

Taller 3: Rosen Roback¹

Economía Urbana
2024 – 20

*Sergio Sandoval
María Fernanda Blanco*

Ejercicio 1

En este ejercicio trabajaremos sobre la elaboración de un índice de precios anual con una submuestra de propiedades en Cook County, Illinois. Para ello, utilice los datos provistos en Bloque Neón (db_ejercicio1.Rds).

1. Construya el índice de precios anual bajo las siguientes metodologías:
 - a) Use la metodología de índice hedónico.
 - b) Use la metodología del estimador de ventas repetidas explicado en clase y utilizado para construir el IPVU.
 - c) Use un estimador de efectos fijos con los errores agrupados (clustered) al nivel de propiedad.
2. Presente los resultados en una gráfica.
3. Compare el desempeño de cada uno.
4. Discuta las ventajas y desventajas de cada enfoque.

Solución:

Medir los precios de las viviendas a lo largo del tiempo es crucial tanto para diseñar políticas públicas como para la toma de decisiones en el sector privado. Desde una perspectiva de política pública, conocer cómo evoluciona el valor de la vivienda permite entender mejor el impacto de políticas de acceso a la vivienda, planificar programas de subsidios o diseñar regulaciones para mitigar crisis en el mercado, como la crisis inmobiliaria de 2008. Por otro lado, desde el sector privado, disponer de estimaciones confiables del valor de la vivienda permite a inversores y desarrolladores tomar decisiones informadas sobre dónde y cuándo invertir. Sin embargo, para que estas decisiones sean adecuadas, es fundamental que las estimaciones sean precisas, ya que cualquier sesgo podría generar decisiones incorrectas, tanto en términos de política como de inversión. Un desafío clave es controlar los efectos de factores fijos, como la ubicación de las propiedades o las amenidades locales, que pueden influir en los precios a lo largo del tiempo. Si, por ejemplo, no controlamos adecuadamente por el efecto fijo de una zona específica, podríamos atribuir cambios en el precio a factores erróneos, cuando en realidad se deben a características permanentes de esa área. De igual manera, es importante aislar el efecto de las amenidades, que pueden influir en los precios de manera heterogénea dependiendo del año o la región. Dado que es imposible observar y controlar todas las variables que pueden afectar simultáneamente el precio de la vivienda y el año de venta, se requieren metodologías robustas que minimicen este tipo de sesgos. Para abordar este desafío, en este análisis se consideran tres metodologías: el índice hedónico, el estimador de ventas repetidas y un estimador de efectos fijos con errores agrupados a nivel de propiedad.

¹Link del repositorio: https://github.com/setosandoval/Econ_Urbana_Grupo4_2024.git

Índice Hedónico

Para empezar se el usa marco teórico del modelo hedónico, en particular, se pretende estimar el índice hedónico. Esta es una metodología ampliamente utilizada en economía urbana para medir cómo cambian los precios de bienes heterogéneos, como las propiedades inmobiliarias, descomponiendo el precio total en los valores implícitos de sus atributos individuales. La idea detrás de este enfoque es que los precios de las viviendas no solo dependen del mercado en general, sino también de las características específicas de cada propiedad, como el tamaño, el número de habitaciones, la ubicación, la antigüedad, y otros atributos físicos o de localización. Para estimar un índice hedónico, se utiliza una regresión donde el precio de las viviendas (normalmente en logaritmos) se explica mediante una combinación de variables que describen estas características. Al incluir todas las características relevantes de las viviendas, el modelo permite construir un índice que captura la evolución del precio puramente debido a factores de mercado, eliminando las variaciones en las características del inventario de viviendas a lo largo del tiempo. Este enfoque se vuelve especialmente útil en mercados donde las viviendas varían considerablemente entre sí, permitiendo obtener una medida ajustada y precisa de los cambios de precios.

Esta metodología presenta varias ventajas, una de sus principales fortalezas es que permite incluir una gran cantidad de variables de control, lo que permite descomponer el precio de las viviendas y analizar el impacto de cada atributo individualmente. Esto es particularmente útil cuando se tiene una base de datos con muchas observaciones y con gran variabilidad en las características de las propiedades. Además, este enfoque permite mantener la mayor cantidad posible de observaciones, ya que no es necesario restringir el análisis a las propiedades que se vendieron más de una vez (como en el caso del estimador de ventas repetidas). Sin embargo, este método presenta*desventajas importantes. Una de las más críticas es que no controla bien por efectos fijos, es decir, por características inobservables de las propiedades que no cambian a lo largo del tiempo pero que afectan su valor, como puede ser el entorno o la calidad intrínseca de la construcción. Esto introduce el riesgo de sesgo si estas características inobservables están correlacionadas con las variables que se incluyen en la regresión. Además, debido a la inclusión de un gran número de controles, existe el riesgo de sobreajustar el modelo y perder grados de libertad, lo que podría reducir la significancia estadística de los estimadores. Finalmente, una desventaja particular en datos de panel es la posibilidad de una doble contabilidad de los efectos fijos de las propiedades que se venden varias veces, lo que puede distorsionar las estimaciones.

Para el presente caso, se plantea hacer uso de la metodología del índice hedónico utilizando la base de datos proporcionada con ventas de viviendas en el condado de Cook, Illinois, entre los años 2000 y 2020. El objetivo es estimar el precio de venta de las propiedades en función de diversas amenidades. Además, se incluye el año de venta como una variable clave, lo cual permite capturar el efecto fijo de tiempo, es decir, las variaciones anuales en los precios que afectan a todas las viviendas por igual, lo que nos da una visión ajustada de la evolución del mercado inmobiliario en este período. La especificación se presenta a continuación:

$$\log(\text{PrecioVenta})_{it} = \alpha + \sum_{k=2001}^{2020} \beta_k \mathbb{1}[\text{Año}_t = k] + X'_i \gamma + \epsilon_{it}$$

La ecuación especificada regresa el logaritmo del precio de venta de la propiedad i en el año t contra una serie de variables indicadoras para cada año desde 2000 hasta 2020 (con año 2000 como base), las cuales indican en que año fue vendida cada propiedad. El vector de controles X'_i contiene una serie de variables asociadas a amenidades de la propiedad, que busca aislar lo más posible el cada efecto de cada año. Estas incluyen: año de construcción, pies cuadrados del edificio, pies cuadrados del terreno, número de dormitorios, número de habitaciones, número de baños con y sin ducha, número de chimeneas, tipo de residencia, calidad de la construcción, si tiene garaje adjunto y su tamaño, si tiene calefacción central, si tiene aire acondicionado y si ha tenido renovación en los últimos 3 años. Cada coeficiente β_k asociado a las variables de año captura la variación en los precios de las propiedades en ese año específico, comparado con el año base. El coeficiente de interés es precisamente el conjunto de estos coeficientes de año, ya que al construir el índice hedónico, se busca evaluar cómo han evolucionado los precios de las propiedades a lo largo del tiempo, manteniendo constantes las características físicas y locativas de las propiedades.

El modelo hedónico planteado requiere una serie de supuestos para que esté correctamente identificado y para que los resultados obtenidos sean válidos y confiables. El supuesto más importante es el de exogeneidad, que implica que las variables explicativas (como las amenidades de las propiedades y el año de venta) no estén correlacionadas con el término de error ϵ_{it} . Esto asegura que los coeficientes estimados capturen el verdadero efecto de las amenidades y del tiempo sobre el precio de la vivienda, sin sesgos causados por variables omitidas o factores inobservables. También, es crucial que las amenidades de las propiedades y los efectos de cada año sean exógenos, es decir, que no estén influenciados por factores inobservables que afecten simultáneamente el precio y las variables incluidas. Sin embargo, este supuesto puede ser difícil de cumplir completamente, especialmente en un mercado inmobiliario dinámico, donde variables inobservables como las decisiones de urbanización o cambios en las preferencias de los compradores podrían influir tanto en los precios como en la selección de características de las propiedades. Además, otro supuesto relevante en este contexto es que las características inobservables de las propiedades que no varían a lo largo del tiempo, como la ubicación exacta o la calidad de construcción intrínseca, se consideran constantes y se controlan a través de los efectos fijos de tiempo (año de venta). No obstante, este modelo no incluye efectos fijos a nivel de propiedad, lo que podría implicar que algunas de estas características invariables afecten los resultados. Por último, es importante asumir que no existe autocorrelación en los errores, lo que significaría que el término de error de una propiedad en un año determinado no esté correlacionado con los errores de años anteriores para la misma propiedad. Este supuesto es crucial en modelos de datos panel para garantizar la validez estadística de los estimadores, y podría ser violado si existen factores inobservables persistentes que afectan los precios de las propiedades a lo largo del tiempo.

