öven, V. A., & Pekdemir, D. (2006). Office Rent Determinants Utilising Factor Analysis—A Case Study for İstanbul. *The Journal of Real Estate Finance and Economics, 33*(1), 51-73.

Pace, R. K., Barry, R., Clapp, J. M., & Rodriquez, M. (1998). Spatiotemporal Autoregressive Models of Neighborhood Effects. *The Journal of Real Estate Finance and Economics, 17*(1), 15-33.

Pagourtzi, E., Assimakopoulos, V., Hatzichristos, T., & French, N. (2003). Real estate appraisal: a review of valuation methods. *Journal of Property Investment & Finance, 21*(4), 383-401.

Peterson, S., & Flanagan, A. B. (2009). Neural Network Hedonic Pricing Models in Mass Real Estate Appraisal. *Journal of Real Estate Research, 31*(2), 147-164.

Pivo, G., & Fisher, J. D. (2011). The Walkability Premium in Commercial Real Estate Investments. *Real Estate Economics, 39*(2), 185-219.

RICS. (2017). The Future of Valuations: The Relevance of Real Estate Valuations for Institutional Investors and Banks - Views from a European Expert Group. *RICS Insight*.

Rijksdienst voor het Cultureel Erfgoed. (2013). Kantoorgebouwen in Nederland 1945-2015: Cultuurhistorische en Typologische Quickscan.

Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy, 82*(1), 34-55.

Ryan, S. (2005). The Value of Access to Highways and Light Rail Transit: Evidence for Industrial and Office Firms. *Urban Studies, 42*(4), 751-764.

Schekkerman, C. (2004). *Nauwkeurigheid in Taxaties: Een onderzoek naar nauwkeurigheid van taxaties en (on)mogelijkheden om de betrouwbaarheid van taxaties te vergroten*. (MSRE Masterthesis), Amsterdam School of Real Estate, Amsterdam.

Sivitanidou, R. (1996). Do Office–Commercial Firms Value Access to Service Employment Centers? A Hedonic Value Analysis within Polycentric Los Angeles. *Journal of Urban Economics, 40*(2), 125-149.

Slade, B. (2000). Office Rent Determinants During Market Decline and Recovery. *Journal of Real Estate Research, 20*(3), 357-380.

Smith, C. (2017). *Decision Trees and Random Forests: A Visual Introduction For Beginners*. Canada: Blue Windmill Media.

Stevenson, S. (2004). New empirical evidence on heteroscedasticity in hedonic housing models. *Journal of Housing Economics, 13*(2), 136-153.

Tay, D. P. H., & Ho, D. K. H. (1992). Artificial Intelligence and the Mass Appraisal of Residential Apartments. *Journal of Property Valuation and Investment, 10*(2), 525-540.

Tobler, W. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography, 46*, 234-240.

Tu, Y., Yu, S. M., & Sun, H. (2004). Transaction-Based Office Price Indexes: A Spatiotemporal Modeling Approach. *Real Estate Economics, 32*(2), 297-328.

van Assendelft, N. (2017). *Rent Premiums and Vertical Sorting in Amsterdam's Tall Office Towers.* (masterthesis), Delft University of Technology.

van Gool, P., & Have, G. G. M. t. (2006). Luchtbellen in vastgoedwaarderingen door verkeerd gebruik taxatiemethoden? *AASRE research papers*.

Varian, H. R. (2014). Big Data: New Tricks for Econometrics. *Journal of Economic Perspectives, 28*(2), 3-28.

Webb, R. B., & Fisher, J. D. (1996). Development of an Effective Rent (Lease) Index for the Chicago CBD. *Journal of Urban Economics, 39*(1), 1-19.

Wheaton, W. C., & Torto, R. G. (1994). Office Rent Indices and Their Behavior over Time. *Journal of Urban Economics, 35*(2), 121-139.

White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica, 48*(4), 817-838.

Winson-Geideman, K., Krause, A., Lipscomp, C., & Evangelopoulos, N. (2018). *Real Estate Analysis in the Information Age: Techniques for Big Data and Statsitical Modeling*. London.

Witten, I. H., & Eibe, F. (2005). *Data Mining: Practical Machine Learning Tools and Techniques* (Vol. Second Edition): Morgan Kaufmann Publishers.

Worzala, E., Lenk, M., & Silva, A. (1995). An Exploration of Neural Networks and Its Application to Real Estate Valuation. *Journal of Real Estate Research, 10*(2), 185-201.

Ziermans, B. (2016). *De determinanten van incentives op de Amsterdamse Kantorenmarkt.* (MRE MRE thesis), Amsterdam School of Real Estate, Amsterdam.

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.

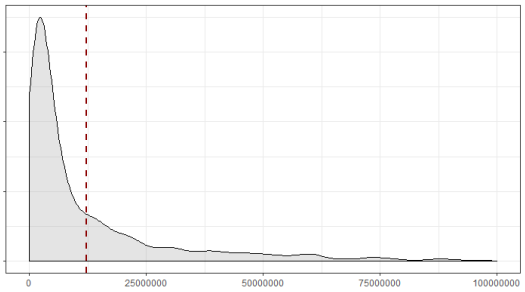
Appendix A: Exploratory Data Analysis – Numeric Variables

Bivariate Analysis

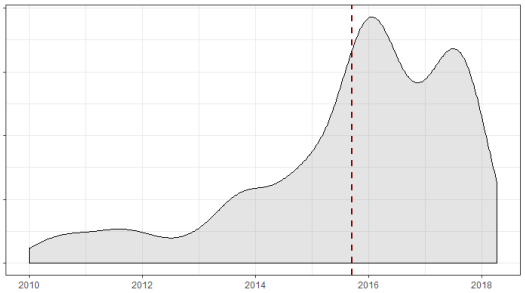
Multivariate Analysis

Note: Figures are before transformations and exclusion of outliers.

