





Determinantes del precio de la vivienda nueva en Bogotá para el año 2019: una aproximación a través de un modelo semiparamétrico de regresión espacial

 Jurgen D. Toloza Delgado¹,  Oscar O. Melo-Marínez² y Juan S. Azcarate-Romero³

Recepción: 13-12-2020 - Aceptación: 18-11-2021 - En línea: 01-12-2021

MSC: 62G08, 91B72

doi:10.17230/ingciencia.17.34.2

Resumen

Este trabajo toma como punto de partida los recientes avances en el campo de la econometría espacial para desarrollar un modelo de regresión semiparamétrico que permite la inclusión de no linealidades y el modelamiento de la heterogeneidad espacial a través de una función bidimensional que depende de las coordenadas geográficas. La metodología se aplica en un modelo hedónico para el precio de la vivienda nueva en Bogotá donde se obtiene un ajuste destacable, en términos del error cuadrático medio y el R^2 . El resultado empírico muestra que el estrato, la condición de entrega y el estado constructivo afectan el precio de manera lineal, mientras que el área, y las distancias a parques, vías y estaciones de Transmilenio presentan resultados no lineales; además se logró modelar

¹ Universidad Nacional de Colombia, jdtolozad@unal.edu.co, Bogotá, Colombia.

² Universidad Nacional de Colombia, oomegom@unal.edu.co, Bogotá, Colombia.

³ Universidad Nacional de Colombia, jsazcarater@unal.edu.co, Bogotá, Colombia.

la tendencia espacial que representa la ubicación sobre el valor de la vivienda, evidenciando un incremento hacia el nororiente de la ciudad. Así, se concluye que el modelo estimado permite medir de manera flexible la relación entre las variables explicativas y la dependiente, estableciéndose como una buena alternativa para entender la formación de los precios en el mercado inmobiliario.

Palabras clave: Modelos hedónicos; econometría espacial; precios de vivienda; regresión semiparamétrica.

Determinants of New Housing Prices in Bogotá for 2019: an Approach Through a Semiparametric Spatial Regression Model

Abstract

This document uses the recent advances in the field of spatial econometrics to develop a semi-parametric regression model that allows the inclusion of non-linearities and the modeling of spatial heterogeneity through a two-dimensional function that depends on geographic coordinates. The methodology is applied in a hedonic model for the price of new housing in Bogotá where a remarkable fit is obtained, in terms of the mean square error and the R^2 . The empirical result shows that the housing delivery condition, stratum, and construction state affect the price in a linear way, while the area, and the distances to parks, roads and Transmilenio stations present non-linear results, additionally, it was possible to model the spatial trend that represents the location on the value of the house where an increase is appreciated towards the northeast of the city. Thus, it is concluded that the estimated model allows the relationship between the explanatory variables and the dependent variable to be measured flexibly, establishing itself as a good alternative to understand the formation of prices in the real estate market.

Keywords: Hedonic models; spatial econometrics; housing price; semiparametric regression.

1 Introducción

El mercado de lo habitacional se comporta de forma particular y diferente a otros mercados de bienes y servicios. La vivienda es un bien compuesto y heterogéneo; por un lado, es compuesto porque su precio está determinado por sus atributos tanto físicos como de localización; por otro lado, es un bien heterogéneo porque cada unidad es diferente a las demás y no se puede

reproducir una vivienda idéntica a otra. Por esto, entender la conformación y dinámica de los precios en el mercado de vivienda sigue siendo un tema de interés para investigadores, planificadores urbanos, constructores y agentes inmobiliarios.

En la actualidad se pueden encontrar numerosos trabajos que abordan la determinación del precio de las viviendas desde distintos enfoques, siendo el modelo de precios hedónicos uno de los métodos más utilizados en la literatura [1]. Estos modelos permiten conocer el precio implícito de cada uno de los atributos que componen los inmuebles donde las características físicas como el área, el número de habitaciones, número de baños y las zonas comunes, entre otras, juegan un papel fundamental en la conformación del precio de la vivienda [2]. Sin embargo, es importante vincular dentro del modelo hedónico el campo espacial para tener una lectura más amplia del comportamiento de los precios en el mercado y de esta manera solucionar los posibles problemas que se pueden presentar cuando se trabaja con datos indexados en el espacio.

La idea de involucrar el componente espacial para el estudio de los precios de la vivienda encuentra su razón en dos ejes fundamentales de la econometría espacial: la dependencia y la heterogeneidad. El primero de ellos, la dependencia espacial, se refiere a la relación funcional que tiene una variable o fenómeno determinado en una región con el comportamiento de la misma variable o fenómeno en regiones vecinas [3]; en el caso de la vivienda se trataría de la relación existente con su entorno, el cual puede generar externalidades positivas o negativas a cada una de las unidades habitacionales de manera distinta. Por su parte, la heterogeneidad espacial indica la variación de las relaciones en el espacio [4] y puede existir en dos vías: la primera se da en los estimadores y se denomina inestabilidad estructural, y la segunda se encuentra en los errores y se denomina heterocedasticidad. En ese sentido, la heterogeneidad estructural se concibe como un fenómeno que indica la ausencia de estabilidad del comportamiento humano en un espacio determinado, es decir, los parámetros varían de acuerdo con la localización geográfica [5], mientras que la heterocedasticidad es causada por la omisión de variables o errores de especificación del modelo.

De esta manera, los modelos espaciales autorregresivos (SAR) popularizados por [3] permiten incorporar fácilmente la dependencia

espacial dentro de la predicción y han sido ampliamente utilizados en el análisis de precios hedónicos alrededor del mundo [1]. No obstante, estos modelos se quedan cortos al no tener en cuenta la heterogeneidad espacial estructural que puede presentarse en los datos indexados en el espacio y el hecho de asumir únicamente relaciones lineales entre las variables explicativas y la dependiente.

En este sentido, algunos autores [6] han extendido el modelo SAR básico para que permita la incorporación de no linealidades entre las variables explicativas y el predictor, así como la heterogeneidad y la dependencia espacial. De esta forma, este modelo es mucho más flexible y obtiene mejores resultados, en términos de predicción, que aquellos que se restringen únicamente al ámbito paramétrico, siendo considerado como uno de los más recientes avances en el campo de la econometría espacial [7].

En consecuencia, el presente trabajo busca aplicar el modelo SAR semiparamétrico en el entendimiento y análisis del mercado de vivienda nueva para la ciudad de Bogotá, con el fin de evidenciar la dependencia espacial, las no linealidades y la posible heterogeneidad que pueda observarse de acuerdo con la información que recolecta Camacol B&C en el marco de su sistema de información Coordinada Urbana.

El motivo de escoger a Bogotá como ciudad de estudio es que la capital del país, según Camacol B&C [8], aporta cerca del 20 % de las unidades de vivienda vendidas a nivel nacional. Además, moviliza más de 8 billones de pesos anuales en la adquisición de unidades habitacionales nuevas; es decir, es una de las plazas más importantes de comercialización y construcción de vivienda en el país.

Para abordar esta problemática, el presente artículo se desarrollará de la siguiente manera: primero se presenta una revisión bibliográfica con los trabajos más destacados que han abordado esta temática en la capital del país. Posteriormente, se presentan los métodos estadísticos que permiten la estimación del modelo SAR clásico y del SAR semiparamétrico, los cuales son aplicados en el modelamiento del precio por metro cuadrado para la oferta de vivienda nueva en Bogotá, a noviembre de 2019. Por último, se presentan las conclusiones y se establece el trabajo futuro a desarrollar para seguir avanzando en esta temática que puede ser bien aprovechada por diversos actores del sector inmobiliario.

2 Revisión de literatura

El estudio del mercado de vivienda por medio de modelos de precios hedónicos ha sido amplio y diverso. Lancaster [9] estableció que algunos bienes podrían ser concebidos como una cesta de características individuales y que su utilidad estaría derivada de la contribución de cada característica. Sin embargo, fue años más tarde que [10] formuló, gracias a los aportes de Lancaster, el modelo formal de precios hedónicos.

De igual manera, [11] determinó que la función de precios hedónica es un precio de equilibrio en función de sus características y que la estimación de los parámetros de cada característica es el precio marginal de la i -ésima característica frente al precio total del inmueble. Además, resaltó la importancia de incluir el enfoque espacial en estos modelos para no incurrir en especificaciones o resultados erróneos y estableció el modelo con rezago espacial de la variable endógena y el modelo con rezago espacial en el error (SEM por sus siglas en inglés), entre otros.

Para la región, los modelos de precios hedónicos se han aplicado para entender la valoración de los individuos a distintos atributos en los mercados de Chile, Argentina, Colombia, entre otros. Por ejemplo, en Chile, [12] aplicó un modelo estimado por medio de MCO donde concluyó que atributos como la antigüedad, la superficie construida y del terreno tiene efectos significativos en el precio de la vivienda, así como la distancia a estaciones del metro que tiene efectos no lineales. Por su parte, [2] por medio de la metodología de precios hedónicos estimó un modelo MCO para concluir que las preferencias de los atributos de una vivienda en renta cambian significativamente de acuerdo con el grupo socioeconómico, principalmente en lo relacionado con baños y alcobas, no obstante, encontró que las distancias a establecimientos comerciales y educaciones no tienen incidencia en la formación del precio de alquiler.