IPVU

El IPVU (Índice de Precios de la Vivienda Usada) es una metodología utilizada para calcular la evolución de los precios de las viviendas basándose en las ventas repetidas de la misma propiedad a lo largo del tiempo. Este enfoque utiliza únicamente las propiedades que se venden más de una vez durante el período de estudio, lo que permite comparar directamente el precio de una propiedad en diferentes momentos, controlando automáticamente por las características invariables de la vivienda (como ubicación, tamaño o calidad de construcción). Al observar las mismas propiedades en distintas fechas, se puede aislar el efecto de los cambios en el mercado inmobiliario o en las condiciones económicas generales, evitando los sesgos que podrían surgir al comparar diferentes tipos de viviendas con características heterogéneas. El IPVU es particularmente útil cuando se busca medir la apreciación o depreciación de los precios de viviendas sin que las características físicas de las mismas influyan en las estimaciones.

El IPVU se destaca por utilizar directamente los valores observados de las propiedades que se han vendido al menos dos veces durante el período de estudio. Esto lo convierte en un índice de precios de calidad constante, ya que las propiedades seleccionadas no deberían haber sufrido modificaciones significativas entre las transacciones, lo que permite controlar con mayor precisión las características de la vivienda. A diferencia del índice hedónico, que debe estimar el valor implícito de cada atributo de la vivienda, el IPVU utiliza información directa de las ventas repetidas, lo que reduce la dependencia en estimaciones y aumenta la precisión de los resultados. Además, su reproducibilidad es una ventaja clave: diferentes analistas, con los mismos datos de ventas, llegarán a los mismos estimadores de cambio de precios, lo que lo convierte en un método más robusto y confiable.

Sin embargo, este método tiene desventajas importantes. En primer lugar, solo utiliza una parte de las transacciones disponibles: aquellas propiedades que se han vendido más de una vez. Esto deja de lado muchas propiedades que pueden ser igualmente representativas del mercado, como las viviendas que solo han tenido un único dueño. Además, el método de ventas repetidas no controla por cambios en la vivienda que pueden ocurrir con el tiempo, lo que viola el supuesto de que las características de la propiedad no cambian entre transacciones. De hecho, la antigüedad de la vivienda es una fuente de sesgo en este método, ya que con el tiempo, una vivienda podría depreciarse o renovarse, afectando su precio. El IPVU tampoco captura tendencias del mercado que influyen en la demanda que pueden no estar presentes en las propiedades de ventas repetidas, pero que son factores clave en el mercado inmobiliario.

En este caso, se utilizó el método de ventas repetidas para construir el índice de precios de vivienda. A diferencia del modelo hedónico, no se incluyeron controles para las amenidades de las propiedades, ya que el enfoque se basa en comparar el precio de la misma propiedad en dos o más transacciones, lo que controla implícitamente por las características invariables de la vivienda. El proceso de estimación del IPVU es más complejo y no se puede expresar con una simple especificación de regresión, ya que se realiza en varias etapas. Primero, se identifica el subconjunto de propiedades que se han vendido más de una vez. Luego, en una segunda etapa, se comparan los precios de venta de cada propiedad en diferentes momentos, ajustando por el tiempo transcurrido entre las transacciones. Finalmente, se estiman los coeficientes de apreciación o depreciación de las propiedades para cada año, a partir de las cuales se construye el índice general. Esta metodología permite obtener una medida más directa de la evolución de los precios de las propiedades a lo largo del tiempo, sin necesidad de introducir amenidades o características específicas como variables de control.

El método de ventas repetidas se basa en una serie de supuestos que son cruciales para garantizar la validez de las estimaciones. Uno de los más importantes es que las características de las propiedades no cambian significativamente entre las transacciones. Este supuesto es esencial para que el modelo controle correctamente los efectos fijos de cada vivienda. Si una propiedad sufre renovaciones, mejoras o deterioros entre ventas, el cambio en su precio podría atribuirse erróneamente al mercado en lugar de a las modificaciones en la propiedad. Sin embargo, en la práctica, las propiedades pueden cambiar con el tiempo. Otro supuesto es que las ventas repetidas son representativas del mercado inmobiliario en general. Esto puede no ser siempre cierto, ya que las propiedades que se venden más de una vez podrían tener características particulares (por ejemplo, ser más atractivas para inversores o estar ubicadas en áreas más dinámicas). Además, se asume que el valor de las propiedades no se ve afectado por factores temporales externos que varíen entre transacciones, como cambios en las preferencias del mercado por características específicas (como más baños o garajes) o políticas públicas que afecten ciertos barrios. Finalmente, otro supuesto importante es la exogeneidad del momento de venta, es decir, que la decisión de vender una propiedad en un momento particular no esté correlacionada con variables no observadas que también afecten su precio. Si los propietarios deciden vender debido a condiciones económicas particulares que no están observadas en el modelo, esto puede introducir sesgos en las estimaciones.

Efectos Fijos

Los modelos de efectos fijos son una herramienta poderosa en econometría para controlar por todas las características invariables de una unidad de observación a lo largo del tiempo, ya que permiten eliminar cualquier sesgo causado por variables no observadas que son constantes. En lugar de tener que especificar todas las características de las propiedades que podrían influir en los precios, los efectos fijos permiten capturar todas esas características intrínsecas (como la ubicación o la calidad de la construcción) que no varían entre las diferentes observaciones de una misma unidad. Además, cuando los errores de las observaciones no son independientes dentro de una misma unidad de observación, se utilizan errores agrupados cluster, que permiten corregir la correlación dentro de los grupos (en este caso, propiedades), garantizando que las inferencias estadísticas sean robustas frente a la autocorrelación o la heterocedasticidad dentro de esos grupos.

En el contexto del presente análisis, el uso de efectos fijos a nivel de propiedad es particularmente ventajoso porque elimina el sesgo relacionado con las características fijas de las viviendas que no cambian con el tiempo, proporcionando una estimación más precisa del cambio en los precios. Al mismo tiempo, los errores agrupados por propiedad aseguran que cualquier correlación entre las múltiples observaciones de una misma vivienda no distorsione los resultados. Sin embargo, una desventaja clave de este enfoque es que puede reducir la variabilidad en los datos, dado que solo utiliza las variaciones dentro de las propiedades a lo largo del tiempo, y no entre diferentes propiedades. Además, como sucede en el modelo hedónico, la necesidad de controlar por un gran número de características temporales (como el año de venta o las condiciones del mercado) sigue siendo necesaria, lo que puede complicar la especificación del modelo y limitar el número de observaciones útiles. La especificación es casi la misma que para el índice hedónico, solo que se contempla el efecto fijo θ_i por unidad i .

$$\log(\text{PrecioVenta})_{it} = \theta_i + \sum_{k=2001}^{2020} \alpha_k \mathbb{1}[\text{Año}_t = k] + Z'_i \delta + \mu_{it}$$

Se incluyen los mismos controles que permiten aislar el efecto de las características de las propiedades. La diferencia clave es que ahora se agregan efectos fijos a nivel de propiedad, lo que significa que cualquier característica constante en el tiempo (como la ubicación o la calidad de construcción) será capturada por el término de efectos fijos. En la práctica, esto implica que si alguna de estas variables incluidas en el modelo es fija (por ejemplo, una característica constante de la propiedad), será omitida automáticamente en la programación debido a que los efectos fijos ya la capturan. Incluir estos controles sigue siendo importante para asegurar una comparación directa y justa con el modelo hedónico, dado que ambas metodologías están intentando capturar el mismo fenómeno de variación de precios a lo largo del tiempo.

En cuanto a los supuestos necesarios, la mayoría son los mismos que en el modelo hedónico. Se sigue asumiendo exogeneidad, lo que significa que las variables incluidas en el modelo (amenidades y años) no deben estar correlacionadas con inobservables u observables no incluidas. Sin embargo, el supuesto clave adicional aquí es que las características fijas de las propiedades no cambian a lo largo del tiempo, lo que es fundamental para que los efectos fijos sean efectivos. Este supuesto puede ser razonable en muchos casos, pero no siempre es plausible si las propiedades experimentan renovaciones o mejoras significativas que no son observadas, lo cual podría generar sesgos en las estimaciones. Por otro lado, el uso de errores agrupados (cluster) a nivel de propiedad implica que asumimos que las correlaciones entre las observaciones de una misma propiedad están debidamente corregidas, lo cual es plausible si las observaciones dentro de cada propiedad están bien medidas y los errores no se correlacionan entre propiedades diferentes.

