Determinants of house prices using spatial analysis: the case for Bulawayo

Ndava Constantine Mupondo[1], Bukhosi Ncube[1], Abigail Mupondo[2] and Success Clifford Nemahwe[2]

[1]*Department of Statistics and Operations Research, National University of Science and Technology, Bulawayo, Zimbabwe.*
[2]*Department of Science Mathematics and Technology Education, National University of Science and Technology, Bulawayo, Zimbabwe.*

Email: constantine.mupondo@nust.ac.zw

ABSTRACT

The factors affecting house prices are crucial to Zimbabwe's property organisation, and they necessitate an understanding of market trends and patterns in the housing industry. The primary goal of this research is to investigate the correlations between house prices and the factors that influence them to develop a model that can forecast house prices in Bulawayo. This study uses exploratory data analysis and spatial regression approaches to analyse factors affecting house prices in Bulawayo to understand how much housing costs are influenced by the availability of health services and retail stores. How does the distance to schools and the central business district (CBD) affect property prices, as well as the size of the land and the physical environment? To attain these goals, spatial analysis and local regression parameter estimates were used. The study found that many variables have both positive and negative effects on house prices across space and that the spatial lag model is the best fit for predicting house values in Bulawayo.

Key words: Spatial analysis, property, model, forecast, house prices.

Received: 09.02.23 **Accepted:**16.02.24

1 INTRODUCTION

1.1 Background of the study

It is a fact that house prices are primarily determined by the supply and demand for housing. The number of households that form and changes in real incomes are both positively connected with underlying housing demand (OECD, 2022). The effect of increased demand on property prices and rentals will be determined by the ratio of increased demand to changing supply. Numerous factors influence house prices, and they vary by location. They may be influenced by the infrastructure of that location of study, and property values typically reflect the interplay of various factors such as environment, age, and size of the building (Oloke et al., 2013).

Moreover, McKinsey & Company (2022) argue that users' willingness to pay higher prices for property that is properly endowed with infrastructure is high because they can derive more benefits from the infrastructure provided than those without. Also, people are prepared to pay high prices for properties that are properly serviced and have good infrastructure. Infrastructure resources are intermediary goods that add social value and are critical components of the city's production systems. These characteristics provide insight into how infrastructure influences the price of a home across the country and show that house values tend to remain high in areas or locations where infrastructure development is occurring. Furthermore, after a discussion made with one of the property sales agents in one of the biggest and best-known property companies, John Pocock, it was discovered that property prices or values are evaluated through discussions, and the most important factor that is used is the location of the property to be evaluated, and the other factors come after or depend on the location of the property. This information shows that the location of the house to be evaluated on its value is one of the most important factors that should be considered, and it brings out other factors to be evaluated on. Furthermore, one possible reason for the relative increase in housing prices is a simple income effect or nonhomogeneity in preferences (Kahn, 2008). As people become wealthier, they may prefer to spend more of their money on housing services, the cost of which tends to rise due to its reliance on land, a fixed factor. The share of housing services tends to increase in gross domestic product (GDP) when the number of people's incomes increases because these consumers tend to spend more, hence increasing expenditure. Another driver of housing prices could be differing technical progress trends in housing services versus other goods (CMHC, 2022). The relatively large share of land and structures in the value of housing, two inputs traditionally thought to be less amenable to embodied technical progress, lends credence to this story.

Housing is extremely important to both society and the economy. On the one hand, it satisfies one of the most basic human needs by providing shelter, and on the other side, it has an impact on people’s well-being. Housing, on the other hand, is a long-term investment that accounts for a significant amount of household wealth. As a result, the single most valuable asset in a household's portfolio is often housing (Chetty, Sandor, and Szeidl, 2016). The real estate market is a marketplace where sellers and buyers exchange real estate. The market can be classified by sector (residential, commercial, agricultural, recreational, or industrial); location (local, regional, national, or worldwide); or type of demand (occupation, ownership, investment, speculation, or development). The residential sector accounts for a higher percentage of the market (CMHC, 2023). Due to inherent qualities such as heterogeneity, large sums of capital involved, and significant transaction costs, the property market is imperfect. Its nature, the techniques for performing transactions in it, and the lack of information about transactions due to their private nature all contribute to its flaws (Pepper et al., 2013). Market participants may be confused by different estimation results for a given variable, particularly disagreements about the direction of the effect. Furthermore, there is reason to suppose that dwelling features are not valued equally across a particular house price distribution. Kamloops (2020), recognise the difficulty of assessing particular housing features and point out that their impact on pricing is difficult to quantify. European Banking Authority (2022), also points out that various consumers may place varying values on certain dwelling attributes.

Furthermore, when dealing with geographical phenomena that must be analysed, geographic location is known to play a significant role in the occurrence of spatial effects such as spatial autocorrelation and heterogeneity. Spatial autocorrelation, which was defined by Costa and Tokuda (2022), was referred to as the ‘coincidence of value similarity with location similarity'. This explains how housing markets operate around the world. With a nearby location, homeowners tend to follow their neighbours' improvements, which results in the similarity of designs and other structural characteristics. Also, spatial autocorrelation arises from the shared locational amenities of houses in the nearby location and neighbourhood (Chau et al., 2022), such as school districts, police stations, green spaces, transportation nodes, shopping centres, and other facilities. On the other hand, spatial heterogeneity describes a housing market operational process in which the same set of housing characteristics may obtain different housing prices in the different parts of the studied region (Xiang, Tang and Yao, 2022).

Finally, home prices may also be influenced by demographic factors such as population, ageing, and migration. According to Takats (2012), a larger population is linked to higher actual housing costs. Furthermore, if the proportion of the elderly population to the working population grows, property values may be impacted. Nonetheless, Chen et al. (2012) conclude that population ageing is unlikely to be the primary driver of housing costs. To summarise, the most important determinants of home prices are economic and financial variables; nonetheless, one can observe that demographic characteristics have a smaller but still considerable influence in the long run.

Currently, house price determination is at the heart of Zimbabwe's property organisation, and the price of a home is thought to be transacted between the buyer and seller following their concession through the property agents. However, there is a lack of understanding among buyers and potential investors as to how a property's value is fully determined. The house price determination is based on the location of that particular property to be valued, as well as some physical qualities included in or found in the surrounding areas, using information from real estate agents. Most purchasers agree that location, availability of essentials, and affordability are the most important aspects; however, Tobler (1970), a well-known geography researcher, has had his ideas collected and referred to as the First Law of Geography: "Everything is related to everything else, but close things are more related than distant things." This concept allows us to further investigate how physical things in space are related. Is it a quantifiable fact that in Bulawayo the predominant property price driver determination is done through comparisons made during discussions, with location as the principal factor considered? For this and other questions in the same vein, it is superlative to rigorously investigate and come up with a mathematical or statistical model that can be used to predict property value in the Zimbabwean context. In this study, the researcher rides on the preceding assumption to come up with such a model for housing properties in Bulawayo. The primary purpose of this research paper is to identify the factors that influence house prices in Bulawayo and to develop a spatial regression model that can be used to forecast property value. To facilitate the attainment of the stated aim, the research will try to address the following objectives: (i) to identify the spatial trends of house prices and their drivers; (ii) to identify the hot and cold spots areas in Bulawayo; (iii) to determine the factors that influence house prices in Bulawayo; and (iv) to develop a model to predict house prices in Bulawayo.

To our knowledge, Bulawayo, Zimbabwe has never before seen the conduct of a preliminary study like this. Without going into detail on the main drivers of real estate prices for the Metropolis of Bulawayo, Chingweya's (2019) research focused mostly on housing for low-income people concerns and lower-income accommodation remedies: chances and challenges in Bulawayo. Magwaro-Ndiweni (2011) conducted a study titled "Contestation in the occupancy of housing space: Dwelling typologies and dwelling sites in Bulawayo, Zimbabwe" that was comparable. There is no direct connection between the results of these investigations and the ongoing study.

