Commercial Real Estate Pricing Dynamics:
A Machine Learning Analysis

Benedict von Ahlefeldt-Dehn[a], Juergen Deppner[a], Eli Beracha[b] and Wolfgang Schaefers[a]

Author Information

[a]IRE|BS International Real Estate Business School, University of Regensburg, Regensburg, Germany; [b]Hollo School of Real Estate, Florida International University, Miami, United States

Benedict von Ahlefeldt-Dehn, PhD Candidate and Research Assistant, University of Regensburg, +49-157-8459-4063, benedict.vonahlefeldtdehn@irebs.de

Juergen Deppner, PhD Candidate and Research Assistant, University of Regensburg, +49-172-289-5333, juergen.deppner@irebs.de

Eli Beracha, Professor of Real Estate, Florida International University, 305-779-7906, eberacha@fiu.edu

Wolfgang Schaefers, Professor of Real Estate Management, University of Regensburg, +49-943-941-5071, wolfgang.schaefers@irebs.de

Acknowledgments

This article received generous support and funding by the Real Estate Research Institute (RERI), a part of the Pension Real Estate Association (PREA). Data was kindly provided by the National Council of Real Estate Investment Fiduciaries (NCREIF). The authors sincerely thank Jeffrey Fisher for data access, and the RERI mentors James Chung, Heidi Learner, Mark Roberts, and Timothy Savage for their helpful comments.

Abstract

Machine learning (ML) algorithms that provide the analytical core of automated valuation models (AVMs) have demonstrated thus far unprecedented accuracy in the estimation of house prices, given an abundance of data recorded by multiple listing services and the development of increasingly sophisticated methods in recent years. Despite the vast potential that AVMs offer, their implementation in the institutional sector is progressing slowly as practitioners are reluctant to rely on these techniques. In contrast to the residential sector, little is known about the usefulness of such data-driven methods in commercial real estate markets where the availability of structured data is still very limited due to market intransparency and property heterogeneity. Moreover, the adoption of these techniques for institutional use is hampered as their mechanisms are black boxes in the sense that an inherent comprehensibility of their predictions is impeded by the complexity of their architectures. This is problematic as regulatory bodies and authorities require transparency. The objective of this study is to propose a holistic framework for the practical use of AMVs in a commercial real estate context that considers both the accuracy and interpretability of the estimation method. For this purpose, we train a deep neural network (DNN) on a unique sample of more than 400,000 property-quarter observations from the NCREIF Property Index (NPI) and perform model-agnostic analysis using “Shapley Additive exPlanations" (SHAP) to provide ex-post comprehensibility of the algorithm’s prediction rules. In doing so, we furthermore assess to which extent the inner workings of the DNN follow an economic rationale and set out how the proposed methods can add to the understanding of pricing processes in institutional investment markets. By addressing the caveats and illustrating the potential of ML in the field of commercial real estate, the contribution of this study represents another important pillar in the practical implementation of AVMs.

Keywords: automated valuation models; commercial real estate; interpretable machine learning

BACKGROUND

The estimation of real estate prices and identification of relevant price determinants in property markets remain a complex task due to the inherent heterogeneity of properties and the multiplicity of factors that influence their values. As stated by Quan and Quigley (1991), market mechanisms are obfuscated by "[…] a noisy signal, reflecting incomplete information as well as the conditions of sale", given that real estate markets are illiquid, opaque, and individual agents in the market are only infrequently engaged in transactions. Appraisers are tasked with extracting important information (i.e., the signal) from irrelevant data (i.e., the noise) by using their expert knowledge about the market, which is justified by their individual experience from observing past transactions. Consequently, pricing processes must be disentangled based on limited information and subjective judgements of price determinants that a valuer considers relevant, resulting in imprecise and biased valuations (Dunse and Jones, 1998; Cannon and Cole, 2011).

This gave rise to hedonic pricing models introduced by Rosen (1974) as the prevalent framework to analyze the mechanisms behind property pricing processes more objectively from an econometric point of view. Parametric hedonic models, such as proposed by Mills (1992), Sirmans and Guidry (1993), or Lockwood and Rutherford (1996), utilize linear regression methods to estimate property prices based on intrinsic property characteristics (e.g., location, size, amenities). Literature has demonstrated the efficiency and ease of interpretability of hedonic models in revealing relevant property price determinants. However, they are also built on strict assumptions which are unlikely to hold. In addition, the models require a fixed additive functional form between the property value and the explanatory variables that needs to be specified a-priori and entails a high risk of misspecification. As explained by Dunse and Jones (1998), hedonic prices may vary across space and time and can thus not be assumed to be constant. Other concerns refer mostly to the non-linearity of pricing processes that cannot be adequately captured with linear models. Studies by Grether and Mieszkowski (1974), Do and Grudnitski (1993), and Goodman and Thibodeau (1995) identify significant non-linearities between property prices and the building age as well as the square footage, demonstrating that complex relationships between property prices and features cannot be simply reduced to a single, invariant beta coefficient.

As data becomes more readily available and artificial intelligence (AI) continues to advance, both industry and academia have witnessed a shift towards the consideration of more adaptable machine learning (ML) techniques for determining property values. This shift has become evident in the form of automated valuation models (AVMs), which have gained importance in the sector, particularly in residential real estate, given the increased flexibility in the underlying models. In the literature, ML-based AVMs have repeatedly demonstrated a so far unprecedented level of accuracy in their predictions and do not require any judgement concerning the model's functional form as they are designed to autonomously find complex non-linear relationships in the data and optimize model fit.

Yet the adoption of ML in the industry in general, and the institutional sector in particular, is facing some critical issues. First, ML techniques are heavily data-dependent and therefore rely on large amounts of data to produce reliable and consistent results, as demonstrated by Worzala et al. (1995). In contrast to the residential domain, data availability is still limited in the commercial sector, which is particularly problematic due to the high heterogeneity of commercial property types (Deppner et al., 2023). Second, the models are criticized for lacking an economic justification and do not foresee any form of intrinsic interpretability (e.g., Din et al., 2001; McCluskey et al., 2013; Valier, 2020). This refers to the fact that these models are purely data-driven, allowing them to make predictions from any combination of data (Rico-Juan and Taltavull de la Paz, 2021), while their complex and opaque architectures impede understanding of how the algorithm arrived at a particular valuation, and how the input factors have affected the outcome. This hampers comprehensibility of the models and prohibits to draw inference on price determinants, making it difficult for practitioners to trust and rely on AVMs, particularly given that regulators and authorities demand transparency in the process of estimating market values.

The current state of research suggests three ways to address this problem. The first is to reduce the complexity of the applied models to such an extent that their interpretability is preserved. However, this makes the models more sensitive to changes in the data and increases the tendency of overfitting, resulting in poor out-of-sample performance (Kok et al., 2017; Pace and Hayunga, 2020; Lorenz et al., 2022). Second, ML can be used to provide constructive criticism, such as in variable selection, model specification (e.g., Yoo et al., 2012; Perez-Rave et al. 2019), or model selection (e.g., Pace and Hayunga, 2020), which can help to improve upon traditional models. However, this means giving up the flexibility and accuracy of ML models for the sake of interpretability. The third alternative is to apply model-agnostic interpretation techniques that can decipher the black box of ML models, thus enabling post-hoc interpretability while maintaining accuracy and precision, as shown by Levantesi and Piscopo (2020), Rico-Juan and Taltavull de la Paz (2021), Lorenz et al. (2022) as well as Potrawa and Teterava (2022).

The aim of this study is to expand upon this discussion by proposing a novel and comprehensive framework for utilizing AVMs in commercial real estate that balances both precision and comprehensibility. To achieve this, we train four deep neural networks (DNNs) on a large data sample comprising over 400,000 property-quarter observations from the asset sectors apartment, industrial, office, and retail, and apply model-agnostic analysis using “Shapley Additive exPlanations” (SHAP) to provide clear insight into the prediction rules of the algorithms. In doing so, we furthermore assess to which extent the inner workings of the DNNs follow economic principles and set out how the proposed methods can add to a deeper and more nuanced understanding of pricing mechanisms in institutional investment markets by revealing non-linear and three-dimensional relationships in the value drivers of commercial real estate.