Estimación

Las estimaciones de los coeficientes y del índice de precios anual se presentan en las Figuras 1 y 2. En la Figura 1 se muestran los coeficientes estimados para cada año de venta utilizando las tres metodologías: el índice hedónico, el IPVU y el modelo de efectos fijos con errores agrupados. Cada punto en el gráfico representa el coeficiente estimado del año correspondiente, lo que refleja cómo ha variado el precio de las viviendas respecto al año base (que no se muestra en el gráfico). Las líneas conectan estos puntos para facilitar la visualización de las tendencias temporales en las tres metodologías. En la Figura 2, se representa el índice de precios calculado a partir de estos coeficientes, que refleja la evolución de los precios de las viviendas a lo largo del tiempo para las tres metodologías. El índice se estandariza en un punto de referencia, por lo general el año base, para mostrar la variación relativa en los precios.

Se observa que los coeficientes siguen una tendencia ascendente en las tres metodologías, aunque con algunas variaciones en ciertos años. Las tres líneas muestran incrementos pronunciados en algunos periodos, como entre 2005 y 2007, justo antes de la crisis inmobiliaria, y una caída visible en los años inmediatamente posteriores, entre 2008 y 2010. A partir de 2012, las estimaciones muestran una recuperación gradual de los precios, con los coeficientes manteniéndose en una tendencia positiva. Las metodologías del índice hedónico y del modelo de efectos fijos parecen seguir trayectorias similares, mientras que el IPVU muestra más fluctuaciones en ciertos años. En la Figura 2, el índice de precios sigue una dinámica similar, con aumentos pronunciados hasta 2007, una fuerte caída en los años posteriores a la crisis y una recuperación sostenida después de 2012, lo que refleja un comportamiento consistente en las tres metodologías a lo largo del tiempo.

Figura 1: Comparación coeficientes modelos (2001-2020)

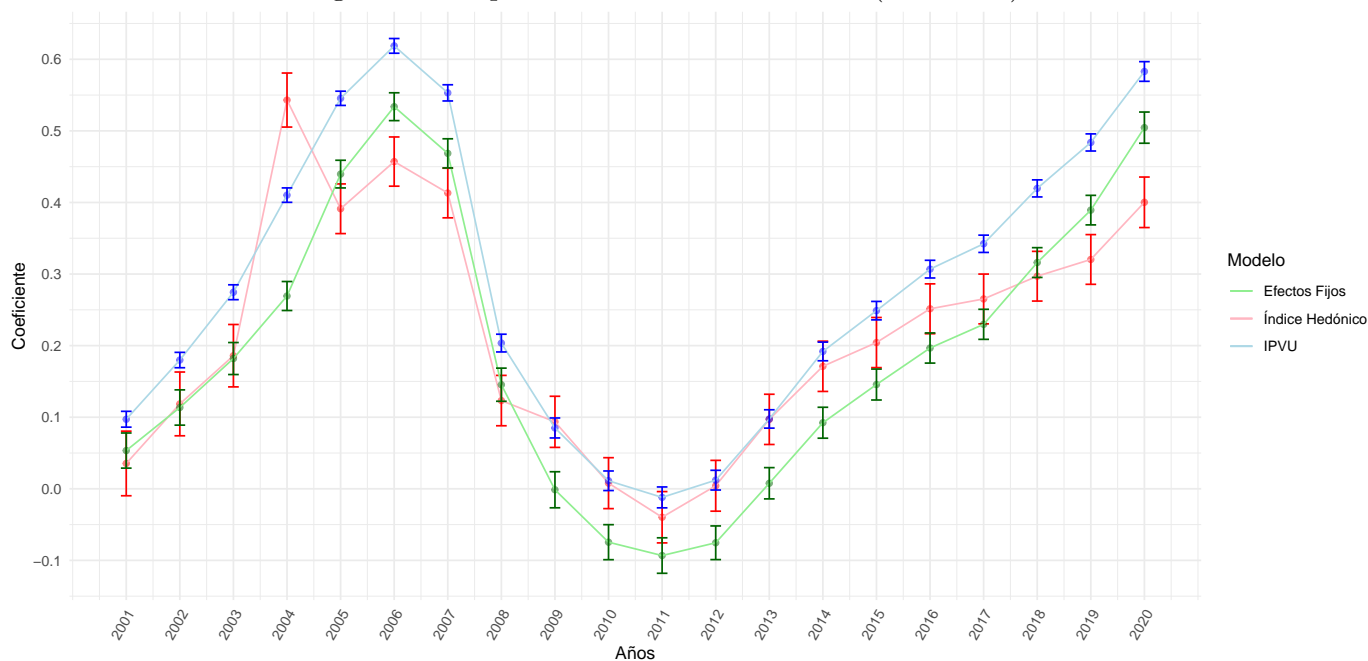
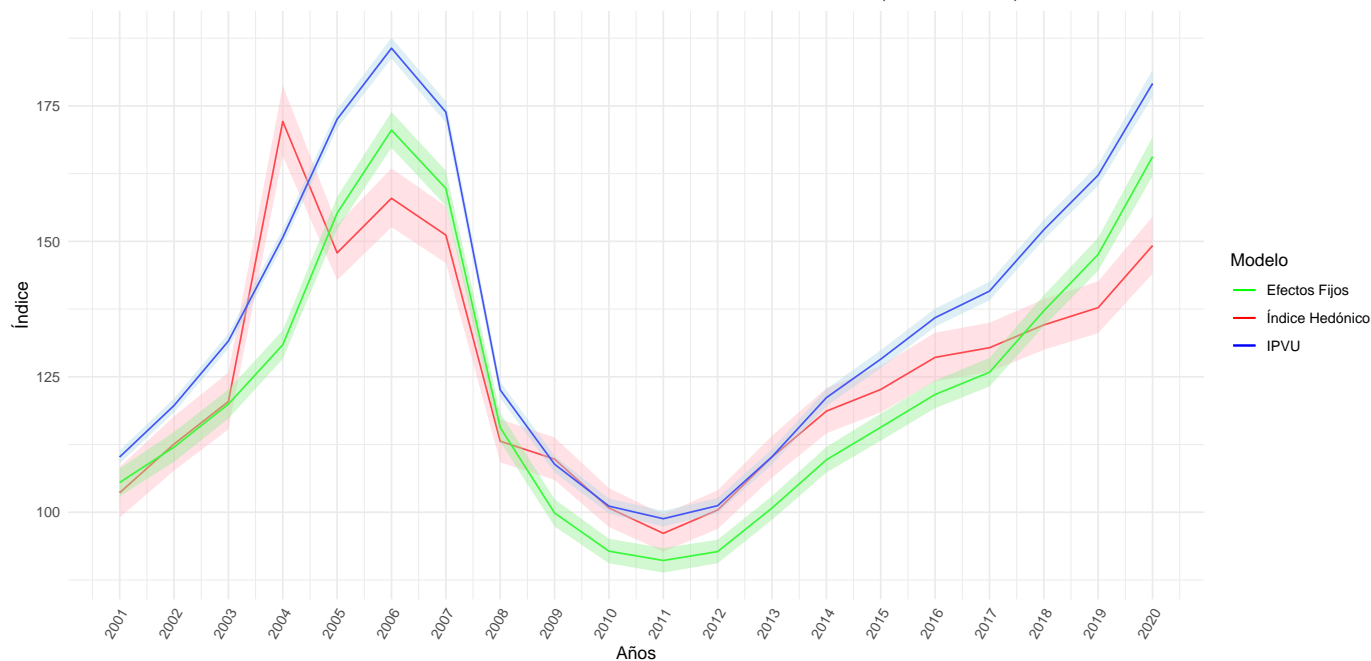


Figura 2: Comparación coeficientes modelos (2001-2020)



En las estimaciones mostradas en la Figura 1, observamos que los coeficientes correspondientes a los diferentes años reflejan la variación en los precios de las viviendas en relación con el año base. Un coeficiente positivo indica que el precio de las viviendas fue mayor en ese año en comparación con el año base, mientras que un coeficiente negativo sugiere que los precios fueron menores. En las tres metodologías, la mayoría de los coeficientes son positivos, lo que indica que los precios de las viviendas experimentaron un incremento sostenido a lo largo del tiempo, excepto durante la crisis financiera entre 2008 y 2010, donde observamos caídas o estancamientos en los precios.

Para el índice hedónico, los coeficientes son positivos en la mayoría de los años, lo que sugiere un efecto positivo en los precios respecto al año base. Sin embargo, los coeficientes para los años 2001, 2010, 2011 y 2012 no son significativos, ya que los intervalos de confianza incluyen el valor cero, lo que indica que no se puede afirmar con certeza que los precios en esos años fueran significativamente diferentes del año base. De manera similar, en el IPVU, los coeficientes son mayoritariamente positivos y significativos, pero los años 2010, 2011 y 2012 no muestran un efecto claro, pues tampoco son significativos. Finalmente, en el modelo de efectos fijos, los coeficientes de 2009 y 2012 no son significativos, lo que nuevamente indica que no hubo un cambio estadísticamente claro en los precios respecto al año base en esos años. Este comportamiento general sugiere que, tras la caída de los precios entre 2008 y 2010, hubo una recuperación paulatina en los años posteriores, con coeficientes positivos que reflejan un aumento en los precios en la mayoría de los años restantes. El hecho de que algunos coeficientes no sean significativos en ciertos años sugiere que durante esos periodos, el mercado inmobiliario podría haber estado estancado o mostrando variaciones menores que no fueron estadísticamente detectables.

