[SMU Data Science Review](https://scholar.smu.edu/datasciencereview)

Volume 3
Number 3 *Fall 2020*

Article 5

2020

Analysis of the Commercial Real Estate Market in a post COVID-19 World

Brandon Croom
Southern Methodist University, bcroom@smu.edu

Sean Kennedy
Southern Methodist University, seank@smu.edu

Sandesh Ojha
Southern Methodist University, sandesho@smu.edu

Justin Sparks
justinssparks@outlook.com

Analysis of the Commercial Real Estate Market in a post COVID-19 World

Brandon Croom[1], Sean Kennedy[1], Sandesh Ojha[1], Justin Sparks[2]
[1] Master of Science in Data Science, Southern Methodist University,
Dallas, TX 75275 USA
[2] WorldVentures Holdings, 5100 Tennyson Pkwy,
Plano, TX 75024 USA
{bcroom, seank, sandesho}@smu.edu
JustinSSparks@outlook.com

Abstract. – The volatility in the commercial real estate market has been greatly influenced by the new societal practices brought about by the COVID-19 pandemic. The COVID-19 pandemic has added additional factors to already complex modeling to value and predict commercial real estate prices. Although multiple methodologies have been applied to commercial real estate valuation, these methods have not yet taken the COVID-19 pandemic factor into account. The main contribution of this article lies in developing an application for commercial real estate valuation which includes the COVID-19 pandemic factor. Thought this article a Hedonic model was developed to compare the impacts of commercial real estate pricing with and without COVID-19 impacts. The model showed no significant changes in the geographic areas analyzed when COVID-19 parameters were input into the model.

1 Introduction

The United States (U.S.) real estate market, both commercial and residential, is highly impacted by various factors. Demographics, interest rates, economic growth, unemployment, and government policies are just a few of these factors. The complexity of modeling these factors by themselves can be difficult. In addition to these factors the COVID-19 pandemic has severely impacted the market and will have lasting economic repercussions.

The COVID-19 pandemic has had a sudden and significant impact on all aspects of people's lives. Although the full economic impact of this pandemic has not been felt yet, it is speculated that these impacts will be felt for years to come. Many of the societal changes that are currently in place will likely remain intact for the foreseeable future, even if an effective vaccine is found [1]. With appropriate protocols, such as social distancing, portions of the economy may be able to restart, but other portions, especially those that involve large scale gatherings, may not be able to restart for some time.

The implications of COVID-19 on the commercial real estate markets will add another factor that must be taken into consideration when assessing individual properties, entire market segments, and company strategies. This single factor will have an impact to various actors in the economy, such as buyers, sellers, banks and governments. The response to this factor by the actors will influence the market overall. As examples, buyers could ask for lower pricing on larger buildings knowing that capacity cannot be reached. Local governments may need to plan differently for how to influence large commercial businesses to come to their jurisdictions. Supply of commercial real estate may also become over saturated due to businesses shutting their doors resulting in financial ripple effects across the entire economy. The ultimate impact is still largely unknown, but these examples provide an early warning of sorts that could be actioned upon if they hold. Future investment dollars will be tightly controlled resulting in financial and economic impacts far into the future. The impact of COVID-19 will reach far beyond the typical boom and bust nature of the cyclical real-estate market resulting in large swings in investments, profitability, availability, and supply chain.

In this article, machine learning techniques are used to predict potential long-tail considerations for investors, builders, and/or consumers interested in future real-estate market. Over the course of the analysis the goal will be to model impacts to the commercial real estate market due to COVID-19. Hedonic models will be used to build out a commercial real estate pricing model in order to predict pricing in various US cities. Hedonic models are often used to build out real estate market models as these models treat the property as a tangible asset based on the property attributions. In addition, difference-in-difference regression will be used to identify the impacts of COVID-19 on specific property locations. Difference-in-difference regression is typically used to measure the impacts of policy or property changes between properties. Coupling these modeling approaches together will allow for understanding the impacts to the commercial real estate market due to COVID-19 and allow for predicting future impacts to the commercial real estate market due to the on-going societal changes due to COVID-19. With the predictions, the intent is to be able to provide actors in the commercial with a means to better operate in this environment.

2 Literature Review

The future real-estate market will be severely impacted following COVID-19 lockdowns and societal changes. Future investment dollars will be rigidly controlled while financial and economic impacts will resonate far into the future – causing some degree of downturn in this sector. The impact of COVID-19 will reach far beyond the typical boom and bust nature of the cyclical real-estate market resulting in large swings in investments, profitability, availability, and supply chain. A review of literature in the area of real estate modeling provides an understanding of what has been accomplished and how it can be built upon for further analysis

2.1 Hedonic Models

Of the multitude of possible valuation metrics available for real estate assets, two of the most prominent are the hedonic model proposed by Lancaster (1966) [2] and the traditional Discounted Future Cash Flow Analysis (DFCF) typically carried out in computational finance. Both methods differ in their predicted valuations of real estate assess. Efforts have been made to bridge this gap in valuation with marginal success [3].

The hedonic model treats the valuation of a tangible asset as a weighted sum of the intrinsic attributes of the asset (square footage, amenities, etc.) and the influence from external attributes (distance to landmarks such as museums and subways). Often this approach is inaccurate for standard linear regression and is typically carried out as log-linear regression in practice.

As it pertains to our analysis of the effect of COVID-19 on real estate pricing, the hedonic model seems very suitable to overlaying external features to the properties we wish to value. Adding covariates which describe the COVID-19 spread (infection rates, hospitalization rates, recovery rates etc.) in the area of a given property will allow us to measure the impact, if any, they have on real estate pricing.

A reasonable assumption would be that commercial real estate is likely to suffer in areas that are dis-proportionally affected by the virus. Many employers are choosing to employ a work-from-home (WFH) strategy to combat the spread [1]. We expect that technologies such as Zoom will be leveraged by corporations as a cost cutting measure that also increases safety. As such, if there are fewer companies willing to purchase or lease real estate for office space, the value of those assets can only decrease.

Understanding valuation and appraisal methods and their shortcomings is particularly important for us within this study, especially since we are attempting to measure the impact of COVID-19 on the commercial real estate market. The difficulty however, using traditional approaches that involve comps from similarly valued properties is that they are never truly comparable, and appraisals can lag the market by as much as 12% above or below the actual transaction price [5]. Automated valuation models (AVM) however, have been shown to be much more accurate (9% absolute error). These models may help to ascertain the true value of a property more accurately than traditional approaches. We could use these models to layer in COVID-19 features and predict future transaction prices of commercial real estate for a specific cross-section of organizations or industries to evaluate the impact of the pandemic. We could leverage accuracy measures, R2, and RMSE metrics to evaluate the models' performance. Moreover, given the rapidly changing environment we are currently facing with COVID-19, traditional valuation methods will likely not keep pace with the market so models such as AVM would be a beneficial to determine accurate valuations.

Financial footnotes can be feature-rich with regards to understanding an organization's financial position following the COVID-19 pandemic. Depending on the organizations selected for our analysis, we will be able to review these sections of the financial statements to identify information (such as risks, future outlooks, and other advisory information) that might be otherwise missed in the raw numbers alone. Given the unstructured nature of the footnotes though, they can be difficult to quickly read without heavy manual intervention. Further, that effort increases as we add more organizations to our research list. Automating the ingestion of this information and speeding up the time to insight would be a beneficial move on our part, especially if we're reviewing multiple organizations. Supervised and unsupervised machine learning approaches have met with minimal success however due primarily to the inherent unstructured nature of footnotes in the financials [6]. Given this, we will likely opt for a manual review of footnotes and financial statements for selected organizations – with the goal of looking for impacts from COVID-19 on real estate, capital expenditure, and financial considerations.