The contributions of the study are relevant and timely for both academia and practice for several reasons. While we do not believe that AVMs are developed to the point where they can substitute manual appraisers in the foreseeable future, the underlying technology still exhibits a high disruptive potential and is likely to reshape the multi-billion-dollar valuation industry in the years to come (Kok et al., 2017). Especially in the commercial domain, where valuations are more complex and need to be executed frequently, these techniques can generate valuable insights to support data-driven decision-making, and thus leverage efficiency in both markets and business processes by increasing the speed and scale of valuations, reducing the cost of transactions, and ultimately increasing transparency in pricing processes. Market participants that incorporate such technologies into their business processes earlier than their competitors will be able to considerably streamline their processes and gain a competitive edge.

DATA

The principal study data comprises quarterly, property-level observations of all properties included in the NCREIF Property Index (NPI) from the first quarter of 1978 to the first quarter of 2021, provided by the National Council of Real Estate Investment Fiduciaries (NCREIF). The NPI is the oldest and most widely followed commercial real estate investment index in the United States and covers institutionally owned commercial real estate properties across the asset sectors apartment, hotel, industrial, office, and retail. The properties included in the index fluctuate over time as properties enter the database upon purchase and leave the database upon sale. This constitutes an initial unbalanced sample of 648,098 property-quarter observations across 30,254 individual properties, for which we record the corresponding market values, a series of structural and physical attributes, as well as cash flows. Due to limited data availability, we exclude non-operating properties and hotels from the initial sample.

We account for missing and erroneous data as follows. Observations with market values, square footages and construction years reported as less than or equal to zero are regarded as data errors and are dropped. Likewise, observations with occupancy rates taking values below zero or higher than one are removed. Furthermore, observations with missing values for the square footage, the construction year, the occupancy rate, the net operating income (NOI), the capital expenditures (Capex), and the property subtype were omitted, as these represent the main explanatory variables from the raw NCREIF dataset. After scaling market values, NOI, and Capex by the property's square footage, we note that remaining errors and anomalies in the data seem to be concentrated at the tails of the distribution of market values per square foot. For this reason, we follow Calainho et al. (2022) and cut off the lower and upper percentile of the distribution for each property type.

We subsequently enrich the cleaned data with a set of new variables. First, we calculate the building age as the difference between the valuation date and the construction date as well as the cumulative sum of a property's capital expenditures scaled by square footage as a proxy for building quality. We also note that NOIs can fluctuate materially over the holding period and in individual quarters. Since the average property in our sample has a five-year holding period, we use the eight-quarter moving average of the properties' NOIs as a proxy for stabilized income.

As demonstrated repeatedly in the literature, location is an important determinant of real estate values. We geocode our sample using the property addresses to retrieve the distances to relevant points of interest (POIs). Around 12.1% of the addresses could not be geocoded because of missing or incomplete addresses, so we drop those observations. For the remaining properties, we source a set of relevant POIs that are expected to cause either a premium or a discount to their surrounding area. For optimal data coverage, we use both Google Places and Open Street Maps (OSM) to retrieve the data and calculate the shortest distance from each property to the respective POIs. We subsequently cluster POIs that are similar into categories. This helps to avoid missing data and to reduce the dimensionality of the regressor matrix, making the models more interpretable and more efficient. **Exhibit 1** provides a summary of the POI clusters.

EXHIBIT 1

Clustering of POIs

Category	POI	Source
Public Transport	Bus Station	Google
	Subway Station	Google
	Light Rail Station	Google
	Train Station	Google
	Public Transport	OSM
Negative Externalities	Prison	OSM
	Graveyard	OSM
	Gas Station	Google, OSM
Food Establishments	Restaurant	Google, OSM
	Cafe	Google, OSM
Healthcare Provider	Pharmacy	Google, OSM
	Doctor	Google
Retail Stores	Shopping Mall	Google, OSM
	Department Store	Google, OSM
Food Stores	Supermarket	Google, OSM
	Convenience Store	Google, OSM
Nightlife Venue	Bar	Google, OSM
	Nightclub	Google, OSM
Educational Institutions	Kindergarten	OSM
	School	Google, OSM
Cultural Institutions	Museum	OSM
	Attraction	OSM
Service Establishments	Bank	Google, OSM
	Post Office	Google, OSM
Fitness	Gym	Google, OSM
	Fitness Centre	OSM
Park	Park	Google, OSM

In addition, we collect macroeconomic data to control for market cycles and varying economic conditions. This includes the ten-year government bond yield as well as the four-quarter percentage change in the gross domestic product (GDP) at the state-level retrieved from the database of the Federal Reserve Bank of St. Louis, the four-quarter percentage change in construction costs by region retrieved from the U.S. Census Bureau, and the four-quarter percentage change in employment at the county-level retrieved from the U.S. Bureau of Labor Statistics. We also collect quarterly real estate market data by property type from NCREIF, that is, market value cap rates, market vacancy rates, and market rental growth rates. Furthermore, we include a dummy indicator for different market cycles during the sample period to better control for shocks and the effect of cyclical movements in the overall market. Market cycles are defined as periods of consecutive positive (i.e., rising markets), or negative (i.e., falling markets) quarterly capital appreciation returns derived from the NCREIF Property Index (NPI).

In a last step, we exclude CBSA codes with fewer than ten properties of the same property type to prevent overfitting. The final study sample consists of 400,370 quarterly market value observations across 18,868 individual properties and is balanced across 30 explanatory variables that are presented in the summary statistics in **Exhibit 2** and **Exhibit 3**. Missing and erroneous data seem to be concentrated in the early years of the initial sample, as the final study data is ranging from the first quarter of 1991 to the first quarter of 2021, covering a period of 30 years.

EXHIBIT 2
Descriptive Statistics of Numerical Variables
All Property Types (N = 402,490)

Variable	Unit	Mean	Sd	Min	1st Q.	Median	3rd Q.	Max
Market Value	[$/SqFt]	189.54	198.54	18.57	71.60	125.40	229.63	2,634.53
SqFt	[k]	283.08	371.09	1.50	109.50	200.64	341.25	22,119.56
Building Age	[Years]	20.77	16.78	0.00	10.00	17.00	27.00	156.00
Occupancy	[%]	0.92	0.12	0.00	0.90	0.96	1.00	1.00
NOI	[$/SqFt]	2.62	2.45	-48.58	1.13	1.90	3.43	73.74
Stabilized NOI	[$/SqFt]	2.60	2.28	-19.69	1.14	1.89	3.39	56.26
CapEx	[$/SqFt]	0.77	2.91	0.00	0.00	0.14	0.59	311.02
CapEx Cumulative Sum	[$/SqFt]	13.20	40.51	0.00	0.41	3.34	11.65	1,802.37
Longitude	[°]	-96.14	17.66	-158.12	-117.53	-93.24	-80.36	-68.75
Latitude	[°]	36.85	5.27	19.63	33.58	37.48	40.72	61.56
Public Transport	[km]	1.70	2.00	0.00	0.32	1.06	2.29	12.99
Negative Externalities	[km]	0.76	0.59	0.00	0.36	0.62	1.00	7.95
Food Establishments	[km]	0.36	0.44	0.00	0.07	0.22	0.50	7.20
Healthcare Provider	[km]	0.42	0.65	0.00	0.08	0.22	0.51	11.93
Retail Stores	[km]	0.92	1.05	0.00	0.24	0.61	1.23	12.93
Food Stores	[km]	0.61	0.55	0.00	0.21	0.46	0.84	8.45
Nightlife Venue	[km]	0.78	0.95	0.00	0.20	0.51	1.06	12.36
Educational Institutions	[km]	0.49	0.52	0.00	0.17	0.35	0.63	8.25
Cultural Institutions	[km]	2.12	1.96	0.00	0.77	1.65	2.84	12.96
Service Establishments	[km]	0.70	0.74	0.00	0.18	0.47	1.00	8.16
Fitness	[km]	0.69	0.84	0.00	0.19	0.44	0.90	12.85
Park	[km]	0.79	0.84	0.00	0.30	0.59	1.00	12.85
GDP yoy	[%]	0.02	0.03	-0.11	0.01	0.02	0.04	0.22
Gov. Bond Yield	[%]	0.03	0.02	0.01	0.02	0.03	0.04	0.08
Construction Cost yoy	[%]	0.03	0.05	-0.10	0.01	0.04	0.05	0.20
Employment yoy	[%]	0.01	0.03	-0.50	0.00	0.01	0.03	1.10
Market Cap Rate qoq	[%]	0.06	0.01	0.04	0.05	0.06	0.07	0.10
Market Vacancy qoq	[%]	0.08	0.03	0.03	0.06	0.07	0.10	0.17
Market NOI Growth qoq	[%]	0.01	0.03	-0.32	-0.01	0.01	0.02	0.14

