

**INCIDENCIA DE LA ESTACIÓN DEL METRO EN SAN JAVIER EN  
LOS PRECIOS DE LAS VIVIENDAS DE LA ZONA**

**Tesis para optar al título de Magister en Economía**

**Por:  
Jorge Enrique Agudelo Torres**

**Tutores:  
Juan Carlos Duque Cardona  
Ermilson Velásquez Ceballos**

**2011  
Universidad EAFIT**

## **Agradecimientos**

En este breve espacio quiero agradecer a mis asesores Dr. Juan Carlos Duque y Dr. Ermilson Velásquez por la valiosa ayuda que me proporcionaron en el desarrollo de este trabajo, de igual manera al Lincoln Institute of Land Policy que patrocinó el presente trabajo con una de las cinco becas que ofrecen para Latinoamérica.

## Resumen

El análisis de las externalidades en el sector inmobiliario ha atraído desde hace varios años la atención de los investigadores suscitando una gran cantidad de estudios al respecto. En este artículo se examina la influencia que tiene en los precios de las viviendas la existencia de una estación del metro en San Javier ubicada en el centro occidente de la ciudad de Medellín, empleando la técnica de regresión geográficamente ponderada. Este trabajo muestra que la cercanía a la estación del metro en San Javier, tiene una influencia positiva en los precios de las viviendas de la zona que se encuentra determinada por los accidentes geográficos del terreno.

**Palabras Clave** Sector Inmobiliario · GWR · Regresión Geográficamente Ponderada · Metro de Medellín

**JEL Classification:** C21; O18; R32

## **Incidencia de la estación del metro en San Javier en los precios de las viviendas de la zona.**

### **1. INTRODUCCIÓN**

El comportamiento del sector inmobiliario ha sido estudiado desde hace más de sesenta años. En el siglo pasado, los trabajos pioneros sobre propiedad raíz estaban enfocados a explicar que la propiedad raíz era una buena inversión dadas las rentas que producía mensualmente y el crecimiento de precio de los inmuebles (Dewey y DeTuro, 1950; Thomson, 1950; Beaty, 1952). Posteriormente en los primeros años de la década del setenta, la atención se centró en los modelos de precios hedónicos que reflejaban la interacción entre la oferta y la demanda del mercado, a través de especificaciones en las que se establecían relaciones funcionales entre las características de los bienes, y los precios de venta de los mismos (Rosen, 1974). En las décadas siguientes, entre los años ochenta y noventa, algunos autores consideraron como factores determinantes en la explicación del precio de la vivienda de factores externos como el vecindario, la cercanía a algún lugar importante o una vista agradable, información que se introducía en la especificación de un modelo econométrico tradicional mediante variables dummy para indicar la existencia externalidades que afectan el precio de las viviendas (French y Lafferty, 1984; Dubin y Goodman ,1982; Portney, 1981; Quang Do y Grudnitski, 1995; Carroll, Claretie y Jensen, 1996).

En los modelos de precios hedónicos, se consideran como determinantes del precio de la vivienda las características de la misma. Esto hace particularmente importantes los problemas de especificación del modelo relacionados con la omisión de variables relevantes o los ligados a la fuerte correlación entre las mismas. Por tal motivo autores como Can (1992) agruparon las especificaciones de las viviendas en ítems como características estructurales, características del vecindario y fecha en la que se recogieron los datos. Otros autores como Basu y Thibodeau (1998) discriminaron las características de las viviendas en siete ítems: las asociadas al lote, a las mejoras, la accesibilidad, la proximidad de externalidades, uso de la tierra y época en la que se recolectan los datos. Todo esto con el fin de evitar sesgos de especificación o de omisión de variables relevantes.

$$\text{Precio} = f(\text{Características estructurales}, \text{Características del vecindario}, \text{Fecha})$$

La metodología econométrica tradicional utilizada comúnmente no tiene en cuenta la correlación espacial de los datos, es decir, se omite el espacio como fuente de variabilidad de los datos. Y en segundo lugar, no tiene en cuenta la heterogeneidad espacial de los datos, lo que implica que los valores de los parámetros pueden variar en el espacio (Anselin, 1988). Estos aspectos llegan a representar un gran problema debido a que la econometría arroja parámetros que son promedios obtenidos a partir de los datos y no reflejan de modo alguno la variabilidad o la asociación espacial de los datos empleados.

Paralelamente a este proceso, en el campo de la ciencia regional se venían desarrollando nuevos métodos de análisis espacial, que permitiesen manejar el incremento en la disponibilidad de datos georreferenciados y a la necesidad de tener en cuenta aspectos relacionados con las heterogeneidades intraurbanas. Una de las técnicas más recientes en este contexto es la de la Regresión Geográficamente Ponderada, GWR por sus siglas en inglés, la cual fue desarrollada por Stewart Fotheringham, Chris Brunsdon y Martin Charlton (2002).

Buena parte de estos aportes se han desarrollado en países en donde la información de precios de vivienda es de acceso público y su uso no presenta restricciones, como es el caso de Estados Unidos, Japón y algunos países de la Unión Europea. En Colombia dicha información es privada, debido a problemas de seguridad existentes desde hace varias décadas, lo cual ha generado un escaso volumen de trabajos académicos sobre el tema, la mayoría de los cuales han estado asociados a precios de oferta de inmuebles más que a valores reales de venta, así como a unas cuantas características del inmueble, lo que tiende a generar sesgos en los coeficientes estimados.

El presente trabajo es pues, la primera aplicación en el país sobre modelos de precios hedónicos que utiliza GWR a precios de vivienda. Concretamente, en este trabajo se busca cuantificar la existencia de un mayor valor en los precios de las viviendas del barrio San Javier en virtud de su cercanía a la estación del metro, lo que permitirá delimitar la influencia de la misma en los valores de las viviendas. La importancia de este trabajo radica en que permite observar como una externalidad positiva generada por una intervención pública, benefició la comunidad de una zona deprimida de la ciudad y con altos índices históricos de violencia y pobreza, a través del crecimiento de su patrimonio personal generado por el aumento del precio de sus viviendas.

En tal sentido, este trabajo permitirá cuantificar el mayor precio que tiene una vivienda por su cercanía a una estación del metro, lo que implícitamente involucra la valoración una mejora en el bienestar de la sociedad creada por un aumento en el precio de sus activos, que a su vez es generada por la intervención estatal en una zona con altos niveles de pobreza y violencia en la ciudad. Adicionalmente, este estudio ayudará a demostrar que la ejecución de obras públicas puede generar un mayor precio de los inmuebles, que puede distribuirse de forma no concéntrica, irregular y discontinua en el espacio. Lo que implica que los mismos pueden ser sujetos de cobros por concepto de valorización<sup>1</sup> por la ejecución de una obra pública y por último este trabajo demostrará la necesidad de que se avance en el país en el uso de las técnicas de análisis espacial para el tratamiento de datos que por su naturaleza así lo requieren, pues las técnicas que se emplean actualmente distan mucho de lo deseable.

El resto del artículo está estructurado como sigue: La sección 2 contiene una breve explicación de la metodología GWR y provee algunas aplicaciones en el contexto de precios de vivienda. En la sección 3 se identifica el caso de estudio, se especifica el modelo y se presentan los resultados. En la sección final se presentan las conclusiones del estudio.

---

<sup>1</sup> Según la Ley de Metros la contribución por valorización en este tipo de obra no puede ser menor al diez por ciento del valor total del proyecto.

## 2. FUNDAMENTACIÓN ECONÓMICA

Desde una perspectiva económica, los modelos de precios hedónicos que son aquellos en los cuales el precio de un bien está determinado por sus características, comenzaron a gestarse a mediados del siglo pasado en la década del sesenta, cuando Lancaster (1966) propuso lo que llamó una “nueva aproximación a la teoría del consumidor”, en la cual consideró que la utilidad de los bienes era derivada de sus características, rompiendo con la idea de que venía dada implícitamente en ellos. Este enfoque contó con tres ideas principales: En primer lugar asumió que cada bien posee un conjunto de características que son las que dan la utilidad al mismo. En segundo lugar, definió que un bien posee más de una característica y muchas características son compartidas por más de un bien. Por último, los bienes en combinación debían tener diferentes características que perteneciesen a los bienes separadamente. De esta manera comenzó a gestarse una nueva aproximación a la teoría del consumidor a través del estudio de los modelos de precios hedónicos.

El problema consiste en que los consumidores derivan una utilidad del consumo de un bien que tiene un vector de  $Z$  de  $J$  características diferentes, más el consumo de un bien compuesto  $Y$ . Los consumidores tienen un ingreso fijo de  $M$  y enfrentan una función de precios  $P(Z)$  que les da el precio del bien heterogéneo como una función de las características  $Z$  del bien y un vector de parámetros  $\alpha$ . Las preferencias son representadas por la función de utilidad:

$$u = u(Z, Y, \alpha)$$

En donde el tercer término del paréntesis corresponde a un vector de parámetros observados y no observados que caracteriza las preferencias de las familias. De la función de utilidad puede derivarse el monto que la familia estaría dispuesta a pagar por un bien como una función de sus características, dados el ingreso y el nivel de utilidad de la familia. Esta familia tiene una función de renta de alquiler definida por  $\beta(Z, M, u, \alpha)$ , entonces la función de utilidad es:

$$u = u(Z, M - \beta, \alpha)$$

La derivada de esta función da la tasa a la cual la familia podría permitirse gastar en un bien cuya característica  $i$  se incrementa, mientras se mantiene constante el nivel de utilidad. Esta es la inversa de la curva de demanda compensada.

La familia escoge un bien con características  $Z$ , y consumo de un bien compuesto  $Y$ , para resolver:

$$\max_{Z, Y} u(Z, Y, \alpha)$$

$$S.A. \_ M \geq P(Z) + Y$$

Aquí la restricción implica que el ingreso de las familias se distribuye entre la compra de un bien compuesto  $Y$  y bienes heterogéneos representados por la función de precios  $P(Z)$ . Las condiciones de primer orden son:

$$\frac{\partial u}{\partial Z_i} = \frac{\partial U(Z, M - \beta, \alpha)}{\partial Z} - \frac{P(z)\lambda}{\partial Z}$$

$$\frac{\partial u}{\partial Y} = \lambda$$

La derivada de  $P$  se refiere usualmente al precio hedónico de la característica  $i$ , y la función  $P(Z)$  es la función de precios hedónicos.

Combinando las condiciones de primer orden:

$$\frac{u_i}{u_y} = P_i$$

Con la diferenciación implícita se obtiene:

$$\frac{\partial \beta}{\partial Z_i} = P_i$$

Se obtiene el resultado de que la elección óptima de un bien está caracterizada por la igualdad entre el intercepto de la oferta de renta y el precio hedónico de cada característica:

$$\frac{\partial \beta}{\partial Z_i} = \frac{u_i}{u_y} = P_i$$

Esta condición de igualdad es parte de la justificación de la aproximación hedónica al análisis de mercados, para ello indicamos que si podemos observar o estimar el precio hedónico de una característica y una elección hecha por el consumidor, entonces bajo la asunción de conducta optimizadora, la observación provee información local acerca de las preferencias o la disposición a pagar por los atributos. En este sentido el problema parece similar al del análisis estándar de la conducta del consumidor, en el cual las elecciones observadas y los precios proveen información acerca de las preferencias de los consumidores y dan suficientes datos que se espera den inferencias acerca de la conducta del consumidor y la determinación de precios.

Años más tarde Rosen (1974) continuando con la línea de Lancaster (1966) definió como conjunto de precios hedónicos al vínculo existente entre los precios de los bienes y las características de los mismos, estableciendo relaciones funcionales entre ambos. A partir de esta teoría formuló el problema de mercado como un problema de equilibrio espacial en el que tanto los consumidores como los productores toman decisiones en el espacio respecto a un bien que cuenta con características estructurales y espaciales definidas. De esta manera comenzó la búsqueda de relaciones funcionales entre las características de los bienes y los precios.

Algunas décadas después, durante los años ochenta y noventa, los modelos de precios hedónicos empezaron a utilizarse frecuentemente en el sector inmobiliario, particularmente en el área de precios de inmuebles, siendo el trabajo de Can (1992) uno de los pioneros en la estandarización de las características de las viviendas al dividir las en dos grupos: uno asociado a

las características estructurales como tamaño del lote, de la construcción y edad de la vivienda y el otro asociado a las características del vecindario como uso del suelo y externalidades.

Sin embargo, la implementación de los modelos de precios hedónicos en el área de precios de vivienda, se desarrolló frecuentemente aplicando la metodología econométrica tradicional a datos de naturaleza espacial, lo que atrajo dos tipos de problemas (Anselin, 1988): El primero, asociado a la correlación espacial que se presenta cuando hay niveles de dependencia espacial entre las variables, lo cual es particularmente válido en el mercado inmobiliario en el que las viviendas de mayor y menor precio tienden a estar aglomeradas en lugares específicos. El segundo, relacionado con la heterogeneidad espacial, según la cual los parámetros y las formas funcionales varían de acuerdo con la localización de los datos, es decir, los coeficientes obtenidos en el modelo son inestables en el espacio. Este inconveniente es particularmente válido en caso del mercado inmobiliario, en el cual una característica de un inmueble, puede ser de vital importancia en una zona con nivel económico alto y no serlo en una zona de estrato bajo, como por ejemplo un parqueadero.

En este contexto de dependencia y heterogeneidad espacial de los datos del mercado de vivienda, surge la técnica GWR que permite tratar estas dificultades al consentir la existencia de coeficientes de regresión que varían en el espacio. En tal sentido, GWR permite construir modelos en los cuales se puede incorporar la heterogeneidad y la dependencia espacial presente en el mercado inmobiliario. Un ejemplo de ello son los vecindarios en los que existen mezclas de casas antiguas, de grandes áreas y precios elevados, con apartamentos nuevos, de áreas menores y precios relativamente más bajos. En este caso la metodología econométrica tradicional arroja coeficientes que son promedios de los datos, mientras que GWR arroja coeficientes ponderados de cada una de las características, entregando una mayor información de la realidad del mercado.

En el contexto de precios de vivienda, se han realizado pocos estudios en ciudades de países que cuentan con mayor flujo de información. Bitter, Mulligan y Dall'érba (2006) en la cual realizaron una comparación entre los métodos de expansión y GWR, mediante un análisis de más de once mil datos de ventas del mercado inmobiliario en Tucson. Para ello especificaron modelos en los que como variable dependiente fue empleado el precio de venta y como independientes el área construida, el área del lote, la calidad de la vivienda, la historia de la vivienda y la edad de la misma, con los cuales especificaron varios modelos empleando distintos tipos de técnicas. Los resultados mostraron que todas las variables empleadas eran significativas y que los precios marginales de los atributos de las viviendas eran variables en el espacio. Por tanto la técnica GWR fue superior al método de expansión en términos de poder explicativo y exactitud de la estimación.

