Big Data in Real Estate? *From Manual Appraisal to Automated Valuation*

Nils Kok, Eija-Leena Koponen, and Carmen Adriana Martínez-Barbosa

Nils Kok
is an associate professor in finance and real estate at Maastricht University and the chief economist at GeoPhy in New York.
n.kok@maastrichtuniversity.nl

Eija-Leena Koponen
is a senior data scientist at GeoPhy in Delft, the Netherlands.
e.koponen@geophy.com

Carmen Adriana Martínez-Barbosa
is a data scientist at GeoPhy in Delft, the Netherlands.
c.barbosa@geophy.com

Real estate is the largest asset class in the world. It received its own global industry classification in August 2016 and makes up, on average, 5.1% of any institutional portfolio (Andonov, Eichholtz, and Kok [2013]). Determining the value of commercial real estate remains elusively hard, however, with a workforce of 74,000 appraisers in the United States alone still manually assessing the value of assets that are sometimes worth billions of dollars.[1]

Appraisals are typically based on the capitalization of the net income of an asset using a yield (or cap rate) that has been inferred from neighboring transactions, or *comps*. However, transactions of comparable properties are never truly comparable in either time or building characteristics, and appraisers are typically anchored on previous valuations or the previous transaction price of a building. The result is that property appraisals typically lag the market and provide smoothed approximations of true market prices, with values that are artificially low in bull markets and high in bear markets (Cannon and Cole [2011]).

These imprecise measures of value are obviously problematic in today's otherwise hyper-efficient capital market: Appraisal-based price indexes, still used by many investors as the benchmark of choice, are incorrect as a result; lenders and investors do not have access to accurate pricing data to underwrite assets; and assessing the risk of real estate assets on the balance sheets of banks and investors is challenging, which was especially problematic during the most recent downturn in 2007. In addition, the real estate sector faces significant cost from the appraisal bureaucracy—the U.S. appraisal market alone has an annual turnover of some USD 90 billion, which is indirectly paid by the ultimate owners of real estate (i.e., public and private pension plans, insurance companies, banks, and their respective constituents and clients).

The precision of appraisals has been a topic of both popular debate and academic study for multiple decades. Most recently, Cannon and Cole [2011] analyzed the accuracy of appraisals for the U.S. commercial real estate sector, using 1984–2010 data from the National Council of Real Estate Investment Fiduciaries (NCREIF). Comparing property appraisals with actual transactions, the authors documented that, on average, appraisals are more than 12% above or below the subsequent transaction price (correcting for the time lag between property transactions and valuations). These results are consistent with the findings of Fisher, Miles, and Webb [1999] for the 1978–1998 period. They documented an average absolute deviation of 9% to 12.5% between appraisals and transaction prices. Outside of the United States, MSCI has published an annual research report for more than a decade, comparing valuations and sale prices (before 2016, this was under the IPD umbrella, the commercial equivalent of NCREIF). The MSCI June 2016 report shows significant variation across countries and years, but the absolute difference between property appraisals and actual transactions ranges between 7.7% (Italy) and 13.9% (Japan).

To circumvent the error inherent in appraisals, the commercial real estate industry has developed alternative property index series, such as the transaction-based indexes marketed by Real Capital Analytics (RCA), as well as sophisticated unsmoothing techniques to deal with otherwise lagged and artificially smooth performance benchmarks (Marcato and Key [2007]). These measures do not resolve the fundamental problem of imprecise measurement of value for other parts of the industry, which may not be interested in index construction, per se, but rather in the precise valuation of an individual real estate asset for investment or financing purposes. "Better" appraisals and indexes also do not address the cost inefficiency stemming from traditional, manual appraisals.

While the imprecision and inefficiencies of property appraisals linger, the real estate industry has witnessed a significant increase in the availability of data and the advent of machine learning techniques, which are now widely used in medical research and in applications such as search algorithms and recommendation engines (examples of machine learning applications vary widely, from fraud detection by PayPal to the personalized online advertising that we have grown used to). In the single-family housing market, some nascent products are using the combination of data abundance and machine learning modeling, including, for example, House Canary's automated valuation model and Zillow's much-discussed "Zestimate." In this study, we provide a practical application of *big data* in combination with sophisticated modeling techniques to develop an automated, machine-based valuation model for the commercial real estate sector. We focus on the multifamily sector, enabled by access to a dataset of some 54,000 U.S. multifamily assets. This dataset is enriched by a wide set of both standard demographic and economic measures and more modern, hyperlocal metrics, such as proximity to music events, bars and restaurants, and green space and local crime incidence. Rather than traditional hedonic models, which are limited both statistically and by a researcher's predisposition toward "standard" explanatory variables, we then apply assisted machine learning models that rely on (stochastic) decision trees. These models can sift through millions of combinations of thousands of variables, training and testing the model on randomly selected parts of the datasets, leading to precise out-of-sample tests of predictive performance.