EXHIBIT 3
Descriptive Statistics of Categorical Variables
All Property Types (N = 402,490)

Variable	N	Percent
Property Type		
... Apartment	88,442	21.97%
... Industrial	151,109	37.54%
... Office	99,271	24.66%
... Retail	63,668	15.82%
Property Subtype		
... Garden	55,566	13.81%
... High-rise	26,889	6.68%
... Low-rise	5,987	1.49%
... Research and Development	6,049	1.50%
... Flex Space	17,054	4.24%
... Manufacturing	729	0.18%
... Other	2,328	0.58%
... Office Showroom	440	0.11%
... Warehouse	124,509	30.93%
... Central Business District	23,114	5.74%
... Suburban	76,157	18.92%
... Community Center	17,757	4.41%
... Theme/Festival Center	167	0.04%
... Fashion/Specialty Center	2,951	0.73%
... Neighborhood Center	23,511	5.84%
... Outlet Center	113	0.03%
... Power Center	6,776	1.68%
... Regional Mall	4,843	1.20%
... Super-Regional Mall	4,319	1.07%
... Single-Tenant	3,231	0.80%
Market Cycle		
... 1991Q1-1994Q1 (Gulf Crisis)	6,324	1.57%
... 1994Q2-2001Q3	47,506	11.80%
... 2001Q4-2002Q2 (Dotcom Crisis)	8,310	2.06%
... 2002Q3-2008Q1	80,138	19.91%
... 2008Q2-2010Q1 (Subprime Crisis)	35,742	8.88%
... 2010Q2-2020Q1	201,418	50.04%
... 2020Q2 (Covid-19 Pandemic)	5,565	1.38%

METHODOLOGY

To understand pricing processes in commercial real estate markets it is important that the selected models adequately capture relationships in the data but are still generalizable enough to predict well out-of-sample. Studies that apply different ML methods show that particularly artificial neural networks (ANNs) tend to produce robust and accurate predictions when applied in combination with sufficient data (e.g., Peterson and Flanagan, 2009; Zurada et al., 2011; Antipov and Pokryshevskaya, 2012; Baldominos et al., 2018; Mayer et al., 2019; Hu et al., 2019).

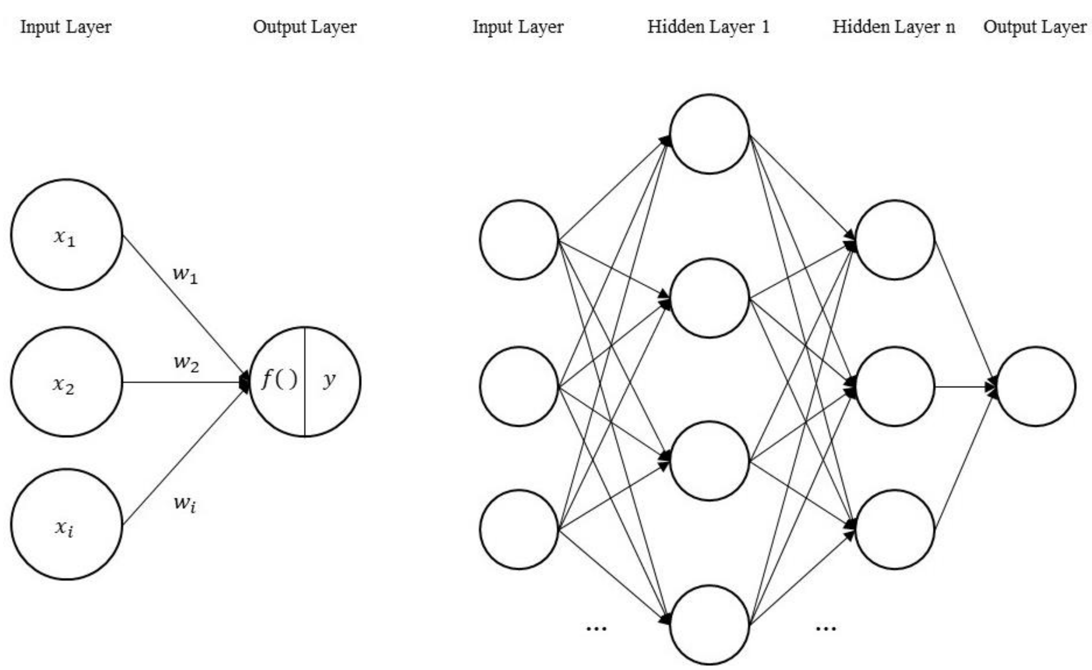
Machine Learning Approach – Artificial Neural Networks

The structure of an ANN is inspired by the human brain and consists of several artificial neurons that form layers which are interconnected. In its simplest form, a single-layer ANN consisting of only one input and one output layer with a linear activation function *f* as depicted in **Exhibit 4, Panel A** can be described as a linear regression. The bias *b* and the weights w_i of the input values x_i can then be compared to the intercept and the beta coefficients in an ordinary least squares (OLS) regression and formulate the prediction *y* as exhibited in **Equation 1** below.

$$y = f\left(\sum_{i=1}^{n} w_i x_i + b\right) \tag{1}$$

EXHIBIT 4
Structure of Neural Networks

Panel A: Single-layer ANN | **Panel B: Multi-layer DNN**



The optimal model fit is found by adjusting the weights of each neuron in an iterative process in the attempt to minimize a loss function that measures the distance from the actual to the predicted values. By adding hidden layers with multiple neurons and choosing other than linear activation functions, the model structure can be made more complex and interaction-effects as well as non-linearity are introduced. This is referred to as a deep neural network (DNN) as depicted in **Exhibit 4, Panel B**.

Model Agnostic Analysis – Shapley Additive Explanations

Interpretable machine learning (IML) methods are model agnostic techniques for interpreting opaque ML models to achieve ex-post transparency. That is, with the help of IML we can fundamentally understand how the model produces a particular outcome. A technique named "Shapley Additive exPlanations" (SHAP) introduced by Lundberg and Lee (2017) is capable of describing the marginal influences of a model's features on the prediction. SHAP is conceptually based on Shapley values, a coalitional game theory approach to determine the marginal contributions of each player to the outcome of a collaborative game (Shapley, 1953). Transferred to a ML context, Shapley values can be thought of as the average marginal contribution of a feature (i.e., "player") in a ML model (i.e., "game") on its prediction (i.e., "outcome"), as described by Molnar (2020). Shapley values are derived by repeatedly simulating different combinations of input features (i.e., "coalitions") and assessing how changes to the coalitions correspond to the final model predictions. This is done for each possible coalition in the model, such that a features' impact on the model prediction is eventually calculated as the average marginal contribution to the overall model score.

Model Estimation

We choose to estimate a separate DNN for each property type due to the peculiarities of the different sectors. The process of model estimation can generally be divided into two parts. The first part involves data transformation, training and tuning of the model. The second part involves out-of-sample performance testing to ensure generalizability of the results.

First, the initial sample is split into 60% training data, 20% validation data, and another 20% test data. Subsequently, all numeric explanatory variables in the three subsamples are z-score standardized by removing each feature values' mean and scaling to unit variance. Each model is trained as a sequential feedforward DNN with a variable number of hidden layers and neurons. To determine the number of hidden layers and neurons as well as the dropout and learning rate, the model is trained and validated on the respective training and validation subsamples. Hyperparameters are tuned with a Bayesian optimization tuner. This is a hyperparameter tuning framework which predicts parameter combinations and generally leads to better configurations with fewer evaluations than grid or random search when dealing with large datasets and a wide range of trainable parameters. Subsequently, the model with the best hyperparameter combination is trained once more on the whole training set (i.e., training and validation data aggregated) and out-of-sample performance is assessed on the remaining 20% test subsample. To evaluate the performance of the DNN in the application context, we estimate a linear regression model as a point of reference.

