Site Selection for Subdivision Development in Guiguinto, Bulacan using GIS and Regression Analysis

Ernest John C. BUGAY, Jolu B. BUNDA, Louie P. BALICANTA, Matthew Oliver Ralp L. DIMAL, and Enrico C. PARINGIT, Philippines

Key words: real estate development, GIS, multiple regression analysis

SUMMARY

Site selection for new residential settlements is a complicated task, dealing with different variables such as accessibility, zoning, proximity to facilities and businesses, and demographic factors. Being quantities of different nature, a different approach is presented in this paper to represent their influence in value using regression analysis. A case study was conducted in Guiguinto, Bulacan, a first-class municipality in the Philippines, about 28 kilometers northwest from Manila. From a developer's perspective, a regression algorithm which calculates for the counter-utility function value was formulated using the operational form of the Bid-Rent economic theory. These values were spatially represented using GIS, showing the utility assessment of Guiguinto for residential use.

Site Selection for Subdivision Development in Guiguinto, Bulacan using GIS and Regression Analysis

Ernest BUGAY, Jolu BUNDA, Louie BALICANTA, Matthew Oliver Ralp DIMAL, and Enrico PARINGIT, Philippines

1. INTRODUCTION

Land development is defined as modifying a previously unsuitable parcel of land to fit as environment for a desired human activity (Dion, et. al, 2002). It has proven to be one of the largest and important industries in the Philippines, and has been considered to be an indicator of the wellness of a nation's economy. Numerous subdivision sites have been constructed in predominantly rural or semi-rural areas in the fringes of Metropolitan Manila, to comply with the need for dwelling units of the immigrating population. Successful ventures of decongesting city-centers through development of the adjacent suburbia have been credited mainly to planners and developers.

One important work of a land developer is finding an appropriate site to construct a subdivision. Site selection in land development requires dealing with interrelated factors simultaneously, which includes finding a location that provides the demands of their target occupants, economically feasible to operate, and contributes to the betterment of the community. Here, the main question arises: How does one deal with these considerations? If a development project team has to compromise some factors over others, they may need to reorganize and set their eyes on the most critical parts of the project. Decision tools that are contextualized will greatly benefit them.

A field of spatial technology closely linked with land development is Geographic Information System (GIS). Since a number of problems that development projects encounter in executing projects are location-dependent, a system that contains the most updated information regarding a location of interest and tools that let decision makers manipulate the information to suit their need for details may be of high utility. The process of site selection for subdivision development may be made easier if development projects will be provided with tools that will help them make decisions based on some understanding of their environment. The decisions regarding where to acquire land for a subdivision with a specific group of target and how much units in that subdivision should cost, will be based on research and updated, true-to-life data, instead of on just the notion of amateur viewers and intuitions of third-party appraisers.

Efforts have been done to support the decision-making process of land development using a more scientific approach; however, many of the studies on acquisition of geographical data have been done using very limited parameters and spatial techniques. The researchers are interested to look for ways to involve the capabilities of Multiple Linear Regression, economics and GIS in providing any residential development project with the necessary information coming from digital data sources, such as GPS coordinates and satellite images.

1.1 Site selection and the developer's role

Basic texts on land development emphasize on the importance of evaluating the market area in assessing the viability of a project. Evaluation of the market area leads various considerations such as potential for profit, utility of a project, and the soundness of expectation that the market may be able to absorb the project. Dion, et. al (2002) describes this evaluation process as the *Macro-* and *Microanalysis* of the possible development site. Macroanalysis determines the viability of the project in terms of demographics and projection of figures in the future, while microanalysis involves maximizing the possibilities offered by the previous analysis. Site selection throughout the market area is involved in the microanalysis.

Upon confirmation of the project feasibility from the macroanalysis, site selection is performed through the minute analysis of the market area. This is done during the development project planning to be able to make sure that the project is appealing to the target market (Kone, 1994). A good residential location, no matter the target market, means that services and security are adequate (*ibid*.). In a Filipino setting, it must be seen that the security and service needs of various age groups represented in an extended family setting, have to be considered, and this must be performed under a strict budget-constrained plan by the development team.

Specific site characteristics to be dissected in the microanalysis vary in nature, and different literatures categorize these factors into subgroups using various schemes. Four main categories can be conceived out of this wide variety of characteristics, first of which is the *site location*. This refers to the relationship of the potential site to adjoining municipal facilities (Dion, et. al., 2002). *Intrinsic conditions* can also be regarded as a major set of factors, describing all the physical characteristics that affect the development plan in terms of design, engineering and construction (Kone, 1994). Examples of intrinsic site properties are soil type, environmental hazards and vegetation on the location. *Financing and marketing* strategies also make up an important set of factors in the consideration of site selection. Finally, those site characteristics concerned with *legal and political practices* of the local and the national situation of the potential site can never be taken for granted, zoning protection as an example is included in this set (Chapin and Weiss, 1962).

1.2 Economics of site selection

To represent the response of future occupants to the target location, an application of Alonso's (1964) *Bid-Rent Theory* has been formulated, wherein the site location factors regarded as significant will be mathematically represented by a Utility function that represents the living condition in that location. With respect to the household, the theory states that there is a marginal (boundary) amount of money that the consumer is willing to pay to possess the location. The Bid-Rent Theory can also be applied with respect to the location, such that there is a minimum amount of money that a household should disburse to be able to enjoy a certain set of characteristics describing a desired living condition. Applying the Bid-Rent Theory to a household's decision making, there are two considerations on its part that must be modeled: the *condition of the site* and the *allowable expenditures*. Expressing the theory in mathematical form by making use of Mäler's (1977) budget constraint inequality, the condition of the site can be formulated as a Utility function of the site U, such that

$$U = U(x_1, x_2, x_3, \dots, x_i), \qquad \text{(Eq. 1)}$$

where the x's are the characteristics describing the site. The expenditures on the other hand are merely the required purchasing power for him to avail of the *x's* as a package, or as a single good *U*, pertaining to the residential site he is interested in. It is established that in the case here, any given *x* characteristic is a *good*, such that it has a positive desirable value (Rosen, 1974). Plugging in the necessary relationship between the condition of the site and the allowable expenditures, the householder should possess the minimum amount of money to be able to obtain his desired location, or

$$U = U(x_1, x_2, x_3, \ldots, x_i) \leq I(y), \qquad \text{(Eq. 2)}$$

where *I* is the wealth of the household coming from its sources of income *y*.

1.3 Preliminary site selection factors derived from previous studies

Urban economic studies found out that residential location choice is mainly affected by the potential of a household location to be accessible to activities important to family members. Of particular interest with respect to the idea of accessibility is the dogma of the importance of travel time to work. In urban economic theory, if a certain parcel of land can yield more benefits to the occupants compared to some other location, desire for this type of land will soar (Rivera and Tiglao, 2006). Therefore, the shorter the travel time is from a place of residence to work, the higher the demand for (and therefore, price or rent of) the location. One of the simplest ways to measure accessibility involves the primitive method of straight-line distances, as recognized by earlier works of Alonso (1964), Carao and Hernandez (1992) and Handy and Niemeier (1997).

A number of studies conducted for urban residential location had also been performed to tie the demand for a particular residential site to a bundle of characteristics. Tamayo and Salvador (1990) views the proximity to market places as a prerequisite for living in comfort, since this means availing of basic commodities at lesser cost. Also, high population density drives people toward the area since they prefer others' company. Good road conditions provide convenience, for they facilitate mobility, although it was concluded in their study that they were insignificant to the results (ibid).

