REVISITING URBAN IMMOVABLE PROPERTY VALUATION: AN APPRAISAL OF SPATIAL HETEROGENEITIES USING BIG DATA IN PUNJAB

Shoaib Khalid, Fariha Zameer,
Muhammad Irfan Gill & Anila Shahzad
(CGP #02-054)

2ND RASTA CONFERENCE
Wednesday 1st & Thursday 2nd June 2022
Marriott Hotel, Islamabad

This document is unedited author's version submitted to RASTA.

RESEARCH FOR SOCIAL TRANSFORMATION & ADVANCEMENT
Pakistan Institute of Development Economics
Islamabad

ABSTRACT

This study undertook the urban immovable property valuation in two major cities of Punjab; Lahore and Faisalabad, using big data and advanced spatial analysis techniques to explore the significant impact of location-specific parameters on the urban immovable property prices. In order to compute the immovable property values, we employed Big Data analytics in Geographic Information System (GIS). The traditional hedonic price models give little importance to the spatial characteristics of individual housing units and revolve around the structural attributes of houses. However, the spatial heterogeneity should be considered while appraising the residential property prices since the house characteristics may vary over space. To address this issue, we established different valuation models based on the ordinary least square regression and the Fast Geographic Weighted Regression (FastGWR) model, a scalable open source implementation of python and Message Passing Interface (MPI) that can process millions of observations. These valuation models estimated the total net worth of the residential real estate market in the study area. The results demonstrate the excellent performance of our valuation models and display the spatial heterogeneity with higher accuracy. The valuation models explained the relationship of explanatory variables to response variable up to 75% for Faisalabad and around 85% for Lahore. Results show that the floor area, proximity of health facilities, recreational sites and market places add premium to, while nearness of educational institutions, worship places and solid waste transfer stations or dumping sites lessen the property values in both the cities however, closeness of industrial units and graveyards capitalize negatively in Lahore but positively in Faisalabad.

PREFACE

The largest source of provincial revenue in Punjab is the taxes levied on property transfers, such as stamp duties, mutation and registration fees, as compared to the other tax sources. Since 1980, the District Collector's (DC) real property valuation rates are used to calculate provincial tax liabilities, such as capital value tax, property tax, and stamp duties which are substantially lesser than the market price. In 2016, the Federal Board of Revenue (FBR) introduced new valuation tables for the collection of capital gains tax and withholding tax at the federal level and kept revising since 2018 on annual basis but the rates of revised valuation tables are still significantly less than the fair market values. Magnitude of under invoicing (the percentage difference between the disclosed price for legal documentation or tax payments and fair market price) is very much high in real estate market of Pakistan in general.

This study investigated the spatial heterogeneity of the real estate property values as a basis for the formulation of a sophisticated and a scientific valuation model. Our motivation for taking up this study is that the current system of valuation of immovable properties by the government agencies (DC rates and FBR rates) is inefficient, non-scientific and inconsistent. Further, the official valuation methods do not account for the spatial attributes of the real estate properties that is why their valuation remains far lower than the fair market values of the immovable properties. Moreover, there is no mechanism to record actual market transactions in Punjab. Due to poor official property valuation system and low regulatory oversight, most of the gains go unreported, which in turn gives rise to black economy practices and loss of revenue for the national exchequer. So, there is a dire need to develop a more sophisticated system of valuation of immoveable property based on spatial variables, that will not only bridge the gap between official rates and market rates for the extended revenue collection but also to help the sellers and buyers to avoid market speculation practices which make the property inflated.

We are grateful to the Pakistan Institute of Development Economics (PIDE) for providing research funding to this study under the Research for Social Transformation and Advancement (RASTA) Competitive Research Grants (Project ID # 02-150).

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATION

CC - Chief Commissioner

DC - District Collector

FBR - Federal Board of Revenue

GS - General Spatial Model

GWR - Geographically Weighted Regression

HPI - House Price Index

OLS – Ordinary Least Square Regression

RTO Regional Tax Office

SEM - Spatial Error Model

SLM - Spatial Lag Model

UIPT - Urban Immoveable Property Tax

INTRODUCTION

1.1 Background

Economic value of some commodity is the value that is estimated on the basis of its benefits for an individual. In other words, the degree of utility of certain product for people is known as its economic value. Although there are certain methods to quantify this pretend value, it is difficult to measure the exact economic value of something. Market prices are set, using estimates of economic values and, by taking into consideration the perceptible and imperceptible features of the commodities (Fisher et al., 2015; Gabrielli & French, 2020; Lovett, 2019). Hedonic valuation is a widely used approach to quantify the economic value of a commodity that employs statistical regression analysis of various attributes of that commodity, on the basis of its past transactions. Attributes of a commodity determine the level of its utility for an individual and, in turn influence its market price. Thus hedonic price models are devised on the basis of the impact of each attribute, contributed in the total market price of the commodity. The basic arrangement of a hedonic price model is a functional relationship between the price of the heterogeneous goods and its quality characteristics (Baranzini et al., 2008, 2010; Bateman et al., 2001).

Certain structural and locational features can be identified which significantly influence the value of urban immoveable properties in a specific area using the hedonic pricing approach. There are evidences of locational factors that strongly influence the value of real estate, such as the area of the property, covered space, structural characteristics of the building, walkability, security, provision of urban amenities like electricity, natural gas, solid waste collection, sewerage system, paved streets, clean drinking water, distances to the markets, schools, work places, recreational sites & health facilities, are among the significant spatial factors which mainly affect the value of a residential property. The Hedonic models for the valuation of real estate properties can be established, using non-spatial techniques like ordinary least squares (OLS) method or spatial techniques, such as geographically weighted regression (Boza, 2015; Erickson et al., 2011; Gilderbloom et al., 2015; Machin, 2011; Pace & Gilley, 1998; Pagourtzi et al., 2003; Shabana et al., 2015). Hedonic valuation of real estate properties are usually based on four sets of explanatory variables; the structural characteristics, locational attributes, environmental features and lastly the neighborhood traits whereas selling prices are considered as response variable in a multiple regression (Freeman, 1981).

Mapping the values of urban immoveable properties is important to understand the dynamics of real estate market and to monitor the unrestrained prices which is critical for housing market appraisals(Brown et al., 2020; Gaffney, 2009).It is also needed to review the master plans of cities for sustainable planning in future (Barreca et al., 2020). Valuation maps may assist sellers and buyers (Goix et al., 2019; Wang et al., 2020) as well as the government agencies where ever they need value assessment of the immoveable properties as in case of real property taxation (Chapman et al., 2009; Larson & Shui, 2020). Taxes levied on immoveable properties on the basis of their fair market values, enhance the fairness of the tax because the urban services provided by the government in a particular area maximize the capital benefit for the owners. As the scale of the provision of urban services positively contributes into the market value of real estate properties while the provision of urban amenities require government funds.

1.2 Housing Market and Policies in Pakistan

A well operational housing market is one of the key features of a robust economy but unluckily, the housing market in Pakistan is not functioning well. The housing demand in urban centers is increasing due to the rapid urbanization and rural to urban migration in Punjab. On the other hand, the supply side is performing not well due to a various factors. Large tracts of precious land at hearts of the city cores are not being utilized wisely and locked up unnecessarily. In the absence of master planning, the policies and processes of urban planning prove counterproductive most of the time. Pace of the of urban infrastructure services provision is inadequate because the authorities run short of funds because of inefficient revenue collection system from property taxes which is the main source of funds for improvements. Property rights are weak and there is a mess of regulations that hinder the land development and construction processes. The supply of finance for land acquisition and development is limited from public as well as from the government. Supply of land for housing is becoming scarce in urban centers of Pakistan like Karachi and Lahore due to obsolete and poor regulations, moreover, rate of housing construction has a great mismatch with population growth rate and urbanization. In Lahore, house prices have increased by 6.5 times in 20 years from 1992 to 2012 but the average household income have not improved at the same pace. Pakistani cities have higher increment ratios of house price and house rent to household income as well as construction cost to income compared to the other Asian countries. This quick upsurge in the market has made the housing unaffordable for a huge number of families living in urban centers, giving rise to informal housing and slums. More than one third of urban population of Lahore and around half of the residents of Karachi are forced to live in katchi abadis (Dowall & Ellis, 2009; Haque, 2015; Wani et al., 2020; Yuen & Choi, 2012).

Housing markets in Punjab do not operate well and the case is even worse in the big cities like Lahore and Faisalabad, where affordable housing is scarce, housing deficit is huge and a large proportion of the population is destined to live in informal housing (Malik et al., 2020; Wajahat, 2012). On the other hand, around 75 percent residential plots, in nearly 50 percent housing colonies were owned by certified speculative investors who never construct houses but only make the market inflated, so, affordable housing becomes a big challenge for low-income masses (Zaman & Baloch, 2011). The reasons behind this practice include the safest investment opportunities in real estate sector, handsome returns on investment, non-levying of any taxes on vacant residential plots by the authorities, and most of all the easy prospects to evade taxes through under-invoicing at the time of sale-purchase because of the flawed valuation system. The price per *Marla* of vacant residential plots rises at a rate of 41 to 140 percent per annum in Lahore (Gul et al., 2018) but the increments in official values do not match the fair market prices.

Residential property prices are highly explosive even for short run such as for monthly periods in Pakistan. The main reason behind sharp upward trend of the house prices is high demand and short supply due to speedy population growth and quick urbanization. This situation suits the investors to reap higher profits but burdens the public at large. Authorities seem not serious to address the issue as housing policies are inconsistent and not published on regular basis. Launching of affordable and large scale housing schemes by the public sector can only stabilize the ever growing house prices (Ahmed et al., 2021).

1.3 Immoveable Property Valuation by FBR

FBR issues circular on annual basis, directing each Regional Tax Office (RTO) across the country to constitute committees comprising Chief Commissioner (CC) along with three more officers from the respective RTO, a representative of property dealers and a representative of the Association of Builders and Developers of Pakistan (ABAD), both the later representatives are to be nominated by the concerned CC. The committee assesses the valuation tables and hence a set of revised valuation tables then notified for the use of tax imposition within the jurisdiction of the RTO. A representative of the CC may physically visit the areas observing major valuation changes (FBR, 2020). All the procedure seems an exercise based upon individual discretion of the CC. Moreover, no scientific method of calculation is considered. Table 1 shows that how much the market rates remained higher than the FBR valuation table rates.

Table 1. Price per Marla (PKR) of FBR and the Property Portal of Zameen.com

DHA Lahore	Aug-2016			Feb-2019			Jul-2019			Jan-2020		
	FBR	Zameen	Difference	FBR	Zameen	Difference	FBR	Zameen	Difference	FBR	Zameen	Difference
Phase I	672000	1908900	184%	806400	2140200	165%	880000	2190375	149%	860000	2144250	149%
Phase II	552000	1935900	251%	662400	2240325	238%	880000	2124000	141%	800000	2183400	173%
Phase III	552000	3243600	488%	662400	2909250	339%	960000	3082950	221%	800000	3176325	297%
Phase IV	525525	2095875	299%	630630	2482650	294%	1080000	2591550	140%	850000	2478150	192%
Phase V	420000	2783250	563%	504000	3072150	510%	1240000	3221100	160%	900000	3313350	268%
Phase VI	405000	2184975	440%	486000	2400750	394%	1100000	2461950	124%	850000	2450025	188%
Phase VIII	-	-	-	378000	2522700	567%	840000	2577600	207%	500000	2557350	411%
Rahbar	-	-	-	405600	2403450	493%	510930	2403450	370%	400000	2395125	499%

Source: Zameen.com and FBR.

1.4 Immoveable Property Valuation by District Collector (DC Rates)

District Collector's valuation (DC rates) remained even poor than FBR valuation as DC rates have had based on just average of the past disclosed property transaction prices by the tax payers, of the localities till 2019. Sellers and buyers do not reveal the actual transaction amounts but get registered minimal values to avoid taxes and scrutiny by the authorities (Dowall & Ellis, 2009). DC rates notified in 2020 for Lahore seem some better than the past. We assume that this office have adopted the method, no more different than the FBR, details of the latest procedure is not in our knowledge yet but the rates are still far lower than the fair market prices. The following table may tell the whole story. Percentage difference in the table shows how much fair market rates remained higher than DC rates.

Table 2. Price per Marla (PKR) of DC Rates and the Property Portal of Zameen.com

DHA Lahore	2016			2017			2020		
	DC Rate	Zameen	Difference	DC Rate	Zameen	Difference	DC Rate	Zameen	Difference
Phase I	560000	1773000	217%	600000	1778400	196%	962500	2144250	123%
Phase II	460000	1967850	328%	520000	2026800	290%	962500	2183400	127%
Phase III	460000	3138975	582%	520000	3086550	494%	1100000	3176325	189%
Phase IV	420000	2249325	436%	490000	2255400	360%	825000	2478150	200%
Phase V	280000	2606175	831%	450000	2801250	523%	962500	3313350	244%
Phase VI	270000	2141325	693%	420000	2230875	431%	687500	2450025	256%
Phase VIII	-	-	-	320000	2293425	617%	550000	2557350	365%
Rahbar	-	-	-	-	-	-	378000	2395125	534%

Source: Zameen.com and BOR.

1.5 Significance of the Study

Real estate market of Pakistan can be a vibrant source of economic growth, since around 60-70% of the total wealth of the country is stored in its real estate assets. Real estate market contributes around 2% to the Gross Domestic Product of the country while the combined contribution of housing and construction sectors is nearly 9%. Federal Board of Revenue valued the real estate sector of the economy of Pakistan at about US$700 billion. Return on Investment in the real estate market of Pakistan is exceptional, which could be over 100% (Ouattara et al., 2018). Pakistan is home to over 207 million people, and population is growing at a faster pace comparatively, with an average annual growth rate of 2.4%. The average household size in Pakistan is 6.5 persons per household (Pakistan Bureau of Statistics, 2019). The year-on-year housing requisite of the country is 7 lacs units per year but only about half of this demand is met. Around 1 million housing units per year are needed to meet the prevailing prerequisites only. The urban housing deficit is estimated to be around 4 million units, while it is around 6 million units in rural and peri-urban areas, so, the overall housing shortfall is around 10 million units and it is growing. Around 32 million households in Pakistan have no shelter or destined to live in very poor accommodations (S. B. of Pakistan, 2019; Rizvi, 2018). Launching of affordable housing projects is need of the hour in Pakistan. This proposed study can help the planners and policy makers to identify the locations for affordable housing projects in urban centers like Lahore.

Whenever government have to acquire land for new infrastructure building or some particular project, the compensatory payments to the owners are made according to DC valuation tables which are always far lower than fair market values that results in mass protests, demonstrations and unrest (Sabir et al., 2017). So, more realistic valuation of real properties close to the fair market price is essential for a rational compensation to the landowners other than tax purpose as well (Malaitham et al., 2020).

A precise housing property valuation is very much important to facilitate the real estate market stakeholders such as sellers, buyers, property agents, investors, financer banks and insurance companies. An accurate valuation may not be carried out by ignoring the locational attributes and considering only the physical characteristics of houses. In fact, the location of a house sometime capitalizes more in the price than its structural features (Mankad, 2021).

Both the land use change in cities and the urban planning by the authorities are symbiotic (Liang et al., 2018), therefore, a precise prediction of the urban land use growth in rapidly expanding cities like Lahore and Faisalabad is important for an operative, adaptable and sustainable urban planning. A number of public properties such as the Government Officers Residences (GORs), Railway lands, and properties held by evacuee trust and Auqaf department etc., are laying idle or they are not being utilized at their full potential in terms of commercial usage. The public sector alone, owns urban properties worth of trillions of rupees in the entire territory of Pakistan without any use, and consequently forms a locked-up capital since it cannot be used for wealth generation while valuable land in cities remains underutilized. All these insights are useful for the policy makers.

1.6 Objectives

The general objective of this study is to establish a valuation model for urban immovable properties based on the spatial attributes, by utilizing the big data (n ≥ 1.2 million) in two big cities of Punjab, i.e., Lahore and Faisalabad. The specific objectives were;

- To examine the dynamics of urban immovable property values on the basis of location specific parameters using the Big Data.

- To investigate the spatial variations in the urban immoveable property values.
- To undertake the valuation of the government owned real properties (valuation of dead capital).

LITERATURE REVIEW

2.1 Hedonic Real Estate Valuation and Market Characteristics

Hedonic models are employed to estimate the impact of various factors that affect the market price of real estate properties. There are two key assumptions of a hedonic valuation model; first is that the market is perfectly competitive as well as both buyers and sellers are perfectly informed, and the second that there are no discontinuities in the product range available to buyers. There are certain extensions of hedonic pricing model which relax these assumptions, applicable to the data generating process (DGP). Hedonic valuation approach is important, not only to estimate price function in the housing market but also to explore the underlying preferences of the people for housing characteristics in a certain vicinity. (Freeman, 1981; Taylor, 2008). A housing market is somewhat different from a perfectly competitive market because the products in a housing market are distinguished to varying degrees, which means that there are discontinuities present in the product range available to buyers. The information about the quantity and quality of the features that constitute the price of the product in a housing market is difficult and costly to obtain. The selling transection in a housing market becomes more of a process rather than an event. This selling transection process may influences the market price of the houses. The influence of the selling transection process should be estimated through the inclusion of time-on-market variable into the hedonic house price model (Knight, 2008).

Floor area and structural characteristics of a residential property no doubt play a key role in determining its value (Xiao et al., 2017) but locational attributes are also equally important, so, inclusion of this set of variables may yield a true assessment and spatial hedonic valuation models work well in this regard (Helbich et al., 2013; Koschinsky et al., 2012). Whereabouts of residential properties strongly effect their value. House buyers intend to pay more for easy access to transport facilities for example (Seo et al., 2014; Tian et al., 2017). Relevant literature suggests that house values and proximity distances to certain amenities & urban services are strongly correlated but show a mixed affect, either positive; such as nearness of shopping facilities, gyms, metro stations, work places and educational institutions or negative; such as nearness of supermarkets and solid waste dumping & waste transfer stations, . Houses in suburban vicinities fetch less prices as compared to the inner-city areas. Moreover, government policies, street & road conditions, water supply and security affect house prices significantly (De & Vupru, 2017; Xiao et al., 2017; Yang et al., 2018b; Zhang et al., 2018). Real estate valuation methods can be categorized into two classes; one group deals with spatial dependence and the second group deals with spatial heterogeneity (L. Krause & Bitter, 2012).

2.2 Real Estate Valuation and Spatial Dependence

Usage of spatial dependence approach in immoveable property valuation research has long past standing and huge volume of literature can be traced in this regard. The degree of spatial autocorrelation among independent observation values in a geographic space is known as spatial dependence (Crawford, 2009). Spatial autocorrelation means the presence of tendency between the observation values to be more similar when they are situated closer and more dissimilar when they are positioned at distant sites in geographic space. It is considered positive when high observation values cluster near highs and low observation values found near lows while it would be negative when the case is opposite (Griffith, 2004; Hubert et al., 1981). According to the proposition of the hedonic price theory, a number of utility bearing features contribute in the value of heterogeneous commodities and these are the quality features that give rise to, or lessen the utility of the users, not the commodities themselves. Thus, econometrically, hedonic price indexes are constructed by regressing the commodity price on quality features to estimate the implicit prices. Dissimilar combinations of these features affect the preference and utility of the users differently, so, the impact estimation of all possible attributes is important for a good valuation appraisal (Lancaster, 1966; Rosen, n.d.). User's willingness to pay for a certain commodity feature denotes the hedonic price of that feature under the assumption of consumer utility maximization. The price function between attributes of a commodity and its price can be created using historical transactions to estimate the implicit hedonic prices of specific features of the commodity. As the attributes of houses vary significantly, a hedonic price model can be established to value their characteristics separately. Landscape characteristics of houses affect their prices remarkably, so, an appraisal of housing valuation on the basis of location-specific parameters is imperative (Goodman, 1978).