The estimation and performance evaluation of the DNN is then complemented using SHAP. This facilitates the interpretability and comprehensibility of the model's prediction rules.

Performance Evaluation

To assess model performance, we use the mean absolute percentage error (MAPE), the mean percentage error (MPE), the mean absolute error (MAE), the mean squared error (MSE), the root mean squared error (RMSE), the coefficient of determination (R^2) and two error buckets that show the proportion of absolute percentage errors below 10% (PE10) and 20% (PE20) respectively. MAPE and MAE are direct measures of accuracy (i.e., absolute distance) while MSE and RMSE assess the models' performance for very high values in the test data as high errors are penalized more (i.e., squared distance). MPE measures the biasedness of the model (i.e., whether the model's predictions generally tend to be higher or lower than the actual values) and R^2 is utilized as a measure of overall model fit. Additionally, the error buckets show how dependable the models are with respect to certain error thresholds (i.e., errors between 10% and 20% are commonly considered a tolerable error range in valuation practices).

EMPIRICAL RESULTS

This section features the empirical results of the analysis. First, model performance in estimating market values is assessed. With respect to the research objective, we furthermore discuss the results from the model agnostic analysis with SHAP and draw conclusions on the features' functional relationships with the dependent variable.

Model Performance

Exhibit 5 depicts the out-of-sample performance metrics of the DNN as well as the OLS respectively. The DNN proves to be highly accurate in estimating market values per square foot, with the MAPE ranging between 9.29% and 10.98% and the corresponding MAE ranging between 7.56 and 25.54 dollars per square foot. The MSE and RMSE show that the apartment, office, and retail models generally produce higher errors that are penalized more than in the industrial model, as market values are generally lower in this sector. Across all property types, over 85% of the market value predictions of the DNN are estimated within a MAPE of 20%, while in the OLS estimation, only around 55% of predictions are falling within this range. In general, the OLS shows a considerably lower model fit compared to the DNN.

EXHIBIT 5
Model Performance Metrics

Method	R^2	MAPE	MPE	MAE	MSE	RMSE	PE10	PE20
Unit	**[%]**	**[%]**	**[%]**	**[$/SqFt]**	**[$/SqFt]**	**[$/SqFt]**	**[%]**	**[%]**
	Panel A: Apartment							
OLS	0.77	0.26	0.04	43.61	7,959.58	89.22	0.31	0.55
ANN	0.97	0.09	-0.03	18.88	1,177.55	34.32	0.65	0.91
	Panel B: Industrial							
OLS	0.73	0.24	0.06	17.53	659.82	25.69	0.30	0.56
ANN	0.95	0.11	0.04	7.56	128.04	11.32	0.62	0.87
	Panel C: Office							
OLS	0.76	0.32	0.07	64.99	9,351.87	96.71	0.26	0.48
ANN	0.96	0.11	-0.03	25.54	1,490.37	38.61	0.58	0.87
	Panel D: Retail							
OLS	0.81	0.30	0.07	62.19	15,125.86	122.99	0.31	0.54
ANN	0.97	0.10	0.03	22.94	2.139.41	46.25	0.67	0.88

Global Model Interpretability

In traditional property valuation, market values of income-generating properties are determined with the income approach which consists of two primary elements, rental income and the capitalization rate. However, alternative methods such as the sales comparison approach and the residual cost approach consider various other factors, including locational, physical, financial, and macroeconomic characteristics (see Pagourtzi et al., 2003) that are not necessarily reflected in the income approach. Our research focuses on a data-driven methodology grounded in economic theory. We use a comprehensive set of physical and structural property attributes, neighborhood characteristics, cash flows, and macroeconomic as well as real estate market indicators to capture all relevant price-determining attributes.

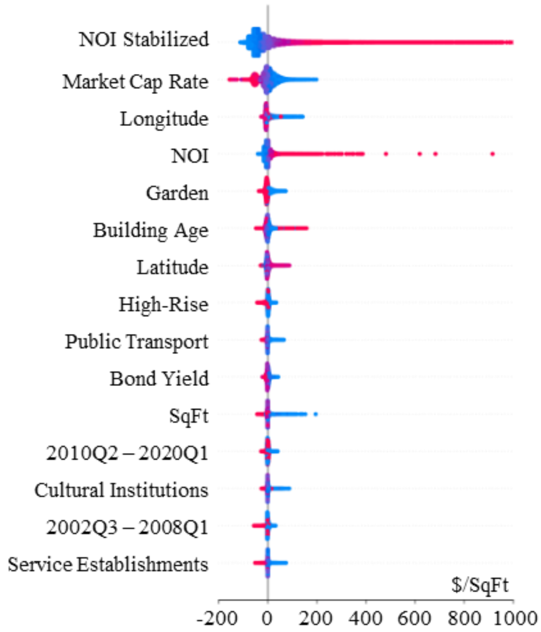
To review the relations of employed features in our models we analyze the features' marginal influences that are presented in **Exhibit 6**. In the respective summary plots, three dimensions can be explored with the features being arranged in a specific order that reflects their relative importance in the model predictions. For all sectors, the stabilized net operating income appears to be the most important feature. The plot also illustrates the characteristics of the features in the second and third dimensions by indicating whether the contribution of a feature to the final prediction is positive or negative and which value the feature takes (i.e., illustrated by color).

We use the SHAP summary plot to identify the key value drivers and relate them to their economic meaning to bridge the gap between economic theory and the data-driven machine learning approach. It is important to note that our models do not incorporate inferential assumptions that can determine causal relationships. That is, the significance of the features is determined solely by the statistical relationships that the model identifies. Ideally, the statistical relationships determined by the model are consistent with economic principles and thus contribute to the understanding of price formation process in commercial property markets. As discussed by Lorenz et al. (2022) a feature importance plot can be utilized to evaluate the relevance of variables for a given predictive task. This method allows insight into the reliability of an algorithmic hedonic model and its ability to capture a plausible understanding of the economic context.

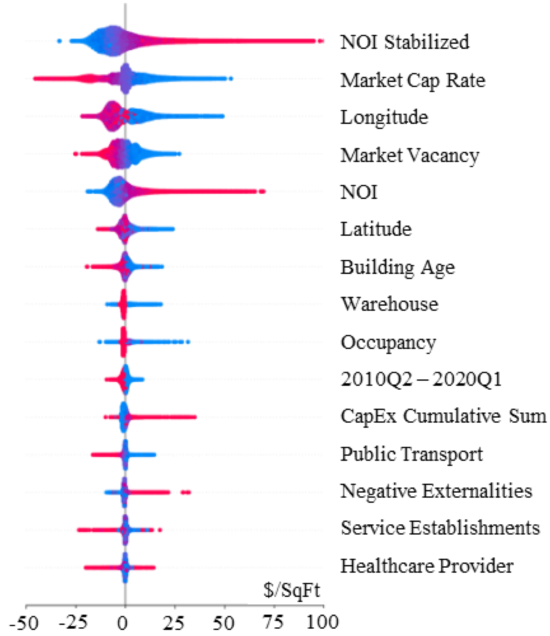
EXHIBIT 6

SHAP Summary Plot (Top 15 Features)