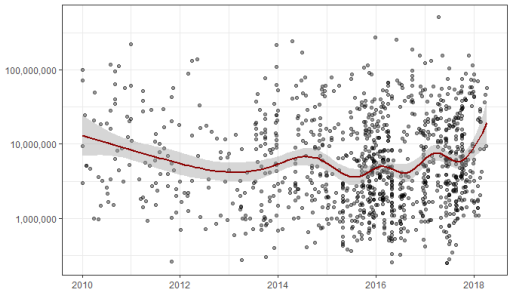
Transaction Price (xlimit: 100mil)



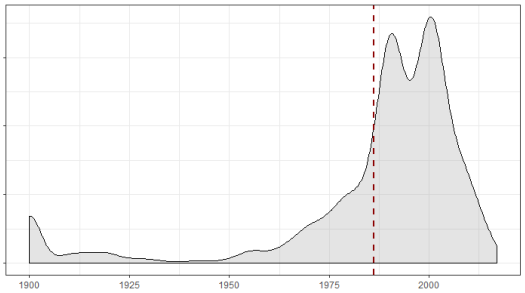
Transfer Date



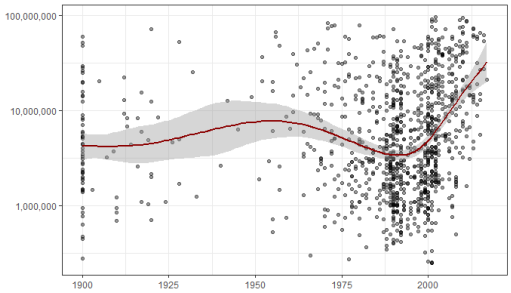
Transfer Date



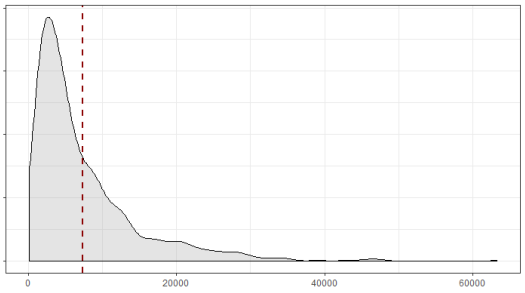
Year Built



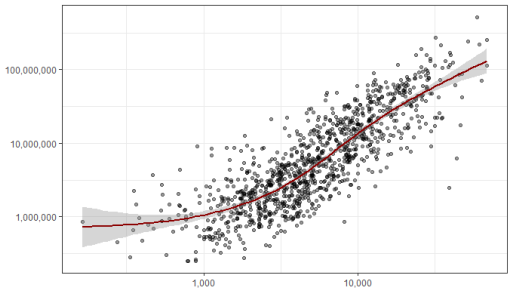
Year Built



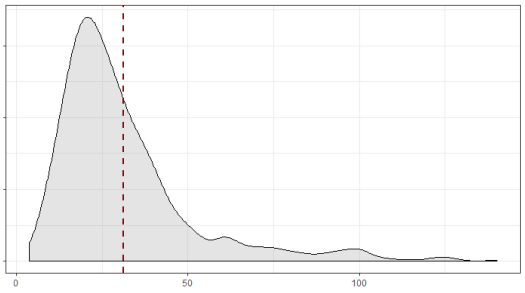
Total LFA (m2)



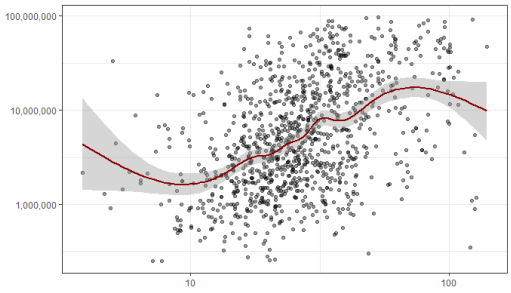
Total LFA (sqm)



Building Height (m1)



