LAND PRICE REGRESSION MODEL AND LAND VALUE REGION MAP TO SUPPORT RESIDENTIAL LAND PRICE MANAGEMENT: A STUDY IN NGHE AN PROVINCE, VIETNAM

(1)**Pham Thi Ha**
School of Agriculture and Resources
Vinh university in Vinh, Nghe An, Vietnam
e-mail: phamha6868@gmail.com

(2)**Nguyen Tran Tuan**
School of Agriculture and Resources
Vinh university in Vinh, Nghe An, Vietnam
e-mail: nguyentrantuandhv@gmail.com

(3)**Nguyen Van Quan**
Faculty of Land Management
Vietnam National University of Agriculture, in Hanoi, Vietnam
e-mail: nvquan@vnua.edu.vn

(4)**Nguyen Van Trung**
Faculty of Geomatics and Land Administration
Hanoi University of Mining and Geology, in Hanoi, Vietnam
e-mail: nguyenvantrung@humg.edu.vn

Abstract

The real estate market in areas with many socio-economic activities needs to be strictly managed due to the difference between the market price of urban land and the price of land set by the state. This study identifies and analyzes the influence of some factors on land prices in peri-urban areas to develop land pricing standards consistent with the price level in Nghe An province. The study surveyed 362 land users and 200 samples of successfully transferred properties in the study area. Based on the multivariate regression method, the study builds a residential land price model and calculates the market price of residential land. The authors also established a map of land value areas to help State agencies manage land prices effectively. The research serves as a basis for State agencies to study the formation and development of the real estate market to develop appropriate land price management measures.

1. Introduction

Since implementing the Doi Moi policy, Vietnam has enjoyed strong economic growth, significantly reducing poverty from 58% in 1993 to 2.8% in 2020 (GSO, 2021). However, Vietnam is still a country with most of the population living in rural areas, and two-thirds of the population working in the agricultural sector; labor productivity remains low (Nguyen, 2021a). In some countries, urbanization has been used as a tool to promote economic growth and reduce poverty (Arouri et al., 2014; Kuddus et al., 2020). If Vietnam maintains a high growth rate, supporting urbanization, in which cities contribute significantly to job creation and Gross Domestic Product, it will be an essential measure (World Bank, 2020). This structural change will cause population and housing demand to increase in cities, whereby good quality affordable housing solutions in reasonable serviced settlements will be

essential (World Bank, 2015). In other words, land price is the bridge between the relationship of land - the market - the management of the state; it is an economic tool for land managers and users to access the market mechanism (Tran Tuan, 2021a). At the same time, it is also the basis for assessing fairness in land distribution.

Vietnam has also gone through various phases of housing policy. Before 1988, Vietnam's formal housing sector was managed through a centralized planning regime, but Doi Moi policy changed to a market-oriented one (Shanks et al., 2004; Tran & Yip, 2019). This change caused the area to proliferate. However, this market orientation offers almost no solution to facilitate the access of the poor and near-poor to housing (Thanh et al., 2013). Growth driven by foreign direct investment and speculation has driven house prices up (Khanh, 2021). One of the important reasons leading to the above shortcomings is that, although Vietnam has a complete information system and land price database, it is not close to market prices (Vietnam has a two land price systems, including the state land price and the market land price). This system has not yet met the reliability necessary to serve the market management and society's development needs. In order to overcome these limitations, it is essential to determine land price reasonably and close to its market value (Tran Tuan, 2021b). In other words, it is necessary to develop a suitable land price list using a land valuation tool. Land valuation is not only intended to deal with individual valuation cases, but it is also used to carry out a mass valuation on a large scale (Demetriou, 2016; Cauto et al., 2021).

According to the international association of appraisers (IAAO, 2013), bulk valuation is the process of valuing a group of properties at a given time, applying common data, standardized methods, and check statistics. Land pricing models apply three approaches to value: a cost approach, price comparison approach, and an income approach (Wincott, 2001). In land valuation, multivariate linear regression analysis (MRA) is one of the best-known statistical approaches with many applications, especially forecasting land prices from regression models (Benjamin et al., 2004). The land price regression model uses factors affecting land price as independent variables to calculate valuation (Alimudin et al., 2017; Karakayaci, 2018). The fundamental factors that affect land prices commonly used in studies include (1) the location of the parcel of land, (2) the distance to important sites, (3) the characteristics of the land plot, and (4) the land plot's environmental and security conditions.Some macro factors, such as the economic and financial situation of countries, also impact real estate prices (Renigier-Bilozoz and Wiśniewski, 2013).

There are quite a few countries with applied methods that use computational models to determine land value in the world. Since 1963, Bailey et al. (1963) introduced a method to determine the value of a property based on a linear regression function. Until 1964, Alonso, a famous researcher in real estate, suggested that the value of a property depends on its location (Alonso, 1964). In 1966, in customer theory, Lancaster proposed a model according to which an asset's value depends on some characteristics of that asset (Lancaster, 1966). By 1974, Rosen applied the ideas of Bailey and Lancaster to come up with a mathematical model called the Hedonic model to determine the value of a product and analyze the equilibrium value in the market for that product (Rosen, 1974). During the following period, many studies applying the Hodenic model to the determination of real estate prices were carried out by researchers in different countries, such as Sweden (Englund et al., 1998), America (Zhou & Sornette, 2008), and France (Gouriéroux & Laferrère, 2009).

