

PROCEEDINGS OF SPIE

SPIDigitalLibrary.org/conference-proceedings-of-spie

Proposal to optimize the results of the new general valuation, application on residential plots in Lakatamia Municipality

Tereza Kyprianou, Thomas Dimopoulos

Tereza Kyprianou, Thomas Dimopoulos, "Proposal to optimize the results of the new general valuation, application on residential plots in Lakatamia Municipality," Proc. SPIE 12786, Ninth International Conference on Remote Sensing and Geoinformation of the Environment (RSCy2023), 127860J (21 September 2023); doi: 10.1117/12.2682036

SPIE.

Event: Ninth International Conference on Remote Sensing and Geoinformation of the Environment (RSCy2023), 2023, Ayia Napa, Cyprus

Proposal to optimize the results of the new general valuation, application on residential plots in Lakatamia Municipality

Tereza Kyprianou^a, Dr Thomas Dimopoulos^a

^a School of Architecture, Land & Environmental Sciences, Department of Real Estate, Neapolis University Pafos, 2 Danais Avenue, 8042 Pafos, Cyprus

ABSTRACT

Property valuation evolved from simple empirical judgements to automated valuation models and their application have extended from single property to mass valuation. Many governments across the world have used AVMs to get a valuation in thousands of properties for tax related purposes. The literature review is extensive and it is growing day by day. The island of Cyprus was introduced to computer assistant mass appraisal (CAMA) in 2013 when the Department of Land and Surveys (DLS) performed a general valuation and then to revaluation in 2018.

The aim of this research is to provide more transparency to the reliability of the data used in the latest general valuation. An automated valuation model was developed, using the MRA method and Hedonic Pricing Model, to test the performance of the data and compare them with the minimum standards a valuation model should have according to the International Association of Assessing Officers (IAAO).

A case study using a holdout sample with data from Lakatamia Municipality was created to observe the reliability of the data but also to improve the accuracy of the Automatic Valuation Model. Three regressions were carried out: a) Basic regression with 503 observations and 10 variables, b) Regression with the previous variables plus 10 nearest neighbors as predictors and c) Regression with the previous variables plus 10 nearest neighbors as predictors, with 450 best observations – deleted outliers based on absolute error. The coefficient of determination (R-squared) measures the goodness of fit of the regression line, in other words, how close the data are to the estimated line. Initially the R-squared was 0.319 which is above IAAO standards but it was increased to 0.765 after the application of the third model. This accuracy showing better performance than the mass valuation system applied by the Department of Land and Surveys in Cyprus with accuracy of 0.384

Concluding the research ends with a critical discussion about the reliability of the data and some suggestions that could be applied by the DLS to improve the performance of the data. It is worth mentioning that the Cypriot data have a limitation due to the high heterogeneity found between properties.

Keywords: AVM, CAMA, Cyprus – General Valuation, Regression Analysis, Land Valuation

1. INTRODUCTION

The greatest proportion of a person's or a county's wealth, in most cases, consists of real estate. However, real estate values fluctuate over time and follow trends and cycles. Considering that significant events, such as the current COVID-19 pandemic or the global financial crisis of 2008, can affect property values significantly, and have the potential to completely change their previous status, the need for the revaluation of all real estate assets is generated. As a task, this would be very difficult, if not almost impossible bearing in mind time, cost and other constraints. Therefore, the development of accurate Mass Appraisal systems is very important and crucial, and their need is especially apparent in current circumstances (Dimopoulos, 2020).

The question appears as to how an applicable, consistent and efficient mass property appraisal model could be formed, in which spatial (locational) attributes can be stored and assessed, taking into consideration the range and variation of all valued characteristics. For a long time, the accurate assessment of property values was and still presented as the great issue. Appraisers and assessors are able to estimate property values on their accumulated knowledge. Nonetheless, the main challenges are, first of all, the consistency and accuracy of the valuation created, referring to the weights provided by the appraiser to define an appraised value, and secondly, the time required for the valuation to take place. Apart from

the property owners and real estate agents, accurate property valuations are also important for local governments whom impose taxes on properties according to their values.

Income that is generated from real property tax is a strong and considerable source for local authorities and the government. The property taxes are an important part of the gross domestic product and in any case, there is a proven evidence of how they contribute. Cyprus is using an 'ad volarem' property taxation system on market values and the first computer assisted mass valuation system (CAMAS) was developed in 2013 to assist the Department of Lands and Surveys to implement general valuation on a frequent basis, as and when they were ordered by the Council of Ministers. Due to the economic crisis and the bail in, the Cyprus government had to sign a memorandum of understanding (MOU) with EU/ECB/IMF in order to be able to get financial assistance, under measure 3.8, and a new General Valuation was ordered by the Council of Ministers as one of the many measures that had to be met.

This study hypothesis that the property mass appraisal systems used by the Department of Land and Surveys in Cyprus to impose property taxes are inaccurate and unreliable. For that purpose, this study tries to compare the results of the prediction model that was developed against the results of General Valuation 1.1.18 used by the Department of Land and Surveys in Cyprus. Finally, provides a critical discussion about the DSL data limitations and the AVM limitations.

2. LITERATURE REVIEW

2.1 Evolution of Mass Appraisal

The origin of mass appraisal dates back to the formal development of hedonics in the early seventies. The first published hedonic study was a master thesis on agricultural land values in 1922 at University of Minnesota as argued by Colwell and Dilmore (Colwell, 1999). Its introduction as an empirical estimation method though can be traced back to Court.