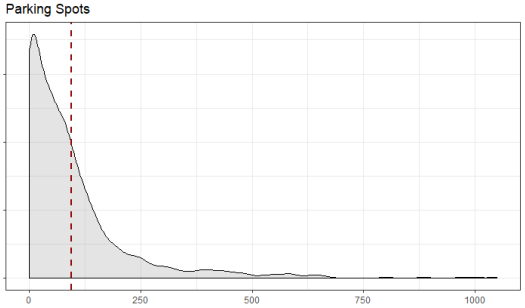
Building Height



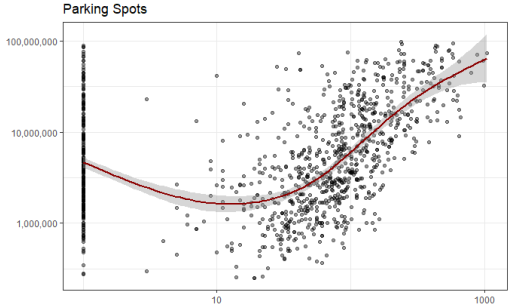
Bivariate Analysis

Multivariate Analysis

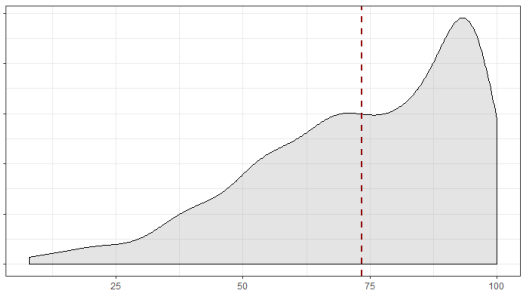
Parking Spots



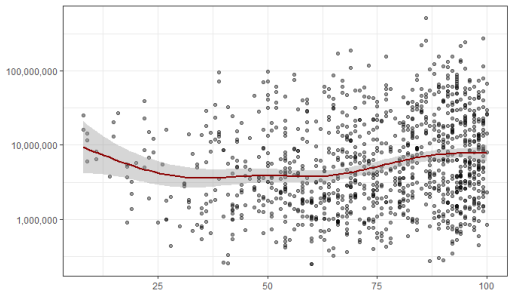
Parking Spots



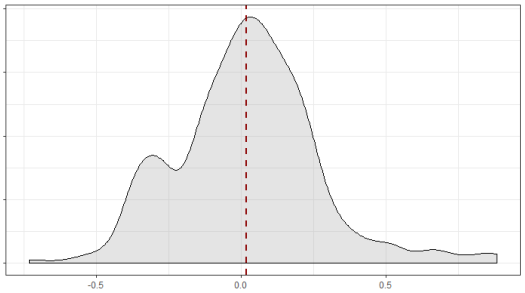
Walkscore



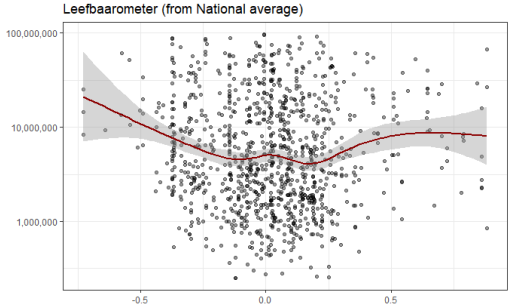
Walkscore



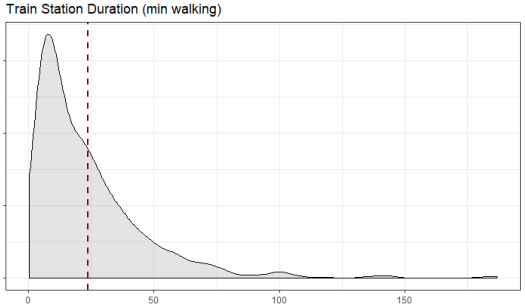
Leefbaarometer



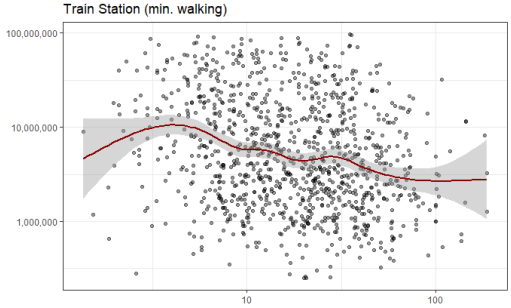
Leefbaarometer (from National average)



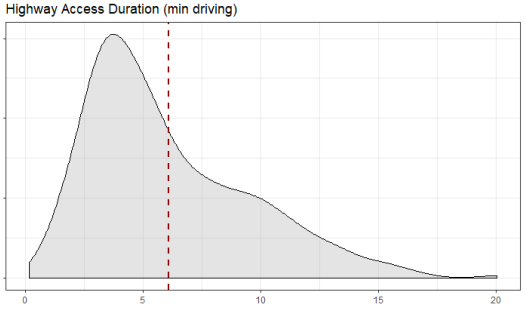
Train Station Duration (min walking)



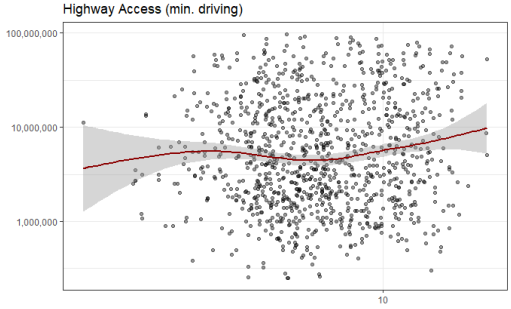
Train Station (min. walking)



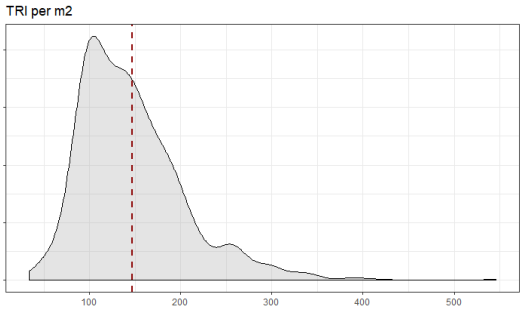
Highway Access Duration (min driving)

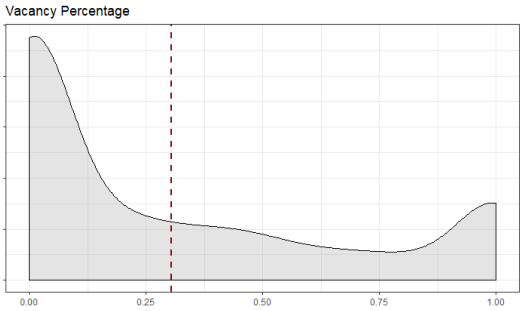


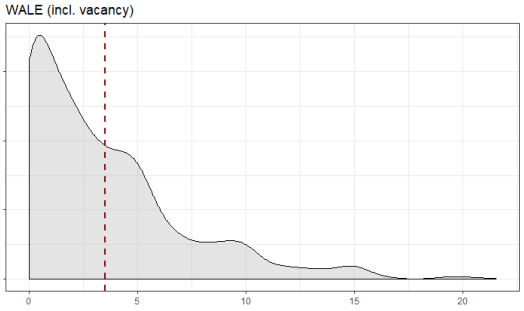
Highway Access (min. driving)