Modeling of interactions between real estate valuations, accrual accounting and corporate valuations has been another approach to model the real estate sector. Miles, Pringle, & Webb argue that the application theoretical modeling to real life situations at the present time is quite challenging due to inadequate information regarding the firm's cost of capital, debt capacity, systematic risks, and operating expenses [7]. Corporations were deciding on real estate based on pure need for additional spaces instead of proper planning and modeling and this study tries to show alternative ways to assess risks and plan when making real estate decisions.

In today's world the challenges faced in 1989 are nonexistent since many corporations can increase their profitability through more effective management of their real estate and the shortage of data is no longer an issue. This entails evaluating real estate on an on-going basis using not only internal factors but also external factors that could have an impact on the value of real estate and its efficiency. In our study we plan to look deeper into the external factor (COVID-19) alongside others to get a deeper understanding of impact on real estate investments now and in the future.

Tezel and Yalpir performed modeling of real estate prices utilizing an Artificial Neural Network (ANN) approach [8]. In this study, an ANN model was structured to simulate the data obtained from the properties of a residential real estate as the area, age, floor condition, physical properties and location of the real estate as 5 input data and the price as 1 output data. The collected data was trained and tested with different hidden layer and neuron number assigning using ANN Toolbox of MATLAB software. The results were presented and the feasibility of ANN in residential real-estate price estimation was discussed. The best performance was achieved with 14-6 neurons in two hidden layer topology. The training and testing errors were obtained as 2.94% and 8,90%, respectively. The market prices were reached by 82% with ANN and R^sup 2^ value was 0,77.

2.2 Real Estate Modeling Through Automated Ensemble Methods

Modeling the real estate market through automated ensemble methods is another machine learning approach that may be taken for analysis. Alfaro-Navarro et al. analyze the approach to using automated ensemble methods for analysis on complex real estate systems [9]. They note that although this method has been leveraged before the focus has been on small geographic locations. Alfaro-Navarro et al. expanded on the small-scale analysis by performing modeling for the entire Spanish market with automatic modelling for each municipality [9]. The results of this study showed good performance of the techniques, when evaluating error measures. The best results from this study were achieved leveraging bagging and random forest techniques [9].

When analyzing real estate geography is also a primary considering for all economic actors involved. Fu et al. evaluate the ways in which real estate purchases are executed [10]. The study evaluates the geographic impact to property valuations. These geographic impacts are related to the individual property characteristics, the values of nearby property, and the prosperity of affiliated latent business areas [10]. The study outlined the design of an algorithm to exploit these geographical dependencies in a probabilistic ranking model. This is performed by extracting known data about these individual property characteristics from various data sources and clustering the affinity of these characteristics. This model was validated against real estate related data in Beijing. The results of the study show this is an accurate prediction method and that these factors have influence on valuation.

2.3 COVID-19 Societal impacts

Moving away from the real estate modeling aspect and beginning to evaluate current literature regarding the societal impacts of COVID-19 information in still sparse from formal studies. Banks et al. provides a good start at evaluating the societal impacts [1]. The article is a conversational article with economists, sociologists, statisticians and operations researchers. The results of this article conclude that society will need to re-engineer activities to allow the return to normal activities without resulting in massive increases of virus infections. The contributors to this article concluded that the economic impacts of the pandemic will be great and will result in long term effects, some of which may still be unknown. Companies will need to change how they approach the societal protocols that need to be in place to prevent virus spread. Those companies that cannot adapt to these societal protocols may no longer exist and thus with respect to areas such as the real estate market have a lasting impact.

Morawska et al. discuss the airborne transmission of COVID-19 and the impacts this has on transmission of the disease [11]. The prevalence of non-contact transmission in cases seen to date indicates that the airborne nature of the virus requires societal precautions to be put into place. These precautions include increased ventilation rates in indoor places, avoiding recirculated air, minimizing the number of individuals co-located, and social distancing [11]. The article further explains that public places such as retail outlets, schools, offices and public transportation are areas that should have high scrutiny of these available precautions. Additionally, the article recommends the use of personal protective equipment as an additive to the above prevention measures [11].

The modeling of social distancing and its impacts are currently being studied in various forms. McMullen attempts to address the social distancing via Coulomb's

Law and a variant of the circle-packing problem [12]. Utilizing an agent-based modeling system the researchers were able to model individuals as particles and mapping the particle affinity to each other. The initial results of this paper suggest that social distancing will be with us for a while and that the effects of social distancing can be applied to future crises that are similar in nature. The paper also suggests that leveraging the current unit square methodology for social distancing may not be adequate to large scale venues such as sporting events and worship services [12]. The author suggests that leveraging differing shapes based on the size of these large venues may result in more optimal capacity while maintaining social distancing. This work may help to inform real estate modeling in that it could suggest an alternative method of maintaining optimal usage of commercial real estate space in the current COVID-19 crisis.

2.3 Pandemic Real Estate Modeling

As the COVID-19 pandemic spread across the globe many correlations were made to the 1918 Flu pandemic. Lilley et al. reviewed the economic impacts of the 1918 Flu pandemic and the mitigation efforts that Non-Pharmaceutical Interventions (NPIs) had on manufacturing employment and manufacturing output [13]. Using difference-in-differences regressions in multiple configurations the team was able to model the impacts of NPIs on 43 US cities evaluating the impacts to manufacturing employment and outputs from 1899 to 1927. [13]. Through this modeling it was determined that there was an apparent positive effect of NPIs at the city level. Although a positive sign the researchers caution against drawing conclusions from this analysis due to confounding factors including divergent growth across cities during the time period. Through further analysis the researchers accounted for the confounding trends, including pre-pandemic growth impacts. With the revised model, the results were statistically inconclusive [13]. Specifically, implementing NPIs ten days earlier than the pandemic or implementing more than one NPI did not provide clear evidence of the effects of non-pharmaceutical interventions on economic growth.

Francke et al. modeled historical outbreaks of the plague in 17th-century Amsterdam and cholera in 19th-century Paris to evaluate the impacts on housing markets [14]. An important note in the research indicates that the 1918 Flu data was not considered due to that pandemic in this geographical region being linked to wars which also impacted the housing market [14]. The researchers built multiple models for each city that tracked the epidemic spread, consumer prices, wages and interest rates. Additionally, some models controlled for rental or housing price changes pre- and post-pandemic for three years to account for unobserved time trends [14]. The results conclude that there are significant but short-lived declines in housing prices, and smaller declines in rental prices [14]. The property price declines are highest immediately after the outbreak and in areas most impacted by the epidemic. The researchers believe this result is due to the increased uncertainty and economic disruption [14]. The lack of long-term impacts is attributed to the resilience of cities to handle the shock brought about through the epidemic and leveraging that shock for significant urban change [14].

Ambrus et al. modeled the cholera epidemic of 1854 in a single urban parish in London to evaluate the economic impacts epidemics inflicted on the economy [15].

Utilizing a regression discontinuity design the researchers were able to conclude that ten years after the epidemic housing prices were significantly lower closest to where the outbreak started [15]. These pricing differences remained for 160 years, thus indicating the long-term impacts that epidemics may have on economics. Furthermore, the research indicates that these economic shocks resulted in a widening of the gap between individuals; less fortunate individuals lived closer to the initial source of the epidemic [15]. The authors further conclude that these economic shocks also provide rationale for interventions such as upgrading of poorer neighborhood's or "urban renewal" [15].