The National Association of Tax Assessing Officers (now the International Association of Accessing Officers [IAAO]) was formed in 1934 for the purpose of improving mass appraisal through education, research and the promotion of uniform appraisal practices. An early, landmark, accomplishment of the association was the publication in 1938 of *Assessment Principles*, which was reprinted nine times, most recently in 1948. This report contains a classic description of the assessment process and sets forth eighty specific recommendations for achieving accuracy and uniformity. The most important recommendation was that all real and personal property be subject to annual revaluation. The report explains that the advantages of annual revaluation, aside from legal requirements, include greater uniformity, better use of resources, and the avoidance of large assessment shifts during revaluations (Eckert, 1990).

Computers were first used in assessment administration in the 1950s, primarily in large jurisdictions that could afford the expense. Second-generation computers, introduced in the 1960s, were responsible for the continued growth of the computer industry. During this time, many jurisdictions began using computers to store property record data and to automate the cost approach. California began to use multiple regression analysis in 1966 (Eckert, 1990).

In 1970, Computerized Assisted Assessments term emerged also from land valuation (Gwartney, 1970). Few years later the concept is introduced also in real estate property valuation (Carbone & Longing, 1977) by the term Model for Automated Assessment almost identical to the used now. During this time, the range of commercially available statistical and mass appraisal software also increased greatly. Many jurisdictions began using multiple regression analysis or other automated applications of the sales comparison and income approaches to value (Eckert, 1990).

Since mass valuation is greatly influenced by statistical models and computer technologies, specifically design systems, using advanced techniques in software and hardware technologies, have started to developed in the 1980's. These Computer Assisted Mass Appraisal systems, can assist and simplify valuation procedures, reducing time and effort required to analyze large numbers of input data in determine land valuation (Labropoulos, Dimopoulou, & Zentelis, 2003). Namely, in some instances the term computer-assisted mass appraisal (CAMA) is used instead of the term AVM. It is observed that these two terms are not synonyms; Computer-assisted mass appraisal is normally applied to estimate using statistical modelling a large number of properties. Normally for this issue are used big quantity of data and the final result is normally used for tax purposes at a specific date. On the other side Automated Valuation Methodology using often the same methodological framework of mass appraisal, uses a statistical model and a large amount of property data to estimate the market value of an individual property or portfolio of properties. This valuation is normally delivered together with a

confidence level to indicate how accurate the valuation is and they are provided essentially for lending purposes. In this case the data is not fixed and coincides with the appraisal date for mortgage lending purposes (Amato, 2017).

2.2 Application of Mass Appraisal in Cyprus - General Valuation 01.01.2018

There was a radical change in the methodology of Data Capture of property characteristics for updating the valuation base due to the development of the DLS Portal (Web Application) that electronically interconnects all building authorities issuing building permits (Municipal Authorities, District Administrations). This application enables automatic updating of the Department’s Database, both for new developments as well as for amendments to existing ones. Regarding the valuation approach, the same base models were applied to value all types of properties as in 2013, which summarized below:

Land Models	Unit Characteristics
<ul style="list-style-type: none"> • Base Residential / Commercial building site • Base Industrial building site • Base Undeveloped field • Base Agricultural/Livestock fields 	<ul style="list-style-type: none"> • Base Residential • Base Commercial • Base Industrial • Base Hotels / Tourist Establishments • Base Schools/Hospitals/Clinics • Base Livestock Units

The 2018 G.V. was refined and improved compared to the 2013 G.V. by adding more property characteristics such as the ‘slope’, ‘view’, ‘subsoil’, ‘High Voltage Tension Lines’, ‘Foreshore Protection Zone’. The following property characteristics were used for the 2013 mass appraisal:

Land Characteristics	Unit Characteristics
<ul style="list-style-type: none"> • Type of Property • Area (sq.m.) • Planning Zone • Location • Kind of access • Relation to the road • Shape 	<ul style="list-style-type: none"> • Type of Unit • Area of Unit (sq.m.) • Year of Construction (Depr. Factor) • Year of substantial renovation • Category (luxury, A, B, C, and D) • Condition (V, good, fair, bad) • View

Furthermore, it became possible to perform automated valuations for parcels that lie within more than one planning zone. In addition, the valuation models were upgraded to be able to perform quantity adjustments for land extent. Also, it became possible to create more planning zone categories under the main ones, thus achieving a much higher degree of accuracy as these parameters were used to adjust the base value (constant) of the properties, according the property characteristics of each.

Another profound change of the 2018 G.V. was the participation of private valuers in the process of determining the land valuation parameters for all municipal administrations on the island. This was made possible through public tendering and the external valuers with their market knowledge and expertise worked together with the respective valuers employed in the Department. The private sector valuers also assisted the Department in tackling specific issues and market inconsistencies that do not follow the usual patterns of the property market.

It was estimated that the total number of properties valued for the 2018 G.V. were about 1.5 million land parcels and 525,000 buildings (458k residential, 67k commercial). The value of all properties for 2018 was estimate at €178,779,262,000. Quality control (ration studied-Median, COD, PRD) tests we are performed regarding the uniformity

of values as well as in order to ensure that horizontal and vertical equity is achieved to meet the compliance of IAAO standards, before the publication of results.