Carao and Hernandez's (1992) subsequent study reveals that influx of people tends to congest an area, leading to increased pollution and competition for amenities. They have also extended the proximity concerns to other types of amenities, such as academic institutions and other establishments, since these factors contribute to the utility of the land to residents. However, this factor was also shown to be insignificant (ibid). The city center, where most establishments and transportation routes occur, has been regarded as a hub of convenient living. Unexpectedly, their study reflected the opposite as the distance to the city center was also not critical with respect to the demand for a residential site.

Kone (2006) also includes other important pointers such as off-site factors that may increase the risk for occupants, like industrial and waste disposal sites. The above characteristics were considered as preliminary factors by the proponents in their study, such as the presence of industrial sites around the location. In the research conducted by de Vor and de Groot (2009), a logarithmic regression analysis of the relationship between the distance from industrial sites and housing prices sheds light to the fact that there comes a point where the transaction price (in Euro) for residential housing varies with the distance from the nearest industrial site up only up to a particular point, due to the asymptotic nature of the relationship. This reveals that increasing distance from industrial sites leads to a gradual diminish of increase in residential lot prices, and further increase in the distance beyond a certain limit poses almost no recognizable residential lot price increase (de Vor and de Groot, 2009).

In spite of this factor being non-location based, the proposed lot sizes of a subdivision is an aspect of the project which is extensively being studied in the development plan (Dion, et. al, 2002). Lot sizes have intangible but nonetheless marketable characteristics. This property is described as *quantity premium*, where in this case larger lots are more expensive per unit area. Tabuchi (1996) sees this situation as stemming from characteristics that go with a large lot, such as the option value of splitting it further or being able to construct a huge structure. Lots are also viewed as *irreversible*, meaning once they are subdivided and sold to a number of purchasers, the subdivided parcels will be very difficult to reassemble because of the intermediary costs and other impediments (Tabuchi, 1996). For the reason that the future revenue and the topographic considerations have to be balanced for a stable and profitable project, the minimum lot size is included as a predictor.

1.4 Assessment of significant site selection factors

In assessing the importance of the individual preliminary factors in the case of their study area, the proponents used Multiple Linear Regression Analysis. Regression is a widely-used tool for such purposes, among other things, because of the simplicity of the method in providing a functional relationship between the significant preliminary factors, or the *predictors*, and the index, or the *dependent variable*. It takes the form

$$y = f(x_1, x_2, x_3, \ldots, x_k) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_k x_k, \qquad \text{(Eq. 3)}$$

where the dependent variable y have k predictors represented by the x's, and the β's are the constant and the coefficients of each predictor. It is straightforward in the sense that the determination of the explicit regression equation is, in a manner of speaking, the final product of the analysis (Chatterjee and Price, 1977), although there are important by-products of the method that are necessary to evaluate the final product.

One of the by-products of regression analysis that can be used to evaluate and validate the regression equation is the *standardized residual* s_i, where

$$s_i = (y_i - \bar{y})/\sigma \qquad \text{(Eq. 4)}$$

where y_i is the ith actual observation while $\bar{y}$ is the mean of the predicted responses and σ is the standard deviation of y's distribution. Standardized residuals are used to check whether the dataset conforms to a number of assumptions of a multiple linear regression model such as the randomness of the errors, or the residuals, of the regression equation. They are also used to gauge whether the model is generally able to predict the dependent variable correctly, as literature refers to standardized residuals less than ±2.00 as acceptable (Chatterjee and Price, 1977). They are also used to assess the assumption of *homoscedasticity*, or constant variance of the residuals. Equal variance of the residuals is a requirement for the model to be able to estimate the k coefficients with acceptable precision. Using the standardized residuals to test for homoscedasticity and zero-mean of the residuals, a scatterplot of the standardized residuals against the predicted response variable $\hat{y}$ known as a *residual plot* is created. When the residuals appear to fluctuate around $s_i = 0$ in a random manner, the dataset contains equal-variance errors.

Another important by-product is the individual *Variable Inflation Factor*, or VIF, of each of the predictors. A predictor's VIF measures the combined effect of dependency of the remaining predictors to the variance of that predictor (Ranjit, 2006). Variable Inflation Factors of all the k predictors are computed from the correlation matrix **X'X** through the following relationship:

$$VIF_k = (\mathbf{X'X})^{-1}_{kk} \qquad \text{(Eq. 5)}$$

When a predictor's VIF is close to unity, it indicates that the predictor does not suffer from severe collinearity with any of the other predictors (ibid).

The *prediction interval* of the dependent variable is also discussed to be able to generate a range of the most plausible predicted response values for the dependent variable based on the model, according to a certain confidence level. (Larsen, 2008) The prediction interval of a dependent variable response is given by the expression

$$\hat{y}^* \pm t_{1-\alpha/2}\,(n-k-1)\ \mathrm{se}[\hat{y}^*] \qquad \text{(Eq. 6)}$$

where $\hat{y}^*$ is the predicted response, $t_{1-\alpha/2}\,(n-k-1)$ is the $(n-k-1)$-quantile of a $t_{1-\alpha/2}\,(n-k-1)$ distribution, and se[$\hat{y}^*$] is the estimated standard error of the predicted response $\hat{y}^*$, where

$$\mathrm{se}[\hat{y}^*] = s\sqrt{(1 + \mathbf{x}^{*T}(\mathbf{x}^T\mathbf{x})^{-1}\mathbf{x}^*)} \qquad \text{(Eq. 7)}$$

with **x*** denoting the matrix of the predictor values $x_1, x_2, x_3, \ldots, x_k$ giving the response $\hat{y}^*$.

2. MATERIALS AND METHODS

2.1 Study Area

Guiguinto, Bulacan is a predominantly agricultural municipality. It has gradually urbanized and developed into a 1st class municipality, with a median income of more than P 35,000 per household. Currently, there are 24 residential subdivisions around the town of Guiguinto. As of 2008, the municipality has a population of 95,092 and consists of 14 barangays with a total land area of 2,512 hectares, a considerable amount of which caters to nurseries of ornamental plants and flowers. Because of its sought-after garden plants, combined with the influx of light commercial and industrial establishments, people from the urban areas of Metro Manila and Malolos congregate to Guiguinto. This influx of industries and households makes Guiguinto an appropriate frontier for subdivision development ventures.

The model that has been generated by this research is limited only within the boundaries of Guiguinto, Bulacan. The study focused on the properties of 20 existing subdivisions in Guiguinto and concentrated on the relationship between the market value predictors and the suitability of a parcel for residential development. Residential development ventures stated were assumed to be commercial in nature. Low cost housing, socialized housing and relocation sites, were not included in this study.

2.2 Conceptual Framework

Finding an appropriate site for subdivision projects is essentially a problem in economics. A development project needs to maximize the potential, or at least the use, of the land that was purchased for the development. Regarding the different types of site characteristics, previous studies regarding determinants of residential property values cited several site selection factors that contribute to the variation among residential lot prices. Majority of the determinants fell under the category of *site location*, such as proximity to city center and presence of amenities. Legal practices, such as zoning ordinances and land use plans, also impacted on the perception of future occupants of the area.