Para Argentina, [13] por medio de variables físicas y de localización investigó sobre la conformación de precios de vivienda en Buenos Aires a distintos niveles geográficos. La autora estimó modelos tanto por MCO como modelos con el efecto espacial concluyendo que estos últimos son los que mejor resultados obtuvieron estadísticamente hablando. Adicionalmente, evidenció que, aunque haya evidencia de autocorrelación y dependencia espacial, los resultados van a ser variados dependiendo

la escala a la cual se realice el análisis como por ejemplo el efecto de la cercanía a vías que en ciertas localidades fue significativo con signo negativo mientras que en otras no lo fue.

Por su parte, [14] realizó un análisis sobre el efecto de riesgos naturales y la calidad medioambiental al precio de la vivienda en Buenos Aires. Su trabajo se basó en reforzar las aplicaciones de econometría espacial en el mercado inmobiliario argentino mediante un modelo hedónico, donde concluye que las aproximaciones mediante MCO se alejan de la realidad e introduce regresiones espaciales (SAR) por cuantiles al modelo siguiente: hedónico. Además, sustenta la utilización de modelos SAR con lo siguiente aunque en general la dependencia espacial en el término de error no afecta la consistencia de los estimadores, puede tener un efecto perjudicial en la eficiencia.

En la misma línea del trabajo de [14] y [15] aplicaron una regresión hedónica por cuantiles para comparar el precio de las viviendas en ciudades de Colombia, empleando además un método complementario de emparejamiento el cual buscó obtener grupos de viviendas idénticas en sus características y establecer comparaciones más homogéneas al momento de estimar los diferenciales de precios.

Para el caso de Medellín, se estimó un modelo de un índice de precios hedónicos incorporando el tema espacial con el objetivo de calcular índice de calidad del vecindario. Los autores construyeron el modelo con el precio del arriendo como variable dependiente, y como variables explicativas el área, el estado del apartamento, la pertenencia en propiedad horizontal, el número de habitaciones y baños, entre otros. Adicionalmente, incorporaron la dependencia espacial en el modelo a través de un autoregresor (SAR), en los errores (SEM) y en otro modelo con la incorporación tanto en las variables explicativas, como en el error (SARAR), concluyendo que este último fue el de mejor estimación con el coeficiente ρ positivo y estadísticamente significativo [16].

Otro estudio realizado para Medellín fue el de [17], donde por medio de una regresión geográficamente ponderada estimaron la influencia de tener cerca una estación del metro a la vivienda. En este artículo los autores concluyeron que la estación de metro influye de manera positiva en el precio sobre las viviendas ubicadas a un radio de 600 metros de una estación,

sin embargo, aquellas viviendas ubicadas frente a las estaciones del metro tienen un efecto negativo.

En las ciudades de la costa también se han aplicado modelos espaciales identificando los cambios en el precio de la vivienda con respecto a la distancia a la playa, en este sentido se encuentra el trabajo [18], donde se realiza un modelo aditivo semiparamétrico para estimar la influencia que ejercen las características estructurales, y los atributos de entorno sobre el precio de la vivienda en Cartagena. Los resultados obtenidos muestran que hay una correlación positiva entre el precio y variables como el área construida y del terreno, la calidad estructural, número de alcobas, baños, estrato y la vista al mar. Con respecto a la localización de la vivienda, se encontró que las viviendas ubicadas sobre la playa tienen en promedio un aumento del precio de más de 7 millones de pesos y este valor disminuye a medida que se aleje la vivienda, este resultado es similar con la distancia a la bahía de Cartagena y contrario a lo que ocurre con las viviendas sobre la Ciénaga La Virgen.

Para el caso puntual de Bogotá, se han realizado diversas investigaciones que acogen algunos de los aportes anteriormente mencionados. Específicamente, [19] con un trabajo pionero estimaron las variables que determinan el precio por metro cuadrado para usos residenciales y comerciales por medio de un modelo de mínimos cuadrados ordinarios; los autores concluyeron que entre los atributos más significativos que explican el comportamiento de precios en la ciudad se encuentra el uso del suelo, el índice de ocupación, el área construida, la calificación de la construcción según las fichas catastrales y específicamente para el caso de vivienda, los estratos.

Por su parte, [20] a través de un modelo hedónico espacial de rezago y error espacial (SARMA) evidenciaron el impacto generado en el precio de las viviendas a través de las características de los inmuebles, su localización, variables de seguridad, ambientales y la cercanía a los ejes de Transmilenio, siendo esta última la variable central de la investigación. Allí concluyeron que los precios disminuyen a medida que los inmuebles se acercan a los ejes de Transmilenio, no obstante, la plusvalía (aumento en el valor del suelo) que generan estos desarrollos se evidencia de manera positiva para los inmuebles ubicados entre 200 y 500 metros de esta red de transporte masivo.

De modo similar, [21] planteó dos modelos hedónicos espaciales (SAR y SEM) con el fin de encontrar los determinantes de los precios de las viviendas nuevas en Bogotá. La función hedónica tuvo en cuenta tres elementos cruciales en el precio: el conjunto de características estructurales, la mayoría con relación positiva frente al precio; las características de localización, puntualmente la tasa de homicidios con relación negativa y el número de viajes por UPZ que registraron una relación negativa si es el destino final y positiva si es el origen; y finalmente la situación del proyecto al interior del mercado, en donde el autor encontró una relación negativa frente al tamaño del proyecto pero positiva con respecto al estado a la hora de compra. Asimismo, se concluyó que en los precios de vivienda existe heterogeneidad y autocorrelación espacial tanto global como local ¹.

Las metodologías de precios hedónicos con enfoque espacial también se han utilizado para la creación de índices tal como lo demuestran [22], quienes a través de un modelo SEM buscaron controlar la dependencia espacial de los errores del modelo para mejorar el ajuste del índice de precios de vivienda. Los autores concluyen que la metodología propuesta sí mejora la capacidad predictiva frente a los modelos convencionales y que el uso de la localización incluida en los errores podría explicar en mejor medida las variables omitidas.

En síntesis, los trabajos revisados dan cuenta de que el precio de los inmuebles en la capital del país tiene un marcado componente de correlación espacial, el cual debe ser tenido en cuenta para obtener una predicción óptima y una inferencia adecuada que permita dar conclusiones válidas. No obstante, la mayoría de los modelos dentro de la literatura se basan en relaciones lineales entre la variable dependiente y las variables explicativas, lo cual pone en evidencia la ausencia en el uso de modelos semiparamétricos que pueden tener un mejor ajuste y permitir un análisis más adecuado de la dinámica inmobiliaria en la ciudad de Bogotá, donde pueden ser concebidos como una valiosa herramienta dentro de los procesos de planeación urbanística.

¹La autocorrelación global mide el grado de relación de cada unidad con todo el espacio mientras que la autocorrelación local realiza la medición con sus vecinos más cercanos. El principal método de contraste es la I de Morán en donde la hipótesis nula establece que la unidad observada se comporta de manera aleatoria.

3 Metodología

Aunque hay una amplia variedad de formas funcionales que pueden adoptar los modelos econométricos espaciales, este trabajo se enfocará únicamente en el modelo SAR, tanto en su versión lineal o clásica [3] como en su contraparte semiparamétrica [6]. En este sentido, se entiende como un modelo SAR aquel que incorpora un rezago espacial de la variable dependiente (\mathbf{WY}), lo que implica que el valor de la observación Y_i dependerá del valor que tomen las observaciones “vecinas” o “cercanas”. Esta situación cobra especial relevancia en el análisis de mercados inmobiliarios, ya que en teoría el precio de un inmueble dependerá del valor de los inmuebles que tenga a su alrededor.

El modelo SAR clásico se puede expresar como:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \rho\mathbf{WY} + \mathbf{u} \quad (1)$$

donde $\mathbf{u} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$, $\boldsymbol{\beta}$ es un vector de $k \times 1$ parámetros de las $k - 1$ variables explicativas, ρ es un parámetro autorregresivo y \mathbf{W} es una matriz de pesos espaciales. Además, el regresor \mathbf{WY} genera problemas de endogeneidad, lo cual produce estimadores inconsistentes al utilizar mínimos cuadrados ordinarios (MCO), por lo cual es necesario utilizar otras metodologías de estimación como máxima verosimilitud [3].

3.1 Modelo SAR semiparamétrico (PS-SAR)

Planteado originalmente por [6], el modelo se concibió como una generalización del modelo SAR presentado previamente en la ecuación (1), el cual permite incluir tendencias espaciales no paramétricas de una o hasta 3 variables (en el caso de modelos espacio-temporales), lo cual mejora considerablemente la capacidad predictiva y el ajuste. En ese sentido, este modelo se concibe como uno de los más recientes avances en el campo de la econometría espacial [7] y ha sido utilizado en el análisis de crecimiento económico [23], dinámicas asociadas al desempleo [24] y el ajuste de modelos de precios hedónicos para vivienda [25].