Table 1: Ratio Studies - General Valuation 01.01.2018 / Sales Price

District	Property Type	No. of observations	Average Ratio	Median Ration IAAO standard $0.9 \leq R \leq 1.1$.	Coefficient of Dispersion COD IAAO Standard $COD \leq 30$
Nicosia	Land	3745	1.08	1.04	0.29
Nicosia	Units	3321	0.92	0.91	0.16
Larnaca	Land	1871	1.06	1.01	0.29
Larnaca	Units	3473	0.99	0.97	0.20
Famagusta	Land	811	0.96	0.91	0.32
Famagusta	Units	1059	0.86	0.85	0.19
Limassol	Land	3065	1.00	0.94	0.37
Limassol	Units	5478	0.90	0.84	0.28
Pafos	Land	1045	0.96	0.95	0.31
Pafos	Units	4927	0.97	0.94	0.22

It is worth noting that on 22 valuation objections were lodged over the period of six months period under the current legislation. Regarding the correction of error applications, there were also very low, approximately 200. The low number was due to the majority of property information being corrected during the process of the previous G. V. It is important to note that the lodging of both applications was possible only electronically through the DLS Portal. The burden of cost (time and administrations) was minimized for both citizens and the Department. All the values of the general valuations are freely available to be viewed on the DLS Portal of the Department as well as their respective property characteristics (Varnavas, Charalambous, & Savvidou, 2015).

3. DATA

3.1 Data used in the research and data cleansing

The data used in this research come from the DLS of Cyprus and represent past sales of properties in Lakatamia Municipality from January 2012 to December 2019. The data had information about houses, apartments, shops and many more but research was going to focus only on plots and thus all non-plot data were removed.

One of the most significant information provided by the DLS of Cyprus are the transactional prices, known as the ‘declared’ and ‘accepted’ price. The declared price is what is actually declared between the buyer and the seller. The accepted price is what the Land Registry believes the sale price should have been based on their own valuations. Sometimes both are the same, however there are many occasions where the Land Registry may take a view that the declared price was very low. The reason for starting this accepted price is to combat fraud and off-the-record payments, stop tax evasion and account for any distressed sales. It was decided that the study would focus on the accepted prices only, which are also commonly used as comparable evidence by property valuers.

In addition, data that falls within non residential zone, or those that had values per sq.m. over €700 were removed. In many cases, properties that registered as plots, on spot may include a building, so for those properties that has values over €700/sq.m. we assume that they do not concern to a vacant land.

Another issue with the transaction data was the fact that since 2016, there have been numerous debt-to-asset (D2A) transactions which have resulted from the Bank’s seizing of properties because of non-performing loans (NPLs). These are also registered as sale in the Land Registry’s sale database, but are flagged accordingly in order to be identified by the readers of the datasets. The majority of these sales were carried out at “Forced Sale Values” estimated at 70% of the Market Value, however there is no definite way to know circumstances of each and every sale and therefore that transactions were removed. The final sample was 503 transactions.

3.2 Accepted price adjustments

The data now were only for residential plots within Lakatamia with transactional prices from 2012 to 2019. This raised an issue of reliability because the model needed to have prices with the same date to compensate from price fluctuations over time.

In Cyprus there are two authorities that publish real estate property indexes and these are the RICS and the Central Bank of Cyprus (CBC). Neither index has data for vacant land, however, because these are residential plots that are going to be developed mainly in houses, we assumed that the trend would be similar to the housing index. The Property Price Index from the Central Bank of Cyprus was used to bring all prices to the date 01.01.2018 because it was deemed as more reliable source mainly because the index is based on actual data and not hypothetical. The property price index used can be seen follow. The accepted prices from DLS were bought forward based on their data. If an accepted price was in 2013, it as multiplied by the price change from 2013 until 2018.

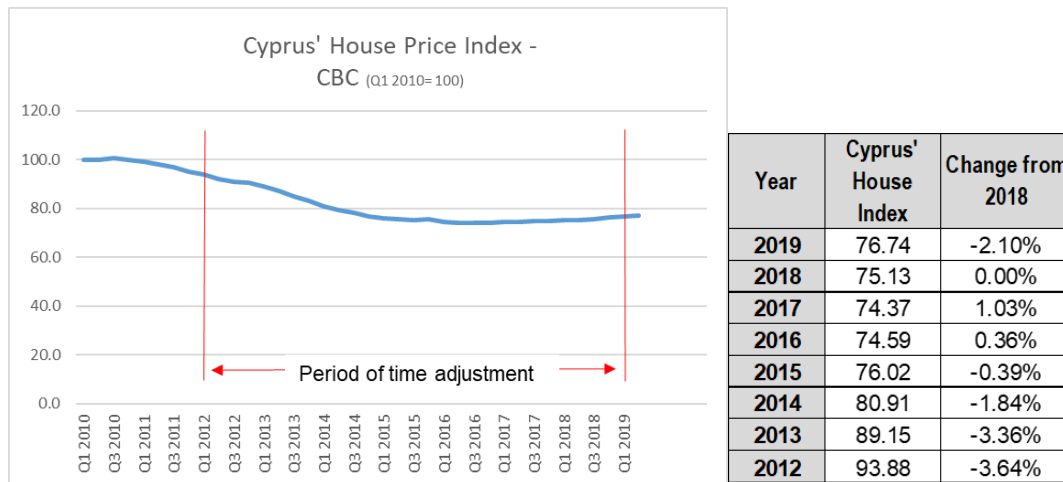


Figure 1. Cyprus’ House Price Index, CBC

3.3 Converting the arithmetic degree to geographic

The rotation for each plot was necessary as it was going to be one of the independent variables. The ArcGIS extract the rotation in an arithmetic degree. Then, through the attributes on Microsoft Excel all the arithmetic degrees convert into geographic as north, north-east, east, south-east, etc.

