

# Relación entre precios de alquiler en portales inmobiliarios y precios de mercado. Evidencias para la Comunidad de Madrid

**David Rey-Blanco.** Universidad Nacional de Educación a Distancia (UNED), Madrid, España.

**Julio González.** Universidad Nacional de Educación a Distancia (UNED), Madrid, España.

**David Sánchez.** Universidad Católica Santa Teresa de Jesús de Ávila, Ávila, España.

**RESUMEN** | La creciente necesidad de información de los agentes económicos no siempre se ve satisfecha por las fuentes de datos abiertas de la Administración Pública. La digitalización de la sociedad, en particular en el sector inmobiliario, ofrece un gran número de alternativas informativas, como los precios de los portales para la búsqueda de vivienda en internet. Este estudio analiza la relación entre los precios de oferta de los portales y los precios de mercado de alquiler, a partir de los datos del Ministerio de Transportes, Movilidad y Agenda Urbana de España, para 177 áreas geográficas en la Comunidad de Madrid. Los resultados demuestran empíricamente que las series de oferta y de alquiler guardan un alto grado de correlación; por tanto, se confirma que los datos de portales son un candidato válido para el análisis detallado y actualizado del mercado del alquiler.

**PALABRAS CLAVE** | vivienda, mercado inmobiliario, economía urbana.

**ABSTRACT** | *The growing need for information of economic agents is not always satisfied by public administration's open data sources. Furthermore, the increasingly changing conditions require updated and frequent control. The digitization of society, particularly in the real estate sector, offers many information alternatives, such as the prices featured in portals used to search for housing on the Internet. Based on data from the Spanish Ministry of Transport, Mobility and Urban Agenda, this study analyzes the relationship between portals' prices and rental market prices, for 177 geographical areas of different characteristics in the Community of Madrid. The results empirically demonstrate a high degree of correlation between the offer and rental series, thus the data featured in these portals is a valid source candidate for conducting detailed and updated analyses of rental markets.*

**KEYWORDS** | *housing, real estate market, urban economy.*

Recibido el 25 de enero de 2023, aprobado el 7 de junio de 2023.

E-mails: drey7@alumno.uned.es | jglez@cee.uned.es | sanchezcabreradavid@gmail.com

## Introducción

En el mercado de la vivienda existen numerosas fuentes de información pública, generalmente proporcionadas por la Administración Pública, pero que raramente están actualizadas y cuyo nivel de desglose funcional y geográfico es limitado. En España, para la compraventa de vivienda residencial, el Instituto Nacional de Estadística (INE, 2022a) publica trimestralmente el índice de precios de la vivienda (IPV) por comunidad autónoma. A nivel nacional, no existe un equivalente para el mercado del alquiler. La fuente más actualizada de la que se dispone es el Sistema Estatal de Índices de Referencia del Precio del Alquiler de Vivienda, desarrollado por el Ministerio de Transportes, Movilidad y Agenda Urbana (MITMA, 2020), que se calcula con frecuencia anual y se publica con un retraso de, aproximadamente, dos años. Esta base de datos cuenta con un mayor nivel de desglose zonal que el IPV, ya que se ofrece desagregada hasta sección censal.

Recientemente, el INE (2021) ha desarrollado, de forma experimental, el índice de precios de la vivienda en alquiler (IPVA), basado en datos registrados en los impuestos de la renta de las personas físicas (IRPF). Este índice tiene un desglose zonal máximo municipal, en municipios con más de 10.000 habitantes, y sobre distritos en las capitales de provincia. Al igual que la fuente del MITMA, estos datos se publican de forma anual y con un retraso de dos años.

La ausencia de información actualizada dificulta las tareas de vigilancia del comportamiento detallado y actualizado de los mercados, el que, por otra parte, es materia obligada ante un escenario económico y social altamente cambiante, y de manera particular para una economía como la española, habitualmente expuesta al comportamiento de los mercados inmobiliarios (Montalvo, 2011).

Los fenómenos de digitalización en casi todos los aspectos de la vida personal y profesional ofrecen un incremento de registros electrónicos de las operaciones en diversos ámbitos. Para el sector inmobiliario, la búsqueda de vivienda no se puede disociar de las plataformas en internet (Pangallo & Loberto, 2018) y puede representar cifras cercanas al 80% del total de la demanda (Chapelle & Eyméoud, 2022). Sin embargo, el aprovechamiento de estas fuentes no excluye importantes retos, entre los que se encuentran el control de la calidad de la información y la corrección de los sesgos, como el de no respuesta y los de sobre- e infrarrepresentación.

El uso de datos de internet para construir índices alternativos a los oficiales ha empezado a tomar cada vez más importancia, especialmente tras la crisis del Covid-19. Al respecto, diversos centros de estudios y agencias estadísticas nacionales han comenzado a replantearse el uso exclusivo de fuentes tradicionales (Biancotti et al., 2020).

Aunque es habitual que la Administración y el sector privado utilicen la información de portales inmobiliarios, existen pocos estudios que hayan profundizado en la validez de los datos de alquiler de los portales de internet en España. Para otros países, como Francia, encontramos los trabajos de Chapelle y Eyméoud (2022), que comparan los datos oficiales y de internet para el alquiler, llegando a la conclusión de que los anuncios *online* reflejan una visión insesgada del mercado inmobiliario. Sin embargo, no existe un consenso total acerca de la validez del precio de oferta

para estudiar el comportamiento del mercado real. Por una parte, algunos autores defienden su equivalencia (Chapelle & Eyméoud, 2022; Kokot & Bas, 2015; Pangallo & Loberto, 2018), mientras otros son mucho más cautelosos (Ardila et al., 2021; Eerola & Lyytikäinen, 2015; Han & Strange, 2016; Shimizu et al., 2016), o hay autores como Kolbe et al. (2021), que lo desaconsejan. Aun así, incluso los autores más escépticos reconocen la utilidad de esta fuente, los portales inmobiliarios, por su capacidad de reflejar dinámicas en la formación de los precios finales.

El presente estudio analiza la adecuación de los datos de alquiler residencial procedentes de portales inmobiliarios como fuentes de información que reflejen la situación de mercado. Además, evalúa el nivel de interacción entre los precios de oferta y mercado con los niveles de demanda y renta de las zonas. Esta investigación pretende la demostración empírica, a través de medidas estadísticas, de la validez de los datos de oferta de portales de internet para el análisis de los precios de oferta de la vivienda de alquiler, centrandolo para ello el caso de estudio en la región de Madrid.

Como complemento al artículo, se ha desarrollado y publicado un conjunto de datos abiertos, completo y documentado (Rey, 2023), que se acompaña con el material de análisis, siguiendo las recomendaciones de Páez (2021) para trabajos de investigación abierta y reproducible, y las relativas a distribución de datos abiertos (Arribas-Bel et al., 2021).

### **Marco teórico y revisión de literatura**

Shimizu et al. (2016) recogen los cuatro precios que toma una vivienda en el proceso de puesta en el mercado: i) primer precio al que se publica un inmueble (precio de oferta), ii) el precio al que se retira el anuncio, iii) el precio al que se formaliza la operación, y iv) el precio al que se registra la transacción en registros oficiales. Aunque muestren diferencias entre ellos, los cuatro están relacionados y son medidas comparables, si bien esta relación depende en buena medida de la situación del mercado (Han & Strange, 2014, 2016; Kolbe et al., 2021; Shimizu et al., 2016).

