Box 2
Spatial Analysis of New Home Prices in Bogota, Medellín, and Cali Using a Geostatistical Approach

María Fernanda Meneses
Camilo Eduardo Sánchez*

The housing market has been one of the sectors of the economy that has performed favorably despite the health crisis: new home sales have reached historic levels so far in 2021. However, the momentum observed in demand has not been accompanied by a similar response on the supply side. This, in turn, has generated pressure on prices.[1]

Monitoring the level of housing prices in Colombia has focused on analyzing this market in different dimensions: for example, by considering the type of housing (LIH and non-LIH), the city and the socio-economic bracket, among other factors. However, this type of analysis may hide some of the heterogeneities within a city. Therefore, this box outlines a spatial analysis tool that uses georeferenced information for the first time to estimate the growth in new home prices in those areas of Bogota, Medellín and Cali where no information is available.[2] The objective is to represent the dynamics of the annual variation in home prices, in their entirety, on maps of the three cities.[3]

1. Methodological Approach

This exercise in geostatistical analysis is based on new housing project information from *Galería Inmobiliaria* to September 2021. Spatial location data (measured in coordinates) and the real annual percentage change in prices per square meter in the urban areas of Bogota, Medellin and Cali are used. From this database, we extract the projects that were in the sales phase, both in September 2020 and September 2021. In other words, we work with the same projects in both periods.[4]

The change in price at a point without information is calculated as the weighted average of the closest observed data. This spatial interpolation technique is known as Kriging and the weights are obtained based on three elements: 1) the distance between observed points, 2) the prediction location and 3) the defined spatial structure. Accordingly, the prediction function corresponds to:

$$z^*(s_0)=\sum_{i=1}^{n}\lambda_i z(s_i)$$

where $z^*(s_0)$ is the predicted growth of price per square meter at coordinates s_0 and λ_i is the weight of observation $z(s_i)$. The variance of the prediction error is minimized [5] to obtain parameters λ_i, which ensures the predictor is unbiased and of the lowest possible variance among the linear predictors. The expression that is minimized to find λ_i is, in turn, a function of the spatial covariance of the data, which comes from the spatial structure in each city.

1.1. Spatial Structure of the Data

It is assumed that the growth in price level follows a stationary process, that it, the mean is constant and the covariance is a function of the spatial distance between the observations. Therefore, the location or address of a project does not add information about the variation of its price. Under this assumption, there is a relationship between spatial covariance and a function known as ***semivariance***, which is defined as the variance of price growth spread between two points:

$$\gamma(s_i,s_j)=\frac{1}{2}\,var(z(s_i)\text{-}z(s_j))$$

Where $z(s_i)$ is the price growth per square meter observed at point s_i; and $\gamma(.)$ is the semivariance function. In this paper, the spatial structure of each city is identified from the semivariance function, and then the covariance function used in the prediction exercise is cleared.

* Los autores pertenecen al Departamento de Estabilidad Financiera del Banco de la República. Sus opiniones no representan las del Banco de la República ni las de los miembros de su Junta Directiva. Los autores agradecen a Eduard Felipe Mora por la facilitación de los polígonos y conversión de coordenadas necesarios para este trabajo.

1 See the special report: *Análisis de la cartera y del mercado inmobiliario en Colombia*, first half of 2021.

2 These are the three main cities in Colombia. Together, they account for 68.9% of the total sales of *La Galería Inmobiliaria's* base at September 2021.

3 This exercise does not estimate a new home price index.

4 Working with the same active projects in both periods prevents results on home price growth from being due to the launching of new projects.

5 The prediction error variance is defined as: $Var(z(s_0)-z^*(s_0))$.

The theoretical semivariance models have the ***nugget effect***, the ***range*** and the ***saddle*** as parameters. The first refers to the semivariance that two points in very close proximity should have. The second refers to the distance beyond which two observations are no longer considered to have a spatial relationship, while the latter indicates the value of the semivariance (or covariance) beyond that distance.[6] Theoretical models with different initial parameters were estimated in this way with different econometric techniques to determine, through the mean squared error, those that best fit the data (Table B2.1).