Meanwhile, a study on land price forecasting in India using neural network techniques and multiple regression sheds light on the spike in land prices in the southern and western regions of Chennai Urban Area (Sampathkum et al., 2015). In Indonesia, a study aimed at predicting house prices in Malang city with regression analysis and Particle Swarm Optimization (PSO) was optimal with the lowest error (Alfiyatin et al., 2017). In Guatemala, a multivariable regression model has shown that the land value in this city has a difference of up to 253% after approaching the central land (Morales et al., 2019). In Kuwait, land prices were determined using traditional spatial regression methods showing the effects of population density, educational facilities, and air pollution levels (Mostafa, 2018).The use of regression analysis, in which independent variables were analyzed and used to predict real estate sales prices in the Belmont and Eastside neighborhoods of Pueblo, proved a success because it estimated the neighborhood's exact selling price in 2017 (Mize, 2017). Kolebe et al. 2019 also estimated land prices by regression method in Germany. They concluded that the method could generate land value estimates that were consistent with estimates from the experts' point of view. From the above analysis, it can be seen that the use of a regression model is a possible way to establish a better estimate of the actual value of land (Mize, 2017). Therefore, this article will develop a

land price regression model and land value region map to assist authorities in managing residential land prices in a locality in Nghe An province.

2. Data and Methodology

Nghe An is a province located in the center of the North Central region with a 419 km long border with the Lao People's Democratic Republic in the west and 82 km in the east with a coastline. Nghe An is also located in the economic corridor connecting Myanmar - Thailand - Laos - Vietnam - East Sea to Cua Lo port. This position gives Nghe An a vital role in domestic and international economic exchanges. It also creates many favorable conditions for Nghe An in calling for investment in socio-economic development.

With the main purpose of building a regression model of residential land prices in Nghi Tan ward - Cua Lo town - Nghe An to create a map of residential land value areas, the main research methods used include:

2.1. Secondary data collection method

Secondary data such as land conditions, land use and management status, and land prices in Cua Lo Town were collected from 2015 to 2020. Some data on the socio-economic development report was provided by the Department of Natural Resources & Environment of Nghe An province.

2.2. Primary data collection method

Data collection was carried out for land users in the study area to collect land prices traded on the market and their influence on people's psychology. In order to evaluate the factors affecting the price of residential land in urban areas in Nghi Tan ward, the study calculates the number of samples according to research by Barbara G Tabachnick and Linda S Fidell (2013) with the formula n ≥ 8 x 39 + 50 = 362 samples (39 is the number of independent variables in the model shown in Table 1). In addition, in order to build a forecasting model to determine urban land prices in the area, the study conducted a survey of 200 samples of successfully transferred real estate in Nghi Tan ward, with a sample number based on a formula by Barbara G Tabachnick and Linda S Fidell (2013) n ≥ 8 x 18 + 50 = 194 samples (18 is the number of independent variables shown in Table 5). These surveys were conducted based on a combination of two methods, cluster and random. The cluster method was applied in selecting properties located on main roads and evenly distributed over the study area. For samples with the same route, the sample was selected based on factors affecting land prices, such as road width and land width. After conducting cluster classification, the authors randomly selected households for the survey.

Meanwhile, to assess the factors affecting the price of residential land, the research team conducted a staff interview survey using the "semi-structured" method. These interviews aimed to assess factors (39 factors in Table 1) on residential land prices of 32 land lots traded in the Nghi Tan ward. An additional five interviews were conducted with staff from the Department of Natural Resources and Environment, the Department of Natural Resources and Environment, the land registration office, a real estate investor, and a land broker.

2.3. Data analysis methods

- **Factor analysis:** Factor analysis is carried out in two stages.
  Stage 1: Building and verifying the quality of the scale.
  Stage 2: Exploratory Factor Analysis (EFA) includes the following processes: Checking the appropriateness of the model, extracting factors, rotating factors and making decisions to keep, name the factor.
  The suitability test for factor analysis is based on the following criteria. KMO (Kaiser-Meiser-Olkin) criterion is an indicator to consider sufficient sample size and correlation between variables. If $0.5 < KMO < 1$, factor analysis is appropriate. The Bartlett test is a statistical quantity used to test the hypothesis that the variables do not correlate in the population, the analysis is only used when the hypothesis is rejected ($p < 0.05$) and there is a correlation between the variables. The extraction of factors is usually based on Eigenvalues. Precisely, the study only retains factors with Eigenvalue>1.
- **Regression analysis method:** The land price regression model used in the study has the form: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + ... + \beta_n X_n$.

Where: Y is the dependent variable representing the price of the land plot; X1; X2; .... Xn are independent variables that affect land prices; β1; β2; ..... βn are regression coefficients showing the impact of factors affecting land prices.

3. Empirical results

3.1. Identifying some factors affecting the price of residential land

Through research, study, and a survey of land prices on the market in Nghi Tan ward, the results show that the price of residential land in this area is affected by many factors that cause the transaction price on the market to always diverge greatly from the price regulated by the State. The impacts ofthese factors' are not the same in terms of scale and level, but each factor affects a different aspect. Most of the factors in each group of factors impact the land price level in the ward. Although the rate of impact varies among factors, they contribute to creating the difference between the market price and the price regulated by the State in the locality. Based on the factors affecting land prices in the Town, this study presents a model consisting of 7 scales representing the factors affecting land prices, with 39 observed variables as shown in Table 1.