Can, (1990) reported that inclusion of spatial attributes spillover yields more accurate appraisal of the urban immoveable properties. Can & Megbolugbe, (1997) maintained that the spatial dependence, hidden in the residential real estate data sets, affects the accuracy of value estimation considerably. Liao & Wang, (2012) noted a u-shaped pattern of spatial dependence among the house prices in the city of Changsha, China which meant that proximity of high and low priced houses tend to contribute more positively in the implicit house prices while proximity of medium priced houses tend to contribute comparatively less. They informed that the house prices tend to be lower in densely built-up surroundings that was a surprising result because it contradicted with the results of the western city market appraisals where housing units in central business districts found always expensive than the houses at distant locations.

2.3 Expansion Method and Spatial Heterogeneity

Occurrence of spatial heterogeneity is a recognized issue in real estate datasets and a common solution to this issue is to segregate the geographic area of interest into homogeneous regions with specified functions while undertaking value estimation (Kauko, 2003). Spatial heterogeneity can be defined as the variability of a qualitative or measurable value in a distribution in space (Dutilleul & Legendre, 1993). So, for the purpose of this study, the phenomenon of spatial heterogeneity can be referred to when real estate properties with similar characteristics are valued in a different way in different localities across the space of interest. Seeking to examine the spatial heterogeneity in real estate valuation has also a long history that started with the development of expansion techniques by Casetti, (1972). The expansion method contains a procedure for reassembling a simple initial mathematical model to broaden its scope by redefining a few or all of its parameters. A spatial version of a non-spatial model can be created through this method by redefining the parameters into functions of spatial variables that index a fundamentally pertinent framework of that initial model (Casetti, 1997).

Can, (1990) employed the spatial expansion method to construct a hedonic regression model for the apprehension of spatial variation of residential property prices with respect to the influence of neighborhood attributes. Anselin, (1995) formulated local indicators of spatial association (LISA) to account for spatial heterogeneity along with spatial dependence and declared that LISA is a statistic capable of explaining any observation value with respect to nearby spatial clustering of similar values where the sum of LISAs for all observation values may be proportionate to an indicator of global spatial association. Brunsdon et al., (1996) designed geographically weighted regression GWR to capture spatial non-stationarity using the expansion technique and redesigning the parameters of Kernel Regression (KR). It is a local type of spatial statistic that estimates and maps the actual parameters for each observation location in space and thus allowing the visual analysis, instead of estimating the parameters fitted to a trend surface as in the case of global techniques. Fik et al., (2003) used interactive variable approach to capture spatial heterogeneity in house prices and argued that the magnitude of the effect contributed in the house prices by different externalities is always unique at different locations, so, the absolute location should always be regarded interactive with other contributing variables. Bitter et al., (2007) inspected the spatially varying relationship between house prices and housing attributes in Tucson, Arizona, USA. They compared the working of global spatial expansion method with that of GWR and conveyed that GWR bears more prognostic accuracy and explanatory power than that of spatial expansion method. Results of their empirical study strengthened the notion that the degree of the effect of housing attributes on house price differs over space. The matter of spatial heterogeneity can only be resolved through some model with spatially varying coefficients such as submarket based dummy variable models, spatial expansion methods or geographically weighted regression (GWR) models. GWR offers higher accuracy and has an exclusive benefit of providing the visual spatial distribution of implicit values (Wen et al., 2017).

Capturing spatial heterogeneity is critical for precise house price appraisals because real estate property prices differ from place to place within urban centers and are subjective significantly to the provision of urban facilities, however, the magnitude of the effect of these urban facilities is not the same at every location in space (Redfearn, 2009; Yuan et al., 2020). Wu et al., (2020) reported that housing submarkets can better be identified using spatial heterogeneity approach instead of spatial dependence. As spatial dependence approach overlooks the complexity of the urban space. Jiang, (2018) advocated the superiority of spatial heterogeneity approach based on scaling law over spatial dependence approach based on Tobler's law and argued for a required paradigm shift from the use of Euclidean geometry to the fractal geometry for geospatial analyses in cases of working with big data specially.

2.4 Using Big Data for Real Estate Market Appraisal

A data that is very large in size, complex and continuously piling up over time at some media platform which is difficult to be processed and interpreted using the usual traditional techniques is referred to as 'big data' such as the data from social media, transactional data of organizations and machine generated data. When this data is geo-referenced too, it can be called spatial- big data or geo-big data (Dalton & Thatcher, 2015; Gao et al., 2017; Goodchild, 2013; Guo et al., 2014). Wu et al., (2016) used big data in the form of the recorded check-in data from Sina Visitor System (a social media platform) to analyze the willingness of home purchasers to pay for different urban amenities in a Chinese city of Shenzhen. Yang, Chu, et al., (2020) scrutinized the capitalization effect of proximity as well as accessibility to a bus rapid transit (BRT) system on residential property value within a 1.5 kilometer corridor on both sides of BRT lanes in Xiamen Island, China. They fetched the big data through a web crawling program from Fang.com which is a main stream property portal in China. Singh et al., (2020) extracted the big data regarding housing sale prices and other variables from various internet platforms using five software packages of data mining, manipulation and classification such as gdata, tidyverse, stringr, lubridate and caret. Ma et al., (2020) argued that the use of big data for real estate market appraisal is advantageous as it can provide large number of factors, offering the researcher more variety of features to select from, and a chance to have deeper insight to mine more veiled information that might be disregarded while using traditional datasets and methods. Furthermore the analysis of big data help to model the issues more accurately because the field-specific algorithms seem robust enough in identifying the relationships as compared to the traditional statistical techniques.

2.5 Real Estate Valuation Models

Ordinary least square (OLS) is a type of linear regression, used to estimate the unknown parameters and to test the nature of the relationship between a dependent and explanatory variable/s, in hedonic valuation models (Dismuke & Lindrooth, 2006; Frost, 2019). OLS being a global regression technique, assumes that the relationship between regressend and regressors remain stationary and constant on all the observed locations across the whole area under study, so, it uses a single equation to estimate the relationship (Wooldridge, 2016). A number of hedonic pricing studies used OLS for parameter estimation but most often, the regression residuals create spatial autocorrelation, which is against the optimal assumptions of randomness (Pace & Gilley, 1998). Measuring spatial association is very much important while working on a spatial model because spatial autocorrelation may lead to critical errors in model interpretation. Two global techniques, Moran's I and GI* are best choices among many other tools to identify spatial patterns of the dataset and degree of spatial association, they also enable a researcher to detect local "pockets" of dependence when used jointly. A combination of both the techniques (Moran's I and GI*) may better help a researcher, working with a spatial model (A. Getis, 2008; A. Getis & Ord, 2010a).

Further, an asymptotically normally distributed statistic such as Oi, can be employed to test for a local spatial autocorrelation and hot spot analysis, in the presence of global autocorrelation because type-1 errors may occur without considering the global autocorrelation structure in the dataset. Type 1 error is mostly assimilated with false positives, it happens in a hypothesis testing when the null hypothesis is rejected, even if being true. So, a local spatial statistic must be interpreted according to the degree of global spatial association present in the data. Oi is best suitable tool for hotspot analysis with larger datasets however it requires the assumption of spatial stationarity (Ord & Getis, 1995, 2001). Spatial non-stationarity within a spatial data can be explored using spatial cluster analysis with the help of a combination of two tools; GI* and K-means. K-means is a multivariate cluster identification technique that divides data into homogeneous groups, taking into account of the geographical location of features and their spatial relationships. Spatial cluster analysis assesses the degree of spatial autocorrelation between features and quantifies the statistical significance of identified clusters (Peeters et al., 2015).

Presence of spatial dependence is a regrettable necessity in any real estate dataset which can be modeled through a spatial weight matrix and then incorporated into either a Spatial Lag Model (SLM) or a Spatial Error Model (SEM) (Kim et al., 2003). Spatial Lag Model (SLM) also known as Spatial Autoregressive Model (SAM) and Spatial Error Model (SEM) are two widely used basic models to capture spatial dependence in real estate market analysis. SLM or SAM considers a linear combination of real estate property values in nearby space while SEM considers the impact of location-specific variables on dependent variable and the resulting spatial autocorrelation in the error terms. A General Spatial Model (GSM) capable of yielding better estimation results may be established by combining both the SLM and SEM models of spatial dependence that can be developed by imposing restraints on the Spatial Durbin (SD) model (Brasington & Hite, 2005; LeSage, 2008).

Geographically Weighted Regression (GWR) is being employed extensively in hedonic valuation modeling to explore spatial heterogeneity. It is a computation intensive tool which estimates location-specific parameters. GWR is a weighted least-squares regression that assigns weight some observations more than others in computing the regression coefficients. The weights change at each point location, which is why GWR is known as a local model technique. GWR is a "distance-decay" model of spatial association, which follows Waldo Tobler's first law of geography. The influence of the calibrated data points on weights, keep decreasing with the increasing distance around, from a certain location of interest. Most of the time, the distances are taken in a straight-line from one location to another, known as Euclidean distance (ED) in GWR (Brunsdon et al., 1996; A. S. Fotheringham, 1997; A. S. Fotheringham et al., 2015; Scott & Janikas, 2010).The coefficients are taken as functions of a specific spatial location in GWR. A coefficient surface (as opposite to a trend surface in global models) of the whole area under study can be created with a set of continuously varying, estimated coefficients at every location, to explore the spatial heterogeneity, using GWR where each set of regression coefficients is estimated by weighted least squares (Brunsdont et al., 1998; Lu et al., 2014). Gollini et al., (2015) introduced GW model and its implementation in R-package to calibrate and map the localized relationships of covariates using a moving window weighting (MWW) technique.

Traditional GWR techniques seem fail in case of larger datasets with millions of observations. The processing time increases exponentially with increasing number of calibration locations and GWR scales may take more than two weeks to compute results on a model with one hundred thousand observations (Harris et al., 2010). Yu, (2007) and Feuillet et al., (2018) had to split their datasets into a number of sub-sets of geographic units due to the computational limitation of traditional GWR tools at the expense of compromised results and reduced utility of their study because sub-setting created bias under boundary and zoning effects which posed hurdle in capturing the true spatial heterogeneity. Fotheringham et al., (2017) attempted to improve the scalability power of the traditional GWR by relaxing its assumption that all the coefficients being estimated in the model, function at a solo spatial scale, and assigning each relationship of covariates a unique spatial scale instead, to minimize concavity and mitigate over-fitting. They named the new improved extension as multiscale geographically weighted regression (MGWR), it basis on Bayesian spatially varying coefficients (SVC) model and uses a back-fitting algorithm in bandwidth vector selection and model calibration. Oshan et al., (2019) introduced a Python-based mgwr software package, for the deployment of MGWR model as an alternative of the original R-based package with enhanced efficiency in terms of kernel selection. They incorporated an adaptive kernel option whose width changes according to the spatial distribution of observations automatically. Z. Li & Fotheringham, (2020) upgraded the computing ability of MGWR to make it applicable for a dataset with observations up to one hundred thousand, by extending the parallelization framework through a method of splitting the calibrations block-wise. They compared this new version with the original one and found the improved version five hundred times faster in calibrating the model.

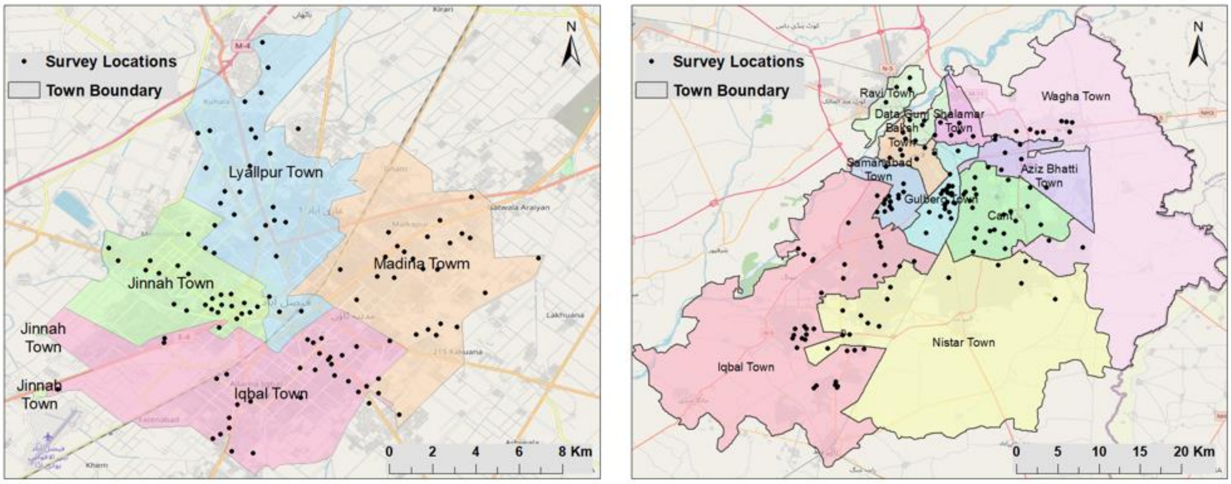
Li et al., (2019) developed FastGWR which was reported to be a more sophisticated extension of GWR, capable of computing and calibrating a data with millions of observations. They compared the computation time performance of four open-source GWR software packages, FastGWR, MGWR (PySAL), GWmodel, and spgwr, with subsets of different sizes from a total of 1.28 million house price observations. They found FastGWR, 12 to 3400 times faster than the rest of the implementations when applied to a dataset of 10 thousand to 15 thousand observations while on a dataset with 20 thousand observations, only FastGWR was able to successfully calibrate a GWR model. FastGWR reduces the memory constraint significantly from $O^{(n2)}$ to $O^{(nk)}$, here 'n' denotes the number of observation locations, and 'k' denotes the number of explanatory variables, further, it uses parallel model diagnostic calculation procedures which considerably decreases the required computation time for a GWR model calibration.

RESEARCH METHODOLOGY

3.1 Study Area

This study covered two major cities of Punjab province i.e., Lahore and Faisalabad.

Figure 1: The Study Area Faisalabad and Lahore



3.2 Data and Sources

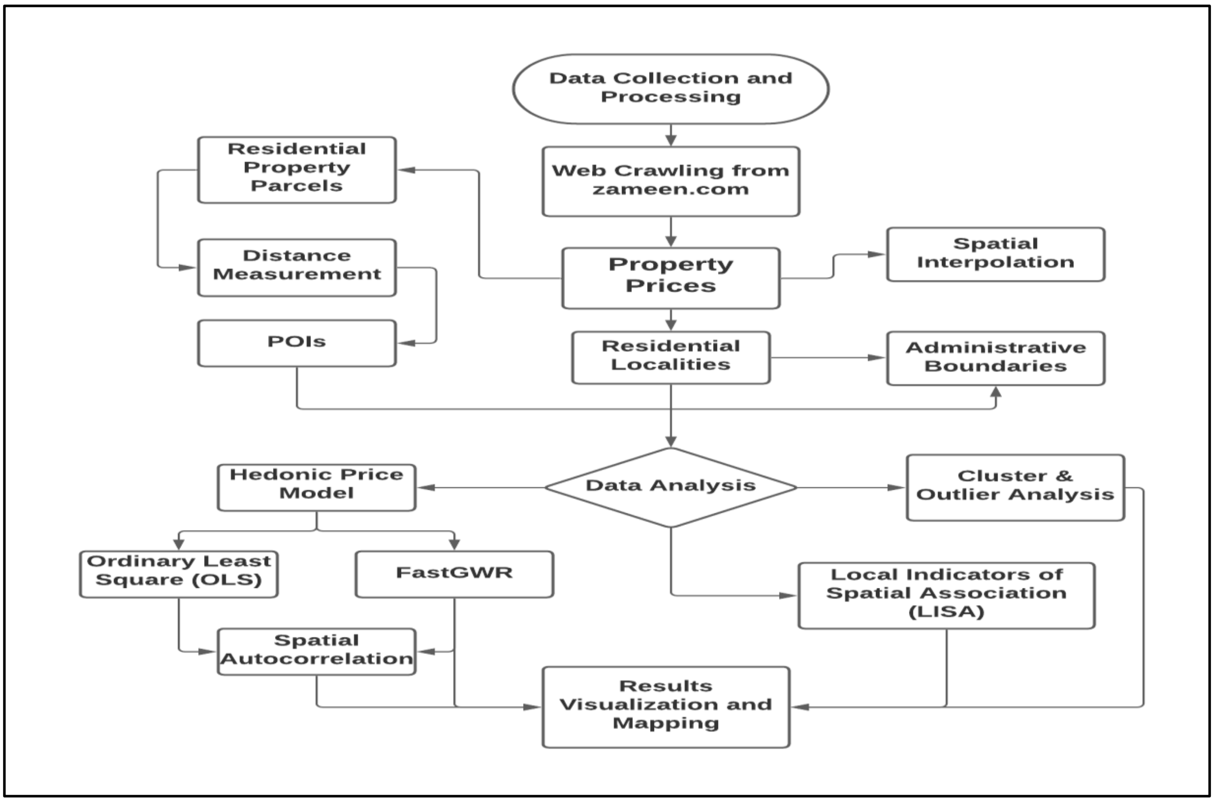
The data used in this study were collected from different sources including governmental, non-governmental organizations and the web-scraping programs. A variety of the data types have been used in this research work such as property price data, house parcels, parcel types, road network, urban land use and location of important places. There are several online portals for property buying and selling in Pakistan including, www.prop.pk, www.zameen.com, www.homespakistan.com, www.realproperty.pk, www.pakrealestate.com, www.aarz.pk, www.realproperty.pk, www.graana.com and www.propertylink.pk. Among these property portals, Zameen.com is the biggest online portal, which provides the services for buying, selling, and renting all type of residential and commercial properties. The services of this online portal are available in more than ninety cities in Pakistan. It records the location, price, and the structural attributes of houses including the area, number of bedrooms, bathrooms, and number of stories of houses. The information about the house price and the location of property provides us the opportunity to analyze the spatial distribution of property values across the city. We obtained this information from zameen.com through a web-scraping program since the website does not offer any Application Programming Interface (API) to download the information. The web-scraping program can download the required information through the browser at a low cost (Li et al., 2017). The data was downloaded for the total number of 3526 houses in Faisalabad city and geocoded the properties on map. It is noteworthy to mention that these are not the transaction costs but the price tags of the residential properties advertised by the homeowners or the real estate agents. We used these property listings which are usually 3 to 5 percent higher than the actual transactions as a proxy of actual market prices since there is no mechanism to record the real estate transactions in Pakistan. Although the selling price is disclosed at the time of property transfer, for tax payments but the level of under-invoicing is very high and the official records only account for 5 to 50 percent of the fair market prices of the real estate properties (Wahid et al., 2021). We acquired the property listing data through web scraping for the year 2019. It is noted that though inclusion of the temporal stability of the transactions would have been an ideal situation, but such information is typically not available in most of the developing countries and Pakistan is no exception. The geometry for house parcels was generated through various methods such as digitizing of detailed property maps, satellite images and a part of parcel dataset were obtained from the Urban Unit Faisalabad, an organization dedicated to provide urban sector planning and management services. The dataset contains the information about the type of property, taxable or exempted, and the location of the property. The locational and environmental attributes were generated in the study area using GIS and the mapping services of google maps.