We find strong evidence on the superiority of automated valuation models (AVMs) over traditional appraisals—the median absolute error of the automated model we develop is 9%, which compares favorably against the accuracy of traditional appraisals, and the model can produce an instant value at every moment in time at a very low cost. We also provide evidence on the importance of using hyperlocal information on the location of an asset. Although the use of economic and demographic data at the census tract and ZIP code level are standard practice in real estate modeling, new information layers gleaned from a wide variety of sources, including social media data, police records, and amenities related to economic vibrancy, add significant value to pricing models.

This study is important for multiple reasons. First, precise, timely estimations of property values are critical for real estate investors and lenders to make informed underwriting decisions, where systematic errors or biases in valuations may have adverse effects on the provision of equity or debt. Second, as Cannon and Cole [2011] noted, "investors, regulators, and others rely upon appraised values to assess returns on the USD 11 trillion U.S. commercial real estate market," and automated property valuations can provide a more accurate reflection of both the real estate stock (i.e., the value of real estate on the balance sheet) and flow (i.e., real estate returns from changes in capital values). Third, AVMs can be used for stress testing under adverse economic scenarios, which remains a much-needed tool for regulators, banks, rating agencies, and investors. Fourth, the availability of an instant, accurate property value may spur financial innovation in the real estate sector, such as automated loan origination by banks, defined contribution products for private real estate investments, and arbitrage products for commercial real estate (comparable to emerging products for the single-family sector, such as Opendoor.com).

METHODOLOGY

An AVM is a service that uses mathematical models to provide the estimated value of a property at a specific point in time. AVMs can take different approaches to compute a property's value. Importantly, the model does not depend on the use of a capitalization rate (the *cap rate*, in jargon terms), which is critical in traditional property valuation techniques. Such a cap rate is simply derived from the net operating income (NOI) of an asset divided by its transaction price. For valuation purposes, the cap rates of three to five recently transacted, nearby buildings are typically used, adjusting for differences between the appraised property and the transacted properties. In contrast, an AVM incorporates all transactions in a given market, assuming a relationship between value and independent variables that is consistent across locations. Most AVMs today use regression (hedonic) models (e.g., O'Neill [2004]; Schulz, Wersing, and Werwatz [2014]); however, some more modern algorithms use adaptive estimation models or neural, self-learning network models (Crosby et al. [2016]).

Hedonic Approach

Hedonic AVMs are well recognized in the real estate sector. Following Rosen [1974], hedonic pricing models require information on basic property attributes, such as construction year, size, and the extent of capital improvements, as well as locational characteristics. These models typically include a search engine that compares the attributes of the subject property with comparable properties using a radius search pattern or other logical search parameters over a predetermined period (see Eichholtz, Kok, and Quigley [2010] for an application). Hedonic AVMs are common because they are based on simple regression models that are easy both to implement and understand. The main disadvantage is that regression models are *global* models, which means that they generate a single predictive formula that is constant across the entire range of variables (e.g., property attributes, amenities, and contextual data.). However, many variables have a nonlinear relationship relative to the predicted value. Consequently, one single model might not be successful at predicting property value in the most accurate way.

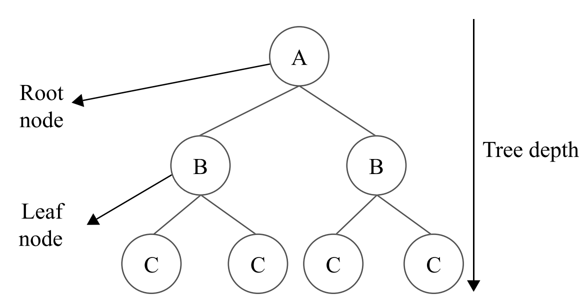
Machine Learning Approach

Instead of traditional hedonic models, we can also apply decision tree models. Such models sequentially divide the dataset into subsets in order to apply a regression model to each subset. There are two types of decision trees, depending on the type of target variable: classification trees, which are aimed at predicting categorical variables, and regression trees, which predict continuous variables (Breiman et al. [1984]). Because the value of an asset is continuous, we use regression trees for the AVMs developed in this article.