The unequal-environment assumption, which has been the subject of extensive discussion and investigation for many years, is the primary premise of this study. According to the unequal-environment-assumption, environmental factors that are pertinent to the trait under research are unequally correlated, meaning that various residential locations encounter distinct sources of environmental variation. Furthermore, unbiased would only arise from environmental disparity if the characteristic of interest is impacted by the environmental factors that differ between the places.

2. MATERIALS AND METHODS

2.1 Data Collection

Secondary data was employed more frequently than primary data. The data for this study was gathered through the Bulawayo City Council's housing department, the property websites and some interviews with property sales agents in the Bulawayo area. The data set used in the study was collected month-by-month from January 2012 to June 2023. Spatial analysis can be done in several packages. In this research, the researcher used packages that include Geoda, R Studio, Microsoft Excel, and Google Earth.

2.2 Steps in Spatial Analysis

The steps of the spatial analysis process include problem formulation, data gathering, exploratory analysis, modelling and testing, and finally reporting. The techniques for geographic statistical modelling discussed in the following subsections are exploratory spatial data analysis, spatial regression, and testing.

Spatial analysis entails the examination of numerical spatial data as well as the development and validation of mathematical models of spatial processes. All of these actions are designed to improve our understanding of spatial processes. The focus of the investigation is on the structure of the housing market and the interrelationships among the variables that influence its functioning, either directly or indirectly.The spatial analysis can be represented using Fig 3.1.

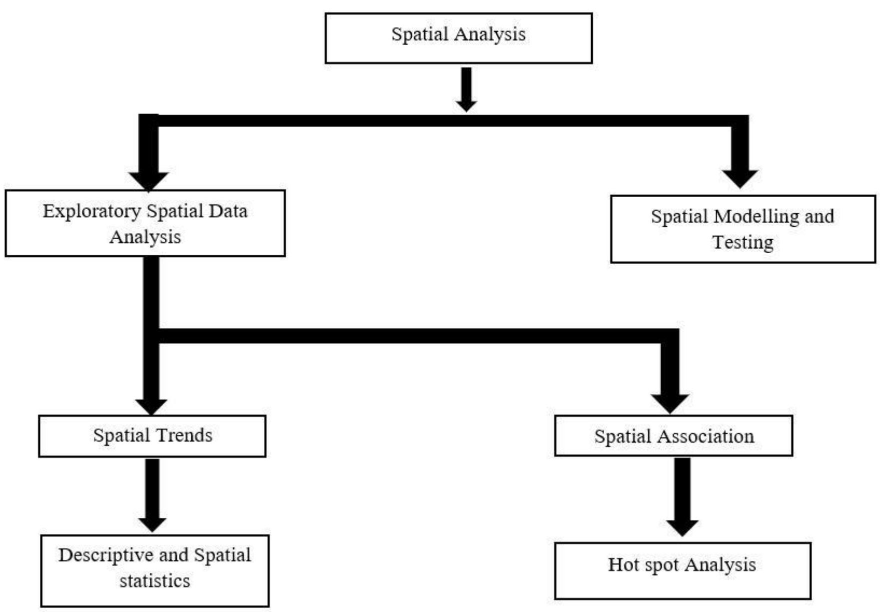

Figure 3.1: The steps in spatial analysis

2.3 Quantitative Spatial Data Analysis

Quantitative geographic data analysis was utilised to uncover spatial trends in house prices as well as hot and cold spot locations on house prices by searching for patterns and probable explanations, as well as geo visualisation through data visuals and maps. Exploratory data analysis (EDA) is a method for analysing data sets, learning about them, and generating ideas. Exploratory spatial data analysis (ESDA) is the application of exploratory data analysis (EDA) to geographic data. The detection of patterns using data visualisation and statistics is a crucial part of ESDA. It entails employing strategies to summarise and visualise data in the quest for significant patterns, as well as identifying troublesome and unusual data known as outliers. Maximising insight into a data set, finding underlying structures, extracting significant variables, detecting outliers and anomalies, validating underlying assumptions, and constructing parsimonious models are some of the goals of ESDA.

Correlation is the most well-known strategy for explaining a trend in a data set. Spatial auto-correlation is a type of spatial analysis that looks at how the magnitude of a variable at one location correlates with nearby locations. The extent of the effects can be calculated using a variety of statistics that represent geographical variation in terms of a function that illustrates how spatial autocorrelation diminishes as distance increases. (Tobler, 1970) captures these key concepts in his "First Law of Geography," which states, "All things are related, but adjacent things are more related than distant things."

A formal characterization of the spatial structure is required to test for spatial auto-correlation and incorporate it into a model. A spatial weight matrix and spatial lags have been used to accomplish this. The strength of the spatial association between each of the n spatial observations is described by each element W_{ij} of a spatial weight matrix, which is an $n \times n$ square matrix known as *W*. The analyst's knowledge and assumptions about the study area and data within the model are embodied in the weight matrix.

When defining the spatial structure, a distance weight matrix considers the distance between observations. Because accurate distances can be collected, it works well with point data. A spatial weight matrix based on distance can be created, with units inside a particular radius having a spatial weight of 1 (neighbours), and units outside of that radius having a spatial weight of 0.

$$W_{ij} = \begin{cases} 1 \text{ if the distance between } i \text{ and } j \leq D \text{ and } i \neq j \\ 0 \text{ otherwise} \end{cases}$$

(1)

where, *i* and *j* are two locations.

Contiguity-based spatial weights are another useful method for defining spatial autocorrelation. In this method, one generated the weight matrix by determining if two locations share a boundary or not; if they do, we put a 1 in the matrix; if they don't, we put a 0.

$$W_{ij} = \begin{cases} 1 \text{ if location } i \text{ is contiguous to } j \text{ and } i \neq j \\ 0 \text{ otherwise} \end{cases} \quad (2)$$

where, *i* and *j* are two locations.

The weight matrix quantifies the strength of the relationships; therefore, two points close in space should have a higher weight than those further apart, with the weight matrix often dropping to zero at some cut-off point. The neighborhood around each observation is formalized by the spatial weight matrix. The weight matrix is multiplied by a variable measured for each observation to produce a spatial lag.

2.3.1 Global Spatial Autocorrelation.

Moran's I, Moran (1950), was used to statistically test for global spatial autocorrelation, and this method is widely employed in spatial analysis. The value of I can lie within an interval of +1 to -1, with zero indicating a null result for autocorrelation, and, like a correlation coefficient, The following equation is used to calculate the Moran's I value for the outcome variable y:

$$I = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij}} \times \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij}(y_i-\mu)(y_j-\mu)}{\sum_{i=1}^{n}(y_i-\mu)} \quad (3)$$

where n = number of spatial units, y is the house price for location i, or j, μ = mean, and i, j represents two neighboring locations. Positive spatial autocorrelation is indicated by an I value that is significantly greater than zero. This is a set of regions in which places with high (or low) values are clustered together by areas with similar high (or low) values, rather than being scattered randomly across space. A value of I that is much smaller than zero, on the other hand, shows negative spatial autocorrelation. This can happen when the high and low values of an outcome variable are more widely scattered than a random spatial distribution would predict.