El hecho de no restringir las relaciones entre las variables explicativas y la variable dependiente al ámbito lineal es quizás el mayor beneficio que se

obtiene de estos modelos. Además, la inclusión de las tendencias espaciales permite tener en cuenta la dependencia y la heterogeneidad espacial [25], siendo una alternativa a los modelos de enfoque local que se fundamentan en el uso de regresión geográficamente ponderada [26].

Sin embargo, los modelos semiparamétricos presentan una limitante a considerar y es que, al dejar atrás el supuesto de linealidad en los parámetros, se pierde la interpretabilidad económica que se puede abstraer a través de las derivadas parciales respecto a cada una de las variables, imposibilitando así el análisis de semielasticidades que generalmente se utiliza en la estimación de modelos hedónicos. En este sentido, el comportamiento no lineal podrá entenderse desde el gráfico generado en el proceso de estimación y las derivadas deberán obtenerse utilizando otras funciones como las propuestas en la librería `gratia` de R [27], aunque estas últimas solo son compatibles con los resultados del paquete `mgcv` [28].

El modelo PS-SAR se puede expresar como:

$$\mathbf{Y} = \rho \mathbf{W}\mathbf{Y} + f(x_{lat.}, x_{lon.}) + \mathbf{X}\boldsymbol{\beta} + \sum_{j=1}^q f_j(\mathbf{x}_j) + \boldsymbol{\omega} \quad (2)$$

donde $|\rho| < 1$ es un parámetro espacial autorregresivo que permite incorporar la interdependencia entre las observaciones, \mathbf{W} es una matriz de pesos espaciales definida previamente, $f(x_{lat.}, x_{lon.})$ es una función que depende de la longitud y la latitud; es decir, la ubicación de los inmuebles, \mathbf{X} es una matriz que incorpora las relaciones lineales entre las variables explicativas, $\boldsymbol{\beta}$ es un vector de parámetros, $\sum_{j=1}^q f_j(x_j)$ corresponde a funciones univariadas no paramétricas de variables explicativas que no están contenidas en \mathbf{X} y $\boldsymbol{\omega} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$.

De acuerdo con [29] las funciones no paramétricas univariadas $f_j(x_j)$ pueden expresarse como una combinación de funciones base o B-splines que incorporen una penalización dentro del proceso de optimización que controle el trade off entre el sesgo y la variabilidad.

$$f(\mathbf{x}) = \sum_{j=1}^c \mathbf{B}_j(\mathbf{x})\boldsymbol{\theta}_j, \quad j = 1, 2, 3, \dots, c$$

con \mathbf{B}_j como funciones base dadas por un B-spline y $\boldsymbol{\theta}_j$ un vector

de coeficientes de regresión de tamaño c , el cual depende del número de nodos utilizados en la construcción de las funciones base. De este modo, incorporando la penalización (asumiendo que no hay dependencia espacial) los estimadores de θ_j pueden obtenerse a partir de:

$$\|\mathbf{Y} - \mathbf{B}\boldsymbol{\theta}\|^2 + \lambda\boldsymbol{\theta}'\mathbf{P}\boldsymbol{\theta}$$

con $\mathbf{P} = \mathbf{D}'\mathbf{D}$, donde \mathbf{D} es una matriz de diferencias de segundo orden y λ es el parámetro de suavizamiento que controla el equilibrio entre el sesgo y la varianza.

Por su parte, de acuerdo con [30] las funciones bidimensionales también pueden expresarse como una combinación de funciones base. En ese sentido, [24] representan el término de interacción $f(x_{(lat.)}, x_{(lon.)})$ incorporado en la ecuación (2) de la siguiente forma:

$$f_{1,2}(\mathbf{x}_1, \mathbf{x}_2) = \sum_{j=1}^{c_i} \sum_{l=1}^{c_k} \mathbf{B}_j(\mathbf{x}_i) \mathbf{B}_l(\mathbf{x}_k) \boldsymbol{\theta}_{jl}, \quad j = 1, \dots, c_1 \quad l = 1, \dots, c_2$$

donde $\mathbf{B}_j(\mathbf{x}_i) \mathbf{B}_l(\mathbf{x}_k)$ es el producto tensorial de dos B-splines marginales y $\boldsymbol{\theta}_{jl}$ es un vector de coeficientes de tamaño $c_1 c_2 \times 1$. En este caso, la penalización está dada por:

$$\mathbf{P} = \lambda_1 \mathbf{D}'_1 \mathbf{D}_1 \otimes \mathbf{I} + \lambda_2 \mathbf{I} \otimes \mathbf{D}'_2 \mathbf{D}_2$$

Para la estimación, se propone utilizar la equivalencia entre modelos mixtos y P-splines, lo cual permite la estimación conjunta de los parámetros de suavizamiento [6]. De este modo, el modelo se puede reexpresar de forma matricial como:

$$\begin{aligned} \mathbf{A}\mathbf{Y} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\omega}, \quad \text{con } \boldsymbol{\alpha} \sim N(\mathbf{0}, \mathbf{G}) \quad \boldsymbol{\omega} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}) \\ \mathbf{A}\mathbf{Y} &= (\mathbf{I} - \rho \mathbf{W})\mathbf{Y} \end{aligned}$$

donde $\boldsymbol{\alpha}$ es un vector de parámetros de efectos aleatorios, $\boldsymbol{\beta}$ es un vector de parámetros de efectos fijos y los parámetros de suavizamiento se obtienen a partir del cociente entre la varianza del error y la varianza de los

efectos aleatorios $\left(\lambda_j = \frac{\sigma^2}{\sigma_{\alpha_j}^2}\right)$. De este modo, la función de verosimilitud restringida (REML) está dada por:

$$\ell_R(\lambda_j, \sigma^2, \rho) = -\frac{1}{2} \ln |\mathbf{V}| - \frac{1}{2} \ln |\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}| + \ln |\mathbf{A}| - \frac{1}{2} \mathbf{Y}'\mathbf{A}' \left(\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\right) \mathbf{A}\mathbf{Y}$$

Esta función de máxima verosimilitud restringida puede ser maximizada mediante métodos numéricos, donde los vectores $\hat{\boldsymbol{\beta}}$ y $\hat{\boldsymbol{\alpha}}$ están dados por:

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \left(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X}\right)^{-1} \mathbf{X}'\hat{\mathbf{V}}^{-1}\hat{\mathbf{A}}\mathbf{Y} \\ \hat{\boldsymbol{\alpha}} &= \hat{\mathbf{G}}\mathbf{Z}'\hat{\mathbf{V}}^{-1}(\hat{\mathbf{A}}\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \\ \text{con } \mathbf{V}^{-1} &= \frac{1}{\sigma^2} \left(\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z} + \sigma^2\mathbf{G}^{-1})^{-1}\mathbf{Z}'\right) \end{aligned}$$

Mínguez, Basile & Durbán [23] adaptaron el algoritmo SAP [31] para ajustar modelos con componentes multidimensionales, permitiendo la incorporación del parámetro autorregresivo espacial en el proceso de estimación, el cual es necesario para la obtención de la matriz \mathbf{A} y los parámetros $\hat{\boldsymbol{\beta}}$ y $\hat{\boldsymbol{\alpha}}$.

4 Resultados

A partir del censo de edificaciones realizado por Camacol B&C en el marco de su sistema de información georreferenciada, Coordenada Urbana, se obtuvo la información de 715 proyectos que al mes de noviembre de 2019 contaban con viviendas disponibles para la venta. A dichos proyectos se les calculó la distancia a estaciones de Transmilenio, parques y vías principales, con el fin de determinar si estas variables tienen una relación no lineal que impacta positiva o negativamente el precio de la vivienda. Adicionalmente, la base cuenta con información de la condición de entrega, el estrato, el número de baños y alcobas en cada una de las viviendas (ver Tabla 1 para más información de las variables).

Tabla 1: Definición de variables explicativas

Variable	Definición
TM	Distancia a la estación de Transmilenio más cercana (metros)
Estrato	Variable categórica asociada al estrato socioeconómico del proyecto (estrato 2, estrato 3, estrato 4, estrato 5 y estrato 6)
Habitaciones	Variable categórica asociada al número de habitaciones que tiene la vivienda (1 habitación, 2 habitaciones, 3 o más habitaciones)
Área	Área de la vivienda (metros cuadrados)
Parques	Distancia al parque más cercano (metros)
Vías	Distancia a vías principales (metros)
Baños	Variable categórica asociada al número de baños que tiene la vivienda (1 baño, 2 baños, 3 o más baños)
Condición entrega	Corresponde a los terminados de la vivienda (Obra negra, obra gris y acabados)
Estado constructivo	Variable categórica asociada al estado constructivo (preventa, construcción o terminado)

Antes de iniciar con el análisis de la información, se depuró la base de datos para eliminar observaciones con datos incompletos, quedando finalmente 708 registros. En este sentido, la Tabla 2 muestra las estadísticas descriptivas para cada una de las variables, donde se aprecia que el precio por metro cuadrado de la vivienda en Bogotá se ubica en promedio en 5.8 millones de pesos, con un área media de $72 m^2$.