Ling et al. built upon the works in an examination of COVID-19 impacts on commercial real estate prices [16]. The researchers focused on commercial real estate (CRE) assets owned by listed US equity real estate investment trusts (REITs) [16]. Utilizing a multivariate analysis, the researchers modeled the daily COVID-19 impacts to REITs, accounting for the differences in exposure to COVID-19 based on locations of their properties [16]. Early-stage pandemic analysis, January 2020 through April 2020, indicate a large variation in performance across REIT property types. The researchers concluded that REITs with a heavy emphasis on retail, office and residential properties performed the worst, while health care and technology performed the best [16]. Factoring in NPI into the analysis, the researchers also concluded that firms that rely heavily on face-to-face communication or close physical proximity may find it less attractive to locate to high-density areas post-pandemic [16].

Milcheva focused in on the slope of the security market line (SML) for real estate equities [17]. The SML describes return and risk relationships. Utilizing Fama-MacBeth regressions, Milcheva finds that the across countries and across sectors there are large differences in the average firm sensitive to COVID-19 risk [17]. Real estate sectors such as retail have the highest sensitivity to COVID-19 risks while healthcare has the lowest sensitivity. Milcheva also concludes that firms with high financial leverage ratios would be negatively impacted by COVID-19 in comparison to those firms with sound fundamentals [17]. Additionally, the research concluded that in geographic locations that had prior experience to similar pandemics firms maintained higher returns [17].

Similar REIT modeling has been performed in Hong Kong by Xie et al. [18]. Utilizing a difference-in-differences approach to examine the role of COVID-19 infection proximity to returns of REITs, Xie et al. concluded that there are significant negative impacts on stock returns. A REIT with property within two miles of a COVID-19 case had a lower daily stock return of 0.02% immediately following with COVID-19 case disclosure [18]. Accounting for property types, REITs that maintained a high number of residential properties had a lower daily stock return of 0.1% [18]. These results are like those described in Ling et al. and provide global similarity to REIT impacts as related to COVID-19.

Continuing to build on the overall response to financial impacts related to COVID-19, Gerding et al. evaluated stock prices and the pandemic impacts [19]. The study evaluated 29,000 firms across 100 countries [19]. Utilizing cross-sectional regression, the researchers concluded that the stock market expects larger economic losses in countries with higher debt-to-GDP ratios [19]. The research findings indicate that countries with better fiscal capacity help mitigate the pandemic's negative effect on economic activity [19].

Baker et al. evaluate the impacts of the COVID-19 pandemic on the overall stock market response [20]. Utilizing newspapers articles and corporate stock prices the researchers were able to attribute a large portion of stock market reaction for COVID-19 cannot be explained by the lethality of the virus [20]. The researchers compared stock market trends for both the 1918 Flu epidemic and the 1957-58 influenza pandemic to the current stock market trends during the COVID-19 pandemic. There is also little influence on stock price changes due to the speed with which information moves in the current time. In conclusion the researchers argue that the impact of commercial activity due to government restrictions relating to NPI and social distancing are largely the influence for the stock market responses [20]. With less ability to travel, school closures, stay-at-home orders, public gathering bans, and non-essential business closures, the economy will continue to be impacted [20].

3 Data

For this analysis multiple data sources were leveraged. The first data set discloses detailed location information of COVID-19 cases in the US. The second data set consists of property listing information

3.1 COVID-19 Dataset in US

COVID-19 is one of the most severe pandemics in recent history. Its impact globally is tremendous. In tracking information related to COVID-19 John Hopkins University (JHU) has made available a data set that aggregates COVID-19 information from multiple sources around the globe. These sources range from global sources such as the World Health Organization to individual US state feeds.

The data set contains information relating to the number of COVID-19 cases and the status of those cases. From the data set the number of deaths, active cases, recovered cases and confirmed cases may be obtained. The data set granularity is at the latitude and longitude level based on geographic centroids.

In order to build the dataset for this analysis, the COVID-19 data is scraped daily, filtering down the data set to US only data points. The data set represents all historical data that the US entities have provided as well as current date information. Data has been pre-cleaned and thus requires minimal manipulation.

3.2 Commercial Real Estate Sales and Leases

The other key data set is commercial real estate sales and lease information. Data at this level allows for an understanding of the patterns that are occurring in commercial real estate around the US. LoopNet.com is the primary source of data.

LoopNet.com provides listings of commercial real estate sales and leases based on geographic region and property type.

The LoopNet dataset contains information relating to the individual properties for sale or lease in a geographic location. Each property listing contains information relating to the general building information (age, property type), building dimensions, lot dimensions, zoning information, and pricing information. This information is at the latitude and longitude level.

The LoopNet.com website is scraped daily to obtain all property listings that have been listed. The data is then stored in raw format in a database specifically for this analysis. Historical property listing data is only mapped as far back as COVID-19 data to maintain timeframe consistency.

3.3 Merged Dataset

By leveraging the latitude and longitude in both the LoopNet dataset and the COVID-19 dataset a combined dataset can be build out. This dataset contains the joint values of the commercial real estate available for lease or for sale coupled with the COVID-19 data.

With these merged datasets, there are two key location variables for analysis. The first is the location of the property for sale or lease. The second is the coordinates of the COVID-19 case. In addition, the time variables for COVID-19 case detection and COVID-19.

4 Methods

4.1 Understanding Hedonic Models

Since the effects of COVID-19-related variables are likely to be non-linear and highly dependent on interactions, an approach like that taken in Ishijima may be appropriate [3]. Leveraging the Box-Cox transformation will allow us to assess the degree of non-linearity in our response variable.

Another problem with traditional valuation models is how to account for the differences in valuation for properties that differ in their height above the ground (i.e., the floor of a particular property in a building can have a profound effect on the value). Unfortunately, whether there is a Z variable to be measured depends largely on the (X, Y) coordinates of the location. Not all locations have multiple floors. Modeling these differences can be a challenge.

Since the Z coordinate is largely dependent on the (X, Y) coordinates, adopting an approach like that undertaken in Mimis, 2016 [4] will be very helpful. By modeling the (X, Y) spatial effects separately from the Z effects, we hope to increase the accuracy of our model. Unlike Mimis, we will use a grid search technique to pinpoint the optimal values of **a and b** which are used to construct the Z-dimension affinity matrix (**H**).

The mathematical methods behind this approach are rather straightforward and involve decomposing the affinity matrix into two components, one that considers Euclidean distance in (X, Y) space and another that has a custom definition of distance which applies to the Z (floor) space.

After decomposing the matrix into two matrices, they are then recombined through the **Hadamard Matrix Product** (HMP), which is an elementwise, distributed multiplication of an m x n matrix with another matrix of the same dimension [5]. Unlike the traditional dot/matrix product, this operation returns elements as follows and is undefined for matrices where dimensions differ:

$$(A \cdot B)_{i,j} = (A)_{i,j}(B)_{i,j}$$

In this way, we should be able to account for floor differences in our valuation process. By setting a small, fixed minimum distance value (to locate spaces that are in the same building) we are effectively weighting their traditional Euclidean affinity matrix as follows (for two places in the same location with one floor of separation):

$$S_{i,j} = 1$$

$$H_{i,j} = 1/(b \cdot 1^a)$$

which gives the values for the affinity matrix (**W**) as follows:

$$W_{i,j} = 1/b$$

4.2 Understanding Difference-in-Difference Models

The difference-in-difference (DD) technique is an experimental design that is typically used in policy analysis. DD techniques evaluate the impacts a policy has on a control group and treatment group across the time periods before and after the treatment were applied. This technique provides a more accurate way of verifying that the average differences between treatment and control groups across time are significant. DD also eliminates fixed factors that could impact treatment and control groups.