2.3.2 Local Spatial Autocorrelation.

The level of autocorrelation may vary due to the presence of spatial heterogeneity. A single value is insufficient to explain the variance in the variables in this scenario. As a result, local autocorrelation approaches were also used in the research. A variety of techniques can be used to create a local spatial relationship (Anselin, 1995). To detect spatial clusters of so-called "hot spots" (areas with high house prices) and "cold spots" (areas with low house prices) and to establish which regions have spatial relationships (areas with low values of house prices). The Moran I scatterplot was employed, as well as statistics (Ord and Getis, 2001). The latter method, which is represented by the equation below, was utilised for hot spot analysis to locate geographic groupings.

$$G_{*i} = \frac{\sum_{i=1}^{n} W_{ij}Y_{ij} - \mu\sum_{i=1}^{n} W_{ij}}{S\sqrt{\frac{n\sum_{i=1}^{n} W_{ij}^{2} - \left(\sum_{i=1}^{n} W_{ij}\right)^{2}}{n-1}}} \quad (4)$$

where n = number of spatial units, y is the house price for location i, or j, μ = mean, and w_{ij} spatial weight for two neighboring locations i, and j.

Table 1: Descriptions of the explanatory variables.

Variable	Description
House price	Log of house price (thousand US$)
Land size	Log of total stand area (square meter)
Schools	Whether the house is near schools (distance)
Retail	Accessibility of retail outlets (ranked between 0 and 10)
Health	Accessibility of health service facility (ranked between 0 and 10)
CBD	Distance from the house to the central business district area (meter)

Physical Env Quality of roads, vegetation and closeness to industries (ranked 0 and 10)

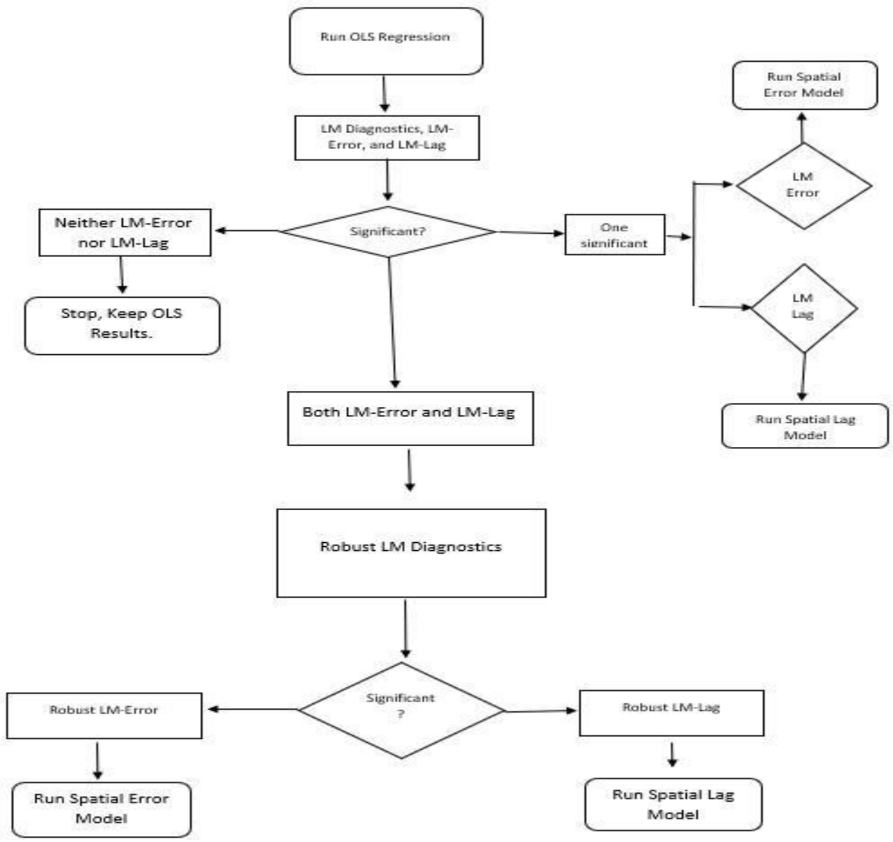

Figure 2: The decision criterion in comparing model

2.3.3 Four steps In Spatial Regression

The spatial regression analysis involves four steps that are used in selecting the best mode that can be used in predicting values of the variable of interest; in this case, the mode to be chosen will be used in predicting house prices in Bulawayo. These four steps include model specification, model estimation, model diagnostics, and model prediction.

3. RESULTS

3.1 Non-Spatial Statistics

In this section non-spatial characteristics are discussed. Table 2 summarizes the statistics of variables. On house prices, a **minimum value of $8000 in high-density** areas (for example in Cowdray Park and Emganwini) was observed, while the maximum value observed is $130000 in low-density areas (for example in Selbourne Park). Pikora (2002) created the Systematic Pedestrian and Cycling Environmental Scan (SPACES), a trustworthy audit tool that assesses the physical environmental elements affecting cycling and walking in nearby communities. To quantify the phenomenon of the physical environment, this study used the same metric. The minimum value for the size of the land is 200 square meters and the maximum value is 1195 square meters, distance to schools and CBD has minimum values of 815 and 8983 meters and the maximum values are 1071 and 15794 meters respectively. Retail outlets accessibility, health services accessibility, and physical environment which include the road network, and quality of the road have minimum values of 10%, 5% and 5% respectively, and the maximum values are 60% , 60% , 65% respectively. From the summary statistics it can be observed that values for a home in areas like Cowdray park Emganwini in high density are significantly low as compared to areas like Selbourne Park and Bradfield and the average house prices of $41242 was noted to be observed most in low and medium areas. Moreover, on explanatory variables, the high values occur in areas with relatively high values for homes as compared to ones with low values for homes.

Table 2: Summary of Variables

Statistic	House Price	Land size	Schools	Retail	Health	CBD	Physical Env
Min	8000	200	815	10	5	1071	5
1st Qu	16442	243	2174	21	20	5108	25
Median	25436	289	3310	35	37	8294	39
Mean	41242	479.8	3794	40.6	47.58	9292	40.19
3rd Qu	64714	790.8	4860	52	54	13507	50
Max	130000	1195	8983	60	60	15794	65

3.2 Spatial Characteristics

Figure 3 shows a quantile map of house prices around the city of Bulawayo. A very clustered spatial distribution or spatial house prices in different residential areas can be observed, with a lower range of house prices in high-density areas (e.g Cowdray Park, Njube, and Nkulumane areas), this implies that house prices in these areas are of less value compared to areas with a higher range of house prices in medium, and low-density areas (e.g Richmond, Selbourne Park, and Bradfield areas). Figure 3 allows us to observe the clusters for house values across the Bulawayo region, with clusters for high values, observed in areas around Richmond, Selbourne Park, and Bradfield areas and low clusters for low values observed in areas around Cowdray Park, Njube, and Nkulumane areas, therefore using this Figure 3 one can easily identify where to invest.