Property datasets, containing seller asking price of 26031 geo-referenced real estate properties of the city of Lahore and 3526 properties of Faisalabad city were selected for the purpose of price surface generation, advertised on Zameen.com, retrieved through a web-scraping program (Thomas & Mathur, 2019; Yang, Chu, et al., 2020). We used this big data comprising residential real estate listings as a proxy of fair market prices (Wu et al., 2016) because the recorded transactions data from FBR and DC tables suffered by under-invoicing, that comes up to only 5 to 50 percent of the market prices, so, the data from government agencies may not produce appropriate estimations. Although, the listings contain the seller prices, they may be assumed very close to actual transaction values, since the sellers do not decrease the prices by more than 3 to 5 percent on average. Previous relevant literature proposes that the actual transactional price of real estate properties is always significantly correlated with seller asking price, so, using seller asking price as a proxy of the actual transactional price in a hedonic appraisal do not distort the estimates considerably until unless one or more independent variables under study are systematically associated with one of the two prices. Moreover, when demand of the real estate properties exceeds supply, the real estate market turns into seller oriented market and the seller becomes price-setter that is why the actual transactional prices come very close to the seller asking prices in such markets (Ibeas et al., 2012; Salon et al., 2014; Yang, Chau, et al., 2020). Moreover, there is a possibility that a real estate listed on a property portal for sale, does not necessarily mean it was sold, it may be later withdrawn from the listings and not sold but it may give an idea of the seller perception. Relevant literature suggests that these type of listings may be used where there the actual transaction records are not readily available as in the case of the study conducted by Katsalap, (2008) in Ukraine and Yang, Chu, et al., (2020) in China. Landuse parcels datasets of Lahore and Faisalabad were acquired from the Urban Unit Lahore and deficiencies in property parcels dataset of the Urban Unit were met through digitization and geo-referencing.

3.3 Processing Operations

All the following processing operations were executed using ArcGIS 10.4.1 software. The selected properties were displayed using coordinate values for the sake of creating price surface for the whole cities. Raster price surfaces have been generated using the geo-referenced property points as input, through Inverse Distance Weighted (IDW) interpolation with barriers. The IDW is a deterministic method of randomly distributed points, which computes the values of unknown points with the weighted average of known values (Hu et al., 2013; L. Li & Revesz, 2004; S. Li et al., 2017). The points close to the center of cell being estimated gets more weight in the averaging process and it assumes that the values closer to each other are similar than the values at far distances. The raster price surface layers then converted to point layers. Price field was shifted from point feature price surface layers to property parcels and the resultant parcels were converted to points. Near tables were generated within attribute tables of property point layers for all the selected spatial amenities. The near tables contain Euclidean distance in meters from a property point to the nearest relevant spatial amenity point. We used planar method as our study areas were not extended to a large geographic space, moreover the data layers were projected. No search radius was defined, so that every property point may get some value of the distance because the selected spatial amenities were well distributed all across and not confined to a specific part of the cities.

Figure 2: Methodological Framework of the Research Project



3.4 The Spatial Hedonic Valuation Model

A hedonic house property valuation model, based on various attributes has been constructed as following:

$$y = \beta_0 + \beta_1(\mathrm{A}) + \beta_2(\mathrm{d.WP}) + \beta_3(\mathrm{d.HR}) + \beta_4(\mathrm{d.Rec}) + \beta_5(\mathrm{d.Mar}) + \beta_6(\mathrm{d.Ind}) + \beta_7(\mathrm{d.HF}) + \beta_8(\mathrm{d.Gy}) + \beta_9(d.\mathrm{Edu}) + \beta_{10}(d.\mathrm{Ban}) + \beta_{11}(d.\mathrm{Com}) + \beta_{12}(d.\mathrm{SCom}) + \beta_{13}(d.SW) + \beta_{14}(d.\mathrm{AF}) + \epsilon \quad \text{3-1}$$

Where y is the estimated value of the house property, β_0 is the intercept, A is the floor area of a house property, d.WP is the proximity distance to a nearest worship place, d.HR is the proximity distance to a nearest hotel or restaurant, d.Rec is the proximity distance to a nearest recreational site such as park, playground or other recreational site, d.Mar is the proximity distance to a nearest market, shopping center or supper store, d.Ind is the proximity distance to a nearest industrial unit; d.HF is the proximity distance to a nearest health facility such as a hospital or clinic, d.Gy is the proximity distance to a nearest graveyard, d.Edu is the proximity distance to a nearest educational institute such as a school, college, university or a technical training institute, d.Ban is the proximity distance to a nearest bank or automated teller machine (ATM), d.Com is the proximity distance to a nearest commercial building, d.SCom is the proximity distance to a nearest semi-commercial building, d.SW is the proximity distance to a nearest solid waste dumping site / collection or transfer station, d.AF is the proximity distance to a nearest animal farm and ϵ represents error term.

3.5 Variable Selection

Based on an extensive literature review, we initially select fourteen co-variates to be included in the analysis. We included the prime structural attribute of housing properties, the floor area, which explains the value of the properties the most (Gluszak & Zygmunt, 2018), rest of the regressors were spatial in nature. The variables were eliminated one after another on a rolling basis until there was no multicollinearity exited in the dataset. The spatial variables such as commercial places, semi-commercial buildings, banks and ATM machines, restaurants and the animal farms were found multicollinear and these variables were thus excluded from the model. This exclusion results into nine potential explanatory variables qualifying for the final model.

Table 3. Description of the Variables for Spatial Hedonic Valuation Model

Category	Features	Description	Faisalabad	Lahore
Parcel counts	Parcels	Number of total parcels in the study area	416,168	808710
House counts	House Parcels	Number of Residential properties in the study area	268,911	780178
House Attribute	Area Sq. M	Total area of the residential properties in square meters	30.22 M	116.04 M
	Area Marla	Total area of the residential properties in Marla	1.20 M	5.55 M
Valuation	Total Worth	Total Worth of Residential Properties	PKR 2.97 T ($ 17.52 B)	PKR 11.36 T ($ 66.8 B)
	Average Price	Average Price per square meters (per Marl	PKR 98,279 (PKR 2.47 M	PKR 97,897 (PKR 2.04 M)
Amenities	Solid Waste	Number of solid waste facilities and transfe stations	70	1091
	Graveyard	Number of graveyards	72	166
Cultural	Worship Places	Number of worship places (i.e., mosque, church, and temples)	1,409	2,281
Education and Health Facilities	Institutes	Number of educational institutions (schools colleges, and universities)	1,705	4329
	Health Facility	Number of health facilities (hospitals, clinic and dispensaries)	318	2,902
Recreation	Parks and Recreation	Number of public parks and recreational sit	368	1,381
Industrial and Commercial	Industries	Number of industrial units	9,319	8,901
	Market Places	Number of market places	141	5,439
	Commercial	Number of Commercial buildings	66,769	96,685
	Semi-Commercia	Number of semi-commercial buildings	2,679	68,967
	Bank and ATMs	Number of banks and automated teller machines	229	2,153
	Restaurants	Number of Restaurants and Cafes	612	2,793
	Animal Farms	Number of animal farms (poultry and dairy farms)	1,224	2,015

Table 4. Explanatory Variables

Variables	Faisalabad	Lahore
Mean Area (Marla)	4.46	7.11
Mean Area (Sq. Meter)	111.87	148.62
Mean Distance to Worship places (m)	127.15	352.23
Mean Distance to Solid Waste Facilities (m)	798.70	696.50
Mean Distance to Parks (m)	520.96	397.55
Mean Distance to Markets (m)	1,769.26	495.77
Mean Distance to Institutions (m)	129.66	287.90
Mean Distance to Industrial Units (m)	165.34	408.30
Mean Distance to Health Facilities (m)	388.75	440.37
Mean Distance to Graveyards (m)	860.15	978.85
Mean Distance to Commercial buildings (m)	31.18	1464.49
Mean Distance to Semi Commercial buildings (m)	1343.56	1654.77
Mean Distance to Banks & ATMs (m)	1,004.45	811.83
Mean Distance to Hotel / Restaurant / Café (m)	434.17	443.11
Mean Distance to Animal farms (m)	283.94	2332.94

3.6 Interpolation of Property Values

Before computing the property valuation model, the inverse distance weighting (IDW) interpolation is performed in order to obtain the predicted surface of property prices for the entire study areas. The IDW interpolation with barriers and without barriers is applied on the dataset to avoid under or over prediction. The results of the interpolation with barriers and without barriers for property prices is shown in Figure 4 and Figure 6.

Figure 3. Interpolation of Property Prices (A) IDW Interpolation (B) IDW Interpolation with Barriers (Faisalabad)

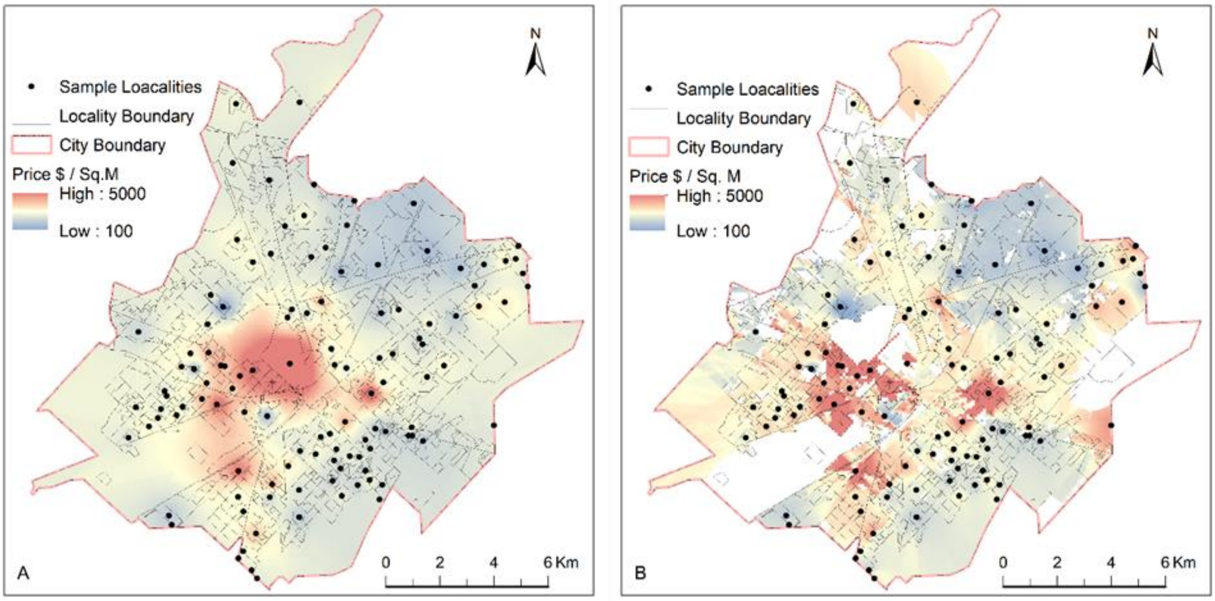


Figure 4: Transferring of Interpolated Values of House Prices to the Locality Boundary and to the Individual House Parcel

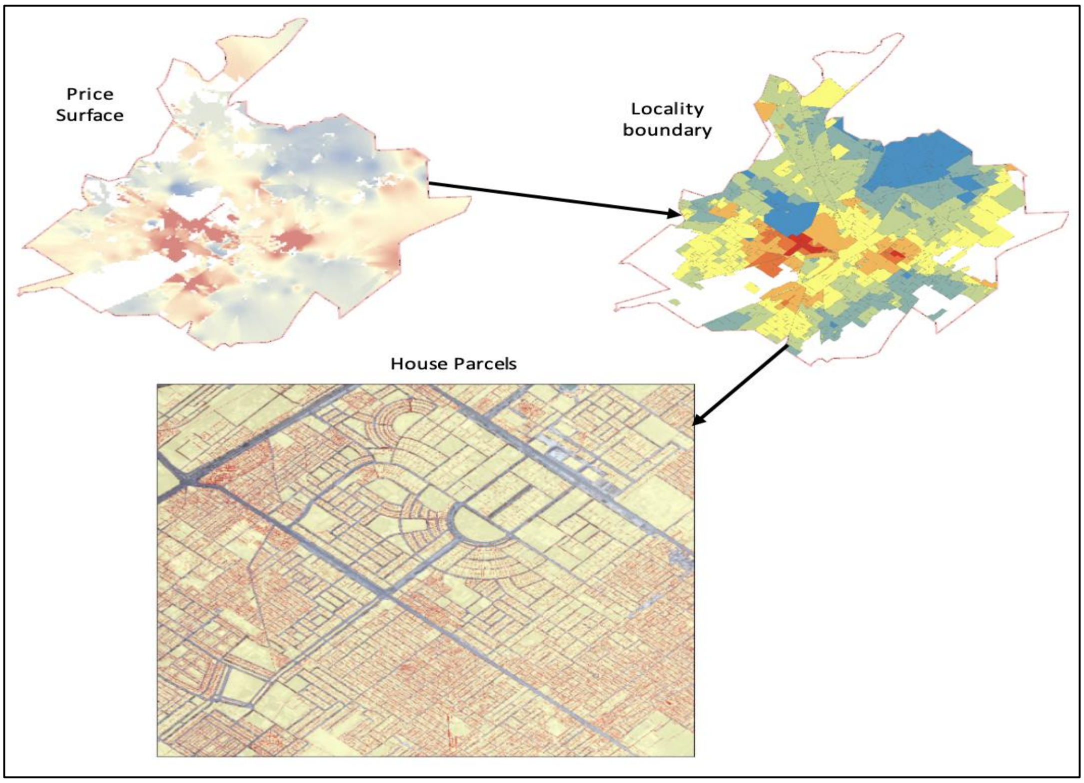
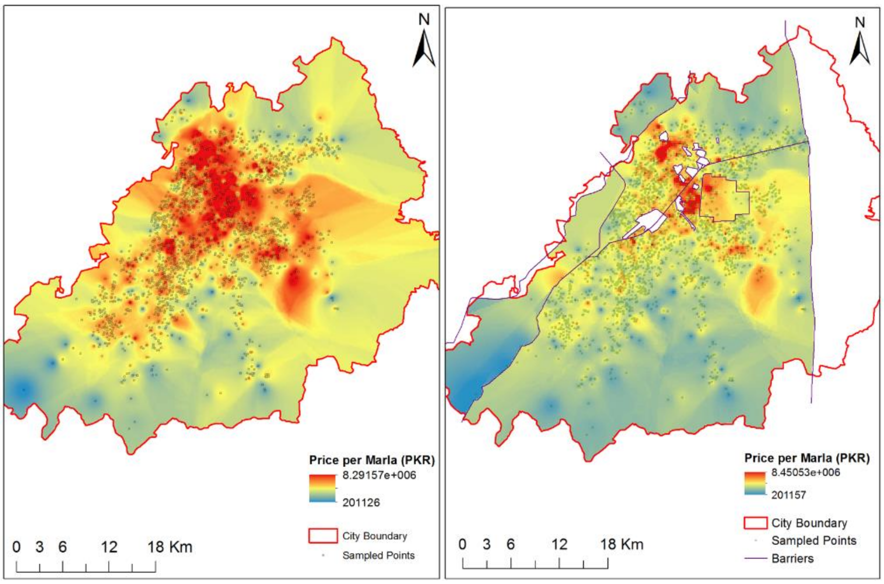


Figure 5. Interpolation of Property Prices (A) IDW Interpolation (B) IDW Interpolation with Barriers (Faisalabad)



3.7 Analyses

Running the OLS- a Linear Global Model

We first performed the hedonic valuation analysis using OLS- a linear global model. OLS is considered the best technique among all the regression methods and used as a proper initial procedure before conducting any of the other regression based spatial analysis. OLS regression tool in spatial statistics toolset of ArcGIS desktop can be used to discover, inspect and model the linear spatial relationship between a dependent and one or more explanatory variables with a global approach. This means that OLS computes the relationship between the variables using a single equation for the whole area under study and works on the basis of the assumption that the relationships remain consistent and stationary at all locations. OLS tool in ArcGIS Desktop, computes the following statistics by default, other than calculating the coefficients of regression, a t-test to compare the means of variables, R^2 and adjusted R^2 to assess the model performance, Variance Inflation Factor (VIF) of each explanatory variable to evaluate the issue of multicollinearity, Akaike's Information Criterion (AICc) to measure the model effectiveness, Joint F test and Joint Wald test to show the model significance and measure the effectiveness of regressors, Koenker (BP) test to inform on the level of stationarity of the relationships and a Jarque-Bera test to assess the bias or validity of the specified model. The value of R2 and adjusted R^2 could be any number between a zero and 1. R^2 value usually rises by including an additional explanatory variable into the model while the adjusted R^2 considers the overall complexity of the data and its value do not increase with a mere addition of regressors, so, its value is always lesser than R^2. A VIF value for any regressor in the model must be less than 7.5 to keep the redundancy within acceptable limit. A model specification with comparatively smallest AICc value is considered the best effective among all the considered versions. The null hypothesis for both the Joint F and Joint Wald tests is that none of the regressors in the model are effective and test results are interpreted on the basis of the significance of the associated p-values but the Joint F test is considered reliable when the p-value of Koenker test of the model is not statistically significant, otherwise Joint Wald test is referred for the purpose. The Koenker test determines the consistency of the relationships spatially and within the dataset, if the associated p-value is significant, there would be sufficient evidence to reject the null hypothesis that the relationships are consistent and static and the alternative hypothesis will be adopted that spatial heterogeneity or heteroscedasticity in the dataset exists. The associated p-value of Jarque-Bera test must not be significant statistically to confirm that the data is normally distributed, if the result is opposite, it can be concluded that the assumption of data normality is being violated (Mitchell, 2005; Scott & Janikas, 2010).

Checking the Level of Spatial Autocorrelation

Global Moran's I test then applied to check the degree of spatial auto-correlation of regression residual values in a global environment. A hotspot analysis was carried out to discover the significant high and low value clustering, using residuals values. Due to the spatial nonstationarity found in the data distributed across the geographic space, the global models such as ordinary least square models are unable to explain the relationships between some sets of variables correctly and produce bias (Fotheringham, et al., 1996). The nature of model must alter over space to reflect the structure within the data. One of the main objectives in spatial analysis is to identify the nature of relationships that exist between variables. Typically, this is undertaken by calculating statistics or estimating parameters with observations taken from different spatial units across a study area. Then Geographically Weighted Regression (GWR) was run which allows the actual parameters for each location in space to be estimated and mapped as opposed to having a trend surface fitted to them. Variations in relationships over space such as those described above are referred to as spatial nonstationarity (Hu et al., 2016). Currently, the global regression models are the leading methods to study the relationships among the geographical or environmental factors and the spatial distribution of a component. These models estimate the average regression variable for each explanatory variable in the entire study area, regardless of location and orientation. Because of the complex urban systems and the strong interaction between different elements, the assumption of spatial stationary relationships may be violated, especially at large geographic scales. Therefore, the local regression model can effectively complement the global model for inferring the relationship at a more precise level (A. Fotheringham et al., 2002).

Cluster and Outlier Analysis

The Cluster and Outlier analysis is applied to this dataset to identify the clustering pattern of property prices across the study area. This analysis identifies the spatial clustering of features with high/low values and spatial outliers using Anselin local Moran's I statistic, if any (Anselin, 1995; Anselin & Le Gallo, 2006). It also computes the z-score, p-values, and a code representing the type of the cluster for the features (house parcels in our case) that are statistically significant (p = .05). While the z-score tells about the strength of spatial autocorrelation, the p-values express its statistical significance. The positive value of 1 specifies that a feature has a nearby feature with similarly high values and this feature is the part of a cluster while the negative value shows that the feature has neighbors with dissimilar values and the feature is an outlier. The p-values for the feature must be small enough for the cluster or outlier to be considered statistically significant. The Cluster and Outlier Analysis is given as;

$$I_i = \frac{x_i - \bar{X}}{S_i^2} \sum_{j=1, j \neq i}^{n} w_{i,j} \left(x_j - \bar{X}\right) \qquad 1$$

where x_i is an attribute for feature i, $\bar{X}$ is the mean of the corresponding attribute, S_i^2 is the variance of feature i and $w_{i,j}$ is the spatial weight between features i and j.