Distintos estudios evalúan el nivel de correlación entre las poblaciones de oferta y las de transacción en alquiler; así, Chapelle y Eyméoud (2022) para el mercado francés, y Kokot y Bas (2015) en el caso de Polonia, estudian el nivel de correlación de transacciones y oferta, obteniendo un coeficiente de correlación de 0,95 y 0,99, respectivamente. Kokot y Bas afirman, además, que las series temporales están cointegradas y que muestran un retraso temporal medio de cinco meses de la serie de transacciones con respecto a la de oferta. En este estudio se argumenta por qué en muchas ocasiones es mejor tomar la serie de oferta, ya que las transacciones ofrecen una visión del mercado mucho más inestable desde un punto de vista estadístico. A este respecto, sin embargo, no hay un consenso claro en cuanto a si el dato de portales inmobiliarios puede sustituir el dato de transacciones (Ardila et al., 2021; Hill & Trojanek, 2022; Kolbe et al., 2021), principalmente porque, aunque las poblaciones estén fuertemente correlacionadas, la distribución de los valores de oferta y la de aquellos de mercado no es idéntica (Ardila et al., 2021; Kolbe et al., 2021; Shimizu et al., 2016). Estas diferencias en la distribución de precios se pueden deber a una sobrerrepresentación en oferta de los inmuebles

menos líquidos y la infrarrepresentación de los más líquidos, o a la presencia de anuncios en régimen de subasta (Han & Strange, 2016). En sentido contrario, sin embargo, Ardila et al. (2021) destacan que el dato de oferta tiende a infraestimar la magnitud de los cambios de tendencia.

Desde un punto de vista teórico, en la relación entre las dos magnitudes de precios se asume, con frecuencia, un valor de oferta superior al final (Ardila et al., 2021; Eerola & Lyytikäinen, 2015; Galesi et al., 2020), aunque este argumento es discutido por autores como Han y Strange (2016), Kolbe et al. (2021), Han y Strange (2014) o Allen y Dare (2004). Finalmente, una tercera tesis defiende que la relación entre ambas magnitudes varía en el tiempo, según el ciclo del mercado inmobiliario, como defienden Shimizu et al. (2016), De Wit et al. (2013) y Han y Strange (2016).

En todo caso, existe un consenso generalizado en cuanto a la utilidad de los datos de oferta debido a su influencia estratégica en el proceso de búsqueda (Han & Strange, 2016), por ejemplo en su capacidad de señalar los precios de reserva del vendedor (Horowitz, 1992; Merlo & Ortalo-Magne, 2004; Merlo et al., 2015). Ello aparte de otros aspectos positivos, como, por ejemplo, la existencia de una relación entre el incremento de precios de oferta y un mayor número de operaciones (De Wit et al., 2013).

Aunque el uso de fuentes alternativas para el cálculo de índices de la vivienda no es algo nuevo, una de las primeras referencias de un índice basado en fuentes alternativas es el de compraventa y alquiler calculado sobre una base de datos de anuncios clasificados en el periódico *El Mercurio*, construido para un barrio de Santiago de Chile con una serie de datos mensuales desde 1998 a 2002 (Desormeaux & Piguillem, 2003). Dicha información fue fundamental en un estudio retrospectivo sobre la evolución de precios en Santiago entre 1982 y 2012, y en el que su autor, Trivelli (2012), anticipa los principales retos cuando se utilizan anuncios clasificados en la construcción de índices de precios de la vivienda. Posteriormente, Anenberg y Laufer (2017) desarrollan un índice de precios altamente actualizado para la Reserva Federal de Estados Unidos, sobre datos de oferta de múltiples MLS,<sup>1</sup> como en transacciones inmobiliarias. La ventaja de utilizar esta fuente, según los autores, es que, por un lado, recoge las condiciones actualizadas y detalladas del mercado; y, por otra parte, adelanta comportamientos de los índices oficiales, como el Case-Shiller,<sup>2</sup> con varios meses de antelación.

Más recientemente, han aparecido numerosos trabajos que justifican el uso de datos de portales inmobiliarios para el desarrollo de índices de precios. En este sentido, Sun et al. (2015) justifican su uso en que el 90% de los compradores y el

---

1 MLS (Multiple Listing Service), o Servicio de Listado Múltiple (SLM), refiere a una bolsa para compartir la cartera de inmuebles de una inmobiliaria con el resto de inmobiliarias. Así, cualquier asociado a esa bolsa tiene la posibilidad de comercializar las propiedades que hayan sido conseguidas por otra agencia o profesional inmobiliario. Véase Timku, "Qué es y cómo funciona MLS (Multiple Listing Service) para vender inmuebles", en <https://tinku.es/que-es-y-como-funciona-mls-multiple-listing-service-para-vender-inmuebles/> [N. de E.]

2 El índice Case-Shiller es un índice de precios mensual de la vivienda de los Estados Unidos de América, calculado en sus diez principales áreas metropolitanas.

92% de los vendedores usan internet para la búsqueda y comercialización de inmuebles. Por otra parte, Chappelle y Eyméoud (2022), en su estudio sobre inmuebles en alquiler en Francia, indican que, aunque su uso es mayoritario, hay un 22% que usa canales alternativos.

Los retrasos de meses en índices oficiales limitan la capacidad de reacción ante situaciones de choque en el mercado. Este retraso también genera un desequilibrio de información entre lo que se conoce del mercado y la situación real. Por ejemplo, en Estados Unidos, la publicación de los datos de los índices de precios al consumo tiene un efecto inmediato en los mercados de valores de las compañías cotizadas, a pesar de que esta información proceda de meses pasados (Anenberg & Laufer, 2017). Sin embargo, esta medida inicialmente positiva, orientada a la eliminación de asimetrías de información y consecuencia de la publicación de información más adecuada de oferta, puede afectar el comportamiento de los precios. Eerola y Lyytikäinen (2015) argumentan que esta información beneficia más a los propietarios que infravaloran su vivienda que a aquellos que la sobrevaloran, por lo que más información conduce generalmente a un incremento de los precios. Los mismos autores indican que con dicha información también se produce una reducción de los tiempos medios de búsqueda. En todo caso, un buen grado de información, tanto de vendedores como de arrendadores, mejora las condiciones de competencia a través de precios más adecuados, reduciendo la posibilidad de monopolios informativos (Lacerda, 2018).

Varios bancos centrales, como los de Italia, Reino Unido o Francia, han desarrollado análisis sobre el potencial de esta información. En Italia, Loberto et al. (2018) utilizan información del portal Immobiliare.it, de frecuencia semanal, para construir un índice de la vivienda alternativo. En este estudio se observa que los índices de oferta tienen una alta correlación con los índices basados en transacciones, con un coeficiente de correlación de Pearson del 0,96 (entre los datos del portal y del Ministerio de Hacienda). También evidencian problemas al trabajar con este tipo de información, como que una misma propiedad pueda estar anunciada más de una vez, o la ausencia de valores en ciertos campos. En el artículo del Banco de Francia, Bricongne et al. (2023) realizan un seguimiento de los precios con frecuencia diaria, sobre datos de cinco portales en el Reino Unido. Mediante técnicas de aprendizaje automático, desarrollan un modelo de correspondencia entre precios de oferta y los registrados por los notarios. En Asia, Wang et al. (2020) elaboran un índice de precios de la vivienda para 274 ciudades de China, basándose en datos de portales inmobiliarios.

Ahora bien, no solamente existen referencias basadas en datos de internet inmobiliario para elaborar índices de precios. Chauvet et al. (2013) construyen un índice basado en las búsquedas más habituales del servicio Google Trends sobre el mercado inmobiliario y su regulación, y demuestran que un índice de “sentimiento” del mercado puede estar altamente correlacionado con la evolución de los precios.

Históricamente, las guías oficiales de las oficinas de estadísticas se han centrado en el tratamiento de fuentes estadísticas primarias; por ejemplo, el Informe de Calidad para la Encuesta Social Europea (2014). Sin embargo, el uso de las técnicas habituales para control de calidad estadístico no es directamente extrapolable a grandes fuentes de datos (Anenberg & Laufer, 2017). Debido al creciente interés

en incorporar fuentes alternativas (Biancotti et al., 2020), en los últimos años estas guías se han adaptado para incluir el tratamiento de fuentes secundarias de información no estadística, como el manual de UNECE (United Nations Economic Commission for Europe, 2015), los enfoques de Struijs y Daas (2014) o la guía de Eurostat (2017) para la incorporación de fuentes de tipo Big Data.<sup>3</sup> Aunque de forma experimental, las agencias estadísticas empiezan a incorporar indicadores basados en fuentes masivas alternativas, como por ejemplo el estudio piloto de movilidad publicado por el INE (2022b), o el uso de indicadores en tiempo real en la gestión de la crisis del Covid-19 en el Reino Unido (Rosenfeld, 2022).