4. RESEARCH METHODOLOGY (REGRESSION ANALYSIS AND RESULTS)

Initially, simple regression was used, the results of which were unsatisfactory and then a nearest neighbor algorithm incorporating enhanced distance metrics. Finally, the application involves a prediction of the dependent variable, i.e. predicted value 1.1.18 in order to compare with the Cyprus’ General Valuation 1.1.18 and the Accepted Prices from the Cyprus DLS.

After analyzing and processing all the data, the final result was an excel file which include the coordinates of centroid (X,Y) for each observation, the share, the land size, the frontage, and the rotation. This excel file was imported into the Python software where through code was performed the regressions.

4.1 1st Model: Basic Regression

The adjusted accepted property prices and the variables will be used in a regression analysis. The analysis will return with regression statistics and coefficients such as the intercept (α) and the β for each variable. The coordinates of the centroid (X, Y), the land size, the share, the frontage and the rotation are the variables that are used in the Hedonic Price Model to estimate the price. The general equation that will be used to estimate the price is the additive. By cleansing data from the dataset, 503 observations were finally registered in order to carry out the first regression analysis ‘Model 1’, the basic regression. The basic regression registered an R^2 and an adjusted R^2 of 0.319 and 0.306 respectively. The Prob(F-statistic) tells the overall significance of the regression. This is to assess the significance level of all the variables together unlike the t-statistic that measures it for individual variables. The null hypothesis under this is “all the regression coefficients are equal to zero”. Prob(F-statistics) depicts the probability of null hypothesis being true. As per the below results, probability is close to zero. This implies that overall the regressions is meaningful.

Prior to carrying out any additional analysis, it was important to carry out checks for high correlations between the different variables that ideally should be completely independent. This phenomenon is known as ‘multicollinearity’ and it occurs when the various independent variables in the regression model are correlated amongst themselves. The greater amount of variables used in a regression typically results in a higher chance of multicollinearity occurring. However, collinearity statistics were also identified in the below table as well (P). The regression output below shows that the most variables are statistically significant because their p-values equal 0.000 or less than 0.05. On the other hand, share and frontage are not statistically significant because its p-value (0.069 and 0.842) is greater than the usual significance level of 0.05. We believe that this is not true, because actually the frontage of a plot is an important factor.

Table 2: OLS Regression Results of Model 1

Coefficients							
Model 1		Coef.	Std. Error	t	P > t	[0.025	0.975]
(Const)		233.1648	10.626	21.942	0.000	212.286	254.043
X ₁	X (°)	0.0210	0.002	10.261	0.000	0.017	0.025
X ₂	Y (°)	0.0195	0.002	12.355	0.000	0.016	0.023
X ₃	Share	19.9638	10.952	1.823	0.069	-1.554	41.481
X ₄	Land size (m ²)	-0.0173	0.008	-2.076	0.038	-0.034	-0.001
X ₅	Frontage (m)	0.0393	0.197	0.199	0.842	-0.349	0.427
X ₆	E (1 or 0)	55.9211	11.362	4.922	0.000	33.597	78.245
X ₇	S-E (1 or 0)	36.3540	7.333	4.957	0.000	21.945	50.763
X ₈	N-W (1 or 0)	40.8039	5.376	7.590	0.000	30.241	51.367
X ₉	N-E (1 or 0)	48.0813	4.812	9.992	0.000	38.626	57.536
X ₁₀	N (1 or 0)	52.0046	11.713	4.440	0.000	28.991	75.018

4.2 2nd Model: Regression with the previous variables plus 10 nearest neighbors as predictors.

In the second model we keep the same variables as the first model and add weights to the ten nearest neighbors. The k nearest neighbor algorithm is one in a family of distance-based algorithms. The basic model consists of a number, k, representing the number of neighbors to be retrieved, a set of comparables, a distance metric, weights for individual comparables and a prediction technique once the nearest neighbors have been retrieved. Given a subject property, whose

price is to be predicted, its distance from each comparable is calculated using the distance metric. The k comparables with the smallest distance are retrieved and used to predict the price of the target property using a prediction technique. The following part of the dissertation will explore the effect of these various aspects of the k-NN on its efficacy within the mass appraisal domain.

The value of k is the number of neighbors or closest comparables to the subject property that will be retrieved and on which the prediction is based. As the k value increases, the predicted value of the subject property is based on a larger number of properties and would therefore be more robust to noise within the comparable set, for example, non-arm's length sales, special purchasers etc. However, taking a very large value of k could have the effect of over generalizing and therefore, the price of the subject property would be based on properties that could be irrelevant to the pricing of the of the subject. The ideal judge of which comparables should be taken into account would be the expert (valuer); however, given the requirements of mass appraisal in terms of the large number of valuations that need to be undertaken in a short period of time, this would be an almost impossible task. Observing the figure below, which shows all the comparables, we judged that the ten closest comparables (k=10) are satisfactory number without generalizing (McCluskey & Anand, 1999).