El análisis del R^2 revela una clara distinción entre los tres métodos. El modelo de efectos fijos presenta el R^2 más alto, con un valor de 0.759, lo que indica que explica una gran proporción de la variabilidad en los precios de las viviendas. Esto es consistente con el hecho de que los efectos fijos permiten capturar todas las características invariables de las propiedades, eliminando gran parte del sesgo causado por factores no observados, y el clustering asegura que las inferencias sean más robustas, especialmente cuando hay múltiples observaciones por propiedad. Por otro lado, el índice hedónico tiene un R^2 de 0.342, significativamente más bajo, lo que refleja que este modelo no controla directamente por los efectos fijos, y como consecuencia, es menos eficiente al explicar la variabilidad en los precios, particularmente en momentos de cambios bruscos como la crisis de 2008-2010. Finalmente, el IPVU muestra un R^2 de 0.257, el más bajo de los tres, lo que puede explicarse por el hecho de que se basa en un subconjunto de propiedades (aquellas con ventas repetidas), lo que limita su capacidad para capturar toda la dinámica del mercado.

En cuanto a las tendencias, el IPVU y los efectos fijos muestran trayectorias similares, con una tendencia ascendente hasta 2007, una fuerte caída durante la crisis inmobiliaria de 2008-2010, y una posterior recuperación gradual a partir de 2012. Estas dos metodologías capturan de manera consistente las variaciones temporales en los precios, ya que ambos modelos logran controlar adecuadamente por las características fijas de las propiedades. En particular, el IPVU tienden a tener intervalos de confianza mas reducidos, lo cual se explica en parte por menores errores estándar. En contraste, el índice hedónico presenta una mayor variabilidad en los coeficientes y no sigue de manera tan precisa la trayectoria marcada por los otros dos métodos. Esto es particularmente notable durante la crisis financiera y en los años inmediatamente posteriores (2008-2012), donde el índice hedónico muestra fluctuaciones más pronunciadas que los otros dos. Esta mayor variabilidad es esperable, ya que el índice hedónico no controla por efectos fijos de las propiedades, lo que lo hace más vulnerable a cambios en la composición del inventario de viviendas vendidas en diferentes años.

Al observar los resultados, se destacan diferencias notables en las tendencias y niveles entre los tres métodos de estimación. Tanto el modelo de efectos fijos como el IPVU presentan trayectorias similares, capturando de manera coherente los patrones de crecimiento y caída en los precios a lo largo del tiempo, especialmente durante la crisis financiera de 2008-2010 y la recuperación posterior. Sin embargo, el modelo de efectos fijos tiende a estar en niveles más bajos que el IPVU en varios años, lo que puede deberse a la capacidad de los efectos fijos de controlar de manera más estricta las características invariables de cada propiedad, eliminando cualquier sesgo relacionado con variaciones de calidad entre propiedades. El IPVU, al depender de ventas repetidas, refleja propiedades que podrían ser más dinámicas en el mercado, lo que puede explicar por qué sus estimaciones de precios tienden a ser más altas en algunos periodos. En contraste, el índice hedónico presenta más variabilidad en los coeficientes, en parte porque no controla por efectos fijos de propiedad, lo que lo hace más sensible a cambios en la composición de las propiedades vendidas en cada año. En términos generales, las tendencias capturadas en los tres métodos muestran un crecimiento constante en los precios hasta la crisis de 2008, seguido por una caída y una recuperación gradual. El índice de precios en la Figura 2 sigue esta misma dinámica, aunque con diferencias en los niveles según el método, lo que refleja cómo cada modelo trata la evolución de los precios.

En conclusión, el modelo de efectos fijos parece ser la opción más sólida para estimar la evolución de los precios de las viviendas en este contexto. Su capacidad para controlar por características fijas y manejar la correlación dentro de las propiedades lo convierte en la opción más robusta, como lo refleja su R^2 más alto. Aunque el IPVU también controla por características fijas al utilizar ventas repetidas y tiene menor error estándar, sus fluctuaciones en ciertos años y su menor R^2 sugieren que es menos eficiente a la hora de capturar la variación total de los precios, posiblemente debido a la selección limitada de propiedades. Por otro lado, el índice hedónico, aunque tiene la ventaja de utilizar todas las observaciones, es más vulnerable a sesgos, especialmente en periodos de alta volatilidad, como la crisis de 2008-2010. Dependiendo del objetivo del análisis, el modelo de efectos fijos parece ser la mejor opción para precisión y estabilidad, mientras que el IPVU podría ser útil en análisis donde las ventas repetidas tienen un interés particular, y el índice hedónico sería más adecuado cuando se busca aprovechar una mayor cantidad de observaciones.

Ejercicio 2

En este ejercicio derivaremos y estimaremos la importancia de la amenidad A en términos reales en el país de Tangolandia.

1. Derive paso a paso la ecuación (5) de Roback (1982).
2. Con la ecuación del paso anterior y la base (db_ejercicio2.Rds) que contiene información ficticia de arriendos, salarios y amenidad A a nivel ciudad, estime la importancia de la amenidad A en términos reales. (Asuma que la proporción de la tierra en el presupuesto del consumidor es 1/3). Interprete el resultado obtenido.

En su análisis sea cuidadoso en explicitar las regresiones que va a estimar. Explicita los supuestos e interprete los coeficientes. Acompañe todas las estimaciones con sus debidos errores estándar.

Solución:

Roback (1982)

La ecuación (5) de Roback (1982) expresa el precio implícito de una amenidad, el cual refleja el monto de ingresos que compensaría a los individuos por un cambio marginal en la calidad de la amenidad. Esta ecuación combina los efectos en los salarios y en los alquileres de la tierra debido a variaciones en las amenidades locales. Para llegar a esta ecuación, es necesario primero plantear el modelo. Este contempla el mercado de vivienda y el laboral, en donde hay 2 sectores: los consumidores o trabajadores y las firmas. Cada uno se enfrenta a un problema de maximización, en el caso del trabajador representativo:

$$\max U(x, l^c; s) \quad \text{sujeto a} \quad w + I = x + l^c r \quad (1)$$

Donde, l^c es el consumo de tierra, x representa el consumo del resto bienes, s es la cantidad de la amenidad urbana en una ciudad específica, w es el salario, I es el ingreso no laboral, y r es el costo del alquiler de la tierra. Entonces, el problema para el trabajador representativo es, dado la cantidad de s en su ubicación, elegir las cantidades de x y l^c de manera que satisfaga la restricción presupuestaria.

Por su parte, la firma representativa minimiza su costo sujeto a su función de producción. La producción de un bien X depende del uso de tierra productiva l^c y del número de trabajadores N , a través de una función de producción $X = f(l^c, N; s)$.

$$V(w, r; s) = k \quad (2)$$

$$C(w, r; s) = 1 \quad (3)$$

Las condiciones de equilibrio se presentan en la ecuación (2) y (3) del paper. La condición para el trabajador se establece cuando su utilidad indirecta, $V(w, r; s)$, es la misma en todas las ubicaciones, lo que garantiza que no haya incentivos para migrar entre ciudades. Esta ecuación refleja que los salarios w y las rentas r se ajustan de manera que el trabajador sea indiferente entre vivir en distintas ciudades, ajustando por las amenidades s . En el caso de las firmas, la condición de equilibrio se deriva de la minimización de costos. Aquí, $C(w, r; s)$ es la función de costos, y la ecuación refleja que las empresas ajustan los salarios w y las rentas r para igualar sus costos en todas las ciudades en función de la amenidad s .