Una aplicación similar fue desarrollada por Yu (2004) quién analizó mercado inmobiliario en Milwaukee utilizando la técnica GWR. Para ello empleó 3,437 datos seleccionados aleatoriamente del registro de propiedades maestras de 2003 de la ciudad. El modelo especificado incluyó las variables precio de la vivienda como variable dependiente y como variables independientes el área de la misma, su edad, el número de baños, el hecho de tener aire acondicionado y chimenea, el tipo de piso y las imperfecciones de la misma.

Los resultados del modelo global de Yu (2004) muestran que el modelo es significativo globalmente y que las variables empleadas son estadísticamente significativas. En el caso de las regresiones geográficamente ponderadas, especifica el siguiente modelo:

$$\text{Precio} = f(\text{Área}, \text{Edad}, \text{Aire Acondicionado}, \text{Número de Baños}, \text{Suelos Imperfectos})$$

A partir de él obtiene varios resultados importantes: En primer lugar, con excepción del área de la vivienda las relaciones entre las características de las viviendas o del vecindario no son significativas en todas las zonas de la ciudad, contrario a lo que puede deducirse del modelo global. En segundo lugar, la cercanía de las viviendas al lago Michigan hace que los precios de las viviendas sean mucho más sensibles a sus atributos, por lo que posiblemente sea necesario incluir en el modelo una medida de distancia al lago. Por último, en algunas zonas la edad agrega valor a las viviendas, presumiblemente porque son un indicativo del valor histórico de las mismas.

Sin embargo, hasta el momento los estudios realizados con la técnica GWR en el área de precios de vivienda han sido pocos como puede verse en la Tabla 1 y en términos generales han utilizado en las ecuaciones especificadas variables intrínsecas de la vivienda como área, número de baños, edad de construcción y garajes, entre otros y adicionalmente han empleado medidas de distancia a zonas verdes, instituciones educativas, supermercados, etc. así como variables dummy con el fin de verificar la relación existente entre el medio que rodea la vivienda y su precio, obteniendo resultados significativos en la mayoría de ellos. En su mayoría, las variables más empleadas en estos trabajos son las relacionadas con el área de la construcción, la edad del inmueble y medidas de distancia del inmueble a infraestructura relevante en la zona. En todos los casos, las dos primeras variables resultaron ser significativas al 99%, mientras que para el caso de la variable distancia resultó ser relevante en casos relacionados con la cercanía al mar y las escuelas, mientras que en casos como la cercanía a zonas verdes, resultaron no serlo.

Este tipo de trabajos ponen sobre la mesa la importancia de las regresiones locales desarrolladas a través de la técnica GWR, porque permiten identificar zonas en las cuales determinados parámetros del modelo no son significativos o se presentan cambios muy fuertes en las magnitudes de los mismos indicando una mayor o menor sensibilidad de las viviendas a los parámetros.

De igual manera, los estudios citados muestran las variaciones en los coeficientes de determinación obtenidos de las múltiples ecuaciones locales, exponiendo las zonas para las cuales el modelo global no representa adecuadamente la realidad del mercado.

<b>Título</b>	<b>Autores</b>	<b>Año</b>	<b>VARIABLES EMPLEADAS</b>	<b>Zona</b>
Statistical analysis of the relationship between public transport accessibility and flat prices in Riga,	Dimitry Pavlyuk	2009	Precio de la vivienda, distancia al lago, distancia a las zonas verdes, distancia a supermercados, distancia a inmuebles destinados a educación superior, disponibilidad de transporte, ubicación en primer piso, nuevo proyecto de apartamentos, proyecto que separa nuevos edificios, proyecto que separa edificios nuevos de antiguos	Riga
Measuring the contribution of water and green space amenities to housing values: an application and comparison of spatially - weighted hedonic models	Seong - Hoon Cho, J.M. Bowker, William M. Park	2006	Precio de la vivienda en el año 2.000, área de la construcción, área del lote, edad, número de cuartos, garaje, chimenea, calidad de la construcción, calidad de la estructura, densidad de la población, tiempo requerido para ir al trabajo, ingreso per cápita, desempleo, cercanía a escuelas, distancia a vías del tren, distancia a parque local, distancia al río, distancia al lago, tamaño del parque más cercano, tasa de interés, temporada de venta, zona (urbana, rural), área de inundación	Knox County, Tennessee
Modeling housing market dynamics in the city of Milwaukee: A geographically weighted regression approach	Danlin Yu	2004	Precio de la vivienda, tamaño de la construcción, edad de la vivienda, chimenea, aire acondicionado, número de baños, suelos, superficies impermeables	Milwaukee, Wisconsin
Incorporating spatial variation in housing attribute prices: a comparison of geographically weighted regression and the spatial expansion method	Christopher Bittr, Gordon F. Mulligan, Sandy Dall'erba	2006	Precio de venta, patio, área construida, área del lote, presencia de aire acondicionado, presencia de piscina, número de cuartos, calidad de la construcción, edad de la construcción, número de propietarios, número de baños, dummy si la construcciones de antes de 1.940, garaje, calidad interior de la vivienda	Tucson, Arizona

Tabla 1. Aplicaciones GWR y mercado inmobiliario.

### 3. METODOLOGÍA

#### 3.1 Preliminares matemáticos y econométricos MCO y MCO Espacial

Para la explicar la variabilidad de una variable en función de otras variables se ha utilizado tradicionalmente el modelo lineal general. Para el caso de variables con algún tipo de dependencia espacial, esta especificación es la correcta solo en el caso en el que este efecto esté explicado en su totalidad por las variables exógenas del modelo.

La especificación del modelo es la siguiente:

$$Y = X\beta + \varepsilon$$

En donde  $\varepsilon$  se distribuye normal con media cero y varianza  $\sigma^2 I$

Para estimar el vector de parámetros  $\beta$  es comúnmente utilizado el método de Mínimos Cuadrados Ordinarios (MCO), el cual se dice que  $X$  tiene una distribución normal  $m$ -variante si su función de densidad de probabilidad es:

$$f_X(x) = (2\pi)^{-m/2} (\det \Sigma)^{-1/2} e^{-(1/2)(x-\mu)'\Sigma^{-1}(x-\mu)}$$

Donde  $\Sigma$  es definida positiva.

Para demostrar lo anterior puede escribirse  $\Sigma$  como  $\Sigma = CC'$  donde  $C$  es una matriz de  $m \times m$  no singular. Sea:

$$X = CU + u$$

Donde  $U_{m \times 1}$  es un vector de variables aleatorias independientes que se distribuyen  $N(0,1)$ , es decir,  $U$  se distribuye  $N_m(0, I_m)$ . La función de densidad de probabilidad conjunta de  $U_1, U_2, \dots, U_m$  es:

$$f_U(u) = \prod_{i=1}^m (2\pi)^{-1/2} e^{-(1/2)u_i^2}$$

$$f_U(u) = (2\pi)^{-m/2} e^{-(1/2)u'u}$$

La transformación inversa es  $U=B(X-\mu)$ , con  $B=C^{-1}$ , y el jacobiano de esta transformación es:

$$\det \begin{bmatrix} \frac{\partial u_1}{\partial x_1} & \dots & \frac{\partial u_1}{\partial x_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial u_m}{\partial x_1} & \dots & \frac{\partial u_m}{\partial x_m} \end{bmatrix} = \det \begin{bmatrix} b_{11} & \dots & b_{1m} \\ \vdots & \ddots & \vdots \\ b_{m1} & \dots & b_{mm} \end{bmatrix}$$

$$\det(B) = \det(C^{-1}) = (\det C^{-1}) = (\det(CC^{-1}))^{-1/2} = (\det(\Sigma))^{1/2}$$

De esta forma:

$$f_x(x) = (2\pi)^{-m/2} (\det \Sigma)^{-1/2} e^{(-1/2)(x-\mu)'C^{-1}C^{-1}(x-\mu)}$$

Y como

$$\Sigma^{-1} = C^{-1'}C^{-1}, \text{ entonces}$$

$$f_x(x) = (2\pi)^{-m/2} (\det \Sigma)^{-1/2} e^{(-1/2)(x-\mu)\Sigma^{-1}(x-\mu)}$$

Por otra parte, considérese una variable aleatoria endógena:

$$Y_i = f(X_{i2}, X_{i3}, \dots, X_{ik}, U_i)$$

y la especificación:

$$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \dots + \beta_K X_{ik} + U_i, \text{ con } i=1,2,\dots, N.$$

Donde  $X_{i2}, X_{i3}, \dots, X_{ik}$  son variables aleatorias explicativas de la variabilidad de  $Y_i$  y  $U_i$  es una variable aleatoria ruido blanco.

Para  $i=1,2,\dots,N$  la información del sistema de ecuaciones generado se puede representar en forma matricial como  $Y=X\beta+U$ , donde:

$$Y_{N \times 1} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{bmatrix}; \quad X_{N \times k} = \begin{bmatrix} 1 & X_{12} & \dots & X_{1k} \\ 1 & X_{22} & \dots & X_{2k} \\ \vdots & \vdots & \dots & \vdots \\ 1 & X_{N2} & \dots & X_{Nk} \end{bmatrix}; \quad B_{k \times 1} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_N \end{bmatrix}; \quad U_{N \times 1} = \begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_N \end{bmatrix}$$

$Y_{N \times 1}$ : Vector de observaciones de la variable endógena.

$X_{N \times k}$ : Matriz de diseño.

$B_{k \times 1}$ : Vector de parámetros.

$U_{N \times 1}$ : Valor de perturbación aleatoria.

Por otra parte, el criterio de mínimos cuadrados ordinarios está relacionado con la solución del problema de optimización:

$$\min_{\beta} \sum_{i=1}^N U_i^2 = \min_{\beta} U^T U$$

El cual permite el estimador del vector de parámetros  $\beta$ :

$$\hat{\beta}_{MCO} = (X'X)^{-1} X'Y$$

Por otra parte, como lo menciona Agudelo (2010), Ord (1975) propuso una especificación diferente para modelar la dependencia espacial, la cual se caracteriza por incluir un solo parámetro que recoge la relación existente entre las agregaciones de regiones vecinas y las propias regiones. Para ello utilizó la primera ley de la geografía formulada por Tobler (1979), la cual indica que .todo tiene que ver con todo, pero las cosas cercanas están más relacionadas entre sí que las cosas lejanas.

Ord materializó este principio incluyendo en sus modelos econométricos una matriz de retardos espaciales o matriz de contigüidades  $W$ , la cual representa la materialización de la primera ley de Tobler en términos de proximidad o lejanía. La matriz resultante es entonces del tipo binario, donde 1 representa la proximidad entre la región correspondiente a la fila y la columna y 0 el caso contrario (la diagonal principal estará compuesta por ceros al no tener sentido el concepto de proximidad al región consigo misma).

Al ser difícil la obtención de información más precisa que denote proximidad o lejanía entre un par de regiones, es comúnmente utilizada la vecindad como expresión de proximidad.

Como lo explica Agudelo (2010), esta matriz es también conocida como Matriz de Retardos Espaciales pues se asemeja al operador de retardos utilizado en series temporales. En el contexto temporal, el operador de retardos provoca un desplazamiento en el eje del tiempo equivalente a la potencia  $j$ . Análogamente, la matriz de contigüidades provoca desplazamientos de masas en el espacio, obteniendo así la suma de las magnitudes de la variable en las regiones vecinas a la rezagada.

$$\begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \\ Y_7 \\ Y_8 \\ Y_9 \end{bmatrix} = \begin{bmatrix} Y_2 + Y_4 \\ Y_1 + Y_3 + Y_5 \\ Y_2 + Y_6 \\ Y_1 + Y_5 + Y_7 \\ Y_2 + Y_4 + Y_6 + Y_8 \\ Y_3 + Y_5 + Y_9 \\ Y_4 + Y_8 \\ Y_5 + Y_7 + Y_9 \\ Y_6 + Y_9 \end{bmatrix}$$

### *Efecto del operador de retardo en una serie espacial ( $WY$ )*

El resultado anterior crea un problema de interpretación con la dimensión de la variable resultado. Si se desea preservar la dimensión, se debe imponer la condición de que la suma de los elementos de cada fila sea 1, por lo que es necesario dividir cada componente de  $W$  entre la

suma de todas las componentes de la fila respectiva; este proceso da lugar a la matriz de contactos estandarizada o normalizada y supone, no la agregación de masas en el espacio, sino hallar el promedio de estas.

La matriz de contigüidades es útil para modelar la dependencia espacial a través de las especificaciones que se verán más adelante, pero también puede ser utilizada para la verificación de la existencia de este fenómeno a través, por ejemplo, del estadístico I de Moran.

El estadístico I de Moran permite verificar o rechazar la dependencia espacial en un conjunto de datos y se define como:

$$I = \frac{N}{S_0} \frac{\sum_{i=1}^N \sum_{j=1}^N w_{ij} (Y_i - \bar{Y})(Y_j - \bar{Y})}{\sum_{i=1}^N (Y_i - \bar{Y})^2}$$

Donde:

$w_{ij}$ : Elemento correspondiente a la posición  $(i,j)$  de la matriz de contigüidades.

$S_0$ : Suma de los pesos espaciales.

$\bar{Y}$ : Valor esperado de la variable  $Y$ .

$N$ : Número de regiones.

Cuando se utiliza la matriz de contigüidades estandarizada, la especificación de la I de Moran se convierte en:

$$I = \frac{\sum_{i=1}^N \sum_{j=1}^N w_{ij} (Y_i - \bar{Y})(Y_j - \bar{Y})}{\sum_{i=1}^N (Y_i - \bar{Y})^2}$$

Pues la suma de los elementos de cada fila dará como resultado la unidad, por tanto, el resultado de:

$$\sum_{i=1}^N \sum_{j=1}^N w_{ij}$$

Será  $N$ .

Las pruebas de hipótesis para la I de Moran son las siguientes:

$H_0: \rho=0$

$H_a: \rho \neq 0$

Con  $Y = \rho WY + \varepsilon$

Y estandarizando se obtiene:

$$Z = \frac{I - E(I)\sqrt{n}}{\sqrt{v(I)}}$$

Una vez comprobada la existencia de dependencia espacial, puede procederse con su modelación a través de alguna de las siguientes especificaciones:

### **Modelo autorregresivo de regresión espacial de orden 1. SAR(1)**

$$Y = \rho WY + \varepsilon$$

En donde  $\varepsilon$  se distribuye normal con media cero y varianza  $\sigma^2 I$ .

Hace referencia a una situación en la cual los valores que adopta una variable están determinados por la localización geográfica de la misma.