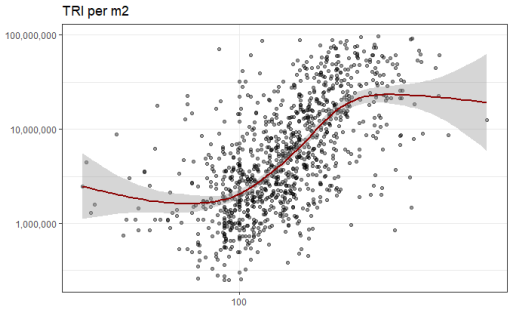
Bivariate Analysis

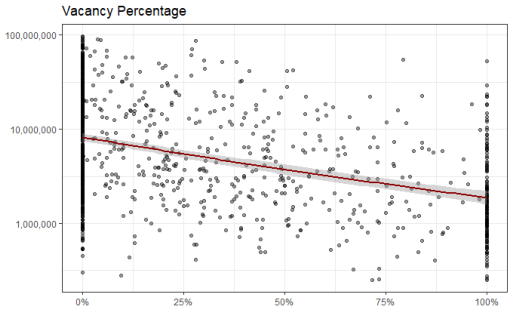


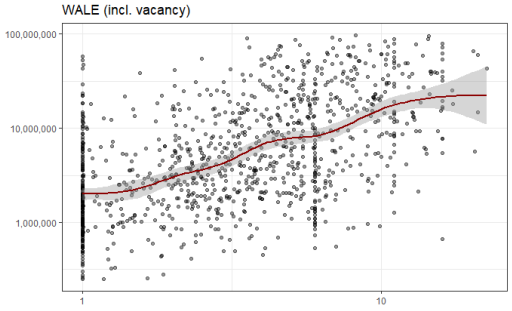




Multivariate Analysis



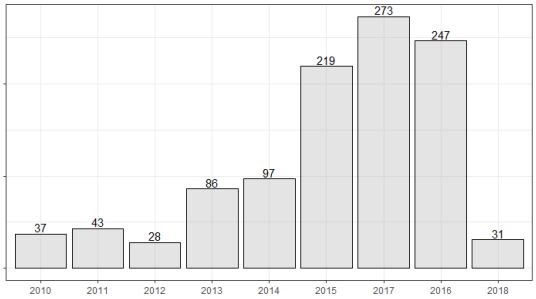




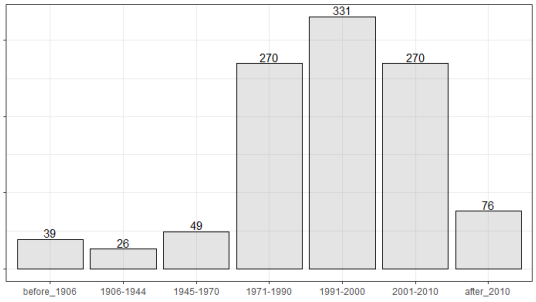
Appendix A: Exploratory Data Analysis – Categorical Variables