Difference-in-difference techniques are essentially regression analysis. The basic formula behind the difference-in-difference technique is:

$$y = \beta_0 + \delta_0 d2 + \beta_1 dT + \delta_1 d2 * dT + other\ factors$$

$$T = \begin{Bmatrix} = 1\ if\ in\ Treatment\ Group \\ = 0\ if\ not\ in\ Treatment\ Group \end{Bmatrix}$$

$$d2 = \begin{Bmatrix} = 1\ if\ Post\ Policy \\ = 0\ if\ Pre - Policy \end{Bmatrix}$$

As it relates to COVID-19 modelling the will use the proximity of COVID-19 cases to a real estate entity as the treatment. If a COVID-19 outbreak is not close to real estate entity, then those will be placed in the control group. This approach is similar to the approach defined in Xie [18].

5 Results

5.1 Exploratory Data Analysis (EDA)

Using the data sources defined in Section 3.3, exploratory data analysis (EDA) occurred on the data. EDA allows for understanding which features in the data set can have impacts to models due to missing or incorrect values. EDA also provides a way to understand the structure of the data. Given the need to ensure the property data was accurate additional time was spent in EDA. In Figure 1 the property price by state is evaluated for the state of New York. The information obtained from Figure 1 provides a baseline to work against as it relates to property pricing for this state. Figure 1 shows that the most common property pricing is less than $20 million dollars.

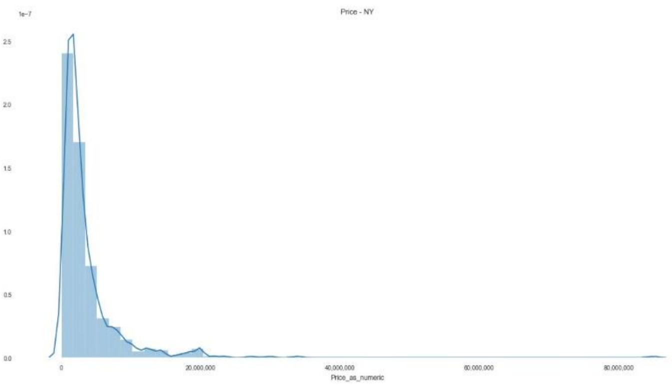

Fig. 1. Property Price by State for New York

Contrasting Figure 1 with Figure 2, which is property price for the state of Michigan, the most common property pricing in Michigan is less than $5 million dollars. The geographical implications of the location of the commercial real estate are observed with these figures.

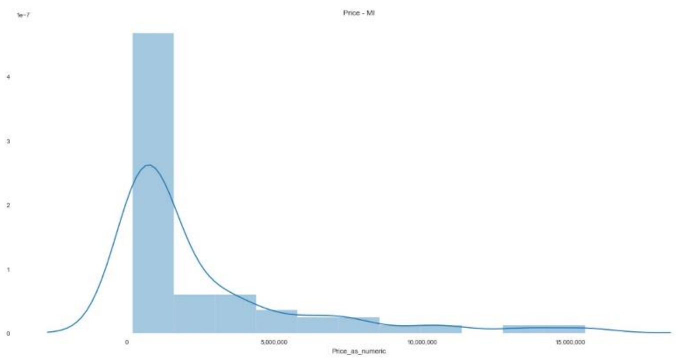

Fig. 2. Property Price by State for Michigan

Another key feature that needs to be understood for appropriate property valuation is the square footage of the property. Similar to the price feature, the square footage feature was evaluated in a similar fashion at the state level. Figures 3 and 4 show the square footage for properties in the states of New York and Michigan respectively.

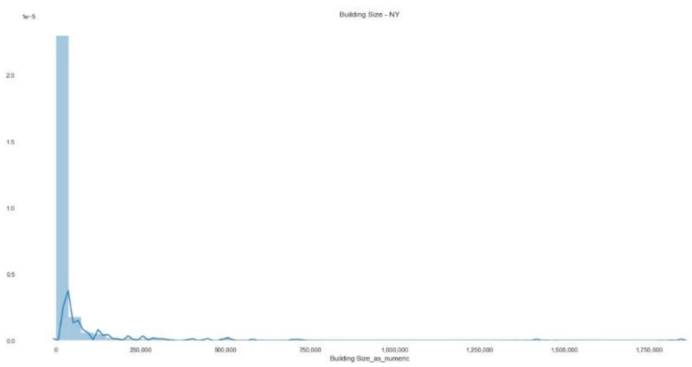

Fig. 3. Building Square Footage by State for New York

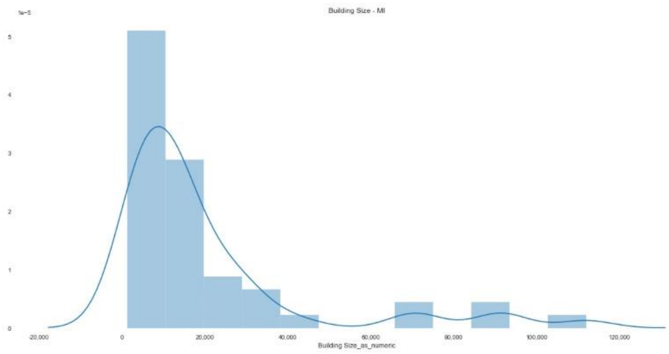

Fig. 4. Building Square Footage by State for Michigan

From Figures 3 and 4 the geographical implications of building square footage become apparent. In New York, the majority of the properties for sale have square footage less than 250,000 square feet. While in Michigan, the majority of the properties for sale have less that 40,000 square feet and no properties for sale are as large as the average property in New York.

Combining both the pricing and square footage features allows for creation of a new feature for price per square foot. As seen in Figures 1 – 4, the geographical differences in pricing and square footage alone make comparisons across states difficult. Evaluating models with a price per square foot feature allows for a more easily understandable feature to do comparisons against. Figures 5 and 6 show the price per square foot feature for New York and Michigan.

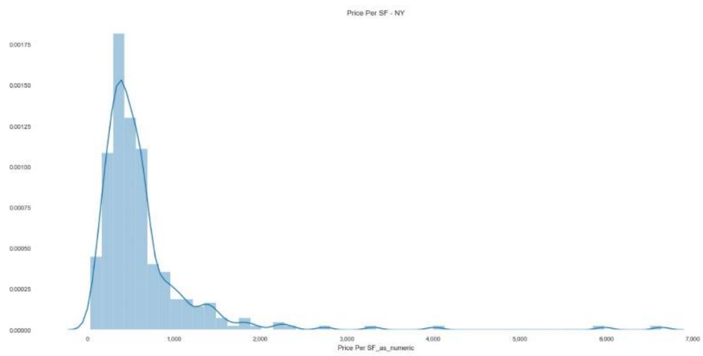

Fig. 5. Price Per Square Foot for New York

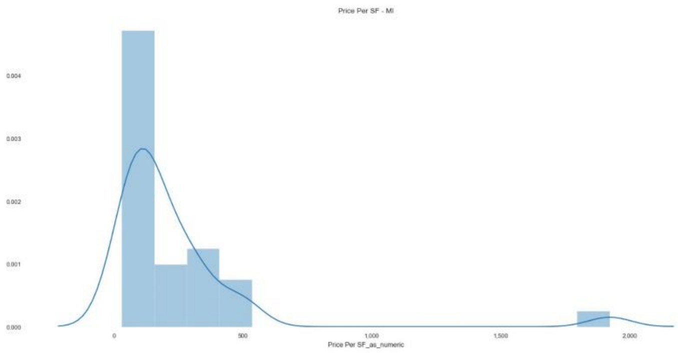

Fig. 6. Price Per Square Foot for Michigan

EDA for other states followed the same pattern as shown in Figures 1 – 6. These values were then able to feed into the models for analysis.