### **Modelo Básico de Regresión lineal.**

$$Y = X\beta + \varepsilon$$

En donde  $\varepsilon$  se distribuye normal con media cero y varianza  $\sigma^2 I$ .

Como se dijo anteriormente, esta especificación será correcta para modelar una variable con dependencia espacial, solo en el caso en el que este efecto esté explicado en su totalidad por las variables exógenas del modelo.

### **Modelo del error espacial**

$$Y = X\beta + U$$

$$U = \lambda WU + \varepsilon$$

En donde  $\varepsilon$  se distribuye normal con media cero y varianza  $\sigma^2 I$ .

Esta es la especificación más utilizada cuando el modelo básico de regresión lineal resulta ineficaz para explicar el efecto espacial presente en la variable endógena, trasladando entonces este efecto hacia los términos de error.

### **Modelo mixto autorregresivo de regresión espacial o modelo del retardo espacial**

$$Y = \rho WY + X\beta + \varepsilon$$

En donde  $\varepsilon$  se distribuye normal con media cero y varianza  $\sigma^2 I$ .

El modelo del retardo espacial también puede ser conveniente para explicar la dependencia espacial de la variable endógena; basta con rezagar espacialmente esta variable y agregarla como explicativa del modelo junto con otras variables exógenas que puedan explicar apropiadamente la variabilidad de la endógena.

### 3.2 Preliminares Matemáticos y Econométricos GWR

La explicación de la metodología de GWR sigue la descrita por Fotheringham, Brunsdon y Charlton (2002). La técnica GWR es un aporte reciente al campo de la econometría espacial que permite modelar procesos espaciales heterogéneos. La esencia de esta metodología radica en estimar una ecuación en función de un conjunto de variables que reciben diferentes pesos dependiendo de su ubicación en el espacio. En este sentido GWR es un modelo de regresión obtenido por el método de mínimos cuadrados ponderados, en donde cada una de las ponderaciones está asociada a su localización espacial.

Habitualmente los modelos especificados a través de la econometría tradicional son de la forma:

$$Y_i = \beta_1 + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + \varepsilon_i$$

Con  $i=1,2,\dots,N$ , donde:

$$Y_1 = \beta_1 + \beta_2 X_{12} + \dots + \beta_k X_{1k} + \varepsilon_1$$

$$Y_2 = \beta_1 + \beta_2 X_{22} + \dots + \beta_k X_{2k} + \varepsilon_2$$

$$Y_N = \beta_1 + \beta_2 X_{N2} + \dots + \beta_k X_{Nk} + \varepsilon_N$$

De manera similar podemos escribir:

$$Y_i = \beta_1 + \sum_k \beta_k x_{ik} + \varepsilon_i$$

Es decir:

$$Y = X \beta + \varepsilon$$

$N \times 1$     $N \times K$     $K \times 1$     $N \times 1$

Con:

$$Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_N \end{bmatrix}; \quad X = \begin{bmatrix} 1 & X_{12} & \dots & X_{1k} \\ 1 & X_{22} & \dots & X_{2k} \\ \vdots & \vdots & \dots & \vdots \\ 1 & X_{N2} & \dots & X_{Nk} \end{bmatrix}; \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}; \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix}$$

En donde el primer término de la derecha corresponde al intercepto de la ecuación, el segundo término corresponde a la sumatoria de variables exógenas que explican la variabilidad de la variable independiente  $y$ . Mientras que  $\epsilon_i$  es el término de error del modelo.

A partir de este modelo, es posible obtener los parámetros de la regresión geográficamente ponderada asignando coordenadas a cada uno de los coeficientes  $\beta$ , de manera que a partir del modelo tradicional:

$$y_i = \beta_1 x_1(\beta) + \beta_2 x_2(\beta) + \dots + \beta_j x_j(\beta) + \beta_1(u_1, v_1)x_1 + \beta_2(u_2, v_2)x_2 + \dots + \beta_j(u_j, v_j)x_j + \epsilon_j$$

$$y_i = \sum_{j=1} \beta_j x_j(\beta) + \sum_{j=1} \beta_j(u_j, v_j)x_j + \epsilon_j$$

Se obtiene:

$$Y_i = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i)x_{ik} + \epsilon_i$$

Con  $k=1,2,\dots,N$

La ecuación clásica obtenida a través de una regresión es de la forma:

$$Y = X\beta + \epsilon$$

En donde el vector de parámetros  $\beta$  es constante en el espacio se estima de la siguiente manera:

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

En términos de regresiones geográficamente ponderadas, su equivalente se obtiene de la siguiente manera:

Inicialmente se considera un modelo lineal general:

$$Y = X\beta + U$$

Con  $U$  vector de perturbaciones aleatorias ruido blanco.

Ahora, pre multiplicando por una matriz de ponderaciones  $T$  no singular se obtiene:

$$TY = TX\beta + TU$$

Como  $T$  no es variable aleatoria, entonces:

$$E(TU) = TE(U) = 0$$

Se tiene también que la matriz de varianzas y covarianzas de  $TU$  es:

$$\sum_{TU} = E[(TU) - E(TU)(TU - E(TU))^T]$$

Como

$$E(TU) = 0$$

entonces:

$$\begin{aligned}\sum_{TU} &= E[TUU^T T] \\ \sum_{TU} &= TE[UU^T] T^T \\ \sum_{TU} &= TE[UU^T] T^T\end{aligned}$$

En los modelos espaciales se introduce naturalmente la heteroscedasticidad. Supongamos entonces que el modelo presenta este fenómeno y que por lo tanto:

$$\sum_U = E[UU^T] = \begin{bmatrix} \sigma_1^2 & & \\ & \sigma_2^2 & \\ & & \sigma_N^2 \end{bmatrix} = \sigma^2 W^{-1} \text{ con } W^{-1} \neq I$$

De acuerdo a lo anterior,

$$\begin{aligned}\sum_{TU} &= TE[UU^T] T^T \\ \sum_{TU} &= T\sigma^2 W^{-1} T^T \\ \sum_{TU} &= \sigma^2 T W^{-1} T^T\end{aligned}$$

Por lo tanto, para que no exista heteroscedasticidad es necesario que  $TW^{-1}T^T = I$ .

El estimador del parámetro  $\beta$  de un modelo geográficamente ponderado es entonces:

$$\hat{\beta} = (X^T W(i) X)^{-1} X^T W(i) Y$$

Pues, al suponer que  $TW^{-1}T^T = I$ , se obtiene que:

$$W^{-1} = T^{-1} (T^T)^{-1}$$

O lo que es lo mismo:

$$W = T^T T$$

Por desarrollo del Modelo Lineal General, se sabe que el estimador del modelo modificado  $TY = TX\beta + TU$  es:

$$\hat{\beta}_{MCTG} = [(TX)^T TX]^{-1} (TX)^T TY$$

$$\hat{\beta}_{MCTG} = [X^T T^T TX]^{-1} X^T T^T TY$$

Como  $W = T^T T$ , entonces:

$$\hat{\beta}_{MCTG} = (X^T W X)^{-1} X^T W Y$$

Donde  $i$  representa una columna de la matriz y  $W(i)$  es una matriz de pesos espaciales de  $n$  por  $n$  de la forma:

$$W = \begin{bmatrix} w_{i1} & \dots & 0 \\ \dots & w_{i2} & \dots \\ 0 & \dots & w_{in} \end{bmatrix}$$

La técnica GWR permite generar superficies de coeficientes (superficies en las cuales los valores de los diferentes coeficientes son similares) para cada variable del modelo, reflejando el cambio de los mismos a través del espacio como puede verse en el gráfico 1.

$$Precio_i = \beta_0 + \beta_1 Distancia_i + \beta_2 \text{Área}_i + \varepsilon_i$$

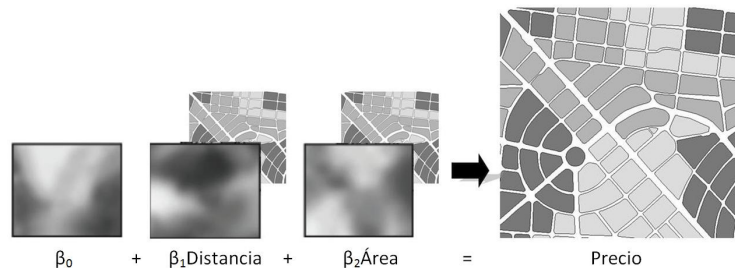


Gráfico 1. Superficies de Coeficientes

En la ilustración se define:

*Distancia:* Es el trayecto existente entre una vivienda y la estación del metro, medida en kilómetros

*Área:* Es la superficie construida de la vivienda medida en metros cuadrados (m<sup>2</sup>)

También puede ocurrir que en algunas situaciones sólo algunos de los coeficientes del modelo varíen geográficamente, por lo que deben tenerse en cuenta modelos en los cuales algunos coeficientes son locales y otros globales, los cuales representan por ejemplo la tasa de cambio Euro/Dólar o el PIB de un país, por lo que el modelo especificado adquiere la forma:

$$y_i = \alpha_1 x_1(\alpha) + \alpha_2 x_2(\alpha) + \dots + \alpha_j x_j(\alpha) + \beta_1(u_1, v_1) x_1 + \beta_2(u_2, v_2) x_2 + \dots + \beta_j(u_j, v_j) x_j + \varepsilon_j$$

$$y_i = \sum_{j=1} \alpha_j x_j(\alpha) + \sum_{j=1} \beta_j(u_j, v_j) x_j + \varepsilon_j$$

En donde  $y_i$  es la variable dependiente. El primer término del lado derecho representa los coeficientes globales y el segundo término representa los coeficientes locales. Este modelo es conocido como modelo mixto.

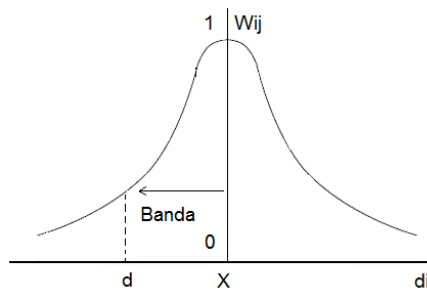


Gráfico 2. Función Kernel y ancho de banda

Uno de los aspectos cruciales en el uso de la técnica GWR es el cálculo de la matriz espacial de pesos  $W(i)$ , la cual asume implícitamente la existencia de correlación espacial y emplea una función kernel que da un mayor peso a las localizaciones cercanas en el espacio antes que a las más distantes, permitiendo calcular el ancho de banda, entendido en el gráfico 2 como la distancia entre el punto “ $x$ ” y el punto “ $d$ ”. Existen dos tipos de kernel: uno fijo que emplea un mismo ancho de banda para cada una de las observaciones utilizadas en el modelo y uno adaptativo que emplea anchos de banda distintos para cada una de las observaciones dependiendo de la densidad de los puntos. En la práctica el más empleado es el kernel adaptativo debido a que permite tener anchos de banda más amplios en regiones donde hay muchos datos y un poco más estrechos en donde la densidad de puntos es menor, evitando los problemas de sesgo que presentan los kernels fijos al asignar un solo ancho de banda independientemente de la densidad de los puntos.

Existen diferentes tipos de funciones kernel que pueden emplearse para determinar los pesos de las observaciones, siendo la más común aquella en la cual se asigna un valor igual a uno a los datos cercanos y cero a los más alejados:

$$W_{ij} = \begin{cases} 1 & \text{si } d_{ij} \leq d^* \\ 0 & \text{en otro caso} \end{cases}$$

En donde  $d$  es la distancia de la observación, en caso de que sea cercana o menor a  $d^*$  es igual a 1 y en caso contrario es cero.

Sin embargo este tipo de función no es muy empleada por ser una función con un gran nivel de subjetividad y porque genera un problema de discontinuidad espacial. Para solucionar estos problemas se emplean otro tipo de funciones como la Gaussiana que reduce las ponderaciones de los pesos en la medida en que las distancias aumentan acuerdo con una función del tipo:

$$w_{ij} = \exp\left[-1/2(d_{ij}/b)^2\right]$$

En donde  $d_{ij}$  es la distancia entre dos puntos y  $b$  es el ancho de banda.

Una función de kernel alternativa puede ser del tipo *bi cuadrada* que provee una distribución cercana a la gaussiana con la particularidad de que produce pesos similares a los de esta distribución en la medida en que se encuentren en el dominio de la distancia  $b$  y de 0 si están por fuera de la misma.

$$W_{ij} = \begin{cases} \left[1 - \left(\frac{d_{ij}}{b}\right)^2\right]^2 & \text{si } d_{ij} \leq b \\ 0 & \text{en otro caso} \end{cases}$$

Existen varias formas de generar anchos de banda adaptativos con este tipo de funciones, una de ellas consiste en agrupar los datos en términos de la distancia de cada punto  $i$  al punto  $j$ , de manera que los puntos más cercanos tienen un peso igual a 1 y su peso decrece en la medida en que el rango aumenta de acuerdo con una función de este tipo:

$$w_{ij} = \exp(-R_{ij}/b)$$

Esta función ajusta los anchos de banda dependiendo de los rangos de puntos y de sus distancias. También es posible asegurándose de que la suma de los pesos sea igual a una constante, lo que matemáticamente se expresa como:

$$\sum_j w_{ij} = c \quad \text{Para todo } i$$

Aquí la constante  $c$  puede escogerse libremente o empleando algún tipo de procedimiento.

Otros tipos de kernel que pueden también emplearse son los siguientes:

KERNEL	FUNCIÓN DE DENSIDAD
Epanechnikov	$K(u) = \begin{cases} \frac{3}{4\sqrt{5}} \left(1 - \frac{1}{5}u^2\right) & \text{si }  u  < \sqrt{5} \\ 0 & \text{en otro caso} \end{cases}$
Bionderado	$K(u) = \begin{cases} \frac{15}{16} (1 - u^2)^2 & \text{si }  u  < 1 \\ 0 & \text{en otro caso} \end{cases}$
Triangular	$K(u) = \begin{cases} (1 -  u ) & \text{si }  u  < 1 \\ 0 & \text{en otro caso} \end{cases}$
Parzen	$K(u) = \begin{cases} \frac{k_1 - u^2}{k_2} &  u  \leq C_1 \\ \frac{u^2}{k_3} - k_4 u  + k_5 & C_1 <  u  \leq C_2 \\ 0 &  u  > C_2 \end{cases}$
Normal	$K(u) = \frac{1}{2\sqrt{\pi}k_6} \exp\left(\frac{-u^2}{2k_6^2}\right)$

Tabla 2. Tipos de funciones Kernel

Sin embargo, la escogencia de la función kernel no es tan crítica como la calibración de la matriz de pesos espaciales, o lo que es lo mismo el cálculo del ancho de banda, debido a que puede generar problemas de sesgo en los coeficientes si no está correctamente especificada. Para ello se han desarrollado varios tipos de criterios dentro de los cuales se encuentran:

*El criterio de validación cruzada:* A través de un proceso iterativo busca el ancho de banda del kernel que minimiza el cuadrado del error de predicción de la variable  $y$ . Tiene la propiedad deseable de que en la medida en que  $b$  se hace muy pequeño el modelo se calibra solo en muestras cercanas a  $i$ .

$$CV = \operatorname{argmin} \sum_{i=1}^n \left[ y_i - \hat{y}_{(i)}(b) \right]^2$$

*El criterio de Akaike:* Busca un intercambio entre bondad de ajuste y grados de libertad, puede ser usado en regresiones lineales, logísticas y del tipo Poisson y está definido como:

$$AIC_c = 2n \log_e(\sigma) + n \log_e(2\pi) + n \left\{ \frac{n + \operatorname{tr}(S)}{n - 2 - \operatorname{tr}(S)} \right\}$$

Donde  $n$  es el tamaño de la muestra,  $\sigma$  es el error estándar estimado del término de error y  $\operatorname{tr}(S)$  es la traza de la matriz sombrero, que es función del ancho de banda.