Bivariate Analysis

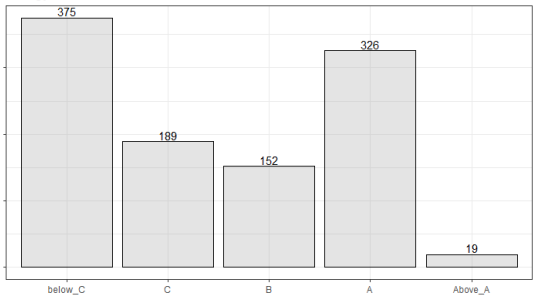
Year Transaction



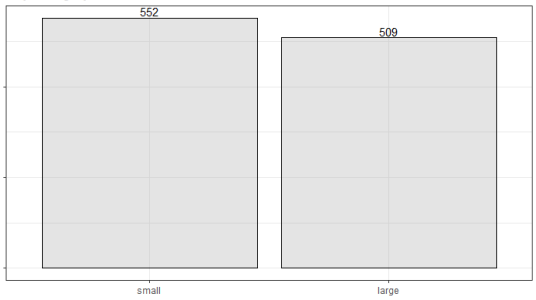
Year Built



Energy Label

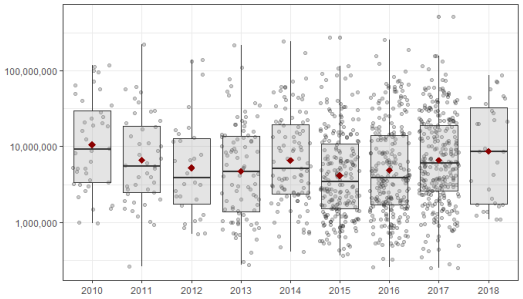


City Category

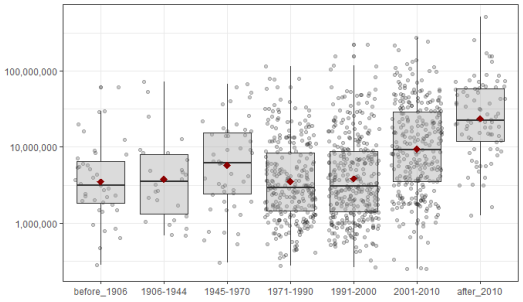


Multivariate Analysis

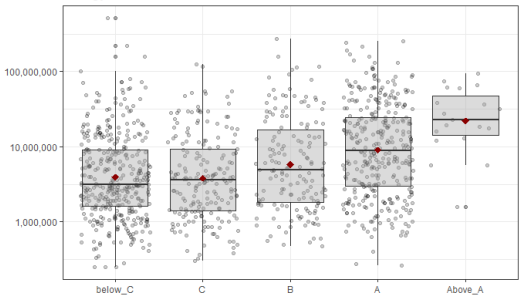
Year Transaction



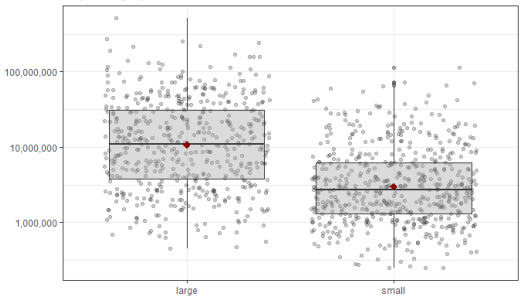
Year Built/Last Renovation



Energy Label

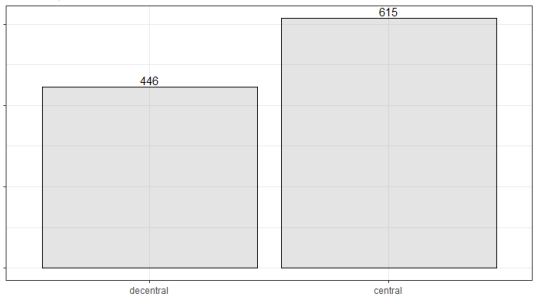


City Category

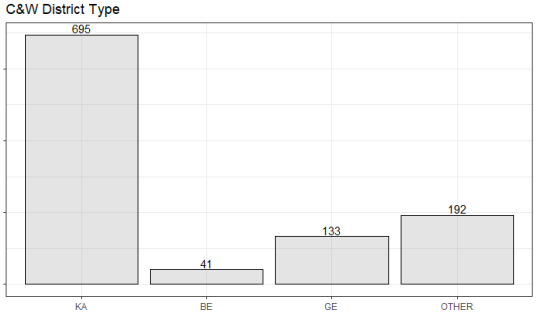