Para obtener la ecuación (4) del paper, se diferencian las ecuaciones (2) y (3) con respecto a s y se resuelve para las derivadas parciales $\frac{\partial w}{\partial s}$ y $\frac{\partial r}{\partial s}$. Para la ecuación (2) se tiene:

$$\begin{aligned} \frac{\partial V}{\partial w} \frac{\partial w}{\partial s} + \frac{\partial V}{\partial r} \frac{\partial r}{\partial s} + \frac{\partial V}{\partial s} &= 0 \\ V_w w_s + V_r r_s + V_s &= 0 \end{aligned}$$

Resolviendo para w_s se obtiene:

$$w_s = -\frac{V_r r_s + V_s}{V_w}$$

Haciendo el mismo ejercicio para la ecuación (3):

$$\begin{aligned}\frac{\partial C}{\partial w} \frac{\partial w}{\partial s} + \frac{\partial C}{\partial r} \frac{\partial r}{\partial s} + \frac{\partial C}{\partial s} &= 0 \\ C_w w_s + C_r r_s + C_s &= 0 \\ w_s &= -\frac{C_r r_s + C_s}{C_w}\end{aligned}$$

Igualando ambos resultados se obtiene r_s :

$$\begin{aligned}-\frac{V_r r_s + V_s}{V_w} &= -\frac{C_r r_s + C_s}{C_w} \\ V_r C_w r_s + V_s C_w &= V_w C_r r_s + V_w C_s \\ V_s C_w - V_w C_s &= V_w C_r r_s - V_r C_w r_s \\ V_s C_w - V_w C_s &= r_s (V_w C_r - V_r C_w) \\ r_s &= \frac{V_s C_w - V_w C_s}{V_w C_r - V_r C_w} \\ r_s &= \frac{-V_w C_s + V_s C_w}{V_w C_r - V_r C_w}\end{aligned}$$

Para obtener w_s , se reemplaza el r_s encontrado en cualquiera de los w_s despejados de las diferenciaciones, en este caso para la diferenciación de la ecuación (2):

$$\begin{aligned}w_s &= -\frac{V_r r_s + V_s}{V_w} \\ -V_w w_s &= V_r r_s + V_s \\ -V_w w_s &= V_r \left(\frac{-V_w C_s + V_s C_w}{V_w C_r - V_r C_w} \right) + V_s \\ -V_w w_s &= \left(\frac{-V_r V_w C_s + V_r V_s C_w}{V_w C_r - V_r C_w} \right) + V_s \\ -V_w w_s &= \frac{-V_r V_w C_s + V_r V_s C_w + V_s V_w C_r - V_s V_r C_w}{V_w C_r - V_r C_w} \\ -V_w w_s &= \frac{-V_r V_w C_s + V_s V_w C_r}{V_w C_r - V_r C_w} \\ -V_w w_s &= \frac{-V_w (V_r C_s - V_s C_r)}{V_w C_r - V_r C_w} \\ w_s &= \frac{V_r C_s - V_s C_r}{V_w C_r - V_r C_w} \\ w_s &= \frac{-V_s C_r + V_r C_s}{V_w C_r - V_r C_w}\end{aligned}$$

Con lo anterior, se llega a los resultados de la ecuación (4) del paper:

$$\begin{aligned}w_s &= \frac{\partial w}{\partial s} = \frac{1}{\Delta} (-V_s C_r + V_r C_s) \\ r_s &= \frac{\partial r}{\partial s} = \frac{1}{\Delta} (-V_w C_s + V_s C_w)\end{aligned} \tag{4}$$

Donde $\Delta = V_w C_r - V_r C_w$.

Ahora bien, la ecuación (5) tiene 2 componentes, el precio de la amenidad p_s y el costo compensatorio de la amenidad, es decir, la derivada parcial C_s . Para hallar esta última se despeja de la diferenciación de la ecuación (3) realizada anteriormente:

$$\begin{aligned} C_w w_s + C_r r_s + C_s &= 0 \\ C_s &= -C_w w_s - C_r r_s \\ C_s &= -(C_w w_s + C_r r_s) \\ C_s &= -\left(C_w \frac{\partial w}{\partial s} + C_r \frac{\partial r}{\partial s}\right) \end{aligned}$$

Notese que, como menciona Roback, $C_w = \frac{N}{X}$ y $C_r = \frac{l^c}{X}$, así:

$$C_s = -\left(\frac{N}{X} \frac{\partial w}{\partial s} + \frac{l^c}{X} \frac{\partial r}{\partial s}\right)$$

Reescribiendo como en el paper:

$$C_s = -\left(\theta_w \frac{\partial \log w}{\partial s} + \theta_r \frac{\partial \log r}{\partial s}\right)$$

Donde θ_i es la proporción del factor i en el costo de X .

El segundo componente es el precio de la amenidad, en este caso Roback lo define como $p_s^* = \frac{V_s}{V_w}$. Así, el precio es igual a la tasa marginal de sustitución entre el salario y la amenidad, lo que indica cuánta compensación salarial es necesaria para mantener el nivel de utilidad constante frente a un cambio marginal en la amenidad. Este cociente define el precio implícito de la amenidad, es decir, cuánto estarían dispuestos a aceptar los individuos en términos de salario por una mejora marginal en la amenidad. Así pues, encontrar el precio se reduce a hallar la expresión $\frac{V_s}{V_w}$. Para esto, de la diferenciación de la ecuación (2) se despeja V_s :

$$\begin{aligned} V_w w_s + V_r r_s + V_s &= 0 \\ V_s &= -V_w w_s - V_r r_s \\ V_s &= -V_w \frac{\partial w}{\partial s} - V_r \frac{\partial r}{\partial s} \end{aligned}$$

Al dividir por V_w :

$$\begin{aligned} \frac{V_s}{V_w} &= \frac{-V_w \frac{\partial w}{\partial s} - V_r \frac{\partial r}{\partial s}}{V_w} \\ \frac{V_s}{V_w} &= -\frac{\partial w}{\partial s} - \frac{V_r}{V_w} \frac{\partial r}{\partial s} \end{aligned}$$

Sabiendo que la cantidad de consumo de tierra de equilibrio es la función $l^c(p_x = 1, r, w)$, se puede aplicar entonces la identidad de Roy, de manera que:

$$l^c = -\frac{\frac{\partial V}{\partial r}}{\frac{\partial V}{\partial w}} = \frac{V_r}{V_w}$$

Al reemplazar, se llega al precio definido por Roback:

$$\begin{aligned} p_s^* &= \frac{V_s}{V_w} = -\frac{\partial w}{\partial s} + l^c \frac{\partial r}{\partial s} \\ p_s^* &= \frac{V_s}{V_w} = l^c \frac{\partial r}{\partial s} - \frac{\partial w}{\partial s} \end{aligned}$$

Además, dividiendo por w , Roback obtiene el precio en términos reales:

$$\frac{p_s^*}{w} = k_l \frac{\partial \log r}{\partial s} - \frac{\partial \log w}{\partial s}$$

Donde k_l es la proporción del presupuesto del consumidor destinada a la tierra.

Juntando todos los resultados obtenidos la ecuación (5) es:

$$\begin{aligned} p_s^* &= \frac{V_s}{V_w} = l^c \frac{\partial r}{\partial s} - \frac{\partial w}{\partial s} \quad o \quad \frac{p_s^*}{w} = k_l \frac{\partial \log r}{\partial s} - \frac{\partial \log w}{\partial s} \\ C_s &= - \left(\frac{N}{X} \frac{\partial w}{\partial s} + \frac{l^c}{X} \frac{\partial r}{\partial s} \right) = - \left(\theta_w \frac{\partial \log w}{\partial s} + \theta_r \frac{\partial \log r}{\partial s} \right) \end{aligned} \quad (5)$$

Estimación

Con esto en mente, ahora se quiere estimar la importancia de la amenidad A en terminos reales a partir de los datos proporcionados, es decir, estimar $\frac{p_s^*}{w}$. Para ello es necesario conocer k_l , $\frac{\partial \log w}{\partial s}$, y $\frac{\partial \log r}{\partial s}$. El primero, la constante, lo proporciona el enunciado $k_l = \frac{1}{3}$, sin embargo, hay que estimar las derivadas parciales del logaritmo del salario y el arriendo con respecto a la amenidad. Para ello, se plantean dos regresiones de ambas variables con la amenidad $s = A$:

$$\begin{aligned} \log w &= \beta_0 + \beta_1 A + u_i \\ \log r &= \alpha_0 + \alpha_1 A + \epsilon_i \end{aligned}$$

Las dos especificaciones planteadas descansan sobre un supuesto de identificación fundamental: que la amenidad A es exógena con respecto a los salarios w y las rentas r . Esto implica que la variación en A no está correlacionada con factores no observados que también puedan influir en los salarios y los precios de las viviendas, como características no medidas de las ciudades o shocks económicos locales. Si este supuesto no se cumple, los coeficientes estimados en las regresiones podrían estar sesgados debido a la endogeneidad, y no reflejarían el verdadero impacto de la amenidad sobre los salarios y las rentas. Para que el análisis sea válido, es crucial que la amenidad esté suficientemente desvinculada de otros determinantes económicos que podrían afectar simultáneamente los resultados en cada ciudad de Tangolandia. Sin embargo, la plausibilidad de este supuesto depende del tipo de amenidad que se esté midiendo. Si A corresponde a una característica natural o fija, como el clima o la geografía, es más probable que el supuesto de exogeneidad se cumpla, ya que estas variables tienden a ser independientes de las decisiones económicas locales. En cambio, si la amenidad está relacionada con características de las viviendas, como el número de habitaciones, el tamaño o la antigüedad de la construcción, es más difícil argumentar exogeneidad, ya que estas variables pueden estar correlacionadas con factores socioeconómicos locales, como el nivel de ingresos o la inversión en infraestructura, que a su vez afectan tanto los salarios como las rentas, lo que complicaría el cumplimiento del supuesto de identificación.