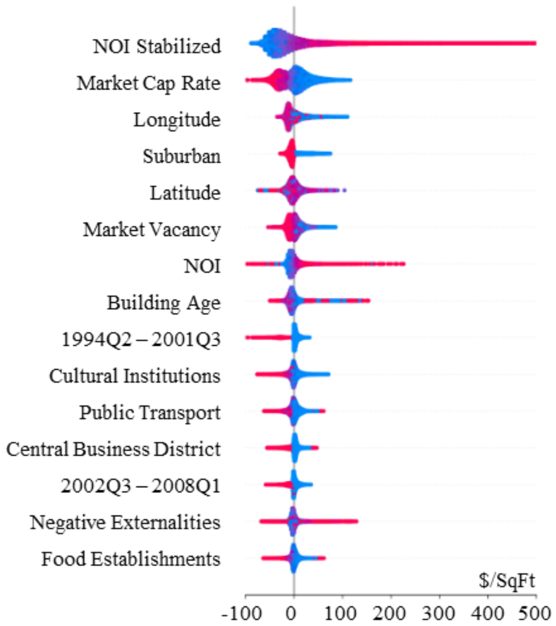
Panel A: Apartment



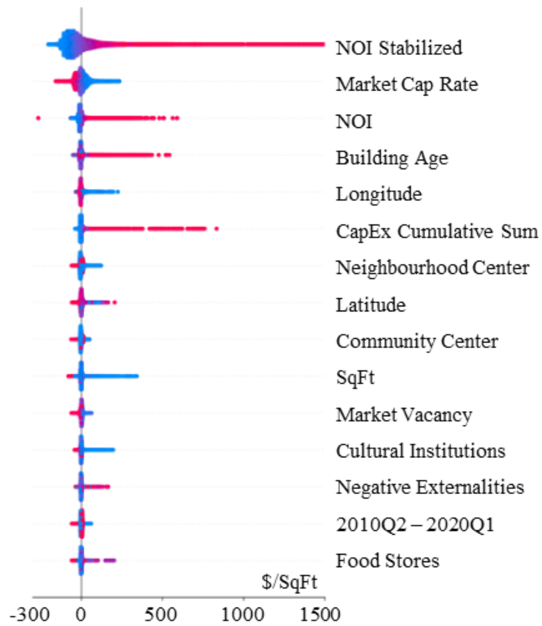
Panel B: Industrial

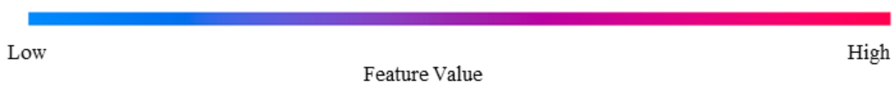


Panel C: Office



Panel D: Retail



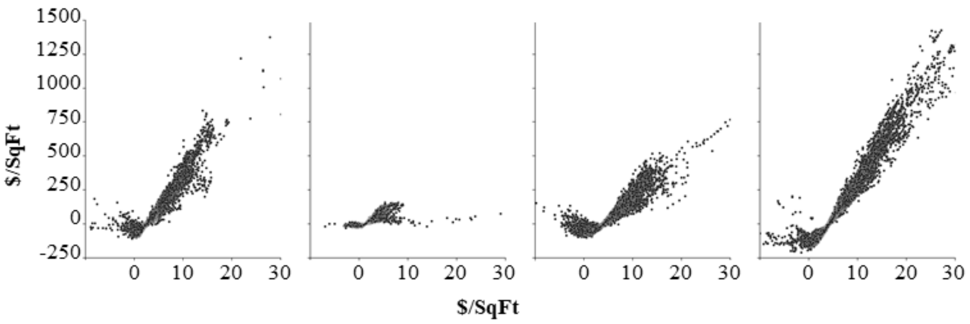


In line with economic theory, **Exhibit 6** depicts the stabilized NOI and the market capitalization rate as the most important feature in the prediction process of the model across all property types. Furthermore, the location expressed by the geo-coordinates, the physical condition proxied with building age, and the current NOI appear to be equally important across all asset sectors and have a strong influence on the model predictions. Moreover, it becomes clear that each property sector has individual value drivers, such as the presence of a garden in the case of apartment properties or the location of an office building in the central business district (CBD). As alluded to previously, SHAP can be used to draw conclusions about the functional relationship between explanatory variables and the dependent variable. This is particularly beneficial in the context of real estate valuation, where understanding of pricing processes is of paramount importance. **Exhibit 7** shows the relationships of four explanatory variables with SHAP partial dependence plots.

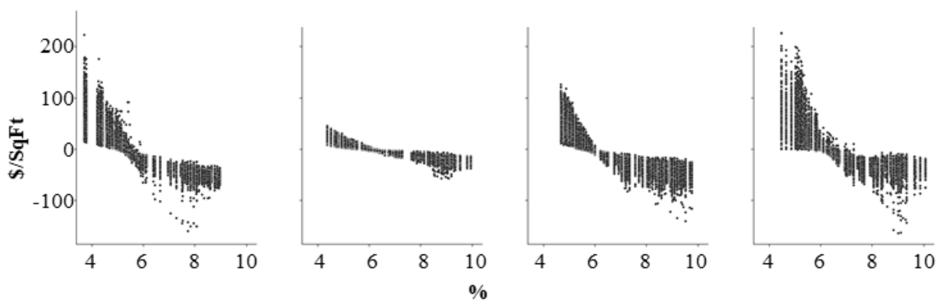
EXHIBIT 7
SHAP Partial Dependence

	Apartment	Industrial	Office	Retail

Panel A: Stabilized NOI



Panel B: Market Cap Rate



Panel C: Building Age

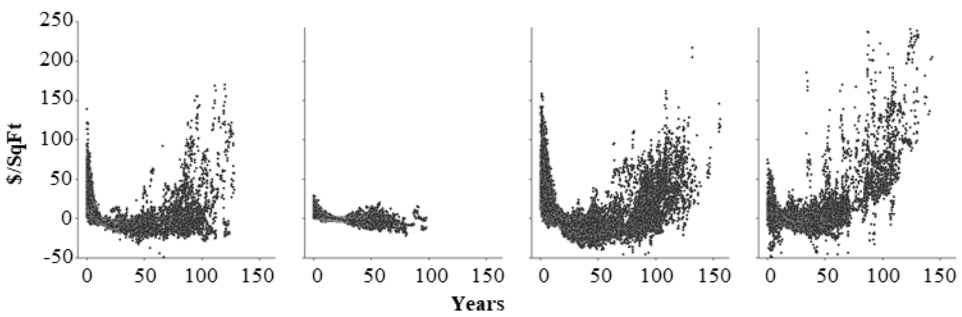


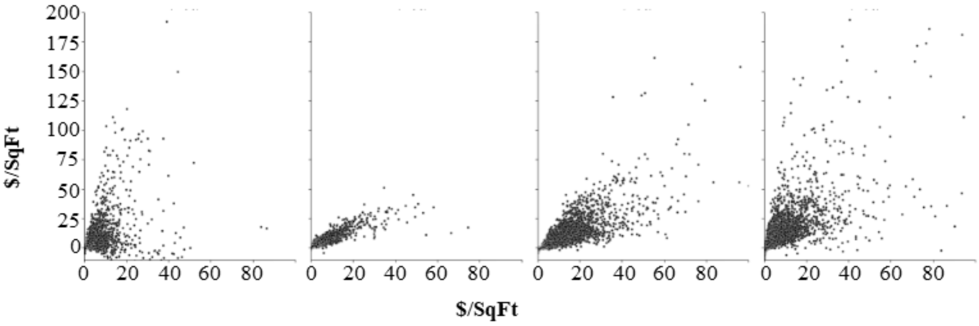
Exhibit 7, Panel A depicts the dependence plots of stabilized NOI and its impact on the prediction of the market value. Across all asset sectors a positive linear relationship for values greater than zero can be observed, as expected market values increase with an increasing stabilized NOI. A negative stabilized NOI shows a non-linear pattern that will be interpreted with further analysis carried out below. The second most important feature in the prediction of market values is the market capitalization rate. **Exhibit 7, Panel B** depicts the relation of this feature to the impact on the market value and it takes the expected relationship in all four property types.

As the capitalization rate is a proxy of risk and return in the real estate market, market values generally decrease with increasing capitalization rates. Notably, the plot for industrial properties deviates from the other property types, but this is due to the mean value of industrial properties in the sample being significantly smaller. With respect to a property's physical condition, we focus on the impact of property age. Lorenz et al. (2022) show that, in line with economic theory, the age of an apartment exhibits a U-shaped pattern, that is both newest and oldest buildings generate highest rents. In **Exhibit 7, Panel C** we observe that this is also the case for the apartment sample alongside with the office and retail properties. Only for industrial properties this U-shape seems to be less pronounced. The plot of industrial properties generally shows a lower building age, which can be attributed to the nature of heavy industry use and the limited usability by third parties.

