Literature Review

How Has Economic Displacement Been Addressed in the Past?

(Zuk 2015)

A Review of Mass Appraisal Techniques

Much of the research on predicting real estate values has been in service of creating mass appraisal models. Mass appraisal models share many characteristics with predictive machine learning model modeling. Mass appraisal models are data-driven, standardized methods that employ statistical testing (Eckert 1990).

(Quintos 2013) Attempts to measure latent variables through a random effect regression model to predict income and expense of non-filers. Difference between the Assessed Value and the Market Value.

New York City annually values commercial properties by the income approach. Commercial properties with an assessed value greater than $40,000 are required to file income and expense statements with the Department of Finance. Some of these required filers may apply for exclusion from filing or they may choose not to file and instead pay a penalty. There are also voluntary filers, who are not required to file but nevertheless submit statements. The filings received are used to formulate income and expense regression models. These models are used to develop comparable rental models and to formulate assessment guidelines based on location and physical characteristics.

For models of income and expense, however, we are not aware of a model in a random effects (panel data) framework—most likely due to the lack of property-level data of income and expense filings.

(d'Amato 2017) Great Lit review in first chapter on the evolution of the Automated Valuation Model. Walks through all different kinds of spatial models: OLS, Heirarchichal, spatial lag, spatial error, etc. Explains COD (coeficient of dispersion). Dodd Frank Act implements financial regulatory reform after the financial crisis of 2008. In particular the title XIV subtitle F distinguishes appraisal process from automated valuation modelling, reorganizing both. In particular it was stressed how the role of valuation (appraisal) cannot be replaced by AVM. Our point of view is coherent with the Dodd Frank act (and Appraisal Methods and the Non-Agency Mortgage Crisis 29 thereby also Pugh's view but not Woodward's): automated valuation modelling is increasingly adaptable in describing real estate market behaviour without succeeding in replacing local information and human inspection in the valuation (appraisal) procedure.

(Koschinsky 2012) This is a recent and thorough discussion of parametric hedonic regression techniques. Some of the variables included are derived from nearby properties, similar to my technique, and these variables are found to be predictive. Methodology section (2) contains a brief but robust literature review of hedonic price modeling applied to real estate marginal willingness to pay (MWTP) for locational attributes. The basic hedonic model assumes that the utility of a household or an individual is a function of a composite good x; a vector of structural characteristics S; a vector of social and neighborhood characteristics N; and finally a vector of locational characteristics L. This study adds to a small body of existing literature that extends this research by addressing the valuation of a property's locational attributes from a spatial perspective.

In the model, for a spatial lag, they use a "We specify W as a queen contiguity weights matrix." The second set of locational attribute data represents a new way of measuring attributes of neighboring properties that is fully exogenous since it is derived from a different dataset than the sales data: It is based on structural characteristics of all residential properties built before 1997 that are not for sale but are within 1,000 feet of a 1997 sale. ... The variables included for neighboring properties within 1,000 feet of a sale are average age, poor condition (%), with electric heating source (strongly correlated with older age) (%), poor construction grade (1–5) (%), high construction grade (10–13) (%), and detached single-family homes (%). The spatial parameter lambda is positive and significant in all cases, i.e. the relation between a home's price and the average price of its neighboring homes is characterized by positive spatial autocorrelation where, for instance, high-price homes are surrounded by houses with high prices

In short, for the data in this study locational characteristics are valued at least as much as (if not more) than important structural characteristics.

In this case the correct welfare measure should be the direct effect since there is a strong argument in the literature (e.g. Pace and Gilley 1998) that spatial autocorrelation in house prices is related to the practice of realtors, appraisers and home owners of using nearby comparable sales to determine the sales price of a property. Therefore it is to be expected that a house which is in a neighborhood where the sales price of recently sold houses is high will be higher than a similar house surrounded by houses recently sold at a low price. This will lead to spatial autocorrelation in house prices, but the origin for such autocorrelation is a pecuniary externality

(Fotheringham 2015) Explores the use of GWR to forecast prices.Explores the combination of time-series forecasting (in the Holt-Winters tradition) to geogaphically weighted regression (GWR). GWR is a variation on OLS that allows for "adaptive bandwidths" of local data to be included, i.e., for each estimate, the number of data points included varies (optimized using CV). In addition, the data points are weighted according to distance. This is known as a "local" model

Has Machine Learning Been Applied to this Problem Before?