Cuadro 1: Estimación del efecto parcial de la amenidad sobre el salario y arriendo en logaritmo

	<i>Variable Dependiente:</i>	
	Log Salario	Log Arriendo
	(1)	(2)
Amenidad	0.262** (0.103)	-0.097 (0.173)
Constante	6.588*** (0.378)	6.917*** (0.635)
Observaciones	75	75
R ²	0.082	0.004

Nota: Errores estándar en paréntesis
*p<0.1; **p<0.05; ***p<0.01

Los resultados muestran que la amenidad tiene un efecto positivo y significativo sobre los salarios. Se tiene que un aumento en una unidad de la amenidad, manteniendo todo lo demás constante, está asociado con un incremento del 26,2% en el salario promedio. Este resultado es significativo al nivel del 5%, lo que sugiere que la amenidad es un factor importante en la determinación de los salarios. En contraste, el efecto de la amenidad sobre los arriendos no es significativo, con un coeficiente de -0.097 y un p-valor que no alcanza los niveles usuales de significancia. Esto implica que, aunque la amenidad parece tener un impacto claro sobre los salarios, su efecto sobre los precios de arriendo es más incierto o posiblemente nulo en este contexto. El R^2 bajo en ambos modelos sugiere que las variaciones en la amenidad explican solo una pequeña fracción de la variación total en salarios y arriendos.

Ya sabiendo los coeficiente, es posible calcular la importancia real de la amenidad reemplazando los valores en la formula del precio de la ecuación (5):

$$\begin{aligned}\frac{p_A^*}{w} &= k_l \frac{\partial \log r}{\partial A} - \frac{\partial \log w}{\partial A} \\ \frac{p_A^*}{w} &= \frac{1}{3}(-0,097) - 0,262 \approx -0,294\end{aligned}$$

El valor obtenido indica que la amenidad tiene un impacto negativo sobre el bienestar relativo de los individuos en términos de salario. Este valor refleja la compensación implícita que los trabajadores requieren por cada unidad adicional de la amenidad, es decir, los trabajadores estarían dispuestos a aceptar una disminución del 29,4% en su salario para mantener constante su utilidad si la amenidad se incrementa. Desde una perspectiva intuitiva, aunque la amenidad parece aumentar los salarios (coeficiente positivo), el resultado negativo en el precio implícito sugiere que, dado que los precios de arriendo no aumentan significativamente con la mejora de la amenidad, el beneficio neto para los consumidores es negativo en términos de compensación salarial. Esto puede interpretarse como una indicación de que los trabajadores valoran más la amenidad en términos de salarios perdidos, ya que no están pagando por ella a través de mayores arriendos, sino aceptando salarios más bajos.

Finalmente, se calcula el error estándar del precio $\frac{p_A^*}{w}$:

$$\begin{aligned}\text{var} \left(\frac{p_A^*}{w} \right) &= \text{var} \left(k_l \frac{\partial \log r}{\partial A} - \frac{\partial \log w}{\partial A} \right) \\ \text{var} \left(\frac{p_A^*}{w} \right) &= k_l^2 \text{var} \left(\frac{\partial \log r}{\partial A} \right) + \text{var} \left(\frac{\partial \log w}{\partial A} \right) - 2 \text{cov} \left(\frac{\partial \log r}{\partial A}, \frac{\partial \log w}{\partial A} \right) \\ \text{var} \left(\frac{p_A^*}{w} \right) &= \frac{1}{3} (0,173)^2 + 0,103^2 - 2 \text{cov} \left(\frac{\partial \log r}{\partial A}, \frac{\partial \log w}{\partial A} \right) \\ \text{var} \left(\frac{p_A^*}{w} \right) &\approx 0,0139 - 2 \text{cov} \left(\frac{\partial \log r}{\partial A}, \frac{\partial \log w}{\partial A} \right)\end{aligned}$$

El problema es que no se conoce el valor de $\text{cov} \left(\frac{\partial \log r}{\partial A}, \frac{\partial \log w}{\partial A} \right)$. Entonces, se estima por medio de bootstrap (*ver script*). Del procedimiento, se obtiene que el valor de la covarianza es 0.001088153. Así:

$$\text{var} \left(\frac{p_A^*}{w} \right) = 0,0139 - 2(0,001088153) \approx 0,0117$$

El error estándar² es:

$$ee \left(\frac{p_A^*}{w} \right) = \frac{\sqrt{\text{var} \left(\frac{p_A^*}{w} \right)}}{\sqrt{n}} = \frac{\sqrt{0,0117}}{\sqrt{75}} \approx 0,0125$$

Se puede concluir que la estimación del precio implícito de la amenidad es muy precisa. El pequeño valor del error estándar sugiere que hay muy poca variabilidad en la estimación de p_A^* , lo que significa que el resultado de -0.294 es bastante confiable. Dado que el tamaño de la muestra es de 75 observaciones, este bajo error estándar indica que las fluctuaciones en los datos no introducen mucha incertidumbre en la estimación, lo cual es un buen indicio de que el efecto de la amenidad está siendo medido robustamente.

²También era posible estimar directamente el error estándar por medio de bootstrap, generando una distribución para p_A^* . Sin embargo, se optó por hacerlo con la covarianza para que se evidenciara todo el desarrollo teórico del cálculo y teniendo en cuenta que el resultado solo varía una milésima.

Ejercicio 3

En este ejercicio estimaremos la disposición a pagar para vivir cerca de un espacio abierto en Bogotá a partir de los datos provistos en el taller 2.

- Obtenga información de parques y plazas de OSM, que será la medida de espacio abierto.
- Definiendo como “cerca” 200 metros del parque y/o plaza más cercano, estime la disposición a pagar por esta cercanía.
- Al escribir sus resultados:
 - Escriba la especificación de la regresión a estimar explicando cada uno de los términos incluidos, resaltando el coeficiente de interés con su interpretación.
 - Estime por OLS e interprete la especificación. Sea preciso con los supuestos que hace sobre los errores estándar.
 - Explícite los supuestos necesarios para que esta interpretación sea causal y discuta por qué es posible que estos sí (o no) se cumplan.
 - Estime la regresión anterior:
 - Con el estimador de diferencias espaciales.
 - Utilizando los errores estándar de Conley.

Compare los resultados entre las distintas especificaciones, comente si estas dos estrategias mitigan o exacerban las certezas o preocupaciones de su interpretación causal. Para ello, sea explícito en los supuestos de cada estrategia y cómo se comparan con los de su estimación base.

Solución:

En este ejercicio, estimaremos la disposición a pagar por vivir cerca de un espacio abierto, como parques o plazas, en Bogotá. Utilizaremos datos de OpenStreetMap (OSM) para obtener información precisa sobre la ubicación de estos espacios abiertos y definiremos cercanía como vivir a menos de 200 metros del parque o plaza más cercano. Nos interesa analizar dos efectos principales: el impacto general de vivir cerca de cualquier espacio abierto sobre el precio de las viviendas y un efecto heterogéneo que diferencia entre parques y plazas, dado que es posible que estos espacios generen diferentes niveles de atractivo para los residentes. Teóricamente, la proximidad a estos espacios abiertos actúa como una amenidad, lo que podría mejorar la calidad de vida y reflejarse en un aumento de los precios de las viviendas. Sin embargo, según modelos como el de Roback (1982), también es posible que la presencia de una buena amenidad, pueda reducir los costos de vivienda, ya que las personas estarían dispuestas a pagar menos por vivir cerca de estos espacios debido a la compensación que reciben en términos de salarios o calidad de vida. Por lo tanto, es plausible que en algunos casos la proximidad a parques o plazas pueda estar asociada a una reducción en los precios, dependiendo de cómo se equilibren estos factores en el mercado inmobiliario.

Metodología

En este caso particular, se busca estimar el efecto de tener un parque o una plaza cerca sobre el precio de las viviendas, diferenciando entre estos dos tipos de espacios abiertos para capturar un efecto heterogéneo. Además, se controlará por una serie de amenidades adicionales de las propiedades para aislar mejor el impacto específico de la cercanía a estos espacios. La especificación se estimará utilizando el método de Mínimos Cuadrados Ordinarios (MCO), y es la siguiente:

$$\log(\text{Precio})_i = \beta_0 + \beta_1 \mathbb{1}[\text{EspacioAbiertoCerca}] + \beta_2 \mathbb{1}[\text{PlazaCerca}] + \sum_{k=1}^4 \alpha_k x_{ki} + \epsilon_i$$

La especificación presentada modela el logaritmo del precio de la vivienda i en función de varias variables. La primera variable, $\mathbb{1}[\text{EspacioAbiertoCerca}]$, es un indicador que toma el valor de 1 si la vivienda está a menos de 200 metros de un espacio abierto, ya sea parque o plaza. La segunda variable, $\mathbb{1}[\text{PlazaCerca}]$, indica si ese espacio abierto es específicamente una plaza, diferenciando entre parque y plaza. Además, se incluyen un conjunto de variables representadas por x_{ki} , que controlan por amenidades de la vivienda, como el número de baños, el número de habitaciones, la superficie total y la superficie cubierta de la propiedad. Finalmente, el término ϵ_i representa el componente de error que recoge todas las variaciones en el precio de la vivienda no explicadas por las variables incluidas en el modelo.