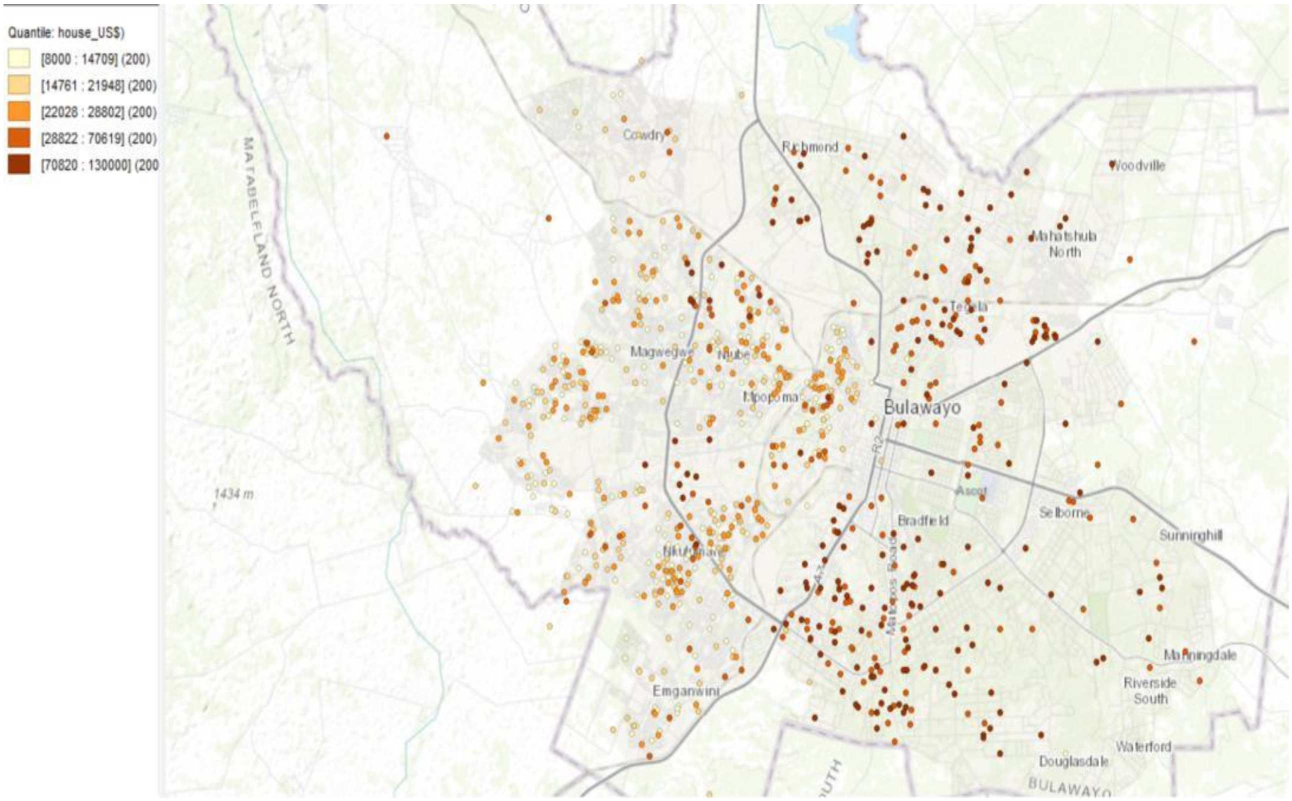

Figure 3: The map of the distribution of house prices

3.3 Global Spatial Autocorrelation

In this section, the researcher used Moran's I to statistically test for global spatial autocorrelation and this method is widely employed in spatial analysis. As Tobler (1970) mentioned, "everything tends to be related to everything else, but things close to each other tend to be more related", this allows us to further investigate how variables are spatially correlated with the residential areas in the city of Bulawayo. The use of Global spatial autocorrelation enables us to achieve our first objective on spatial trends of house prices and to obtain relationships of these variables in different residential areas in the city of Bulawayo. Therefore, Moran's I have been employed to uncover the presence of global spatial autocorrelation by measuring a single value of autocorrelation level across the study area for each variable. The positive values of Moran's I mean that a high value tends to attract other high values but repel low values and the negative ones are just the opposite (Huang et al., 2017), and the Monte-Carlo simulation to select significant variables to be included. Table 3 shows the output from Monte-Carlo simulation statistics for each variable.

Table 3 summarizes the values of calculated Moran's I for each variable. The positive values of Moran's I suggest that all variables are positively correlated to nearby locations: house prices and land size present a strong correlation whereas the distance to school has a weaker correlation this implies that the prices of houses and the size of land are influenced by nearby areas to a higher extent as compared to other variables. Therefore, it can be concluded that all variables are significantly positively auto correlated and it implies that there are significant relationships between these variables and residential areas in Bulawayo. For example, when one look at the Size of Land in Cowdray Park and Luveve which are nearby areas, the global autocorrelation statistics of 0.735 it implies that these residential areas have familiar values on the sizes of land as noted by (Santos Laurent, 2020).

Table 3: Global Moran's I statistics, output from Monte-Carlo simulation.

Variable	Moran's I	Observed rank	*p*-value
House price	0.658	1000	0.001
Land Size	0.735	1000	0.001
Distance to Schools	0.151	1000	0.001
Retail outlets access	0.426	1000	0.001
Health Service access	0.247	1000	0.001
Distance to CBD	0.286	1000	0.001
Physical environment	0.386	1000	0.001

3.4 Local Spatial Autocorrelation

Local spatial autocorrelation focuses on finding the relationships between each observation and its nearby areas. In this section Moran scatterplot, Local Moran's I and Getis and Ord's Gi and Gi* methods were employed in the study, it was employed to discover hot spots and cold spots in house prices.

The Moran scatterplot in Figure 4 identifies clusters of high and low values in house price data. As indicated from the plot on the left-hand lower bottom of the graph it is associated with low values surrounded by low values, this implies that areas with lower house prices are surrounded by similar areas with lower house prices. It also presents clusters of high values surrounded by high values on the upper right-hand side of the graph, this means that areas with higher house prices are surrounded by areas with higher house prices (Assam, 2022). Their value of 0.658 indicates a strong correlation in the city of Bulawayo on house prices with nearby locations.

Figure 5 represents the outputs of Getis and Ord's Gi and Gi* analysis, and indicates the distribution of raw Gi* values across the Bulawayo area. The figure also displays the distribution of hotspots and cold spots and the significant analysis, this technique was employed to achieve the objective of identifying the cold spots and hotspots. The medium and low-density for example Richmond, Selbourne Park, and Bradfield areas of Bulawayo stand as a cluster of high property prices while some parts of high-density areas are identified as a cluster of the relatively low property price, this implies that the areas around Richmond, Selbourne Park, and Bradfield areas have physical attributes which drive the price for homes to high as compared to the areas around Cowdray Park, Njube, and Nkulumane areas, which have physical attributes of less value which drives the low price for homes.