A regression tree algorithm finds the best predictors from the set of independent variables by first minimizing the variance of a regression between each combination of the dependent and an independent variable. This yields the order of importance of the variables. Each of the explanatory variables then represents a node in the decision tree. The first node, also called the *root node*, contains the entire dataset (or at least the part of the dataset that is used for training the model). The split point within the root node is based on the best predictor and is determined by applying a cost function that minimizes the variance in a regression line for both subsamples. The order of importance is calculated again for the set of independent variables, based on the sample in the *leaf node* (the second layer of nodes in the tree). The best-performing independent variable is selected, and the split is repeated. This procedure is recursive in all branches of a tree and is repeated until a limit is reached (i.e., the desired depth of the tree). Exhibit 1 shows a stylized example of the structure of a decision tree.

More generally expressed, for each explanatory variable (or predictor), the data are split at one of the nodes of a decision tree, where the node's level in the tree is based on the strength of its relation with the dependent variable. For a split point m, representing a subsample R_m with N_m observations (values of the target variable), the node is selected under the following condition:

EXHIBIT 1
Machine Learning Models—Example of Decision Tree with Four Nodes



$$\text{Node} = \min(\text{Var}_1, \text{Var}_2, \ldots, \text{Var}_m) \quad (1)$$

where Var_m is the variance (or standard deviation) of the explanatory variable in subsample R_m.

Types of Models

Decision trees have several advantages over hedonic regression models and other machine learning techniques. For instance, decision trees are simple to understand and to interpret, and statistical significance can easily be calculated. Decision trees are also able to handle categorical variables, so there is no need to (manually) create dummy variables. Decision trees can make predictions in a very short computing time, even with large amounts of data.

Despite these advantages, decision trees have some limitations. First, they can have unlimited vertical growth, until the nodes have a sample of just one observation. These deep trees produce very complex models that take all the noise from the training data. This effect, which is called *overfitting*, generates models that yield poor predictions for unseen data. Second, deep decision trees are nonrobust in the sense that a small change in the training data can result in a big change in the tree and thus a big change in the final prediction. If decision trees are too shallow, on the other hand, they produce neither good models nor good predictions. This effect is called *underfitting*.

To overcome the limitations of decision tree models, one can create multiple decision trees and then obtain the average prediction. This is called an *ensemble of trees*, and it is a powerful method to obtain better predictive performance as compared to what would be obtained from any of the constituent decision trees alone. One of the most popular ensembles of trees is the *random forest model*, which consists of bootstrapped samples of the original training dataset. Using the bootstrapped sample, a decision tree is fully developed such that each of its nodes is calculated by selecting random explanatory variables. The whole process, each time starting with a new bootstrapped sample, is repeated many times. This process ensures that each decision tree generated will be different and uncorrelated with the others. The final prediction of the target variable is a weighted *vote*, or the average of the predictions of all trees in the collection. Random forest models are quite good at reducing variance; therefore, they are suitable to reduce the overfitting generated by single, deep decision trees.

Another well-known ensemble method is *stochastic boosting*. In gradient boosting models, many small decision trees are built sequentially from residual-like measures from the previous tree. At each iteration, a tree is built from a random subsample of the dataset (selected without replacement), producing an incremental improvement of the model. Ultimately, all of the small trees are stacked together as a weighted sum of terms. The overall model accuracy progressively improves with each additional term. Because of their structure, boosting models are aimed at decreasing the selection bias in the data. For the AVM modeling in this article, we use both the random forest model and two gradient boosting methods: gradient boosted trees and XGBoost (eXtreme Gradient Boosting is an advanced implementation of the gradient boosting algorithm).

Model Performance

Data science offers several key performance indicators (KPIs) that allow for precise measurement of the performance of a model. Comparable to financial decision-making, one KPI does not necessarily tell the truth, and it is therefore desirable to exploit multiple indicators. For the models developed in this article, we deploy two measurements: robustness, measured by explained variance or R^2, and accuracy, measured by the mean squared error (MSE) or the mean or median absolute percentage error (MAPE and MdAPE, respectively).

$$R^2 = 1 - \frac{\text{Residual sum of squares}}{\text{Total sum of squares}} = \frac{\text{SSE}}{\text{TSS}} = \frac{n * \text{MSE}}{\text{TSEE}} \quad (2)$$

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (3)$$

$$\text{MAPE} = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \quad (4)$$

The optimal model maximizes explained variance while optimizing the accuracy. Obviously, there is a trade-off between these two KPIs—a robust model is not necessarily accurate, and vice versa.