5.1 Outlier Detection/Cleanup

As data cleanup progresses, initial modeling will be focused on properties that are **for sale** - properties that are **for lease** will be analyzed separately. Due to the inexact science of parsing each listing, the modeling will focus on only those properties that have prices/sq ft less than 100 USD. Typical real estate costs, for furnished office spaces in Manhattan, (Figure 7) are less than 100 sq/ft for most buildings.

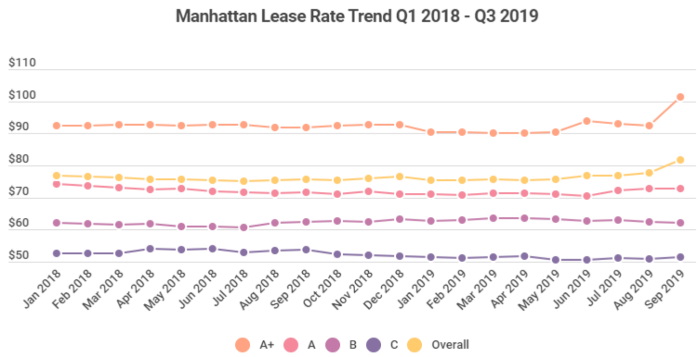

Fig. 7. Manhattan Lease Rate Trend Q1 2019 – Q3 2019

All numerical data will be normalized to mean zero and unit variance. All missing categorical data will be imputed to **Unknown** wherever the values are missing.

5.2 Basic Hedonic Model

A basic hedonic model was constructed on properties observed where both the square footage and price were observed. Square footage and price are calculated fields based on a variety of observed features for each property. The parsing logic yields sufficient data for preliminary analysis. The baseline model is a standard OLS regression on a subset of featured with representative coverage using statsmodels.api. The training set consists of 7, 131 observations. The results of the OLS regression model are shown in Figure 8.

Fig. 8. Regression Results

```
                            OLS Regression Results
==============================================================================
Dep. Variable:            Calcd_Price_SF   R-squared:                       0.002
Model:                               OLS   Adj. R-squared:                 -0.002
Method:                    Least Squares   F-statistic:                    0.5372
Date:                   Sun, 04 Oct 2020   Prob (F-statistic):              0.982
Time:                           07:59:38   Log-Likelihood:                -72507.
No. Observations:                   7131   AIC:                         1.451e+05
Df Residuals:                       7100   BIC:                         1.453e+05
Df Model:                             30
Covariance Type:               nonrobust
==================================================================================
```

Without outlier cleanup, the model is rather poor and shows that additional feature engineering and data cleanup is required. Evaluating the model on R-squared yields poor results. Digging deeper in the parameter estimate shows serious problems with outliers in certain states (Figure 9)

Parameter	**Coef.**	**Std.Err.**	**t**	**P>|t|**	**[0.025**	**0.975]**
Square Footage	-5.89	73.75	-0.08	0.94	-150.10	138.33
Arizona	-77.87	369.47	-0.21	0.83	-802.16	646.40
California	220.60	208.32	1.05	0.03	-187.77	628.99
Colorado	-28.79	628.88	-0.05	0.96	-1261.58	1204.01
Florida	13.47	290.96	0.05	0.96	-553.91	583.86
Georgia	-27.92	321.66	-0.09	0.93	-658.48	602.63
Illinois	-163.75	297.58	-0.55	0.58	-747.10	419.60
Michigan	-7.10	678.68	-0.01	0.99	-1337.50	1323.34
New York	209.50	272.34	0.77	0.44	-324.38	743.37
Texas	-12.65	243.96	-0.05	0.96	-490.88	465.58
Virginia	99.58	547.48	0.18	0.86	-973.64	1172.80
Washington	4.81	695.71	0.01	0.99	-1357.93	1367.55

Fig. 9. Model with no outlier cleanup

It is entirely expected the California and New York would have larger positive parameter estimates given that the expected cost per square foot is higher in those states relative to others. While these to parameter estimates show the most statistical significance (p-values of 0.29 and 0.442 respectively), they are far from meeting any type of significance level that would result in a viable model. Every other estimate appears to be entirely random: p-values are nowhere near statistical significance. Examining the residuals in Figure 10 indicates that a few poor predictions are driving the poor performance of the model.

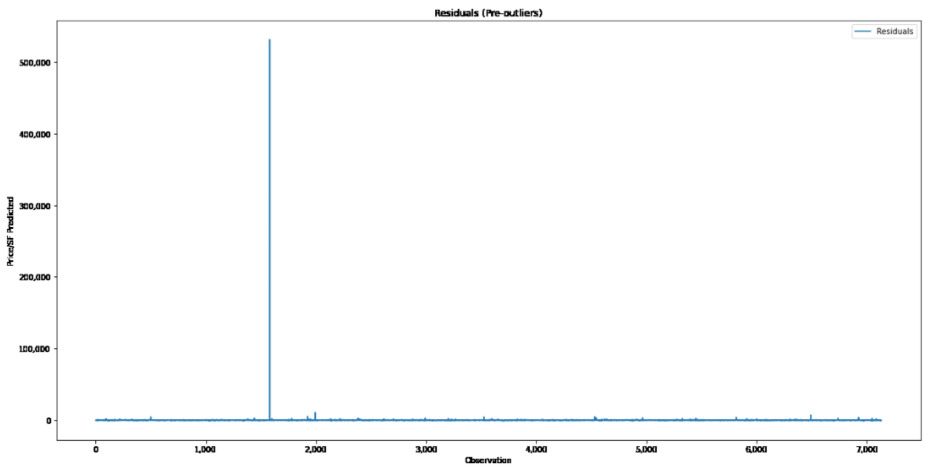

Fig. 10. Residual Estimates

A 20% holdout set (1,783 observations) was used to make predictions and evaluate the out-of-sample performance of the model. Predictions without outlier cleanup are predictably poor. The mean squared error (MSE) of this model is poor at MSE = 104,265)