Using the concept of assessing which among the above determinants affected variations in residential lot values, Multiple Regression Analysis can be employed to evaluate the individual contribution of these *site location* factors, *legal practices* and *intrinsic condition* factors to the prediction of Market Values of feasible residential parcels in Guiguinto. The preliminary factors and the market value per unit area of the subdivision lots form the linear using Eq.3, where the predicted Market Value per Unit Area was represented by the dependent variable y, and the β's were the constant, and the coefficients of each Market Value predictor. The result of the regression process highlighted the site characteristics that significantly affected the variation among market values of the different subdivision lots.

Tools for measuring distances, counts and statistical figures regarding the site characteristics of interest were manipulated using the GIS. These significant site characteristics could then be able to predict the market value $\hat{y}$ at which a particular lot with a given set of predictor attributes could be sold. The predictive power of the resulting regression equation was then evaluated for its applicability and predictive accuracy for the entire Guiguinto area.

To be able to create a model of the feasibility of a subdivision development project, the results of the Regression Analysis has to be perceived as an operational form of the Bid-Rent Theory, which was adapted to suit the point of view of the development project. From Eq. 2, the aggregated utility function of the site Un, is given by the following equation:

$$U_n = U(x_{1n}, x_{2n}, x_{3n}, \ldots, x_{in}), \qquad \text{(Eq. 9)}$$

where j is the total site characteristics and the x's describing the first i ($i < j$) site characteristics. Then, every site location and/or legal and political factors that had been confirmed as significant by the regression analysis in predicting the Market Value per Area of subdivision parcels, were transposed to the Utility expression side of the inequality. The Utility expression (in monetary units), for any location n in Guiguinto contained location-based characteristics that described living conditions abounding the residential site. The expenditure terms were on the other side of the inequality, which included characteristics such as incurred cash flows to the development project. These characteristics are intrinsic to the development, since they affect construction and design of the development plan.

Since the expression was in monetary value, which is essentially a part of the development project's revenue from the sale of a residential unit (house and lot package), it has to be classified as a *negative incurred cost* of development. The negative sign is due to the costs being attributed to the perspective of the developer; hence any cash inflow was marked as a negative cost. Formulating the expenditure terms from the remaining ($j - i$) characteristics:

$$\begin{aligned} E_n &= E(x_{(j-i)n}, x_{\{j-(i-1)\}n}, x_{\{j-(i-2)\}n}, \ldots, x_{jn}, MVA_n) \\ &= \beta_0 + \beta_{(j-i)}x_{(j-i)n} + \beta_{\{j-(i-1)\}}x_{\{j-(i-1)\}n} + \ldots + \beta_j x_{jn} - MVA_n \qquad \text{(Eq. 10)} \end{aligned}$$

Using the relationship between the condition of the site and the allowable expenditures, the inequality (Eq. 11) showing the minimum amount of money to be able to obtain the desired parcel n anywhere in Guiguinto is:

$$U(x_{1n}, x_{2n}, x_{3n}, \ldots, x_{in}) \leq E(x_{(j-i)n}, x_{\{j-(i-1)\}n}, x_{\{j-(i-2)\}n}, \ldots, x_{jn}, MVA_n), \qquad \text{(Eq. 11)}$$

where the x values in the Utility function were sets of "goods," such that development project

and the occupants alike will put positive marginal price on those characteristics while their value increases. In this situation, higher values for the Utility function meant better living condition for the future occupants.
With the help of the data representation capability of GIS, the Utility function value UTILITY of a certain location,

UTILITY = $U(x^*_1, x^*_2, x^*_3, \ldots, x^*_i)$ (Eq. 12)

was computed for applicable parcels in the entire Guiguinto. Spatial patterns produced by the fluctuation of the values of predictors making up the Utility function were depicted by contour maps. These contour maps could be overlaid with the Guiguinto parcels to be able to visualize the movement of predictor values throughout the area. This overlay function and the computational capability of GIS facilitated in the assignment of Utility function terms to the parcels in aiding with the computations of UTILITY.

2.3 Data Requirement

To model the site selection criteria that suit the willingness-to-pay for residential sites, data for the dependent variable which is the *market price* of subdivision lots and for the preliminary factors, as discussed in the Related Literature portion, should be available for input in the Regression Analysis. A base map of the entire Guiguinto, Bulacan area is necessary for the geographic database platform. Through the assistance the Municipal Planning and Development Office (MPDO), a shapefile of the parcels of the study area was obtained. For the market prices of subdivision lots in Guiguinto, an updated list of the current subdivision projects was obtained also through MPDO. To gather figures on the *Market Values per Unit Area* (MVA) field inspections and interviews to lot owners and brokers were conducted.
Locations of the pertinent points in the study such as the input subdivisions, establishments and the road network around the Guiguinto area were plotted as GIS points and polylines on the base map using satellite images from handheld GPS readings observed during site inspection and online satellite images of the municipality. Population records per barangay dated 2008 were taken from the records of the Municipal Planning and Development Office.
GIS software was utilized to generate the tables of data necessary for the variable inputs. Most of the data are proximity values, and others involved queries of *how many* or *how much*.
Dataset for the *minimum lot size* (LA) was completed by highlighting the parcels that comprise a subdivision and choosing the *Statistics* tool to display all the statistical figures regarding the selection.
The *population densities per barangay* (PDB) were computed using the data extracted from the land areas of each barangay in square meters and converting to hectares. Therefore, population density values were given as the number of persons per hectare of the barangay in the year 2008. Each of the subdivisions was then associated with their respective barangay for their population density values.
The datasets for the variables *distances to nearest establishments* such as *schools* (DS), *marketplaces* (DMP), the *Guiguinto Municipal Hall* (as the municipal center) (DMC), *hospitals* (DH), *banks* (DB), *churches* (DC), *gasoline stations* (DGS) and *restaurants* (DR) as well as the closest *main road* (DMR) were generated through the use of the *Near* function of the Proximity Analysis tools by ArcMap.

A recent study conducted by de Vor and de Groot (2009) of the Netherland's Tinbergen Institute reveals that the proximity of industrial sites to a residential area has a negative logarithmic relationship to the site's market value. Therefore, a count of the *number of industrial sites* within a four- (4) kilometer radius from each of the subdivisions (IND4) was generated using the *Buffer* function and the *Select by Location* capability of GIS.

2.4 Model Generation and Representation

In determining the influence of each of the preliminary factors as a residential site selection criterion, the proponents utilized Multiple Regression Analysis to come up with an explicit equation that displays the coefficient for the predictors. Generating the regression equation also evaluates which among the preliminary factors significantly affects the dependent variable. In this procedure, the twenty (20) subdivision samples in Guiguinto, are used in the modeling set for the selection of preliminary factors to be included by performing Forward Selection wherein only one factor was included in the regression at the first run, noting that its *t-statistic* should exceed ± 1.00 (Chatterjee and Price, 1977). Addition of new preliminary factors should be according to this fashion until the last factor's t-ratio failed this criterion. Transforming the data using simple mathematical functions was also deemed necessary to achieve linearity of the dataset and to follow certain assumptions such as homoscedasticity (constant error variance). Upon following the guideline, five (5) preliminary factors were chosen to be included in the model as the Market Value (MVA) predictors: minimum lot size (LA), natural logarithm and square root of the distance to the nearest marketplace (*ln* DMP) and bank ($\sqrt{\text{DB}}$), number of industrial sites within 4 Km (IND4) and natural logarithm of the population density (*ln* PDB).