Running the GWR: Local Model

GWR is a commonly used tool for studying the spatial variability of phenomena in a geographic area. It calculates site-specific parameter estimates, making the calibration process more extensive. The GWR explores the spatial non-stationarity, a condition where a linear regression model does not have the capability to explain the relationship effectively between variables over a geographical area (A. Getis & Ord, 2010b). In order to check the spatial dependence, the local geographically weighted regression analyzes the residential property value for each locality. This allows us to incorporate the regional variation in the regression model and explore relationship in much more details (A. Getis & Ord, 2010b; Mitchel, 2005). GWR works by allowing model coefficients to vary regionally. Essentially, the regression model is run for each location rather than the whole study area. A major limitation of the current open source GWR softwares is that the maximum number of data points such software can process is around 15,000 values on a standard desktop computer. In the age of big data, this places severe restrictions on the use of GWR (Li et al., 2019) for larger areas (i.e. entire cities) such as this study intended to do. To overcome this restriction, an open-source Python based application namely FastGWR developed by Li et al., 2019 was used. This application utilizes the Message Passing Interface (MPI), a standard communication protocol for parallel computers whose goal is high performance, scalability, and portability. Through parallelization, the FastGWR optimizes the memory usage to boost the computing performance significantly that can process millions of observations. It also outperforms the existing software packages available for GWR computations. In order to overcome the memory restrictions, the FastGWR algorithm improves the linear algebra within the GWR calibration. The memory requirements are reduced from $O(n^2)$ to $O(nk)$, where n is the number of observations and k is the number of covariates. Given that k is far smaller than n in most GWR application, this approach saves a significant amount of memory. The FastGWR offers parallel model diagnostic computation methods that can be applied to a very large datasets consisting of millions of observations and it considerably decreases the time required for GWR calibration by factors up to thousand times faster than the contemporary existing techniques (Li et al., 2019). While there are various MPI implementations available, in this research the *OpenMPI* is used, which is an open source implementation. Besides the *OpenMPI,* a python wrapper "*mpi4py*" is used to integrate the MPI into the FastGWR algorithm as described by Li et al., (2019). The call to the FastGWR program is as follows:

$$mpiexec - np\ 32\ python\ fastgwr - mpi.py - data\ input.csv - out\ gwr.csv - a - bw\ 1000 \tag{2}$$

where *mpiexec* is the command to execute an MPI-based program; argument – *np 32* indicates the number of processors to allocate; – *data input.csv* is the name of the input data table containing coordinates and associated dependent and independent variables; -*out gwr.csv* is the file containing the GWR outputs that include an ID for each calibration location, predicted values, residuals, local parameter estimates, and local standard errors; -a (-f) indicates the use of an adaptive or fixed bandwidth using a bisquare (Gaussian); and – *bw 1000* indicates a user-defined bandwidth for calibrating the GWR model. In this model the adaptive kernel is used with 1000 meters of bandwidth.

FINDINGS AND DISCUSSION

4.1 Global Model Implementations

Linear model implementation produced adjusted R^2 values of 0.85 for Lahore and 0.75 for Faisalabad, slightly lesser than R^2 values and there were not big differences, which indicates that the models were properly specified because the adjusted R^2 takes into account the overall model complexity and consequently gives more accurate result regarding model performance. We could not trust the Joint-F test result for the reason that the associated p-value of Koenker BP test was significant statistically, so, the Joint-Wald test determines here the overall model significance. Since the p-value of Joint-Wald statistic was statistically significant and far smaller than 0.05 at 95% confidence level, there was enough evidence to reject the null hypothesis and adopt the alternative that the regressors in the model were effective. P-value associated with Koenker BP test was also much lesser than 0.05 at 95% confidence level which pointed toward the conclusion that the relationship among the response and explanatory variables was not consistent and stationary across the study area and spatial heterogeneity exited. That result was expected as the relevant literature already confirmed that the occurrence of spatial non-stationarity in the housing data was a common real life phenomenon because the degree of the effect contributed in the house prices by different externalities is always unique at different locations. Spatial heterogeneity can be reported using a global model like OLS but could further be measured using some local model such as GWR (Bitter et al., 2007; Brunsdon et al., 1996; Fik et al., 2003; Wen, Jin, et al., 2017). A statistically significant p-value of Jarque-Bera test is the indication of the presence of non-normality in the residuals distribution and that the model is not bias free and could not be trusted fully but the relevant literature suggests that a model with a significant p-value of Jarque-Bera test can be trusted when working with a large dataset because it is a proven fact that the assumption of normality for response as well as explanatory variables is not required in an OLS linear model when the sample size is too large. Since distribution of regression residuals depends on the distribution of regression variables, normality assumption can be ignored for residuals distribution as well in case of large sample size (Lumley et al., 2002). When a regression model comes with non-normally distributed residuals, one should look up robust standard errors and robust p-values to conclude if the variable coefficients are significant statistically instead of standard errors and probabilities (Mitchell, 2005; Scott & Janikas, 2010). Since Koenker statistic were noted significant statistically in the model diagnostics, the robust p-values (probabilities) were to be consulted only which were smaller enough than 0.05 for all explanatory variables, so, there were sufficient evidences to reject the null hypotheses of T-Statistic and adopt the alternative that all the coefficients were significant and helping the model. The coefficient values reflect the nature and strength of the relationship of each regressor to the response variable.

The results of linear model for the entire cities of Faisalabad and Lahore are presented in .

Total floor area (in Marla unit for Lahore and, in square meter unit for Faisalabad) showed a positive relationship with the house price which indicates that every additional Marla in floor area, increases the house price by PKR 2202216.09 in Lahore while every additional square meter in floor area, increases the house price by PKR 86256.45 in Faisalabad. Among the spatial variables, distance to nearest health facility, market place, and recreational facility such as park have shown a negative relationship with the response variable, they lower the house price by PKR 2614.52, 1149.4, 1105.2 respectively with a one meter increase in the Euclidean distance, so, it is imperative that people value these amenities and like to live near to them. On the other hand, distance to nearest industrial unit, educational institution, graveyard, solid waste dumping site / transfer station and worship place have shown a positive relationship with the response variable and the house price increases with increasing these distances at a rate of PKR 1921.57, 647.83, 563.58, 429.79 and 234.92 per meter respectively, it indicates that the residents do not prefer to reside near these features in Lahore City. In Faisalabad, the dynamics however are somewhat different from Lahore. Distance to nearest health facility, park, market place, industrial unit and graveyard have shown a negative relationship with the response variable, they lower the house price by PKR 1495.64, 1460.89, 517.08, 130.66 and 45.87 respectively with a one meter increase in the Euclidean distance while distance to nearest educational institution, solid waste dumping site / transfer station and worship place have shown a positive relationship with the response variable and the house price increases with increasing these distances at a rate of PKR 504.57, 393.37, 355.84 per meter respectively. Unlike Lahore, proximity to industrial units and graveyards capitalize positively in the house prices in Faisalabad City. A possible reason of this contradiction may be the abundant occurrence of small industrial units such as power looms and graveyards within residential areas, especially in central parts of Faisalabad city, this phenomenon is not found on such a large scale in Lahore.

Table 5. The adjusted R squared value explains the relationship of explanatory variables to the house prices up to 75% for Faisalabad and around 85% for Lahore in the linear model. However, the robustness of the model is improbable since the residential property prices are less normally distributed. Most of the coefficients for the explanatory variables are as expected and explain the relation between the response variable and explanatory variables.

Total floor area (in Marla unit for Lahore and, in square meter unit for Faisalabad) showed a positive relationship with the house price which indicates that every additional Marla in floor area, increases the house price by PKR 2202216.09 in Lahore while every additional square meter in floor area, increases the house price by PKR 86256.45 in Faisalabad. Among the spatial variables, distance to nearest health facility, market place, and recreational facility such as park have shown a negative relationship with the response variable, they lower the house price by PKR 2614.52, 1149.4, 1105.2 respectively with a one meter increase in the Euclidean distance, so, it is imperative that people value these amenities and like to live near to them. On the other hand, distance to nearest industrial unit, educational institution, graveyard, solid waste dumping site / transfer station and worship place have shown a positive relationship with the response variable and the house price increases with increasing these distances at a rate of PKR 1921.57, 647.83, 563.58, 429.79 and 234.92 per meter respectively, it indicates that the residents do not prefer to reside near these features in Lahore City. In Faisalabad, the dynamics however are somewhat different from Lahore. Distance to nearest health facility, park, market place, industrial unit and graveyard have shown a negative relationship with the response variable, they lower the house price by PKR 1495.64, 1460.89, 517.08, 130.66 and 45.87 respectively with a one meter increase in the Euclidean distance while distance to nearest educational institution, solid waste dumping site / transfer station and worship place have shown a positive relationship with the response variable and the house price increases with increasing these distances at a rate of PKR 504.57, 393.37, 355.84 per meter respectively. Unlike Lahore, proximity to industrial units and graveyards capitalize positively in the house prices in Faisalabad City. A possible reason of this contradiction may be the abundant occurrence of small industrial units such as power looms and graveyards within residential areas, especially in central parts of Faisalabad city, this phenomenon is not found on such a large scale in Lahore.

Table 5. Results of the Linear Models estimates for the study area

Variables	**Lahore**				**Faisalabad**			
	Linear Model Adjusted R^2 =0.85				**Linear Model Adjusted R^2 =0.75**			
	Coeff.	t-Stats	p-Value	VIF	Coeff.	t-Stats	p-Value	VIF
Intercept	356220	20.36	0.000*	-----	3280399	199.14	0.000*	----
Floor Area	2202216	2050.22	0.000*	1.04	86256.45	881.22	0.000*	1.04
D_Worship Places	234.92	6.95	0.000*	3.14	355.84	5.99	0.000*	1.11
D_SolidWaste Site	429.79	24.8	0.000*	2.82	393.37	34.48	0.000*	1.62
D_Parks	-1105.2	-48.86	0.000*	1.85	-1460.89	-119.19	0.000*	1.18
D_Market	-1149.4	-46.33	0.000*	4.68	-517.08	-94.11	0.000*	1.55
D_Institutes	647.83	15.85	0.000*	3.39	504.57	8.3	0.000*	1.2
D_Industries	1921.57	73.14	0.000*	1.54	-130.66	-3.04	0.002*	1.23
D_Hospitals	-2614.52	-109.91	0.000*	3.1	-1495.64	-73.75	0.000*	1.3
D_Graveyards	563.58	39.45	0.000*	1.26	-45.87	-4.51	0.000*	1.51

The model produced results are near to our expectations which were made on the basis of extensive literature review and personal civic experience of the study area. Research outcomes are in-line with the findings of relevant previous studies. Some of the supporting references are given in the following lines. Floor area of a real estate property is the most operative factor to determine its price (Ma et al., 2020). Proximity of shopping facilities augment premium to the house prices (De & Vupru, 2017; Xiao et al., 2017; Yang et al., 2018b; Yang, Chau, et al., 2020; Zhang et al., 2018).

Worship places of every religion contribute in the house prices positively, however, the housing properties facing this feature may experience negative capitalization because of noise and higher number of visitors that create disturbances for the residents (Brandt et al., 2014; Thompson et al., 2012). Urban green spaces, public parks, playground and other recreational sites add premium to the housing properties (Crompton & Nicholls, 2020; Liao & Wang, 2012; Shabana et al., 2015). Residential properties close to graveyards fetch lesser values due to superstitions linked to the burial grounds (Hassan et al., 2021). An industrial neighborhood proved to be negatively influencing factor for residential property prices in most of the research findings (Grislain-Letrémy & Katossky, 2014; Munshi, 2020).

4.2 Local Model Implementations

Table 6. Result of the FastGWR Model estimates for the Entire Cities

Predictors	**Lahore**		**Faisalabad**	
	R^2 = 0.9		**R^2 = 0.78**	
	Coeff.	SE	Coeff.	SE
Area	2202216	2050.22	0.0048	0.00015
D_Worship Places	234.92	6.95	0.00052	0.00041
D_Solid Waste Facilities	429.79	24.8	-0.00038	0.00089
D_Parks	-1105.2	-48.86	0.00094	0.00064
D_Markets	-1149.4	-46.33	0.00443	0.00065
D_Institutes	647.83	15.85	0.00042	0.00043

D_Industry	1921.57	73.14	0.00057	0.00049
D_Hospitals	-2614.52	-109.91	0.00039	0.00062
D_Graveyards	563.58	39.45	0.00283	0.00076

The results of FastGWR model implementation produced adjusted R squared of 78%, for Faisalabad and around 99% for Lahore which show a strong relationship between the house values and the predictor variables. The model produced results are near to our expectations which were made on the basis of extensive literature review and personal civic experience of the study area. Research outcomes are in-line with the findings of relevant previous studies. Some of the supporting references are given in the following lines. Floor area of a real estate property is the most operative factor to determine its price (Ma et al., 2020). Proximity of shopping facilities augment premium to the house prices (De & Vupru, 2017; Xiao et al., 2017; Yang et al., 2018b; Yang, Chau, et al., 2020; Zhang et al., 2018).

Worship places of every religion contribute in the house prices positively, however, the housing properties facing this feature may experience negative capitalization because of noise and higher number of visitors that create disturbances for the residents (Brandt et al., 2014; Thompson et al., 2012). Urban green spaces, public parks, playground and other recreational sites add premium to the housing properties (Crompton & Nicholls, 2020; Liao & Wang, 2012; Shabana et al., 2015). Residential properties close to graveyards fetch lesser values due to superstitions linked to the burial grounds (Hassan et al., 2021). An industrial neighborhood proved to be negatively influencing factor for residential property prices in most of the research findings (Grislain-Letrémy & Katossky, 2014; Munshi, 2020).

4.2 Local Model Implementations

Table 6 presents the results of FastGWR model estimates for the entire cities. The coefficients of FastGWR are positively correlated except the distance to solid waste facility which deceptively indicates that the house values decrease as the distance increases from the solid waste facility. Since these coefficients are the average values that are affected by the high negative values in the results, so we have to examine the results locally. Figure 6c depicts the significant parameter estimates for the predictor variable of distance to solid waste facility. The map shows that the house parcels colored in blue have a converse effect of distance to solid waste facility, meaning that the values of these houses decrease as the distance from solid waste facility increases. One possible reason for this inverse coefficient is that earlier the solid waste facilities were established away from the settlements but the settlements have grown around these facilities presently and the land prices nearby these facilities also raised with time. The low variance inflation factor (less than 7.5 for each regressor) indicates that there is no multicollinearity as we had already eliminated the muiltcollinear explanatory variables while the regression residual are random and not spatially auto-correlated.

Figure 13 shows the distribution of regression residuals with very low R^2 values. In the OLS results table, we need to understand the t-statistics value that evaluates the statistical significance of the explanatory variables. Higher the t-statistics value, the more significant the variable is. This value explains that the area of the house is the most important structural variable for house price in the entire city while the other significant variables are distance to solid waste facility, distance to worship places, and distance to educational institutes respectively. These accessibility variables have the positive coefficients indicating that the house price increases as the distance from these features increases. The other significant variables with the negative coefficients are distance to parks, distance to markets, distance to hospitals, distance to graveyards, and distance to industries respectively. The negative coefficients suggest that the residential property prices are likely to decrease as the distance from locational features increase. These finding are similar to the study of Li et al., (2019) in the city of Los Angeles, California.

The results of ordinary least square (OLS) model for FD-I rating area are presented in Table 7. The semi-log model explains the relationship up to 77% while the linear model explains it up to 80%. The coefficients are as expected but the distance to worship places, parks, markets, educational institutes, and hospitals are negative. This indicates that the house price reduces as the distance from these variables increase. The worship places are a key cultural features in the city that appear to impact the prices of residential houses in a positive way (Brandt et al., 2014; De & Vupru, 2017). In FD-I zone, the average price per square meter is US$875 and the average house price is US$81,320 with an average area of 95 square meters. The t-statistics is suggestive of the order of significance for these negative coefficients, which indicates that the distance from places of worship, distance from parks, distance from health facilities, and distance from market are the most significant influencing locational features, respectively. The coefficient of distance to solid waste facility is negative in the linear model while in the semi-log model, this coefficient is positive, but they are not statistically significant.

Table 7. Results of the Linear Model Estimates for Rating Area FD-I, FD-II and FD-III

Rating Area FD-I	Semi-Log OLS Model R^2 =0.77			Linear Model R^2 =0.80			
	Coeff.	t-Stats	p-Value	Coeff.	t-Stats	p-Value	VIF
Intercept	10.3558	1307.87	0.0000*	14688.1	25.88	0.0000*	----
House Area	0.00985	301.19	0.0000*	793.34	338.31	0.0000*	1.06
D_Worship Places	-0.0006	-30.3	0.0000*	-47.65	-31.82	0.0000*	1.12
D_SolidWaste Facility	2E-06	0.27	0.7859	-0.24	-0.41	0.6831	1.22
D_Parks	-0.0004	-30.17	0.0000*	-30.39	-33.18	0.0000*	1.13
D_Market	-5E-05	-10.15	0.0000*	-6.38	-17.09	0.0000*	1.18
D_Institutes	-0.0001	-5.19	0.0000*	3.95	2.35	0.0186*	1.12
D_Industries	0.00028	18.32	0.0000*	25.08	22.32	0.0000*	1.14
D_Hospitals	-0.0002	-13.58	0.0000*	-16.28	-18.45	0.0000*	1.1
D_Graveyards	0.00016	30.3	0.0000*	12.22	31.22	0.0000*	1.3
Rating Area FD-II	Semi-Log Model R^2 =0.78			Linear Model R^2 =0.81			
	Coeff.	t-Stats	p-Value	Coeff.	t-Stats	p-Value	VIF
Intercept	10.1637	852.36	0.0000*	-10914	-10.54	0.0000*	----
House Area	0.0106	187.93	0.0000*	1015.18	207.08	0.0000*	1.07
D_Worship Places	-7E-05	-1.82	0.068	-2.52	-0.72	0.4705	1.11
D_SolidWaste Facility	-0.0002	-27.3	0.0000*	-21.46	-29.28	0.0000*	1.53
D_Parks	0.00015	7.21	0.0000*	9.03	4.98	0.0000*	1.44
D_Market	0.00005	7.27	0.0000*	0.72	1.2	0.2305	1.54
D_Institutes	0.00016	4.77	0.0000*	19.52	6.7	0.0000*	1.31
D_Industries	-0.0003	-11.34	0.0000*	-22.13	-8.9	0.0000*	1.59
D_Hospitals	0.00026	16.44	0.0000*	16.89	12.07	0.0000*	1.32
D_Graveyards	0.00047	49.54	0.0000*	38.72	46.48	0.0000*	1.69
Rating Area FD-III	Semi-Log Model R2 =0.71			Linear Model R2 =0.79			
	Coeff.	t-Stats	p-Value	Coeff.	t-Stats	p-Value	VIF
Intercept	10.3926	5323.04	0.0000*	17254.9	135.38	0.0000*	----
House Area	0.00761	726.77	0.0000*	612.54	896.17	0.0000*	1.04

D_Worship Places	0.00014	22.37	0.0000*	8.88	21.66	0.0000*	1.12
D_SolidWaste Facility	4.8E-05	40	0.0000*	2.66	34.25	0.0000*	1.6
D_Parks	-0.0001	-98.16	0.0000*	-8.76	-106.34	0.0000*	1.15
D_Market	-2E-05	-32.97	0.0000*	-1.72	-42.69	0.0000*	1.53
D_Institutes	1.1E-05	1.79	0.0741	1.03	2.47	0.0134*	1.21
D_Industries	-5E-05	-10.94	0.0000*	-1.25	-4.15	0.00003*	1.24
D_Hospitals	-8E-05	-37.37	0.0000*	-7.11	-50.95	0.0000*	1.31
D_Graveyards	-5E-05	-41.54	0.0000*	-1.77	-24.89	0.0000*	1.53

Table 8. Results of FastGWR Model estimates for Rating Area FD-I, FD-II and FD-III

Predictors	FD-I (R^2=0.61)		FD-II (R^2=0.59)		FD-III (R^2=0.80)	
	Coeff.	SE	Coeff.	SE	Coeff.	SE
Area	0.00511	0.0002	0.00524	0.00021	0.00475	0.00015
D_Worship Places	0.00072	0.00053	0.00181	0.00056	0.00049	0.00039
D_Solid Waste Facility	0.00516	0.00043	0.00025	0.00065	-0.001	0.00093
D_Parks	0.00044	0.00062	0.00515	0.00043	0.00078	0.00064
D_Market	0.00498	0.0005	0.00205	0.00057	0.00449	0.00066
D_Institutes	0.00047	0.00054	0.00086	0.00053	0.00038	0.00041
D_Industry	0.0022	0.00061	0.00109	0.00063	0.00038	0.00048
D_Hospitals	0.00165	0.0005	0.00442	0.00048	0.00006	0.00064
D_Graveyards	0.00539	0.00053	0.01	0.00049	0.00209	0.00079

These values indicate that in FD-I rating area, the distance to solid waste facilities does not influence the prices of residential properties while all other explanatory variables have an impact on the house prices positively or negatively. The results of FastGWR are presented in

Table **8**. The R-squared value for FD-I is 0.61 indicating that the model is able to explain 60% variance based on the explanatory variables. As suggested by the coefficients of all the predictive variables, there exists a positive correlation. This zone comprises of the central business district (CBD) where most of the properties are commercial and semi-commercial and there are only 28,090 residential properties.