(Schernthanner H. 2016) Paper compares traditional linear regression techniques to more advnaced techniques such as krigging (stochastic interpolation) and random forrest; finds that more advanced techniques are sound and more accurate. The research findings indicate that the analysis results achieved by any of the new methods, ranging from stochastic interpolation to the "random forest" method of machine learning, are more valid than results obtained from traditional statistical methods

(Guan et al. 2014) Uses three different approaches to defining comps, all using euclidean distance; a radius technique, a k-nearest neighbors technique using only distance and a k-nearest neighbors technique using all attributes. Interestingly, the location-only KNN neighborhood performed best, although by a very slim margin (potentially meaningless). The MRA [Multiple Regression Analysis] method, although widely used in mass appraisal, has been criticized for its inability to model data features typically found in real estate data. Common problems with MRA assessment of real estate properties are well known and they include nonlinearity, multicollinearity, and heteroscedasticity (Antipov and Pokryshevskaya 2012; Kilpatrick 2011; Mark and Goldberg 1988; Peterson and Flanagan 2009). In recent years, data mining methods have been proposed as an alternative, and have been tested with very mixed results.

(Fu 2014) Prediction model for real estate in Beijin, China. They do a clustering, then do a rank-ordered prediction of investment returns segmented into categories: 4>3>2>1>0

While a number of estate appraisal methods have been devel- oped to value real property, the performances of these meth- ods have been limited by the traditional data sources for es- tate appraisal

the geographic dependencies of the value of an estate can be from the characteristics of its own neighborhood (individual), the values of its nearby estates (peer), and the prosperity of the affiliated latent business area (zone)

ClusRanking is able to exploit geographic individ- ual, peer, and zone dependencies in a probabilistic ranking model. Specifically, we first extract the geographic utility of estates from geography data, estimate the neighborhood popularity of estates by mining taxicab trajectory data, and model the influence of latent business areas via ClusRank- ing.

From related works: Recent works [8, 21] study the automated valuation models, which aggregate and analyze physical characteristics and sales prices of comparable properties to provide property valuations

(Rafiei 2016) Fascinating paper which employs a Restricted Boltzmann Machine (neural network with back propagation) to predicted the sale price of residential condos in Tehran, Iran. The paper focuses on computational efficiency. A non-mating genetic algorithm is used for dimensionality reduction. The paper concludes that two primary strategies help in this regard: Sales which happened closer in time to a prediction are more important, and it also uses a learner to accelerate the recognition of important features. The paper compares this technique to several other common NN approaches and finds that while not necessarily the only way to get the best answer, it is definitely the fastest way to get to the best answer. The lit review sections walks through several recent and notable papers specifically on the topic of sales price prediction of real estate. There is also mention of a paper which characterizes a real estate market as supply inelastic which may be worth investigating further.

(Helbich 2013) This is a very recent paper which contains a brief but robust literature review in the introduction. Great quote: hedonic pricing models "can be improved in two ways: (a) Through novel estimation techniques (e.g. Brunauer et al., 2010; Koschinsky, Lozano-Gracia, & Piras, 2011) and (b) by ancillary structural, locational, and neighborhood variables on the basis of Geographic Information System (GIS) algorithms (e.g. Hamilton & Morgan, 2010)"

Let's follow up on the sources mentioned. I believe my micro-neighborhood technique falls into the "unique estimation" bucket, so it would be wise to position it that way

(Kontrimasa 2011) Mass appraisal is commonly used to compute real estate tax. Study uses an n = 100 (very small) and compares accuracy of linear regression vs other ANN techniques like SVM.