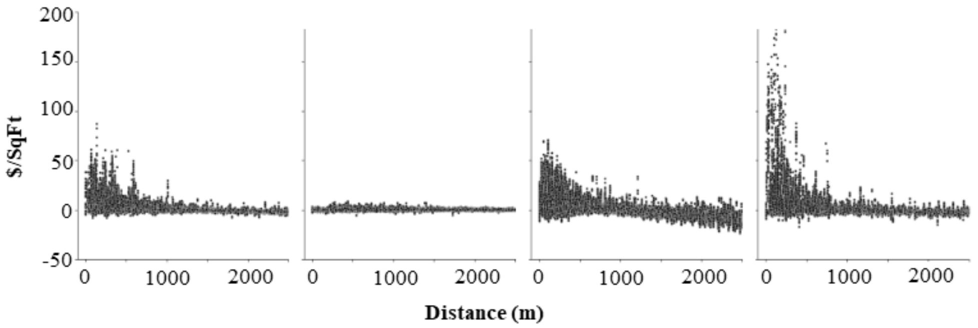
EXHIBIT 8
SHAP Partial Dependence

	Apartment	Industrial	Office	Retail

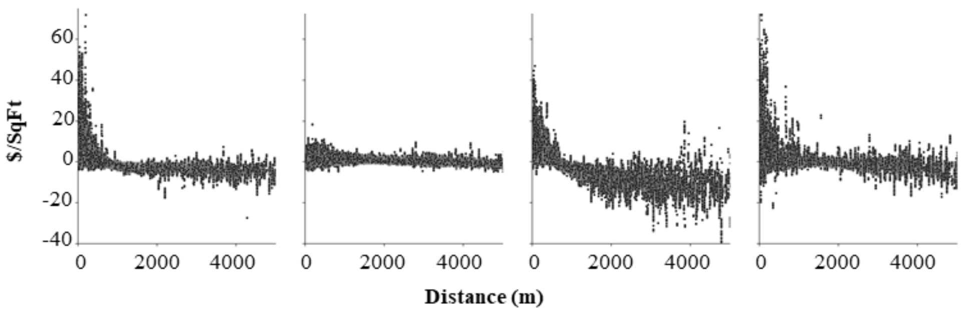
Panel A: Capital Expenditures



Panel B: Cultural Institutions



Panel C: Public Transport



While **Exhibit 7** shows features that have similar impacts across the four property types **Exhibit 8** depicts features that behave differently in their relation to market values across the property types. **Exhibit 8, Panel A** illustrates the relation of Capex and its model's impact on the market value. Generally, Capex lead to an increase in market values whereby the marginal effect varies across property types. A dollar of Capex per square foot appears to have the strongest impact on the market value per square foot for apartment properties whereas industrial properties exhibit the lowest marginal effect. **Exhibit 8, Panel B** depicts the impact of the proximity to a cultural institution (i.e., museum, entertainment facilities or attractions) on the model's prediction of the market value. It is interesting to observe that retail properties in close proximity to cultural institutions experience a higher premium than all other property types. This could be related to increased pedestrian flows generated by cultural institutions which drive market values of retail properties. In contrast, the proximity to cultural institutions seems to pose no effect on the market value of industrial properties. **Exhibit 8, Panel C** shows the impact of a property's proximity to public transport on the market value. Whereas for industrial properties the impact seems considerably low, retail, apartment and office properties show strong relations to this POI. Interestingly, retail and apartment properties experience a positive impact on the market values when in close proximity to public transport but barely see negative impacts for larger distances. However, in the office sector, public transport seems to be of particular interest as bigger distances are related to negative impacts on the predictions. Hence, there seems to be a sweet spot up to which the presence of POIs matters.

Exhibit 7 and Exhibit 8 present multiple instances where a feature can take values that result in both a positive and negative model impact. The factors contributing to such attributions can be examined more closely with the interaction effects for the respective variable. For example, the stabilized NOI in **Exhibit 7, Panel A** shows negative values leading to both positive and negative model impacts. We expect such behavior to be related to structural characteristics of the corresponding properties and thus analyze the interaction effects of the stabilized NOI with both capital expenditures as well as occupancy, illustrated in **Exhibit 9**.

EXHIBIT 9

SHAP Partial Dependence with Interaction Effects (Financial)

Panel A: Stabilized NOI vs. Occupancy	**Panel B: Stabilized NOI vs. Cumulative Capex**

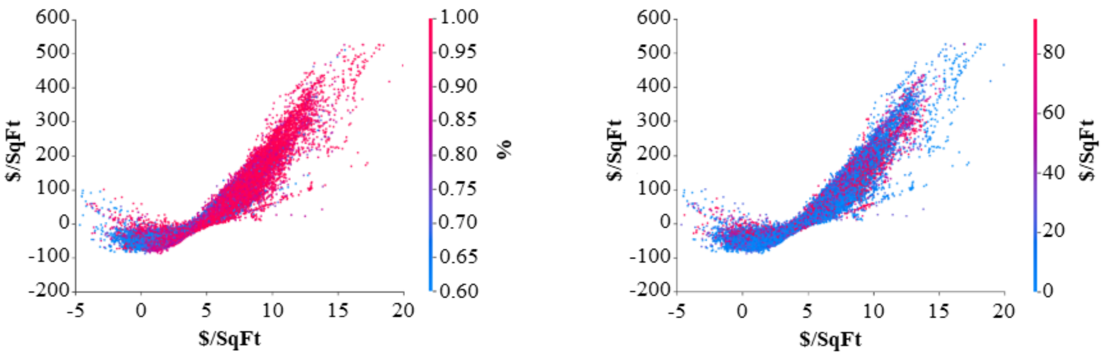
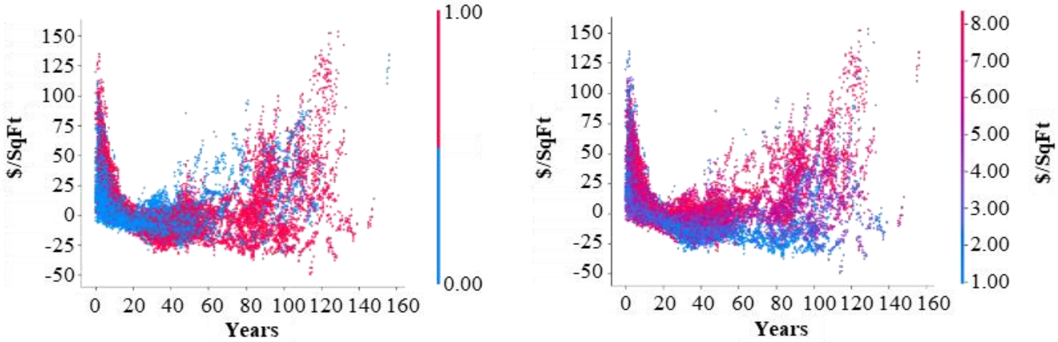


Exhibit 9, Panel A displays the interaction effect between occupancy and stabilized NOI, while **Exhibit 9, Panel B** shows the interaction effect between cumulative Capex and stabilized NOI. The blue color on the graphs indicates low interaction feature values, while red color indicates high interaction feature values. We observe that in cases where negative NOI contributes negatively to the model prediction and thus leads to the expectation of lower market values, both occupancy and capital expenditures tend to be low, indicating high vacancy and potentially lower building quality in comparison to other properties. On the other hand, observations with negative NOI that contribute to the model's prediction positively are characterized by higher occupancy and high capital expenditures that increase a buildings quality and thus the value. In **Exhibit 10**, we analyze the observed U-shaped pattern in the building age by inspecting interaction effects with both location (Panel A) and income (Panel B). In suburban areas, the building age generally shows a negative relation as seen in **Exhibit 10, Panel A.** That is, older properties that are located in suburban areas tend to have lower market values. From **Exhibit 10, Panel B** we can deduct that properties for which high building ages are positively related with market value and high NOIs tend to be clustered in CBDs. Osland (2010) summarizes the main rationale behind early land economic theories and concludes that overcoming space in any form is costly and therefore needs to be economized. Thus, highest centrality in the CBD of a city creates high demand that generally leads to high values.