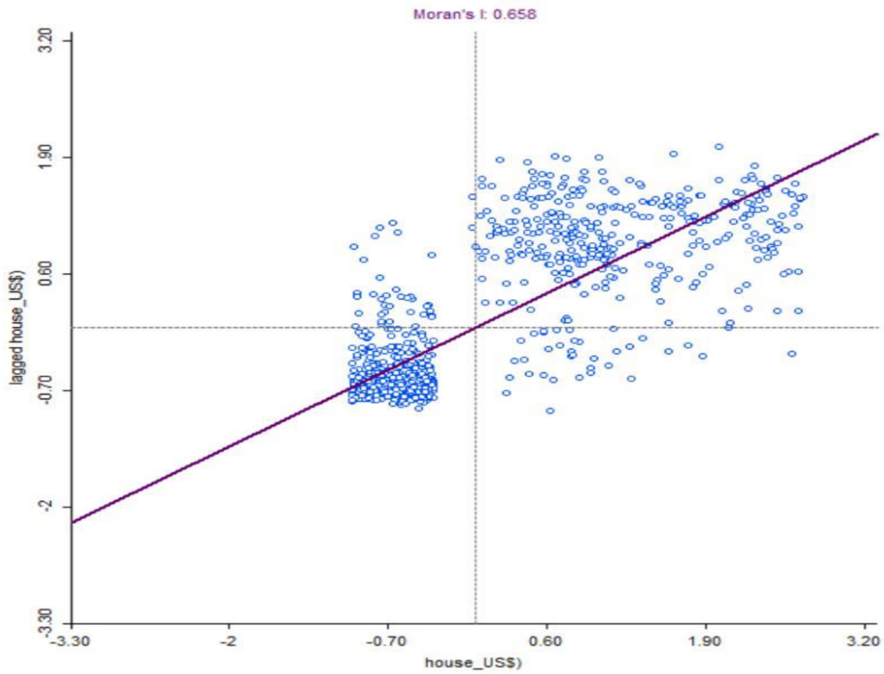

Figure 4: Moran Scatterplot of House Prices

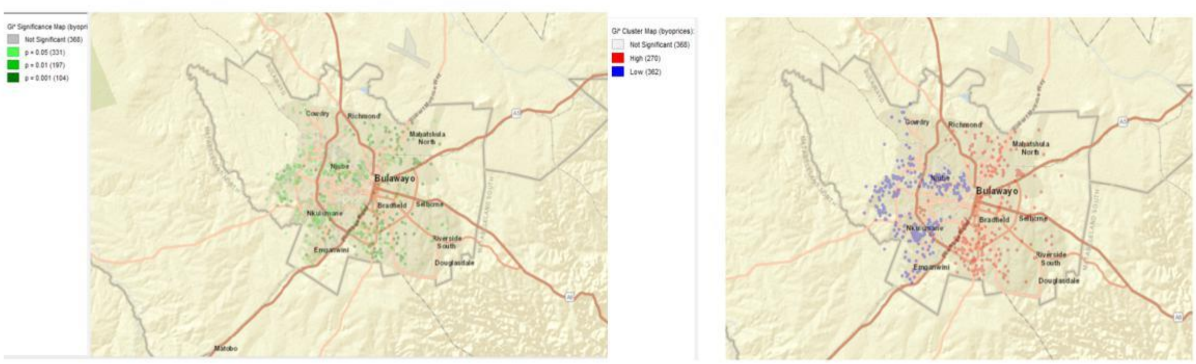

Figure 5: The Output of Getis and Ord's Gi*

3.4.1 Local Moran's I

Using the Local Moran approach, the areas with significant clusters have been uncovered. Consistent with the results calculated by Getis and Ord's Gi and Gi *

(Santos Laurent, 2020), there are areas in Bulawayo associated with high-high clusters in the medium and low density for example Richmond, Selbourne Park, and Bradfield areas this implies that they are associated with higher values of house prices and with high autocorrelation surrounded by areas with similar higher values of houses prices and autocorrelation, and low-low clusters have been found in Cowdry Park, Njube, and Nkulumane areas this means these areas are associated with lower values of house prices and autocorrelation surrounded by areas with similar lower values of house prices and autocorrelation. Figure 6 summarises local Moran's I, and represents the clusters and significant clusters in the city of Bulawayo.

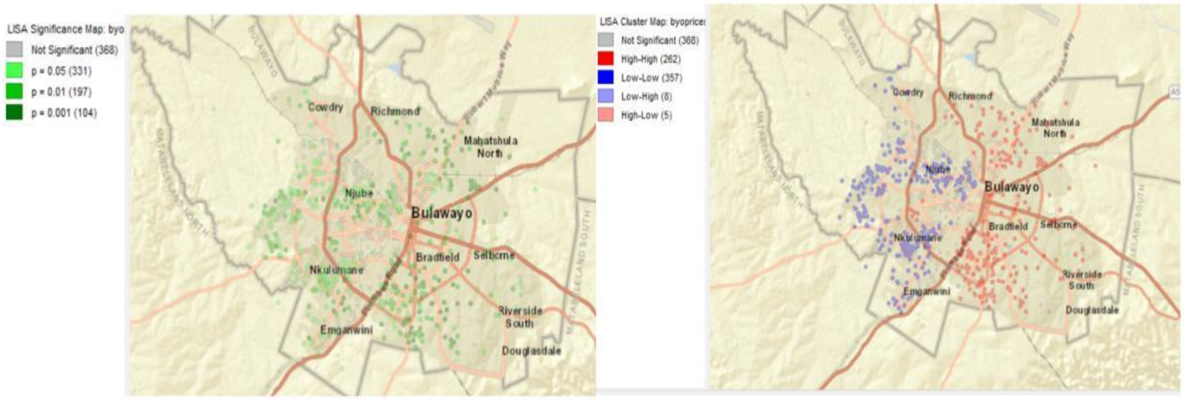

Figure 6: Local Moran's I

3.4.2 Multiple Linear Regression

Multiple linear regression was conducted, and this procedure was employed to obtain an estimated model. The log transformation was also employed on the variables so that highly skewed highly skewed variables in the data set are set to follow a normal distribution and to obtain more interpretable patterns. Table 4 shows the results of the ordinary least squares (OLS) regression.

From Table 4 the size of land, accessibility to retail outlets, and quality of physical environment seem to be statistically significant since p-values are less than 0.05 this implies that statistically, we have evidence that these variables contribute more value in predicting the value for a home. However, the distances to schools and CBD as well as accessibility to health facilities exhibit statistical insignificance with their respective p-values exceeding 0.05. This suggests that our statistical analysis does not provide enough evidence to support the notion that these variables significantly impact a home's value. From the study done by Maclennan (2012) he argues that search and matching mechanisms drive housing markets therefore, several drivers of house prices can be considered to select the most significant ones with higher influence. The value of multiple R-squared (0.7414) indicates that the multiple linear regression explains 74.14% of the variance in the data, it implies that 74.14% of the variance in house prices is explained by the size of land, accessibility to health and retail outlets, quality of the physical environment, and distance to schools and CBD. The statistically significant p-value in F-test with a p-value less than 0.05, also explains a significant fitting of the regression model with explanatory variables included in the model. Using the multiple regression results from Table 4, the researcher will use these results to determine a model for predicting house prices.

Table 4: Results of Multiple Linear Regression

Variable	Coefficient	Standard Error	t-value	p-value
Intercept	4.458	0.431	10.35	< 0.001
Land Size	0.944	0.032	29.39	0
Distance to schs	-0.041	0.024	-1.709	0.088
Retail	0.129	0.042	3.035	0.002
Health	0.014	0.025	0.567	0.571
Distance to CBD	-0.039	0.024	-1.597	0.111
Physical environment	0.099	0.025	3.992	< 0.001
Residual standard error:	0.3946	on 993 degrees of freedom		
Multiple R-squared:	0.7414	Adjusted R-squared: 0.7398		
F-statistic: 474.4	on 6 and 993 DF	*p*-value: 2.2e-16		

3.5 Model Building

In this section the researcher begins the analysis for model building, in this analysis will identify the best model that will best suit the price prediction of houses in the city of Bulawayo.

3.5.1 Model Specification

To find out what affects house prices in Bulawayo, we used model specification to decide which variables to use in the regression model. Using Table 4.3, the size of land, accessibility to retail outlets, and quality of physical environment seem to be statistically significant since p-values are less than 0.05, this implies that these explanatory variables should be included in the model. Whilst, the distance to schools, distance to CBD, and accessibility to health facilities are statistically insignificant since their p-values are greater than 0.05 hence these explanatory variables should not be included in the model. However, these can include a variable even if it is insignificant due to economic intuition (Panas, 2021). Therefore, we can include these variables in the predictions of house prices even if other variables can be statistically insignificant.

3.5.2 Model Estimation

The estimated model from OLS regression was extracted from Table 4, since the data set was normalized using the log transformation to come up with estimated exact values we anti-log the coefficients obtained from the OLS. After anti-logging the coefficients for the estimated model of house prices is written as follows:

$$Y_{houseprice} = 28707 + 8.79X_{land} - 1.10X_{schools} + 1.35X_{retail} + 1.03X_{health} - 1.09X_{CBD} + 1.26X_{physical} \quad (6)$$

From the estimated OLS regression model, the coefficients indicate a positive relationship between the size of land, accessibility to retail outlets and health facilities and physical environment with house price (Santos Laurent, 2020) in the study of spatial analysis of house price determinants whilst, there is a negative relationship in distance to schools and to CBD with house prices. A positive coefficient on the size of land, accessibility to retail outlets and health facilities and physical environment implies that an increase in each of the variables by a unit also increases the house price by a unit and also this means that the predicting the value for the home with highest values on these variables led in the highest value for a home. However, with negative coefficients for distance to schools and distance to CBD, this means increasing each of these variables by a unit reduces the house price by a unit. The intercept of the model is equal to 28707.8 this value sets the explanatory variables to zero, on average the price of a home is approximately $28707.80 before one can include other external drivers of house prices. However, price determination can be based on economic theory, which says that prices are influenced by the forces of demand and supply. This implies that when there is more demand for homes then house prices can be relatively high, and when one has more supply for homes then the prices for homes may be relatively low. Therefore, house price determination can also relate to the economic theory where supply and demand is a theory in microeconomics that offers an economic model for price determination. The researcher can use this theory so that the unit price for homes no matter how prices vary, the forces of demand and supply for houses settle it until a point of economic equilibrium is reached. By collecting and investigating information on the demand and supply for houses, then these forces might generalize the prices for a home to be approximately $28707.80 holding other things constant.