Sin embargo, aunque los orígenes de la información de internet están incrementando su popularidad en países como Estados Unidos (Boeing, 2020), en muchos de los casos se desconoce con precisión la representación del mercado (Besbris et al., 2018; Boeing & Waddell, 2017; Schachter & Besbris, 2017) y, por tanto, es difícil estimar su grado de representatividad. Los sesgos de asumir una composición distorsionada del mercado real impactan en la fiabilidad de información y en los costes de búsqueda, a la vez que lo hacen en las conclusiones que los investigadores y reguladores extraen de las fuentes masivas (Arribas-Bel & Bakens, 2019; Folch et al., 2018), y que pueden influir en el propio mercado. Desde el ángulo de la demanda, como plantean Mclaughlin y Young (2018), una mayor eficiencia en el proceso de búsqueda debida a más información, no está exenta de efectos negativos; entre ellos, un incremento de la heterogeneidad espacial de los precios residenciales provocada por dos factores: i) la elección se expande hasta zonas más diversas que las buscadas; y ii) el aumento de la demanda por vivir cerca de ciertos servicios sin un acompañamiento de la oferta, que provoca una expulsión de los futuros residentes a través del aumento de los precios.

Por los motivos anteriores, el uso de datos procedentes de fuentes de internet requiere un tratamiento específico (Loberto et al., 2018), puesto que, a pesar de que las plataformas dispongan de departamentos y sistemas de control de calidad de los anuncios, los datos han sido generados por usuarios y, por tanto, están sujeto a la subjetividad, la disparidad de criterios a la hora de interpretar los campos del anuncio, o a diversos errores de entrada de datos. Además, existen prácticas fraudulentas por parte de ciertos profesionales, que deben eliminarse. Por último, los portales han ido transformándose a lo largo del tiempo (inicialmente solo operaban como página web en PC, después como página para teléfonos móviles, y en los últimos años se han hecho accesibles también a través de una aplicación para dispositivos móviles), lo que modifica la forma en la que se registra e interpreta la información. Para evitar los problemas anteriores, es necesario un tratamiento previo de la información, consistente en tres pasos: deduplicación de datos (proceso que elimina las copias excesivas de los datos), eliminación de casos atípicos e imputación de atributos (Loberto et al., 2018).

Dada la naturaleza heterogénea de los activos inmobiliarios, muchos análisis aplican el modelado de precios hedónicos, definido como “el método mediante el

---

3 El término Big Data se refiere al manejo de grandes volúmenes de datos, heterogéneos y actualizados frecuentemente.

cual se calculan los precios implícitos de los atributos o características que componen a un bien compuesto...” (Rosen, 1974). Dicha aproximación es especialmente eficaz en la construcción de índices de la vivienda, especificados mediante un gran número de variables y altamente influidos por su ubicación. En España se pueden encontrar artículos recientes que lo aplican; por citar algunos, Alfaro-Navarro et al. (2020), Rico y Taltavull (2021) y Rey-Blanco et al. (2023).

### Fuentes de datos

El presente análisis compara dos aspectos para el precio del alquiler residencial: la perspectiva de oferta y la de mercado. La primera representa el precio por el que se anuncian los inmuebles; la segunda, las rentas que pagan los inquilinos. Para ambos casos se han utilizado datos de anuncios del portal de internet Idealista entre los años 2015 al 2019, correspondientes a la Comunidad de Madrid (España), el área administrativa en la que se encuentra la capital del país. Esta región, con una densidad de población de 844 habitantes/km<sup>2</sup> y 6,75 millones de habitantes (INE, 2022c), es una de las zonas con mayor actividad inmobiliaria de Europa.

Desde el punto de vista de áreas geográficas de estudio, se ha realizado un desglose zonal doble: a nivel de barrio para Madrid, basado en la división administrativa del municipio; y a nivel de área municipal para el resto de la provincia. La denominada “área municipal” es o bien un municipio o una agrupación de municipios contiguos, que son inmobiliariamente similares.<sup>4</sup> El número total de áreas geográficas empleadas es de 177, sumados barrios (112) y áreas municipales (65).

El dato de oferta procede del portal Idealista, que es el principal medio de anuncios clasificados inmobiliarios en España en internet (Similarweb, 2022). Debido a las características de esta fuente, ha sido necesario aplicar procesos de limpieza y reponderación de la población, por dos motivos: i) corregir sesgos poblacionales de respuesta, sobre- e infraponderación de segmentos; y ii) eliminar errores y duplicidades de datos, para evitar las distorsiones poblacionales de oferta mencionadas por Boeing (2020). Para el cálculo de los elevadores muestrales, se ha llevado a cabo un proceso de correspondencia estadística (D’Orazio et al., 2006; Leucescu & Agafitei, 2013) mediante calibración (Deville, 2000; Deville & Särndal, 1992) de pesos, usando como datos auxiliares la encuesta de Presupuestos Familiares (EPF) y el Censo de Población y Viviendas desarrollados por el INE (2006).

Por otra parte, el dato de mercado procede del Sistema Estatal de Índices de Referencia del Precio del Alquiler de Vivienda, del MITMA. Esta fuente recoge varias tablas estadísticas del mercado de la vivienda, procedentes de las declaraciones de renta de las personas físicas, y cuenta con series de precios anuales desde 2015 a 2020, a nivel de sección censal. Entre las medidas que recoge se encuentran el número de viviendas por zona y el precio mediano, y los percentiles 25 y 75 de precios, totales y por metro cuadrado. Para nuestro caso, se toman las medianas por sección censal y se agregan, para las zonas de trabajo, como medias ponderadas

---

4 La agrupación de varios municipios se realiza en los municipios con un número de registros insuficiente.

según el número de inmuebles del MITMA. Adicionalmente, se incluyen los ingresos anuales medios por hogar (*RENTA*) para cada zona de estudio<sup>5</sup> y año (INE, 2018).

El diccionario de datos de la tabla de datos sobre la que ha trabajado se describe en la Tabla 1.

**TABLA 1 | Diccionario de datos de la muestra**

VARIABLE	DESCRIPCIÓN
TIPO	Tipo de vivienda: Unifamiliar (casa independiente o adosada) o Plurifamiliar (piso).
YEAR	Año al que se refiere el dato.
CÓDIGO	Código de zona (basado en código INE/Municipal de barrio o municipio).
NOMBRE	Nombre de la zona.
CAPROV	Indica si la zona está en capital de provincia (valor 1) o no (valor 0).
ASKING_UNITPRICE	Media ponderada del precio de oferta para la zona en €/m <sup>2</sup> construidos.
FACTOR_MITMA	Factor de elevación poblacional MITMA.
FACTOR_ASKING	Factor de elevación poblacional dato Idealista.
LEADS	Número medio normalizado de contactos al mes por zona.
RENTA	Renta media por hogar.

FUENTE: ELABORACIÓN PROPIA

La variable LEADS contiene la demanda relativa de la zona, calculada como el número medio de contactos a propietarios por las personas que buscan inmuebles.<sup>6</sup> Esta medida permite comparar zonas por el nivel de interés de sus demandantes de vivienda.

Se trabajará con 1770 registros procedentes de MITMA, correspondientes a 168.385 hogares de 177 zonas para cinco años. Se consideran 65 zonas municipales y 112 barrios en Madrid, correspondientes al 79,54% y 8,79% de los hogares en viviendas plurifamiliares, respectivamente. Para viviendas unifamiliares, la muestra representa el 11,19% de los hogares para el resto de la Comunidad, y el 0,48% para la capital. Debido al bajo número de hogares unifamiliares en Madrid, se omite este estrato del análisis por insuficiencia estadística.