EXHIBIT 10

SHAP Partial Dependence with Interaction Effects (Structural)

Panel A: Building Age vs. CBD | **Panel B: Building Age vs. Stabilized NOI**



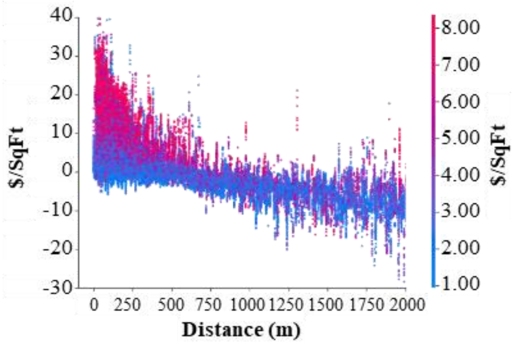
Of course, the centrality of a property cannot only be described by its location in the CBD or a suburban area, but can also be formulated as the sum of multiple characteristics that define the location of a property. Can (1992) mentions neighborhood effects which refer to characteristics that drive demand for real estate in a certain location (i.e., neighborhood) and should materialize in the price function.

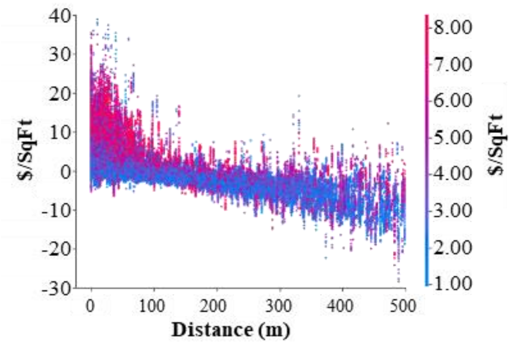
Such trends are not only seen for the market value but generally for the price level when observing the interaction effect of the stabilized NOI and the proximity to public transport or food establishments. This is demonstrated in **Exhibit 11** – the bigger the distances to public transport or food establishments the lower the stabilized NOI that is paid for that property. It is noteworthy that the turning points for the positive effects on the models diverge between the two POIs. **Exhibit 11, Panel A** shows that public transport links located within approximately 750 meters of a property show a positive impact, while food establishments only show positive neighborhood characteristics within a radius of approximately 150 meters as depicted in **Exhibit 11, Panel B.**

EXHIBIT 11

SHAP Partial Dependence with Interaction Effects (POIs)

Panel A: Stabilized NOI vs. Public Transport	**Panel B: Stabilized NOI vs. Food Establishments**





Local Model Interpretability

Shapley values do not only offer the possibility to analyze the relations of variables on a global (i.e., aggregated) level but also provide a local interpretation as Shapley values are calculated for each observation. Thus, the individual contribution of features to the predicted market value can be explained on the property-level. From the SHAP summary plots and partial dependence plots each colored dot represents a single observation that can be explained on its own. SHAP force plots can visualize each single observation to make it comprehensible on the smallest level possible. The contributions of all features are shown as the difference between the actual prediction and the mean prediction (base value) for that very observation. This breaks the expected market value down to the contributing characteristics. It is important to note that feature effects can behave differently for different observations.

EXHIBIT 12
SHAP Force Plot

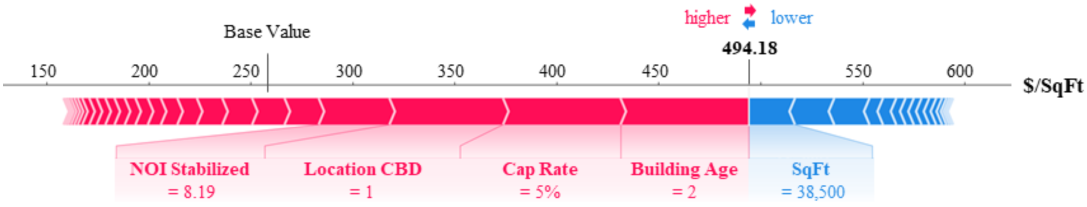


Exhibit 12 shows the composition of a market value prediction for an office property located in Boston, Massachusetts. The prediction for this office property's market value is estimated to be 494.18 $/SqFt. The mean prediction (base value) of office market values in the sample is 258.53 $/SqFt and can be thought of as the "best guess" to predict the market value without knowing anything about the specific property. The features that mainly drive the prediction from the base value of 258.53 $/SqFt to the estimated 494.18 $/SqFt are the stabilized NOI, location, market cap rate, and the building age. In this example the square footage of the property reduces the prediction as it contributes negatively. The property is newly built (building age = 2 years), located in the CBD and has a stabilized NOI of 5.23 $/SqFt, well above the sample average of 2.50 $/SqFt, thus increasing the prediction relative to the base value. The positive contribution of the stabilized NOI to the prediction increases the base value by 149.58 $/SqFt. Additionally, the building age of the property contributes 46.85 $/SqFt, while its CBD location contributes 38.65 $/SqFt, and the market value cap rate of 5% in the quarter of observation contributes 54.99 $/SqFt. In total, these four features result in a contribution of 290.07 $/SqFt to the base value of 258.53 $/SqFt, leading to a predicted value of 548.59 $/SqFt. However, this is not the predicted 494.18 $/SqFt as so far, the negative contributions have been left aside. As highlighted in blue color, the size of the property has a negative impact and pushes against the other features, thus reducing the final prediction. The size of the property, with 38,500 square foot is smaller than the sample

average of 277,124 square foot resulting in a negative impact 30.63 $/SqFt. In sum all other features add up to a negative 23.78 $/SqFt and lead to an expected market value of 494.18 $/SqFt.

CONCLUSION

The objective of this study was to introduce an effective and comprehensive framework for the practical utilization of ML-based automated valuation models (AMVs) in the domain of commercial real estate that seeks to strike a balance between accuracy and interpretability of the estimation method without compromising neither one. To illustrate this, we trained a deep neural network (DNN) using a unique sample of more than 400,000 property-quarter observations from the NCREIF Property Index (NPI) and applied model-agnostic "Shapley Additive exPlanations" (SHAP) to shed light on the algorithm's prediction rules. The proposed method enabled ex-post interpretability of the models' prediction rules and could disentangle value drivers on an aggregated global level as well as on a disaggregated local level, that is for each property individually.

The proposed methodological framework demonstrates high accuracy in the estimation of commercial real estate market values across all four asset sectors. SHAP demonstrates that the inner workings of data-driven techniques are generally consistent with economic theory and follow predominantly the traditional income approach by using the net operating income and market capitalization rates as the key explanatory features. Moreover, the location expressed by the geo-coordinates and the distance to points of interest as well as the properties' physical condition proxied with building age showed a strong influence on the models' predictions. Deviations in the feature importance across property types were observed predominantly in sector specific characteristics. Furthermore, non-linear and three-dimensional relationships between market values and features were revealed and confirm previous findings in the literature. For instance, it could be shown that the relation between market value and building age follows a U-shaped function, which can be explained by the bid-rent curve, as older buildings tend to be concentrated in city centers and CBDs, as well as a sample selection bias as good-quality buildings prevail while outdated or stranded assets leave the market to make room for new developments. On the local level of interpretation, SHAP furthermore showed that the effect of individual features can differ significantly across properties due to non-stationarity across space and time. This is perhaps one of the main advantages of machine learning techniques compared to linear hedonic models, as the latter reduce feature effects to a single, fixed beta coefficient that does not allow to differentiate complex interactions between regressors.

In summary, our study demonstrates that machine learning algorithms can obtain both the estimation accuracy and interpretability, while following and economic logic and being consistent to the existing understanding of pricing processes in the literature. The use of these techniques can moreover add to the existing knowledge by providing a deeper and more nuanced understanding of pricing processes in institutional investment markets.