*El criterio de Schwartz:* es utilizado principalmente con muestras pequeñas, pues cuando las muestras son grandes tiende a identificar pocos parámetros como óptimos, se define como:

$$BIC = -2\log_e(L) + k \log_e(n)$$

Donde L es la verosimilitud del modelo, este criterio es óptimo para pocos parámetros.

Un aspecto importante son las pruebas de bondad de ajuste, que permiten comparar el modelo econométrico global con el modelo de regresión geográficamente ponderado que permite obtener tantas ecuaciones como datos hay en el modelo a través de medidas como:

*Coefficiente de determinación*: mide la proporción de la variación en la variable dependiente que es explicada por el modelo de regresión, este coeficiente varía entre 0 y 1, mientras más cercano es a la unidad mejor es el ajuste.

*Coefficiente de determinación ajustado*: mide la proporción de la variación en la variable dependiente que es explicada por el modelo de regresión ajustada por el número de variables. Este coeficiente varía entre 0 y 1, mientras más cercano es a la unidad mejor es el ajuste, pero penaliza el valor del coeficiente en la medida en que se agreguen más regresores.

*Criterio de información de Akaike*: Es una medida que permite observar la “distancia relativa” entre el modelo ajustado y el “modelo verdadero”, este criterio igual que el coeficiente de determinación ajustado impone una penalidad en la medida en que agrega más regresores. Los modelos con menores valores del criterio de Akaike se consideran con un mejor ajuste respecto al “modelo verdadero”.

## 4. CASO DE ESTUDIO

### 4.1 Contexto

San Javier es una comuna de estrato bajo y tamaño medio ubicada en el centro occidente de Medellín, Colombia, la cual cuenta según la subsecretaría de catastro municipal, aproximadamente con 29,500 predios de uso habitacional repartidos en veinte barrios, cerca del 94% corresponden a los estratos uno (ubicados en la periferia y bastante alejados de la estación del metro), dos y tres y el 6% restante a viviendas de estrato cuatro.

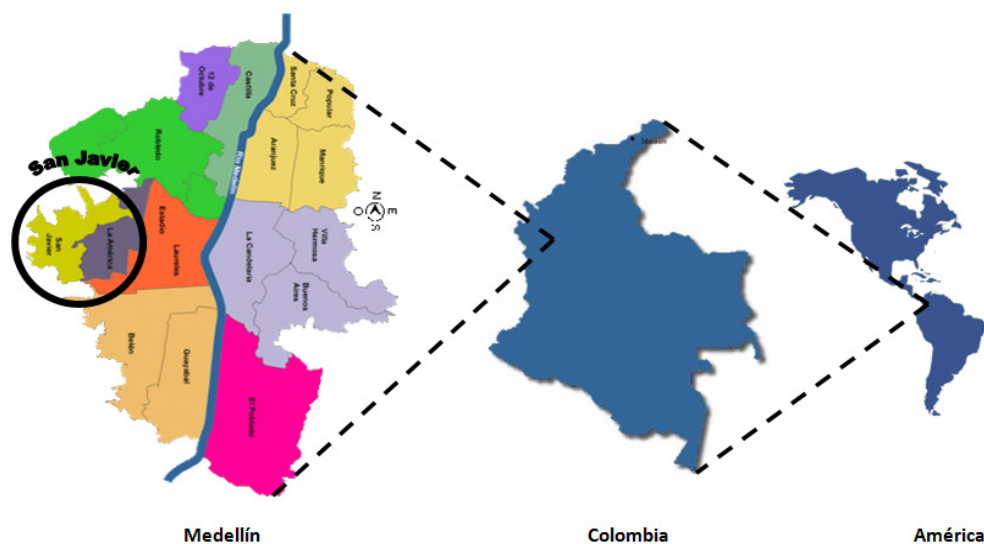


Gráfico 3. Ubicación espacial de San Javier

Según la encuesta de calidad de vida de la ciudad realizada en 2008 su población es de casi ciento sesenta y seis mil personas, la mayoría de los cuales son mestizas y cerca del uno por ciento son de raza negra. Aproximadamente noventa mil habitantes de la zona tienen un nivel educativo igual o inferior a educación primaria. La tasa de desempleo en el lugar es de aproximadamente el 12.5%, casi tres puntos por encima del promedio de la ciudad.

La estructura urbana de San Javier es homogénea con predominancia de inmuebles residenciales unifamiliares, por encima de los multifamiliares, de los comerciales y de los servicios, lo que ha hecho que su infraestructura vial se encuentre en regular estado y sus especificaciones luzcan inferiores frente a la cantidad de vehículos que transitan diariamente por ella. A pesar de ello, la zona cuenta con un buen sistema de transporte público integrado con el sistema metro, que permite a sus habitantes viajar rápidamente de un extremo a otro de la ciudad sin generar mayores presiones en las vías.

El sistema de transporte masivo cuenta con una estación ubicada en las inmediaciones de la carrera 99 con la calle 45, es la última estación de la línea B, la cual comunica la zona con el centro de la ciudad. En ella confluyen todas las rutas del sistema integrado de transporte SIT que permite a los habitantes de las zonas más alejadas viajar desde sus casas en un bus hasta el

sistema. La estación del metro se ha convertido en un punto de referencia en la zona debido al aumento de la calidad urbanística asociada a la construcción de la estación y reflejada en mejoras de andenes, semaforización de las vías y construcción de bahías de parqueo, entre otros, así como al continuo flujo de pasajeros que transformó la zona cercana a la estación pasando de ser netamente residencial a ser una mezcla de uso residencial con comercial.

La estación del metro de San Javier comenzó a operar junto con el resto de la línea B en 1996. A partir de entonces la zona ha sufrido cambios importantes convirtiendo en corredores comerciales zonas que antes eran residenciales gracias al importante flujo de personas que se dirigen diariamente a la estación. En la ciudad ha habido algunas discusiones importantes acerca del impacto que ha tenido el metro en la localidad, pero siempre referidos al deterioro que ha generado en el centro de la ciudad la línea A del metro que se caracteriza por estar cercana a comercio e industria, pero no a la línea B que tiene su área de influencia en zonas residenciales quizá porque su construcción se hizo sobre una canalización lo que permitió que se insertara más fácilmente en los ambientes residenciales de la zona.

Cerca de la estación se encuentran ubicados un parque infantil, una escuela pública primaria, un parque biblioteca, una iglesia y la línea del metro cable, así como la calle San Juan que es un referente vial en la ciudad.

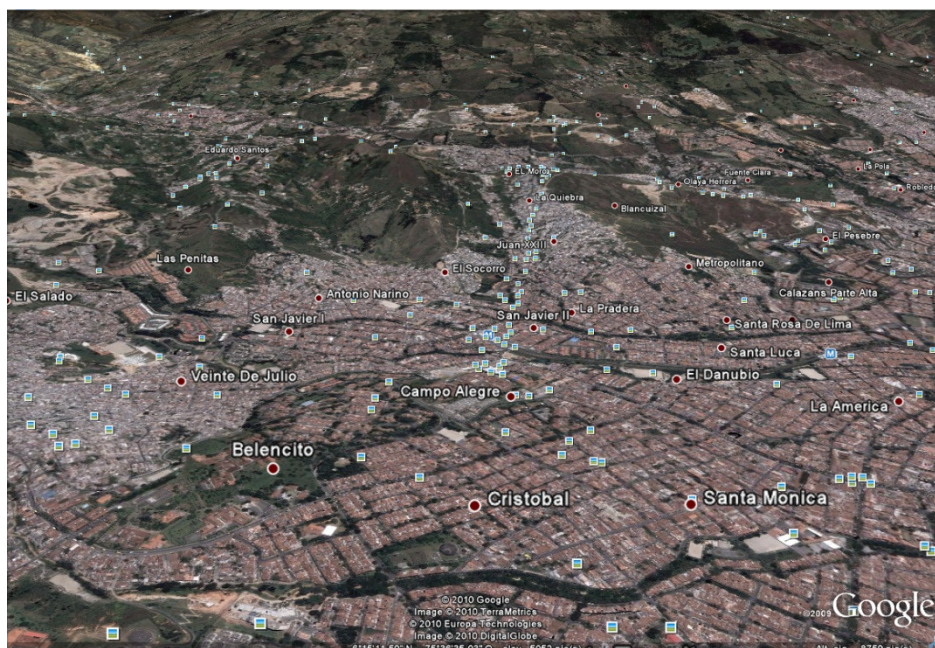


Gráfico 4. Fotografía San Javier

En cuanto al equipamiento urbano, la zona presenta, como en muchos otros lugares de la ciudad, déficits de espacios culturales y recreativos, sin embargo, cuenta con amplias zonas verdes en las partes altas como se observa en el gráfico 4, las cuales pueden emplearse como reservas forestales o parques ecológicos en un futuro.

Como en la mayoría de las comunas de Medellín, el plan de ordenamiento territorial en San Javier permite unos mayores índices de construcción en las zonas bajas y de topografía plana y menores en las zonas altas y pendientes como una manera de regular la densificación de la ciudad. Sin embargo, los datos de las encuestas de calidad de vida de la ciudad muestran que las zonas con mayor número de viviendas por hectárea de esta comuna están ubicadas en los barrios altos como Las Independencias, La Pradera y El Pesebre, que tienen un menor nivel de desarrollo arquitectónico y urbano, mientras que algunos de los barrios con las menores densidades como San Javier No. 1 y San Javier No.2 se ubican en las zonas bajas que tienen una calidad urbanística superior, es decir, se presenta todo lo contrario a lo que busca el plan de ordenamiento territorial.

En términos del mercado inmobiliario residencial, San Javier se ha caracterizado por representar entre el cinco y el ocho por ciento del total de negocios de vivienda realizados en la ciudad y ha participado activamente en los auges y depresiones que ha sufrido el mercado inmobiliario local. Es así como durante 2002, un año regular para el mercado inmobiliario, se negociaron 1,200 viviendas, mientras que en 2006, año de auge, se transaron más de 2,200, mostrando que es una zona articulada con el desarrollo inmobiliario y comercial de la ciudad.

La motivación del estudio radica en presentar evidencia de que el desarrollo de obras públicas en zonas menos favorecidas de la ciudad contribuye a mejorar el bienestar de sus habitantes a través de un aumento de su patrimonio expresado en términos del aumento de precio de sus viviendas. De igual forma este estudio permite clarificar el sustento económico de los cobros por valorización que realizan las administraciones locales cuando se ejecuta una intervención urbanística o se desarrolla una obra pública, los cuales usualmente generan polémica pues no es claro si valorizan o desvalorizan las propiedades, ni es clara la magnitud del impacto que generan en los precios de las viviendas.

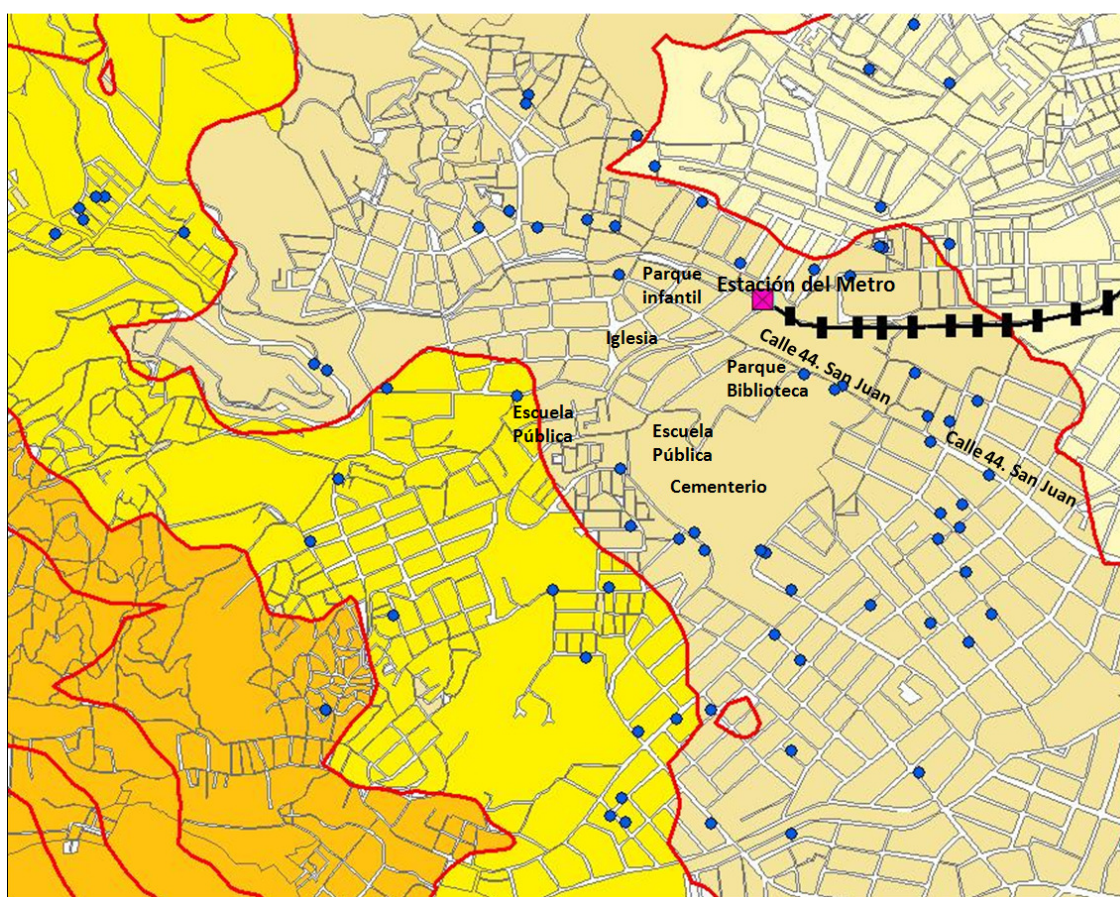
#### **4.2 Datos Utilizados**

Para desarrollar el estudio se emplearon datos correspondientes a 89 precios de viviendas ubicadas en los alrededores de la estación del metro en San Javier, localizados en once barrios que cuentan con 21,503 viviendas, detallados en avalúos realizados por 4 peritos distintos entre 2008 y 2009. En total la muestra representa el 0.41% del total de viviendas de los barrios estudiados, la cual es inferior a la empleada en otros de estudios de este tipo como el desarrollado por Yu(2004) que es del 0.86% en Milwaukee, pero similar a las utilizadas en Colombia dada la escasez y dificultad para conseguir los datos. Por ejemplo la muestra empleada por Medina y Morales (2006) en Bogotá es del 0.45% a pesar de que corresponde a valores catastrales y no de mercado como los empleados en este trabajo, los cuales son más adecuados por ser más precisos y confiables.