En esta especificación, el coeficiente β_1 captura el efecto de tener un espacio abierto (ya sea parque o plaza) a menos de 200 metros de la vivienda sobre el precio de la propiedad. Este es el coeficiente de interés principal, ya que nos indica si la cercanía a cualquier tipo de espacio abierto afecta los precios de las viviendas. Por otro lado, β_2 captura el efecto diferencial de que ese espacio abierto sea específicamente una plaza en lugar de un parque. La inclusión de β_2 en la regresión permite identificar si hay diferencias en cómo los residentes valoran estar cerca de una plaza frente a un parque. Al agregar esta variable, se busca capturar posibles efectos heterogéneos, ya que plazas y parques pueden ofrecer distintos niveles de amenidad en términos de estética, uso recreativo o valor cultural, lo que podría reflejarse en diferencias en los precios de las viviendas cercanas a estos espacios.

Para que la estimación mediante MCO sea válida, es necesario que se cumplan varios supuestos. Uno de los más importantes es el de exogeneidad, que implica que las variables explicativas (cercanía a un parque o plaza, y las amenidades) no estén correlacionadas con el término de error ϵ_i , lo que aseguraría que no haya variables omitidas que afecten tanto el precio de las viviendas como la cercanía a los espacios abiertos. Sin embargo, en este caso es probable que este supuesto no se cumpla completamente, ya que podría haber factores no observados, como la calidad del vecindario o la presencia de otras amenidades, que estén correlacionados con la ubicación de parques y plazas, lo que introduce un sesgo en la estimación. Además, otro supuesto crucial es que los errores estándar sean homocedásticos, es decir, que la varianza de los errores sea constante a lo largo de las observaciones. Esto también implica que los errores no estén correlacionados entre sí, lo que es menos plausible en datos espaciales, ya que propiedades cercanas pueden compartir características no especificadas en el modelo. La violación de este supuesto podría llevar a una subestimación de los errores estándar, lo que aumenta el riesgo de cometer errores tipo I o tipo II al evaluar la significancia de los coeficientes. Por ello, es probable que la estimación base con MCO no capture adecuadamente la correlación espacial y la heterocedasticidad presentes en los datos.

Por lo anterior, además de la estimación básica con MCO, se realizarán estimaciones utilizando errores estándar agrupados (cluster), lo que permitirá corregir la correlación entre observaciones dentro de un mismo grupo (en este caso, viviendas cercanas). Esta metodología es útil en situaciones donde las observaciones dentro de un grupo pueden no ser independientes, como en este caso donde las propiedades cercanas pueden compartir características no observadas que afectan los precios. Al utilizar errores agrupados por distancia o vecindario, se mejora la robustez de los estimadores, haciendo las conclusiones más confiables, aunque supone que las propiedades dentro del mismo cluster son comparables.

Adicionalmente, se emplearán errores estándar de Conley, una metodología diseñada para corregir la correlación espacial entre las observaciones, que es especialmente relevante en contextos como este, donde las características no observables de los vecindarios pueden afectar tanto el precio de las viviendas como la presencia de espacios abiertos. Este método ajusta los errores estándar al permitir correlación espacial hasta una cierta distancia, en este caso 1 km, lo que significa que las propiedades dentro de este rango pueden compartir características no incluidas en el modelo (como la calidad de los servicios públicos, la seguridad o la infraestructura del vecindario), afectando de manera conjunta el precio de las viviendas. El uso de los errores de Conley corrige no solo la heterocedasticidad, sino también la autocorrelación espacial, proporcionando una estimación más precisa de la significancia estadística de los coeficientes, especialmente en entornos donde la proximidad física influye en las características no observables de las propiedades. Sin embargo, este método también tiene sus limitaciones, ya que asume que las propiedades dentro del radio de 1 km son comparables entre sí, lo cual puede ser un supuesto fuerte en un contexto urbano como Bogotá, donde los vecindarios pueden variar significativamente dentro de un mismo radio. Por ejemplo, dentro de un kilómetro podrían coexistir barrios de estratos socioeconómicos muy diferentes,

con características geográficas, de infraestructura y sociales marcadamente distintas. Esto podría hacer que las propiedades dentro de un mismo cluster no sean realmente comparables, lo que puede introducir cierto sesgo si estas diferencias no se controlan adecuadamente. A pesar de esta limitación, el método de Conley proporciona una mayor robustez en las estimaciones al corregir la correlación espacial que la metodología MCO no logra abordar.

Además de las metodologías previamente mencionadas, también se utilizará el estimador de diferencias espaciales para abordar las preocupaciones relacionadas con la endogeneidad derivada de heterogeneidades no observadas entre las observaciones y la correlación espacial. La idea detrás de este enfoque es comparar cada observación con su vecino geográfico más cercano, bajo el supuesto de que las unidades adyacentes son más comparables entre sí que las unidades más alejadas. Este enfoque restringe la comparación a propiedades vecinas y permite diferenciar el impacto de las variables no observadas que son comunes a las propiedades cercanas, eliminando así parte del sesgo generado por la omisión de variables que pueden estar afectando a las viviendas dentro de un mismo entorno. Esta metodología es útil porque, a diferencia del MCO tradicional, no supone que todas las unidades dentro de un rango geográfico amplio sean comparables. La especificación es la misma que antes pero con las variables rezagas:

$$dif(\log(Precio)_i) = \beta_0 + \beta_1 dif(\mathbb{1}[EspacioAbiertoCerca]) + \beta_2 dif(\mathbb{1}[PlazaCerca]) + \sum_{k=1}^4 \alpha_k dif(x_{ki}) + \epsilon_i$$

El supuesto central de esta metodología es la Local Conditional Independence Assumption, que establece que las unidades adyacentes comparten los mismos valores en factores no observables, lo que permite aislar el efecto causal de la proximidad a espacios abiertos. Este enfoque es particularmente útil en contextos urbanos donde las características de un barrio (infraestructura, servicios, entorno social) tienden a ser similares entre vecinos cercanos. No obstante, el gran desafío de este método es que depende en gran medida de la calidad y proximidad de los vecinos. En algunos casos, la vivienda vecina puede no estar en el mismo barrio o zona homogénea, lo que podría introducir diferencias no observadas significativas que el método no logra corregir completamente. Aun así, la diferencia espacial es una herramienta poderosa para mitigar el sesgo por variables omitidas cuando los vecinos son efectivamente comparables, lo que puede proporcionar estimaciones más cercanas al efecto causal verdadero que el MCO tradicional o incluso los errores estándar de Conley.

Estimación

El cuadro 2 muestra los resultados de las estimaciones para los diferentes modelos. En los primeros 2, los de Mínimos Cuadrados Ordinarios (MCO), se observa que para el precio de venta, la variable de proximidad a un espacio abierto (parque o plaza) tiene un efecto significativo y negativo. El coeficiente de -0.135 indica que, al estar ubicado a menos de 200 metros de un espacio abierto, el precio de venta de la vivienda disminuye en promedio un 13.5%, manteniendo constantes las demás variables. Este resultado podría parecer contradictorio a la expectativa de que la cercanía a un parque o plaza aumenta el valor de las propiedades. Una posible explicación para este efecto negativo es que los parques y plazas podrían estar ubicados en áreas menos atractivas en términos de otros factores no observados, como mayor ruido, inseguridad o falta de servicios, lo que contrarrestaría el valor de la amenidad de estar cerca de un espacio abierto. También podría sugerir que, aunque los espacios abiertos son una amenidad, hay efectos negativos asociados con estar demasiado cerca de ellos, como la congestión pero principalmente la delincuencia. O, en línea con Roback, que el valor positivo de la amenidad se compensa con salario más bajos

Por otro lado, el efecto heterogéneo de estar cerca de una plaza en lugar de un parque es positivo y también significativo, con un coeficiente de 0.061. Esto indica que, aunque estar cerca de un espacio abierto reduce el precio de las viviendas, el efecto es menos negativo si ese espacio abierto es una plaza en lugar de un parque. En otras palabras, las viviendas cercanas a plazas tienden a perder menos valor que las cercanas a parques. Este resultado sugiere que los residentes pueden percibir las plazas como más valiosas o menos problemáticas en comparación con los parques, lo que podría estar relacionado con su uso, mantenimiento o el tipo de actividades que se desarrollan en ellas. Es importante destacar que este efecto positivo no implica que vivir cerca de una plaza aumente los precios, sino que mitiga

el impacto negativo de la cercanía a un espacio abierto. En cuanto a la significancia estadística, los resultados para el precio de venta son robustos, mientras que en el caso del precio de arriendo, ni la variable de proximidad a un espacio abierto ni el efecto heterogéneo de estar cerca de una plaza son estadísticamente significativos, lo que sugiere que en el mercado de arriendos, la cercanía a estos espacios no parece influir de manera notable en los precios. Además, el R^2 del modelo de venta es de 0.452, lo que indica una buena capacidad explicativa, mientras que el R^2 para arriendos es ligeramente más alto, pero sin efectos significativos en las variables clave.