Model Testing

It is important to estimate the parameters that reduce the error in the model prediction. This can be achieved through model testing, which is often referred to as *cross-validation*. Cross-validation is a technique that is used to evaluate model performance on unseen data. In general, a dataset is divided into train and test datasets. The train dataset is used to build the prediction model, and the test dataset is used to evaluate the model. The goal of cross-validation is to define a dataset to test the model during the training phase to limit either underfitting or overfitting. We calibrate the model based on 70% of the data (randomly selected) and use the remaining 30% as the test set.

DATA

For any valuation of a commercial asset, manual or automated, information on both NOI and actual property transactions represents critical input. We concentrate on multifamily assets to develop the automated valuation model, focusing on the states of California, Florida, and Texas. The decision to focus on multifamily assets in these states is based on data availability rather than any preference or prior. We collect information on NOI and basic property characteristics from a large multifamily lender for the 2011–2016 period. We expand the initial dataset of 12,958 multifamily assets by collecting information on 4,154 multifamily assets from commercial mortgage-backed securities (CMBS) pools, which are available in the public domain. Additional data on 26,922 multifamily assets are sourced from RCA, a leading global provider of property transaction data. In total, these three datasets contain 36,586 unique properties and 7,448 duplicates (these duplicates follow from the fact that RCA also records sales transactions on properties financed by the multifamily lender that supplied the NOI data, and some assets change ownership during the sample period). It is important to note that building data are not equally available across datasets and locations (see Exhibit 2 for a review). Most importantly, the RCA database includes information on NOI for just 27.7% of the transactions. We therefore develop three separate AVMs: one model including NOI, one model excluding NOI, and one model including modeled NOI (using an NOI model).

For the NOI model, we use the dataset that includes the combined information from the large multifamily lender, CMBS data, and the RCA dataset on multifamily assets. This dataset has a total of 84,305 observations in 36,586 unique properties, with 26,354 missing NOI values or NOI values measured before the year 2011, which are removed. During the data preparation, we also remove outliers based on the construction year and the target variable (NOI) by using a cutoff of 1.5 or 5 times the interquartile range (IQR).[2] We also remove observations in the dataset that have a year of construction that is later than 2017 (in the future) or date from the 19th century. We obtain a final dataset of 57,127 observations for 28,145 unique properties.

The dataset to model asset value comprises just 5,164 transactions between the years 2011 and 2016, which limits the performance of the model. Outlier analysis and removal, similar to the NOI model, is applied to the dataset: The year built in the future is removed, as are obvious data errors, reducing the dataset by 40 observations. We also exclude properties that have over 524 units (1.5 times the IQR), around 100 properties total. This leaves a dataset of 5,018 unique transactions between 2011 and 2016.

We enhance the dataset with detailed information on property characteristics (*attributes*), such as the number of units, unit size, presence of amenities, and so on. Property characteristics are obtained from property tax data, which are available through the county assessor's website as a download, or an online form that uses address or parcel number as input. Information on tax-relevant property characteristics is available either at the unit level or the parcel level. For the former, data are summed across all units on a parcel.

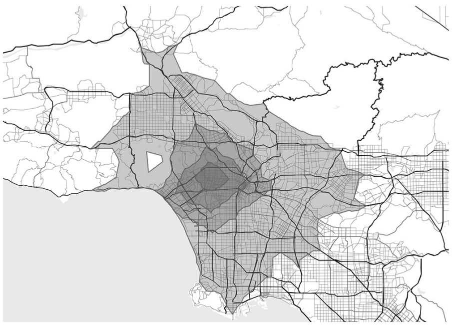
Although the set of property characteristics is quite exhaustive, we dedicate significant resources to obtain detailed locational data, following the real estate mantra of "location, location, location." To be able to document

EXHIBIT 2
Data Sources

Source	Type	Total (unique)	CA	TX	FL
Multifamily Lender	NOI	29,870 (12,958)	21,253 (9,400)	6,199 (2,792)	2,519 (1,099)
CMBS	NOI	11,804 (4,154)	4,737 (1,682)	4,526 (1,614)	2,541 (907)
RCA	NOI/ Transaction price	42,631 (26,922)	24,425 (10,862)	9,696 (3,496)	8,510 (3,935)

EXHIBIT 3
Data Enrichment Process

Collection of Catchment Data



all the different types of locational amenities that are relevant for property value, we first calculate *catchment areas* for each property, which are based on real driving times using the surrounding road network. A catchment area is a polygon that represents how far a person can drive within a certain amount of time. The catchment area is then intersected with three main data categories: amenities (e.g., restaurants, bars, green space, music events, schools), market data (e.g., construction permits, vacancy rates, crime data, local mortgage delinquency rate), and census data (e.g., population, income, employment). In addition, we include U.S.-wide macroeconomic data (e.g., the S&P 500, 30-year mortgage interest rate). See the Online Appendix for the complete list of variables. Each of the catchments is calculated for 3, 5, 10, 15, and 30 minutes' driving time, which, when multiplied by the wide range of independent variables, leads to a dataset that includes thousands of variables. Exhibit 3 illustrates the catchment area concept for a property in Los Angeles, based on different driving times.