Limiting the scope of the analysis to only those properties with the calculated price per square foot less than $100 and the calculated square footage less than or equal to 500k square feet yields far better results though it limits the sample size to 1,765 observations in the training set and 442 observations in the test set. The results of this OLS model are shown in Figures 11 – 14.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:            Calcd_Price_SF   R-squared:                       0.648
Model:                               OLS   Adj. R-squared:                  0.642
Method:                    Least Squares   F-statistic:                     118.3
Date:                   Sun, 04 Oct 2020   Prob (F-statistic):               0.00
Time:                           10:19:21   Log-Likelihood:                -7722.3
No. Observations:                   1765   AIC:                         1.550e+04
Df Residuals:                       1737   BIC:                         1.565e+04
Df Model:                             27
Covariance Type:               nonrobust
```

Fig. 11. Revised OLS Model (Post Outlier Cleanup)

Parameter	Coef.	Std.Err.	t	P>|t|	[0.025	0.975]
Square Footage	-7.35	0.80	-9.17	0.00	-8.93	-5.78
Arizona	0.67	2.67	0.25	0.80	-4.58	5.92
California	14.71	1.64	8.97	0.00	11.49	17.93
Florida	4.55	2.38	1.91	0.06	-0.13	9.23
Georgia	1.20	1.95	0.61	0.54	-2.63	5.02
Illinois	-1.37	2.25	-0.61	0.54	-5.79	3.05
Michigan	4.67	4.02	1.16	0.25	-3.22	12.56
New York	15.06	4.95	3.04	0.00	5.33	24.78
Texas	0.85	1.86	0.46	0.65	-2.80	4.50
Virginia	-0.96	3.87	-0.25	0.80	-8.56	6.63
Washington	10.69	4.93	2.17	0.03	1.00	20.38

Fig. 12. Revised Model Output (Post Outlier Cleanup)

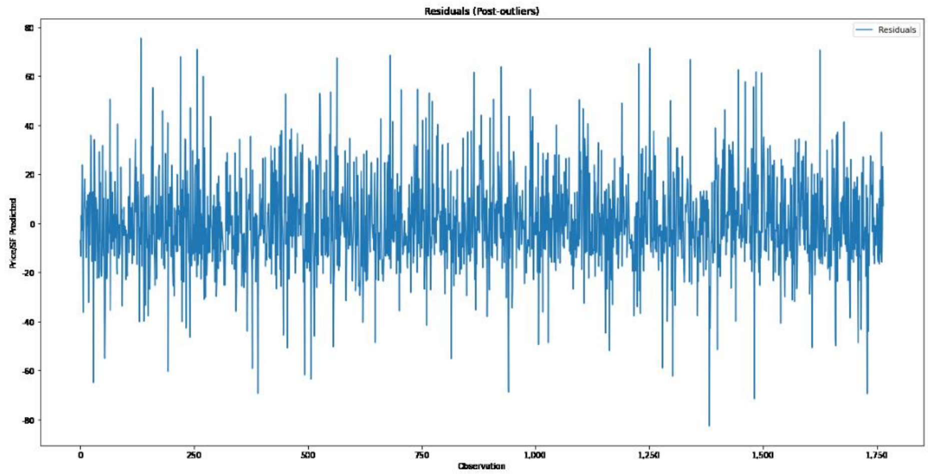

Fig. 13. Residual Estimates (Post Outlier Cleanup)

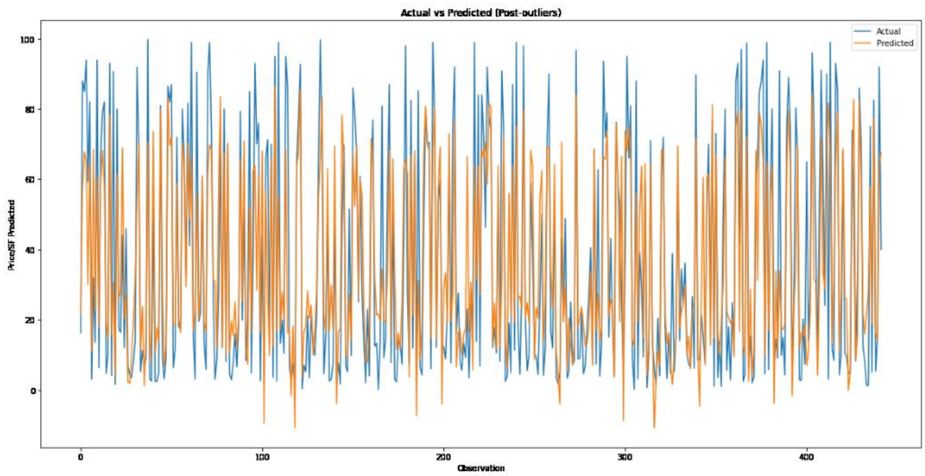

Fig. 14. Actual vs. Predicted (Post Outlier Cleanup)

After adjusting for outliers, the parameters for each state start to make more sense. New York and California still show a premium per square foot relative to the other states. California real estate is, on average, $15 per square foot more expensive, whereas Illinois is $1.40 per square foot less expensive than the average. Five states that did not meet the threshold alpha of 5% (p-values less than 0.05) were Georgia, Illinois, Texas, Virginia and Arizona. These states are the least represented in the sample. It is expected that the parameter estimates will tend more toward statistical significance as additional data is collected. There was also a marked improvement in mean squared error after adjusting for outliers (MSE = 325).

5.2 Hedonic Model (with COVID-19 Data)

Adding a feature to our dataset that tracks the percentage increase in COVID-19 related deaths from July 1st, 2020 to September 23rd, 2020 (roughly the start and end dates of our data collection) yields the feature shown in Figure 15.

state	Percent Death Increase: July-Sept
NY	0.022963
MI	0.131494
IL	0.222595
CO	0.267105
WA	0.554054
VA	0.729563
GA	1.395826
CA	1.496552
AZ	2.212209
FL	2.775890
TX	5.097944

Fig. 15. Percentage increase in COVID-19 Deaths

New York has had the lowest number of deaths increase at 2.2% whereas Texas has seen a 500% increase in deaths over the same period. Simply adding this parameter to the model has little to no effect on the r-squared values or the mean squared error ($325). In the following sections, a more meaningful application of the COVID-19 time series will be applied to a subset of locations which had multiple observations of price changes - these listings will serve as the panel for our time series. Matching the COVID-19 data to the paneled observations should be more effective at determining the effect, if any, of COVID-19 death rates on the effect of real estate pricing. Due to the downward biased nature of price listings over time - logic dictates that the longer a property is listed - the more likely it is that the price of said property will be declining - properties that hit the bid are not likely to be listed long - the model will look to isolate whether there exists a larger downward trend in locations where the COVID-19 infection has taken a stronger hold.

5.2.2 Other Categorical Predictors:

Property Type

Parameter	Coef.	Std.Err.	t	P>|t|	[0.025	0.975]
Flex	17.26	6.52	2.65	0.01	4.45	30.06
Industrial	12.62	3.06	4.12	0.00	6.61	18.63
Land	-35.08	3.77	-9.31	0.00	-42.48	-27.68
Multifamily	16.60	3.39	4.90	0.00	9.94	23.25
Office	18.81	3.27	5.75	0.00	12.39	25.23
Retail	16.05	2.73	5.89	0.00	10.70	21.40
Specialty	-6.92	5.35	-1.29	0.20	-17.43	3.59
Sports And Entertainment	0.00	0.00	0.82	0.41	0.00	0.00
Unknown	10.73	5.26	2.04	0.04	0.41	21.06

Fig. 16. Other Categorical Predictors

Land: Clearly a drag on price per/sf. Undeveloped land will always be cheaper relative to its counterparts.

Multifamily: Multifamily housing is typically more expensive/SF due to the capacity to use the space for rentable income as well as personal living space.

Industrial: Expected to be the lowest premium of the non-land categories, and it is. Industrial spaces are typically larger and cheaper per square foot relative to office spaces and retail spaces.

Office/Retail: Office and retail space should be expected to exhibit similar pricing behaviors as they occupy similar functions within the economy.
Sure enough - these two parameters are nearly identical and are also more expensive than their industrial counterparts.

Flex/Specialty/Unknown: More analysis needs to be carried out to assess the premiums associated with these classes. While the p-values and parameter estimates are all within expected tolerances - more data mining needs to be done to assess the properties that fall into the classes.

Sale Type

Parameter	Coef.	Std.Err.	t	P>|t|	[0.025	0.975]
Investment	10.51	2.91	3.61	0.00	4.79	16.22
Investment NNN	4.40	6.57	0.67	0.50	-8.50	17.30
Investment or Owner User	8.81	3.00	2.94	0.00	2.92	14.69
Owner User	5.81	3.21	1.81	0.07	-0.50	12.12
Unknown	20.53	11.65	1.76	0.08	-2.34	43.40

Fig. 17. Sale Type Indicators

Investment/Investment NNN/Investment or Owner User: From the table above, it seems safe to combine *Investment* and *Investment or Owner User* into one category: Investment. After the combination, parameter estimates for Investment NNN are close but vary somewhat with respect to standard errors (and hence upper/lower confidence intervals). More analysis should be undertaken to analyze the characteristics for listings that fall in the Investment NNN category as it is not immediately clear what NNN represents.

Owner User: This category seems to garner a lower premium for price per square foot than the other categories (other than Unknown). The parameter has a lower CI value near zero, further analysis should be done to determine what properties fall into this category.

Unknown: As the dummy variable, it is encouraging that the model does not deem the parameter estimate to be significant. Further refinement of the data scraping should limit the occurrences of this category in future analyses.

Building Class: The model is not picking up the Building Class parameter well. This is most likely due to many values being imputed into the Unknown category. In theory, Building Class is an ordinal categorical variable (i. e. Class A should be treated as being more valuable than class B). That is not how the model is treating the category at all. While all the parameters show statistical significance, Class A should logically have the highest estimate of the group. That is diverges significantly indicates further refinements are required. The statistical results of the Building Class variable are shown in Figure 18.

Parameter	Coef.	Std.Err.	t	P>|t|	[0.025	0.975]
Building Class A	8.79	8.91	0.99	0.32	-8.70	26.29
Building Class B	16.44	3.25	5.05	0.00	10.05	22.82
Building Class C	12.50	3.06	4.08	0.00	6.49	18.52
Building Class F	0.00	0.00			0.00	0.00
Building Class Unknown	12.33	4.23	2.91	0.00	4.02	20.64

Fig. 18. Building Class Analysis

Opportunity Zone

Parameter	Coef.	Std.Err.	t	P>|t|	[0.025	0.975]
Opportunity Zone - No	25.95	1.98	13.09	0.00	22.06	29.84