Cuadro 2: Resultados de los Modelos

	Variable Dependiente: Log Precio							
	OLS		Cluster		Dif. Espaciales		Conley	
	(1) Venta	(2) Arriendo	(3) Venta	(4) Arriendo	(5) Venta	(6) Arriendo	(7) Venta	(8) Arriendo
Espacio Abierto Cerca (200m)	-0.135*** (0.025)	0.036 (0.032)	-0.135* (0.073)	0.036 (0.090)			-0.135** (0.068)	0.036 (0.073)
Efecto Heterogéneo Plaza Cerca (200m)	0.061** (0.024)	-0.035 (0.031)	0.061 (0.073)	-0.035 (0.089)			0.061 (0.062)	-0.035 (0.071)
Diferencia Espacio Abierto Cerca (200m)					-0.148*** (0.024)	0.041 (0.031)		
Diferencia Efecto Heterogéneo Plaza Cerca (200m)					0.071 (0.023)	-0.039 (0.030)		
R^2	0.452	0.494	0.452	0.494	0.442	0.481		
Observaciones	38,167	22,628	38,167	22,628	38,166	22,627	38,167	22,628

Nota 1: Errores estándar en paréntesis
*p<0.1; **p<0.05; ***p<0.01

Nota 2: Todas las especificaciones incluyen 4 controles: número de habitaciones, número de baños, superficie total y superficie total cubierta.

En los modelos estimados con errores agrupados (cluster), específicamente los modelos 3 y 4, observamos que los coeficientes de las variables de interés se mantienen en los mismos niveles que en los modelos OLS, pero la significancia estadística cambia. En el caso del modelo 3, que estima el precio de venta, el coeficiente para la proximidad a un espacio abierto sigue siendo negativo (-0.135), pero su significancia disminuye, pasando a ser significativa solo al 10 % en lugar del 1 % como en el modelo OLS. Esto sugiere que al permitir la agrupación de los errores por vecindario o área geográfica, la robustez de la estimación se ve afectada y el margen de error aumenta, lo que podría reflejar la existencia de correlación entre las propiedades dentro de un mismo grupo que no se capturaba en el modelo OLS. Para el precio de arriendo, el coeficiente para la proximidad a un espacio abierto sigue sin ser significativo, lo que refuerza la idea de que, en los alquileres, esta variable no tiene un impacto claro. El R^2 no cambia en comparación con los modelos OLS, lo que indica que la capacidad explicativa del modelo no se ve afectada por la corrección de los errores estándar, pero la significancia de los coeficientes muestra una mayor sensibilidad a la correlación espacial dentro de las áreas.

Por su parte, en los modelos con errores estándar de Conley, los resultados presentan un patrón similar al de los modelos cluster. En el modelo 7, que estima el precio de venta, el coeficiente para la proximidad a un espacio abierto sigue siendo negativo y significativo al 5 %, lo que sugiere que la corrección de la correlación espacial a una distancia de 1 km proporciona estimaciones más robustas que las obtenidas con OLS, aunque algo menos precisas que en el caso de los errores agrupados. El efecto heterogéneo de las plazas también sigue siendo positivo, pero pierde su significancia estadística, lo que indica que la correlación espacial podría estar influyendo más de lo que se capturaba con OLS. En el modelo 8, para los precios de arriendo, los coeficientes siguen sin ser significativos, lo que refuerza aún más la conclusión de que la cercanía a espacios abiertos no parece tener un impacto claro en los precios de alquiler. En general, los errores estándar de Conley permiten un ajuste más preciso en cuanto a la correlación espacial, pero los resultados sugieren que, en este caso, las correcciones no cambian de manera sustancial las interpretaciones generales en comparación con los modelos OLS y cluster, aunque ofrecen una mayor confianza en la robustez de las estimaciones.

En los modelos estimados con diferencias espaciales, que comparan cada vivienda con su vecino geográfico más cercano, los resultados presentan una dinámica diferente respecto a los modelos anteriores. En el modelo 5, que estima el precio de venta, el coeficiente para la proximidad a un espacio abierto es -0.148, lo que significa que el impacto de vivir a menos de 200 metros de un espacio abierto es una reducción del precio de la vivienda en un 14.8 %, una magnitud ligeramente mayor que la obtenida en los modelos OLS

y cluster. Este coeficiente es altamente significativo al 1 %, lo que sugiere que al comparar propiedades cercanas, el efecto negativo de vivir cerca de un parque o plaza se refuerza. Esto puede deberse a que las diferencias espaciales eliminan parte de la heterogeneidad no observada entre las viviendas, haciendo que el impacto negativo se revele con más claridad. El efecto heterogéneo de estar cerca de una plaza también es positivo en este modelo, con un coeficiente de 0.071, aunque no es estadísticamente significativo, lo que indica que, si bien vivir cerca de una plaza parece tener un impacto menos negativo que estar cerca de un parque, no hay suficiente evidencia para afirmar que esta diferencia es significativa. En el modelo 6, que estima el precio de arriendo, los resultados son consistentes con lo observado en los modelos anteriores: ni la proximidad a un espacio abierto ni el efecto de estar cerca de una plaza son significativos. Esto refuerza la conclusión de que en el mercado de arriendos, la cercanía a un parque o plaza no tiene un impacto claro en los precios, a diferencia de lo que ocurre con los precios de venta. Así, las diferencias espaciales confirman que estar cerca de un espacio abierto tiende a reducir el precio de venta de las viviendas, y aunque vivir cerca de una plaza puede mitigar este efecto en comparación con estar cerca de un parque, el efecto heterogéneo no es concluyente.

Los resultados obtenidos revelan una dinámica interesante en la relación entre la proximidad a espacios abiertos y los precios de las viviendas en Bogotá, que desafía algunas expectativas intuitivas sobre el valor de estas amenidades. El hecho de que la cercanía a un parque o plaza esté asociada con una reducción en el precio de las viviendas, especialmente en los modelos de venta, puede sugerir que estos espacios abiertos, en ciertos contextos urbanos, no siempre son percibidos como una mejora directa en la calidad de vida. Una posible explicación es que algunos parques y plazas pueden estar ubicados en áreas donde existen otros factores negativos, como mayores niveles de ruido, inseguridad o sobrecongestión debido al tránsito y uso público de esos espacios. Esto podría reducir el atractivo de vivir cerca de ellos, particularmente en zonas donde los espacios abiertos no están bien mantenidos o generan externalidades negativas. Además, la percepción de la seguridad alrededor de estos espacios puede ser un factor determinante, ya que en algunos contextos, parques o plazas pueden ser vistos como focos de inseguridad, lo que influiría directamente en la disposición a pagar por vivir cerca de ellos.

Por otro lado, el hecho de que el efecto heterogéneo de estar cerca de una plaza sea positivo pero no siempre significativo sugiere que las plazas pueden ser percibidas como espacios más controlados.^o más valiosos en comparación con los parques, quizás por su diseño más orientado al encuentro social o su ubicación en áreas más céntricas y organizadas. Sin embargo, la falta de significancia en algunos modelos sugiere que esta percepción no es uniforme y depende mucho del contexto específico de cada plaza o parque. Es posible que algunas plazas estén mejor integradas en la estructura urbana y sean más valoradas por los residentes, mientras que otras no ofrecen los mismos beneficios. Además, la falta de impacto en los precios de arriendo podría indicar que los inquilinos valoran menos las amenidades a largo plazo o que los factores que influyen en los precios de alquiler, como la accesibilidad o la demanda temporal, tienen más peso que la proximidad a un espacio abierto. Este comportamiento diferencial entre ventas y arriendos sugiere que los compradores de viviendas tienden a ser más sensibles a la presencia de externalidades negativas asociadas a los espacios abiertos, mientras que los arrendatarios, con horizontes temporales más cortos, pueden no percibir estos efectos de la misma manera.