The FD-II rating area is characterized by the small and medium scale industries, timber market, and wholesale businesses. Although the average house area in this zone is smaller but the average price per square meter is 22.6% more than the FD-I rating area and the average house price is also 20.32% more (i.e., US$102063). The higher prices of small houses in this area are due to the ease of access to workplaces and being close to the city center. The results of the valuation model for this region are dissimilar from FD-I rating area. All the explanatory variables are statistically significant except the distance form worship places. Although the average distance to the places of worships is much smaller (93.4 meters), worship places do not seem to influence the residential property prices in this particular zone. This effect is possibly due to the socio-economic conditions of the area, since the average house size and the income level is lower than the other residential districts. The worship places give the impression of being less important for the local residents by any means possibly due to the degree of adherence to religion and the level of noise from the loud speakers of mosques. Researchers have found the negative effects of places of worships on adjacent house prices while this effect declines with the increasing distance and diminishes after 300 meters (Brandt et al., 2014; Do et al., 1994).

Figure 6. Maps of Significant Parameter Estimates for the Predictive Variables, Area of House (A), Distance to Worship Places (B), Distance to Solid Waste Sites (C), and Distance to Parks (D)

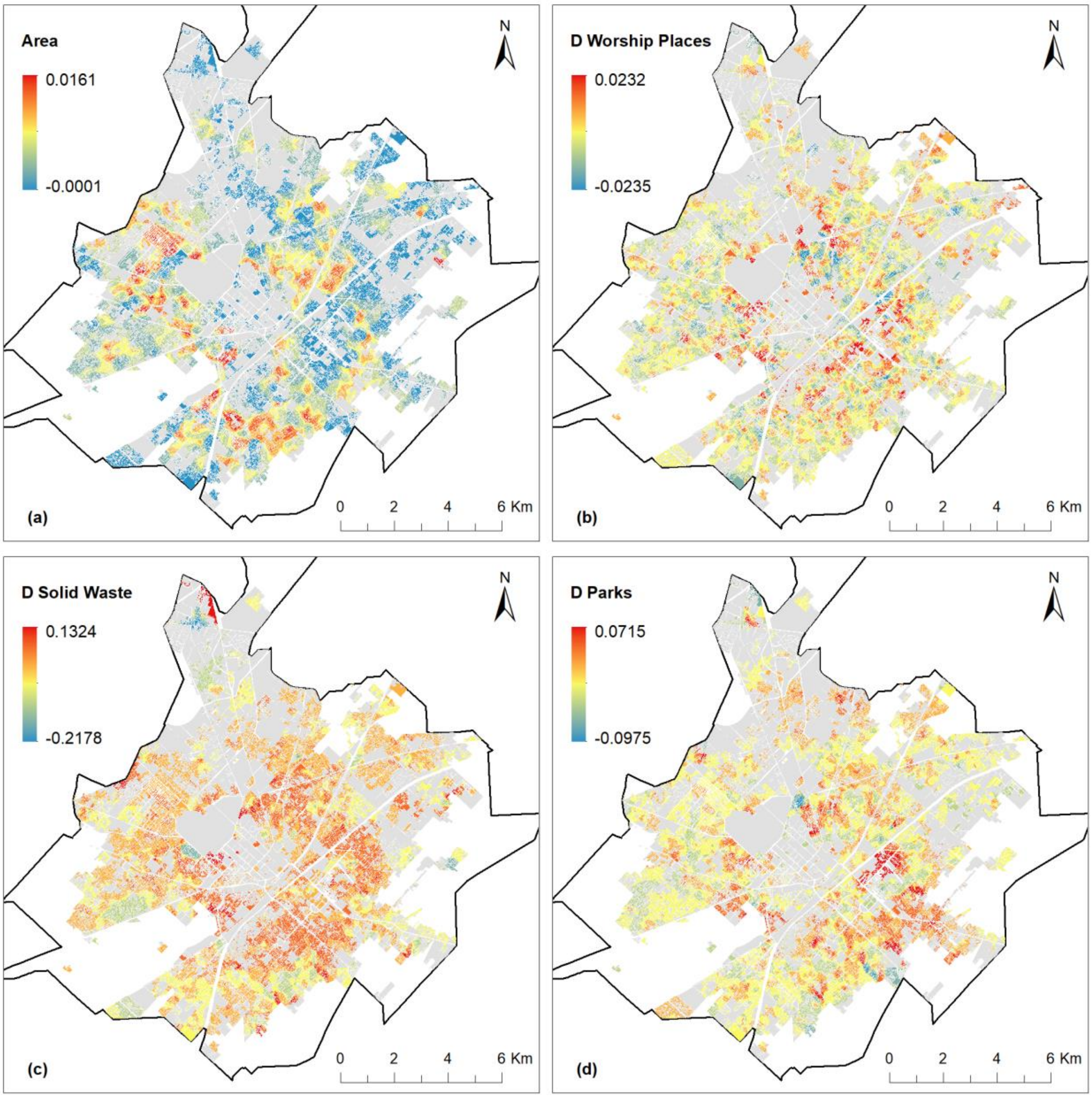
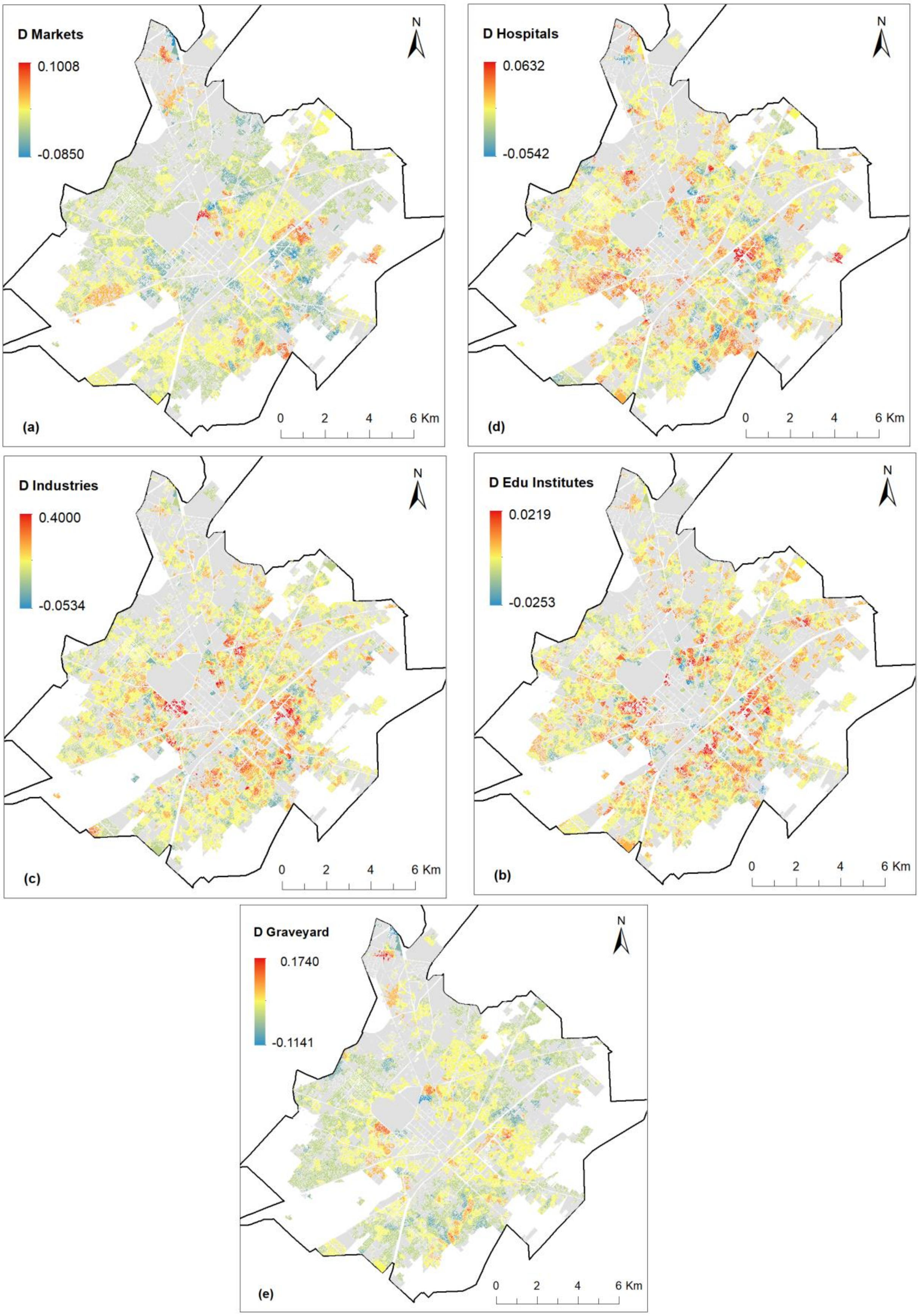


Figure 7. Maps of Significant Parameter Estimates for the Predictive Variables, Distance to Markets (A), Distance to Educational Institutions (B), Distance to Industries (C), Distance to Hospitals (D), Distance to Graveyards



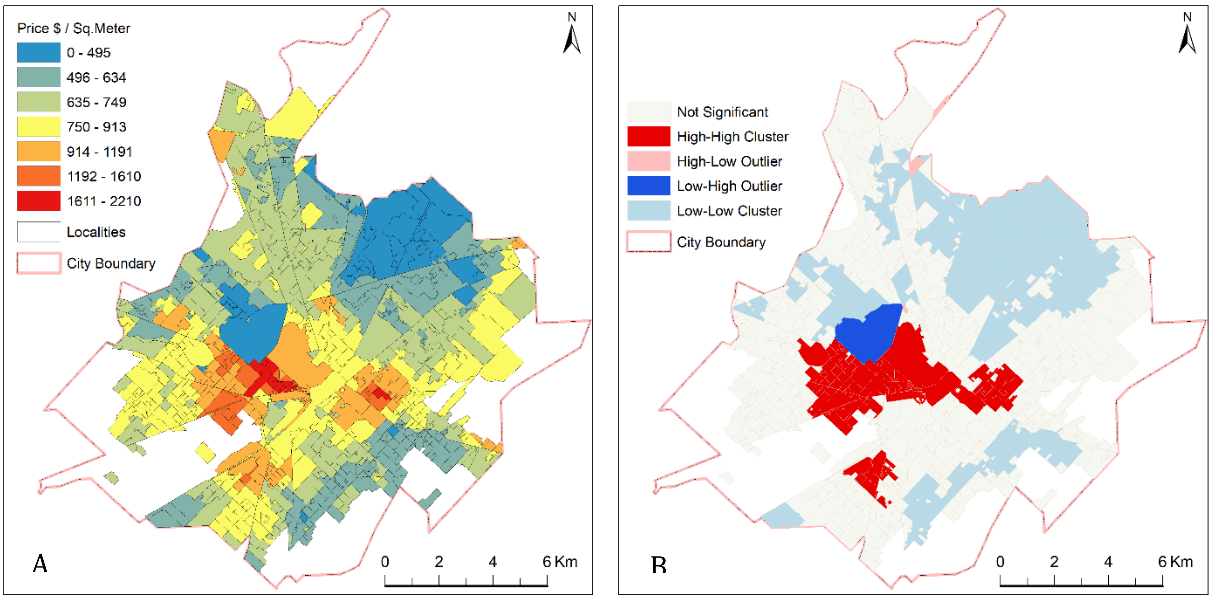
The coefficients for distance to solid waste facilities and distance to industries is negative and statistically significant (99% confidence). Solid waste transfer stations provide the opportunities for scavengers and scrap dealers to collect the recyclable materials to earn their livings. Small and medium scale industries like power looms, garment factories, embroidery units, plastic products, leather factories, paper and chemical factories exist in this area, thus, offering the employment opportunities to the local residents. Distance to market is statistically significant in the semi-log model but the results of linear model are not statistically significant. There are only 4.13 % (n = 11107) houses present in this zone whereas the 75.7% of the houses in FD-II are exempted from the property tax as per the policy of revenue department, and 24.27% houses are accountable for the property tax. Results of the semi-log model explain the relationship of house price and explanatory variables up to 78% while the linear model explains it up to 80%. In this zone the results of FastGWR are also positively correlated and the R squared value is 0.59, which shows an intermediate to higher performance by the model. Though the FastGWR model explains relatively smaller variance as compared to the linear models (59% vs 80%), the consideration of spatial aspects in FastGWR makes it more reliable when applied to analyze geographical disparities. One possible reason for this weak relationship is that there are only 11,107 residential building situated in this zone while most of the building are commercial and semi-commercial.

The third zone, FD-III rating area is the largest among all other zones in the city, which holds a total number of 339420 properties, while 229714 (85.42%) are residential. In this zone, only 22% of the residential properties are liable to pay the property tax while 78% houses are exempted from any kind of property tax. The results for FD-III rating area are significant for all the explanatory variables except the distance from educational institutions, which is insignificant under the semi-log model but significant under the linear model. Several studies have demonstrated the effects of schools and educational institutions on the property prices, (Sah et al., 2016; Wen, Xiao, et al., 2017; Yang et al., 2018a). However, in this zone the educational institutions do not impact the residential property prices. One explanation for this is the fact that the schools do not have the strict zone boundaries and the students can travel to longer distances in order to get admission based on the education quality of certain educational institutes. Distances to parks, hospitals, markets, graveyard, and industries have the negative coefficients. The results of FastGWR demonstrate that all the predictive variables have positive coefficients while the distance to solid waste facilities has a negative effect, suggesting that the residential parcels closer to these facilities have higher values while the residential properties away from solid waste facility have lower values.

4.3 Cluster and Outlier Analysis

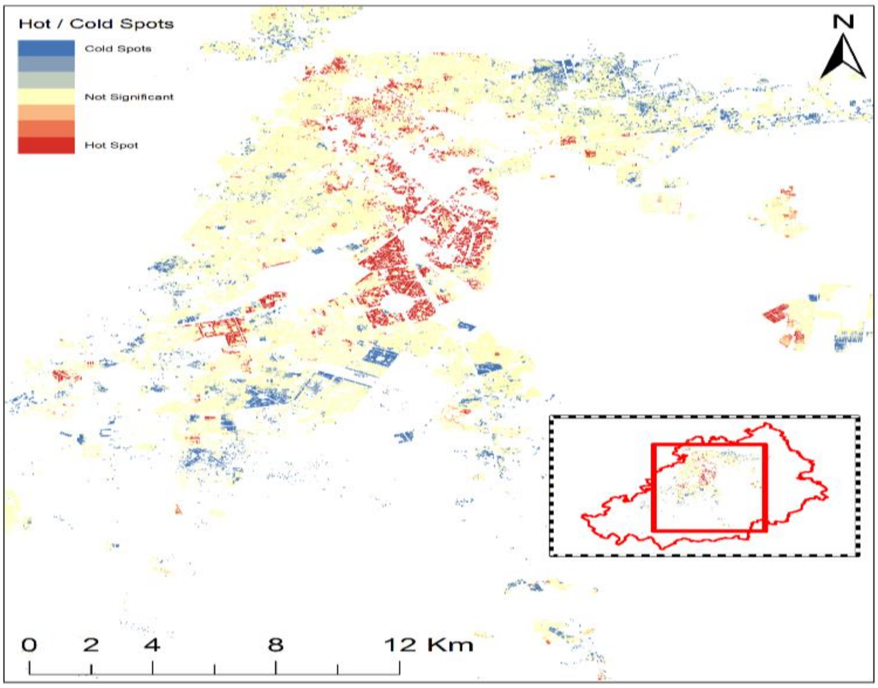
Figure 8 demonstrates the cluster and outlier analysis of property prices within the residential localities of Faisalabad city. The localities highlighted in red color are the high-high cluster, having the high residential property values. One major high values cluster is around the city center and the localities of *Clock Tower,* Jinnah colony, Madina Town, Peoples colony, Civil lines, Susan road, and Jaranwala road. Another high-high cluster is found in the localities of *Samanabad*. This high-high cluster is isolated from the main cluster of high residential property values. The localities colored in light blue color are the low-low clusters while the grey colored localities do not have any significant clustering. One high-low outlier, colored in pink, is also found in the area of Muslim town, which indicates that one particular locality has the higher property values and that locality is surrounded by the localities having the lower property values. The blue colored locality is the low-high outlier showing that this locality is surrounded by the localities with relatively higher property values. This particular locality is the University of Agriculture, Faisalabad, a public university having an area of ~345 hectares.

Figure 8. Average Price (US $) Per Square Meters in Different Localities of Faisalabad (A) Cluster and Outlier Analysis for the Property Values within the Residential Localities (B)



Spatial autocorrelation report of the Hotspot analysis showed a significant p-value statistically, so, the null hypothesis suggesting the randomness of regression residuals stands rejected and z-score being 767.76 points toward the significant clustering. Further, the Hotspot analysis output map figure-9 tells the story of the nature and distribution of high and low value clusters. Most of the high value clusters can be traced in, and around the Walled City area (the CBD), posh areas of Model Town, Gulberg, DHA and other elite-class residential colonies while the low value clusters were found in the peripheries.