Opportunity Zone - Yes	24.11	2.48	9.72	0.00	19.24	28.98

Fig. 19. Opportunity Zone

An opportunity zone is a designated geographic area identified by census tract. Historically, these zones have been lower income zones where the US government has tried to stimulate economic activity by offering businesses certain tax benefits. There are approximately 8,700 opportunity zones in both rural and urban areas throughout the US. Because these zones offer tax benefits, it is not surprising that they garner a premium per sf relative to zones that do not offer the same benefit. The strong statistical significance of the dummy variable needs to be investigated in further analyses.

6 Discussion

Evaluation of the impact of COVID-19 on the commercial real estate market is one of the many topics that is foremost on the collective members of society. The implications of this topic on local governments, businesses and the general population have impacts that may last for years to come. Understanding which sectors of commercial real estate are impacted due to the social changes required by COVID-19 can help to provide guidance on how specific commercial properties may be utilized in the future.

Over the course of this paper the modeling of real estate impacts due to COVID-19 have been addressed. The modeling started with a basic hedonic model of real estate prices without any COVID-19 factors. This model provided a baseline for the analysis and allowed for understanding how real estate pricing in general was performing over the time period. The results from the basic hedonic model with no COVID-19 overlay showed that the values that intuitively have an impact on real estate prices statistically do impact those prices, specifically the price per square foot and property type. Additionally, the initial hedonic model identified California, Illinois, and New York as states that were statistically significant. This result highlights the impact that states with larger populations have on real estate pricing when viewed on a total US basis. This result is expected intuitively, however it is interesting that not all large states, such as Texas, show as significant to the model. This could indicate the overall geographic differences when it comes to commercial real estate within those states. Just because the state has a large population does not imply that it has a large amount of commercial real estate for sale.

Enhancing the Hedonic model with the COVID-19 data, the results indicate that the increases in number of COVID-19 did not have any significant impact on real estate pricing. This is to be expected in a hedonic model as each value is treated as independent.

6.1 Ethical Implications

The ethical implications of this research are driven largely by the data that was gathered. At noted in Waller et al. there are numerous ethical questions faced by real estate brokers, such as dual agency and broker owned properties [21]. These ethical dilemmas also manifest themselves in the pricing of real estate. Understanding that there is the potential for the commercial real estate data to contain some bias as it relates to these factors forces further comparison of the data sources to attempt to mitigate bias. At the same time, economic market forces also assist in mitigating some of this bias.

With respect to the COVID-19 data, there could also be bias in that data. Given the speed with which the pandemic hit and quick need to track the spread multiple data sources were gathered as part of the John Hopkins data set. According to Angelopoulos et al. there are possible estimation criteria that much be compensated for as they relate to deaths, cases and recoveries when evaluating COVID-19 data [21]. The imperfections in the data reporting for COVID-19 could result in bias as underrepresented populations could not be included in the data. The imperfections in the data reporting could also be due to delayed reporting by hospitals and other healthcare entities.

The political climate around COVID-19 imparts bias into the data sets. COVID-19 has become a very political topic at the federal level. Finding a trusted data set that was not politically biased was a key consideration over the course of this research. As the political landscape changed, especially between the Center for Disease Control (CDC), Health and Human Services (HHS) and the White House, data sets changed or stopped all together. The Johns Hopkins data was not as impacted by these changes.

With the combination of location-based data and COVID-19 data the bias in the individual data sets could be magnified and other bias characteristics may emerge. For example, if a specific set of buildings is in a geographically depressed area but the COVID-19 information is underreported it may be that the pricing methodology becomes biased towards these areas as they could look attractive.

6.2 Challenges and Limitations

Over the course of the research there were challenges encountered. As noted in section 6.1, the political climate around COVID-19 was a concern. The data leveraged for COVID-19 was impacted by some of these political issues. On July 15, 2020, COVID-19 data reporting was changed under US Federal Government decision. This resulted in the data layout changing for multiple COVID-19 data reporting that was being used as supplemental information to the John Hopkins COVID-19 data. Luckily, the primary data source used was unchanged over this time however there was a concern that these reporting changes would impact the amount and timing of data available for research.

Another challenge during the research was the sheer size of commercial property data when looking at the US in totality. The size of this data resulted in a decision being made to evaluate commercial real estate data for certain states only. Scaling down to this smaller data set allowed for more applicable models to be built and refined. Expanding the models to other states would not be complex but could result in the need for an individual model by state to paint a more accurate picture of the COVID-19 impacts on commercial real estate.

A final challenge that was encountered during the research was timeliness of data. The imperfections in the COVID-19 data reporting and the availability of that data in a timely fashion had some impact on the speed by which models could be evaluated and analyzed. There were gaps that showed in the COVID-19 data at times that would then be completed at later dates. This indicated delayed reporting by healthcare entities. This delayed reporting impacted models in that results continued to change until the data became more stable. This is a challenge with any attempt to model real-time data.

6.3 Future Research

The applicability of this research on future research is vast. This research geographically focused on larger cities where there were large commercial real estate markets. Expanding the research scope to midsize and small cities could provide additional understanding of how those geographic areas commercial real estate markets have been impacted by COVID-19. Understanding how these dynamics play out could be critical for the businesses located in these areas as recovery may not be as fast as larger cities.

Enhancing these models to include other factors from the COVID-19 data may also provide additional results not seen in the research. Factoring in hospital bed capacity to handle any additional outbreaks could result in further refinement of the impacts to commercial real estate market in that location. A location with a high hospital bed capacity that can handle a widespread outbreak may have less impact on the commercial real estate market than a location with low bed capacity.

Another research area that the models in this research could be focused on would be the residential real estate market. This research focused solely on the commercial real estate market, leveraging the same modeling on residential real estate may provide some interesting results. Given that COVID-19 results in social distancing, understanding if there is a decrease in close quarters living populations, such as a decrease in apartment and condominium dwellers, would be an interesting research expansion. As work-from-home options continue to remain more permanent for companies and the need to be physically present in an office reduces is there an outflux of individuals from urban dwellings or in higher priced real estate markets to more suburban dwellings or lower priced real estate markets.