The *standardized residuals* of the modeling set were computed for to be used in the detection of erroneous prediction of Market Value per Area of subdivision lots, seeing that no standardized residual exceeded ± 2.00 (Chatterjee and Price, 1977). The resulting regression equation must also be checked against serious violations of the assumptions for precise determination of the coefficients, such as multicolinearity. This was performed by computing for the *Correlation Matrix* and the individual *Variable Inflation Factors* (VIF) using Ranjit's (2006) formulas and criteria, seeing that no correlation value between any two predictors was above *0.70* (no strong correlation) and that no VIF was above *10.0*. Another assumption to be checked for serious violation would be homoscedasticity and this was done with the use of residual plots. Standardized residuals should appear to scatter around the plot randomly centering at $s_i = 0$ if the dataset manifests constant error variance.

The 20 subdivisions were then divided into two groups, one group for generating the regression equation and the other for testing of the prediction acceptability. This was done by systematically dividing the subdivisions into two, 10 for each group. The model will use the five variables that were chosen earlier from the Forward Selection. The first 10 subdivision market values were used to generate the coefficients for the model. The remaining 10 subdivision market values were predicted by the regression equation to assess the accuracy of the Market Value prediction capability. This was done by calculating the prediction intervals of the response for each remaining subdivision to have a grasp of whether the Market Value per Area prediction intervals were able to capture the actual Market Values at a certain confidence level. This cycle ran for 10 times. Upon confirmation of the acceptability of the predictive power of the 10 generated regression equations from the different sampling systems, the coefficients of the predictors from those 10 models were averaged to produce the coefficients of the final model.

2.5. Model representation

To be able to easily visualize areas which are feasible subdivision development sites, the proponents mapped first the predicted Market Value per Unit Area (MVA) of parcels throughout the municipality.

In assigning the values of the predictors to each of the parcels, the proponents employed Gao and Asami's (2005) idea of a "contour map" of the spatial characteristics. The distances of each parcel to the nearest marketplace (DMP) and bank (DB) were simulated by a contour map, where "contour lines" represented distance from the closest market or bank. The distance value of the contour line within which the centroid of a parcel falls was assigned as the parcel's DMP and DB. Meanwhile, the number of industrial sites within 4 kilometers (IND4) of the parcels was also counted by employing the method, wherein lines formed boundaries around areas with a particular count of industrial sites within 4 Km, and assigning the industrial site count to a parcel the centroid of which falls inside the bounds.

The minimum lot size (LA) value, however, is considered in the profitability study of the project. It is not a spatial characteristic, and therefore, cannot be mapped. Hence, the proponents assumed an outstanding value for this predictor to be assigned as the LA of all the parcels in Guiguinto. With the mean minimum lot size value of $58.7m^2$ for the fifteen model subdivisions, an LA of $60\ m^2$ was assigned to all the parcels. This assumption would charge an estimated amount of ±*P1, 380* in the MVA.

Chatterjee and Price's (1977) warned in estimating a linear relationship deduced from the regression analysis, which may not hold for values outside the range of observations. Parcels which extremely deviate from the predictor values would then have to be marked off from the model prediction.

To be able to obtain the predicted MVA of all the other parcels, the part of the price contributed by each of the predictors were computed for, using the functional form of the regression equation

$$MVA = f(LA, \ln DMP, \sqrt{DB}, IND4, \ln PDB), \qquad \text{(Eq. 13)}$$

where $f(LA, \ln DMP, \sqrt{DB}, IND4, \ln PDB) =$

$$\beta_0 + \beta_1(LA) + \beta_2(\ln DMP) + \beta_3(\sqrt{DB}) + \beta_4(IND4) + \beta_5(\ln PDB) \qquad \text{(Eq. 14)}$$

The regression equation was transformed to represent the Bid-Rent Model applicable to the Guiguinto area. A *Budget constraint* equation was derived upon isolating the development costs, including the Market Value, from the Utility function terms:

$$- \beta_2(\ln DMP) - \beta_3(\sqrt{DB}) - \beta_4(IND4) - \beta_5(\ln PDB) = \beta_0 + \beta_1(LA) - MVA. \qquad \text{(Eq. 15)}$$

The UTILITY value proposed by the researchers was based on the Utility function value, or the evaluation of the entire expression on the left side of the equation, derived from the regression equation for the predicted Market Value per Unit Area of each parcel.

With the use of GIS, the UTILITY values of all the valid parcels in Guiguinto, were calculated and depicted in a map to visualize centers of actual and feasible residential areas with good and poor utility values.

3. RESULTS AND ANALYSIS

3.1. Multiple regression analysis and significant Market Value predictors

Subdivisions used in the generation of the regression equation and assessment of the confidence level of the predictive power of the equation are shown in **Table 1**:

Table .1. List of subdivisions involved in the study

Subdivisions Used in the Market Value Prediction			
Cyberville	Harcelville	North Bel-Air	Sta. Clara
Dona Pilar	La Aldea	Oro Villa	Sta. Rita
Felville	Masagana Phase 3	Rockaville	Sta. Rita Village
Goldridge	Maunlad Homes	Rosaryville	Tierra Dayao
Green Estates	Mira Verde	St. Agatha	Violeta

Data gathered were used as input for regression using spreadsheet program by employing the Forward Selection method. The results are shown on **Table 2**.

Table 2. The initial regression results. Listed down are the estimated coefficients with their t-ratios, the R^2 of the regression line and the F-statistics of the predictors.

	Coefficients	*Standard Error*	*t Stat*
Intercept	21100.663	2816.667	7.49136
Industrial (IND4)	-584.451	78.248	-7.469199
Minimum lot size (LA)	43.874	5.702	7.695001
Nearest bank ($\sqrt{}$ DB)	-43.434	13.639	-3.184626
PD (pers/ha) (ln PDB)	-726.019	341.306	-2.127181
Nearest market (ln DMP)	-180.468	162.637	-1.109636

Regression Statistics	
Multiple R	0.955954771
R Square	0.913849525
Adjusted R Square	0.883081498
Standard Error	486.4760073
Observations	20

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	5	35145352.27	7029070	29.70127169	5.576E-07
Residual	14	3313224.679	236658.9		
Total	19	38458576.95			

Examining the t-statistics of the individual Market Value predictors, all were within *0.10* level of significance except for the variable *ln DMP*. The R^2 value of *0.91* proved the significance of the predictors, as they justified a high percentage of the variation among predicted Market Values. The F-value of *29.70* adds evidence to the linear relationship of the combined set of predictors, explaining empirically that the estimated coefficients above are not equal to zero. This F-statistic reports with above 99% confidence level that the set of predictors are collectively significant.

After the selection process, the dataset was tested for violations of the linear regression assumptions. Upon confirmation that none of the basic assumptions of Linear Regression were severely violated, the regression analysis revealed five significant predictors in the determination of the Market Value per Unit Area MVA of the subdivisions. These five predictors were then used to formulate 10 different equations, then each coefficients generated were averaged to produce the final regression coefficient. To test whether the mean coefficients were significantly different from the coefficients with the 20 subdivisions, the t-statistics of each coefficient was compared at 95% confidence. The two models would be considered distinct if the t-statistics of the coefficients would have been greater than 2.262. **Table 3** shows the coefficients of the regression results and their t-statistics:

Table 3. The final regression results. Since all t-statistics are less than 2.262, the mean coefficients are not significantly different from the model coefficients.