Los datos empleados fueron aportados por la Lonja de Propiedad Raíz de Medellín y Antioquia de su base de datos particular, debido a que las cifras de precios de inmuebles transados en el mercado no quedan inscritos ni en las oficinas de registro de instrumentos públicos ni en las oficinas de catastro debido a las condiciones de seguridad existentes en el

país. Por tal motivo, los precios fijados en los avalúos constituyen una buena aproximación a lo que son los precios de mercado de los inmuebles y mejores que los precios de oferta que continuamente se utilizan para desarrollar este tipo de trabajos en Colombia, debido a que los precios de oferta normalmente están sobrevaluados con el fin de dejar un margen de negociación en el caso de que pueda darse un trato.

Cada dato de avalúo contiene información correspondiente a la fecha en que se realizó el avalúo, el precio de la vivienda, el tipo de inmueble avaluado (casa o apartamento), la edad de la construcción, el área construida, el área del parqueadero si lo tiene, el número de baños, el nombre del edificio o la unidad residencial en que está ubicada la vivienda si es el caso y una serie de características de la unidad residencial o del edificio como son tener ascensor, portería, tanque de agua, parqueadero de visitantes, juegos de niños, piscina, sauna, turco, entre otros.



*Gráfico 5. Topografía e infraestructura de San Javier*

En el gráfico 5 se observa el plano de la zona, en donde cada punto representa un dato de avalúo. El cuadrado oscuro representa la estación del metro y la línea oscura la vía del tren. Las franjas de colores que van de izquierda a derecha representan la topografía del terreno que es más inclinada en las zonas más oscuras y más plana en las zonas más claras.

Las estadísticas descriptivas muestran la diversidad de las viviendas de la zona, que incluyen viviendas con edades de construcción que van desde uno hasta los sesenta años, áreas entre los veintiséis hasta los doscientos treinta y ocho metros cuadrados, con distancias a la estación del metro que oscilan entre los 0.09 y los 1.8 kilómetros, con estratos que varían entre dos y cuatro:

	<b>Máximo</b>	<b>Mínimo</b>	<b>Promedio</b>
Valor total	198,420,000	18,988,188	78,612,317
Edad (años)	60	1	21
Construcción (m2)	238	35	104
Estrato	4	2	
No. Baños	4	1	2
Características	6	0	
Distancia (km)	1.845	0.0989	0.778
Parqueadero (m2)	14	0	0.4

*Tabla 3. Estadísticas de las variables utilizadas.*

Los datos muestran una que la zona tiene algún nivel de redesarrollo. En las franjas cercanas a la estación del metro existen casas antiguas con grandes áreas y precios elevados mezcladas con apartamentos de pequeñas áreas y precios altos, mientras en los lugares más alejados y ubicados mucho más al occidente hay apartamentos y casas con áreas construidas menores, de estrato más bajo, algunas de ellas ubicadas en unidades cerradas con algunas características importantes como parqueaderos de visitantes, juegos de niños y portería. Mientras que las viviendas ubicadas un poco más al oriente tienen una mezcla de casas antiguas de áreas grandes, con apartamentos relativamente nuevos, de áreas pequeñas y altos precios.

Las variables empleadas para desarrollar el estudio y los signos esperados de las mismas se detallan a continuación:

Variable	Definición	Signo Esperado
Fecha	Indica la fecha en que se realizó el avalúo, se cuenta como el número de meses hacia atrás a partir del último avalúo registrado.	Negativo. En la medida en que el avalúo sea más antiguo, el valor de la vivienda deberá ser menor.
Inmueble	Es una variable dicótoma que toma el valor de 0 si es un apartamento y 1 si es casa.	Positivo. Una casa tiene un mayor valor que un apartamento.
Construcción	Es el área construida de la vivienda medida en metros cuadrados (m2).	Positivo. A mayor área mayor precio.
Distancia	Es la distancia entre la vivienda y la estación del metro medida en kilómetros (km).	Negativo. En la medida en que la vivienda esté más alejada de la estación debe tener un menor valor.
Estrato2	Es el estrato socioeconómico de la vivienda. Toma el valor de 1 si es estrato dos, en caso contrario toma el valor de cero.	Negativo. En la medida en que una vivienda es estrato dos debe tener un menor precio que las de los estratos más altos. La categoría base es el estrato 4.
Estrato3	Es el estrato socioeconómico de la vivienda. Toma el valor de 1 si es estrato tres, en caso contrario toma el valor de cero.	Negativo. En la medida en que una vivienda es estrato tres debe tener un menor precio que las de los estratos más altos. La categoría base es el estrato 4.
Baños	Es el número de baños de la vivienda. Es una variable cuantitativa.	Positivo. A mayor número de baños mayor precio.
Característica	Es el número de beneficios que tiene una vivienda ubicada en unidad cerrada.	Positivo. A mayores características de la unidad mayor precio.
Edad	Indica los años de construcción de la vivienda. Medidos en años.	Negativo. A mayor edad menor valor.
Edad2	Indica el cuadrado de los años de construcción de la vivienda. Medidos en años.	Negativo. A mayor edad menor valor.
Unidad	Indica si el inmueble está en unidad cerrada.	Positivo. En términos generales las viviendas ubicadas en unidad cerrada tienden a tener mayor precio que las ubicadas a borde de calle.
Parqueadero	Indica el área del garaje de la vivienda, medido en metros cuadrados (m2).	Positivo. A mayor área del parqueadero mayor precio

Tabla 4. Definición de las variables empleadas.

## Resultados

En la siguiente sección se desarrolla inicialmente el análisis econométrico con la regresión por OLS, luego se emplean modelos OLS espaciales y finalmente se utiliza el modelo de regresión geográficamente ponderada que permite tratar datos correlacionados espacialmente y obtener coeficientes que varían en el espacio.

### 4.3 .1 Modelos OLS.

Inicialmente con el fin de detectar si existe correlación espacial utiliza el estadístico I de Moran, que arroja una un valor de 0.5019 con un p-valor de 0.0010, es decir el test muestra la existencia de correlación espacial positiva en la variable precio.

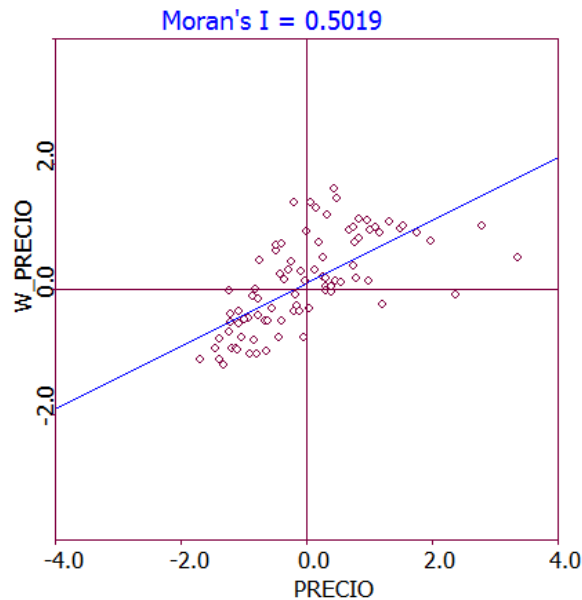


Gráfico 6. I de Moran

Luego con los 89 datos disponibles y empleando el programa GWR<sup>3</sup> se obtuvo un modelo econométrico tradicional a partir de la siguiente especificación:

$$\text{Precio} = f(\text{fecha}, \text{Inmueble}, \text{Edad}, \text{Edad2}, \text{Construcción}, \text{Parqueadero}, \text{Distancia}, \text{Estrato2}, \text{Estrato3}, \text{Baños}, \text{Unidad}, \text{Características}, \text{error})$$

Los resultados fueron los siguientes: En primer lugar, el modelo ajustado explica cerca del 83% de la variación en los precios de las viviendas. En segundo lugar todas las variables del modelo tienen los signos esperados y ocho de las once variables incluidas son significativas: Construcción, Parqueadero, Distancia, Estrato2, Estrato3, Baños y Características con niveles de significancia superiores al 95%.

---

2. Desarrollado por Stewart Fotheringham, Martin Charlton y Chris Brundson. Disponible en: <http://ncg.nuim.ie/ncg/GWR/>

Parameter	Estimate	Std Error	T
Intercepto	40,691,640	10,030,115	4.056
Fecha	-589,336	274,707	-2.145**
Inmueble	2,360,994	4,626,913	0.510
Edad	-749,909	470,639	-1.593
Edad2	10,183	8,872	1.147
Construccion	599,105	70,703	8.473*
Parqueadero	664,742	256,424	2.592*
Distancia	-15,334,198	5,138,936	-2.983*
Estrato2	-30,390,019	6,307,967	-4.817*
Estrato3	-21,810,221	3,831,272	-5.692*
Banos	11,830,123	2,829,071	4.181*
Unidad	-103,771	4,804,863	-0.0215
Caracteristicas	2,184,874	1,205,624	1.812*

\* Significativo al 99%

\*\* Significativo al 95%

Tabla 5. Coeficientes OLS

Los mayores valores de coeficientes se presentan en su orden en las variables: *distancia* que mide la distancia de la vivienda al metro, lo cual implica que a mayor cercanía hay mayor facilidad de transporte y mejoras urbanísticas asociadas a la construcción de la estación. La variable *baños* que indica el número de baños que tiene la vivienda y es una muestra del estatus y la calidad de la misma, debido a que las viviendas de estrato más bajo tienen sólo un baño conocido como “baño social”, las de un nivel medianamente superior que tienen dos baños: un “baño social” y un baño en la alcoba del servicio, y las de estrato más alto tienen tres o más baños que incluyen los dos anteriores más un baño en la alcoba principal.

Las medidas de bondad de ajuste muestran unos valores altos y adecuados para el modelo desarrollado por OLS:

Criterio de información de Akaike	3218.3
Coefficiente de Determinación	0.8379
r2 Ajustado	0.8099

El coeficiente de determinación muestra que las variables independientes del modelo explican el 83.26% de la variación del precio de las viviendas. El coeficiente de determinación muestra que el modelo especificado ajustado por el número de variables, explica cerca del 80.62% de los precios de las viviendas.

Sin embargo, el modelo especificado presenta problemas de heteroscedasticidad, como lo muestra el test de heteroscedasticidad de White:

Test de Heteroscedasticidad de White:

	Valor	Probabilidad
Estadístico F	2.09956	0.013493
Observaciones*R <sup>2</sup>	32.60452	0.026689

Por tal motivo se procedió a corregir dicho problema a través de una transformación logarítmica:

Variable dependiente: Log(Precio)

	Estimado	Std. Error	t-Statistic
Unidad	0.041630	0.068920	0.604
Parqueadero	0.009110	0.003678	2.476**
Inmueble	0.056912	0.066367	0.857
Fecha	-0.007948	0.003940	-2.017**
Estrato3	-0.330093	0.054955	-6.006*
Estrato2	-0.574976	0.090480	-6.354*
Edad2	0.000132	0.000127	1.035
Edad	-0.006631	0.006751	-0.982
Distancia	-0.200627	0.073712	-2.721*
Construccion	0.006033	0.001014	5.949*
Caracteristica	0.037349	0.017293	2.159**
Baño	0.143794	0.040580	3.543*
Constante	1.767.610	0.143870	1.228*

\* Significativo al 99%

\*\* Significativo al 95%

Tabla 6. Coeficientes OLS

El modelo presenta los signos esperados para todas las variables. Sin embargo, la interpretación del mismo corresponde al de una elasticidad, por ejemplo, si se incrementa en un metro cuadrado el área construida de la vivienda, en promedio, el precio total de la vivienda aumenta en un 0.6033%, manteniendo todo lo demás constante. De igual forma, el modelo global permite corroborar la hipótesis de la influencia positiva de la estación del metro en los valores de las viviendas: un incremento de un kilómetro en la distancia a la estación del metro genera, en promedio, una reducción de un 20% en el precio total de la vivienda manteniendo todo lo demás constante.

Las medidas de bondad de ajuste del modelo muestran unos valores altos para el modelo: presenta un valor de  $r^2$  igual a 0.81 y un coeficiente de determinación ajustado de 0.78. Adicionalmente el modelo significativo globalmente como lo muestra el valor de la F que es igual al 27.04. De igual forma se ha corregido el problema de heteroscedasticidad del modelo como lo muestra el test de White:

Test de Heteroscedasticidad de White:

	Valor	Probabilidad
Estadístico F	1.308129	0.207281
Observaciones*R <sup>2</sup>	23.56890	0.213212

### 4.3.2 Modelos con error espacial

A través del programa Geoda<sup>3</sup> se especificaron dos tipos de modelos espaciales uno con rezago espacial y otro con error espacial para el correcto tratamiento de los datos espaciales. Inicialmente, la especificación para el modelo con rezago espacial fue la siguiente:

$$\text{Precio} = f(W*\text{Precio}, \text{fecha}, \text{Inmueble}, \text{Edad}, \text{Edad2}, \text{Construcción}, \text{Parqueadero}, \text{Distancia}, \text{Estrato2}, \text{Estrato3}, \text{Baños}, \text{Unidad}, \text{Características}, \text{error})$$

Para los modelos con error espacial,  $W$  es la matriz de pesos espaciales generada con el método k-Nearest Neighborhood igual a 4, que es el número de viviendas que se asume tienen características similares en la zona, relacionadas con edades, tipos y características.

Los resultados del muestran que las variables fecha, construcción, parqueadero, distancia, estrato2, estrato3 y baños son significativas al 99%, mientras que la variable características lo es al 90%.