Figure 9. Distribution of High and Low Value Clusters



4.4 Local Indicators of Spatial Association

The Local Indicators of Spatial Association (LISA) demonstrates that the high-priced residential properties are clustered together, and the low-priced houses are clustered together, as one might expect. **Error! Reference source not found.** shows the results of cluster and outlier analysis (local oran's I) in the study area (95% significance). In the entire city, 21% of the high value residential houses are clustered together while the cluster of low values account for 40% of the total residential properties. This is apparently a strong indication of the distribution of the real estate values across the city and the associated income class. The costly residential houses are located in the east and southeastern part of the city including the localities of Madina Town and Iqbal Town. The localities in Lyallpur town have a mixed type of values ranging from high-cost to low-cost residential houses since this town holds the older residential communities near the city center and newly built communities in the north of the city. Higher cost of living in this area also gives a rise to the informal settlements such as slums and squatters. The results regarding low-high outliers from LISA indicate that the low-cost houses are surrounded by the high cost residential properties while in some areas, we see that the high cost houses are surrounded by the low-cost houses, which is a typical sign of topophilic adherence to the place of local residents. For instance, if some of the residents become wealthier enough to buy a high cost property, they prefer to build the luxurious house in the same poor community instead of buying a new house in the posh areas of the city, thus making it an outlier among other houses.

Figure 10. Cluster Map of Residential Properties in Entire City (A), FD-I (B), FD-II (C), And FD-III (D)

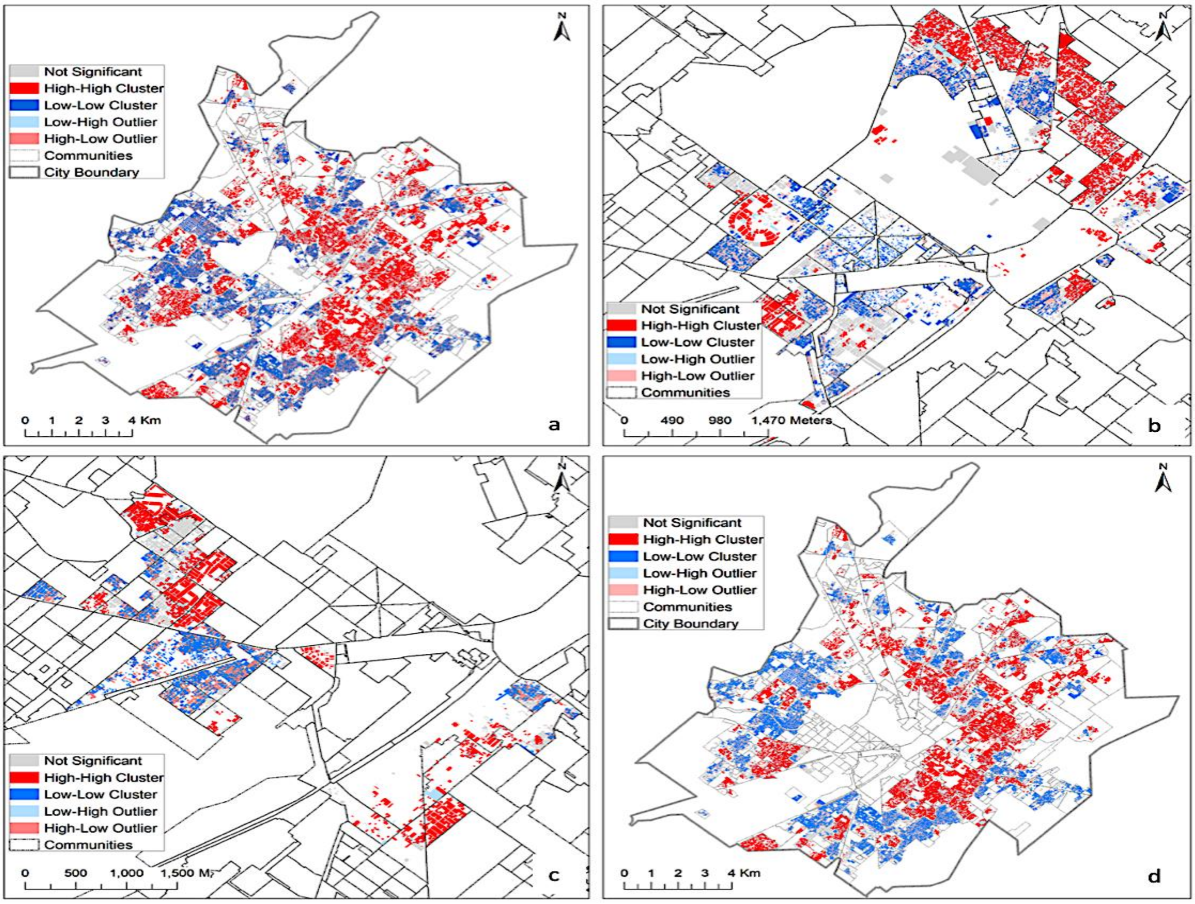


Figure 11. Percentage Distribution of Clustered Properties in the Study Area

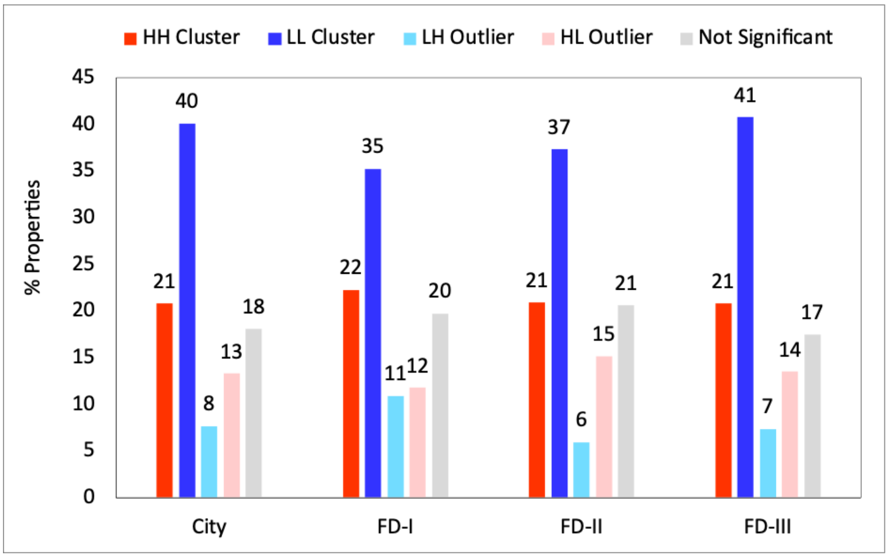


Figure 12. Estimated House Prices in Faisalabad City

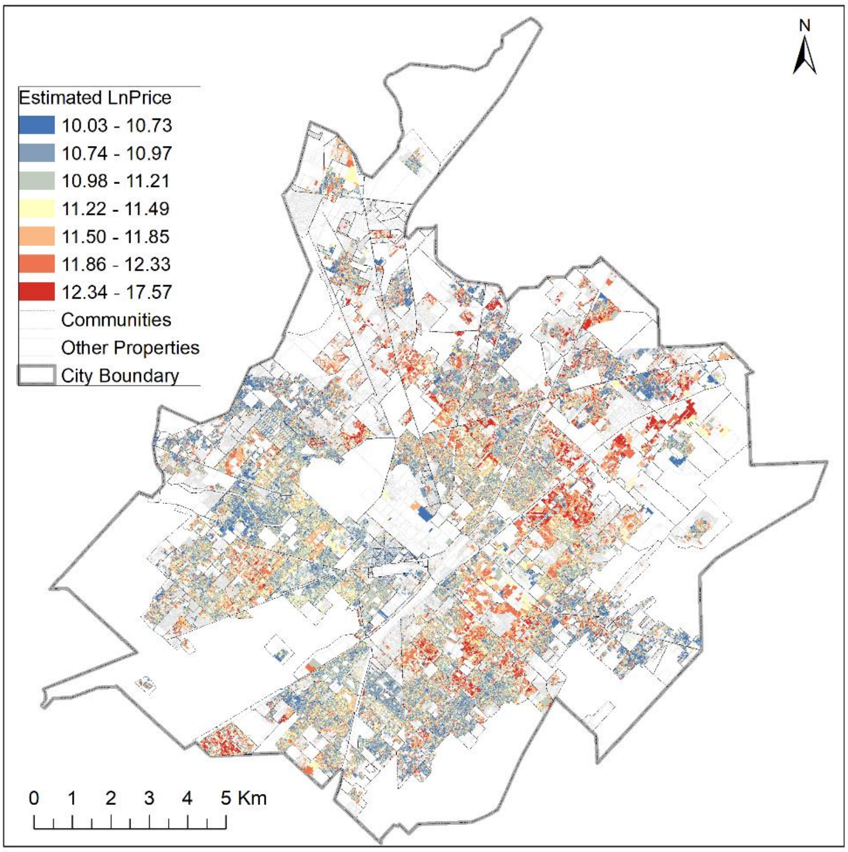


Figure 13. Residuals of the OLS Semi-Log Model for Entire City (A), FD-I (B) FD-II (C) And FD-III (D)

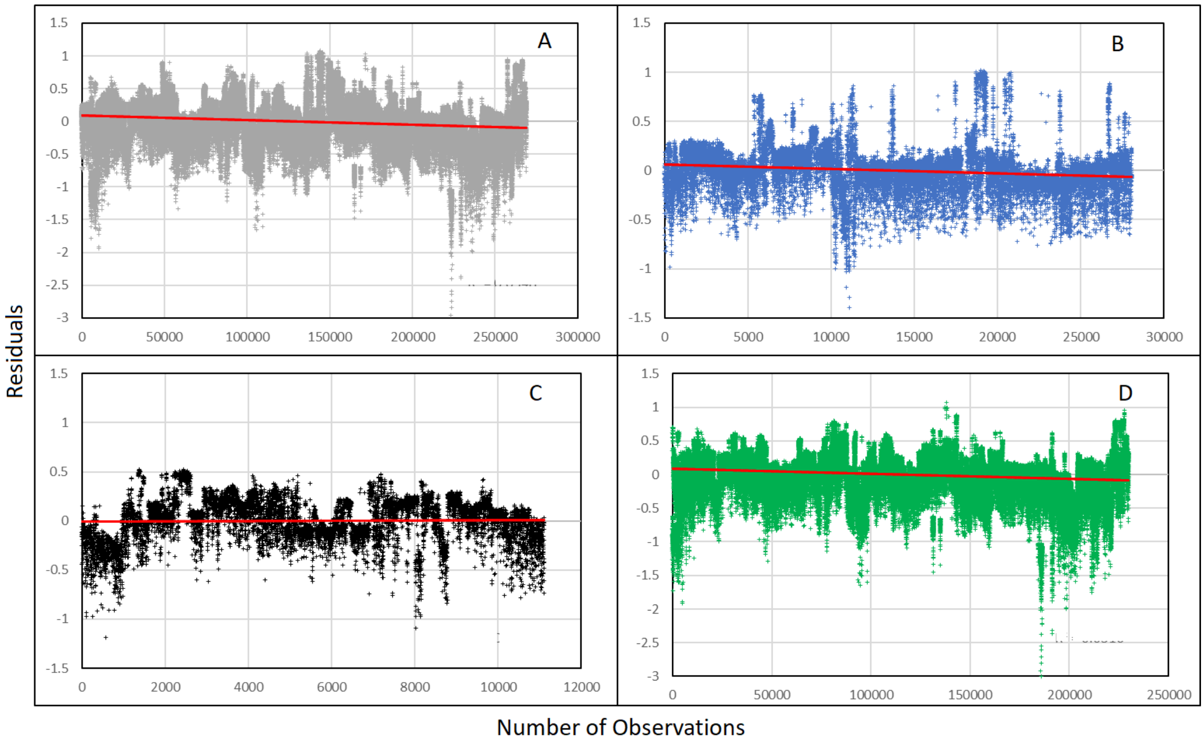
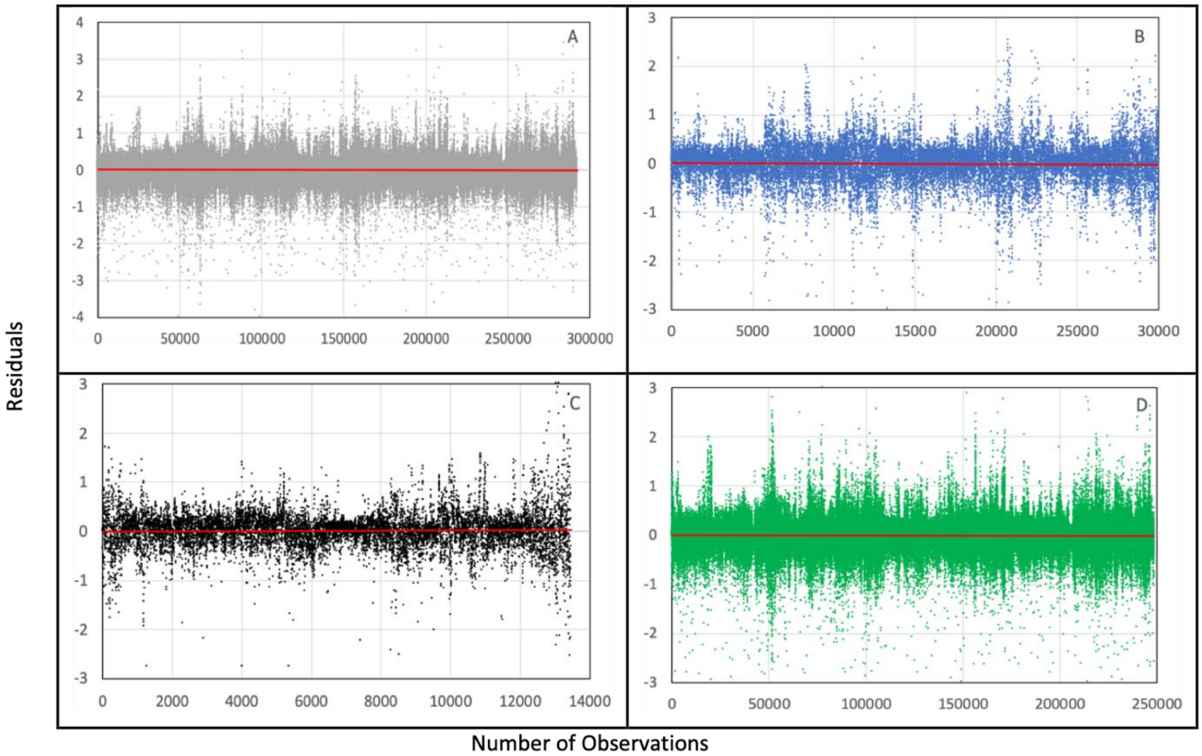


Figure 14. Residuals of the FastGWR Semi-log Model Entire City (a), FD-I (b), FD-II (c), and FD-III (d)



CONCLUSION

This paper examined the spatial determinants of prices for residential properties in the cities of Lahore and Faisalabad using a spatial hedonic approach. The spatial hedonic models, OLS and the FastGWR regression models, were used to analyze the association between several explanatory variables and the urban immoveable property prices. Nine locational features within four categories (i.e. amenities, cultural, educational and health facilities, and recreation) were selected to explore the correlations between the spatial determinants and the housing prices. While the performance of the two models vary slightly, the results showed the positive and negative statistically significant correlations between different locational features and the residential property prices. The principal contributing factors in the price was the area of house—strong positive coefficient for all the results. Other positively correlated variables were distance to worship places, distance to solid waste facility, and distance to educational institutions. The distance to public parks, markets, hospitals, graveyards, and distance to industries have a negative association with house prices. The ongoing development and the accelerated urban expansion have evolved the cities in Punjab from a single center to multiple centers pattern, and there is a need of reappraisal of properties in order to increase revenues from property taxes. Conclusively, the spatial determinants of housing valuation are significantly important, as presented in this study, and hence, must be integrated into policy formulation and future urban planning and design. There are several limitations in the study at this moment, for instance, it focuses on the locational features. However, there also exist socio-economic determinants that might influence the housing prices at different spatial scales. We could not include them due to two particular reasons, first, the housing level data on these socio-economic variables is not only a huge constraint in Pakistan, but also in many other developing countries. Therefore, we did not include them in the model. Second, the aim of this study was to particularly analyze the locational features and highlight their influence on housing prices, which has been neglected in the previous studies in Pakistan’s context.

RECOMMENDATIONS AND POLICY IMPLICATIONS

The results of this study have some general implications for policy-makers, investors, real estate developers and the urban planners. Our model produced the meaningful and reliable estimates which are appropriate to inform about the residential property values. The spatial diversity of the coefficients are much important for the decision-makers, requiring explicit knowledge of the local or regional housing markets. This helps to refine policies and have better understanding about the local house price variations. The approach applied is flexible and can be applied to different geographical locations in Pakistan. Since the record of past market transactions plays a key role in factual valuation of immoveable properties, we suggest the formulation of a system to record the fair market prices of real estate properties. For this purpose, the following steps may be taken; firstly, the property transfer fee may be waived off to attract the sellers and buyers towards disclosing the factual deal prices of a properties and the deficit created by this relinquishment may be bridged through the increase in annual property tax. Secondly, an immoveable property valuation desk may be established at Revenue Departments to serve the masses by assessing the market price of their properties at a nominal fee. This desk can record the geo-locations and structural attributes of properties, told by the assessment seekers as well as the assessed values and further, it may collect a reasonable earning for the public exchequer. A uniform and scientific method of immoveable property valuation as demonstrated in this study, may be adopted by the Federal Board of Revenue as well as provincial revenue departments considering, not only the structural attributes but also the spatial amenities, that must be updated in real time at regular intervals. The residential property data should be accessible to the researchers in order to explore the different aspects of real estate market which may help in building the suitable policies.

REFERENCES

Ahmed, R., Jawaid, S. T., & Khalil, S. (2021). Bubble Detection in Housing Market: Evidence From a Developing Country. SAGE Open, 11(2). https://doi.org/10.1177/21582440211006690

Anselin, L. (1995). Local indicators of spatial association—LISA. Geographical Analysis, 27(2), 93–115.

Anselin, L., & Le Gallo, J. (2006). Interpolation of Air Quality Measures in Hedonic House Price Models: Spatial Aspects. Spatial Economic Analysis, 1(1), 31–52. https://doi.org/10.1080/17421770600661337

Attarwala, F. S. (2020). Indicators pointing south for real estate prices. Dawn, The Business and Finance Weekly.

Baranzini, A., Schaerer, C., Ramirez, J., & Thalmann, P. (2008). Basics of the Hedonic Price Model. In A. Baranzini, C. Schaerer, J. Ramirez, & P. Thalmann (Eds.), Hedonic Methods in Housing Markets, Pricing Environmental Amenities and Segregation. Springer Science+Business Media, LLC, 233 Spring Street, New York,NY 10013, USA. https://doi.org/10.1007/978-0-387-76815-1

Baranzini, A., Schaerer, C., Ramirez, J., & Thalmann, P. (2010). Feel it or measure it-perceived vs. measured noise in hedonic models. Transportation Research Part D: Transport and Environment, 15D(8), 473–482.

Barreca, A., Curto, R., & Rolando, D. (2020). Urban vibrancy: An emerging factor that spatially influences the real estate market. Sustainability (Switzerland), 12(1), 346. https://doi.org/10.3390/su12010346

Bateman, I. J., Day, B., & Lake, I. (2001). The Effect of Road Traffic on Residential Property Values: A Literature Review and Hedonic Pricing Study. Bateman, Ian Day, Brett Lake, Iain Lovett, Andrew, January, 208.

Bitter, C., Mulligan, G. F., & Dall'erba, S. (2007). Incorporating spatial variation in housing attribute prices: A comparison of geographically weighted regression and the spatial expansion method. Journal of Geographical Systems, 9(1), 7–27. https://doi.org/10.1007/s10109-006-0028-7

Boza, E. (2015). Investigation of Housing Valuation Models based on spatial and non-spatial techniques. Middle East Technical University (METU), Ankara, Turkey.