EMPIRICAL RESULTS

We develop separate models for property value and NOI. For the NOI model, we estimate both a Los Angeles (LA) model and a three-state model, whereas we estimate only a three-state model for the value model (the limited number of transactions simply does not allow for more granular analysis). Before and during the modeling process, some variable transformations are conducted. For example, NOI and property value are scaled by the number of units in the property to properly compare smaller- and larger-unit properties and to reduce the variance in the distributions of NOI and value.

EXHIBIT 4
Model Performance—OLS versus Machine Learning Models

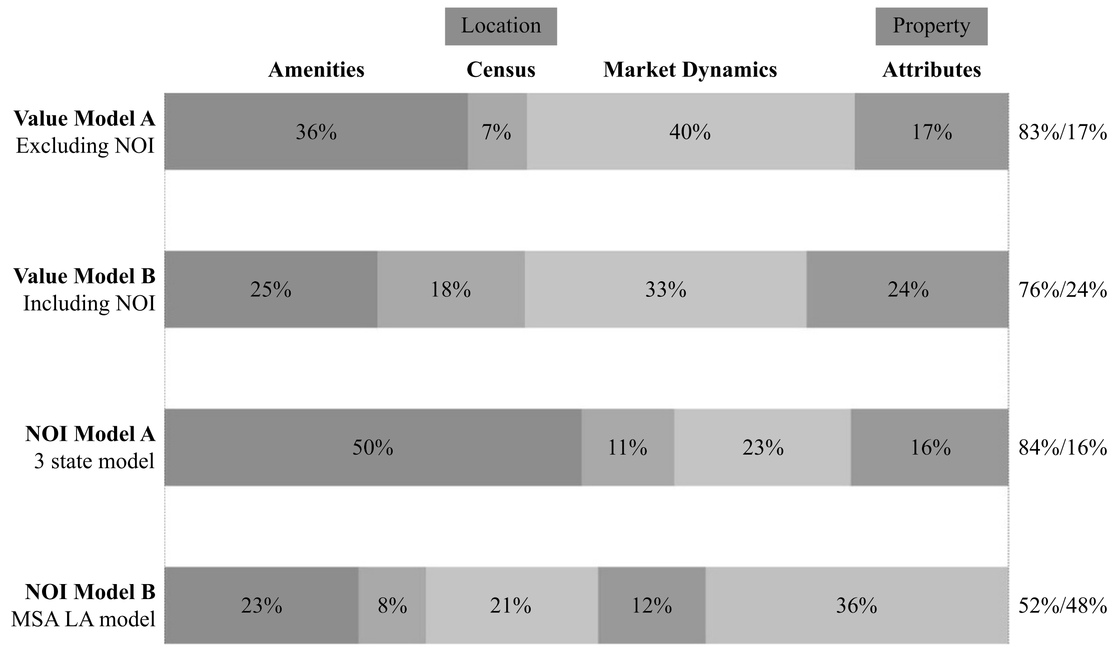
		OLS	Random Forest	Gradient Boost	XG Boost
Panel A: Transaction Price/Unit (excluding NOI)					
	MdAPE	22.6	22.3	18.8	19.1
	R^2	0.58	0.64	0.74	0.73
Panel B: Transaction Price/Unit (including NOI)					
	MdAPE	10.8	13.4	9.3	9.6
	R^2	0.84	0.80	0.89	0.92
Panel C: NOI/Unit					
LA MSA	MdAPE	20.8	15.3	9.9	9.2
	R^2	0.25	0.62	0.79	0.78
Three-State Model	MdAPE	12.8	9.8	8.5	8.6
	R^2	0.73	0.86	0.89	0.88
Panel D: Transaction Price/Unit (including modeled NOI)					
	MdAPE	15.6	15.2	14.3	13.9
	R^2	0.76	0.78	0.80	0.82

Model Performance

Exhibit 4, Panel A provides the KPIs for the model that predicts transaction price per unit. Note that NOI is not included as an explanatory variable in this model. The key statistics in evaluating the models are robustness, as measured by the R^2, and the accuracy of the prediction, as measured by MdAPE. The results provide some important insights. First, the three machine learning models have significantly better performance compared to the simple hedonic model (where the hedonic model include the same set of explanatory variables). Using hedonic modeling, the R^2 of the value/unit model (excluding NOI) is 0.58, whereas the R^2 of the gradient boost machine learning model is much higher at 0.74. This finding indicates that more advanced modeling techniques can be quite beneficial to real estate market analysis. Second, the MdAPE of the best performing model is 18.8%, which is slightly worse than the average manual appraisal, indicating the importance of including NOI in the model.