Bivariate Analysis

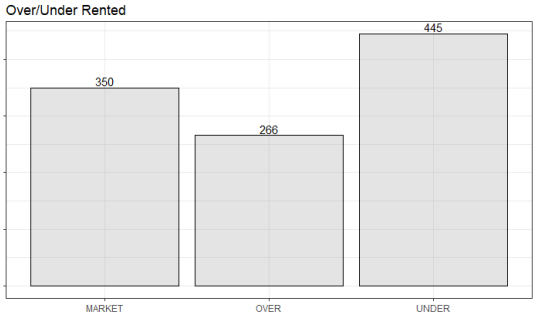
Centrality



C&W District Type

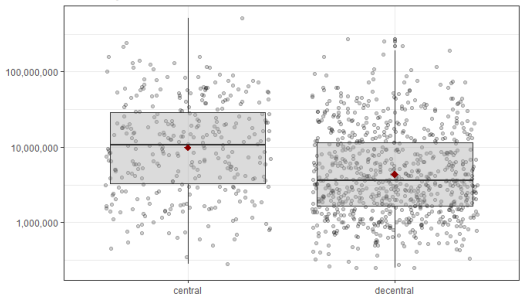


Over/Under Rented

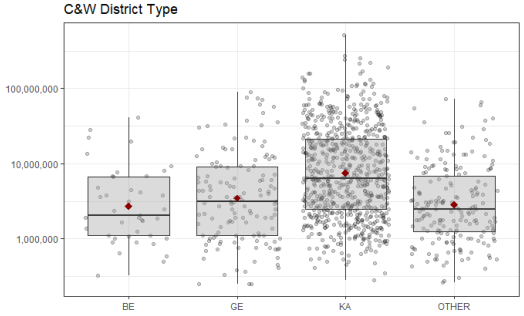


Multivariate Analysis

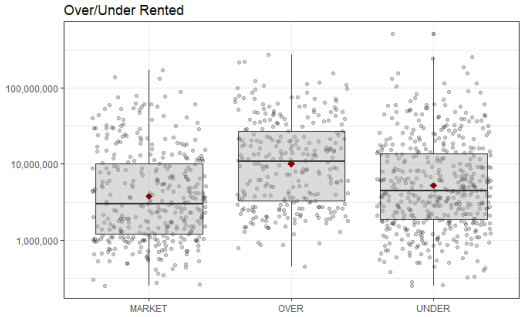
Centrality



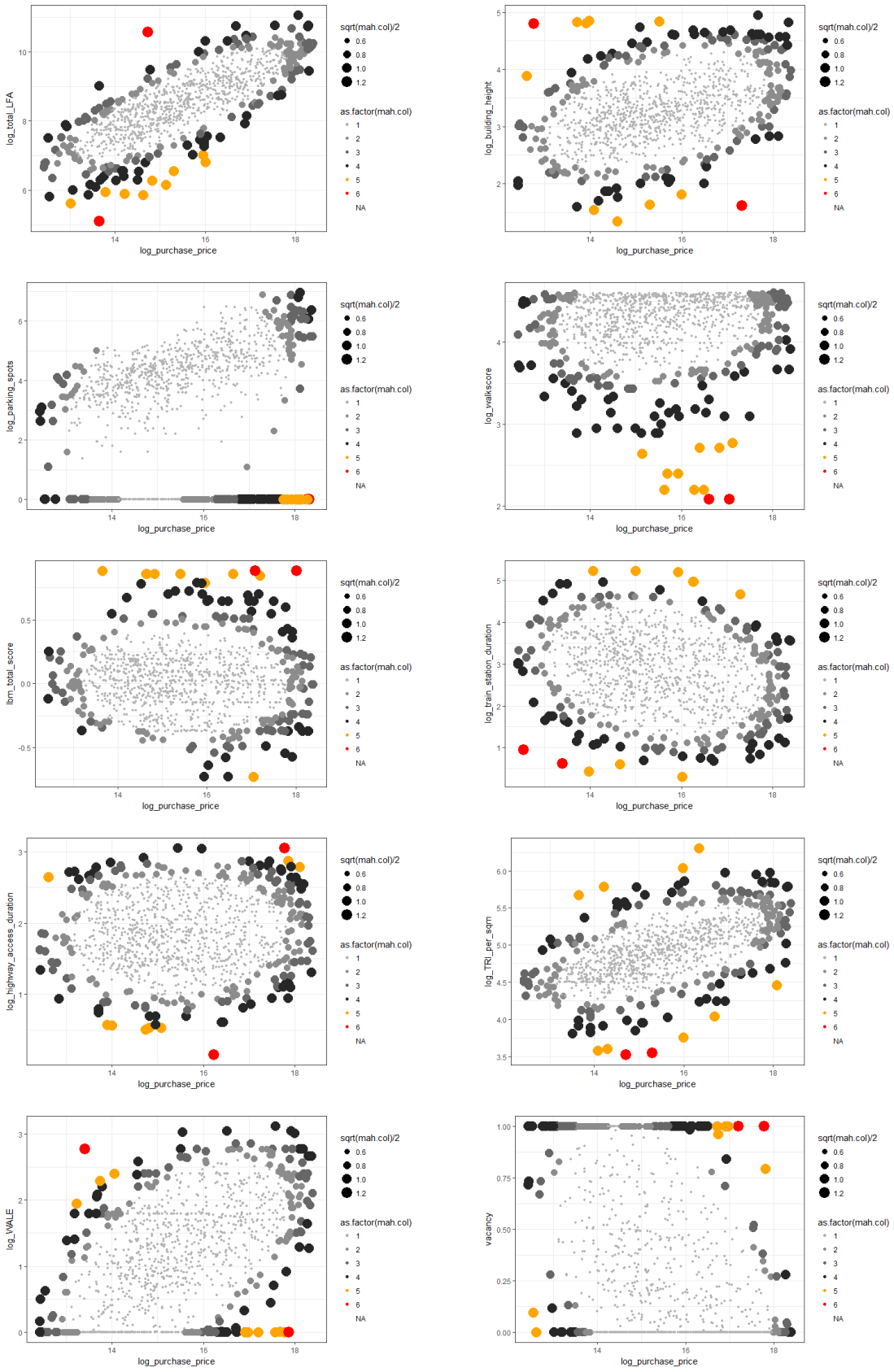
C&W District Type



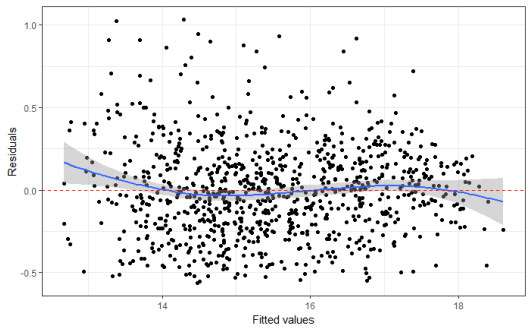
Over/Under Rented

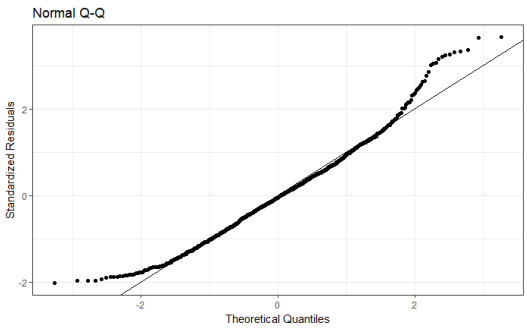


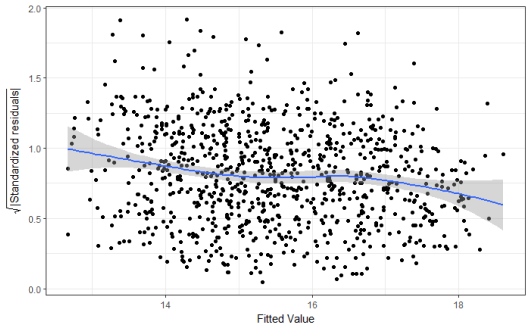
Appendix B: Outlier Analysis – Mahalanobis' Distance