Por otra parte, el 40% de los proyectos ofrecen viviendas de más de 3 habitaciones, el 49.3% tienen dos baños, el 45.8% de los proyectos se encontraban en preventa a noviembre de 2019 y el 88.3% se entregan con acabados.

Tabla 2: Estadísticas descriptivas de la base de datos.

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Área	708	72.484	48.616	17	47	81	403
Precio de venta (miles de pesos)	708	464,598.900	528,842.200	69,041	214,925	504,250	4,490,174
Precio por m^2 (miles de pesos)	708	5,866.301	2,343.246	1,469	4,088.5	7,196.5	17,000
TM	708	1,105.774	911.801	33.900	424.750	1,555.100	5,474.200
Vías	708	313.322	278.743	4.140	115.498	425.392	2,447.130
Parques	708	108.400	90.034	0.000	44.300	143.328	630.848
1 Habitación	708	0.267	0.443	0	0	1	1
2 Habitaciones	708	0.333	0.472	0	0	1	1
3 o más habitaciones	708	0.400	0.490	0	0	1	1
1 baño	708	0.329	0.470	0	0	1	1
2 baños	708	0.493	0.500	0	0	1	1
3 o más baños	708	0.178	0.383	0	0	0	1
Preventa	708	0.458	0.499	0	0	1	1
Construcción	708	0.199	0.400	0	0	0	1
Terminado	708	0.343	0.475	0	0	1	1
Obra negra	708	0.004	0.065	0	0	0	1
Acabados	708	0.883	0.322	0	1	1	1
Obra gris	708	0.113	0.317	0	0	0	1
Estrato 2	708	0.086	0.281	0	0	0	1
Estrato 3	708	0.322	0.468	0	0	1	1
Estrato 4	708	0.268	0.443	0	0	1	1
Estrato 5	708	0.185	0.389	0	0	0	1
Estrato 6	708	0.138	0.346	0	0	0	1

En términos de los estratos, se observa una mayor preponderancia de los estratos 3 y 4 con ausencia del estrato 1 en la muestra actual. Basados en las características mencionadas previamente, se concluye que en su mayoría los proyectos corresponden a vivienda con precios que superan el tope de los 150 salarios mínimos. Lo anterior no significa que el mercado inmobiliario bogotano se concentre en este tipo de viviendas, sino que, al analizar la oferta, este tipo de inmuebles son los que mayor peso tienen dado que los VIS² se comercializan más rápido y los saldos no se mantienen por mucho tiempo.

Ahora bien, al analizar el comportamiento de los precios por metro cuadrado de los proyectos en la capital, a partir de la Figura 1, se evidencia una posible dependencia espacial entre los proyectos, ya que hacia la periferia de la ciudad el precio es más bajo mientras que en las localidades de Chapinero y Usaquén, éste tiende a ser más alto, principalmente en el barrio Chicó el cual se ubica en la zona nororiental de la ciudad.

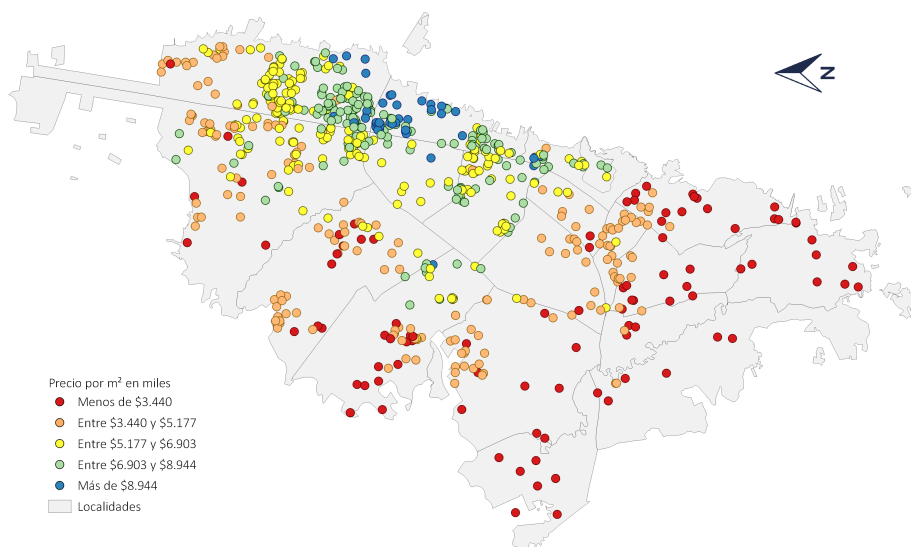


Figura 1: Precios por metro cuadrado de los proyectos con oferta en Bogotá.

²Se define la vivienda VIS como aquella que tiene un precio mayor a 150 salarios mínimos mensuales legales vigentes

Teniendo en cuenta que la oferta se concentra principalmente en unidades con menos de 3 habitaciones y baños, estas variables se categorizaron para trabajarlas como factores, donde se tendrán 3 niveles con los siguientes valores: 1 habitación/baño (base), 2 habitaciones/baños y 3 o más habitaciones/baños.

De este modo, se propone estimar 4 modelos (1 lineal y 3 semiparamétricos), los cuales tienen las siguientes especificaciones y permitirán evidenciar el aumento en la capacidad predictiva que se obtiene al incorporar los componentes no paramétricos y las funciones de suavizado dentro de la estimación de los precios de vivienda nueva:

Modelo - 1 SAR:

$$\ln(\text{Precio por } m_i^2) = \rho \sum_{j=1}^n w_{ij} \ln(\text{Precio por } m_i^2) + \beta_0 + \beta_1(2 \text{ habitaciones}_i) + \beta_2(3 \text{ o más habitaciones}) + \beta_3(2 \text{ baños}_i) + \beta_4(3 \text{ o más baños}_i) + \beta_5(\text{Área}_i) + \beta_6(\text{Parques}_i) + \beta_7(\text{Vías}_i) + \beta_8(\text{obra gris}_i) + \beta_9(\text{Acabados}_i) + \beta_{10}(\text{construcción}_i) + \beta_{11}(\text{Terminado}_i) + \beta_{12}(\text{Estrato } 3_i) + \beta_{13}(\text{Estrato } 4_i) + \beta_{14}(\text{Estrato } 5_i) + \beta_{15}(\text{Estrato } 6_i) + \beta_{14}(\text{TM}_i) + \varepsilon_i \quad (3)$$

donde Precio por m_i^2 corresponde al precio por metro cuadrado del i -ésimo proyecto, ρ es un parámetro autorregresivo para estimar, w_{ij} son los elementos de la matriz de pesos espaciales, habitaciones hace referencia al número de habitaciones y al ser una variable categórica se tiene como base 1 habitación mientras que dentro del modelo se expresan 2 habitaciones y 3 o más habitaciones, baños que también es una variable categórica e indica la cantidad de baños dentro del inmueble siendo la base 1 baño y dejando en la ecuación 2 baños y 3 o más baños, Área corresponde al tamaño en metros cuadrados del inmueble, Parques corresponde la distancia al parque más cercano, Vías la distancia a la vía arterial más cercana, Obra gris y acabados son variables dummy para la condición de entrega de los inmuebles (obra negra es la base de comparación), Construcción y Terminado son también variables dicotómicas que indican el estado del proyecto (Preventa es la categoría base para la comparación), TM es la distancia del proyecto a

la estación de Transmilenio más cercana, Estrato es un factor que tiene 5 niveles donde el estrato 2 es la base de comparación $\varepsilon_i \sim N(0, \sigma^2)$ para $i = 1, \dots, 715$.

Modelo 2 - PS-SAR con términos no paramétricos unidimensionales:

$$\begin{aligned} \ln(\text{Precio por } m_i^2) = & \rho \sum_{j=1}^n w_{ij} \ln(\text{Precio por } m_j^2) + \beta_0 + \beta_1 (2 \text{ habitaciones}_i) + \\ & \beta_2 (3 \text{ o más habitaciones}) + \beta_3 (2 \text{ baños}_i) + \beta_4 (3 \text{ o más baños}_i) + \\ & f_1(\text{Área}_i) + f_2(\text{Parques}_i) + f_3(\text{Vías}_i) + \beta_5 \text{obra gris}_i + \\ & \beta_6 \text{Acabados}_i + \beta_7 \text{construcción}_i + \beta_8 \text{Terminado}_i + \beta_9 \text{Estrato } 3_i + \\ & \beta_{10} \text{Estrato } 4_i + \beta_{11} \text{Estrato } 5_i + \beta_{12} \text{Estrato } 6_i + f_4(\text{TM}_i) + e_i \quad (4) \end{aligned}$$

donde a diferencia de (3), se incorpora el componente no paramétrico mediante una función del área $f_1(\text{rea})$, una función de la distancia al parque más cercano $f_2(\text{Parques})$, una función de la distancia a la vía arterial más cercana $f_3(\text{Vas})$, una función que depende de la distancia de la vivienda a la estación de Transmilenio más cercana $f_4(\text{TM})$ y $e_i \sim N(0, \sigma^2)$ para $i = 1, \dots, 715$.