3.1.1. Checking the reliability of the scale (Cronbach's Alpha coefficient)

According to the analysis of testing the scale, the scale's overall Cronbach's Alpha value is guaranteed according to the set standards (Cronbach's Alpha > 0.6). In addition to the Cronbach Alpha standard, we also consider the correlation coefficient of the total variable (Corrected Item - Total Correlation). According to the standard, any coefficient < 0.3 is discarded. The specific test results are as follows:

Table 1

Cronbach's Alpha test results

No	Categories	Code	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
1	Distance to center	VT1	29.920	20.944	0.025	0.647
2	Distance to Market	VT2	30.406	18.724	0.422	0.670
3	Distance to School	VT3	29.696	17.182	0.461	0.652
4	Distance to Bus station	VT4	30.381	20.835	0.080	0.630
5	Distance to Hospital	VT5	30.434	22.606	-0.133	0.652
6	Distance to Beach	VT6	30.185	18.628	0.441	0.667
7	Type of road adjacent to the land plot	VT7	27.826	20.449	0.207	0.607
8	Distance to People's Committee	VT8	30.381	19.073	0.385	0.678
9	Distance to police station	VT9	29.646	16.728	0.507	0.640
10	Distance to post office	VT10	30.506	21.314	0.130	0.616
11	Distance to bus stop	VT11	30.080	19.753	0.234	0.602
12	Distance to shopping mall	VT12	30.204	18.734	0.440	0.668
13	Distance to landfill	VT13	28.246	18.884	0.204	0.614
14	Economic growth rate	KT1	13.116	14.097	0.449	0.801
15	Supply - demand for land in the market	KT2	13.398	11.758	0.679	0.747
16	Income and consumption of the population	KT3	15.227	18.048	0.115	0.838
17	Animal price volatility	KT4	13.669	11.513	0.792	0.717
18	Interest rates	KT5	13.525	12.305	0.698	0.743
19	Business environment	KT6	13.276	12.871	0.586	0.770
20	Urbanization	XH1	22.843	10.327	0.230	0.750
21	Real estate speculation	XH2	22.387	9.257	0.591	0.666
22	Population density	XH3	22.572	9.459	0.476	0.690
23	Health and education issues	XH4	22.602	8.534	0.616	0.652

24	Education level of the population	XH5	22.845	8.547	0.596	0.656
25	The problem of feng shui	XH6	22.732	8.934	0.561	0.668
26	Social Security	XH7	22.390	11.712	0.052	0.771
27	Air quality	MT1	7.519	2.394	0.448	0.704
28	Water quality	MT2	7.721	1.825	0.604	0.509
29	Sound environment	MT3	7.727	2.127	0.529	0.609
30	Electricity - water system	HT1	6.564	1.887	0.574	0.601
31	Communication systems	HT2	6.873	1.757	0.539	0.625
32	Transportation system	HT3	6.530	2.017	0.497	0.671
33	Shape of the land	CB1	9.196	12.103	0.915	0.892
34	Area of the land	CB2	9.575	12.356	0.870	0.908
35	Depth of the land	CB3	9.544	13.711	0.847	0.915
36	Width of the land	CB4	9.541	15.113	0.776	0.938
37	Legal status of the land	PL1	8.193	1.913	0.528	0.636
38	Policy on land use	PL2	8.290	1.641	0.751	0.632
39	Restrictions of land use planning	PL3	8.334	2.495	0.361	0.812

Source: Calculated by the authors.

According to the analysis of the scale test, the overall Cronbach's Alpha value of all the scales has a Cronbach's Alpha coefficient of the population greater than 0.6. However, the correlation coefficient of the total variable Corrected Item - Total Correlation of 10 observed variables is not qualified when the value < 0.3 is: VT1, VT4, VT5, VT7, VT10, VT11, VT13, KT3, XH1, XH7. The remaining 24 variables have the standard correlation coefficients of all variables > 0.3. Thus, the scales formed according to the above analysis results ensure good quality of the research.

3.1.2. Exploratory Factor Analysis (EFA)

After testing the scale's reliability, the residential land value scale with 24 observed variables in the previous section continued to be analyzed for factors to determine the relevant factors and variables. By analyzing and verifying the scale's quality and the EFA model's tests, we identified seven groups of factors representing 24 measurement variables for factors affecting land prices in the ward. The results are summarized in Table 2.

Table 2

Adjusted model through Cronbach's Alpha and EFA testing

No.	Code of category groups	Characteristic variables	Element group name	Cronbach's Alpha coefficient after correction
1	X_1	VT2, VT3, VT6, VT8, VT9, VT12	Location	0.904
2	X_2	KT1; KT2; KT4; KT5; KT6	Economy	0.923
3	X_3	XH2; XH3; XH4; XH5; XH6	Society	0.875
4	X_4	MT1; MT2;MT3	Environment	0.843
5	X_5	HT1; HT2; HT3	Infrastructure	0.798
6	X_6	CB1; CB2; CB3; CB4	Particular	0.799
7	X_7	PL1; PL2; PL3	Legal policy	0.902

Source: Calculated by the authors.

3.1.3. Evaluating the influence of factors by regression model

The variables with statistical significance through the regression model tests include X1, X2, X3, X4, X5, X6, X7. These variables have the theoretical ability to influence the determination of land prices. Based on the normalized regression coefficients, these variables can be converted to percentages and arranged in order of priority from highest to lowest, as shown in Table 4.