3.5.3 Testing for Multicollinearity

Multicollinearity in regression analysis occurs when two or more predictor variables are highly correlated to each other, such that they do not provide unique or independent information in the regression model if the degree of correlation is high enough between variables, it can cause problems when fitting and interpreting the regression model. (Matthews, 2006), the GeoDa diagnostic that may point out a potential problem is called a "condition number" as a rule of thumb, the values of the condition number greater than 30 are considered a suspect. Variance Inflation Factor (VIF) is another common way to detect multicollinearity it starts at one and using the rule of thumb means no correlation and if between 1 and 5 it has moderate correlation hence there is no severe multicollinearity (Santos Laurent, 2020). Table 5 shows the results which indicate that there is no severe multicollinearity

.

Table 5: Summary of Multicollinearity Analysis

Variables	Land size	Schools	Retail	Health	CBD	Physical Env
VIF value	1.209	2.379	2.07	1.445	1.64	1.546
GeoDa condition number	14.819					

3.5.4 Testing for Normality Errors (Jarque-Bera Test)

A normality test was conducted in this study to find out whether the errors follow a normal distribution. Some tests can be conducted for assessing residual normality, this is done by performing hypothesis testing where the null hypothesis is that the errors have a normal distribution. Below shows the hypothesis test that was done using the Jarque-Bera test, and Table 6 shows the results.

H_0: **There is Normality in the errors.**

H_1: **There Non-Normality in the errors.**

Table 6: Test on Normality of Errors

Test	df	Value	p-value
Jarque-Bera	2	105.3102	1E-05

From Table 6 we have a *p*-value of 0, which is less than 0.05 hence we reject H_0, and conclude that there is non-normality in the errors and that follows a normal distribution. This implies that some transformation should be employed in the data to make it follow a normal distribution.

3.5.5 Testing for Heteroskedasticity (Breusch-Pagan)

To understand the relationship between several explanatory variables and the response variable house price, multiple linear regression is employed. In this analysis, systematic changes in the variance of residuals over a range of measured values may occur. To determine whether this takes place or not, we perform by using the Breush-Pagan test. Table 7 shows the results of the Breusch-Pagan test.

H_0**: There is Heteroskedasticity in the error terms**

H_1**: There is Non-Heteroskedasticity in the error terms.**

Table 7: Test for Heteroskeasticity

Test	df	Value	p-value
Breush-Pagan	6	14.888	0.02114

From Table 7, we have a p-value of 0.02114 which is less than 0.05 hence we fail to reject H_0, therefore there is the existence of heteroskedasticity.

This is not necessarily a surprise because the error variance could well be affected by the spatial dependence in the data, therefore, rejecting the null hypothesis.

This is because of spatial dependence, therefore there is heteroskedasticity in the errors.

3.5.6 Model Diagnostics

In this section, ideally diagnostic aid in dictating and distinguishing between substantive (Lag), and nuisance (Error)

3.5.7 Lagrange Multiplier test for spatial lag and spatial error dependence

From the output above, the lag test returns a chi-squared statistic of 49.05 with a p-value of 0.000015, while the error test returns a chi-squared statistic 3.5024 of with a p-value of 0.06128.

There is a high value of chi-square statistic along with the small p-values on the LM-Lag model indicating that the Lag regression models can be used as the regression model. Indicated by the higher value of chi-squared statistic in the spatial lag test might be used as a better interpretation of the spatial autocorrelation. Table 8 shows the results for the Lagrange Multiplier.

Table 8: Lagrange multiplier diagnostics for spatial dependence

Test	Df	Value	p-value
LM Error	1	3.5024	0.061
LM Lag	1	49.05	2E-05

3.6 Spatial Regression Models

After running model diagnostics using the Lagrange Multiplier, the spatial lag model was found to be the best fit for our spatial data for house prices. In this section, we model spatial regression models, the spatial error model and the spatial lag model. Spatial regression models were used to investigate which variables explain their location in different areas in the city of Bulawayo, this case house prices are investigated to find out which spatial regression model can be used to predict house prices in Bulawayo. The use of spatial regression models is to account for the presence of spatial autocorrelation in these models, and this autocorrelation is important because the researcher need to ascertain whether a spatial distribution is significantly different from the outcome of a random process so that we do not make the mistake of attributing pattern to what is a random distribution. Spatial Error and Spatial Lag models were employed in this study and the results for the two models are shown in this section.

3.6.1 Spatial Error Model

After employing the spatial error model, which is used for correcting for spatial autocorrelation in the data. This is done by including the spatially corrected errors = W + to the model which arise from unobserved features or omitted variables associated with the location under study (Anselin et al., 2001).

After employing the spatial error model obtained similar results with OLS, where the size of land, accessibility to retail outlets, and quality of the physical environment are statistically significant since *p*-values are less than 0.05 while the distance to schools, distance to CBD, and accessibility to health facility are statistically insignificant since their p-values are greater than 0.05. Anti-logging the coefficients will obtain the estimated spatial error model gives as follows:

$$Y_{houseprice} = 33728.73 + 8.50X_{land} - 1.09X_{schools} + 1.33X_{retail} + 1.04X_{health} - 1.09X_{CBD} + 1.27X_{physical} \tag{7}$$

From estimated model 7 for the spatial error model we have an intercept of 33728.73 with a p-value less than 0.05, this implies that by setting all the explanatory variables to 0 we have a house price of $33728.73 which means that the model tells us that on average before one can take into account other external drivers of the house price. This price may be accounting for all the internal factors that a house has internal for example number of bedrooms and bathrooms, and the material used in the construction of the whole physical structure. However, not forgetting the economic theory which states that prices of certain goods or services are derived by the forces of demand and supply. Therefore, using this concept, the amount of $33728.73 may indicate that there is more supply for houses on the market driving it to cost that much or there is demand for houses driving the prices to cost $33728.73.

Positive coefficients of the size of land, accessibility to retail outlets, and quality of the physical environment in model 7 as agreed by (Santos Laurent, 2020) it implies that adding a unit to these variables increases house price by a unit since there are positive relationships on the variables while the negative coefficient on the distance to schools, distance to CBD, and accessibility to health facility imply that adding a unit to these variables reduces the house price by an until because of negative relationships arising from the variables. In the next section, we use the spatial lag model and discover how it differs from the spatial error model.

3.6.2 Spatial Lag Model

In this section spatial lag model was employed to capture the interactions of the dependent variable, in this case, the researcher would like to quantify the structure of the spatial relationships of house prices around the residential areas in the city of Bulawayo. The spatial lag model is a spatial autoregressive model that includes a spatially lagged dependent variable given by W y where W y represents the weighted house price of a neighbour and is the spatial dependence parameter.