Modelo 3 - PS-SAR 2 (con función bidimensional):

$$\begin{aligned} \ln(\text{Precio por } m_i^2) = & \rho \sum_{j=1}^n w_{ij} \ln(\text{Precio por } m_j^2) + \beta_0 + \beta_1 (2 \text{ habitaciones}_i) + \\ & \beta_2 (3 \text{ o más habitaciones}) + \beta_3 (2 \text{ baños}_i) + \beta_4 (3 \text{ o más baños}_i) + \\ & \beta_5 (\text{Área}_i) + \beta_6 \text{Parques}_i + \beta_7 \text{Vías}_i + \beta_8 \text{obra gris}_i + \beta_9 \text{Acabados}_i + \\ & \beta_{10} \text{construcción}_i + \beta_{11} \text{Terminado}_i + \beta_{12} \text{Estrato } 3_i + \beta_{13} \text{Estrato } 4_i + \\ & \beta_{14} \text{Estrato } 5_i + \beta_{15} \text{Estrato } 6_i + \beta_{16} \text{TM}_i + f(x_{\text{long}}, x_{\text{lat}}) + \epsilon_i \quad (5) \end{aligned}$$

este modelo sigue el mismo planteamiento explicado en (3), aunque incorpora una función bidimensional $f(x_{\text{long}}, x_{\text{lat}})$ que depende de las coordenadas geográficas y $\epsilon_i \sim N(0, \sigma^2)$ para $i = 1, \dots, 715$.

Modelo 4 - PS-SAR 3 (con función bidimensional y términos no paramétricos univariados):

$$\ln(\text{Precio por } m_i^2) = \rho \sum_{j=1}^n w_{ij} \ln(\text{Precio por } m_i^2) + \beta_0 + \beta_1 (2 \text{ habitaciones}_i) + \beta_2 (3 \text{ o más habitaciones}) + \beta_3 (2 \text{ baños}_i) + \beta_4 (3 \text{ o más baños}_i) + f_1 (\text{Área}_i) + f_2 (\text{Parques}_i) + f_3 (\text{Vías}_i) + \beta_5 \text{obra gris}_i + \beta_6 \text{Acabados}_i + \beta_7 \text{construcción}_i + \beta_8 \text{Terminado}_i + \beta_9 \text{Estrato 3}_i + \beta_{10} \text{Estrato 4}_i + \beta_{11} \text{Estrato 5}_i + \beta_{12} \text{Estrato 6}_i + f_4 (TM_i) + f(x_{\text{long}}, x_{\text{lat}}) + \vartheta_i \quad (6)$$

este modelo se concibe como una combinación de (4) y (5) debido a que toma las funciones no paramétricas de (4) y la función bidimensional que depende de las coordenadas de (5). Además, ϑ_i es un error que se distribuye normal con media cero y varianza σ^2 con $i = 1, \dots, 715$.

Como es habitual en este tipo de modelos se utiliza una transformación logarítmica para que los estimadores puedan ser interpretados como semielasticidades. Además, este tipo de transformaciones ayuda a que los datos se ajusten a una distribución normal, principalmente cuando se supone que la distribución de la variable dependiente corresponde a una log-normal [32].

Por su parte, para incorporar la dependencia espacial dentro del modelo se utilizó una matriz de distancia inversa al cuadrado $\left(w_{ij} = \frac{1}{d_{ij}^2}\right)$, la cual se estandarizó para poder tener en cuenta el impacto que generan las viviendas que se encuentran alrededor en el precio de cada unidad habitacional. En otras palabras, el precio por metro cuadrado de una vivienda dependerá de los precios por metro cuadrado a los que se comercialicen los inmuebles cercanos.

Las estimaciones y el desarrollo computacional se llevaron a cabo en el software R Core Team [33] usando las librerías `spdep` [34] y `sptpsar` [35] y el código básico puede ser solicitado por correo electrónico a los autores.

Antes de realizar la estimación de los modelos, se calcula la I de Moran como una medida para establecer la presencia o no de correlación espacial global en el logaritmo de los precios por metro cuadrado. Se obtiene un valor

de 0.6755, el cual resulta estadísticamente significativo y que es consistente con lo evidenciado en el Figura 2. De esta manera, al igual que [21] y [16], se concluye que existe una dependencia espacial positiva, es decir, que viviendas de precios por m^2 altos estarán rodeadas de otros inmuebles con precios altos, mientras que viviendas de precios por m^2 bajos estarán rodeadas de otras con precios bajos.

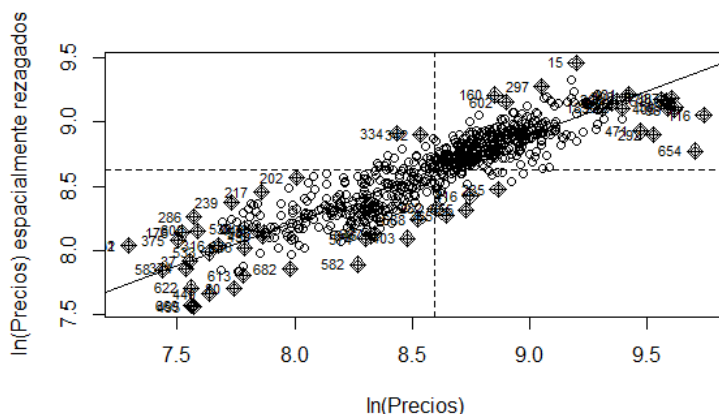


Figura 2: Plot de Moran para $\log(\text{precio}/m^2)$.

Los resultados para los términos paramétricos y las estadísticas de predicción de los modelos se presentan en la Tabla 3, donde se aprecia que la inclusión de funciones no paramétricas unidimensionales de las variables explicativas y bidimensionales de las coordenadas mejora considerablemente la capacidad predictiva en términos del error cuadrático medio (ECM) y el R^2 .

En general los modelos muestran unos resultados similares en cuanto a la magnitud y signo de los estimadores; sin embargo, cuando se incluyen los componentes no paramétricos hay unos ligeros cambios en algunos de los coeficientes, principalmente el asociado a la distancia a parques que es positivo en el modelo SAR y negativo en el PS-SAR 2.

Tabla 3: Resultados de los modelos propuestos

Variable	SAR (3)	PS-SAR (4)	PS-SAR2 (5)	PS-SAR3 (6)
Área	0.001*** (0.0002)		0.0009*** (0.0002)	
TM	-0.00003*** (0.00001)		-0.00002 (0.000019)	
Vías	0.00002 (0.00002)		-0.00005* (0.000028)	
Parques	0.00003 (0.0001)		-0.00005 (0.00007)	
acabados	0.277** (0.095)	0.265*** (0.095)	0.226* (0.088)	0.214** (0.088)
obra gris	0.130 (0.099)	0.084 (0.099)	0.098 (0.092)	0.063 (0.093)
Estrato 3	0.201*** (0.030)	0.164*** (0.031)	0.127*** (0.036)	0.121*** (0.037)
Estrato 4	0.313*** (0.034)	0.291*** (0.033)	0.207*** (0.040)	0.205*** (0.041)
Estrato 5	0.380*** (0.038)	0.371*** (0.035)	0.306*** (0.044)	0.286*** (0.046)
Estrato 6	0.474*** (0.042)	0.457*** (0.038)	0.394*** (0.051)	0.366*** (0.052)
Dos habitaciones	-0.081*** (0.019)	-0.050* (0.022)	-0.045** (0.017)	-0.029 (0.020)
Tres o más habitaciones	-0.111*** (0.023)	-0.093*** (0.027)	-0.060*** (0.022)	-0.040 (0.026)
Dos baños	-0.016 (0.017)	0.002 (0.019)	-0.021 (0.017)	-0.015
Tres o más baños	0.017 (0.029)	0.025 (0.033)	-0.015 (0.027)	-0.003 (0.030)
Construcción	0.014 (0.017)	0.018 (0.018)	0.021 (0.016)	0.023 (0.016)
Terminado	-0.083*** (0.014)	-0.073*** (0.015)	-0.066*** (0.014)	-0.060*** (0.014)
Constante	2.627*** (0.252)	3.095*** (0.126)		
ρ	0.637*** (0.027)	0.601*** (0.037)	0.135* (0.071)	0.111 (0.074)
R^2	0.846	0.851	0.874	0.8793
$ECM(\ln(P))$	0.0264	0.0234	0.0197	0.0177
AIC	-455.42	-1588.69	-1794.89	-1607.15

Nota: *p<0.1; **p<0.05; ***p<0.01

Una situación similar sucede al analizar los términos no paramétricos, donde los resultados son similares entre PS-SAR y PS-SAR 3 (ver Figuras 3 y 4). Aunque en este último modelo las no linealidades presentes en las distancias a vías, estaciones de Transmilenio y a parques no son tan marcadas gracias a la inclusión de la función bidimensional de las coordenadas geográficas, tal como se evidencia en el Figura 4.