Table 3

Results of regression coefficients

Model	Unnormalized regression coefficients		Normalized regression coefficient	Test t	Significance level	Multicollinear Statistics	
	Coefficient B	Error	Beta			Accept	Variance Inflation Factor (VIF)
Constant	0.141	0.079		48.864	0.000		
x_1	0.425	0.008	0.399	41.442	0.000	0.987	1.044
x_2	0.258	0.006	0.254	-11.354	0.000	0.998	1.032
x_3	0.043	0.076	0.101	.3,632	0.000	0.935	1.065
x_4	0.089	0.007	0.089	-.954	0.000	0.976	1.076
x_5	0.401	0.008	0.358	4.098	0.000	0.999	1.034
x_6	0.099	0.023	0.201	-.456	0.000	0.997	1.098
x_7	0.173	0.003	0.168	-3.001	0.000	0.986	1.023

Source: Calculated by the authors.

Table 4

Influence of factors on land prices

Influential group of factors	Normalized regression coefficient	Percentage (%)	Order of influence
X1 – Location	0.425	28.56	1
X5 – Infrastructure	0.401	26.95	2
X2 – Economy	0.258	17.34	3
X7 – Legal policy	0.173	11.63	4
X6 – Particular	0.099	6.65	5
X4 – Environment	0.089	5.98	6
X3 – Society	0.043	2.89	7
Total	**1.488**	**100.00**	

Source: Calculated by the authors.

Through the tests, it is possible to confirm that seven groups of factors affect the land price in Nghi Tan, which are considered the strong points. These groups of factors are statistically significant and ranked in order of importance, as shown in Table 4.

3.2. Building a residential land pricing model in urban areas in Nghi Tan ward

3.2.1. Defining variables for the model

In general, the price of a land parcel (which may include buildings on the land parcel) in the regression model depends on the characteristics of the land plot (like the location compared to the center and near the utility areas) and the value of the buildings on that land parcel (like the house area, number of bedrooms, number of floors). The goal of the model is to determine an equation for the price of the land plot based on the above characteristics that are closest to the market price. Models for determining prices can be simple models, such as linear models, or more complex models, such as exponential and logarithmic models. Thus, the selection of these models will be evaluated and applied according to each data set. In this study, the author used a multivariable linear regression model.

By surveying the current situation, as well as researching and evaluating groups of factors affecting urban land prices in Nghi Tan ward, the authors divided the 24 factors into seven groups. The study evaluates the influence of the selected groups of factors on land prices and excludes the variables that have a minor influence on land prices. Then, the topic proceeds to reformat the variables shown in Table 5.

Table 5

Variable format for land price regression model

No.	Variable symbol	Description	Type	Unit	Estimated sign
A		**Dependent variable**			
1	G_DAT	Land price	Quantitative	Million VND/m2	
B		**Independent variables**			
1	KC_TT	Distance to Center	Quantitative	m	-
2	KCTI_1	Distance to utilities (school, markets, hospital…)	Dummy variable	1 = Distance to utilities is between 0 – 1,000m. 0 = Distance to the utilities in the remainder.	+
3	KCTI_2	Distance to utilities (school, markets, hospital…)	Dummy variable	1 = Distance to utilities is between >1,000m – 2,000m. 0 = Distance to the utilities in the remainder.	+
4	D_TICH	Area of land	Quantitative	m^2	+
5	H_The	Shape of land	Qualitative	1 = good 0 = not good	+
6	CR_MT	Width of frontage of land	Quantitative	m	+
7	CR_DUONG	Road width	Quantitative	m	+
8	CL_DUONG	Road quality	Qualitative	1 = good 0 = normal	+
9	TT_TL	Communities	Qualitative	1 = good 0 = normal	+
10	DIEN_NUOC	Water and electricity system	Qualitative	1 = good 0 = normal	+
11	MTST_TOT	Good ecological environment	Dummy variable	1 = Good ecological environment 0 = Ecological environment in the remainder.	+
12	MTST_BT	Normal ecological environment	Dummy variable	1 = Normal ecological environment 0 = Ecological environment in the remainder.	+
13	AN_TOT	Good security	Dummy Variable	1 = Good security 0 = Security in the remainder.	+
14	AN_BT	Normal security	Dummy Variable	1 = Normal security 0 = Security in the remainder.	+
15	MTKD_TOT	Good business environment	Dummy variable	1 = Good business environment 0 = Business environment in the remainder.	+
16	MTKD_KHA	Normal business environment	Dummy variable	1= Normal business environment 0 = Business environment in the remainder.	+
17	P_LY	Legality of the land	Dummy	1 = Full legal documents	+

		variable	0 = Incomplete	
18	Q_HOACH	Land use planning	Dummy variable	1 = Hanging planning	+
				0 = Not hanging planning	

Source: Calculated by the authors.

3.2.2. Model building

It is possible to build many models to determine the dependent variable Y for each model with different independent variables Xi. The question is which of the proposed models is the best. Usually, choosing the best model is based on the coefficient of determination R2. The higher this index, the better the model. However, it should also be noted that each regression model has many attributes. In order to evaluate model quality, it is necessary to consider those attributes simultaneously.

a. Land price regression model for the first time

Running the regression model for the first time with all 18 independent variables, we get the resuls in Table 6.