	Mean	*Model*	*Std Dev*	*t-stat*
Intercept	**24118.753**	21000.663	5385.935	1.772
Industrial	**-589.173**	-584.451	88.595	-0.169
Lot size	**42.497**	43.874	6.838	-0.637
√ Bank dist	**-58.351**	-43.434	34.461	-1.369
ln Popden brgy	**-1138.752**	-726.019	715.949	-1.823
ln Market dist	**-169.264**	-180.468	212.763	0.167

The estimated linear equation generated was:

$$MVA = 24118.75 + 42.50*LA - 589.17*IND4 - 58.35*\sqrt{DB} - 1138.75*\ln PDB - 169.26*\ln DMP. \quad \text{(Eq. 16)}$$

Most of the expected signs of the MVA predictors were taken on by the regression result. Gathering the negative signs of the square root and the natural logarithm of the *Distances to nearest bank* (DB) *and marketplace* (DMP), it is logical that residential areas farther away from public institutions and amenities be sold at a lower price since occupants of these residential areas pay higher fare and travel longer to get to these establishments. The natural logarithm of the *Population density per barangay* (PDB) also possessed a negative sign, revealing that residential lots in congested barangays have the tendency to sell for less; moreover, as the *Number of industrial sites within 4 Km* (IND4) increased, the lower the MVA was for a residential site. This predictor affected market value because the off-site factors resulting to noise and environmental hazards that deteriorates occupants' perception (Kone, 1994). Also, the *Minimum lot size* (LA) had positive value to the MVA of a subdivision lot. The sizes of parcels in Guiguinto possessed the intrinsic marketable property called *quantity premium*, which incorporated future use for the lot to the size of the parcel such as further subdivision or building large houses (Tabuchi, 1996). This means the bigger the lot, the higher quantity premium it possesses and therefore, the higher its market value per area.

3.2. Mapping predicted Market Values using GIS

Upon assigning values of the significant predictors to each of the parcels in Guiguinto, the regression analysis was applied to the parcels to obtain an estimate of the feasible market values based on the empirical results. A map of the predicted Market Values of parcels in Guiguinto is provided below.

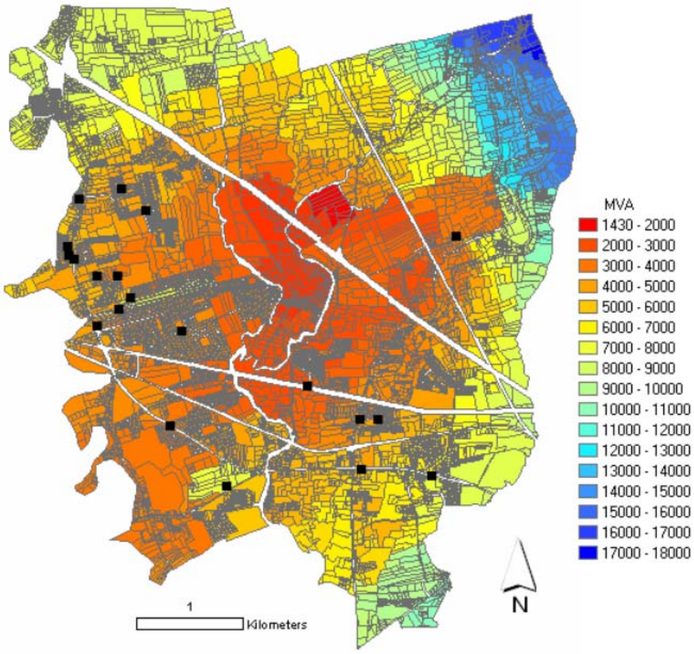
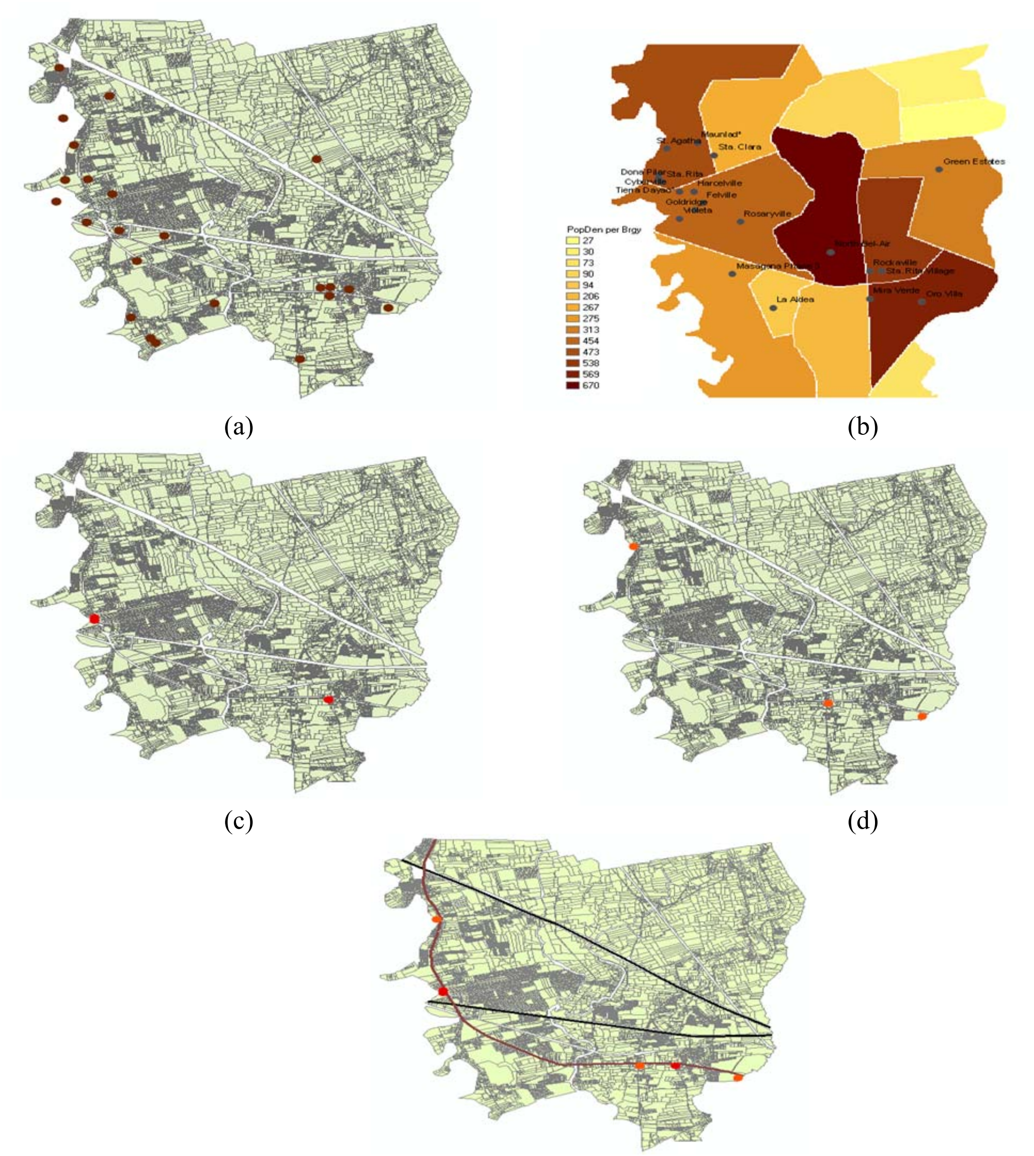


Figure 1. Predicted Market Values per Unit Area of parcels in Guiguinto, Bulacan. Legend shows the range of the prediction, from as low as P1,430 per sq. m. in the central portion, to as expensive as P17,500 in the northeast.