A modo de ejemplo, se muestran tres series de precios en la Figura 1: el barrio del Pilar en Madrid, un municipio del área metropolitana (Getafe) y un municipio fuera del área metropolitana (Mejorada del Campo).

## Metodología

El análisis se ha realizado sobre la relación entre el precio de mercado y el precio de oferta, y se expresa como la ratio mercado-oferta,  $\beta_{m,o}$  y cuya forma funcional lineal sería:

$$P_m = \beta_0 + \beta_{m,o} \cdot P_o + \varepsilon \quad (1)$$

5 El dato se calcula como la media de la renta por hogar en todas las secciones que conforman el área geográfica de análisis.

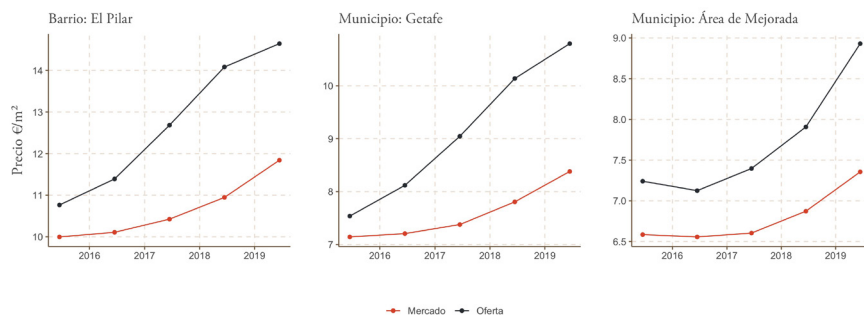
6 La zona que menos contactos recibe toma valor 1 y la que más, 100; el resto de los valores se calcula de forma proporcional entre el valor máximo y mínimo.



donde  $P_m$  representa el precio medio de mercado para la zona,  $P_o$  el precio medio de oferta,  $\beta_o$  la intersección y  $\varepsilon$  un término de error aleatorio.

La ratio general se puede estudiar a través de los coeficientes de correlación entre ambas magnitudes, mostrados gráficamente en la Figura 2. En ella se puede confirmar que ambos valores están fuertemente correlacionados, y que una de las magnitudes se podría expresar como una combinación lineal de la otra.

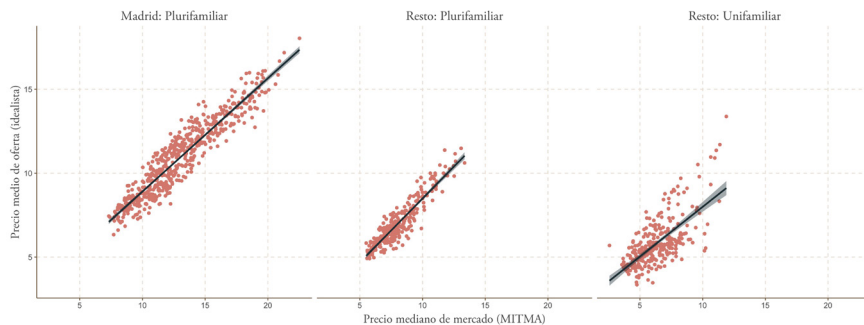
**FIGURA 1 | Series de precios €/m<sup>2</sup>/mes, oferta y mercado**



FUENTE: ELABORACIÓN PROPIA

Para el caso de la vivienda unifamiliar, esta relación muestra un mayor grado de dispersión, particularmente en el segmento de precios más altos.

**FIGURA 2 | Correlación entre precio de mercado y precio de oferta**



FUENTE: ELABORACIÓN PROPIA

Se puede confirmar numéricamente la fortaleza de la relación entre los precios de mercado y los de oferta a través del coeficiente de correlación de Pearson entre precios. Dicha relación es muy fuerte en vivienda plurifamiliar para la capital (0,94), menor (0,92) para el resto de los municipios, y mucho menor en unifamiliares para municipios fuera de la capital (0,62). Estos valores son consistentes con lo observado en otros mercados, como el de Chapelle y Eyméoud para Francia (2022), que logra un coeficiente de correlación del 0,95, o el de Kokot y Bas (2015) para Polonia, con un 0,95.

Para profundizar en el comportamiento de la interacción de los dos precios, se estudia cómo es esta relación teniendo en cuenta los siguientes estratos del mercado: macrozonales, por tipología, y nivel de demanda. Se ha decidido usar una estratificación zonal por la exposición de los precios de la vivienda a fenómenos de heterogeneidad espacial, cuestión que se analiza en profundidad en Hu et al. (2022), Wu et al. (2020), Helbich et al. (2014), Páez et al. (2008) y Kestens et al. (2006).

El primer análisis estima la relación lineal entre las magnitudes, en un panel que realiza un estudio transversal sobre grandes demarcaciones geográficas y tipo de vivienda. Las relaciones entre los precios de mercado y de oferta, expresadas como los coeficientes, se calculan solo sobre las distintas interacciones entre tipo de zona (Madrid, Resto), tipo de vivienda (unifamiliar, plurifamiliar) y nivel de demanda (medio, alto y bajo). La siguiente ecuación expresa la relación funcional entre ambas magnitudes. Por una cuestión de legibilidad, se omiten las interacciones de los precios con los niveles de demanda:

$$P_m = \beta_0 + P_o \cdot [\beta_{p,M} \cdot L_M \cdot T_p + \beta_{p,R} \cdot L_R \cdot T_p + \beta_{u,R} \cdot L_R \cdot T_u] + \varepsilon \quad (2)$$

donde  $\beta_{p,M}$  es el coeficiente para la vivienda plurifamiliar en Madrid,  $\beta_{p,R}$  y  $\beta_{u,R}$  lo son para el resto de la región en los segmentos plurifamiliar y unifamiliar, respectivamente;  $L_M$  es una variable ficticia dicotómica que toma valor 1 si el registro está en Madrid y 0 en el resto de los municipios;  $L_R$  tiene la misma función, pero para el resto de la región; y finalmente  $T_p$  y  $T_u$  son variables dicotómicas que indican si la observación se refiere a un tipo plurifamiliar o unifamiliar.

El panel estudia la relación de las magnitudes de precio para cada una de las zonas. El desglose zonal se divide en tres macroestratos: vivienda plurifamiliar por barrio en la ciudad de Madrid, vivienda plurifamiliar para el resto de la región y vivienda unifamiliar para este último ámbito. El estimador de regresión lineal general se especifica como:

$$P_m = \beta_0 + \sum_{i \in L} \beta_i \cdot P_o \cdot L_i + \varepsilon, \forall i \in L \quad (3)$$

donde  $\beta_i$  será el coeficiente para la zona  $i$  dentro del conjunto de zonas  $L$  del ámbito;  $L_i$  es la variable de tipo dicotómico que toma un valor 1 si el registro está en la zona  $i$  y 0 en caso contrario. La medida de ajuste de los distintos modelos se calcula a través del coeficiente de determinación  $R^2$  ajustado, una medida estadística que recoge la proporción de la varianza de la variable dependiente y se explica por un modelo sobre las variables independientes. Su ventaja sobre otras medidas, como el error cuadrático medio o el error medio absoluto, es que es más informativo (Chicco et al., 2021).