Brandt, S., Maennig, W., & Richter, F. (2014). Do Houses of Worship Affect Housing Prices ? Evidence from Germany. 45(4), 549–570. https://doi.org/10.1111/grow.12066

Brasington, D. M., & Hite, D. (2005). Demand for environmental quality : a spatial hedonic analysis. 35, 57–82. https://doi.org/10.1016/j.regsciurbeco.2003.09.001

Brown, G., Reed, P., & Raymond, C. M. (2020). Mapping place values: 10 lessons from two decades of public participation GIS empirical research. In Applied Geography (Vol. 116, p. 102156). Elsevier Ltd. https://doi.org/10.1016/j.apgeog.2020.102156

Brunsdon, C., Fotheringham, A. S., & Charlton, M. E. (1996). Geographically weighted regression: a method for exploring spatial nonstationarity. Geographical Analysis, 28(4), 281–298. https://doi.org/10.1111/j.1538-4632.1996.tb00936.x

Brunsdont, C., Fotheringham, S., & Charlton, M. (1998). Geographically weighted regression - Modelling spatial non-stationarity. Journal of the Royal Statistical Society Series D: The Statistician, 47(3), 431–443. https://doi.org/10.1111/1467-9884.00145

Can, A. (1990). The measurement of neighborhood dynamics in urban house prices. Economic Geography, 66(3), 254–272. https://doi.org/10.2307/143400

Can, A., & Megbolugbe, I. (1997). Spatial Dependence and House Price Index Construction. Journal of Real Estate Finance and Economics, 14(1–2), 203–222. https://doi.org/10.1023/A:1007744706720

Casetti, E. (1972). Generating Models by the Expansion Method: Applications to Geographical Research. Geographical Analysis, 4(1), 81–91. https://doi.org/10.1111/j.1538-4632.1972.tb00458.x

Casetti, E. (1997). The expansion method, mathematical modeling, and spatial econometrics. International Regional Science Review, 20(1–2), 9–33. https://doi.org/10.1177/016001769702000102

Chapman, J. I., Johnston, R. J., & Tyrrell, T. J. (2009). Implications of a land value tax with error in assesed values. Land Economics, 85(4), 576–586. https://doi.org/10.3368/le.85.4.576

Crawford, T. W. (2009). Scale Analytical. In International Encyclopedia of Human Geography (pp. 29–36). Elsevier. https://doi.org/10.1016/B978-008044910-4.00399-0

Crompton, J. L., & Nicholls, S. (2020). Impact on property values of distance to parks and open spaces: An update of U.S. studies in the new millennium. Journal of Leisure Research, 51(2), 127–146. https://doi.org/10.1080/00222216.2019.1637704

Dalton, C. M., & Thatcher, J. (2015). Inflated granularity: Spatial “Big Data” and geodemographics. Big Data and Society, 2(2). https://doi.org/10.1177/2053951715601144

De, U. K., & Vupru, V. (2017). Location and neighbourhood conditions for housing choice and its rental value: Empirical examination in an urban area of North-East India. International Journal of Housing Markets and Analysis, 10(4), 519–538. https://doi.org/10.1108/IJHMA-10-2016-0072

Dismuke, C. E., & Lindrooth, R. (2006). Ordinary least squares. In E. Chumney & K. N. Simpson (Eds.), Methods and Designs for Outcomes Research (pp. 93–104). American Society of Health-System Pharmacists.

Do, A. Q., Wilbur, R. W., & Short, J. L. (1994). An empirical examination of the externalities of neighborhood churches on housing values. The Journal of Real Estate Finance and Economics, 9(2), 127–136.

Dowall, D. E., & Ellis, P. D. (2009). Urban Land and Housing Markets in the Punjab , Pakistan. Urban Studies, 46(11), 2277–2300. https://doi.org/10.1177/0042098009342599

Dutilleul, P., & Legendre, P. (1993). Spatial Heterogeneity against Heteroscedasticity: An Ecological Paradigm versus a Statistical Concept. Oikos, 66(1), 152. https://doi.org/10.2307/3545210

Erickson, D. L., Lovell, S. T., & Méndez, V. E. (2011). Landowner willingness to embed production agriculture and other land use options in residential areas of Chittenden County, VT. Landscape and Urban Planning, 103(2), 174–184. https://doi.org/10.1016/j.landurbplan.2011.07.009

FBR. (2020). Circular; C.No. 1(121)R&S/2017-15297-R, Constitution of Committees for Updation of Valuation Tables of Immoveable Property under Sub-Section (4) of Section 68 of the Income Tax Ordinance, 2001 (p. 1). Revenue Division, Federal Board Of Revenue, Government of Pakistan.

Feuillet, T., Commenges, H., Menai, M., Salze, P., Perchoux, C., Reuillon, R., Kesse-Guyot, E., Enaux, C., Nazare, J. A., Hercberg, S., Simon, C., Charreire, H., & Oppert, J. M. (2018). A massive geographically weighted regression model of walking-environment relationships. Journal of Transport Geography, 68, 118–129. https://doi.org/10.1016/j.jtrangeo.2018.03.002

Fik, T. J., Ling, D. C., & Mulligan, G. F. (2003). Modeling Spatial Variation in Housing Prices: A Variable Interaction Approach. In Real Estate Economics (Vol. 31, Issue 4, pp. 623–646). John Wiley & Sons, Ltd. https://doi.org/10.1046/j.1080-8620.2003.00079.x

Fisher, B., Naidoo, R., & Ricketts, T. (2015). A Field Guide to Economics for Conservationists. Macmillan Learning.

Fotheringham, A., Brunsdon, C., & Charlton, M. (2002). Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. In John Wiley & Sons (Vol. 13).

Fotheringham, A. S. (1997). Trends in quantitative methods I: Stressing the local. Progress in Human Geography, 21(1), 88–96. https://doi.org/10.1191/030913297676693207

Fotheringham, A. S., Crespo, R., & Yao, J. (2015). Geographical and Temporal Weighted Regression (GTWR). Geographical Analysis, 47(4), 431–452. https://doi.org/10.1111/gean.12071

Fotheringham, A. S., Yang, W., & Kang, W. (2017). Multiscale Geographically Weighted Regression (MGWR). Annals of the American Association of Geographers, 107(6), 1247–1265. https://doi.org/10.1080/24694452.2017.1352480

Freeman, A. M. (1981). Hedonic Prices, Property Values and Measuring Environmental Benefits: A Survey of the Issues. In Measurement in Public Choice (pp. 13–32). Palgrave Macmillan UK. https://doi.org/10.1007/978-1-349-05090-1_2

Frost, J. (2019). Introduction to Statistics: An intuitive guide (First Edit). Statistics by Jim publishing: State College, PA, USA, 2019.

Gabrielli, L., & French, N. (2020). Pricing to market: property valuation methods – a practical review. Journal of Property Investment and Finance. https://doi.org/10.1108/JPIF-09-2020-0101

Gaffney, M. (2009). The role of land markets in economic crises. American Journal of Economics and Sociology, 68(4), 855–888. https://doi.org/10.1111/j.1536-7150.2009.00657.x

Gao, S., Li, L., Li, W., Janowicz, K., & Zhang, Y. (2017). Constructing gazetteers from volunteered Big Geo-Data based on Hadoop. Computers, Environment and Urban Systems, 61, 172–186. https://doi.org/10.1016/j.compenvurbsys.2014.02.004

Getis, A. (2008). A history of the concept of spatial autocorrelation: A geographer's perspective. Geographical Analysis, 40(3), 297–309. https://doi.org/10.1111/j.1538-4632.2008.00727.x

Getis, A., & Ord, J. K. (2010a). The analysis of spatial association by use of distance statistics. In Advances in Spatial Science (Vol. 61, pp. 127–145). Springer International Publishing. https://doi.org/10.1007/978-3-642-01976-0_10

Getis, A., & Ord, J. K. (2010b). The analysis of spatial association by use of distance statistics. In Perspectives on spatial data analysis (pp. 127–145). Springer.

Gilderbloom, J. I., Riggs, W. W., & Meares, W. L. (2015). Does walkability matter? An examination of walkability's impact on housing values, foreclosures and crime. Cities, 42(PA), 13–24. https://doi.org/10.1016/j.cities.2014.08.001

Gluszak, M., & Zygmunt, R. (2018). Development density, administrative decisions, and land values: An empirical investigation. Land Use Policy, 70, 153–161. https://doi.org/10.1016/j.landusepol.2017.10.036

Goix, R. Le, Giraud, T., Cura, R., Le Corre, T., & Migozzi, J. (2019). Who sells to whom in the suburbs? Home price inflation and the dynamics of sellers and buyers in the metropolitan region of Paris, 1996–2012. PLoS ONE, 14(3), e0213169. https://doi.org/10.1371/journal.pone.0213169

Gollini, I., Lu, B., Charlton, M., Brunsdon, C., & Harris, P. (2015). Gwmodel: An R package for exploring spatial heterogeneity using geographically weighted models. Journal of Statistical Software, 63(17), 1–50. https://doi.org/10.18637/jss.v063.i17

Goodchild, M. F. (2013). The quality of big (geo)data. Dialogues in Human Geography, 3(3), 280–284. https://doi.org/10.1177/2043820613513392

Goodman, A. C. (1978). Hedonic prices, price indices and housing markets. Journal of Urban Economics, 5(4), 471–484. https://doi.org/10.1016/0094-1190(78)90004-9

Griffith, D. A. (2004). Spatial Autocorrelation. In Encyclopedia of Social Measurement (pp. 581–590). Elsevier. https://doi.org/10.1016/B0-12-369398-5/00334-0

Grislain-Letrémy, C., & Katossky, A. (2014). The impact of hazardous industrial facilities on housing prices: A comparison of parametric and semiparametric hedonic price models. Regional Science and Urban Economics, 49, 93–107. https://doi.org/10.1016/j.regsciurbeco.2014.09.002

Gul, A., Nawaz, M., Basheer, M. A., Tariq, F., & Raheel Shah, S. A. (2018). Built houses as a tool to control residential land speculation - A case study of Bahria Town, Lahore. Habitat International, 71, 81–87. https://doi.org/10.1016/j.habitatint.2017.11.007

Guo, H., Wang, L., Chen, F., & Liang, D. (2014). Scientific big data and Digital Earth. Chinese Science Bulletin, 59(35), 5066–5073. https://doi.org/10.1007/s11434-014-0645-3

Haque, N. U. (2015). Flawed urban development policies in Pakistan. PIDE Working Papers, 1(119), 1–20.

Harris, R., Singleton, A., Grose, D., Brunsdon, C., & Longley, P. (2010). Grid-enabling geographically weighted regression: A case study of participation in higher education in England. Transactions in GIS, 14(1), 43–61. https://doi.org/10.1111/j.1467-9671.2009.01181.x

Hassan, M. M., Ahmad, N., & Hashim, A. H. (2021). Homebuyers Superstitious Belief: Feng Shui and Housing Property. International Journal of Academic Research in Business and Social Sciences, 11(7). https://doi.org/10.6007/ijarbss/v11-i7/10297

Helbich, M., Jochem, A., Mücke, W., & Höfle, B. (2013). Boosting the predictive accuracy of urban hedonic house price models through airborne laser scanning. Computers, Environment and Urban Systems, 39, 81–92. https://doi.org/10.1016/j.compenvurbsys.2013.01.001

Hu, S., Cheng, Q., Wang, L., & Xu, D. (2013). Modeling land price distribution using multifractal IDW interpolation and fractal filtering method. Landscape and Urban Planning, 110, 25–35. https://doi.org/10.1016/j.landurbplan.2012.09.008

Hu, S., Yang, S., Li, W., Zhang, C., & Xu, F. (2016). Spatially non-stationary relationships between urban residential land price and impact factors in Wuhan city, China. Applied Geography, 68(1111), 48–56. https://doi.org/10.1016/j.apgeog.2016.01.006

Hubert, L. J., Golledge, R. G., & Costanzo, C. M. (1981). Generalized Procedures for Evaluating Spatial Autocorrelation. Geographical Analysis, 13(3), 224–233. https://doi.org/10.1111/j.1538-4632.1981.tb00731.x

Ibeas, ángel, Cordera, R., Dell'Olio, L., Coppola, P., & Dominguez, A. (2012). Modelling transport and real-estate values interactions in urban systems. Journal of Transport Geography, 24, 370–382. https://doi.org/10.1016/j.jtrangeo.2012.04.012

Jiang, B. (2018). Spatial heterogeneity, scale, data character and sustainable transport in the big data era. In ISPRS International Journal of Geo-Information (Vol. 7, Issue 5, p. 167). Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/ijgi7050167

Katsalap, V. (2008). Spatial dependency and heterogeneity in housing prices across Ukrainian cities. National University "Kyiv-Mohyla Academy."

Kauko, T. (2003). On current nueral network applications involving spatial modelling of property prices. Journal of Housing and the Built Environment, 18(2), 159–181. https://doi.org/10.1023/A:1023977111302

Kim, C. W., Phipps, T. T., & Anselin, L. (2003). Measuring the benefits of air quality improvement: A spatial hedonic approach. Journal of Environmental Economics and Management, 45(1), 24–39. https://doi.org/10.1016/S0095-0696(02)00013-X

Knight, J. R. (2008). Hedonic Modeling of the Home Selling Process. In P. T. Andrea Baranzini, José Ramirez, Caroline Schaerer (Ed.), Hedonic Methods in Housing Markets; Pricing Environmental Amenities and Segregation (pp. 39–54). Springer Science+Business Media, LLC, 233 Spring Street, New York, NY 10013, USA. https://doi.org/10.1007/978-0-387-76815-1_2

Koschinsky, J., Lozano-Gracia, N., & Piras, G. (2012). The welfare benefit of a home's location: An empirical comparison of spatial and non-spatial model estimates. Journal of Geographical Systems, 14(3), 319–356. https://doi.org/10.1007/s10109-011-0148-6

L. Krause, A., & Bitter, C. (2012). Spatial econometrics, land values and sustainability: Trends in real estate valuation research. Cities, 29(SUPPL.2). https://doi.org/10.1016/j.cities.2012.06.006

Lancaster, K. J. (1966). A New Approach to Consumer Theory. Journal of Political Economy, 74(2), 132–157. https://doi.org/10.1086/259131

Larson, W., & Shui, J. (2020). Land Valuation using Public Records and Kriging: Implications for Land versus Property Taxation in Cities.

LeSage, J. P. (2008). An introduction to spatial econometrics. In Revue d'Economie Industrielle (Vol. 123, Issue 3, pp. 19–44). https://doi.org/10.4000/rei.3887

Li, L., & Revesz, P. (2004). Interpolation methods for spatio-temporal geographic data. Computers, Environment and Urban Systems, 28(3), 201–227.

Li, S., Ye, X., Lee, J., Gong, J., & Qin, C. (2017). Spatiotemporal Analysis of Housing Prices in China: A Big Data Perspective. Applied Spatial Analysis and Policy, 10(3), 421–433. https://doi.org/10.1007/s12061-016-9185-3

Li, Z., & Fotheringham, A. S. (2020). Computational improvements to multi-scale geographically weighted regression. International Journal of Geographical Information Science, 34(7), 1378–1397. https://doi.org/10.1080/13658816.2020.1720692

Li, Z., Fotheringham, A. S., Li, W., & Oshan, T. (2019a). Fast Geographically Weighted Regression (FastGWR): a scalable algorithm to investigate spatial process heterogeneity in millions of observations. International Journal of Geographical Information Science, 33(1), 155–175. https://doi.org/10.1080/13658816.2018.1521523

Li, Z., Fotheringham, A. S., Li, W., & Oshan, T. (2019b). Fast Geographically Weighted Regression (FastGWR): a scalable algorithm to investigate spatial process heterogeneity in millions of observations. International Journal of Geographical Information Science, 33(1), 155–175. https://doi.org/10.1080/13658816.2018.1521523

Liang, X., Liu, Y., Qiu, T., Jing, Y., & Fang, F. (2018). The effects of locational factors on the housing prices of residential communities: The case of Ningbo, China. Habitat International, 81, 1–11. https://doi.org/10.1016/j.habitatint.2018.09.004

Liao, W. C., & Wang, X. (2012). Hedonic house prices and spatial quantile regression. Journal of Housing Economics, 21(1), 16–27. https://doi.org/10.1016/j.jhe.2011.11.001

Lovett, A. A. (2019). Economic Valuation of Services (C. von H. et al. (eds.) (ed.); Landscape, pp. 315–326). Springer, Dordrecht. https://doi.org/10.1007/978-94-024-1681-7_20

Lu, B., Charlton, M., Harris, P., & Fotheringham, A. S. (2014). Geographically weighted regression with a non-Euclidean distance metric: A case study using hedonic house price data. International Journal of Geographical Information Science, 28(4), 660–681. https://doi.org/10.1080/13658816.2013.865739

Lumley, T., Diehr, P., Emerson, S., & Chen, L. (2002). The importance of the normality assumption in large public health data sets. In Annual Review of Public Health (Vol. 23, pp. 151–169). https://doi.org/10.1146/annurev.publhealth.23.100901.140546

Ma, J., Cheng, J. C. P., Jiang, F., Chen, W., & Zhang, J. (2020). Analyzing driving factors of land values in urban scale based on big data and non-linear machine learning techniques. Land Use Policy, 94, 104537. https://doi.org/10.1016/j.landusepol.2020.104537

Machin, S. (2011). Houses and schools: Valuation of school quality through the housing market. Labour Economics, 18(6), 723–729. https://doi.org/10.1016/j.labeco.2011.05.005

Malaitham, S., Fukuda, A., Vichiensan, V., & Wasuntarasook, V. (2020). Hedonic pricing model of assessed and market land values: A case study in Bangkok metropolitan area, Thailand. In Case Studies on Transport Policy (Vol. 8, Issue 1, pp. 153–162). Elsevier. https://doi.org/10.1016/j.cstp.2018.09.008

Malik, S., Roosli, R., & Tariq, F. (2020). Investigation of informal housing challenges and issues: experiences from slum and squatter of Lahore. Journal of Housing and the Built Environment, 35(1), 143–170. https://doi.org/10.1007/s10901-019-09669-9

Mankad, M. D. (2021). Comparing OLS based hedonic model and ANN in house price estimation using relative location. Spatial Information Research, 1–10. https://doi.org/10.1007/s41324-021-00416-3

Mitchel, A. (2005). The ESRI Guide to GIS analysis, Volume 2: Spartial measurements and statistics. ESRI Guide to GIS Analysis.

Mitchell, A. (2005). The ESRI guide to GIS analysis. Volume 2, Spatial measurements and statistics (First Edit). Esri Press.

Munshi, T. (2020). Accessibility, infrastructure provision and residential land value: Modelling the relation using geographic weighted regression in the city of Rajkot, India. Sustainability (Switzerland), 12(20), 1–16. https://doi.org/10.3390/su12208615

Ord, J. K., & Getis, A. (1995). Local Spatial Autocorrelation Statistics: Distributional Issues and an Application. Geographical Analysis, 27(4), 286–306. https://doi.org/10.1111/j.1538-4632.1995.tb00912.x

Ord, J. K., & Getis, A. (2001). Testing for local spatial autocorrelation in the presence of global autocorrelation. Journal of Regional Science, 41(3), 411–432. https://doi.org/10.1111/0022-4146.00224

Oshan, T. M., Li, Z., Kang, W., Wolf, L. J., & Stewart Fotheringham, A. (2019). MGWR: A python implementation of multiscale geographically weighted regression for investigating process spatial heterogeneity and scale. ISPRS International Journal of Geo-Information, 8(6). https://doi.org/10.3390/ijgi8060269

Ouattara, K., Kim, Y., Raza, H., Hadi, Q. ul A., Rostom, A. M. T., Milyutin, A., Rashid, A. A., Khan, B. N., Haq, E., Grigoryeva, E., Funahashi, J., Chiquier, L., Zaheer, N., Dione, N. T., Hassler, O., Shaikh, S. A., Ahmad, S., Athar, S., & Conde, V. M. O. (2018). PAKISTAN HOUSING FINANCE PROJECT (P162095)Report No: PAD2385.