A number of patterns regarding the fluctuation of lot market prices emerged. Contrary to the conventional outward radial decrease of lot market price with the increase of distance from the geographic center of the municipality, a reversal of trend was observed in the case of Guiguinto. Residential location choice usually assumed a monocentric model of the municipality development, where the municipal center as the hub of activities is assumed as a single point located at the geographical center (Mäler, 1977). A deviation from this assumption was seen in the study area, as the Land Use Plan of Guiguinto revealed residential – commercial zones traversing the McArthur Highway, but mostly concentrated on two key portions: the northwestern and southeastern areas.

Since the minimum lot sizes for parcels to be developed are assumed to take the value of *60 square meters*, only the four other predictors vary in value. Looking at the number of industrial sites within 4 kilometers, the locations of factories and other industries showed a linear pattern running from the northwestern part of Guiguinto down to southwest continuing up to the southeast. These sites appeared to envelop the central portion of the municipality; therefore that portion became surrounded by almost all the industrial sites in the area, bringing down the prices of parcels around the center.



(a) (b)

(c) (d)

(e)

Figure 2. Spatial distribution of the significant Market Value predictors. **a)** Location of industrial sites in brown points. **b)** Population density per barangay. Gray points are the subdivision entrances. **c)** Location of banks in red points. **d)** Location of marketplaces in orange points. **e)** Configuration of NLEX and McArthur Highway with respect to the location of amenities. NLEX is depicted as black, while McArthur is in maroon.

Regarding the distances to the nearest banks and marketplaces, the locations of these establishments were seen to be concentrated along McArthur Highway, specifically at the northwest, west and southeast of Guiguinto. Another observation on the predicted MVA was prices of lots along the national highway (NLEX) were relatively cheaper compared to lots along the McArthur Highway. All of the amenities considered are situated along the McArthur Highway. We can see that the distance of lots along NLEX to these significant amenities are larger than the lots located along the McArthur Highway. Thus, lower predicted MVA are computed for lots on along NLEX than lots located along McArthur Highway.

Also, the population density of the barangay occupying this portion was found to be high,

bringing the lot prices here lower further. The upper right part, with very high lot prices, was seen to have scarcely – populated barangays.
Chatterjee and Price (1977) warned about using the prediction model in areas with values outside of the observation range of the predictors' dataset that was used in the generation of the regression equation. Upon applying the said restrictions regarding the observation ranges, the number of parcels applicable for Market Value prediction went down from 16,029 to 12,808. This step eliminated most of the parcels on the northeastern side of Guiguinto which registered lot prices from values of a little below P10,000/sq. m. to as high as P16,000/sq. m. The range of predicted Market Values of parcels was trimmed down, as shown by **Figure 3**.

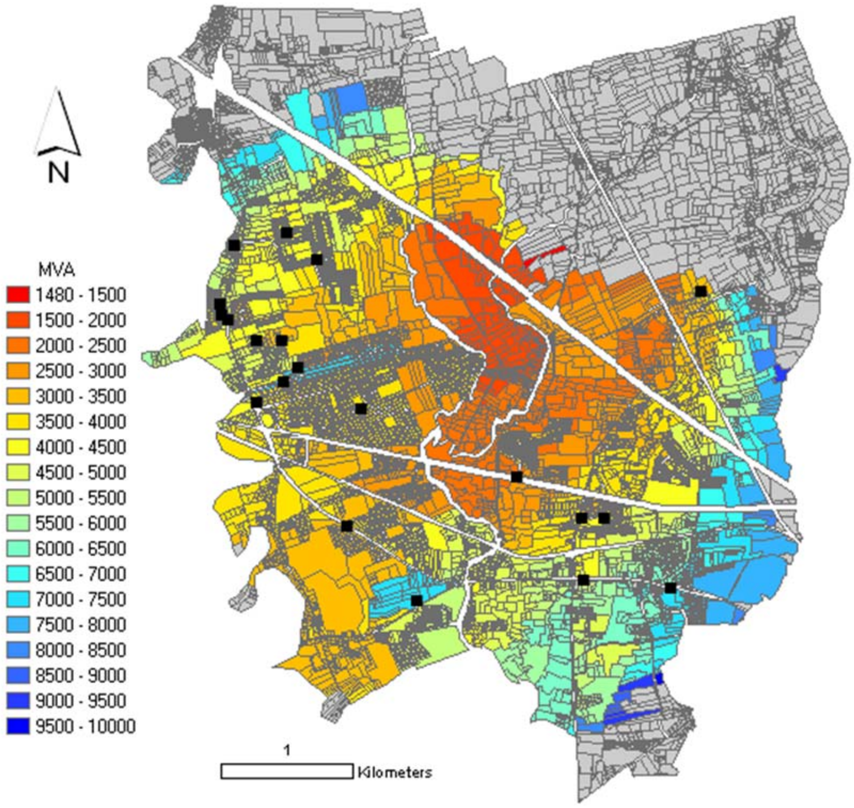


Figure 3. Parcels within the observation range of the predictor values used in generating the regression equation. A large portion of the upper right area of Guiguinto was removed in the process, cutting in half the range of the predicted Market Values (legend).

3.3. Factors involved in the Utility Function

The Multiple Linear Regression for the Market Value prediction were analyzed, considering that the operational form of the Bid-Rent Model will be formulated from the resulting regression equation. Creation of a criterion for feasibility of a subdivision development in an area was made by isolating the location-based predictors from those which are intrinsic to the development. Among the significant predictors, only the *minimum lot size* (LA) can be considered as what can incur development costs. The four other factors were then transposed to the other side of the equation, and *minimum lot size* LA and the *Market Value* MVA are retained together with the intercept. From **Equation 17**, the Bid-Rent Equation becomes

$$589.17*IND4 + 42.50*\sqrt{DB} + 1138.75*\ln PDB + 169.26*\ln DMP \geq 24118.75 + 42.50*LA - MVA \quad \text{(Eq. 17)}$$

where the terms involving the *number of industrial sites* IND4, *distance to nearest bank and marketplace* DB and DMP, and the *population density per barangay* PDB make up what is called the Utility expression for a given parcel with a designated location *n* in the Guiguinto area

$$U_n = U(IND4_n, DB_n, PDB_n, DMP_n). \quad \text{(Eq. 18)}$$

This form can be taken as a functional form of the *Budget constraint* of the development project when choosing among locations for a suitable residential site. However, a reversal of the usual dynamics was seen with *U*. While previous researches established the Utility of a commodity as a *good*, where an increase in the value of its characteristics result to a desirable outcome (Rosen, 1974), a higher value for *U*'s expression means less convenience for the occupants, as more industrial sites nearby and greater distances from amenities negatively affect living conditions in *n* according to literature review. Therefore the function U_n can be more appropriately called the *Counter-Utility Function*. It will be easier to think of the monetary equivalent of the Counter-Utility Value as the aggregate transportation costs that the household spends by traveling to desired locations and mitigating other undesirable conditions assuming the household chooses to live in location *n*.