In Panel B, we therefore add NOI to the model. Given the strong correlation between NOI and transaction price, the performance of the simple ordinary least squares (OLS) model improves quite significantly, with an R^2 of 0.84, but the XG Boost machine learning model is still superior, with an R^2 of 0.92. The MdAPE of the most accurate model is very high, at 9.3%—we

EXHIBIT 5
Model Attributes—Location vs. Property



use this model for a more detailed comparison with traditional appraisals in the next section.

Exhibit 4, Panel C shows the results of the machine learning models applied to NOI data, predicting the NOI/unit both for the LA metropolitan statistical area (MSA) and for the three states of California, Florida, and Texas. The results are consistent with findings for the value models: Machine learning techniques show superior performance relative to traditional OLS. In addition, predicting transaction prices across states provides higher accuracy as compared to predicting transaction prices within an MSA. Although this is counterintuitive—after all, real estate is often perceived to be a local business—it shows that the relationship between explanatory variables such as distance to public transport and presence of schools holds across regions, increasing the sample size for predictive models from three to five local comps to thousands of observations across states. In addition, in a sample that covers a wider geography, the cross-sectional variation of some of the variables (e.g., census data) will increase, leading to a more robust model.

Panel D goes one step further: In the absence of information on NOI for a multifamily asset, we use the predicted NOI as an input to estimate property value. The results of this exercise are encouraging, with strong robustness of the models (the R^2 ranges from 0.80 to 0.82) and an MdAPE of 13.9% at best, not dissimilar to the models that include NOI as an explanatory variable.