Finalmente, dado que también se dispone del dato de rentas, se evaluará si las series de precios están correlacionadas con el nivel de ingresos de la zona. A este respecto, las opiniones se reparten entre los que argumentan que existe cointegración entre precio y fundamentales macroeconómicos (Kishor & Marfatia, 2017; Xu & Zhang, 2022), los que no (Gallin, 2006; Mikhed & Zemčik, 2009), y los

que estiman que esta afirmación depende del submercado inmobiliario (Bangura & Lee, 2022). En nuestro caso se estudiará solo el grado de correlación entre estas magnitudes, para lo cual se calcula la razón precio ingresos ( $RI$ ), en las magnitudes de oferta ( $RI_{o,z}$ ) y de mercado ( $RI_{m,z}$ ), expresadas como:

$$RI_{m,z} = \frac{P_{m,z}}{R_{s,z}} \quad (4) \quad RI_{o,z} = \frac{P_{o,z}}{R_{s,z}} \quad (4)$$

donde  $R_s$  es la renta por hogar anual de la zona  $z$ ,  $P_{o,z}$  y  $P_{m,z}$  los precios de oferta y mercado para esta zona.

### Discusión de resultados

Conforme al objeto del trabajo, los resultados de los modelos permitirán evaluar tres aspectos, mediante una interpretación de la validez y magnitud de los coeficientes obtenidos: i) la existencia o no de un patrón geográfico asociado a la distribución de los coeficientes; ii) la existencia o no de una relación de la proporción precio oferta/mercado y la demanda de cada área; y iii) si el precio de oferta es mejor predictor de los ingresos que el precio final.

La Tabla 2 muestra los coeficientes de regresión para la segmentación por demanda y tipo. Se obtiene un coeficiente de determinación  $R^2$  de 0,93, con un total de 1121 grados de libertad, un p-valor menor a  $2,2 \times 10^{-6}$  y estadístico F de 1628. Puede deducirse que Madrid ofrece mayores niveles de confianza en la estimación de las betas, a la vista que toman valores t más altos, debido a que toda la ciudad tiene un alto nivel de demanda. Por otra parte, las viviendas de tipo unifamiliar para el resto de la región ofrecen una confianza menor, al contener una gran variedad de zonas residenciales (desde un punto de vista morfológico y demográfico), cuyos niveles de demanda son diferentes entre sí. Existe un factor adicional, ligado a la menor representatividad, ya que en estos lugares existe un menor número de observaciones.

**TABLA 2 | Coeficientes regresión precio de mercado y de oferta**

TIPO	ZONA	COEFICIENTE ( $\beta$ )	STD.ERR	T-VALUE	P-VALUE	SIGNIF
(Intercepto)		1.904	0.105	18.194	< 2e-16	***
Plurifamiliar	Madrid	0.693	0.008	85.921	< 2e-16	***
Plurifamiliar	Resto	0.640	0.014	46.690	< 2e-16	***
Unifamiliar	Resto	0.618	0.017	36.098	< 2e-16	***

FUENTE: ELABORACIÓN PROPIA

Son destacables las diferencias entre tipologías para el resto de la Comunidad, con un valor más bajo en la vivienda unifamiliar, ligado a que una parte importante se ubica en zonas de baja demanda (principalmente rurales). Por tanto, una zona de menor interés dará un mayor poder de negociación al inquilino y las diferencias entre el precio de oferta y negociado serán menores; consecuentemente, la función

que relaciona el precio de mercado y el de oferta tendrá una menor pendiente, cuestión que se podrá comprobar gráficamente más adelante en la Figura 5b.

Coefficientes menores indican un menor margen entre la oferta y el precio real, y se asocian a mercados más líquidos, caracterizados por una mayor demanda, tal y como argumentan Han y Strange (2016). Para el caso del alquiler en Madrid, la relación oferta/mercado es inversamente proporcional al número de pujas, y no sigue de forma tan estricta el patrón argumentado por Han y Strange (2016). La causa podría ser la existencia de heterogeneidad espacial en la medida de la demanda, u otros factores específicos de cada mercado no controlados por el modelo (por ejemplo, la rentabilidad de las inversiones inmobiliarias). De la misma manera, un mayor coeficiente de la vivienda plurifamiliar se interpreta como una mayor diferencia entre los precios de oferta y mercado con respecto a la unifamiliar; además, esta última muestra un mayor nivel de variabilidad, representado por un mayor error estándar y debida a una mayor heterogeneidad de la misma. La Tabla 3 y la Tabla 4 representan, con una muestra de tres zonas por quintil, los coeficientes de precios para Madrid y resto de municipios, respectivamente.

**TABLA 3 | Coeficientes regresión precio barrios de Madrid**

QUINTIL	ZONA	COEFICIENTE ( $\beta$ )	STD.ERR	T-VALOR	P-VALOR	SIGNIF.
	(Intercepto)	4,965	0,137	36,3	< 2e-16	***
1	Entrevías	0,234	0,023	10,0	< 2e-16	***
	San Andrés	0,259	0,022	11,9	< 2e-16	***
	Los Ángeles	0,275	0,022	12,3	< 2e-16	***
2	Numancia	0,363	0,019	18,9	< 2e-16	***
	Comillas	0,363	0,018	20,3	< 2e-16	***
	Fontarrón	0,364	0,021	17,1	< 2e-16	***
3	Concepción	0,436	0,016	27,1	< 2e-16	***
	Canillas	0,439	0,017	26,2	< 2e-16	***
	Valdezarza	0,444	0,017	26,2	< 2e-16	***
4	Palacio	0,493	0,013	39,4	< 2e-16	***
	Vallehermoso	0,493	0,012	39,7	< 2e-16	***
	Cortes	0,494	0,012	40,7	< 2e-16	***
5	Ríos Rosas	0,522	0,012	41,9	< 2e-16	***
	El Viso	0,522	0,013	41,6	< 2e-16	***
	Universidad	0,523	0,012	43,6	< 2e-16	***

FUENTE: ELABORACIÓN PROPIA

TABLA 4 | Coeficientes municipios resto de la Comunidad

QUINTIL	ZONA	COEFICIENTE ( $\beta$ )	STD.ERR	T-VALOR	P-VALOR	SIGNIF.
	(Intercepto)	3,599	0,112	32,003	< 2e-16	***
1	Área de San Martín de la Vega	0,256	0,024	10,813	< 2e-16	***
	Área de Collado Mediano	0,264	0,021	12,825	< 2e-16	***
	Área de Guadalix de la Sierra	0,279	0,022	12,621	< 2e-16	***
2	Área de Algete	0,347	0,019	18,534	< 2e-16	***
	Área Meco	0,349	0,023	15,233	< 2e-16	***
	Área de Manzanares el Real	0,359	0,020	18,327	< 2e-16	***
3	El Escorial	0,386	0,019	20,348	< 2e-16	***
	Área de Brunete - Quijorna	0,386	0,018	21,053	< 2e-16	***
	Área de Sevilla la Nueva	0,390	0,022	18,030	< 2e-16	***
4	Área de Cobeña	0,439	0,018	23,961	< 2e-16	***
	Leganés	0,440	0,016	26,730	< 2e-16	***
	Alcalá de Henares	0,446	0,018	24,836	< 2e-16	***
5	Torrelodones	0,503	0,014	35,955	< 2e-16	***
	Alcobendas	0,509	0,013	38,192	< 2e-16	***
	Área de San Agustín de Guadalix	0,518	0,018	28,350	< 2e-16	***

FUENTE: ELABORACIÓN PROPIA

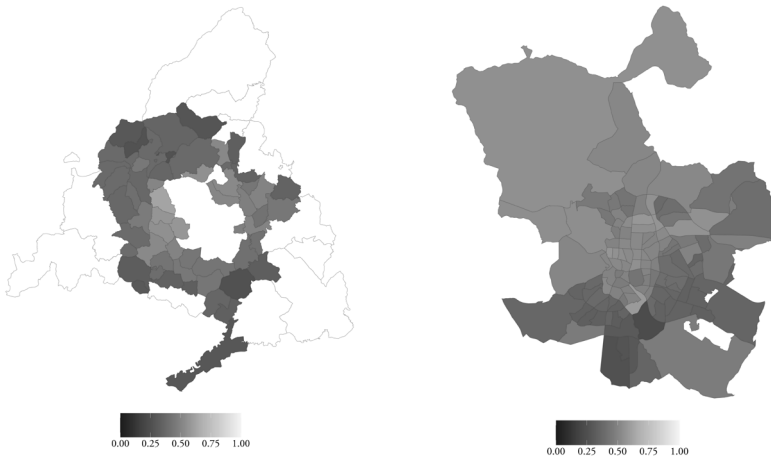
El desglose zonal en la capital sobre los grupos de quintiles muestra una notable variabilidad de los coeficientes, con una media de 0,445 y una desviación típica de 0,081, aunque todos son significativos, con un  $R^2$  de 0,98, un total de 447 grados de libertad, un p-valor menor a  $2,2 \times 10^{-6}$  y un estadístico F de 187,6.