A final research area that the models in this research could be focused on would be the economic impacts COVID-19 has had on cities and their pre and post COVID-19 commercial real estate markets. A focused view of individual cities may provide additional guidance to state and local governments of how economic incentives leveraged prior to COVID-19 may need to change to continue to offer the same level of commercial real estate occupancy as in pre-COVID-19 times, assuming the market will ever recover to that level.

7 Conclusion

Despite the challenges and limitations outlined above, the research presented provides a key understanding of how COVID-19 has impacted areas of society, more specifically the commercial real estate market. Over the course of this research the impacts of COVID-19 on the commercial real estate market are becoming evident. The research presents that COVID-19, based on Hedonic modeling, results in an understatement of the true impact of the pandemic on commercial real estate. The Hedonic modeling showed no sustained difference in commercial real estate pricing in the markets evaluated with or without COVID-19 variables included. This indicates that Hedonic modeling is not the best approach for modeling these pandemic-based impacts on commercial real estate. A time dependent model should show more of the COVID-19 based impacts on commercial real estate when compared to the Hedonic model. As noted in previous research, a Difference-in-Difference model has proven well as another modeling type that can determine COVID-19 impacts.

The models developed over the course of this research provide a framework for understanding the implications of a pandemic on the scale of COVID-19 on the commercial real estate market. This research is useful to commercial real estate buyers and sellers in that it can assist in providing guidance in investment opportunities. These personas will be greatly impacted if they are unable to understand the impacts COVID-

19 has introduced into the market. Traditional approaches to commercial real estate may not be as applicable with COVID-19.

The research presented here is also useful to local governments. The impacts of empty commercial buildings are widely documented in real estate literature. Cities with a high number of empty commercial buildings may have a harder time recovering economically. Understanding how COVID-19 has impacted these commercial real estate offerings in cities allows state and local governments the ability to react and possibly decrease economic impacts to citizens.

The final impacts of COVID-19 on the commercial real estate market are still being played out. The models presented in this research may continue to be built upon as more is understood about how COVID-19 will continue to move through the population. Having models that focus on these pandemic events provide a framework for expanding understand during a pandemic as well as lay groundwork for future understanding if another pandemic is experienced.

Acknowledgments. Jacquelyn Cheun, PhD. – Capstone Professor

References

1. Banks, D., Albert, L., Caulkins, J., Frühwirth-Schnatter, S., Greig, F., Raftery, A., & Thomas, D. (2020). A Conversation about COVID-19 with Economists, Sociologists, Statisticians, and Operations Researchers. Harvard Data Science Review. Retrieved from https://hdsr.mitpress.mit.edu/pub/zp886sce
2. Lancaster, K. (2010). A new approach to consumer theory. IDEAS Working Paper Series from RePEc,
3. Ishijima, H., & Maeda, A. (2012a). Real estate price modeling and empirical analysis. International Journal of Economic Policy Studies, 7(1), 31-51. doi:10.1007/BF03405736
4. Mimis, A. (2016). 3d weight matrices in modeling real estate prices. The International Archives of the Photogrammetry, Remote Sensing and Spatial

Information Sciences, XLII-2/W2(2), 123-125. doi:10.5194/isprs-archives-XLII-2-W2-123-2016
5. Kok, N., Koponen, E., & Martínez-Barbosa, C. (2017). Big data in real estate? from manual apppraisal to automated valuation. Journal of Portfolio Management, 43(6), 202-211. doi:10.3905/jpm.2017.43.6.202
6. Maryam Heidari, & Carsten Felden. (2015). Financial footnote analysis: Developing a text mining approach. Proceedings of the International Conference on Data Mining (DMIN), , 10. Retrieved from https://search.proquest.com/docview/1705019804
7. Miles, M., Pringle, J., & Webb, B. (1989). Modeling the corporate real estate decision. The Journal of Real Estate Research, 4(3), 47.
8. Tezel, G., & Yalpir, S. (2011). Modeling of real-estate prices using artificial neural network (ann) approach. International Journal of Arts & Sciences, 4(15), 335-340.
9. José-Luis Alfaro-Navarro, Cano, E. L., Esteban Alfaro-Córtes, Noelia García, Matías Gámez, & Larraz, B. (2020). A fully automated adjustment of ensemble methods in machine learning for modeling complex real estate systems. Complexity, 2020 doi:10.1155/2020/5287263
10. Fu, Y., Xiong, H., Ge, Y., Zheng, Y., Yao, Z., & Zhou, Z. (2016). Modeling of geographic dependencies for real estate ranking. ACM Transactions on Knowledge Discovery from Data (TKDD), 11(1), 1-27. doi:10.1145/2934692
11. Morawska L, Cao J. Airborne transmission of SARS-CoV-2: The world should face the reality. Environ Int. 2020;139:105730. doi:10.1016/j.envint.2020.105730
12. McMullen, P.R. (2020). Social distancing via coulomb's law doi:10.4236/am.2020.117038
13. Lilley, A., Lilley, M., & Rinaldi, G.Public health interventions and economic growth: Revisiting the spanish flu evidence. SSRN Electronic Journal, doi:10.2139/ssrn.3590008Dheeriya, P. (2009). Modeling volatility in california real estate prices. IUP Journal of Applied Economics, 8(1), 26-38.
14. Francke, M., & Korevaar, M.Housing markets in a pandemic: Evidence from historical outbreaks. SSRN Electronic Journal, doi:10.2139/ssrn.3566909Kontsevaya, N. V. (2016). Modeling real estate market: Forecasting the price of a square. Statistika i Èkonomika, (4), 31-34. doi:10.21686/2500-3925-2016-4-31-34
15. Ambrus, A. (2020). Loss in the time of cholera.110(2), 475-525. Retrieved from http://www.econis.eu/PPNSET?PPN=1696251184Linné, M., & Cirincionne, J. (2010). Integrating geographic information and valuation modeling for real estate. The Appraisal Journal, 78(4), 370-378.
16. Ling, D. C., Wang, C., & Zhou, T.A first look at the impact of COVID19 on commercial real estate prices: Asset level evidence. SSRN Electronic Journal, doi:10.2139/ssrn.3593101Tezel, G., & Yalpir, S. (2011). Modeling of real-estate prices using artificial neural network (ann) approach. International Journal of Arts & Sciences, 4(15), 335-340.
17. Milcheva, S.Volatility and the cross-section of real estate equity returns during COVID-19. SSRN Electronic Journal, doi:10.2139/ssrn.3599211
18. Xie, L., & Milcheva, S.Proximity to COVID-19 cases and REIT equity returns. SSRN Electronic Journal, doi:10.2139/ssrn.3641268
19. Gerding, F., Martin, T., & Nagler, F.The value of fiscal capacity in the face of a rare disaster. SSRN Electronic Journal, doi:10.2139/ssrn.3572839

20. Baker, S. R., Bloom, N., Davis, S. J., Kost, K. J., Sammon, M. C., & Viratyosin, T. (2020). The unprecedented stock market impact of COVID-19. National Bureau of Economic Research Working Paper Series, No. 26945 doi:10.3386/w26945
21. Waller, B. D., Barrett, P. B., & Waller, L. (n.d.). Ethical Dilemmas Facing Today's Real Estate Professional. Academic and Business Research Institute. Retrieved August 9, 2020, from http://www.aabri.com/OC2010Manuscripts/OC10086.pdf
22. Angelopoulos, A. N., Pathak, R., Varma, R., & Jordan, M. I. (2020). On Identifying and Mitigating Bias in the Estimation of the COVID-19 Case Fatality Rate. Harvard Data Science Review. https://doi.org/10.1162/99608f92.f01ee285