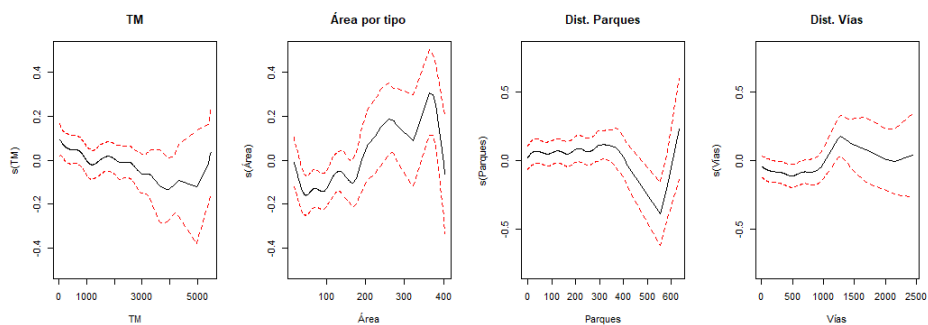


Figura 3: Términos no paramétricos para PS-SAR.

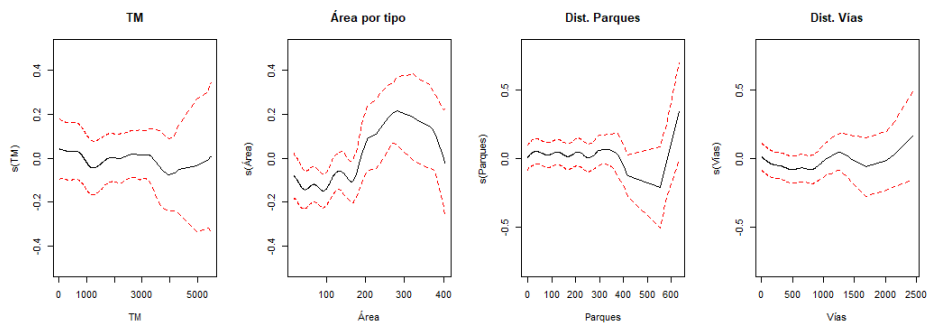


Figura 4: Términos no paramétricos para PS-SAR3.

Por otro lado, la tendencia espacial presentada en la Figura 5 explica una buena parte de la variabilidad del precio por metro cuadrado de la vivienda nueva en Bogotá, donde el pico más alto de la función se obtiene en la parte nororiental de la ciudad, específicamente en el barrio Chicó cerca a los cerros donde hay una gran concentración de oficinas y comercio,

tal como se mostró en la Figura 1. Esta evidencia va en línea con lo encontrado por [13] en Buenos Aires donde los inmuebles que evidenciaron autocorrelación positiva se ubican en una mixtura de zonas comercial y residenciales consolidadas.

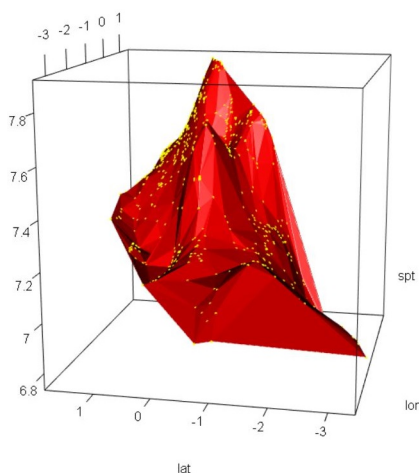


Figura 5: Tendencia espacial.

Se selecciona el modelo PS-SAR3³ (6) por su bajo error cuadrático medio y su elevado R^2 . Los principales resultados que arroja el modelo son que las viviendas en Bogotá que se entregan con acabados tienen un precio por metro cuadrado 21 % más alto a aquellas que se entregan en obra negra. Adicionalmente, los precios por metro cuadrado aumentan a medida que se incrementa el estrato, es decir que una vivienda tiene un precio superior en 12 %, 20 %, 29 % y 37 % si se ubica estrato 3, 4, 5 y 6, respectivamente frente a una vivienda de estrato 2.

El número de alcobas y de baños evidencia una relación negativa frente al precio de la vivienda, aunque, su resultado no es significativo. Este resultado difiere a lo encontrado por [21] y [17] donde la relación es positiva y significativa, pero va en línea con lo encontrado por [13] donde los resultados para una ciudad de Argentina tienen comportamiento

³Sin embargo, en términos del AIC el modelo PS-SAR2 puede llegar a ser una buena opción también, aunque ofrecerá unas predicciones menos precisas respecto a PS-SAR 3.

negativo y no significativo en la estimación SAR, SEM y SAC. Sin embargo, es importante entender el cambio en los patrones sociodemográfico que atraviesa la ciudad de Bogotá donde los hogares tienen una preferencia mayor en el área y la localización frente a otros atributos tal como lo demuestra el estudio de oferta y demanda de vivienda y destinos no habitacionales de Camacol [8].

En esta línea se encuentran los resultados obtenidos en el área de la vivienda. Volviendo a figura 4 es posible evidenciar el comportamiento del precio de la vivienda con respecto al área donde las viviendas pequeñas tienen un precio elevado y a medida que aumenta el área el precio disminuye hasta 100 metros cuadrados, área donde tiene un punto de inflexión y su precio aumenta hasta los 300 metros cuadrados y allí se estabiliza. Este comportamiento se debe a que los apartaestudios, usualmente bien ubicados cerca a los centros de empleo y universidades, tienen un precio más alto por dicho motivo. Adicionalmente, la tendencia que toma el precio a medida que aumenta el área se debe a que las viviendas más grandes se van ubicando más lejos de los centros; no obstante, las viviendas de mayor metraje que tienen un precio más alto son las que se ubican sobre el borde nororiental de Bogotá y usualmente son estrato 6.

La distancia a vías principales tiene una relación en general positiva con respecto al precio por metro cuadrado, esto se explica debido a los factores como el ruido y la contaminación que son percibidos como negativos por los compradores. Por lo tanto, entre más alejada esté la vivienda de vías principales tendrá un precio más alto, sin embargo, esta distancia debe ser prudente ya que más de 1,5 km el precio disminuye, tal como se evidencia gráficamente en la Figura 4.

La distancia a Transmilenio muestra un comportamiento similar al que presenta el precio frente a la distancia a las vías, una tendencia negativa pero no muy pronunciada, es decir que a medida que se aleja de una estación el precio disminuye siendo entre 1 y 2,5 Km la distancia de mayor valor. Este comportamiento se asemeja al evidenciado por [21] donde el resultado al parámetro que acompaña la distancia al Transmilenio es negativo para cada uno de los modelos estimados sin embargo la potencia de la estimación semiparamétrica permite inferir el comportamiento del precio a diferentes distancias tal como se evidencia en la figura 4.

La distancia a parques no tiene efecto sobre el precio por metro cuadrado de las viviendas, en los primeros 300 m; sin embargo, después de dicho valor empieza a evidenciarse una reducción en el precio por m^2 y un posterior crecimiento a partir de los 550 m. Este hecho se debe a que las mayores distancias se dan sobre las viviendas ubicadas sobre los cerros donde tienen otro tipo de servicio como el que puede brindar estar cerca a los cerros orientales de la ciudad. Finalmente es posible evidenciar que la ubicación es un factor importante en la conformación del precio, pues es el aspecto de mayor relevancia entre los compradores a la hora de tomar la decisión de invertir en vivienda, tal como lo demuestra el estudio de oferta y demanda de vivienda y destinos no habitacionales de Camacol [8].

Las predicciones del modelo PS-SAR 3 tienen en promedio una diferencia de 909.000 COP, con respecto a los precios por metro cuadrado reales. En la Figura 6 se aprecia que las predicciones tienen un comportamiento muy similar al evidenciado en la figura 1 con los precios observados.