The right-hand side of the Budget constraint equation quantifies the *Willingness-to-pay value*, or the maximum monetary amount that the development project finds reasonable in acquiring the site. Dissecting the expression into its terms, the combined first and second terms, $24118.75 + 42.50*LA$, reflect the costs of acquiring, marketing and selling the parcel (per unit area) by the development project. This value varies depending on the minimum lot size *LA* planned to be subdivided by the project, and this term is also expected to capture the differences in the design and the social strata of the development, since according to *quantity premium,* larger lot sizes possess additional intangible value, such as the capacity to accommodate opulent structures (Tabuchi, 1996). The – *MVA* term now represents the part of the development costs that the future occupants will be willing to shoulder. With respect to the development project, it represents the revenues coming from the sale of the lot, giving the expression $24118.75 + 42.50*LA – MVA$ the development costs after sales. A limitation of the development costs expression is that this side of the Budget constraint equation is only a part of the gross development costs of developing the parcel. There is no mention of the construction costs for structure in any part of the methodology employed, nor was there any mention of its sales. Another limitation is from the side of the Utility function where there is no way that the Utility function will incur a negative cost.

Now that limitations have been acknowledged and assumptions have been stated, the dynamics of the occupants' willingness-to-pay, *MVA*, depending on the environmental and neighborhood attributes of a residential site *n* can be analyzed. Consider the scenario where the Counter-Utility is apparent, meaning U_n is high and the living conditions are inconvenient. When U_n changes, the value of $24118.75 + 42.50*LA – MVA$ also adjusts. And since the gross development costs $24118.75 + 42.50*LA$ has already been incurred by the project, it remains constant leaving the term – *MVA* the only adjustable term. When living conditions are inconvenient, U_n increases, and the value – *MVA* becomes less negative, resulting to a lower market value that the future occupants will be willing to pay for the lot. Under a different situation where the living conditions become suitable such that U_n decreases, – *MVA* becomes more negative reflecting a higher market price because future occupants will find the place more desirable to live in, they will be willing to pay more for it.

3.4. Mapping Counter-Utility function values

Since it had been established that the Counter-Utility value affects the willingness-to-pay of occupants and that it also dictates the projected development costs incurred by the project, the expression was applied to the parcels to obtain the value of Counter-Utility. Examining the frequency distribution of the Counter-Utility values in **Figure 4**, the 12,808 parcels in Guiguinto have a mean Counter-Utility value of 22,438 in monetary units, and a standard deviation of 1,256. Looking at the maximum value of Counter-Utility, it is said to be at 25,188, which is just over two standard deviations from the mean. The minimum Counter-Utility value on the other hand was at 17,033, more than four standard deviations below the mean value. This presented the skewed character of the parcel population, as what can be seen from the Frequency Distribution graph, with the mean is inclined slightly to the right.

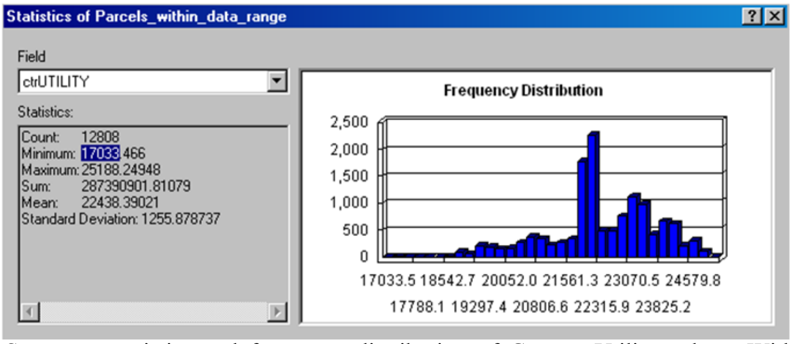


Figure 4. Summary statistics and frequency distribution of Counter-Utility values. With the mean skewed toward the right, Counter-Utility values less than the mean ($\mu_{Counter\text{-}Utility} = 22{,}438$) are spread over a vast range of values.

Presenting the parcels with their Counter-Utility values in a map alone could be misleading, as the spread of the distribution is inclined to the right. Parcels labeled as below $\mu_{Counter\text{-}Utility}$ (mean Counter-Utility Value at mapping is 22,740) would still be above the median value of 22,551. To normalize the frequency of values below $\mu_{Counter\text{-}Utility}$ and maximize the difference among values above $\mu_{Counter\text{-}Utility}$, the map for the Counter-Utility Function Values of parcels in Guiguinto, Bulacan was color-classified using the *Natural Breaks (Jenks)* Method, where the big jumps among values above $\mu_{Counter\text{-}Utility}$ were magnified, and parcels with Counter-Utility values below $\mu_{Counter\text{-}Utility}$ had been more widely – distributed. A map of the Counter-Utility values of parcels in Guiguinto is shown on **Figure 5.**

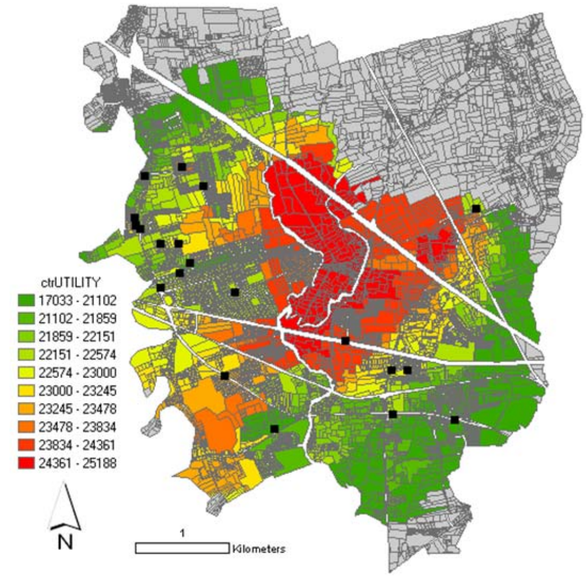


Figure 5. Counter-Utility Function Values of parcels in Guiguinto, Bulacan. Legend shows the Counter-Utility values with monetary units. This monetary value can be thought of as the aggregate transportation costs that a household spends within a certain period by traveling to desired locations and mitigating other undesirable conditions. Green to yellow parcels are viewed as more desirable locations than orange to red areas.

After applying the zoning restrictions regarding the parcels labeled as *Protected agricultural lands*, the following map is produced.

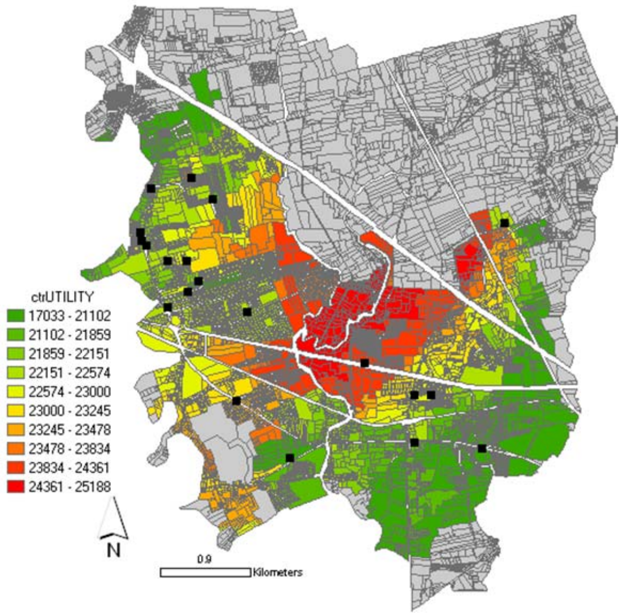


Figure 6. Counter-Utility Function Values of parcels in Guiguinto, with the development restrictions applied.