Para el resto de la Comunidad de Madrid se aprecia un comportamiento equivalente al anterior, donde todas las zonas tienen betas significativas, su media es 0,414 y su desviación típica de 0,086, tiene un  $R^2$  de 0,98, con un total de 219 grados de libertad, un p-valor menor a  $2,2 \times 10^{-6}$  y estadístico F de 226,2. En ambos casos, los valores t indican que aquellas zonas donde existe una mayor proximidad entre el precio de mercado y el de oferta (betas mayores) ofrecen menor variabilidad en la estimación, que puede estar asociada a menor poder de negociación del inquilino, que redundaría en menos cambios de precio por parte de los arrendatarios, lo cual es consistente con la apreciación sobre los valores t descrita para la Tabla 2.

Gráficamente, se observa que las zonas con mayores coeficientes se distribuyen en torno a la capital y su anillo concéntrico inmediatamente exterior, como muestra la Figura 3. Se confirma la existencia de distintos niveles de representación en las áreas de estudio (Boeing, 2020; Boeing & Waddell, 2017), con mejor representación de las zonas urbanas altamente pobladas. Predominan los valores altos en la zona metropolitana noroeste, formada por las zonas con mayor nivel adquisitivo de la región, lo que confirma la hipótesis de Boeing (2020), quien afirma que las

áreas con mayores ingresos tienden a sobrerrepresentarse.<sup>76</sup> En cambio, son las zonas rurales (zonas agrícolas del sur y zona de la Sierra de Madrid) las que tienen los coeficientes más bajos. Esto coincide con el argumento de Boeing et al. (2021), según el cual las zonas minoritarias también adolecen de falta de representación en las fuentes de internet, y ratifica los resultados de Chapelle y Eyméoud (2022), que muestran una relación inversamente proporcional entre la variabilidad de los precios implícitos y la densidad de población.

**FIGURA 3 | Coeficiente mercado-oferta para la comunidad de Madrid**



(a) Resto de la Comunidad de Madrid

(b) Ciudad de Madrid

FUENTE: ELABORACIÓN PROPIA

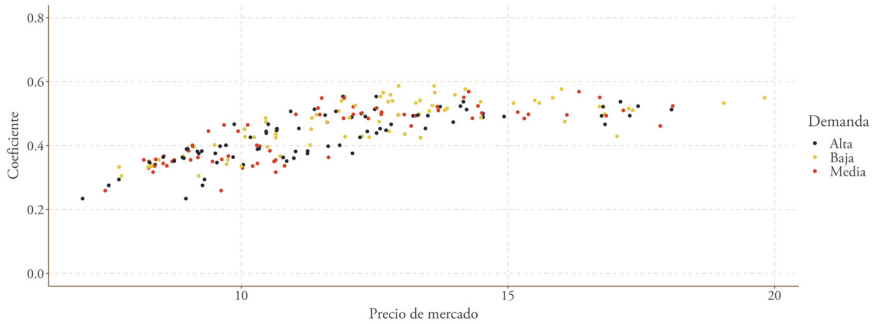
Los coeficientes dentro la ciudad ofrecen una clara diferencia entre el norte y oeste, y el sur y este: en el primer grupo se concentran los coeficientes más altos y en el segundo, los más bajos. Tal diferencia se observa también en los niveles de ingresos entre estos dos grupos: el primero congrega zonas con altos precios y alto poder adquisitivo, mientras que el sur y este abarcan las zonas con menos ingresos y precios medios más bajos.

De Wit et al. (2013) afirman que los mercados más líquidos muestran menores diferencias entre el precio final y el de oferta; por otra parte, Han y Strange (2016) indican que esta diferencia es inversamente proporcional al número de pujas que recibe en inmueble en su comercialización. Para comprobar si la proporción entre la intensidad de la demanda y el precio depende de los niveles de la última, se evalúa la relación entre la diferencia de precios (a través del coeficiente de zona) y el número medio de contactos que reciben los anuncios en cada zona de estudio.

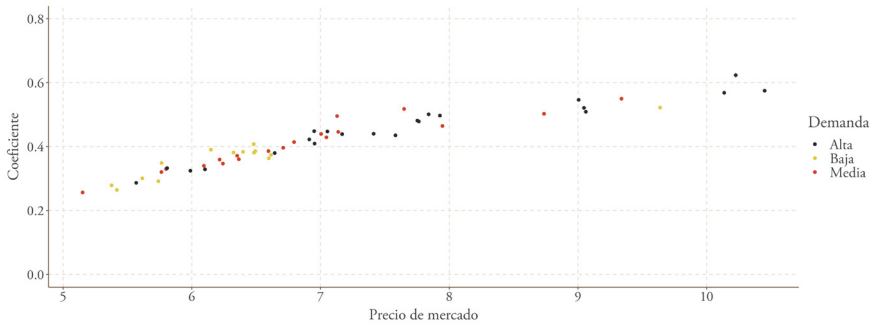
7 Sobre la base de que los estratos de mayores ingresos tienen mayor adopción de medios tecnológicos y, por tanto, mayor probabilidad de sobrerrepresentar el uso de portales inmobiliarios.

Se confirma esta hipótesis, aunque con determinado grado de incertidumbre y diferencias entre zonas, pues se observa un patrón que relaciona mayores precios a coeficientes. En cuanto a la demanda, no se aprecia un patrón fuerte que se pueda relacionar al precio o los coeficientes para las demandas baja y media, lo que probablemente sea consecuencia de la heterogeneidad espacial en el comportamiento del número de contactos<sup>87</sup> (Helbich et al., 2014). Tal y como se aprecia en la Figura 4, las zonas con mayores precios tienen una relación mercado/oferta mucho más alta, lo que confirma que estas zonas tienen un margen más estrecho entre oferta y mercado. Por otra parte, las áreas con precios menores muestran diferencias mucho más amplias. También ocurre que a partir de los 15 €/m<sup>2</sup> no existen diferencias en el coeficiente.