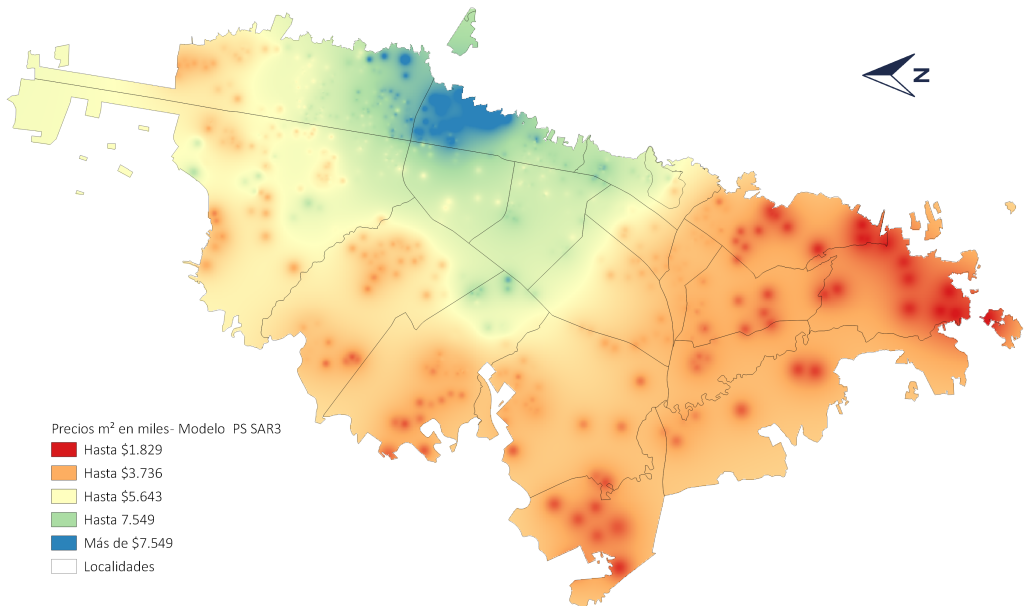


Figura 6: Predicciones del modelo PS-SAR 3.

Adicionalmente, el modelo SAR lineal y los semiparamétricos presentados previamente son útiles para tratar la dependencia espacial que naturalmente tienen los datos, tal como se observa en el Figura 7, donde no se aprecia ningún tipo de relación entre los residuos espacialmente rezagados y los residuos común y corrientes. Este resultado es muy importante y va en línea con lo encontrado por [36], donde la inclusión de $f(x_{long}, x_{lat})$ puede servir como un mecanismo para tener en cuenta la dependencia espacial sustantiva y la heterogeneidad espacial presente en los datos, la cual puede verse claramente en la tendencia de la Figura 5.

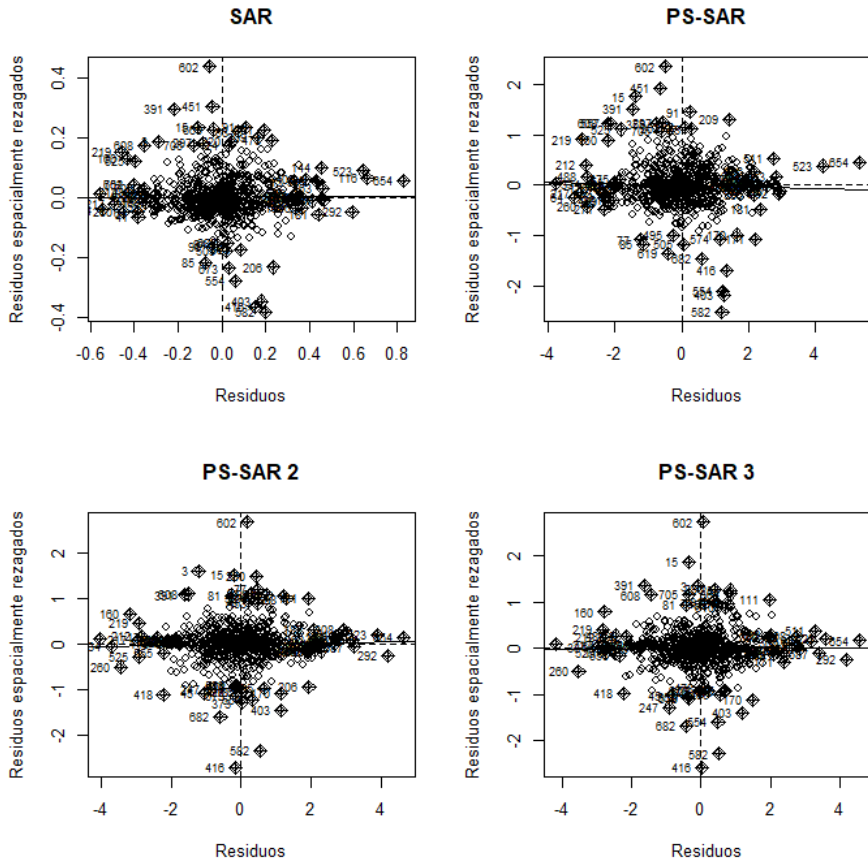


Figura 7: Plot de Moran para los residuos de los modelos.

5 Conclusiones y recomendaciones

Los modelos semiparamétricos permiten incluir relaciones más flexibles entre el predictor y la variable dependiente. En ese sentido, la investigación asociada a esta temática ha avanzado a pasos agigantados y se ha extendido a diferentes áreas del conocimiento donde ha encontrado un espacio para la aplicación en numerosos casos prácticos [37]. Asimismo, los modelos econométricos espaciales también han evolucionado permitiendo la incorporación de relaciones no lineales, lo cual mejora considerablemente la capacidad predictiva y permite incluso que se eliminen los efectos de dependencia espacial propios del modelo SAR habitual [7].

El modelo semiparamétrico de precios hedónicos propuesto para Bogotá en esta investigación da cuenta de las mejoras que se obtienen al utilizar estas metodologías respecto a otras más tradicionales como la planteada por [3]. Puntualmente estos métodos pueden ser utilizados fácilmente por agentes inmobiliarios (constructores, compradores, evaluadores e inversionistas) que piensen en comercializar, adquirir o entender el comportamiento del precio de las viviendas.

El modelo también permitió identificar los atributos que influyen en la determinación del precio de la vivienda nueva desde el lado de la oferta donde se destacan el estrato, la condición de entrega y estado constructivo con un efecto lineal sobre el precio y el área, junto con las distancias a parques, vías y estaciones de Transmilenio con efectos no lineales. Además, la función bidimensional permitió aislar el efecto de la ubicación sobre el precio por metro cuadrado de los inmuebles.

Una de las principales ventajas del método utilizado es que permite identificar las variables con efecto no lineales y estimar el impacto de estas sobre la variable exógena, por lo tanto, la metodología propuesta podrá ser utilizada por diversos actores para mejorar sus análisis, las políticas públicas y las decisiones de inversión en los diferentes mercados inmobiliarios del país.

Aunque los datos utilizados en esta investigación corresponden solo a la vivienda nueva, este tipo de modelos pueden utilizarse también para el mercado de lo usado, donde iniciativas como la planteada por [38] cobran gran relevancia porque permiten obtener una gran cantidad de observaciones y de variables adicionales que no se contemplaron

dentro de este estudio. Además, para casos donde se cuente con datos longitudinales o tipo panel, pueden incluirse tendencias no solo espaciales sino espacio-temporales.

En ese sentido, esta metodología puede ser de gran utilidad en la construcción de índices de precios para la vivienda nueva, siguiendo procedimientos análogos a los propuestos por [32] o los planteados por [22].

A pesar de la consistencia que evidencian los resultados, es importante mencionar que esta especificación no tiene en cuenta la posible endogenidad que se puede incorporar al introducir variables contextuales como las relacionadas con espacios abiertos (parques, plazas, vías, entre otras), variables ambientales (calidad de aire) o aspectos socioeconómicos (tasa de criminalidad, calidad de educación) [39]. Además, no se contempló la correlación que podría existir entre la distancia a vías, a estaciones de transmilenio y a parques con los errores del modelo. Para abordar esta situación se podría plantear una estrategia en dos etapas y el uso de variables instrumentales.

Por último, quedan abiertas varias posibilidades para seguir ampliando estos modelos en futuras investigaciones, permitiendo por ejemplo el modelamiento de la varianza (bajo heterocedasticidad) y el análisis de otras estructuras de tipo jerárquicas o multinivel como las propuestas por [40] donde, para el caso de los inmuebles, se contemplan dos estructuras de vecindad, la primera entre unidades habitacionales (desagregada) y la segunda a partir de las zonas (localidades o UPZ) en las cuales se encuentran dichas viviendas (agregada). Lo anterior podría permitir una mejor evaluación de las interacciones y el control de la heterogeneidad espacial en el mercado viviendista.

Agradecimientos

Los autores agradecen a CAMACOL Bogotá y Cundinamarca por la información de Coordinada Urbana.