That said, the machine learning methods also come with certain limitations that should be considered carefully before their use. Despite their powerful applications, these methods are not a panacea that can solve all real-world problems. However, if applied prudently, they could provide an answer to several problems and may become an indispensable tool for many tasks. With immense amounts of data being recorded every day and the development of quantum computing, machine learning applications are about to experience a steep improvement in terms of scale and efficiency. However, with these advances taking at least another five to ten years to take hold, the application of interpretable AVMs in the commercial real estate sector is a milestone on a path yet to be travelled. By pointing to the caveats and illustrating the potential of these methods, our contribution represents a further step along this path and will hopefully motivate further research in this field.

REFERENCES

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. https://doi.org/10.1016/j.eswa.2011.08.077.

Baldominos, A., Blanco, I., Moreno, A., Iturrarte, R., Bernárdez, Ó., & Afonso, C. (2018). Identifying real estate opportunities using machine learning. *Applied Sciences*, *8*(11), 2321. https://doi.org/10.3390/app8112321.

Calainho, F. D., van de Minne, A., & Francke, M. K. (2022). A machine learning approach to price indices: Applications in commercial real estate. *The Journal of Real Estate Finance and Economics,* Forthcoming. https://doi.org/10.1007/s11146-022-09893-1.

Can, A. (1992). Specification and estimation of hedonic housing price models. *Regional Science and Urban Economics*, *22*(3), 453–474. https://doi.org/10.1016/0166-0462(92)90039-4.

Cannon, S. E., & Cole, R. A. (2011). How accurate are commercial real estate appraisals? Evidence from 25 years of NCREIF sales data. *The Journal of Portfolio Management*, *35*(5), 68-88. https://doi.org/10.3905/jpm.2011.37.5.068.

Deppner, J., von Ahlefeldt-Dehn, B., Beracha, E., & Schaefers, W. (2023). Boosting the accuracy of commercial real estate appraisals – An interpretable machine learning approach. *The Journal of Real Estate Finance and Economics,* Forthcoming. https://doi.org/10.1007/s11146-023-09944-1.

Din, A., Hoesli, M., & Bender, A. (2001). Environmental variables and real estate prices. *Urban Studies*, *38*(11), 1989–2000. https://doi.org/10.1080/00420980120080899.

Do, A. Q., & Grudnitski, G. (1993). A neural network analysis of the effect of age on housing values. *Journal of Real Estate Research*, *8*(2), 253–64. https://doi.org/10.1080/10835547.1993.12090712.

Dunse, N., & Jones, C. (1998). A hedonic price model of office rents. *Journal of Property Valuation and Investment*, *16*(3), 297–312. https://doi.org/10.1108/14635789810221760.

Goodman, A. C., & Thibodeau, T. G. (1995). Age-related heteroskedasticity in hedonic house price equations. *Journal of Housing Research*, *6*(1), 25–42. http://www.jstor.org/stable/24825889.

Grether, D. M. & Mieszkowski, P. (1974). Determinants of real values. *Journal of Urban Economics*, *1*(2), 127–145. https://doi.org/10.1016/0094-1190(74)90013-8.

Hu, L., He, S., Han, Z., Xiao, H., Su, S., Weng, M., & Cai, Z. (2019). Monitoring housing rental prices based on social media: An integrated approach of machine-learning algorithms and hedonic modeling to inform equitable housing policies, *Land Use Policy*, *82*, 657–673. https://doi.org/10.1016/j.landusepol.2018.12.030.

Kok, N., Koponen, E.-L., & Martínez-Barbosa, C. A. (2017). Big data in real estate? From manual appraisal to automated valuation. *The Journal of Portfolio Management*, *43*(6), 202–211. https://doi.org/10.3905/jpm.2017.43.6.202.

Levantesi, S., & Piscopo, G. (2020). The importance of economic variables on London real estate market: A random forest approach. *Risks*, *8*(4), 1–17. doi:10.3390/risks8040112.

Lockwood, L. J., & Rutherford, R. C. (1996). Determinants of industrial property value. *Real Estate Economics*, *24*(2), 257–272. https://doi.org/10.1111/1540-6229.00690.

Lorenz, F., Willwersch, J., Cajias, M., & Fuerst, F. (2022). Interpretable machine learning for real estate market analysis. *Journal of Real Estate Economics*, Forthcoming. https://doi.org/10.1111/1540-6229.12397.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *31st Conference on Neural Information Processing Systems (NIPS)*.

Mayer, M., Bourassa, S. C., Hoesli, M., & Scognamiglio, D. (2019). Estimation and updating methods for hedonic valuation. *Journal of European Real Estate Research*, *12*(1), 134–150. https://doi.org/10.1108/JERER-08-2018-0035.

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. https://doi.org/10.1080/09599916.2013.781204.

Mills, E. S. (1992). Office rent determinants in the Chicago area. *Journal of Real Estate Economics*, *20*(2), 273–287. https://doi.org/10.1111/1540-6229.00584.

Molnar, C. (2020). *Interpretable Machine Learning. A Guide for Making Black Box Models Explainable*. Leanpub.

Osland, L. (2010). An application of spatial econometrics in relation to hedonic house price modeling. *The Journal of Real Estate Research*, *32*(3), 289–320. https://doi.org/10.1080/10835547.2010.12091282.

Pace, R. K., & Hayunga, D. (2020). Examining the information content of residuals from hedonic and spatial models using trees and forests. *The Journal of Real Estate Finance and Economics*, *60*(1-2), 170–180. https://doi.org/10.1007/s11146-019-09724-w.

Pagourtzi, E., Assimakopoulos, V., Hatzichristos, T., & French, N. (2003). Real estate appraisal: Review of valuation methods. *Journal of Property Investment & Finance*, *21*(4), 383–401. https://doi.org/10.1108/14635780310483656.

Pérez-Rave, J., Correa-Morales, J., & González-Echavarría, F. (2019). A machine learning approach to big data regression analysis of real estate prices for inferential and predictive purposes. *Journal of Property Research*, *36*, 59–96. https://doi.org/10.1080/09599916.2019.1587489.

Peterson, S., & Flanagan, A. (2009). Neural network hedonic pricing models in mass real estate appraisal. *The Journal of Real Estate Research*, *31*(2), 147–164. https://doi.org/10.1080/10835547.2009.12091245.

Potrawa, T., & Tetereva, A. (2022). How much is the view from the window worth? Machine learning-driven hedonic pricing model of the real estate market. *Journal of Business Research*, *144*, 50–65. https://doi.org/10.1016/j.jbusres.2022.01.027.

Quan, D. C., & Quigley, J. M. (1991). Price formation and the appraisal function in real estate markets. *The Journal of Real Estate Finance and Economics*, *4*, 127–146. https://doi.org/10.1007/BF00173120.

Rico-Juan, J. R., & La Taltavull de Paz, P. (2021). Machine learning with explainability or spatial hedonic tools? An analysis of the asking prices in the housing market in Alicante, Spain. *Expert Systems with Applications*, *171*. https://doi.org/10.1016/j.eswa.2021.114590.

Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy*, *82*(1), 34–55. http://dx.doi.org/10.1086/260169.

Shapley, L. S. (1953). A value for n-person games, in Kuhn, H. and Tucker, A. (Eds.), *Contributions to the theory of games, Vol. II*, 307–317, Princeton University Press.

Sirmans, C. F., & Guidry, K. A. (1993). The determinants of shopping centre rents. *Journal of Real Estate Research*, *8*(1), 107–15. http://www.jstor.org/stable/44095414.

Valier, A. (2020). Who performs better? AVMs vs hedonic models. *Journal of Property Investment & Finance*, *38*(3), 213–225. https://doi.org/10.1108/JPIF-12-2019-0157.

Worzala, E., Lenk, M., & Silva, A. (1995). An exploration of neural networks and its application to real estate valuation. *The Journal of Real Estate Research*, *10*(2), 185–201. https://doi.org/10.1080/10835547.1995.12090782.

Yoo, S., Im, J., & Wagner J. E. (2012). Variable selection for hedonic model using machine learning approaches: A case study in Onondaga County, NY. *Landscape and Urban Planning*, *107*(3), 293–306. https://doi.org/10.1016/j.landurbplan.2012.06.009.

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*, 349–388. https://doi.org/10.1080/10835547.2011.12091311.