The spatially lag model allows the fitted model to capture the spatial relationship of the house price for example in Cowdray Park and house price in Luveve or Njube. Anti logging the coefficients obtained the estimated spatial lag model gives as follows:

$$Y_{houseprice} = 1693.18 + 5.86X_{land} - 1.07X_{schools} + 1.25X_{retail} + 1.02X_{health} - 1.08X_{CBD} + 1.23X_{physical} \quad (8)$$

From estimated model 8 for the spatial error model we have an intercept of 1693.18, this implies holding all the explanatory variables obtained a house price of $1693.18 therefore, on average before one can take into account other external factors of house price is $1693.18.

Positive coefficients of the size of land, accessibility to retail outlets, and quality of the physical environment in model 8 it implies that adding a unit to these variables increases house price by a unit since there are positive relationships between the variables while the negative coefficient on the distance to schools, distance to CBD, and accessibility to health facility implies that adding a unit to these variables reduces the house price by a unit because of negative relationships arising from the variables.

In the next section, the researcher compared the spatial error model and spatial lag model and come up with a best-fitted model that can be used to predict house values in Bulawayo taking into account the spatial relationship.

3.6.3 Comparing Models.

In this section comparing the two models (spatial lag and spatial error regression models), the log-likelihoods of the maximum likelihood estimation were employed.

The proper measures for the goodness of fit are based on the likelihood function and include the maximized likelihood, the Akaike Information Criterion (AIC), and the Schwartz Criterion (SC).

The model with the highest log likelihood, or with the lowest AIC or SC is chosen as the best mode. The results are summarized in the table below:

Table 9: Comparing Spatial regression models.

Models	Log Likelihood	AIC	SC
Spatial Lag Model	-461.4586	940.9	252
Spatial Error Model	-483.7011	985.4	290

From Table 9 the best model to be chosen is the Spatial Lag model with the highest log-likelihood function and the lowest values for AIC and SC as compared to the Spatial Error model.

3.7 Model Prediction

After spatial regression models were employed, the researcher chose the Spatial Lag model to predict house prices in the Bulawayo region after conducting various tests to compare the spatial models. From the analysis made the Spatial Lag regression model was observed as the best model that can be used to predict house prices in Bulawayo incorporating the spatial dependence.

4. DISCUSSION

The assessments of house price determinants using spatial analysis involved several procedures and tests, where the overview of data was obtained. In achieving the first objective of identifying the spatial trends of house prices and their drivers, based on the spatial autocorrelation, the results of global spatial autocorrelation house price and land size observed the highest level of autocorrelation with Moran's I value, (I= 0.658 and I= 0.735) respectively. On the other hand, distance to schools observed the lowest autocorrelation of I = 0.151. Based on these results of global spatial autocorrelation, one can note that prices of homes in the city of Bulawayo are influenced by nearby locations, and other variables like Land size is also influenced by nearby locations gave similar findings by (Huang et al., 2017) and san-tos2020spatial. This implies that changes in the prices of houses in one location will lead to changes in the nearby areas, which share similar attributes in the determination of house prices. Moreover, based on the results of the global spatial autocorrelation using the Monte-Carlo simulation and positive Moran's I autocorrelation.

To achieve the second objective of the study of identifying the hot and cold spot areas in Bulawayo, the Local spatial autocorrelation method is employed to identify these areas. Hotspot areas refer to areas observed to have significantly the highest house price values and cold spot areas with the lowest house price values. The parts of low and medium-density areas were identified to be the sources of hotspot areas and the parts of high-density areas were identified to be the source of cold spots in the city of Bulawayo. Using Moran's I scatterplot to identify clusters in house prices, indicated areas with similar characteristics were clustered together. The areas with the highest house price values were clustered together in the low and medium-density areas, whereas areas with lower house price values were clustered together in the city of Bulawayo.

In achieving the objective of determining a model for predicting house values in the city of Bulawayo, in the study OLS was employed, size of land, accessibility to retail outlets, accessibility to health outlets, and quality of physical environment had a positive relationship with house prices. On the other hand, the distance to schools and the distance to CBD had a negative relationship with house prices. The value of multiple R-squared (0.7414) indicates that the multiple linear regression explains 74.14% of the variance in the data. After running model diagnostics from the OLS regression spatial lag model was observed to best fit the data using the Lagrange Multiplier.

When dealing with spatial data, conventional multiple linear regression is often not enough to account for the presence of spatial regression (Santos Laurent, 2020). By modifying the OLS regression formulation to include a spatial term, by use of spatial regression models the researcher was able to obtain a model that fitted the data better. After running the spatial regression model, between the spatial error model and spatial lag model, it was observed that the spatial lag model was observed to best fit the data compared to the spatial error model. This was obtained by using, maximized likelihood, the Akaike Information Criterion (AIC), and the Schwartz Criterion (SC), where the lowest values for AIC and SC were observed in the spatial lag model. In the study done by Santos (2020), the spatial error model was observed as the best fit for predicting house values however, this study presented the spatial lag model as the best fit this means that house values in Bulawayo are attracted by nearby house values.

With data extracted from the property websites (www.classified.co.zw and www.property.co.zw) and the use of R-studio and GeoDa software, the researchers managed to achieve the objectives of the study. The study identified clusters for high house price values together in the low and medium density and low values for houses clustered in high-density areas. The chapter also presented cold spot areas (relatively low prices) in High-density areas (Cowdray Park and Emganwini) and hotspot areas (relatively high prices) in low and medium areas (Selbourne Park and Bradfield). In model building, the study was able to identify the spatial lag model as the best fit for determining house values in Bulawayo by comparing OLS regression, spatial error, and spatial lag regression model.

House markets involve individual investors investing in the ownership of property, where these individuals are faced with challenges on where to invest or how much does would a property to invest in can cost. The use of spatial techniques allows individual investors to observe where to invest by looking at the cold pots and hotspots areas with low and high prices of house prices, this allows them to check their budget and identify the areas in which they can invest. More so, since spatial techniques were able to quantify the relationships between property values and their drivers, the individual investors can locate the area in which they want to invest to and also consider the drivers of house prices in that area and compare with other places which can determine favorable ones.

4.1 Restrictions and Potential Considerations for Using Spatial Analysis to Analyse Bulawayo Home Prices

Although geographical analysis is a useful technique for comprehending the factors that influence housing prices in Bulawayo, it's important to recognize its limitations and potential topics for further research:

4.1.1 Data limitations:

- **Data integrity and accessibility:** It may be difficult to obtain current, reliable spatial data on features, property characteristics, and environmental elements, particularly in developing nations like Zimbabwe. Outcomes that are deceptive can be produced by sparse or erroneous data.
- **Spatiotemporal resolution:** The costs in the real estate sector are subject to sudden fluctuations due to its dynamic nature. Though such modifications might not be easily accessible, historical data would prove ideal for the investigation to gather.
- **Data grouping:** The research might hinge on the area- or neighbourhood-level aggregated data, which could obscure changes across these divisions and obscure micro-factors affecting specific property rates, contingent upon the information that is accessible.

4.1.2 Limitations in methodology:

- **Data integrity and accessibility:** It may be difficult to obtain current, reliable spatial data on features, property characteristics, and environmental elements, particularly in developing nations like Zimbabwe. Outcomes that are deceptive can be produced by sparse or erroneous data.
- **Spatiotemporal resolution:** The costs in the real estate sector are subject to sudden fluctuations due to its dynamic nature. Though such modifications might not be easily accessible, historical data would prove ideal for the investigation to gather.
- **Data grouping:** The research hinges on the area- or neighbourhood-level aggregated data, which could obscure changes across these divisions and obscure micro-factors affecting specific property rates, contingent upon the information that is accessible.