Distance is the way mathematicians determine which points are most similar in a n-dimensional space. The intuition is that the smaller the distance between the points the more similar they are. In the present study, is used to calculating distance in a 2-dimensional space, the x, y coordinates of the centroid of each observation. There are several ways to calculate distance but we are going to use the Euclidean distance. The formula works by creating a right triangle between two points and determining the length of the hypotenuse.

$$d = \sqrt{(X1 - X2)^2 + (Y1 - Y2)^2}$$

Equation 1: Euclidean distance

Ten comparables have been extracted from the data set based on their distance from the subject property. The objective being to extract those comparables having the lowest “distance” value. As similarity between the subject and the comparable decreases the distance figures increases. For example, as Comprable 1 is the most similar to the subject in terms of distance, its reasonable to expect it to have the highest influence on value and consequently the highest comparable weigh.

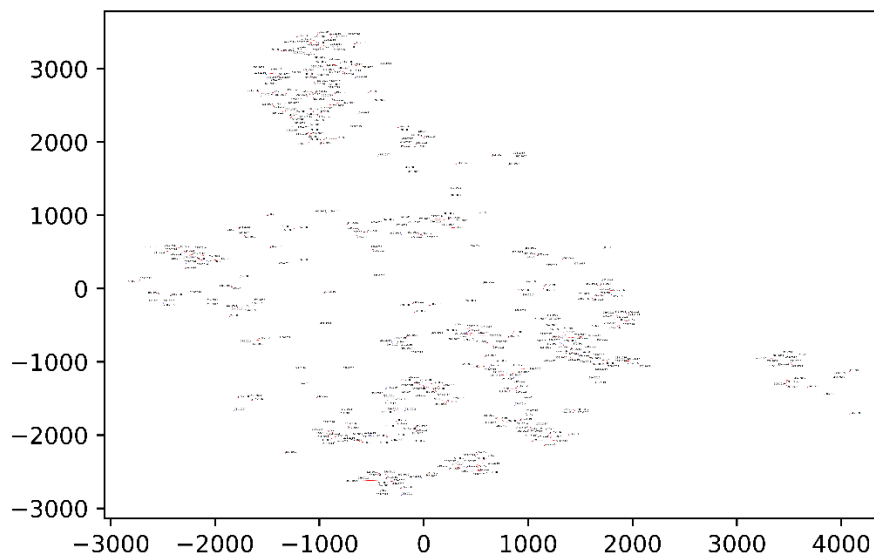


Figure 2: Illustration of the comparative sales

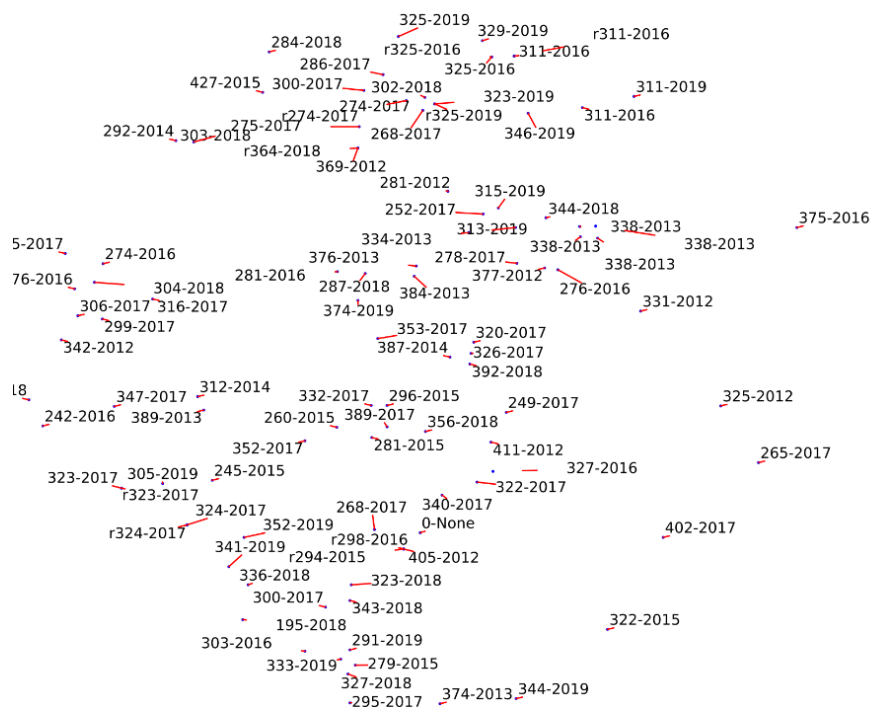


Figure 3: Zoom of the previous figure

The new model comprised of 484 observations (19 observations with a frontage equal to zero were removed) and 20 variables. The first 10 parameters are the same as the basic regression while the additional 10 refer to the prices of the ten closest comparables. ‘Model 2’ registered an R^2 and an Adjusted R^2 of 0.559 and 0.541 respectively. This is approximately 0.24 greater than the ‘Model 1’ (the basic regression). As per the below summary results, the Prob(F-statistics) is close to zero. This implies that overall the regressions is meaningful.