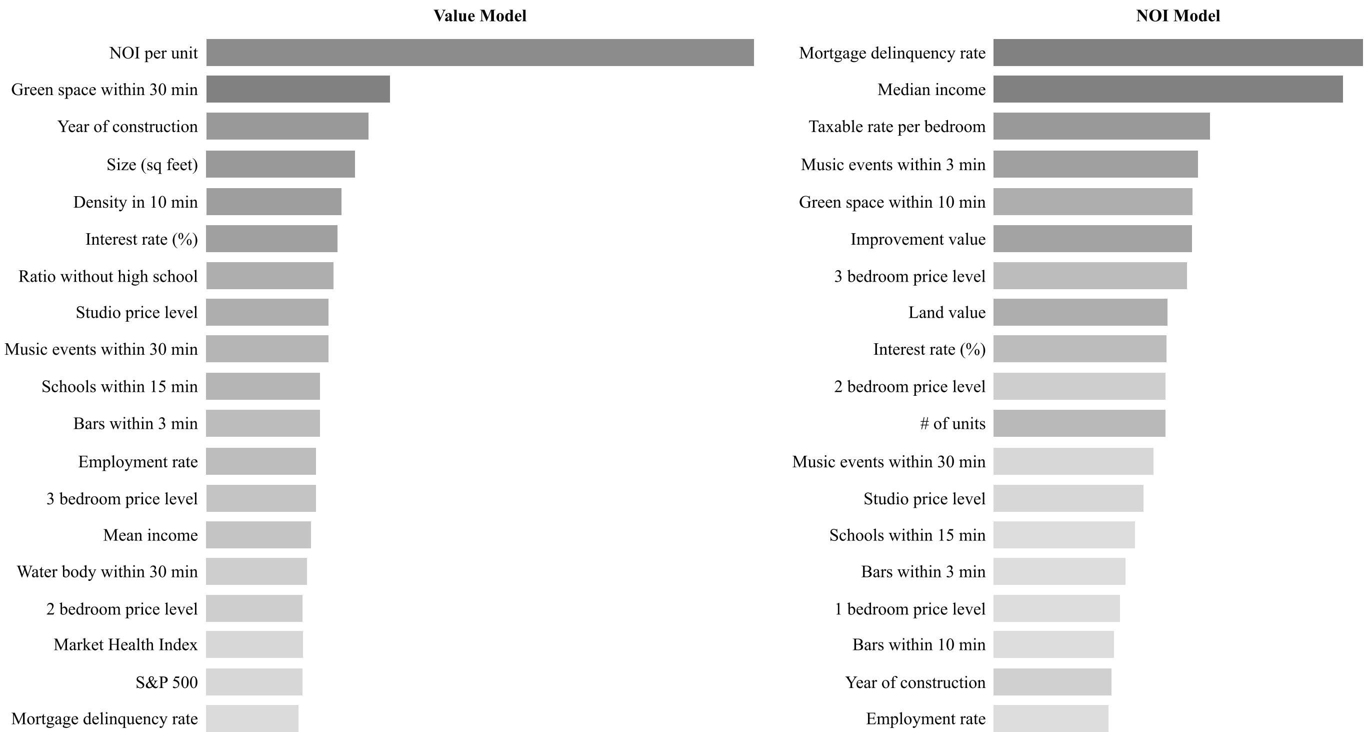
Input Variables

The AVMs estimated in this article are distinct from traditional models not just because of more advanced modeling techniques but also because of the use of nonstandard explanatory variables, including hyperlocal measures of economic activity. To show the relative importance of property characteristics, locational information, and time-varying macroeconomic indicators, Exhibit 5 provides an overview of each category of variables and its importance in the different models.

The results are quite striking: Property characteristics are the dominant characteristics in none of the models. For the model that predicts NOI per unit in the LA market, property characteristics explain less than half of the variance (48%), whereas information on local amenities, census data, and local market dynamics have 52% of the weight. Importantly, 36% of the weight is on property attributes that can be retrieved from tax records, which are open to the public. In the other models, location explains up to 84% of NOI/

EXHIBIT 6

Model Attributes—Explanatory Variables (value/unit)



unit and 83% of value/unit. Although the real estate maxim "location, location, location" is an old adage in the property market, these results empirically confirm the importance of detailed locational information in predicting property prices.

The advantage of machine learning modeling is that a model does not have to be constrained to the 10 to 15 variables that the researcher deems important for property value. Instead, an unlimited number of variables can be used by the model, with computation power being the only constraint. Exhibit 6 shows the most important variables in the value model (using the three-state model). To predict value, NOI per unit is obviously the most important attribute. In addition, the year of construction and average size of the unit are important. Furthermore, locational variables like "green space within 30 minutes" and "music events within 30 minutes" are the variables with the highest weight. These variables can be measured at a very detailed level of granularity, which makes them useful proxies for indicators such as income, employment, and neighborhood vibrancy. Similarly, the most important variable in predicting NOI per unit is the delinquency rate on single-family homes—of course, such a measure is not considered by multifamily tenants and investors directly, but it provides a very granular measure of neighborhood prosperity that cannot otherwise be measured directly.

SUMMARY AND IMPLICATIONS

Assessing the market value of real estate is a daunting task that is still performed manually, typically by comparing an asset to a small set of previously transacted reference buildings that are somehow comparable. We argue that the advent of more advanced modeling techniques, including machine learning, and the growth of data availability, both in frequency and level of observation, allow for the development of AVMs for the real estate sector.

How does the AVM for multifamily stack up against traditional, manual appraisals? Previous research has showed that the difference between appraisals and transactions ranges from 10% to 15% (Cannon and Cole [2011]; MSCI [2016]). In addition an appraisal typically takes about three weeks to obtain and costs some USD 3,000–5,000 for the average commercial asset. Exhibit 4 shows that the automated valuation model beats traditional appraisals on all counts: The MdAPE between property value as predicted by the model and the actual transaction price is 9.3% for multifamily assets in the states of California, Florida, and Texas. Of course, the median error masks a broader distribution: 52% of automated valuations are within 10% of the actual transaction price, and 92% of automated valuations are within 30% of the actual transaction price.

The results are to a large extent determined by the NOI of an asset, as well as granular information on the local market. Importantly, in the absence of information on NOI, our model allows for the inclusion of estimated NOI, leading to a model with a performance equal to the average appraisal—an MdAPE of 13.9%.

The implications of these findings are exciting, although perhaps unnerving to some. For underwriting and refinancing purposes, AVMs can provide an instant indication of property value, which saves significant time and resources for portfolios of both investors and lenders, as well as those interested in a single property. This is especially beneficial on the lending side, where the debt service coverage ratio is a leading indicator, with the loan-to-value as an important, but secondary, input in the underwriting process. Automated appraisals can provide banks, insurance companies, pension funds, and other institutional investors and lenders with an accurate, instant revaluation of the assets on the balance sheet, obviating the need for an expensive and lengthy annual (or quarterly) revaluation process, which regulators increasingly require. Such instant assessment of the market value of the book is especially useful in times of market volatility.