Table 6

Results of the first regression model

Model	Unnormalized regression coefficients		Normalized regression coefficient	Test t	Significance level	Multicollinear Statistics	
	Coefficient B	Error				Accept	VIF
(Constant)	3323.785	5312.333		0.626	0.551		
KCTT	-1.461	2.437	-0.096	-0.599	0.000	0.039	5.449
KCTI_1	-189.078	495.534	-0.016	-0.382	0.000	0.545	1.835
KCTI_2	81.953	741.053	0.006	0.111	0.915	0.305	3.276
CR_DUONG	354.889	76.916	0.811	4.614	0.002	0.033	3.722
CL_DUONG	212.414	554.215	0.020	0.383	0.000	0.363	2.758
TT_LL	-352.214	555.956	-0.031	-0.634	0.547	0.433	2.310
DIEN_NUOC	389,288	588,962	0.015	0.661	0.509	0.768	1.303
D_TICH	4.490	7.762	0.080	0.578	0.581	0.053	8.905
H_THE	234.098	809.739	0.022	0.289	0.000	0.170	5.887
CR_MT	28.439	135.966	-0.027	-0.209	0.840	0.058	7.143
MTST_TOT	282.934	526.701	0.026	0.537	0.000	0.425	2.352
MTST_BT	-314.289	1011.872	-0.030	-0.311	0.000	0.108	9.258
AN_TOT	-901.241	1982.037	-0.085	-0.455	0.663	0.029	4.534
AN_BT	-984,817	1021,522	-0.056	-0.964	0.336	0.121	8.286
MTKD_TOT	2304.939	1618.613	0.145	1.424	0.000	0.096	1.364
MTKD_KHA	341.710	1082.544	0.032	0.316	0.000	0.101	9.934
P_LY	-2306.446	2493.133	-0.213	-0.925	0.386	0.019	2.689
Q_HOACH	608.095	1169.374	0.038	0.520	0.000	0.185	5.409

Source: Calculated by the authors.

Through the analysis results in Table 6, it can be seen that most of the observed variables have signs as expected, and have coefficients R2 = 0.893, especially variables TT_LL, P_LY have negative signs contrary to expectations. The significant level of the variables KCTI_2, TT_LL, DIEN_NUOC, D_TICH, CR_MT, AN_TOT, and AN_BT is greater than the 5% significance level. Therefore, these variables are not statistically significant, and the study will exclude them from the model.

b. Land price regression model for the second time

After removing the variables P_LY, KCTI_2, TT_LL, DIEN_NUOC, D_TICH, CR_MT, AN_TOT, AN_BT from the model and rerunning the model with the remaining ten variables, we find ourselves with the results in Table 7.

Table 7

Results of the second regression model

Model	Unnormalized regression coefficients		Normalized regression coefficient	Test t	Significance level	Multicollinear Statistics	
	Coefficient B	Error				Accept	VIF
(Constant)	812.952	2181.983		0.373	0.004		
KCTT	-0.426	1.352	-0.028	-0.315	0.000	0.085	1.823
KCTI_1	-282.434	357.991	-0.024	-0.789	0.000	0.692	1.445
CR_DUONG	389.834	657.004	0.891	0.593	0.000	0.000	3.184
CL_DUONG	433.166	432.329	0.041	1.002	0.004	0.395	2.533
H_THE	183.337	373.343	0.017	0.491	0.000	0.529	1.889
MTST_TOT	418.175	17280.813	0.039	0.024	0.000	0.000	2.151
MTST_BT	173.978	385.119	0.016	0.452	0.000	0.527	1.898
MTKD_TOT	2863.779	13335.786	0.181	0.215	0.000	0.001	1.824
MTKD_KHA	600.677	564.026	0.055	1.065	0.005	0.246	4.070
Q_HOACH	685.871	917.035	0.043	0.748	0.000	0.199	5.021

Source: Calculated by the authors.

Based on the above results, the authors found that, when excluding the variables P_LY, KCTI_2, TT_LL, DIEN_NUOC, D_TICH, CR_MT, AN_TOT, and AN_BT from the model, the adjusted R2 value was almost unchanged from the original model. The final results show that the model fits perfectly; all ten variables are statistically significant (Sig.<=0.05). The research model has an adjusted R2 coefficient of 0.896; that is, the model's independent variables explain 89.6% of the variation of the dependent variable GIA_DAT. From there, the study gives a regression model of residential land price for Nghi Tan ward as follows:

*GIA_DAT = 812.952 - 0.426 * [KCTT] - 282.434 * [KCTI_1] + 389.834 * [CR_DUONG] + 433.166 * [CL_DUONG] + 183.337 * [H_THE] + 418.175 *[MTST_TOT] + 173.978 * [MTST_BT] + 2863.779* [MTKD_TOT] + 600.677* [MTKD_KHA]+ 685.871 * [Q_HOACH]*

3.2.3. Model comment and verification

a. Check the multicollinearity phenomenon

First, the authors test the phenomenon of multicollinearity using the variance magnification factor VIF. In Table 7, in the column Variance Inflation Factor value, VIF < 10. Otherwise, VIF = 1/(1 - R2) = 1/(1 - 0.896) = 9.61 < 10. Thus, the model exhibits no multicollinearity. Then, the authors test the phenomenon of multicollinearity by conducting sub-regression for each independent variable for the remaining independent variables, and the results are shown in Table 8.

Table 8

The results of determining R2 of the independent variables subregression model

No	Dependent variable	Coefficient R²
1	Distance to center (KCTT)	0.342
2	Distance to nearest utilities (KCTI_1)	0.164
3	Road width (CR_DUONG)	0.214
4	Road Quality (CL_DUONG)	0.023
5	Shape of land (H_THE)	0.145
6	Good ecological environment (MTST_TOT)	0.533
7	Normal ecological environment (MTST_BT)	0.499
8	Good Business environment (MTKD_TOT)	0.222
9	Normal business environment (MTKD_KHA)	0.134
10	Land use planning (QH)	0.245

*Source:*Calculated by the authors.

