Policy Brief

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

Shoaib Khalid, Fariha Zameer & Fariha Zameer

(This document is unedited author's version submitted to RASTA)

INTRODUCTION

This study investigated the spatial heterogeneity of the real estate property values as a basis for the formulation of a sophisticated and scientific valuation model of urban immoveable properties. Our motivation for taking up this study was 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 highly incompatible with the fair market values of the urban immovable properties. Moreover, there is no mechanism to record actual market transactions. 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. 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. 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.

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;

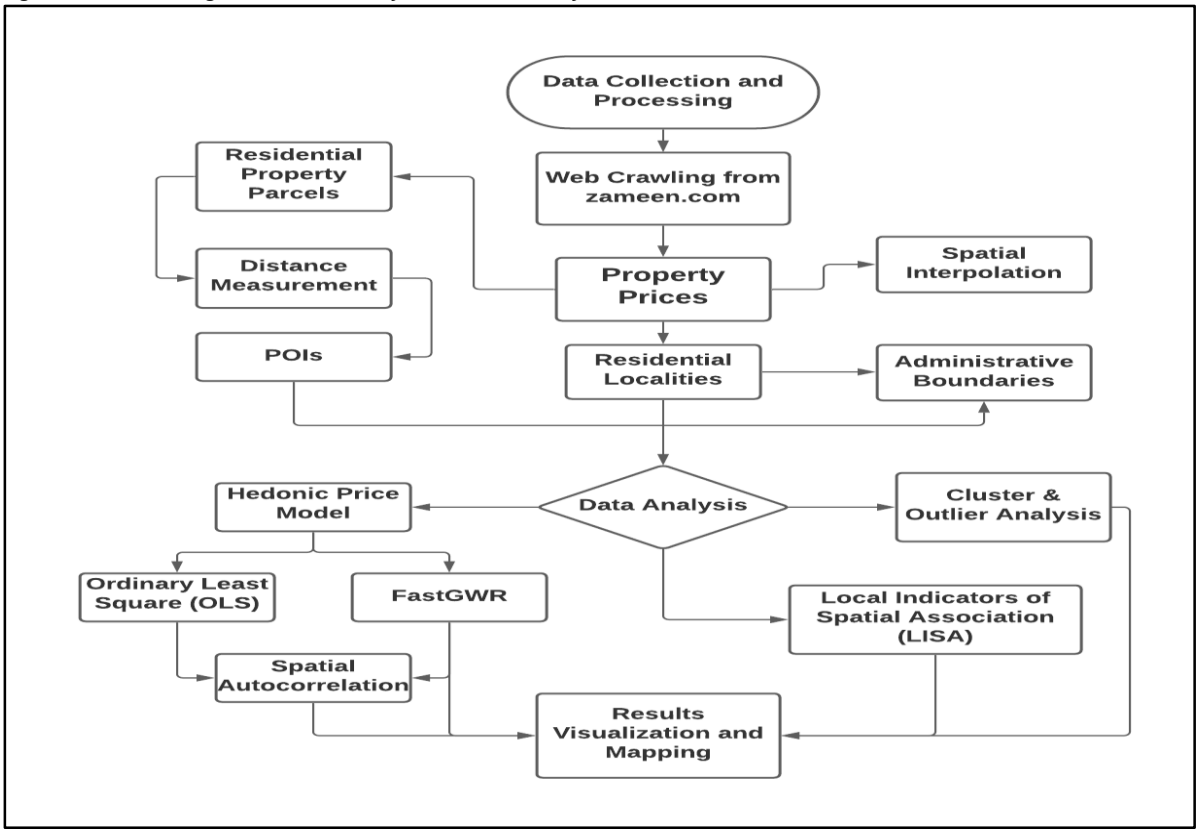
1. To examine the dynamics of urban immovable property values on the basis of location specific parameters using the Big Data.
2. To investigate the spatial variations in the urban immoveable property values.

RESEARCH METHODOLOGY

Policy Brief

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.

Figure 1. Methodological Framework of the Research Project



Data and Sources

The data used in this study were collected from different sources including the primary field survey, governmental and private organizations. 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. This generated a big dataset of 1.2 million residential property parcel dataset. The information about the residential property price and the location of property was obtained from the property portal of zameen.com through a web-scraping program. The data were also collected through an intensive field survey in both the cities. 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 was obtained from the private organizations. The locational attributes were generated in the study areas using Geographic Information System and the mapping services of google maps.

Processing

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 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 Spatial Hedonic Valuation Model

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

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

Where y is the estimated value of the house property, β_0 is the intercept, A is the floor area of a house property, rest of the βs are spatial variables and ϵ represents error term.

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, 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 resulted in nine potential explanatory variables qualifying for the final model.

Table 1. 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 Attributes	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 ($ 66.83 B)

	Average Price	Average Price per square meters (per Marla)	PKR 98,279 (PKR 2.47 M)	PKR 97,897 (PKR 2.04 M)
Amenities	Solid Waste	Number of solid waste facilities and transfer 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, clinics, and dispensaries)	318	2,902
Recreation	Parks and Recreation	Number of public parks and recreational sites	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-Commercial	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

Analyses

We performed the hedonic valuation analysis using OLS a linear regression model on the dataset of Lahore and Faisalabad separately. In order to check the degree of spatial auto-correlation of regression residual values, the Global Moran's I test was applied. A hotspot analysis was carried out to discover the significant high and low value clustering, using residuals values. The Cluster and Outlier analysis was applied to identify the clustering pattern of property prices across the study area. The property datasets were finally analyzed through the Geographic Weighted Regression (GWR) which explored the spatial the spatial variability of property values in our study area.

MAJOR FINDINGS

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 property values up to 75% for Faisalabad and around 85% for Lahore. Results show that the floor area, proximity to health facilities, recreational sites and market places add premium to the property values, while nearness of educational institutions, worship places and solid waste facilities lessen the property values in both the cities. However, closeness to industrial units and graveyards affect the property values negatively in Lahore but positively in Faisalabad.

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 property values.

RECOMMENDATIONS AND POLICY IMPLICATIONS

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. Spatial determinants of housing valuation are significantly important, and hence, must be considered in the policy formulation and future urban planning and design.

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.

LIMITATIONS

There were several limitations in the study, for instance, the structural attributes of the residential properties are not available for the entire study area. The data of socio-economic determinants that might influence the housing prices at different spatial scales is also missing. 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 analyze the property values in connection with the spatial features and highlight their influence on housing prices, which has been neglected in the previous studies. We lack in high computing power which can compute the large dataset of Lahore city, therefore, we had to analyze the data in smaller parts. In order to compute the data of big cities, the High Performance Computing Cluster (HPCC) is required.