Referencias

- [1] L. Anselin and N. Lozano-Gracia, “Spatial hedonic models,” in *Palgrave handbook of econometrics*. Springer, 2009, pp. 1213–1250. https://doi.org/10.1057/9780230244405_26 25, 26
- [2] D. Desormeaux and F. Piguillem, “Precios hedónicos e índices de precios de viviendas,” *Documento de trabajo*, no. 12, 2003. 25, 27
- [3] L. Anselin, *Spatial econometrics: methods and models*. Springer Science & Business Media, 1988, vol. 4. 25, 31, 47
- [4] R. M. Serrano and E. V. Valcarce, *Técnicas econométricas para el tratamiento de datos espaciales: la econometría espacial*. Edicions Universitat Barcelona, 2000, vol. 44. 25
- [5] C. Chasco Yrigoyen, “Modelos de heterogeneidad espacial,” University Library of Munich, Germany, Tech. Rep., 2004. https://www.researchgate.net/publication/23742953_Modelos_de_heterogeneidad_espacial 25
- [6] J. Montero, R. Mínguez, and M. Durbán, “Sar models with nonparametric spatial trends. a p-spline approach,” *Estadística Española*, vol. 54, no. 177, pp. 89–111, 2012. 26, 31, 33
- [7] R. Basile and R. Mínguez, “Advances in spatial econometrics: Parametric vs. semiparametric spatial autoregressive models,” *The economy as a complex spatial system*, pp. 81–106, 2018. https://doi.org/10.1007/978-3-319-65627-4_4 26, 31, 47
- [8] CAMACOL, “Estudio de oferta y demanda de vivienda y no habitacionales en Bogotá y Cundinamarca,” Tech. Rep., 2019. <https://ww2.camacolcundinamarca.co/documentos/EstudiosEspecificos/LAE-CAMACOL2019.pdf> 26, 44, 45
- [9] K. J. Lancaster, “A new approach to consumer theory,” *Journal of political economy*, vol. 74, no. 2, pp. 132–157, 1966. <https://doi.org/10.1086/259131> 27
- [10] S. Rosen, “Hedonic prices and implicit markets: product differentiation in pure competition,” *Journal of political economy*, vol. 82, no. 1, pp. 34–55, 1974. <https://doi.org/10.1086/260169> 27
- [11] L. Anselin, “Spatial regression,” *The SAGE handbook of spatial analysis*, vol. 1, pp. 255–276, 2009. 27
- [12] A. Sagner, “Determinantes del precio de viviendas en la región metropolitana de Chile,” *El trimestre económico*, vol. 78, no. 312, pp. 813–839, 2011. <http://www.scielo.org.mx/pdf/ete/v78n312/2448-718X-ete-78-312-00813.pdf> 27

- [13] S. León, “Precios inmobiliarios en caba: un análisis econométrico espacial, desde lo global a lo local,” 2016. https://www.researchgate.net/publication/311641420_Precios_inmobiliarios_en_CABA_un_analisis_econometrico_espacial_desde_lo_global_a_lo_local 27, 43
- [14] J. I. Zoloa, “Tópicos en el mercado inmobiliario: un análisis desde la econometría espacial,” 2015. <https://doi.org/10.35537/10915/49057> 28
- [15] L. Galvis, B. Carrillo *et al.*, “índice de precios espacial para la vivienda urbana en Colombia: una aplicación con métodos de emparejamiento,” *Revista de Economía del Rosario*, vol. 16, no. 1, pp. 25–29, 2013. <https://revistas.urosario.edu.co/index.php/economia/article/view/2829> 28
- [16] M. A. Urrea Zora, N. Cárdenas Mazo *et al.*, “Un modelo espacial de precios hedónicos y calidad de vecindario: una aplicación para Medellín, Colombia,” B.S. thesis, Universidad EAFIT, 2019. <https://repository.eafit.edu.co/handle/10784/14941> 28, 40
- [17] J. C. Duque, H. V. Ceballos, and J. Agudelo, “Infraestructura pública y precios de vivienda: una aplicación de regresión geográficamente ponderada en el contexto de precios hedónicos,” *Ecos de Economía: A Latin American Journal of Applied Economics*, vol. 15, no. 33, pp. 99–122, 2011. 28, 43
- [18] M. Rodríguez Gómez, “Caracterización hedónica del mercado de finca raíz en la ciudad de Cartagena: una aproximación semiparamétrica,” *Economía y Región*, vol. 1, no. 1, 2004. <https://revistas.utb.edu.co/index.php/economiaayregion/article/view/265> 29
- [19] C. Melo and O. Melo, “Estimación de precios hedónicos para propiedades residencial y comercial en la ciudad de Bogotá,” *Ingeniería*, vol. 8, no. 1, pp. 10–18, 2003. 29
- [20] J. C. Mendieta, J. A. Perdomo *et al.*, *Especificación y estimación de un modelo de precios hedónico espacial para evaluar el impacto de Transmilenio sobre el valor de la propiedad en Bogotá*. Univ. de los Andes, Fac. de Economía, CEDE Centro de Estudios sobre Desarrollo Económico, 2007. <http://hdl.handle.net/1992/8115> 29
- [21] S. Higuera, “Modelación hedónico espacial de los determinantes del precio de la vivienda nueva en Bogotá,” Universidad Nacional de Colombia, Colombia, Tech. Rep., 2010. 30, 40, 43, 44
- [22] W. A. Cabrera-Rodríguez, J. S. Mariño-Montaña, and C. A. Quicazán-Moreno, “Modelos hedónicos con efectos espaciales: una aproximación al cálculo de índices de precios de vivienda para Bogotá,” *Borradores de Economía*, no. 1072, 2019. <https://doi.org/10.32468/be.1072> 30, 48

- [23] R. Mínguez, M. L. Durbán, and R. Basile, “Spatio-temporal autoregressive semiparametric model for the analysis of regional economic data,” *Stat. Methods Appl*, vol. 16126, 2019. 31, 34
- [24] R. Mínguez, R. Basile, and M. Durbán, “An alternative semiparametric model for spatial panel data,” *Statistical Methods & Applications*, vol. 29, no. 4, pp. 669–708, 2020. <https://doi.org/10.1007/s10260-019-00492-8> 31, 33
- [25] J.-M. Montero, R. Mínguez, and G. Fernández-Avilés, “Housing price prediction: parametric versus semi-parametric spatial hedonic models,” *Journal of Geographical Systems*, vol. 20, no. 1, pp. 27–55, 2018. <https://doi.org/10.1007/s10109-017-0257-y> 31, 32
- [26] B. Manganeli, P. Pontrandolfi, A. Azzato, and B. Murgante, “Using geographically weighted regression for housing market segmentation,” *International Journal of Business Intelligence and Data Mining 13*, vol. 9, no. 2, pp. 161–177, 2014. <https://doi.org/10.1504/IJBIDM.2014.065100> 32
- [27] G. Simpson, “Package gratia,” *R package version*, 2021. 32
- [28] S. Wood, “Package mgcv,” *R package version*, 2020. 32
- [29] S. N. Wood, *Generalized additive models: an introduction with R*. CRC press, 2017. 32
- [30] P. H. Eilers, I. D. Currie, and M. Durbán, “Fast and compact smoothing on large multidimensional grids,” *Computational Statistics & Data Analysis*, vol. 50, no. 1, pp. 61–76, 2006. <https://doi.org/10.1016/j.csda.2004.07.008> 33
- [31] M. X. Rodríguez-Álvarez, D.-J. Lee, T. Kneib, M. Durbán, and P. Eilers, “Fast smoothing parameter separation in multidimensional generalized p-splines: the sap algorithm,” *Statistics and Computing*, vol. 25, no. 5, pp. 941–957, 2015. <https://doi.org/10.1007/s11222-014-9464-2> 34
- [32] J. Castaño, M. Laverde, M. Morales, and A. Yaruro, “Índice de precios de la vivienda nueva para Bogotá: metodología de precios hedónicos,” *Temas de Estabilidad financiera*, vol. 78, 2013. 39, 48
- [33] R. C. Team *et al.*, “R: A language and environment for statistical computing,” 2020. 39
- [34] R. Bivand, M. Altman, L. Anselin, R. Assunção, O. Berke, A. Bernat, and G. Blanchet, “Package spdep,” *The Comprehensive R Archive Network*, 2020. <https://cran.r-project.org/web/packages/spdep/spdep.pdf> 39

- [35] R. Mínguez, R. Basile, and M. Durbán, “sptpsar: an R package for the development of semiparametric spatial and spatio-temporal econometric models,” Tech. Rep. 39
- [36] R. Basile, M. Durbán, R. Mínguez, J. M. Montero, and J. Mur, “Modeling regional economic dynamics: Spatial dependence, spatial heterogeneity and nonlinearities,” *Journal of Economic Dynamics and Control*, vol. 48, pp. 229–245, 2014. <https://doi.org/10.1016/j.jedc.2014.06.011> 46
- [37] J. Harezlak, D. Ruppert, and M. P. Wand, *Semiparametric regression with R*. Springer, 2018. <https://link.springer.com/book/10.1007/978-1-4939-8853-2> 47
- [38] J. A. C. Rubio, F. J. C. Guzmán, and J. Otero, “Una base de datos de precios y características de vivienda en Colombia con información de internet,” *Revista de economía del Rosario*, vol. 22, no. 1, pp. 75–100, 2019. <https://doi.org/10.12804/revistas.urosario.edu.co/economia/a.7768> 47
- [39] L. Wang, “Spatial econometric issues in hedonic property value models: model choice and endogenous land use,” Tech. Rep., 2006. https://etda.libraries.psu.edu/files/final_submissions/3844 48
- [40] R. Cellmer, K. Kobylińska, and M. Belej, “Application of hierarchical spatial autoregressive models to develop land value maps in urbanized areas,” *ISPRS International Journal of Geo-Information*, vol. 8, no. 4, p. 195, 2019. <https://doi.org/10.3390/ijgi8040195> 48