**TABLA 5 | Relación entre el coeficiente y el precio de mercado, vivienda plurifamiliar**



(a) Barrios de la ciudad de Madrid



(b) Municipios del resto de la Comunidad de Madrid

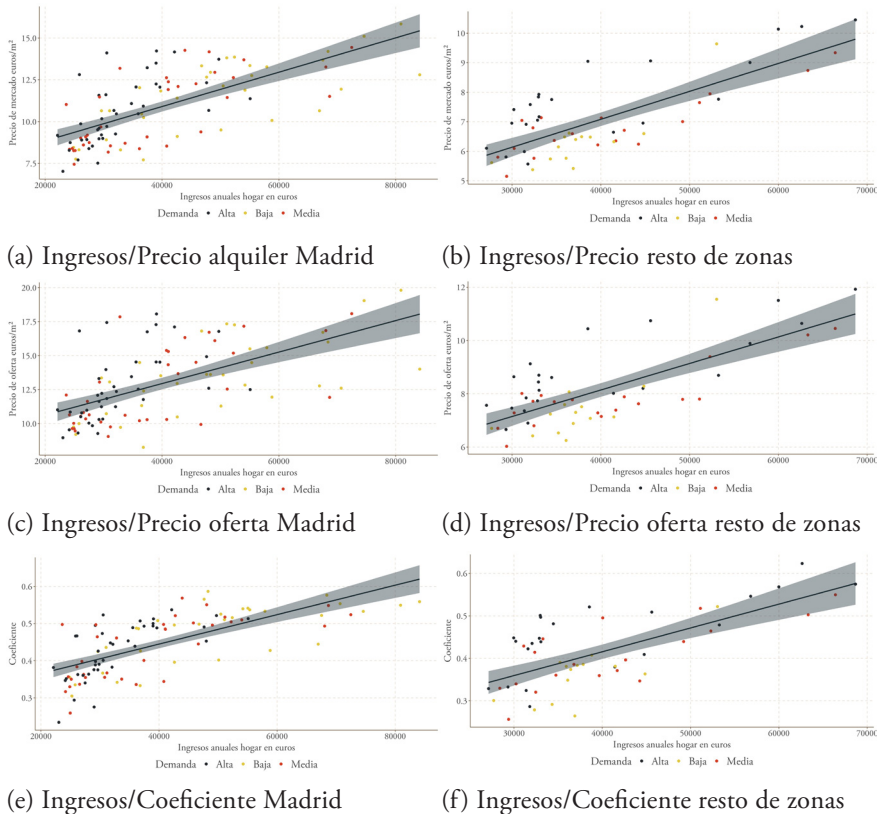
FUENTE: ELABORACIÓN PROPIA

8 La heterogeneidad espacial se refiere a que la proporción entre contactos por anuncio y precio toma valores diferentes en función de la zona.

Para el resto de la Comunidad se aprecia un patrón más claro en el nivel de demanda, donde los valores bajos y medios se corresponden con niveles menores de demanda, que estarían relacionados, de nuevo, con un margen de negociación a favor del propietario (Han & Strange, 2014). En el segmento de los precios más altos existe mayor dispersión, aunque –como en el caso de Madrid– hay una relación positiva entre el precio de mercado y el coeficiente.

Sobre la hipótesis de que el precio de la vivienda está relacionado con los ingresos de la zona, se podría considerar el precio de oferta como un predictor de los ingresos familiares. En este sentido, Gan y Hill (2008) confirman esta tesis, aunque las diferencias específicas de las condiciones locales del mercado requieren un cálculo específico por zona. En el presente trabajo se pretende, además, corroborar que en nuestro caso tanto el precio de oferta como el de mercado son predictores válidos de los ingresos.

**FIGURA 4 | Relación precio de mercado y renta anual por hogar, vivienda plurifamiliar**



FUENTE: ELABORACIÓN PROPIA



El análisis de la ponderación de la oferta ofrece una correlación entre precio de oferta e ingresos similar a la del precio de mercado y los ingresos, como muestra la Figura 5. Dado que la dispersión sobre las líneas de regresión “Ingresos/Coeficiente” son menores que las de “Ingresos/Precio oferta”, se deduce que el proceso de reponderación utilizado sobre la oferta gestiona los potenciales sesgos en el uso de datos procedentes de internet identificados por Boeing (2020).

Se observa que los ingresos medios por hogar están altamente relacionados con los precios de los alquileres, tanto para Madrid como para el resto de la región, como se desprende de la Figura 5, confirmando, por tanto, la tesis de Gan y Hill (2008).

El nivel de demanda, en cambio, no es demasiado informativo, con una mayor correlación en el resto de los municipios que en la capital, aunque en ambos con una gran dispersión de valores. Ello puede indicar que existe una cointegración entre ambas magnitudes, como defienden Kishor y Marfatia (2017) y Xu y Zhang (2022), aunque esta relación no es homogénea en la geografía, debido a los distintos niveles de liquidez de los mercados de estudio.

**TABLA 6 | Coeficientes de determinación entre precio e ingresos**

ZONA	COEFICIENTE SOBRE INGRESOS	COEFICIENTE SOBRE PRECIO DE MERCADO	COEFICIENTE SOBRE PRECIO OFERTA
Madrid	0,769	0,722	0,647
Resto	0,569	0,575	0,561

FUENTE: ELABORACIÓN PROPIA

Los coeficientes de correlación de Spearman entre las series de precios e ingresos (Tabla 5) indican una fuerte relación entre las magnitudes de renta y precio de la vivienda. En Madrid, la relación entre ingresos y coeficientes es ligeramente más fuerte que su equivalente por unidad de superficie.

Resulta importante destacar cómo la relación se diluye cuando se comparan los ingresos con el precio de oferta. Estos resultados son consistentes con la tesis de que existe una relación persistente entre la variable de precios y renta, pero que el grado varía en función del submercado inmobiliario de estudio (Bangura & Lee, 2022; Xu & Zhang, 2022). Por otra parte, un menor grado de correlación entre los coeficientes y el precio de mercado puede estar relacionado con el ruido que mencionan Shimizu et al. (2016), y que es consecuencia de una muestra con mayores sesgos poblacionales, por la sobrerrepresentación de los inmuebles menos líquidos, aquellos que tienden a permanecer más tiempo en el stock de oferta (Ardila et al., 2021).

## Conclusiones

El presente estudio comprueba empíricamente que la información de oferta de alquiler procedente del portal Idealista, correctamente tratada, es una medida válida para estimar los precios de mercado oficiales, a pesar de que no exista una cobertura completa. Esto supone una ventaja importante para los distintos actores públicos y privados del mercado inmobiliario, ya que con ello disponen de un mecanismo actualizado de estimación de la serie oficial.

De forma particular, los resultados del estudio demuestran que los datos de portales inmobiliarios, convenientemente tratados, permiten la construcción de índices del precio fieles a la realidad del mercado de la vivienda. Además, dada la frecuente actualización de la información, permiten que dichos índices se generen con pocas semanas o días de retraso, eliminando el desfase temporal de meses propio de los datos oficiales. Por último, estas fuentes cuentan con un número mayor de atributos, que permiten el estudio en detalle de los fenómenos inmobiliarios.

Además, se ha confirmado que el precio de oferta es un buen predictor de los precios de mercado, en particular para las viviendas de tipo plurifamiliar. El ajuste de los pesos poblacionales de la oferta permite un uso de la fuente *online* sin incorporar medidas importantes de sesgo que, habitualmente, se producen al usar datos de oferta.

El estudio del comportamiento de la variable relación precio de mercado y precio de oferta, y de sus interacciones con la renta por hogar, puede ser una herramienta adecuada para la identificación de submercados de la vivienda, y evitar el uso de unidades de análisis geográfico arbitrarias o con fines diferentes, como la postal o la administrativa.

Una vez demostrada la viabilidad de los análisis en una ciudad con un volumen de actividad muy amplio, como es el caso de Madrid, convendría extender el ámbito del análisis a otras regiones españolas y estudiar los factores que determinan las distintas relaciones entre precio y oferta en el mercado. También existe una segunda línea de investigación, que se centrará en el uso de esta medida para desarrollar una segmentación zonal mucho más granular, ya que a través de modelos hedónicos es posible estimar los precios de mercado y de oferta a nivel de manzana o finca catastral, lo que permitiría construir una delimitación de submercados de la vivienda urbana mucho más apropiada.