The above results show that the correlation coefficient R2 = 0.894 of the land price regression model is larger than the R2 of the sub-regression models. This means that these sub-regression models are meaningless, and there is no autocorrelation between the independent variables. Both ways of

testing multicollinearity above conclude that a multicollinearity phenomenon does not exist in the model.

b. Check for autocorrelation

According to the regression results, we have the value d = 1,878 in the range of conditions 1 < d < 3, so autocorrelation does not occur in the above regression model.

c. Check for the phenomenon of variance change

According to the test results, all significance levels of the independent variables are > 0.05, which means there is no residual variance. From there, it says the model is stable, and the data is reasonable. Thus, the test shows that the residual variance does not change. The variables with statistical significance through the regression model tests include KCTT; KCTI_1; CR_DUONG; CL_DUONG; H_THE; MTST_TOT; MTST_BT; MTKD_TOT; MTKD_KHA; QH.

From the above analysis, the authors conclude that the regression model satisfies the conditions of unbiased linearity; that is, there is no autocorrelation, multicollinearity, and variable variance. On the other hand, the variables in the model are all statistically significant. This means that the proposed model is quite suitable and can be applied in practice. The selected model therefore reads as:

GIA_DAT = 812.952 - 0.426 * [KCTT] - 282.434 * [KCTI_1] + 389.834 * [CR_DUONG] + 433.166 * [CL_DUONG] + 183.337 * [H_THE] + 418.175 *[MTST_TOT] + 173.978 * [MTST_BT] + 2863.779* [MTKD_TOT] + 600.677* [MTKD_KHA]+ 685.871 * [Q_HOACH]

Using the above regression model to calculate land prices by the tool on ArcGIS software, the land price results are shown in the Column "Results calculated by the regression model" as shown in Table 9. Then, comparing the actual survey land price results (these properties are not the investigated points for building the land price regression model) with the land price results calculated by the regression model, these two results are not significantly different and fluctuate by about 10%.

Table 9

Results of checking the accuracy of the residential land price model

No	Address	Land price (thousand VND/m²)		Difference (%)
		Results calculated by regression model	Market price survey results	
1	Land No. 226 is located on Highway 46.	14,361,297 (621.97 USD/m²)	15,000,000 (649.63 USD/m²)	4.3
2	Land No.507 is located on Highway 46.	15,076,263 (652.94 USD/m²)	16,000,000 (692.64 USD/m²)	5.8
3	Land No. 106 corner of the old train line and Nghi Quang road.	9,501,007 (411.48 USD/m²)	10,500,000 (454.74 USD/m²)	9.5
4	Land No.100 clings to Nghi Quang road.	8,494,617 (367.89 USD/m²)	9,000,000 (389.78 USD/m²)	5.6
5	Land No. 206 clings to the concrete road leading to the banks of the Cam River.	4,663,501 (201.97 USD/m²)	5,000,000 (216.54 USD/m²)	6.7
6	Land No. 166 clings to the asphalt road to the market.	8,277,277 (358.48 USD/m²)	9,000,000 (389.78 USD/m²)	8.0
7	Land No. 5 clings to the concrete road from block 2 to block 6.	5,435,411 (235.4 USD/m²)	6,000,000 (259.85 USD/m²)	9.4
8	Land No. 9 sandwiched between concrete road block 2 block 3.	3,258,700 (141.13 USD/m²)	3,500,000 (151.58 USD/m²)	6.9
9	Land No. 215 clings to the concrete road block 8 to block 9.	4,583,103 (198.49 USD/m²)	5,000,000 (216.54 USD/m²)	8.3
10	Land No. 598 clings to the corner of road46 and 535.	16,361,297 (798.59 USD/m²)	17,000,000 (736.25 USD/m²)	3.8

Source: Calculated by the authors.