	<b>Coficiente</b>	<b>Valor Z</b>
W_Precio	0.085947	1.05
Intercepto	33,125,100	2.92
Fecha	-610,346	-2.46*
Inmueble	3,128,074	0.73
Edad	-668,987	-1.57
Edad2	8,640	1.07
Construccion	583,481	8.88*
Parqueadero	638,633	2.77*
Distancia	-14,116,560	-3.01*
Estrato2	-27,706,980	-4.59*
Estrato3	-20,208,430	-5.46*
Banos	11,641,530	4.59
Unidad	-380,480	-0.08
Características	1,990,593	1.79**

\* Significativo al 99%

\*\* Significativo al 90%

Tabla 7. Coeficientes OLS

Los resultados muestran una leve mejora con respecto a la del modelo OLS en cuanto al nivel de ajuste. Esta mejora se sustenta principalmente en la adición del rezago espacial del precio al modelo, lo que mejora el ajuste del mismo. Por otra parte los criterios de Akaike y de Schwartz presentan valores similares, mientras que por el criterio de Sigma cuadrado el modelo OLS presenta un mejor comportamiento con respecto al modelo con rezago espacial. Sin embargo el test Breusch-Pagan sugiere que existe heteroscedasticidad en el modelo, por tanto puede concluirse que la introducción del término de error espacial mejora el ajuste del modelo y no hacen desaparecer los efectos espaciales.

#### Diagnóstico de heteroscedasticidad

Test	DF	VALUE	PROB
Breusch-Pagan test	12	46.81011	0.0000

#### Diagnóstico de dependencia espacial

Test	DF	VALUE	PROB
Likelihood ratio test	1	4.216	0.0400

3. Desarrollado por GeoDa Center. Se encuentra disponible en <http://geodacenter.asu.edu/>

Luego, se desarrolló un modelo con error espacial con la siguiente especificación:

$$\text{Precio} = f(\text{fecha}, \text{Inmueble}, \text{Edad}, \text{Edad2}, \text{Construcción}, \text{Parqueadero}, \text{Distancia}, \text{Estrato2}, \text{Estrato3}, \text{Baños}, \text{Unidad}, \text{Características}, \text{error})$$

$$\text{Error} = W * \text{error} + \mu$$

Los resultados del modelo con error espacial, muestran que las variables construcción, parqueadero, distancia, estrato2, estrato3, baños y el coeficiente Lambda que contiene el retardo espacial son significativos al 99%, mientras que las variables fecha y edad lo son al 95%. La variable edad2 lo es al 90%.

	<b>Coefficiente</b>	<b>Valor Z</b>
Intercepto	40,758,520	4.52
Fecha	-605,942	-2.32**
Inmueble	3,739,654	0.83
Edad	-967,043	-2.33**
Edad2	14,937	1.82***
Construccion	578,749	8.87*
Parqueadero	705,194	2.74*
Distancia	-16,458,080	-4.11*
Estrato2	-34,700950	-6.50*
Estrato3	-23,862,060	-7.32*
Banos	14,458,130	5.18*
Unidad	1,825,563	0.41
Caracteristicas	1,280,379	1.23
Lambda	-0.44512	-2.57*

\* Significativo al 99%

\*\* Significativo al 95%

\*\*\* Significativo al 90%

*Tabla 7. Coeficientes OLS*

Los resultados muestran una leve desmejora con respecto a la del modelo OLS, con un  $r^2$  de 0.82, la cual se sustenta principalmente en la adición del rezago espacial del error al modelo lo que mejora el ajuste del mismo. De igual forma el test Breusch-Pagan sugiere que en este tipo de modelo también existe heteroscedasticidad a pesar de haber introducido el término de error espacial. Y en el test de razón de verosimilitud, no es significativo. Puede concluirse que la introducción del término de error espacial mejora el ajuste del modelo al tiempo que hace desaparecer los efectos espaciales.

**Diagnóstico de heteroscedasticidad**

Test	DF	VALUE	PROB
Breusch-Pagan test	12	38.2018	0.0001

**Diagnóstico de dependencia espacial**

Test	DF	VALUE	PROB
Likelihood ratio test	1	0.4278	0.5130

Al igual que en el caso del modelo con rezago espacial, este modelo presenta unos valores de los criterios de Akaike y Schwartz similares a los del modelo OLS tradicional, aunque el criterio de sigma cuadrado presenta una leve mejora respecto al modelo OLS.

### 4.3.3 Modelo GWR.

A través del programa GWR se especificó la regresión geográficamente ponderada. Los resultados obtenidos empleando un kernel adaptativo, muestran que utilizando el criterio de minimización de Akaike se obtiene el ancho de banda que minimiza dicho criterio luego de ocho iteraciones, como se observa en la tabla 8:

Bandwidth	AICc
34.412	3,2364.333
49.500	3,243.064
64.588	3,231.664
73.912	3,233.748
58.825	3,233.621
68.149	3,233.581
62.386	3,230.729
61.026	3,231.515

Tabla 8. Ancho de banda

El modelo GWR obtuvo unas mejores medidas de bondad de ajuste en comparación con la regresión por OLS, debido al ajuste del ancho de las bandas que depende del número de puntos. En este caso el modelo explica el 92.3% de la variación en los precios de las viviendas y ajustado por el número de variables, el estadístico  $r^2$  es igual a 0.8762.

Criterio de información de Akaike	3,231.515
Coefficiente de Determinación	0.923
$r^2$ Ajustado	0.8762

Por otra parte, el coeficiente sigma (la raíz cuadrada de la suma de los residuales normalizados al cuadrado) presenta un valor inferior al del modelo econométrico tradicional aunque el criterio de información de Akaike es ligeramente superior.

Source	SS	DF	MS	F
OLS_Residuals	18061509068381200	13		
GWR_Improvement	10687452737961900	22.39	*****	
GWR_Residuals	7374056303710100	53.61	*****	3.4693

Tabla 9. ANOVA

La tabla 9 muestra que el modelo GWR tiene un mejor comportamiento en relación con el modelo global obtenido inicialmente y además el estadístico F indica que el modelo es significativo globalmente.

Test de variabilidad espacial de los parámetros	
Parámetro	P-valor
Intercepto	0.54
Fecha	0.86
Inmueble	0.08
Edad	0.88
Edad2	0.84
Construccion	0.08
Parqueadero	0.48
Distancia	0.00*
Estrato2	0.00*
Estrato3	0.00*
Banos	0.30
Unidad	0.68
Características	0.99

\* Significativo al 99%

Tabla 10. Variabilidad espacial de los parámetros

El examen de la variabilidad espacial de los parámetros locales se realiza a través de un test de Monte Carlo (tabla 10). Los resultados muestran que los parámetros *Distancia*, *Estrato2* y *Estrato3* presentan una variabilidad espacial significativa, lo que implica que hay una alta probabilidad de que no haya ocurrido aleatoriamente, mientras que los parámetros restantes presentan un nivel de variabilidad espacial con alta probabilidad de ser obtenida al azar.

El gráfico 7 muestra en la columna izquierda, planos de cómo varían los coeficientes de las variables de cada una de las ecuaciones ajustadas. Los colores más oscuros representan los valores más bajos de los coeficientes y los más claros los más altos. El cuadrado es indicativo de la estación del metro y las líneas rojas que cruzan los planos corresponden a líneas de nivel que muestran las distorsiones en el terreno. En la columna derecha los planos muestran como varía la significancia de los parámetros de las ecuaciones. El color oscuro representa zonas donde los parámetros no son significativos y el color claro representa zonas donde los parámetros son significativos.

Los resultados del estudio muestran que el coeficiente de la variable *Distancia* (gráfico 7h) es significativo en toda la zona que circunda la estación del metro (lo que se refleja en un color más oscuro). Los coeficientes asociados a la variable distancia, presentan grandes valores en las zonas aledañas a la estación del metro representados por las zonas oscuras ubicadas en la zona superior derecha del plano, mostrando una fuerte influencia de la estación en los precios de las viviendas en un radio de aproximadamente 600 metros, lo cual se explica porque la zona recibe una influencia directa del flujo continuo de pasajeros que emplean el metro y de las mejoras urbanísticas desarrolladas en el entorno de la estación, como perfeccionamientos de andenes, bahías para vehículos y ubicación de semáforos. Sin embargo, la influencia de la estación del metro en los precios de las viviendas se reduce en la medida en que se presentan depresiones naturales del terreno como ocurre con la zona más clara de la parte central derecha del plano y en la medida en que la división de las manzanas es menos clara en unos casos y las vías más

estrechas en otros, como ocurre en toda la margen derecha del mapa, razones por las cuales la influencia de la estación del metro tiende a ser menor en estos lugares.

El coeficiente de la variable *Fecha* (gráfico 7b) muestra el cambio de precio de los inmuebles en la medida en que ha transcurrido el tiempo de realización del avalúo. En el gráfico se aprecia que las zonas ubicadas en la parte izquierda del plano mejoraron sus precios más rápidamente que las zonas con mayor valor y de nivel económico superior, situación que se presentó gracias a las mejoras en materia de seguridad observadas durante los años 2008 y 2009 en las zonas más pobres de Medellín. Fue así como por ejemplo en las comunas nor oriental y sur oriental de la ciudad los precios del suelo residencial crecieron sustantivamente durante este período según la Lonja de Propiedad Raíz de Medellín y Antioquia.

El coeficiente de *Edad* (gráfico 7d) muestra que los mayores valores se encuentran ubicados en la parte central y superior del mapa, que es donde se encuentran ubicadas una buena cantidad de casas de la zona, las cuales tienen edades superiores a 30 años, mientras que en las zonas más claras hay una mayor proporción de apartamentos o son zonas que han sido densificadas más recientemente por ser más alejadas de modo que la edad del inmueble no castiga demasiado el precio de la vivienda.

Por otra parte los coeficientes de la variable *Construcción* (gráfico 7f) muestran las mayores contribuciones al precio total en el sentido izquierda derecha, debido a que las calidades y tipos de las construcciones mejoran en esta vía, pasando de viviendas con paredes sin pintura y baños en mal estado en el extremo izquierdo del plano, a viviendas con acabados de buena calidad, en excelente estado de conservación, con buen diseño, bien iluminados y ventilados, en el extremo derecho, los cuales presentan mayores precios. Por otra parte, los mayores valores se presentan en zonas aledañas a la estación del metro debido a que en esta zona se encuentran casas de grandes áreas y altos precios lo que hace que la variable *Construcción* tenga una alta influencia en el precio total del inmueble.

Para la variable *Parqueadero* (gráfico 7g) que muestra el área construida del mismo, presenta los mayores valores de los parqueaderos en lugares correspondientes a las zonas más económicamente más pudientes del lugar, debido a que en estas zonas deprimidas de la ciudad la presencia de un parqueadero es indicativo de que el inmueble es una vivienda con características superiores al promedio, pues implica que sus ocupantes tienen dinero suficiente para comprar un automóvil. De igual manera, el parqueadero debe tener una característica constructiva similar a la de la vivienda de la cual hace parte, por lo que las viviendas de estrato alto tienen parqueaderos más costosos que aquellas con precios bajos.

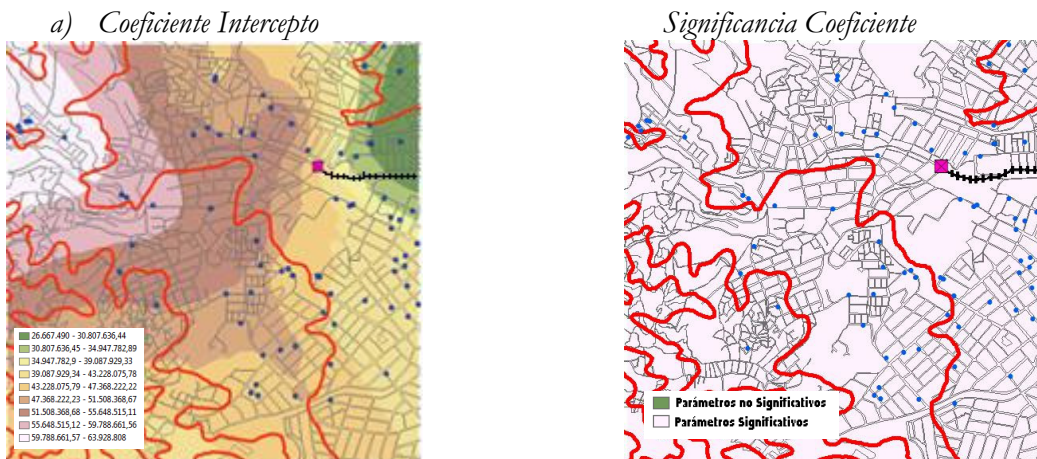
Las variables *Estrato2* (gráfico 7i) y *Estrato3* (gráfico 7j) indican pertenencia de las viviendas a los estratos 2 y 3. La pertenencia a los mismos es determinada por las oficinas de catastro correspondientes para lo cual se califican variables como andenes, vías de acceso y cercanía a focos de contaminación, así como materiales de construcción de fachadas y techos. En la parte izquierda del plano, los coeficientes de las variables muestran una gran influencia en los precios de las viviendas de la zona con coeficientes negativos pues los andenes, las vías, la topografía y los materiales de construcción de las viviendas son de menores especificaciones que las del resto de la zona y este tipo de características exógenas en mal estado como andenes, vías de acceso, etc tienden a reducir el precio de una vivienda.

Los coeficientes de los *Baños* (gráfico 7k) presentan los mayores valores en las zonas superior e inferior izquierda del mapa debido a que en estas zonas el número de baños se constituye en un determinante del estatus de la familia que la habita, pues las viviendas con menor estrato sólo cuentan con un baño llamado baño social, mientras que las viviendas de estrato medio cuentan con baño social y baño en la alcoba de servicio. Por otra parte las viviendas de nivel más alto cuentan con los dos anteriores, más baño en la alcoba principal, por tanto en las zonas más deprimidas el número de baños de una vivienda se convierte en un elemento diferenciador de precios. Caso contrario ocurre con las viviendas ubicadas en la zona centro derecha del plano que son de un nivel económico un poco superior y en dónde el número de baños no se constituye en un determinante importante del precio de las viviendas.