Table 3: OLS Regression Results of Model 2

Coefficients						
Model 1	Coef.	Std. Error	t	P > t	[0.025	0.975]
(Const)	52.6752	18.132	2.905	0.004	17.045	88.306
X ₁ X (°)	0.0031	0.002	1.419	0.156	-0.001	0.007
X ₂ Y (°)	0.0027	0.002	1.454	0.147	-0.001	0.006
X ₃ Share	-3.8879	10.929	-0.356	0.722	-25.364	17.588
X ₄ Land size (m ²)	-0.0218	0.007	-3.130	0.002	-0.035	-0.008
X ₅ Frontage (m)	0.0356	0.179	0.199	0.842	-0.315	0.387
X ₆ E (1 or 0)	18.3404	10.040	1.827	0.068	-1.389	38.070
X ₇ S-E (1 or 0)	-3.5862	7.193	-0.499	0.618	-17.721	10.549
X ₈ N-W (1 or 0)	9.8971	5.551	1.783	0.075	-1.012	20.806
X ₉ N-E (1 or 0)	9.6716	5.019	1.927	0.055	-0.191	19.534
X ₁₀ N (1 or 0)	18.3522	10.102	1.817	0.070	-1.499	38.204

X ₁₁	Adj. Acc. Price of the 1 st nearest	0.3954	0.040	9.786	0.000	0.316	0.475
X ₁₂	Adj. Acc. Price of the 2 nd nearest	-0.0457	0.049	-0.923	0.356	-0.143	0.052
X ₁₃	Adj. Acc. Price of the 3 rd nearest	0.1115	0.044	2.535	0.012	0.025	0.198
X ₁₄	Adj. Acc. Price of the 4 th nearest	0.0902	0.045	2.023	0.044	0.003	0.178
X ₁₅	Adj. Acc. Price of the 5 th nearest	-0.0355	0.047	-0.761	0.447	-0.127	0.056
X ₁₆	Adj. Acc. Price of the 6 th nearest	0.0539	0.048	1.122	0.263	-0.041	0.148
X ₁₇	Adj. Acc. Price of the 7 th nearest	0.1285	0.049	2.643	0.008	0.033	0.224
X ₁₈	Adj. Acc. Price of the 8 th nearest	0.0445	0.050	0.882	0.378	-0.055	0.144
X ₁₉	Adj. Acc. Price of the 9 th nearest	0.0345	0.048	0.724	0.469	-0.059	0.128
X ₂₀	Adj. Acc. Price of the 10 th nearest	0.0591	0.048	1.226	0.221	-0.036	0.154

4.3 3rd Model: Regression with the previous variables plus 10 nearest neighbors as predictors, with 450 best observations – deleted outliers based on absolute error.

The closer the R² is to 1 means that the independent variables explain the changes in the dependent variable. To be able to do that had to remove some outliers. According to the IAAO, 2018 trimming outliers may be acceptable and therefore we deleted outliers based on absolute error. We kept the 450 best observations with the results as follows: ‘Model 3’ registered an R² and an Adjusted R² of 0.765 and 0.755 respectively which is greater than the ‘Model 1’ and ‘Model 2’. Also, the coefficient table is slightly better off in ‘Model 3’ compared to that in ‘Model 2’ in terms of its collinearity statistics (p-values). In the last ten variables it is logical to have multicollinearity since it concerns the value of the nearest comparables that may have the same characteristics, or maybe is a resale of the subject property, see example in the following figure. Something that trouble is that the p-value of some variables still more than 0.05.

Table 4: OLS Regression Results of Model 3

Coefficients							
Model 1		Coef.	Std. Error	t	P > t	[0.025	0.975]
(Const)		58.3328	11.799	4.944	0.000	35.141	81.525
X ₁	X (°)	0.0022	0.001	1.637	0.102	-0.000	0.005
X ₂	Y (°)	0.0034	0.001	2.861	0.004	0.001	0.006
X ₃	Share	-14.1359	6.830	-2.070	0.039	-27.560	-0.712
X ₄	Land size (m ²)	-0.0226	0.004	-5.277	0.000	-0.031	-0.014
X ₅	Frontage (m)	-0.0983	0.115	-0.855	0.393	-0.325	0.128
X ₆	E (1 or 0)	21.6574	6.665	3.249	0.001	8.556	34.758
X ₇	S-E (1 or 0)	-3.8698	4.597	-0.842	0.400	-12.906	5.167

X ₈	N-W (1 or 0)	9.4304	3.522	2.677	0.008	2.508	16.353
X ₉	N-E (1 or 0)	4.2950	3.224	1.332	0.184	-2.043	10.633
X ₁₀	N (1 or 0)	26.8198	6.404	4.188	0.000	14.232	39.408
X ₁₁	Adj. Acc. Price of the 1 st nearest	0.3784	0.034	11.048	0.000	0.311	0.446
X ₁₂	Adj. Acc. Price of the 2 nd nearest	0.0192	0.034	0.566	0.571	-0.047	0.086
X ₁₃	Adj. Acc. Price of the 3 rd nearest	0.1379	0.029	4.782	0.000	0.081	0.195
X ₁₄	Adj. Acc. Price of the 4 th nearest	0.0721	0.028	2.563	0.011	0.017	0.127
X ₁₅	Adj. Acc. Price of the 5 th nearest	-0.0010	0.029	-0.035	0.972	-0.058	0.056
X ₁₆	Adj. Acc. Price of the 6 th nearest	0.0477	0.030	1.565	0.118	-0.012	0.108
X ₁₇	Adj. Acc. Price of the 7 th nearest	0.0715	0.031	2.327	0.020	0.011	0.132
X ₁₈	Adj. Acc. Price of the 8 th nearest	0.0543	0.032	1.710	0.088	-0.008	0.117
X ₁₉	Adj. Acc. Price of the 9 th nearest	0.0450	0.032	1.400	0.162	-0.018	0.108
X ₂₀	Adj. Acc. Price of the 10 th nearest	0.0430	0.030	1.425	0.155	-0.016	0.102



Figure 4: Illustration of the comparative sales with resales. Sales highlighted with red indicates the resales.