3.2.4. Map of residential land value area

After building the land price regression model utilizing multivariate regression analysis, the authors used that regression model to calculate the bulk prices for land plots in the Nghi Tan ward. After interpolating to calculate the land price in the study area and dividing the price range for residential land according to the rules of natural zoning, the result is the formation of 5 sub-regions of residential land value in Nghi Tan ward, shown in Figure 1.
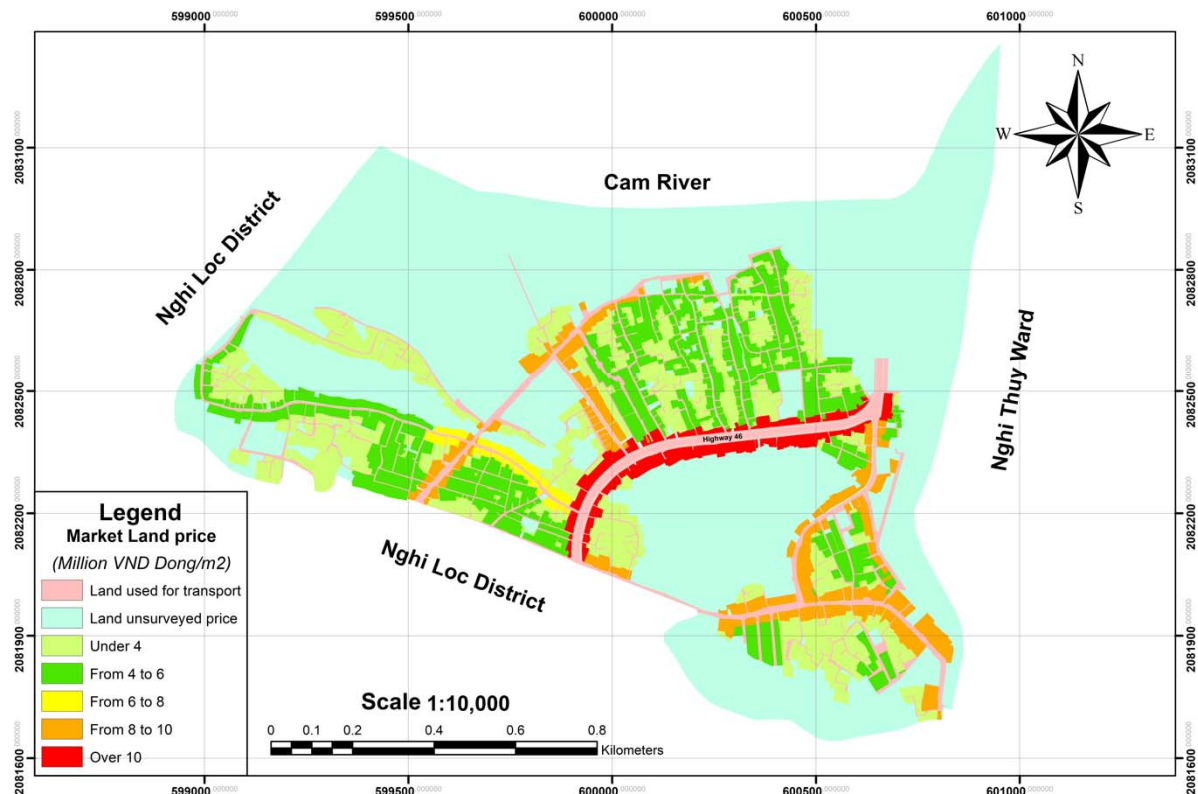
Fig. 1: Map of land prices according to market prices at Nghi Tan Ward in 2020

Based on the built value zone map, the area of Nghi Tan ward is divided into five sub-regions of land value as follows. Sub-region 1 has shallow land value and fluctuates below 4 million VND/m² (<173.24 USD/m²). Sub-region 2 has a 4-6 million VND/m² (173.24 – 259.85 USD/m²), while sub-region 3 has a land price of 6-8 million/m² (259.85 – 346.47 USD/m²); sub-region 4 has a land price of 8-10 million VND/ m² (346.47 – 433.09 USD/m²), whereas the price of land in sub-region 5 is more than 10 million VND/m² (>433.09 USD/m²). It can be seen that the sub-regions with the highest land value belong to areas located along National Highway 46 running through the ward. This is also the main traffic route of Nghi Tan ward.

4. Conclusions

The land price model in the study area consists of 10 independent variables, with the main influencing factors being the business environment, road width, and road quality. The study used 200 survey sample points in the Nghi Tan ward and obtained an $R^2 = 0.89$. According to spatial data, the land price model may give more reliable results because the database can quantify several socio-economic and environmental factors. Research results have shown that the price of residential land built in the ward correctly reflects the significant difference between the price of land plots in problematic street areas and the price of land plots in streets far from the center. The price of land plots on the same street in a favorable location for business and trade will be higher than in less convenient locations. The price of land in a location with an extensive road surface is much higher than in other locations.

In short, mass valuation helps state management agencies in charge of land offer a price close to market price based on factors affecting the land price for each specific area and area. The biggest

challenge and impediment to the batch valuation approach lies in the complexity of the regression analysis technique. In addition, the data must be up-to-date and large enough. Therefore, building a database that stores complete information about the price and characteristics of each property being traded is necessary.

5. References

Alfiyatin, A. N., Febrita, R. E., Taufiq, H., & Mahmudy, W. F. (2017). Modeling House Price Prediction using Regression Analysis and Particle Swarm Optimization. *International Journal of Advanced Computer Science and Applications, 8*(10), 323–326.

Alonso, W. (1964). Location and land Use. Havard University Press. https://doi.org/10.4159/harvard.9780674730854

Alimudin, A., Simangunsong, P., & Wajdi, M. B. N. (2017). The Factors Affecting Land Prices In Housing Location In Sidoarjo Regency. *The Siprit of Society Journal, 1*(1), 37–47.

Arouri, M. E. H., Youssef, A. B., Cuong, N.-V., & Soucat, A. (2014). *Effects of urbanization on economic growth and human capital formation in Africa.* halshs-01068271. https://halshs.archives-ouvertes.fr/halshs-01068271

Bailey, M. J., Muth, R. F., & Nourse, H. O. (1963). A Regression Method for Real Estate Price Index Construction. *Journal of the American Statistical Association, 58*(304), 933–942. https://doi.org/10.1080/01621459.1963.10480679

Benjamin, J. D., Guttery, R. S., & Sirmans, C. F. (2004). Mass Appraisal: An Introduction to Multiple Regression Analysis for Real Estate Valuation. *Journal of Real Estate Pratice and Education, 7*(1), 65–77. Advance online publication. https://doi.org/10.1080/10835547.2004.12091602

Couto, G., Martins, D., Pimentel, P., & Castanho, R. A. (2021). Investments on urban land valuation by real options – The Portuguese case. *Land Use Policy, 107*, 105484. https://doi.org/10.1016/j.landusepol.2021.105484

Demetriou, D. (2016). The assessment of land valuation in land consolidation schemes: The need for a new land valuation framework. *Land Use Policy, 54*, 487–498. https://doi.org/10.1016/j.landusepol.2016.03.008

Englund, P., Quigley, J. M., & Redfearn, C. L. (1998). Improved Price Indexes for Real Estate: Measuring the Course of Swedish Housing Prices*. *Journal of Urban Economics, 44*, 171–196. https://doi.org/10.1006/juec.1997.2062

Gouriéroux, C., & Laferrère, A. (2009). Managing hedonic housing price indexes: The French experience. *Journal of Housing Economics, 18*(3), 206–213. https://doi.org/10.1016/j.jhe.2009.07.012

GSO (General Statistics Office in Vietnam). (2021). Statistical summary book of Vietnam. Statistical Publishing House.