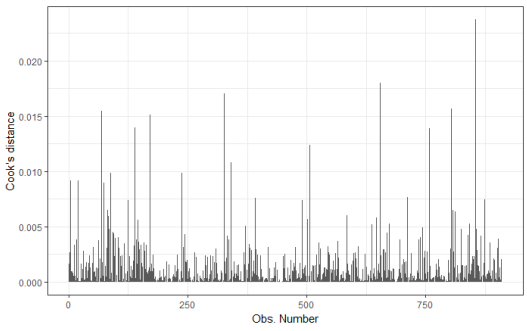


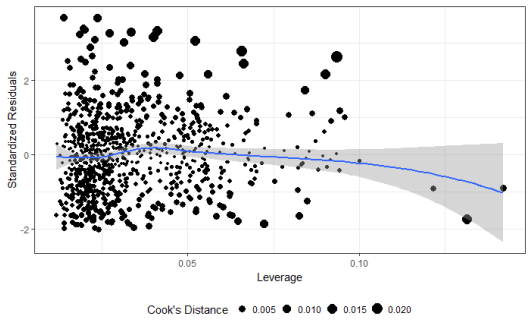
Appendix C: Residual Analysis

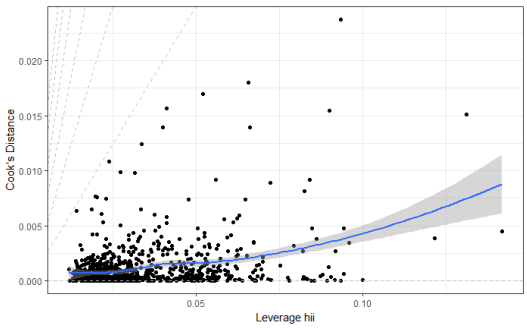








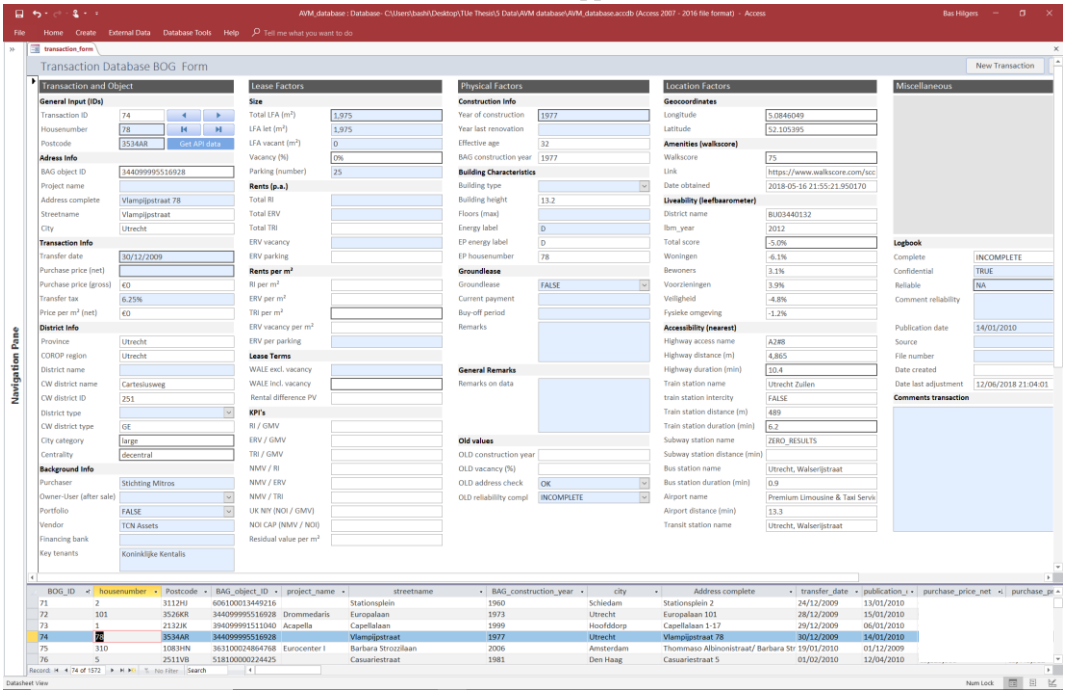




Appendix D: AVM Application

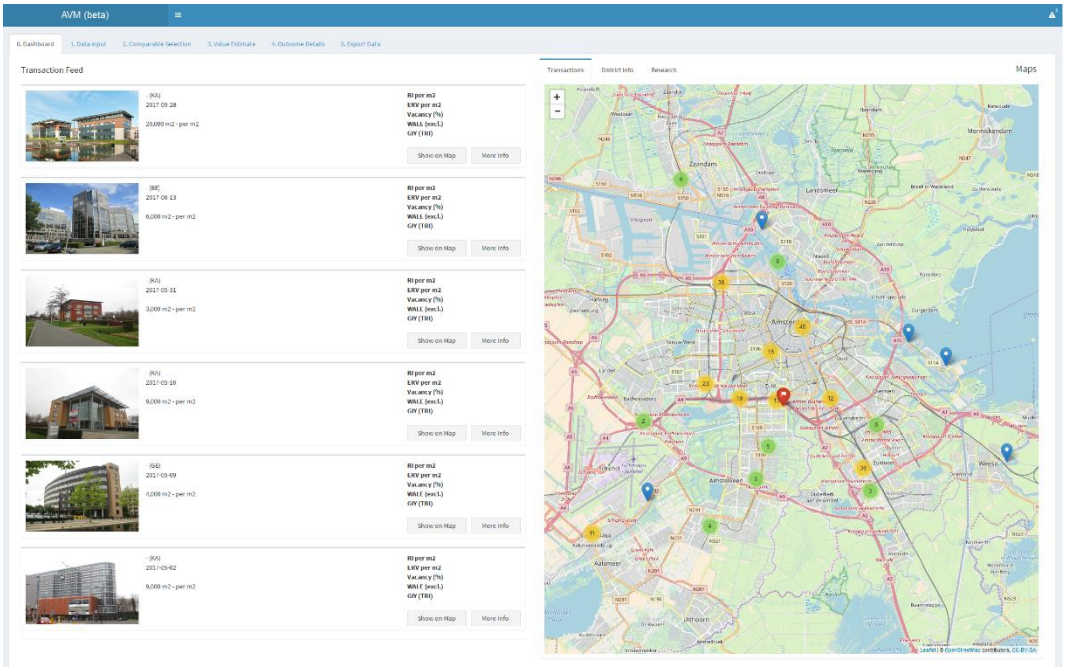
Data Collection

The input and validation of the AVM data is done manually in Access 365 from the office. The data is saved into an SQL database that is linked to the AVM application.