4.5 Summary

The summary of the three models can be found in the table below.

Table 5: Summary of OLS Regression results for the three models.

	Model 1	Model 2	Model 3
Model:	OLS		
Method:	Least Squares		
R²:	0.319	0.559	0.765
Adjusted R²:	0.306	0.541	0.755
F-statistic:	25.61	30.99	73.50
No. Observations:	503	484	449
No. Variables:	10	20	20
DF Residuals:	493	464	429
DF Model:	9	19	19

It is evident that ‘Model 3’ can be considered the most accurate and with the greatest significant levels. Based on all the above, the regression model and forecasting equation is as follows:

$$\text{Plot value per m}^2 = 58.3328 + 0.0022 * X \text{ Coordinate} + 0.0034 * Y \text{ Coordinate} - 14.1359 * \text{Share} - 0.0226 * \text{Land Size} - 0.0983 * \text{Frontage} + 21.6574 * \text{East} - 3.8698 * \text{S-E} + 9.4304 * \text{N-W} + 4.2950 * \text{N-E} + 26.8198 * \text{North} + 0.3784 * \text{Price of the 1st nearest} + 0.0192 * \text{Price of the 2nd nearest} + 0.1379 * \text{Price of the 3rd nearest} + 0.0721 * \text{Price of the 4th nearest} - 0.0010 * \text{Price of the 5th nearest} + 0.0477 * \text{Price of the 6th nearest} + 0.0715 * \text{Price of the 7th nearest} + 0.0543 * \text{Price of the 8th nearest} + 0.0450 * \text{Price of the 9th nearest} + 0.0430 * \text{Price of the 10th nearest}$$

Equation 2: Forecasting Equation of the optimal model

4.5.1 X and Y Coordinates of the centroid

The X and Y coordinates of the centroid are six-digit numbers. For the modelling purposes, these coordinates have been divided by a constant coefficient to produce smaller numbers. In regressions it is good to have smaller numbers – it is also the reason that there is a factor in front of each coordinate.

4.5.2 Share

It was predicted that the share is likely to have a negative impact on price. This is indeed the case since the range of buyers who would choose a share land is limited (lower demand – lower price).

4.5.3 Land Size

The drop in land values once the extent was larger was evident to see, resulting in a negative linear relationship between the price per m² and the total extent of the land (m²). The analysed data indicates that as the land extent increases the price per m² drops by 2,26%.

4.5.4 Frontage

It was predicted that the frontage would have either positive or negative effect on the sale price. In model 1 and model 2 would have a positive effect while in model 3 the effect is negative. The general rule of thumb is that the larger the frontage the higher the sale price. In general, the frontage does not work properly in the model, it seemed from the beginning that it had high p-values.

4.5.4 East, South-East, North-West, North-East, North

The general rule of thumb is that the south rotation is the optimal for residential plots, so the south-east orientation had to have a positive effect on the value and the rest a negative one. In the above equation we have exactly opposite results.

4.5.5 Price of the k-nearest neighbors

The objective being to extract those comparables having the lowest “distance” value. As similarity between the subject and the comparable decreases the distance figures increases. For example, as Comparable 1 is the most similar to the subject in terms of distance, its reasonable to expect it to have the highest influence on value and consequently the highest comparable weigh. The most important thing that was observed in these parameters is that the nearest comparable has the greatest effect on the value while as we move away this effect decreases - except for the fourth comparable which has a negative effect. Nine out of the closest neighbors have a positive effect on value while one has a negative.

5. CONCLUSION, LIMITATIONS AND RECOMMENDATIONS

Ad volarem property tax systems normally require that the tax base be regularly revalued on a three-yearly, five-yearly, or even annual cycle. Working in an environment of cost constraints, limited resources, and reflecting the scale of having to value hundreds of thousands of properties as of a given date, have important implications for government valuations departments with responsibility for property tax assessment. The need to develop and apply computer-based mass appraisal systems has never been greater than it is today.

The main aim of this research was to create a fully working AVM using data from Cyprus Department of Land and Survey in order to compare with the results of the latest General Valuation (1.1.2018). The correlation between the actual prices and the predicted prices determined by the ‘Model 3’ was calculated at 76.5%. This indicated that the ‘Model 3’ provide much more accurate property estimates, since the correlation factor between the actual residential plot prices and the values determined by the Cypriot Authority is much lower with accuracy of 38.4%. The results also verify the hypothesis of the term study, indicating that the effect of the nearest neighbors is fundamentally significant for ensuring the accuracy of price prediction models as regards real estates.