With these semivariance models for each of the cities, their corresponding covariance function is introduced into the prediction error expression, their variance is minimized to obtain each λ_i, and, finally, the prediction exercise is carried out.

2. Results and Considerations

Graph B2.1 shows the estimated real annual growth in prices per square meter for the cities under analysis as of September 2021. The shading in the maps near purple and black represents the areas of the city where housing projects became more expensive, while the yellow areas indicate projects with no major changes in price. The blue shading reflects areas where prices declined.

In the case of Bogota,[7] the increase in prices per square meter was observed to a greater extent in the southern part of the city, although sectors such as Usaquén (the most expensive neighbor, after Chapinero) also observed increases (Graph B2.1, panel A). In general, the largest rise in housing project prices occurred in areas with relatively low home prices, such as Kennedy, Ciudad Bolívar, Tunjuelito, Usme and Rafael Uribe, and some parts of Fontibón.

After the spatial estimation exercise and to complement the study of prices per square meter, a descriptive analysis is performed with variables observed in this market, to infer characteristics by locality or sector. It is found that in the areas of Bogota where prices increased the most, there were also large sales. In the case of Usaquén, there was no increase in demand. On the other hand, in some sectors such as Ciudad Bolívar and Usme the greatest change in the price per square meter was evidenced for LIH-type projects, which account for more than 80% of the housing projects in those areas. The opposite is the case in Usaquén, where LIH real estate accounts for a low percentage.

For Medellín,[8] the acceleration in prices was observed in the northwestern sectors and in some parts of the south. By sector, this occurred in Guayabal,[9] Belén, Robledo and Poblado (the most expensive area of the city; Graph B2.1, panel B). Moreover, 69 active housing projects were registered in Medellín, with less than 5% being LIH projects. Therefore, the rise in prices was mostly for non-LIH projects.

In Cali,[10] the largest increase in prices was observed in the southwest part of the city, particularly in communes 18, 19 and 22, which are characterized as having a socio-economic bracket equal to or higher than bracket four (Graph B2.1, panel C). While the projects in commune 18 maintain relatively lower prices and register a high share of LIH projects, in the case of housing projects in communes 19 and 22, the real estate is mostly non-LIH housing and registers one of the highest prices in the city. On the other hand, more than 30% of the projects in Cali are located in the commune 2 (north zone) and are mostly non-LIH. Finally, although communes 4 and 16 register few housing projects, the most home units were sold in these areas of during the last few months.

Table B2.1
Summary of Semivariogram Models

City	Theoretical model	Mean squared error
Bogota	*Powered exponential*	MCO: 2.05 MCP1: 2.13 MCP2: 2.25 MLE: 10.35 REML: 11.41
Medellin	*Exponential*	MCO: 18.91 MCP1: 19.09 MCP2: 20.44 MLE: 349.67 REML: 349.67
Cali	*Wave*	MCO: 95.92 MCP1: 95.52 MCP2: 115.01 MLE: 114.39 REML: 126.45

OLS: ordinary least squares, WLS1: weighted least squares by the number of pairs of points seen at each distance.
WLS2: weighted least squares by the number of pairs and semivariance seen at each distance.
MLE: maximum likelihood, REML: restricted maximum likelihood.
Source: *Banco de la República*

6 For a better understanding of the theoretical models, see Cressie, N. (1991).

7 At the time of analysis, there are no data on new residential housing projects in Tunjuelito and Sumapaz.

8 At the time of analysis, there are no data on new residential housing projects in communes 1 (Popular), 2 (Santa Cruz), 3 (Manrique), 4 (Aranjuez), 5 (Castilla), 6 (12 Octubre), 8 (Villa hermosa) and 13 (San Javier).

9 In some areas of Medellín, such as Guayabal and Robledo, this increase occurred in large housing projects.

10 At the time of analysis, there are new housing projects only in communes 2, 3, 4, 10, 16, 17, 18, 19 and 22.