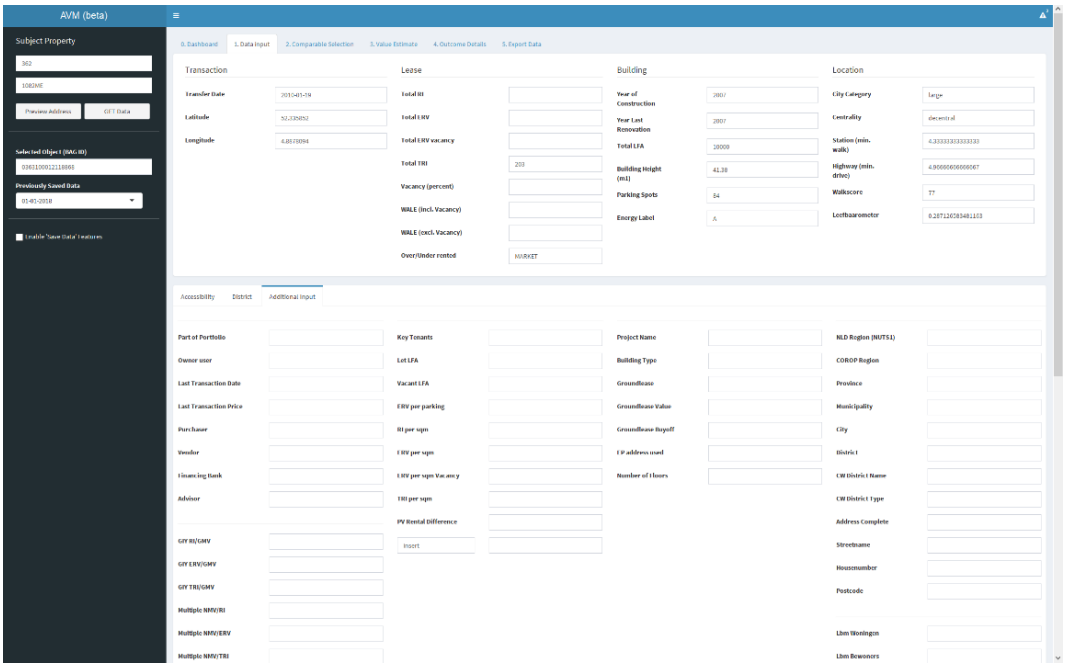
AVM Step 0: Dashboard

Newsfeed with latest transactions and including Spatial and Temporal analysis (e.g. prices, rents and yields in the area of the subject property)



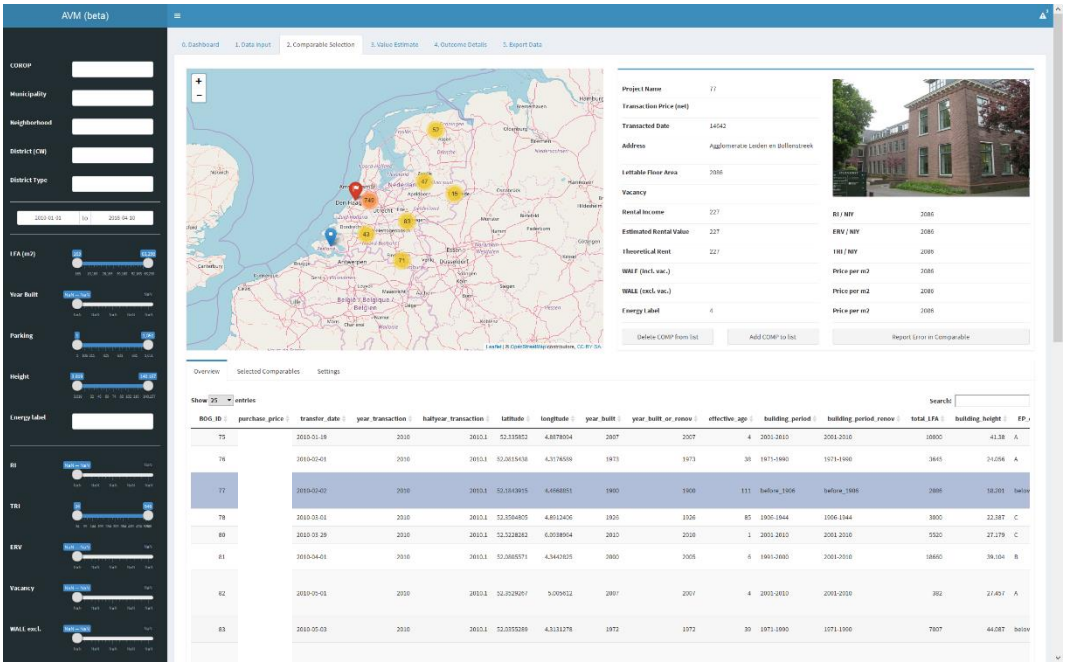
Step 1: Input Data

Input the data of the subject property of which we want to estimate the Market Value. As much data as possible is gathered through databases and API calls.



Step 2: Comparable Selection

The values are more derived from comparable properties than non-comparable properties through a weight matrix.



Step 3: Model Selection and Estimation

Confidential