The timeliness of automated valuations also allows for the development of financial trading strategies and innovations in the underwriting process. For example, instant assessment of the value of a commercial real estate asset can be used for new investment models, comparable to what Opendoor has developed in the single-family market. Instant and accurate assessment of the value of a portfolio of assets could be used for arbitrage trading on real estate investment trusts, not dissimilar to the quant trading strategies used by some hedge funds, and banks can resort to automated origination models that obviate the need for (costly) underwriting processes.

Of course, the model developed in this article needs further work to be ready for broader application in the real estate industry. The model needs extension across both time and space—from three states to the entire United States and from the 2011–2016 period to a period that includes the full market cycle, including a significant downturn. Although the model is calibrated based on real market transactions (which also reflect decreases in prices) and includes general macroeconomic indicators as well as real-estate-specific market indicators (such as the mortgage delinquency rate, local market rents, and vacancy rates), the current model is calibrated against the backdrop of a rising market with decreasing cap rates. An important question is how the model responds to declining markets and increasing cap rates. Including a longer time period would also allow for further investigation of time variation in the importance of certain locational and property attributes in the automated valuation model.

In addition, some may argue that multifamily is a relatively homogeneous real estate type relative to retail or office buildings (although previous research from Cannon and Cole [2011] showed that the difference between appraisals and transactions values is lowest for office and industrial assets). The AVM therefore needs to include other property types, where the relative paucity of property transactions may be a constraint (but probably no more so than for traditional appraisals).

Finally, for the industry to accept an automated valuation over a traditional, manual appraisal will take significant education as well as market adoption by leading lenders and investors. The importance of the appraisal business for the bottom line of traditional real estate advisory firms, such as CBRE, C&W, Colliers, and JLL, remains significant, especially during periodic dry spells with limited brokerage activity, and the real estate appraisal industry is deeply entrenched in the traditional real estate business. Notwithstanding the opportunities that more accurate, instant, and cheaper valuations may bring, the USD 90 billion valuation industry has much to lose from disruption of the decades-old model of manual appraisals.

ENDNOTES

We are grateful to Real Capital Analytics for providing access to data and to Mathijs Nelemans, Ali Ayoub, Philipe Isarin (GeoPhy), and Tom Hemmes for assistance with data collection and enrichment. Nils Kok is supported by a VIDI grant from the Dutch National Science Foundation.

[1]See http://www.appraisalinstitute.org/assets/1/7/U.S._Appraiser_Demos_3_1_16.pdf.

[2]The IQR is also called the *midspread*, or the middle 50% (i.e., the values between Q1 and Q3).

REFERENCES

Andonov, A., P. Eichholtz, and N. Kok. "A Global Perspective on Pension Fund Investments in Real Estate." *The Journal of Portfolio Management*, Vol. 39, No. 5 (2013), pp. 32-42.

Breiman, L., J. Friedman, R. Olshen, and C. Stone. *Classification and Regression Trees*. Belmont, CA: Wadsworth, 1984.

Cannon, S., and R. Cole. "How Accurate Are Commercial Real Estate Appraisals? Evidence from 25 Years of NCREIF Sales Data." *The Journal of Portfolio Management*, Vol. 35, No. 5 (2011), pp. 68-88.

Crosby, H., P. Davis, T. Damoulas, and S.A. Jarvis. "A Spatio-Temporal, Gaussian Process Regression, Real-State Price Predictor." In *Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. New York: ACM, 2016.

Eichholtz, P., N. Kok, and J. Quigley. "Doing Well by Doing Good: Green Office Buildings." *The American Economic Review*, Vol. 100, No. 5 (2010), pp. 2492-2509.

Fisher, J., M. Miles, and B. Webb. "How Reliable Are Commercial Appraisals? Another Look." *Real Estate Finance*, Vol. 16, No. 3 (1999), pp. 9-15.

Marcato, G., and T. Key. "Smoothing and Implications for Asset Allocation Choices." *The Journal of Portfolio Management*, Vol. 33, No. 5 (2007), pp. 85-98.

MSCI. "Private Real Estate: Valuation and Sale Price Comparison." *Research Report*, June 2016.

O'Neill, J.W. "An Automated Valuation Model for Hotels." *The Cornell Hotel and Restaurant Administration Quarterly*, Vol. 45, No. 3 (2004), pp. 260-268.

Rosen, S. "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition." *Journal of Political Economy*, Vol. 82, No. 1 (1974), pp. 34-55.

Schulz, R., W. Wersing, and A. Werwatz. "Automated Valuation Modelling: A Specification Exercise." *Journal of Property Research*, Vol. 31, No. 2 (2014), pp. 131-153.