4.2 Future considerations:

- Taking into account non-spatial elements: Socioeconomic variables that have a substantial impact on the housing sector include household revenue levels, job prospects, and criminal activity. Even though these variables might not directly relate to space, they can nevertheless be included in the investigation by using robust techniques like geographically weighted regression or spatial regression models.
- Dynamic modelling: A more thorough picture of the fluctuations of the housing market can be obtained by including time-varying data and taking temporal variations in those variables influencing home values into consideration. Methods such as dynamic panel data analysis and spatiotemporal models may be used in this.
- Miniature study: Although neighbourhood-level analysis is useful, a closer look at the features of distinct properties and their unique place in the urban landscape can reveal more subtleties while shedding light on pricing differences.
- Incorporating community information and viewpoints into the study can enhance comprehension of regional variables impacting housing affordability and market dynamics, hence guiding the development of additional equitable and welcoming planning approaches.

Geographic data analysis may continue to be extremely important in offering insightful information about the factors influencing property values in Bulawayo, enabling well-informed decision-making for a sustainable and fair housing market, by accepting these limits and investigating potential future research avenues.

ACKNOWLEDGEMENTS:

We would like to thank the Higher Life Foundation, Econet and Mrs Tsitsi Masiyiwa for their continued support and assistance in writing this research.

REFERENCES

Alter, A. and Mahoney, E. M. 2020. Household Debt and House Prices-at-risk: A Tale of Two Countries. IMF Working Papers. WP/20/42.

Anselin, L. 1995. Local indicators of spatial association - LISA. *Geographical Analysis*. 2:94-115.

Assam, A. 2022. Cities have driven house price growth this year, as demand slowed in the suburbs. Director of Halifax.

Bramley, G. 2018. Housing supply requirements across Great Britain: for low-income households and homeless people. National Housing Federation.

Bond-Smith, S. and McCann, P. 2022. The work-from-home revolution and the performance of cities. Socio and economic research council. The productive institute. Uk.

Cevik, S. and Naik, S. 2022. Don't Look Up: House Prices in Emerging Europe. International Monetary Fund. WP/22/236.

Canada Mortgage and Housing Corporation (CMHC). 2022. Canada's Housing Supply Shortages: Estimating what is needed to solve Canada's housing affordability crisis by 2030. https://assets.cmhc-schl.gc.ca/sites/cmhc/professional/housing-markets-data-and-research/housing-research/research-reports/2022/housing-shortages-canada-solving-affordability-crisis-en.pdf?rev=88308aef-f14a-4dbb-b692-6ebbddcd79a0

Canada Mortgage and Housing Corporation (CMHC). 2023. Rental Market Report January 2023 Edition.

Chau, L. D., Wong, K. W., McCord, S. K., & Haran, M. 2022. Factors Affecting Spatial Autocorrelation in Residential Property Prices. Land, 11(6): 1-16. [931]. https://doi.org/10.3390/land11060931

Chen, H., Chiang, R. H., and Storey, V. C. 2012. Business intelligence and analytics:From big data to big impact. MIS quarterly, 1165–1188.

Chetty, R., Sandor, S. and Szeidl, A. 2016. The Effect of Housing on Portfolio Choice. Central European University and CEPR.

Chingwenya, A. 2019. Low income housing problems and low-income housing solutions: opportunities and challenges in Bulawayo. *Journal of Housing and the Built Environment*. 34:927–938 https://doi.org/10.1007/s10901-019-09676-w

Costa, L. F. and Tokuda, E. K. 2022. A Similarity Approach to Cities and Features. S˜ao Carlos Institute of Physics - DFCM University of S˜ao Paulo.

Deloitte. 2022. What is next for residential markets in Europe? Property Index Overview of European Residential Markets 11th edition, August 2022.

Dowans, D. and Wilshaw, R. 2021. A better understanding of the housing market 2021 Places for People Group with research support from the University of Cambridge Centre for Housing and Planning Research.

Englund, P. and Flan, H. 2021. Nordic housing markets and policies: Nordic Economic Policy Review 2021. Mordic council of ministers.

ESRB. 2022. Vulnerabilities in the residential real estate sectors of the EEA countries. European system risk board. The European system of financial supervision.

European Banking Authority. 2022. EBA Thematic note-residential real estate exposures of EU banks risks and mitigants. EBA/REP/2022/24. 10 OCTOBER 2022

Frayne, C., Szczypińska, A,. Vašíček, B. and Zeugner, S. 2022. Housing Market Developments in the Euro Area: Focus on Housing Affordability. Discussion paper 171 | SEPTEMBER 2022. European Commission.

Huang, Z., Chen, R., Xu, D., and Zhou, W. 2017. Spatial and hedonic analysis of housing prices in shanghai. *Habitat International*. 67: 69–78.

Kahn, J. A. 2008. What drives housing prices? FRB of New York Staff Report, (345).

Kamloops. 2020. Kamloops Housing Needs Assessment final report, 2020. https://www.kamloops.ca/sites/default/files/2022-02/scd_27500_kamloopshousing_needs_assessment_final_oct2020.pdf

Lebesmuehlbacher, T. 2022. Housing Markets, Local Constraints, and Monetary Policy in England. Department of Economics Xavier University.

Matthews, S. A. 2006. Geoda and spatial regression modeling. In Center for Spatially In-tegrated Social Science, Population Research Institute. University of California, Santa Bar-bara. http://ibis. geog. ubc. ca/* brian/workshop/GeoDa Spatial Regression. pdf. Santa Barbara, California, USA. p. 2017.

Magwaro-Ndiweni, L. 2011. Contestation in the Use of Residential space: House Typologies and Residential Land in Bulawayo, Zimbabwe. *African Review of Economics and Finance* 3 (1): 40–56.

McKinsey & Company. 2022. Private markets rally to new heights. : McKinsey.com/industries/private-equity-and-principal-investors/our-insights.

Maclennan, D. 2012. Understanding housing markets: Real progress or stalled agendas.

Moran, P. A. 1950. A test for the serial independence of residuals. *Biometrika*, 37 (1/2): 178–181.

Muellbauer, J. 2022. Real Estate Booms and Busts: Implications for Monetary and Macroprudential Policy in Europe. Challenges for monetary policy in a rapidly changing world. European Central Bank.

OECD. 2020. Social housing: A key part of past and future housing policy. Employment, Labour and Social Affairs Policy Briefs.

OECD. 2022 Affordable Housing Database – http://oe.cd/ahd

Oloke, C. O., Simon, R. and Adesulu, A. F. 2013. An examination of the factors affecting residential property values in magodo neighbourhood, lagos state. International *Journal of Economy, Management and Social Sciences*. 2 (8): 639–643.

Ord, J. K., and Getis, A. 2001. ''Testing for Local Spatial Autocorrelation in the Presence of Global Autocorrelation.'' *Journal of Regional Science* 41: 411–32.

Panas, D. 2021. Should we exclude the variable from the model if the p value is greater than the significance level?

Pepper, A., Richardson, R., and Walker, L. 2013. J. coyle-shapiro, k. hoque, i. kessler.

Takats, E. 2012. Aging and house prices. *Journal of housing economics,* 21 (2):131–141.

Santos, L. J. R. 2020. Spatial analysis for house price determinants. Department of Civil, Environmental and Geomatic Engineering, 6 (5).

Tobler, W. R. 1970. A computer movie simulating urban growth in the Detroit region. *Economic geography,* 46:(sup1), 234–240.

Velten, B., Braunger, J. M., Argelaguet, R., Arnol, D., Wirbel, J., Bredikhin, D., Zeller, G. & Stegle, O. 2022. Identifying temporal and spatial patterns of variation from multimodal data using MEFISTO.

Xiang, G., Tang, J. and Yao, S. 2022. The Characteristics of the Housing Market and the Goal of Stable and Healthy Development in China's Cities. *Journal of risk and financial management. China*.

Zhou, Y., Tian, Y., Jim, C. Y., Liu, X., Luan, J. and Yan, M. 2022. Effects of Public Transport Accessibility and Property Attributes on Housing Prices in Polycentric Beijing.