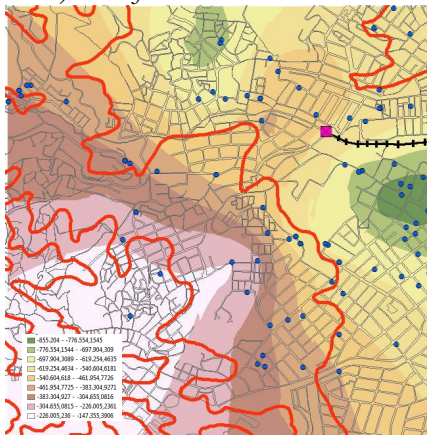
Para el caso de la variable *Unidad* (gráfico 7l), que representa el hecho de que la vivienda se encuentre o no en una unidad cerrada o edificio se presenta una característica importante: en la periferia de la zona, el hecho de estar en unidad cerrada tiende a hacer perder valor a las viviendas pues implícitamente indica que las viviendas pertenecen a la categoría de viviendas de interés social que son productos masivos, de bajas especificaciones y de precios bajos (inferiores a 135 salarios mínimos mensuales legales vigentes), por lo que los coeficientes presentan signos negativos, mientras que en la margen derecha del plano que corresponde a un mayor estrato socioeconómico, el hecho de que una vivienda se encuentre en una unidad cerrada o edificio tiende a agregar valor al precio del inmueble frente a viviendas ubicadas a bordo de vía.

Los coeficientes de la variable *Características* (gráfico 7m) que indican el número de ítems muestran los mayores valores en la zona inferior izquierda del mapa que se observa en color claro, en donde las unidades residenciales están compuestas por casas las cuales poseen juegos de niños, piscinas, canchas de fútbol, entre otros, los cuales tienden a darle mayor valor a las mismas, mientras que en las zonas más oscuras las unidades residenciales corresponden a edificios con características como citófono, ascensor, etc, las cuales aunque son útiles aportan un menor valor a las mismas.

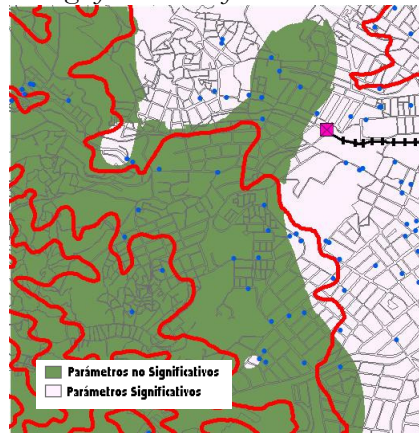
En cuanto al poder explicativo de las ecuaciones, todas las regresiones locales presentan valores entre 86.7% y 91.9%, como consecuencia de la utilización de un kernel adaptativo y de un modelo con variables explicativas adecuadas similares a las empleadas en otros estudios.



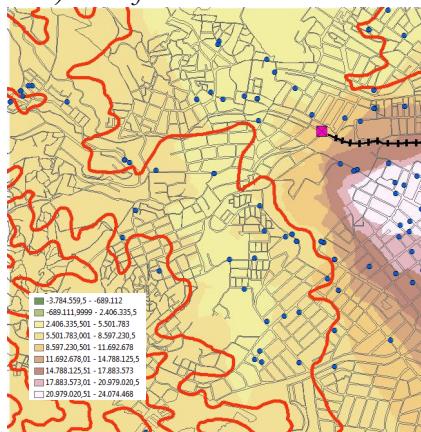
b) *Coefficiente Fecha*



*Significancia Coeficiente*

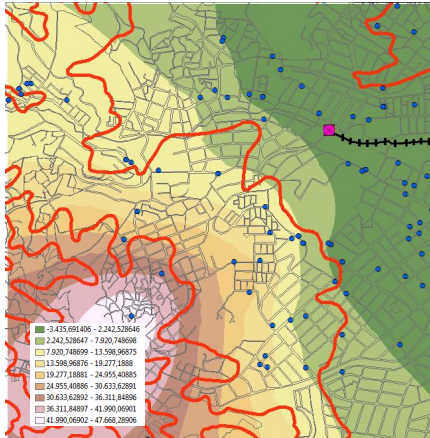


c) *Coefficiente Inmueble*

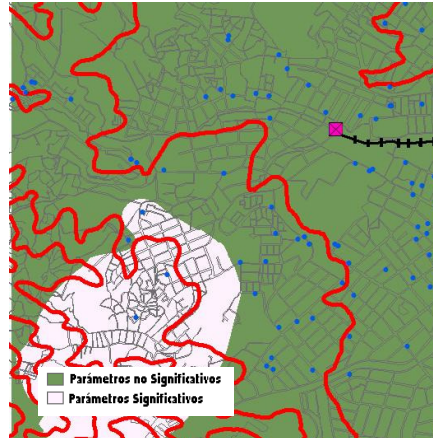


*Significancia Coeficiente*</

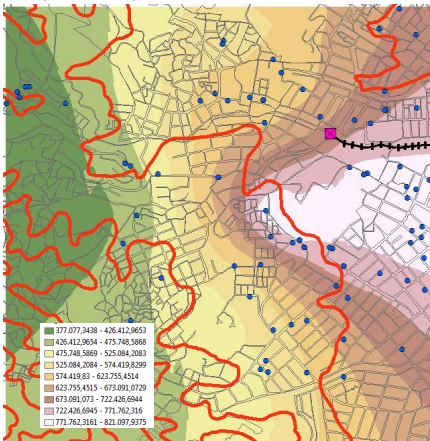
e) *Coficiente Edad2*



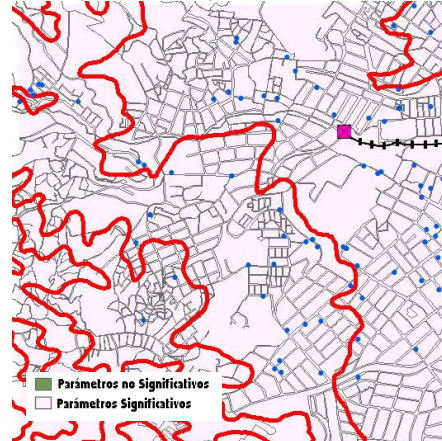
*Significancia Coeficiente*



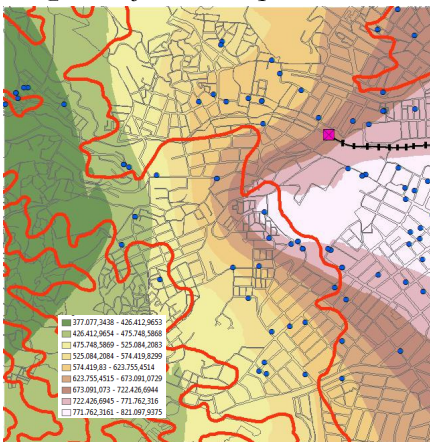
f) *Coficiente Construcción*



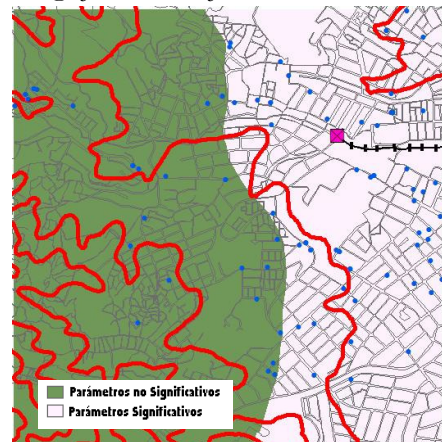
*Significancia Coeficiente*



g) *Coficiente Parqueadero*



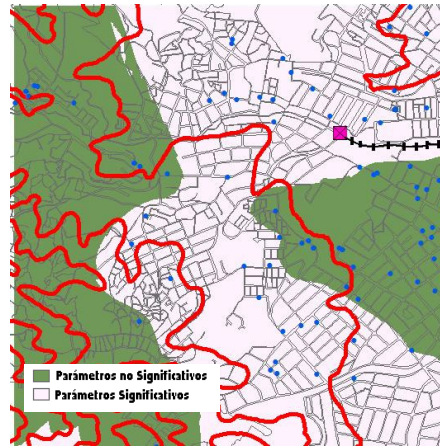
*Significancia Coeficiente*



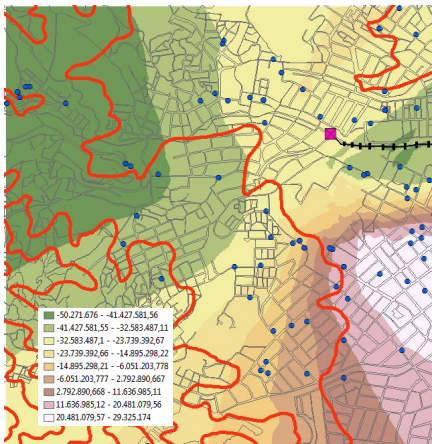
b) *Coficiente Distancia*



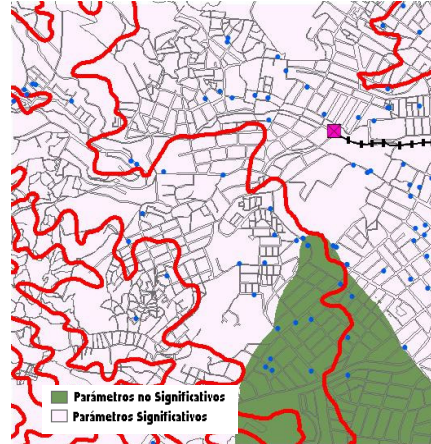
*Significancia Coeficiente*



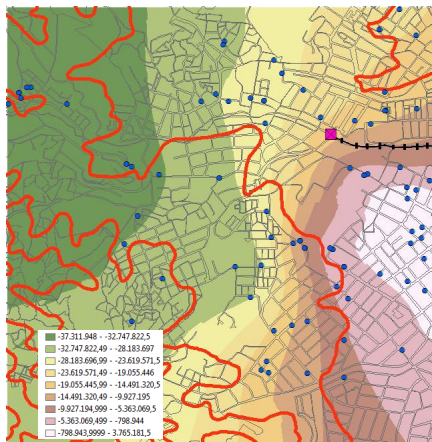
i) *Coficiente Estrato2*



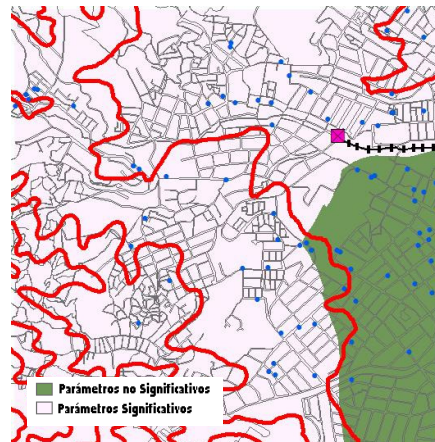
*Significancia Coeficiente*



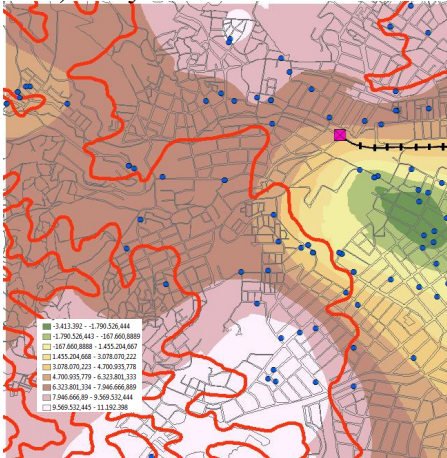
j) *Coficiente Estrato3*



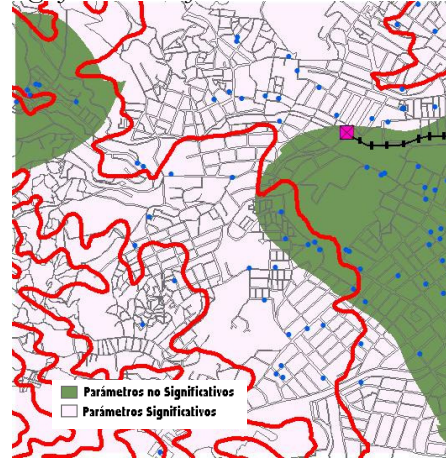
*Significancia Coeficiente*



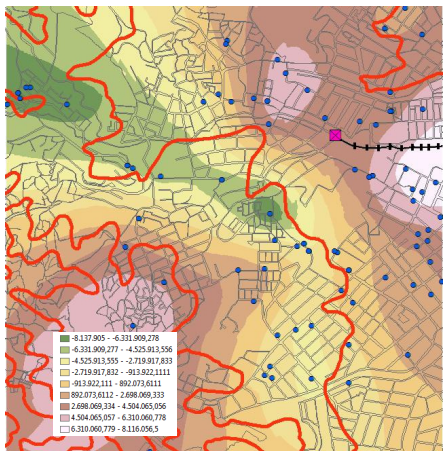
k) *Coefficiente Baños*



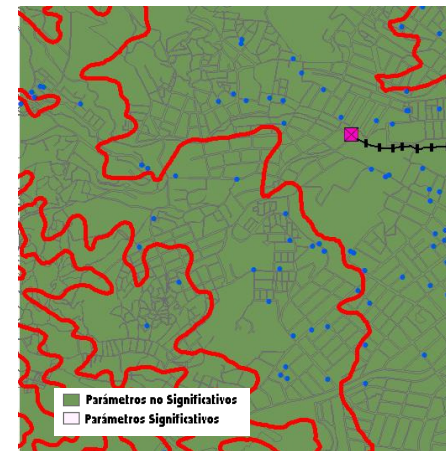
*Significancia Coeficiente*



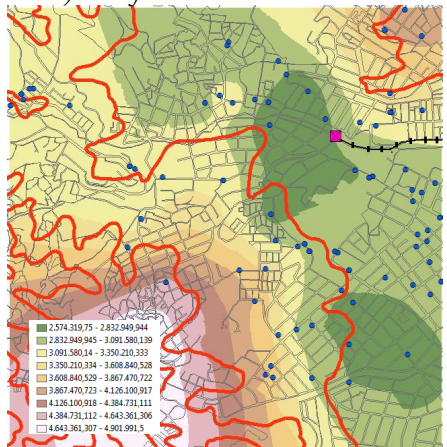
l) *Coefficiente Unidad*



*Significancia Coeficiente*



m) *Coefficiente Características*



*Significancia Coeficiente*

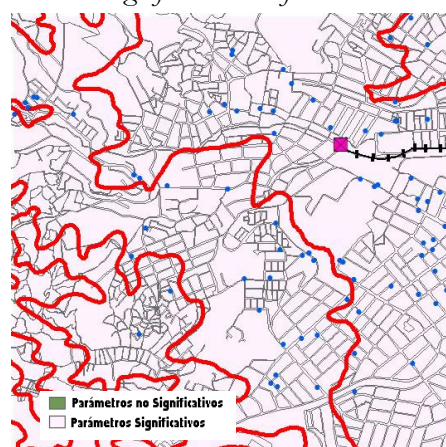


Gráfico 7. Mapas de Coeficientes y de Significancia.