(Dietzell 2014) This paper examines internet search query data provided by "Google Trends", with respect to its ability to serve as a sentiment indicator and improve commercial real estate forecasting models for transactions and price indices. The empirical results show that all models augmented with Google data, combining both macro and search data, significantly outperform baseline models which abandon internet search data

(Gary and D. 2011) Examines the effects of walkability on property values and investment returns. Use data from the National Council of Real Estate Investment Fiduciaries and Walk Score to examine the effects of walkability on the market value and investment returns of more than 4,200 office, apartment, retail and industrial properties from 2001 to 2008 in the United States. On a 100-point scale, a 10-point increase in walkability increased values by 1–9%, depending on property type. We also found that walkability was associated with lower cap rates and higher incomes, suggesting it has been favored in both the capital asset and building space markets

(Park 2015) Machine learning applied to residential real estate price prediction. Developed a housing price prediction model based on machine learning algorithms such as C4.5, RIPPER, Naïve Bayesian, and AdaBoost and compare their classification accuracy performance. The experiments demonstrate that the RIPPER algorithm, based on accuracy, consistently outperforms the other models in the performance of housing price prediction.

sample citations

Sample Citation: (Antipov and Pokryshevskaya 2012) (see: Antipov and Pokryshevskaya 2012, 33–35; also Antipov and Pokryshevskaya 2012, ch. 1 and *passim*)

A minus sign (-) before the @ will suppress mention of the author in the citation. This can be useful when the author is already mentioned in the text:

Antipov says blah (2012).

You can also write an in-text citation, as follows:

Antipov and Pokryshevskaya (2012) says blah.

Antipov, Evgeny A., and Elena B. Pokryshevskaya. 2012. "Mass Appraisal of Residential Apartments: An Application of Random Forest for Valuation and a Cart-Based Approach for Model Diagnostics." *Expert*

Systems with Applications.

Dietzell, Nicole; Schäfers, Marian Alexander; Braun. 2014. "Sentiment-Based Commercial Real Estate Forecasting with Google Search Volume Data." *Journal of Property Investment & Finance,* 32 (6): 540–69.

d'Amato, Tom, Maurizio; Kauko, ed. 2017. *Advances in Automated Valuation Modeling.* Springer International Publishing.

Eckert, J. K. 1990. *Property Appraisal and Assessment Administration.* Chicago, IL.: International Association of Assessing Officers.

Fotheringham, R; Yao, A.S.; Crespo. 2015. "Exploring, Modelling and Predicting Spatiotemporal Variations in House Prices." *The Annals of Regional Science* 54.

Fu, Yanjie; et al. 2014. *Exploiting Geographic Dependencies for Real Estate Appraisal: A Mutual Perspective of Ranking and Clustering.* Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery; data mining.

Gary, Pivo, and Fisher Jeffrey D. 2011. "The Walkability Premium in Commercial Real Estate Investments." *Real Estate Economics* 39 (2): 185–219. doi:10.1111/j.1540-6229.2010.00296.x.

Guan, Jian, Donghui Shi, Jozef M. Zurada, and Alan S. Levitan. 2014. "Analyzing Massive Data Sets: An Adaptive Fuzzy Neural Approach for Prediction, with a Real Estate Illustration." *Journal of Organizational Computing and Electronic Commerce* 24 (1). Taylor & Francis: 94–112. doi:10.1080/10919392.2014.866505.

Helbich, et al., Marco. 2013. "Boosting the Predictive Accuracy of Urban Hedonic House Price Models Through Airborne Laser Scanning." *Computers, Environment and Urban Systems* 39: 81–92.

Kontrimasa, Antanas, Vilius; Verikasb. 2011. "The Mass Appraisal of the Real Estate by Computational Intelligence." *Applied Soft Computing.*

Koschinsky, J. et al. 2012. "The Welfare Benefit of a Home's Location: An Empirical Comparison of Spatial and Non-Spatial Model Estimates." *Journal of Geographical Systems* 10109.

Park, Jae Kwon, Byeonghwa; Bae. 2015. "Using Machine Learning Algorithms for Housing Price Prediction: The Case of Fairfax County, Virginia Housing Data." *Expert Systems with Applications* 42 (6): 2928–34.

Quintos, Carmela. 2013. "Estimating Latent Effects in Commercial Property Models." *Journal of Property Tax Assessment & Administration* 12 (2).

Rafiei, Hojjat, Mohammad Hossein; Adeli. 2016. "A Novel Machine Learning Model for Estimation of Sale Prices of Real Estate Units." *Journal of Construction Engineering and Management* 142 (2).

Schernthanner H., Gonschorek J., Asche H. 2016. "Spatial Modeling and Geovisualization of Rental Prices for Real Estate Portals." *Computational Science and Its Applications* 9788.

Zuk, Miriam; et al. 2015. "Gentrification, Displacement and the Role of Public Investment: A Literature Review."