IAAO. 2013. *Standard on Mass Appraisal of Real Property.* ISBN: 978-0-88329-207-5. https://www.iaao.org/media/standards/marp_2013.pdf

Karakayaci, Z. (2018). Regression Analysis for the Factor Affecting on Farm Land/Urban Land Value in Urban Sprawl. *Turkish Journal of Agriculture –. Food Science and Technology (Campinas), 6*(10), 1357. https://doi.org/10.24925/turjaf.v6i10.1357-1361.1886

Khanh, H. (2021). *Speculation pushes land prices up: Ministry of Construction.* Vietnamnet Global. https://vietnamnet.vn/en/business/speculation-pushes-land-prices-up-ministry-of-construction-715496.html

Kolebe, J., Schulz, R., Wersing, M., & Werwatz, A. (2019). Land value appraisal using statistical methods. FORLand-Working Paper, no. 07, Humboldt- Universität zu Berlin, DFG Research Unit 2569 FORLand "Agricultural Land Markets - Efficiency and Regulation", Berlin.

Kuddus, M. A., Tynan, E., & McBryde, E. (2020). Urbanization: A problem for the rich and the poor? *Public Health Reviews, 41*(1), 1. Advance online publication. https://doi.org/10.1186/s40985-019-0116-0 PMID:31908901

Mize, S. (2017). *Using Regression Analysis to Predict Single Family Home Values/Prices in the Belmot/Eastside Areas of Pueblo.* Colorado State University.

Morales, J., Flacke, J., & Zevenbergen, J. (2019). Modelling residential land values using geographic and geometric accessibility in Guatemala City. *Environment and Planning. B, Urban Analytics and City Science, 46*(4), 751–776. https://doi.org/10.1177/2399808317726332

Mostafa, M. M. (2018). A spatial econometric analysis of residential land prices in Kuwait. *Regional Studies, Regional Science, 5*(1), 290–311. https://doi.org/10.1080/21681376.2018.1518154

Nguyen, T. T. (2021a). Conversion of land use and household livelihoods in Vietnam: A study in Nghe An. *Open Agriculture, 6*, 82–92. https://doi.org/10.1515/opag-2021-0010

Renigier-Bilozor, M. & Wisniewski, R. (2013). The impact of macroeconomic factors on residential property prices indices in Europe. *Aestimum*, 149-166.

Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy, 82*(1), 34–55. https://doi.org/10.1086/260169

Sampathkum, V., Santhi, M. H., & Vanjinatha, J. (2015). Evaluation of the Trend of Land Price using Regression and Neural Network Models. *Asian Journal of Scientific Research, 8*, 182–194. https://doi.org/10.3923/ajsr.2015.182.194

Shanks, E., Luttrell, C., Conway, T., Loi, V. M., & Ladinsky, J. (2004). *Understanding pro-poor political change: the policy progress*. Overseas Development Institute. https://cdn.odi.org/media/documents/3902.pdf

Thanh, H. X., Anh, T. T., & Phuong, D. T. T. (2013). *Urban poverty in Vietnam – a view from complementary assessments*. Human Settlements Working Paper Series Poverty Reduction in Urban Areas – 40. International Institute for Environment and Development.

Tran Tuan, N. T. (2021a). The consequences of expropriation of agricultural land and loss of livelihoods on those households who lost land in Da Nang, Vietnam. *Environmental & Socio-economic Studies, 9*(2), 26–38. https://doi.org/10.2478/environ-2021-0008

Tran Tuan, N. (2021b). Shrinking agricultural land and changing livelihoods after land acquisition in Vietnam. *Bulletin of Geography. Socio-Economic Series, 53*(53), 17–32. https://doi.org/10.2478/bog-2021-0020

Tran, H. A., & Yip, N. M. (2019). *Vietnam's Post-reform Housing Policies: Social Rhetoric, Market Imperatives and Informality*. In book: Housing Policy, Wellbeing and Social Development in Asia, Rebecca Lai Har Chiu and Seong-Kyu Ha (Eds). Routledge, Taylor and Francis Group.

Wincott, D.R. (2001). The myth of three independent approaches to value. *Real Estate Issues*, 26(2).

World Bank. (2015). Vietnam Affordable Housing: A Way Forward. https://openknowledge.worldbank.org/bitstream/handle/10986/22921/Vietnam000Affo0sing000a0way0forward.pdf?sequence=1&isAllowed=y

World Bank. (2020). Urbanization at a Crossroads: Embarking on an Efficient. Inclusive, and Resilient Pathway.

Zhou, W.-X., & Sornette, D. (2008). Analysis of the real estate market in Las Vegas: Bubble, seasonal patterns, and prediction of the CSW indexes. *Physica A, 387*(1), 243–260. https://doi.org/110.1016/j.physa.2007.08.059