Ahora, con el fin de observar la diferencia entre los resultados obtenidos con la técnica GWR y el método OLS, se evaluaron los resultados obtenidos para una casa de 72,69m<sup>2</sup> construidos, edificada hace 14 años, con un estrato socio económico de 3, que está ubicada en unidad cerrada, posee tres baños y tiene un parqueadero de 11.5m<sup>2</sup>. Los resultados obtenidos pueden verse en el gráfico 8.

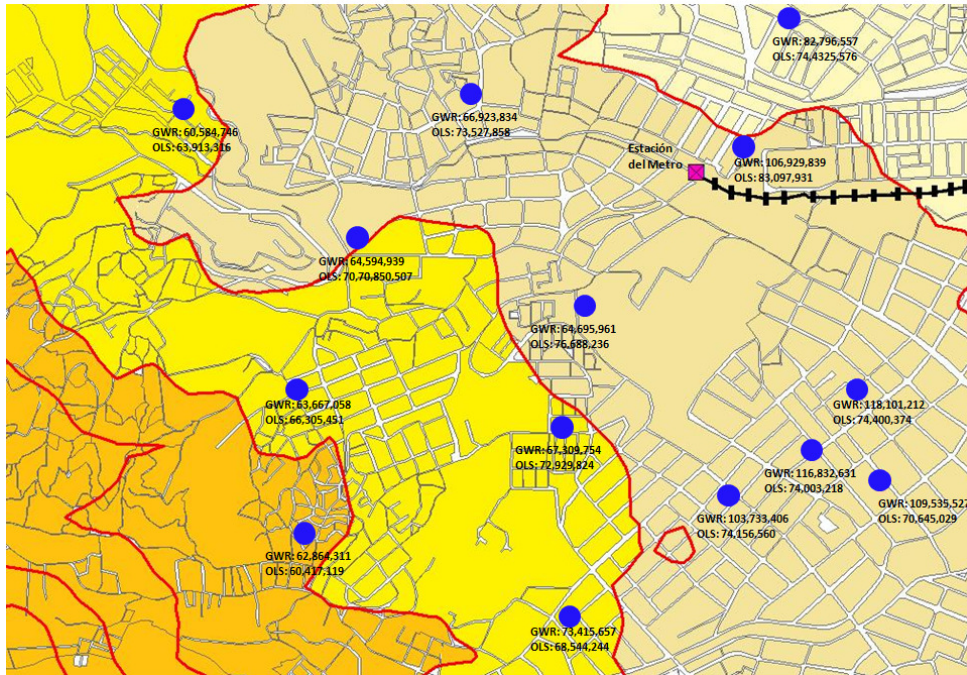
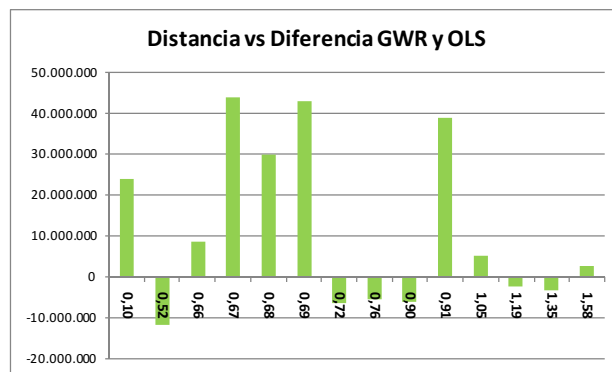


Gráfico 8. Valores de vivienda estimados por GWR y OLS.

El ejercicio consistió en evaluar las características de la vivienda haciendo variar su ubicación en el espacio. En este caso, los valores de la vivienda obtenidos a través de GWR tienden a variar más dramáticamente en el espacio de lo que lo hacen los valores obtenidos por OLS, debido a que con GWR los parámetros cambian continuamente en el espacio, ajustándose a características como el estrato y la edad, mientras que la técnica OLS sólo arroja resultados promedio de las variables, con lo cual los cambios evidenciados son menores. Un ejemplo claro es lo ocurrido con los puntos ubicados en el extremo izquierdo y derecho del mapa. La diferencia entre los dos valores arrojados por GWR es de \$48,950,781, mientras que en el caso del método OLS es de \$6,731,713 lo que ilustra claramente la situación.



*Gráfico 9. Distancia vs diferencia GWR y OLS.*

La diferencia entre los valores obtenidos por GWR y OLS tiende a ser alta en algunos casos como se observa en el gráfico 9, con diferencias cercanas a los cuarenta millones de pesos debido a que en los modelos OLS sólo varía la distancia, mientras en el caso de GWR además de la distancia, varían también los coeficientes. En la tabla 11 pueden verse las diferentes estimaciones y la diferencia entre ambas.

<b>Distancia</b>	<b>Estimación GWR</b>	<b>Estimación OLS</b>	<b>Diferencia</b>
0,10	106.929.839	83.097.931	23.831.907
0,52	64.695.961	76.688.236	-11.992.275
0,66	82.796.557	74.432.576	8.363.981
0,67	118.101.212	74.400.374	43.700.838
0,68	103.733.406	74.156.560	29.576.846
0,69	116.832.631	74.003.218	42.829.413
0,72	66.923.834	73.527.858	-6.604.024
0,76	67.309.754	72.929.824	-5.620.070
0,90	64.594.939	70.850.507	-6.255.568
0,91	109.535.527	70.645.029	38.890.498
1,05	73.415.657	68.544.244	4.871.414
1,19	63.667.058	66.305.451	-2.638.393
1,35	60.584.746	63.913.316	-3.328.570
1,58	62.864.311	60.417.119	2.447.192

*Tabla 11. Estimaciones GWR y OLS.*

## 5. Conclusiones

Este trabajo busca cuantificar la influencia de la estación del metro de San Javier en el precio de las viviendas de la zona, en virtud de su cercanía a la misma. La importancia de este estudio radica en que permite conocer cómo una externalidad positiva creada por una intervención pública generó beneficios a una comunidad con altos índices históricos de violencia y pobreza, además de que permite demostrar que la ejecución de obras públicas puede contribuir a incrementar los precios de los inmuebles, lo que es de importancia para los gobiernos municipales que pueden emplearlo como una herramienta para justificar el uso de políticas públicas como el cobro por concepto de valorización.

A partir de los informes de avalúos de 89 viviendas realizados en la zona, se especificó un modelo hedónico tradicional corregido por heteroscedasticidad que presenta unos buenos resultados reflejados en un valor  $r^2$  elevado de 0.81 y en los signos esperados de las variables, con excepción de la variable unidad que indica si la vivienda se encuentra en unidad cerrada o en edificio la cual presenta un signo negativo, sin embargo como se explicó, las unidades cerradas pueden considerarse un símbolo de estatus en otras zonas de la ciudad, pero en una zona con un nivel económico bajo, la presencia de algunas unidades residenciales está asociada a viviendas de interés social que son productos masivos, de bajas especificaciones y de precios bajos. El modelo hedónico tradicional permite corroborar la hipótesis de la influencia positiva de la estación del metro en los valores de las viviendas: un incremento de un kilómetro en la distancia a la estación del metro genera, en promedio, una reducción de un 20% en el precio total de la vivienda manteniendo todo lo demás constante.

Posteriormente, a través del uso del estadístico I de Moran, se determinó la existencia de autocorrelación espacial en la variable precio, sin embargo, al especificar modelos OLS espaciales los resultados obtenidos en cuanto a poder de predicción son similares a los del modelo hedónico tradicional con valores de  $r^2$  de 0.82 para el modelo de error espacial y de 0.84 para el modelo de rezago espacial, lo que no representa una mejora significativa con respecto al modelo OLS especificado inicialmente. En ambos modelos la variable distancia que mide la influencia de la estación del metro sobre el precio total de las viviendas es significativa, en el caso del modelo con retardo espacial el valor del coeficiente distancia indica que una vivienda ubicada a un kilómetro de distancia de la estación del metro, tiene en promedio, un precio inferior en \$14,116,560 que una que esté ubicada un kilómetro más cerca. En el caso del modelo con error espacial, dicho coeficiente tiene un valor de \$-16,458,080 con igual interpretación que el coeficiente anterior.

Por último se especificaron modelos locales a través de la técnica GWR los cuales presentan valores  $r^2$  altos y superiores a los de las técnicas analizadas, los cuales oscilan entre 0.86 y 0.93 y muestran que los coeficientes de las distintas variables presentan grandes variaciones en su magnitud a través del espacio, justificando el empleo de GWR en el estudio. El coeficiente de la variable distancia resulta ser significativo en el modelo global e indica que una vivienda

ubicada a un kilómetro de distancia de la estación del metro tiende a tener en promedio un precio total inferior en \$15,334,198, que si estuviera ubicada un kilómetro más cerca del mismo. Sin embargo, al aplicar la técnica GWR se observa que para los puntos más cercanos a la estación del metro el coeficiente distancia adquiere valores que oscilan entre \$50,510,286 y \$60,359,631 para las viviendas más cercanas y de entre \$3,779,116 y \$4,265,097 para las más alejadas, mostrando la influencia que tiene en el precio total la ubicación de las viviendas cerca al metro, en virtud de la facilidad de transporte y las mejoras urbanísticas que trae consigo la estación.

Finalmente, otras variables como por ejemplo construcción y parqueadero presentan valores ascendentes en sentido oriente occidente mostrando que en la medida en que las viviendas se alejan de la periferia hacia el oriente estas características tienden a tener un peso mayor en el precio de las viviendas. Mientras que en el caso de la variable unidad la cual indica si una vivienda se encuentra o no en unidad cerrada se presenta el caso contrario, en la periferia presenta valores negativos mientras que en la medida que se aleja de ella los coeficientes comienzan a ser positivos, mostrando que el hecho de estar en unidad cerrada tiende a hacer perder valor a las viviendas en la franja oriental pues implícitamente indica que los inmuebles pertenecen a la categoría de vivienda de interés social que es un producto masivo, de bajas especificaciones y bajo precio, mientras que en la zona occidental que es de mayor estrato el hecho de que una vivienda se ubique en una unidad cerrada tiende a agregar valor al precio del inmueble.

Para futuros estudios a desarrollar en la zona, es importante conseguir un tamaño de muestra mayor, porque aunque los datos son difíciles de conseguir por la situación de seguridad que aún hoy vive el país, esto permitiría detallar con mayor precisión las zonas de isovalores de los coeficientes de manera que los resultados sean más exactos. Por otra parte, quedan abiertas las posibilidades de continuar desarrollando este tipo de estudio tomando como muestra todo el sistema metro o aplicarlo a las viviendas ubicadas cerca de las estaciones de metrocable.

## 6. Bibliografía

Agudelo, G. (2010). *Dependencia Espacial: Detección, Validación y Modelación*. Tesis Maestría en Matemáticas Aplicadas. EAFIT.

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Dordrech: Kluwer.

Anselin, L. (2003). Spatial externalities, spatial multipliers, and spatial econometrics. *International Regional Science*. 26:153–166.

Basu, S., Thibodeau, TG. (1998). Analysis of spatial autocorrelation in house prices. *Journal of Real Estate Finance and Economics*, 17:61-85.

Beatty, J. (1952). Rental real estate often a good investment. *Med Econ*. 5(6): 93 – 94.

Bitter, C., Mulligan, G., & 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.

Brunsdon, C., Fotheringham, S., Charlton, M. (1998). Geographically weighted regression – modeling spatial non-stationarity. *The Statistician*. 47(3): 431-443.

Can, A. (1992). Specification and estimation of hedonic house Price models. *Regional Sciences and Urban Economics*, 22:453-474.

Carroll, T., Clauretje, T., Jensen, J. (1996). Living next to godliness: Residential property values and churches. *Journal of real estate finance and economics*. 12(3): 319 – 330.

Cho, S., Bowker, J., Park, W. (2006). Measuring the Contribution of Water and Green Space Amenities to Housing Values: An Application and Comparison of Spatially Weighted Hedonic Models. *Journal of Agricultural and Resource Economics*. 31(03): 485 – 507.

Dewey, L., DeTuro, P. (1950). Should I invest in real estate?. *Med Econ*. 28(3): 85 – 93.

Dubin, R., Goodman, A. (1982). Valuation of education and crime neighborhood characteristics through hedonic housing prices. *Population and environment*. 5(3): 166 – 181.

Fotheringham, S., Brundson, C., & Charlton, M. (2002). *Geographically weighted regression & associated techniques*. Chichester: Wiley.

Fotheringham, S., Brunsdon, C., Charlton, M. (2002). *Geographically Weighted Regression: the analysis of spatially varying relationships*. Chichester: Wiley.

- Frech, H., Lafferty, R. (1984). The effect of the California coastal commission on housing prices. *Journal of Urban Economics*. 16(1): 105 – 123.
- Grudnitski, G., Do, Q. (1997). Adjusting the value of houses located on a golf course. *The appraisal Journal*. 65(3): 261-266.
- Lancaster, K. (1966), A new approach to consumer theory. *Journal of Political Economy*. 74(1): 132 – 157.
- LeSage, JP. (2004). A family of geographically weighted regression models. *Advances in spatial econometrics. Methodology, tools and applications*. 241-264.
- LeSage J., Peace, K (2004). Spatial statistics and real estate. *Journal of Real Estate Finance Economics*. 29:147–148.
- LeSage, J., Peace, K (2009). *Introduction to Spatial Econometrics*. CRC press.
- Medina, Carlos; Morales, Leonardo y Jairo Nuñez (2010). Quality of life in urban neighbourhoods of Bogotá and Medellín. The Quality Life in Latin American Cities: Markets and Perceptio. Editado por Eduardo Lora, Andrew Powell, Bernard M.S. van Praag y Pablo Sanguinetti, BID y BM.
- Mei, C., He, S., Fang, T. (2004). A note on the mixed geographically weighted regression model. *Journal of Regional Science*. 44(1): 143-157.
- Mennis, J. (2006). Mapping the results of geographically weighted regression. *Cartographic Journal*, 43(2): 171-179.
- Pace, R., Kelley, R., Sirmans, C. (1998). Spatial Statistics and Real Estate. *Journal of Real Estate Finance and Economics*. 17: 5–13.
- Pavlyuk, D. (2009). Statistical analysis of the relationship between public transport accesibility and flat prices in Riga. *Transport and telecommunication*. 10(02): 26 – 32.
- Portney, P. (1981). Housing prices, health effects, and valuing reductions in risk of death. *Journal of environmental economics and management*. 8(1): 72 – 78.
- Rosen, S. (1974). Hedonic prices and implicit markets: product differentiation and pure competition. *Journal of Political Economy*. 82: 34 – 55.
- Yu, D. (2004). Modeling housing market dynamics in the city of Milwaukee: a geographically weighted regression approach. En internet: <http://www.ucgis.org/ucgisfall2004/studentpapers/files/danlinyu.pdf>