Pace, R. K., & Gilley, O. W. (1998). Generalizing the OLS and grid estimators. Real Estate Economics, 26(2), 331–347. https://doi.org/10.1111/1540-6229.00748

Pagourtzi, E., Assimakopoulos, V., Hatzichristos, T., & French, N. (2003). Real estate appraisal: A review of valuation methods. In Journal of Property Investment & Finance (Vol. 21, Issue 4, pp. 383–401). MCB UP Ltd. https://doi.org/10.1108/14635780310483656

Pakistan Bureau of Statistics. (2017). Area & Population of Aadministrative Units. In Population Census.

Pakistan, S. B. of. (2019). State Bank of Pakistan Infrastructure, Housing & SME Finance Policy for Low Cost Housing Finance State Bank of Pakistan Infrastructure, Housing and SME Finance Department (Issue March).

Peeters, A., Zude, M., Käthner, J., Ünlü, M., Kanber, R., Hetzroni, A., Gebbers, R., & Ben-Gal, A. (2015). Getis-Ord's hot- and cold-spot statistics as a basis for multivariate spatial clustering of orchard tree data. Computers and Electronics in Agriculture, 111, 140–150. https://doi.org/10.1016/j.compag.2014.12.011

Rizvi, Z. M. (2018). National Affordable Housing Policy (pp. 1–22).

Rosen, S. (n.d.). Hedonic Prices and Implicit Markets : Product Differentiation in Pure Competition. 34–55.

Sabir, M., Torre, A., & Magsi, H. (2017). Land-use conflict and socio-economic impacts of infrastructure projects: the case of Diamer Bhasha Dam in Pakistan. Area Development and Policy, 2(1), 40–54. https://doi.org/10.1080/23792949.2016.1271723

Sah, V., Conroy, S. J., & Narwold, A. (2016). Estimating School Proximity Effects on Housing Prices: the Importance of Robust Spatial Controls in Hedonic Estimations. Journal of Real Estate Finance and Economics, 53(1), 50–76. https://doi.org/10.1007/s11146-015-9520-5

Salon, D., Wu, J., & Shewmake, S. (2014). Impact of bus rapid transit and metro rail on property values in Guangzhou, China. In Transportation Research Record (Vol. 2452, pp. 36–45). National Research Council. https://doi.org/10.3141/2452-05

Scott, L. M., & Janikas, M. V. (2010). Spatial Statistics in ArcGIS. In M. M. F. A. Getis (Ed.), Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications (pp. 27–41). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-03647-7_2

Seo, K., Golub, A., & Kuby, M. (2014). Combined impacts of highways and light rail transit on residential property values: A spatial hedonic price model for Phoenix, Arizona. Journal of Transport Geography, 41. https://doi.org/10.1016/j.jtrangeo.2014.08.003

Shabana, Ali, G., Bashir, M. K., & Ali, H. (2015). Housing valuation of different towns using the hedonic model: A case of Faisalabad city, Pakistan. Habitat International, 50, 240–249. https://doi.org/10.1016/j.habitatint.2015.08.036

Singh, A., Sharma, A., & Dubey, G. (2020). Big data analytics predicting real estate prices. International Journal of Systems Assurance Engineering and Management, 11(2), 208–219. https://doi.org/10.1007/s13198-020-00946-3

Stewart Fotheringham, A., Charlton, M., & Brunsdon, C. (1996). The geography of parameter space: an investigation of spatial non-stationarity. International Journal of Geographical Information Systems, 10(5), 605–627.

Taylor, L. O. (2008). Theoretical Foundations and Empirical Developments in Hedonic Modeling. In P. T. Andrea Baranzini, José Ramirez, Caroline Schaerer (Ed.), Hedonic Methods in Housing Markets; Pricing Environmental Amenities and Segregation (pp. 15–37). Springer Science + Business Media, LLC, 233 Spring Street, New York, NY 10013, USA. https://doi.org/10.1007/978-0-387-76815-1_1

Thomas, D. M., & Mathur, S. (2019). Data Analysis by Web Scraping using Python. Proceedings of the 3rd International Conference on Electronics and Communication and Aerospace Technology, ICECA 2019, 450–454. https://doi.org/10.1109/ICECA.2019.8822022

Thompson, E., Butters, R. B., & Schmitz, B. T. (2012). The property value premium of a place of worship. Contemporary Economic Policy, 30(2), 215–222. https://doi.org/10.1111/j.1465-7287.2011.00255.x

Tian, G., Wei, Y. D., & Li, H. (2017). Effects of accessibility and environmental health risk on housing prices: a case of Salt Lake County, Utah. Applied Geography, 89, 12–21. https://doi.org/10.1016/j.apgeog.2017.09.010

Wahid, A., Mantell, E. H., & Mumtaz, M. Z. (2021). Under invoicing in the residential real estate market in Pakistan. International Journal of Strategic Property Management, 25(3), 190–203.

Wajahat, F. (2012). Perceptions of tenure security in a squatter settlement in Lahore, Pakistan. In Transforming Asian Cities: Intellectual Impasse, Asianizing Space, and Emerging Translocalities (pp. 137–147). https://doi.org/10.4324/9780203093894

Wang, W., van Noorloos, F., & Spit, T. (2020). Stakeholder power relations in Land Value Capture: comparing public (China) and private (U.S.) dominant regimes. Land Use Policy, 91, 104357. https://doi.org/10.1016/j.landusepol.2019.104357

Wani, S., Shaikh, H., & Harman, O. (2020). Urban property taxes in Pakistan's Punjab. International Growth Centre (IGC).

Wen, H., Jin, Y., & Zhang, L. (2017). Spatial heterogeneity in implicit housing prices: evidence from Hangzhou, China. International Journal of Strategic Property Management, 21(1), 15–28. https://doi.org/10.3846/1648715X.2016.1247021

Wen, H., Xiao, Y., & Zhang, L. (2017). School district, education quality, and housing price: Evidence from a natural experiment in Hangzhou, China. Cities, 66, 72–80. https://doi.org/10.1016/j.cities.2017.03.008

Wooldridge, J. M. (2016). Introductory econometrics: A modern approach (Fifth Edit). Nelson Education, Ltd.

Wu, C., Ye, X., Ren, F., Wan, Y., Ning, P., & Du, Q. (2016). Spatial and social media data analytics of housing prices in Shenzhen, China. PLoS ONE, 11(10), e0164553. https://doi.org/10.1371/journal.pone.0164553

Xiao, Y., Chen, X., Li, Q., Yu, X., Chen, J., & Guo, J. (2017). Exploring Determinants of Housing Prices in Beijing: An Enhanced Hedonic Regression with Open Access POI Data. ISPRS International Journal of Geo-Information, 6(11), 358. https://doi.org/10.3390/ijgi6110358

Yang, L., Chau, K. W., Szeto, W. Y., Cui, X., & Wang, X. (2020). Accessibility to transit, by transit, and property prices: Spatially varying relationships. Transportation Research Part D: Transport and Environment, 85, 102387. https://doi.org/10.1016/j.trd.2020.102387

Yang, L., Chu, X., Gou, Z., Yang, H., Lu, Y., & Huang, W. (2020). Accessibility and proximity effects of bus rapid transit on housing prices: Heterogeneity across price quantiles and space. Journal of Transport Geography, 88, 102850. https://doi.org/10.1016/j.jtrangeo.2020.102850

Yang, L., Wang, B., Zhang, Y., Ye, Z., Wang, Y., & Li, P. (2018a). Willing to pay more for high-quality schools? International Review for Spatial Planning and Sustainable Development, 6(1), 45–62. https://doi.org/10.14246/irspsd.6.1_45

Yang, L., Wang, B., Zhang, Y., Ye, Z., Wang, Y., & Li, P. (2018b). Willing to pay more for high-quality schools? A hedonic pricing and propensity score matching approach. International Review for Spatial Planning and Sustainable Development, 6(1), 45–62. https://doi.org/10.14246/irspsd.6.1_45

Yu, D. (2007). Modeling owner-occupied single-family house values in the City of Milwaukee: A geographically weighted regression approach. GIScience and Remote Sensing, 44(3), 267–282. https://doi.org/10.2747/1548-1603.44.3.267

Yuen, B., & Choi, S. (2012). Making Spatial Change in Pakistan Cities Growth Enhancing (PK 11/12; World Bank Policy Paper Series on Pakistan).

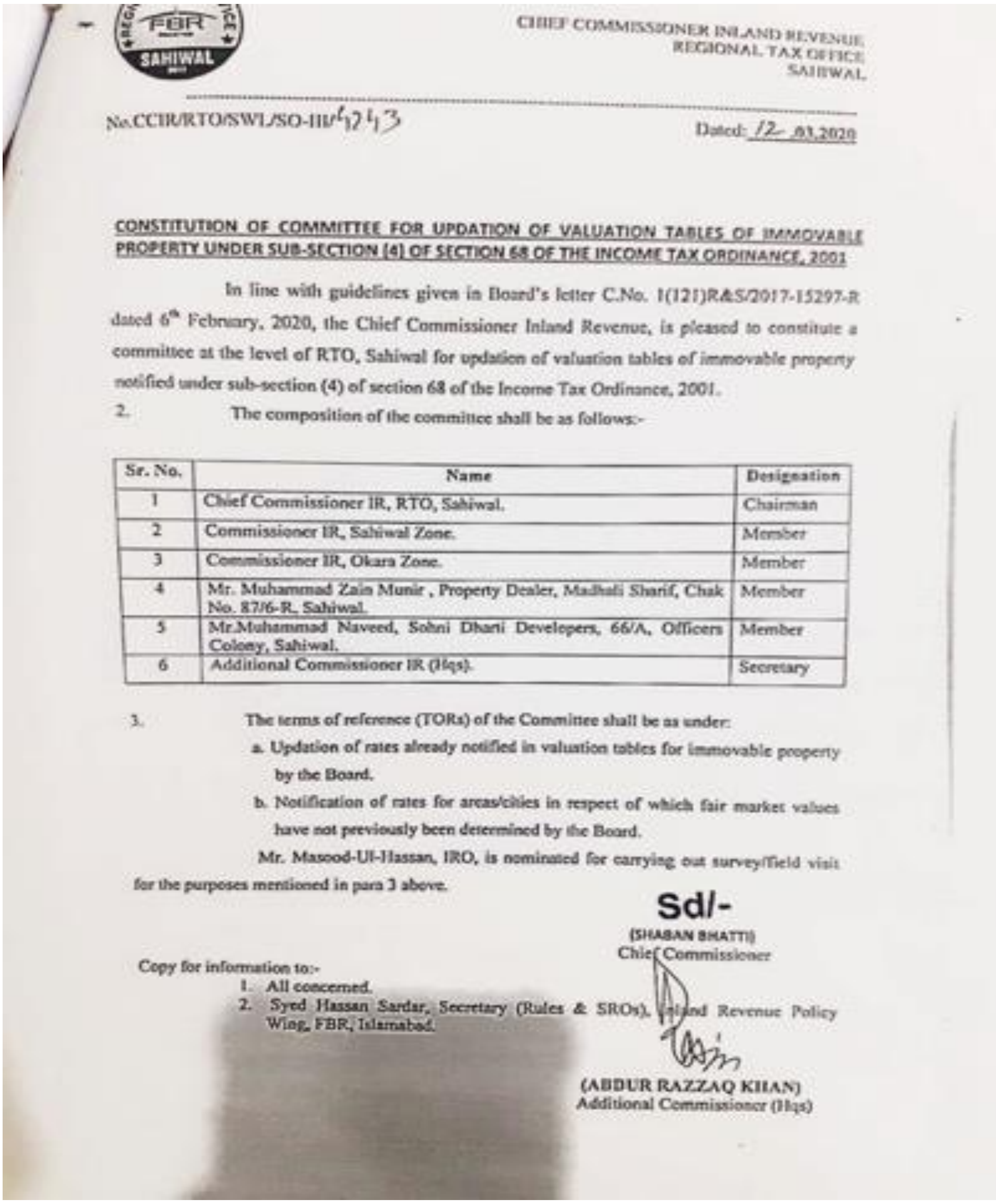
Zaman, K. U., & Baloch, A. A. (2011). Urbanization of Arable Land in Lahore City in Pakistan: A Case-Study. Canadian Social Science, 7(4), 58–66.

Zhang, L., Chen, J., Hao, Q., & Li, C. Z. (2018). Measuring the NIMBY effect in urban China: the case of waste transfer stations in metropolis Shanghai. Journal of Housing and the Built Environment, 33(1), 1–18. https://doi.org/10.1007/s10901-017-9565-2

APPENDICES

Figure 15. Circular about the Valuation of Immovable Properties

Figure 16. FBR Notification about the Constitution of Committee



FBR SAHIWAL

CHIEF COMMISSIONER INLAND REVENUE
REGIONAL TAX OFFICE
SAHIWAL

No.CCIR/RTO/SWL/SO-III/4243

Dated: 12-03.2020

CONSTITUTION OF COMMITTEE FOR UPDATION OF VALUATION TABLES OF IMMOVABLE PROPERTY UNDER SUB-SECTION (4) OF SECTION 68 OF THE INCOME TAX ORDINANCE, 2001

In line with guidelines given in Board's letter C.No. 1(121)R&S/2017-15297-R dated 6th February, 2020, the Chief Commissioner Inland Revenue, is pleased to constitute a committee at the level of RTO, Sahiwal for updation of valuation tables of immovable property notified under sub-section (4) of section 68 of the Income Tax Ordinance, 2001.

2. The composition of the committee shall be as follows:-

Sr. No.	Name	Designation
1	Chief Commissioner IR, RTO, Sahiwal.	Chairman
2	Commissioner IR, Sahiwal Zone.	Member
3	Commissioner IR, Okara Zone.	Member
4	Mr. Muhammad Zain Munir, Property Dealer, Madhali Sharif, Chak No. 87/6-R, Sahiwal.	Member
5	Mr.Muhammad Naveed, Sohni Dharti Developers, 66/A, Officers Colony, Sahiwal.	Member
6	Additional Commissioner IR (Hqs).	Secretary

3. The terms of reference (TORs) of the Committee shall be as under:

a. Updation of rates already notified in valuation tables for immovable property by the Board.

b. Notification of rates for areas/cities in respect of which fair market values have not previously been determined by the Board.

Mr. Masood-Ul-Hassan, IRO, is nominated for carrying out survey/field visit for the purposes mentioned in para 3 above.

Sd/-
(SHABAN BHATTI)
Chief Commissioner

Copy for information to:-

1. All concerned.
2. Syed Hassan Sardar, Secretary (Rules & SROs), Inland Revenue Policy Wing, FBR, Islamabad.

(ABDUR RAZZAQ KHAN)
Additional Commissioner (Hqs)

Figure 17. FBR Immovable Property Valuation Tables

Government of Pakistan
Revenue Division
Federal Board of Revenue
*****

Islamabad, the 23rd July, 2019.

NOTIFICATION
(Income Tax)

S.R.O. ९38 (I)/2019.- In exercise of the powers conferred by sub-section (4) of section 68 of the Income Tax Ordinance, 2001 (XLIX of 2001) and in supersession of its Notification No. S.R.O. 121(I)/2019 dated the 1st February, 2019, the Federal Board of Revenue is pleased to notify the value of immoveable properties in columns (3) and (4) of the Table below in respect of areas or categories of Lahore specified in column (2) thereof.

(2) The value for residential and commercial superstructure shall be —
(a) Rs.1500 per square foot if the superstructure is upto five years old; and
(b) Rs.1000 per square foot if the superstructure is more than five years old.

(3) In order to determine the value of constructed property, the value of open plot shall be added to the value worked out at sub-paragraph (2) above.

(4) This notification shall come into force with effect from 24th July, 2019.

LAHORE

ALLAMA IQBAL TOWN

S. No	Area	Value of Residential property per marla (in Rs.)	Value of Commercial property per marla (in Rs.)
(1)	(2)	(3)	(4)
1	ABDALIAN COOP SOCIETY	852,720	1,309,000
2	ABADI MUSALA MOUZA MUSALA	244,200	447,120
3	ABID GARDEN ABADI MUSALA	332,970	804,540
4	ADJOINING CANAL BANK ALL SOCIETY MOUZA KANJARAN	564,960	1,331,000
5	ADJOINING CANAL BANK ALL SOCIETY MOUZA IN SHAHPUR KHANPUR	746,900	1,326,380
6	AGRICHES COOP SOCIETY	570,900	1,039,500
7	AHBAB COLONY	262,680	392,610
8	AHMAD SCHEME NIAZ BAIG	390,443	800,228

Page 1 of 38

9	AHMAD NAGAR KATCHI KOTHI	441,210	728,640
10	AITCHISON COLLEGE COOP SOCIETY	643,720	887,700
11	AJODIAPUR	446,477	866,861
12	AJODIAPUR ALL SOCIETY AND TOWNS IN AJODIAPUR	526,205	1,016,894
13	AL-HAJAJ COOP SOCIETY	428,890	693,000
14	ALHAMAD PARK NIAZ BAIG	394,020	662,400
15	ALI PARK NIAZ BAIG	582,780	883,890
16	ALI RAZABAD	394,020	834,555
17	ALL SOCIETIES AND TOWN IN RAKH KHAMBA	374,220	695,200
18	ALL SOCIETIES AND TOWN IN AMIR KOT	661,980	917,400
19	ALL SOCIETIES AND TOWN IN MOUZA JULIANA	352,440	878,900
20	ALL SOCIETIES AND TOWN IN KHAMBA	428,890	691,900
21	ALL SOCIETIES AND TOWN IN MOHLANWAL	335,610	594,000
22	AMIR KOT (Old Abaadi)	230,340	363,630
23	AWAN TOWN	548,130	932,800
24	AWAISIA COOP SOCIETY	380,160	799,710
25	AZAM GARDEN	580,030	979,000
26	BADA PURA	422,235	686,550
27	BEHRIA TOWN	638,330	1,384,570
28	BAKAR MANDI	696,300	1,276,000
29	BHAIGILL ALL SOCIETIES	323,730	1,005,158
30	BHULA GARI	394,020	679,305
31	BOR SOCIEITY	692,780	1,320,000
32	MAIN BUD ROAD MOTOR WAY CHOWK TO CHOWK YATEEMA KHANA	–	2,317,020
33	CAMPUS VIEW TOWN	398,640	867,848
34	CANAL BANK ROAD NEAR TWO SIDE CAMPUS TO THOKHAR	820,270	1,890,900
35	CANAL BREEZE SOCIETY	445,830	790,050
36	KANAL BURG SOCIETY	577,170	1,162,305
37	KANAL VIEW	638,220	1,216,470
38	KANAL VIEW SOCIEITY	746,020	1,351,900
39	CHAK MOZANG	516,120	1,176,795
40	CHAK MOZANG ALL SOCIEITIES	596,629	1,247,180
41	CHUNG PUNJGRAIN	318,912	632,903
42	CHUNG PUNJGRAIN ALL SOCIETIES	459,690	760,035
43	DEENANATH AMIR KOT	215,985	372,773
44	DHNA SING WALA	478,665	976,178
45	DHNA SING WALA ALL SOCIETIES	506,880	951,510
46	DHOLANWAL	506,880	794,880

Page 2 of 38