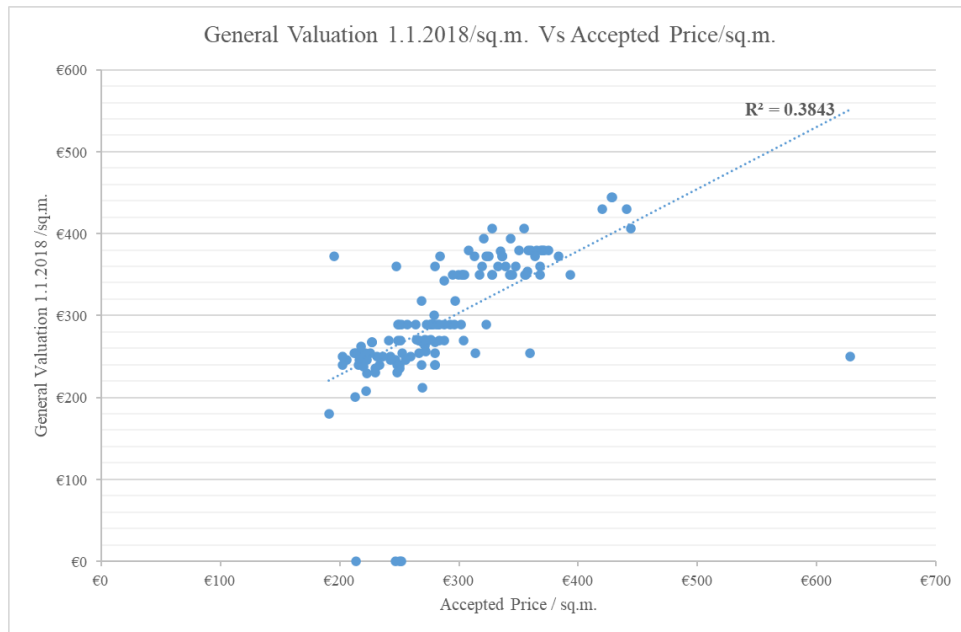


Figure 5: Accuracy of the General Valuation 1.1.2018

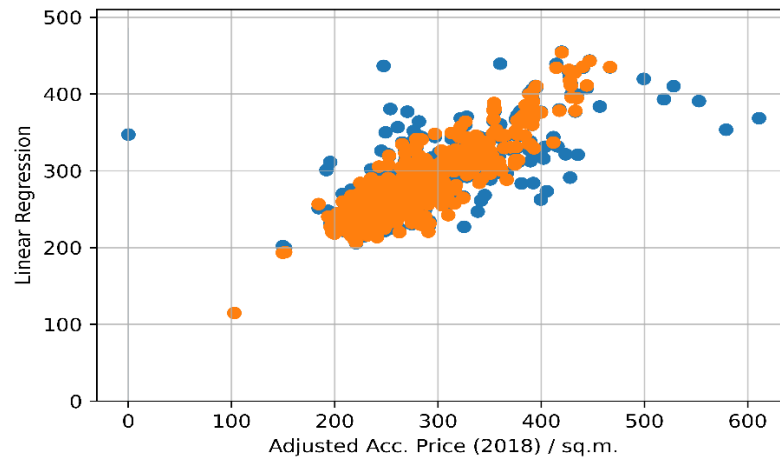


Figure 6: Accuracy of prediction model (Model 3) of this study

At this point it is important to notice that the accuracy of the Cyprus Land Registry has been calculated only for residential plots that have been sold between 2018 and 2019 due to the fact that in General Valuation 1.1.18 were used data only for these two years and therefore no time adjustments were made. While the accuracy of the Model 3 concerns the period 2012-2019 and compares the Predicted Values with ‘Adjusted’ accepted price. Time adjustments have been achieved based on one reliable price index tool (CBC index) that provide a general view of price fluctuations. Thus, the plot prices used in the model as actual values are based on approximations. Therefore, the ‘Model’ has the possibility to be further improved in accuracy by providing property transactions achieved within the same market period.

The extensive research that followed allowed for the identification and quantification of variables associates with residential plot values, where by twenty of them can explain the price per sq.m., with the most significant variables identified as the share and two orientations.

However, the above models and findings are not without any limitations. It is general knowledge that the time period 2012-2019 capture three different economic periods. This is clearly visible by looking at the number of transactions on an annual basis (**Error! Reference source not found.**).

There other variables as well, which were not included in the analysis, as there was lack of information resources to include them. For example, ‘the supply’ in each residential area could also have a bearing on the land values. Basic supply and demand parameters state that the smaller supply (thus the availability of vacant land), the higher the price. Therefore, the availability of land in each location in comparison to the total residential land extent should be looked at. This could be done with GIS or other mapping techniques to identify vacant land. However, as some areas also have a number of derelict buildings with no economic value, on-site visit would also be required to identify the true supply figures.

This study has highlighted and identified a considerable amount of further opportunities for research for those involved in the residential industry. Firstly, it appears that those comparables having the lowest ‘distance’ value, have the highest influence on values and consequently the highest comparable weigh. This could therefore, lead to further research in identifying the optimal number of k-nearest neighbors through process of cross-validation

REFERENCES

- [1] Amato, M. (2017). *Advances in Automated Valuation Modeling* (Vol. 86). <https://doi.org/10.1007/978-3-319-49746-4>
- [2] Carbone, R., & Longing, R. L. (1977). A feedback Model for Automated Real Estate Assessment. *Manage Sci*, 24, 241–248.
- [3] Colwell, P. F. (1999). Who was first? An examination of an Early Hedonic Study. *Land Econ.*, 75, 620–626.
- [4] Eckert, J. K. (1990). Property Appraisal and Assesement Administration, (IAAO).
- [5] Gwartney, T. (1970). A computerized Assessment Progra. *Assess. L. Value*, 125–142.
- [6] IAAO. (2018). *Standard on Automated Valuation Models (AVMs)*.
- [7] Labropoulos, A., Dimopoulou, E., & Zentelis, P. (2003). The necessity of developing a CAMA System for the Real Estate Market in Greece, 1–12.
- [8] McCluskey, W., & Anand, S. (1999). The application of intelligent hybrid techniques for the mass appraisal of residential properties. *Journal of Property Investment & Finance*.
- [9] Varnavas, P., Charalambous, M., & Savvidou, E. (2015). Assessment of the variance between GV assessed values 01.01.2013 and Market Price, (January).