4. CONCLUSION

Using multiple linear regression, the following factors have been analyzed to negatively affect subdivision lot market values in Guiguinto, Bulacan: numerous industrial sites within a 4 –

Km radius from subdivisions, farther distances from amenities such as banks and markets, and higher population density. Meanwhile, larger lot sizes were shown to have positive correlation with subdivision lot market prices. These variables were seen as predictors of market values of subdivision lots with 95% confidence level. This result provides for the capability of the Market Value Prediction equation to model the different variables that influence differences among market values of the subdivisions in this study.

It was also shown in the study that the regression equation for predicting market values can be formulated to follow the Bid – Rent theory. The study was able to produce a Budget constrain equation relating possible factors that incur costs to the development project with factors that describe a lot's environmental and neighborhood qualities. A Counter-Utility Function, which was derived from the environmental and neighborhood side of the Budget constraint equation, is an expression that can be used to assess the level of utility of a site as a residence. The Counter-Utility Function values of individual parcels can be plotted in a map to locate which areas in a study area represent relatively higher utility than others. However, the study was able to show only one area where the applicability of the Counter-Utility Function can be used. To solve this issue, the validity of transforming the regression equation for predicting market values into an operational form of the Bid-Rent theory can be further established by applying the concept to site selection studies in other commercial and industrialized municipalities.

In dealing with these analyses that require spatial – related processes, Geographic Information System has been an indispensable tool for accomplishing numerous tasks. Measuring distances, counting nearby features and summarizing information from geographic datasets are easily done without the mechanical intervention of humans, making such tasks free of bias. These operations are also linked to common data gathering techniques, such as survey plans and satellite positioning system readings, which makes the transformation of data from its raw to processing up to presentation form readily executed in a single platform. Representation of the Market Values per Unit Area and the Counter-Utility Function Values of parcels in maps using GIS made the process of understanding the fluctuation of these values and the patterns emerging from the changes in values easier to visualize and derive conclusions from.

5. RECOMMENDATIONS

To further improve the model, a sensitivity analysis regarding the estimated regression coefficients of the Market Value predictors can be performed to refine the understanding of the effects of the predictors to the regression equation. This can increase the applicability of the model to study areas near and similar to Guiguinto, Bulacan. Since a part of the Budget constraint equation is devoted to the costs incurred by the development project, additional predictors relating to design and construction of the parcel can be included in the regression.

Some other practical application of this capability aside from the specific objective of this study is that the resulting model can facilitate the mass valuation procedure, where the general trend of the lot market values for a specific type of land use can be easily visualized by a person with limited experience in valuation. A more experienced assessor can likewise use the model for checking the results of his assessment for a particular parcel. As the modeling procedure is enhanced through further fine-tuning, more applications in the field of formulation of regulatory practices and other bureaucratic functions can be found.

REFERENCES

Alonso, W. (1964). Location and Land Use. Massachusetts: Harvard University Press.

Carao, R., Hernandez, A. (1992). 'Determinants of Urban Residential Land Values: Makati', Quezon City: University of the Philippines, Diliman.

Chapin, F. S. Jr. and Weiss, S. F. (1962). *Factors influencing land development: an urban studies research monograph*, Chapel Hill: Chapel Hill Press.

Chatterjee, S., Price, B. (1977). Regression Analysis By Example, New York: Wiley.

de Vor, F., de Groot, H. L. F. (2009). 'The Impact of Industrial Sites on Residential Property Values: A Hedonic Pricing Analysis for the Netherlands', *Tinbergen Istitute Discussion Paper*, 35(3).

Dion, T. R. (2002). Land Development for Civil Engineers, New York: Wiley.

Gao, X., Asami, Y. (2005) 'Influence of Spatial Features on Land and Housing Prices', *Tsinghua Science and Technology*, 10(3), pp. 344 – 353.

Kone, D. L. (1994). Land Development, Washington, DC: Home Builder Press.

Larsen, P. V. (2008) 'Inference in Regression', Retrieved 30 July 2010 from the World Wide Web: http://statmaster.sdu.dk/courses/st111/module03/index.html.

Mäler, K. (1977). 'A Note on the Use of Property Values in Estimating Marginal Willingness to Pay for Environmentalt Quality', *Journal of Environmental Economics and Mangement*, 4, pp. 355 – 369.

Ranjit, K. P. (2006) 'Multicollinearity: Causes, Effects and Remedies', Retrieved 20 July 2010 from the World Wide Web: http://www.iasri.res.in.

Rivera, M. A. I., Tiglao, N. C. C. (2006) 'Investigating the Effects of Accessibility on Zonal Land Valuation and Residential Location Choice in Metro Manila', *Symposium on Infrastructure and the Environment*.

Rosen, S. (1974). 'Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition', *The Journal of Political Economy*, 82(1), pp. 34 – 55.

Tabuchi, T. (1996). 'Quantity Premia in Real Property Markets', *Land Economics*, 72(2), pp. 206 – 217.

Tamayo, G. E., Salvador, A. S. (1990). 'A Study on the Determinants of Urban Residential Real Property Values', Quezon City: University of the Philippines, Diliman.

CONTACTS

Mr. Ernest John C. Bugay
Student, Dept. of Geodetic Engineering
MH 224, College of Engineering, University of the Philippines, Diliman
Quezon City
PHILIPPINES
Email: ecbugay@up.edu.ph

Mr. Jolu C. Bunda
Student, Dept. of Geodetic Engineering
MH 224, College of Engineering, University of the Philippines, Diliman
Quezon City
PHILIPPINES
Email: jcbunda@up.edu.ph

Mr. Louie P. Balicanta
Instructor, Dept. of Geodetic Engineering
MH 224, College of Engineering, University of the Philippines, Diliman
Quezon City
PHILIPPINES
Tel +63 2 981-8500 loc. 3126
Fax +63 2 920-8924
Email: lpbalicanta@up.edu.ph or louie_balicanta@yahoo.com
Website: dge.upd.edu.ph

Asst. Prof. Matthew Oliver Ralp L. Dimal
Dept. of Geodetic Engineering
MH 224, College of Engineering, University of the Philippines, Diliman
Quezon City
PHILIPPINES
Tel +63 2 981-8500 loc. 3126
Fax +63 2 920-8924
Email: mldimal@up.edu.ph or matthewdimal@gmail.com
Website: dge.upd.edu.ph

Assoc. Prof. Enrico C. Paringit, D Eng
Dept. of Geodetic Engineering
MH 406, College of Engineering, University of the Philippines, Diliman
Quezon City
PHILIPPINES
Tel +63 2 981-8500 loc. 3147
Fax +63 2 920-8924
Email: ecparingit@up.edu.ph or paringit@gmail.com
Website: dge.upd.edu.ph