### Referencias bibliográficas

- Alfaro-Navarro, J.-L., Cano, E. L., Alfaro-Cortés, E., García, N., Gámez, M. & Larraz, B. (2020). A fully automated adjustment of ensemble methods in machine learning for modeling complex real estate systems. *Complexity*, 2020. <https://doi.org/10.1155/2020/5287263>
- Allen, M. & Dare, W. (2004). The effects of charm listing prices on house transaction prices. *Real Estate Economics*, 32(4), 695-713. <https://doi.org/10.1111/j.1080-8620.2004.00108.x>
- Anenberg, E. & Laufer, S. (2017). A more timely house price index. *Review of Economics and Statistics*, 99(4), 722-734. [https://doi.org/10.1162/REST\\_a\\_00634](https://doi.org/10.1162/REST_a_00634)
- Ardila, D., Ahmed, A. & Sornette, D. (2021). Comparing ask and transaction prices in the Swiss housing market. *Quantitative Finance and Economics*, 5(1), 67-93. <https://doi.org/10.3929/ethz-b-000495140>

- Arribas-Bel, D. & Bakens, J. (2019). Use and validation of location-based services in urban research: An example with Dutch restaurants. *Urban Studies*, 56(5), 868-884. <http://dx.doi.org/10.1177/0042098018779554>
- Arribas-Bel, D., Green, M., Rowe, F. & Singleton, A. (2021). Open data products-A framework for creating valuable analysis ready data. *Journal of Geographical Systems*, 23(4), 497-514. <https://doi.org/10.1007/s10109-021-00363-5>
- Bangura, M. & Lee, C. L. (2022). Housing price bubbles in Greater Sydney: evidence from a submarket analysis. *Housing Studies*, 37(1), 143-178. <https://doi.org/10.1080/02673037.2020.1803802>
- Besbris, M., Schachter, A. & Kuk, J. (2018). The language of segregation: Variation in the description of rental housing across neighborhoods. En *American Sociological Association Annual Meeting*. Philadelphia, PA.
- Beullens, K., Matsuo, H., Loosveldt, G. & Vandenplas, C. (2014). *Quality report for the European Social Survey, round 6*. European Social Survey ERIC.
- Biancotti, C., Rosolia, A., Veronese, G., Kirchner, R. & Mouriaux, F. (2020). Covid-19 and official statistics: a wakeup call? *Questioni di Economia e Finanza (Occasional Papers)*, (605). <https://dx.doi.org/10.2139/ssrn.3828122>
- Boeing, G. (2020). Online rental housing market representation and the digital reproduction of urban inequality. *Environment and Planning A: Economy and Space*, 52(2), 449-468. <https://doi.org/10.1177/0308518X19869678>
- Boeing, G., Besbris, M., Schachter, A. & Kuk, J. (2021). Housing search in the age of big data: smarter cities or the same old blind spots? *Housing Policy Debate*, 31(1), 112-126. <https://doi.org/10.1080/10511482.2019.1684336>
- Boeing, G. & Waddell, P. (2017). New insights into rental housing markets across the United States: Web scraping and analyzing craigslist rental listings. *Journal of Planning Education and Research*, 37(4), 457-476. <https://doi.org/10.1177/0739456X16664789>
- Bricongne, J.-C., Meunier, B. & Pouget, S. (2023). Web-scraping housing prices in real-time: The Covid-19 crisis in the UK. *Journal of Housing Economics*, 59, Part B, 101906. <https://doi.org/10.1016/j.jhe.2022.101906>
- Chapelle, G. & Eyméoud, J. B. (2022). Can big data increase our knowledge of local rental markets? A dataset on the rental sector in France. *PLoS One*, 17(1), e0260405. <https://doi.org/10.1371/journal.pone.0260405>
- Chauvet, M., Gabriel, S. & Lutz, C. (2013). *Fear and loathing in the housing market: Evidence from search query data*. <http://dx.doi.org/10.2139/ssrn.2148769>
- Chicco, D., Warrens, M. J. & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, 7, e623. <https://doi.org/10.7717/peerj-cs.623>
- De Wit, E. R., Englund, P. & Francke, M. K. (2013). Price and transaction volume in the Dutch housing market. *Regional Science and Urban Economics*, 43(2), 220-241. <https://doi.org/10.1016/j.regsciurbeco.2012.07.002>
- Desormeaux, D. & Piguillem, F. (2003). Precios hedónicos e índices de precios de viviendas. *Documento de trabajo*, nº 12.
- Deville, J.-C. (2000). Generalized calibration and application to weighting for non-response. En J. G. Bethlehem & P. G. M. van der Heijden (Eds.), *COMPSTAT* (pp. 65-76). Physica, Heidelberg. [https://doi.org/10.1007/978-3-642-57678-2\\_6](https://doi.org/10.1007/978-3-642-57678-2_6)

- Deville, J.-C. & Särndal, C. E. (1992). Calibration estimators in survey sampling. *Journal of the American statistical Association*, 87(418), 376-382. <https://doi.org/10.2307/2290268>
- D'Orazio, M., Di Zio, M. & Scanu, M. (2006). *Statistical matching: Theory and practice*. John Wiley & Sons.
- Eerola, E. & Lyytikäinen, T. (2015). On the role of public price information in housing markets. *Regional Science and Urban Economics*, 53, 74-84. <https://doi.org/10.1016/j.regsciurbeco.2015.05.006>
- Eurostat. (2017). HICP Recommendation on Obtaining Scanner Data.
- Folch, D. C., Spielman, S. E. & Manduca, R. (2018). Fast food data: Where user-generated content works and where it does not. *Geographical analysis*, 50(2), 125-140. <https://doi.org/10.1111/gean.12149>
- Galesi, A., Mata, N., Rey, D., Schmitz, S. & Schuffels, J. (2020). *Regional housing market conditions in Spain*. Research Memorandum 029, Maastricht University, Graduate School of Business and Economics (GSBE). <https://doi.org/10.26481/umagsb.2020029>
- Gallin, J. (2006). The long-run relationship between house prices and income: evidence from local housing markets. *Real Estate Economics*, 34(3), 417-438. <https://dx.doi.org/10.2139/ssrn.410808>
- Gan, Q. & Hill, R. (2008). *A new perspective on the relationship between house prices and income*. School of Economics University of New South Wales. <https://dx.doi.org/10.2139/ssrn.1313073>
- Han, L. & Strange, W. (2014). Bidding wars for houses. *Real Estate Economics*, 42(1), 1-32. <https://doi.org/10.1111/reec.12015>
- Han, L. & Strange, W. (2016). What is the role of the asking price for a house? *Journal of Urban Economics*, 93, 115-130. <https://doi.org/10.1016/j.jue.2016.03.008>
- Helbich, M., Brunauer, W., Vaz, E. & Nijkamp, P. (2014). Spatial heterogeneity in hedonic house price models: The case of Austria. *Urban Studies*, 51(2), 390-411. <https://doi.org/10.1177/0042098013492234>
- Hill, R. J. & Trojanek, R. (2022). An evaluation of competing methods for constructing house price indexes: The case of Warsaw. *Land Use Policy*, 120, 106226. <https://doi.org/10.1016/j.landusepol.2022.106226>
- Horowitz, J. L. (1992). The role of the list price in housing markets: theory and an econometric model. *Journal of Applied Econometrics*, 7(2), 115-129. <https://doi.org/10.1002/jae.3950070202>
- Hu, L., He, S. & Su, S. (2022). A novel approach to examining urban housing market segmentation: Comparing the dynamics between sales submarkets and rental submarkets. *Computers, Environment and Urban Systems*, 94, 101775. <https://doi.org/10.1016/j.compenvurbsys.2022.101775>
- Instituto Nacional de Estadística (INE), España. (2006). *Metodología de la Encuesta de presupuestos familiares. Base 2006*. <https://www.ine.es/>
- Instituto Nacional de Estadística (INE), España. (2018). *Atlas de distribución de la renta de los hogares. Resultados por municipios, distritos y secciones censales*. [https://www.ine.es/dyngs/INEbase/operacion.htm?c=Estadistica\\_C&cid=1254736177088&menu=resultados&cidp=1254735976608](https://www.ine.es/dyngs/INEbase/operacion.htm?c=Estadistica_C&cid=1254736177088&menu=resultados&cidp=1254735976608).
- Instituto Nacional de Estadística (INE), España. (2021). Índice de Precios de la Vivienda en Alquiler.

- Instituto Nacional de Estadística (INE), España. (2022a). *Datos del Índice de Precios de la Vivienda - Diciembre 2022*. [https://www.ine.es/prensa/ipv\\_tabla1.htm](https://www.ine.es/prensa/ipv_tabla1.htm)
- Instituto Nacional de Estadística (INE), España. (2022b). Estudio piloto de movilidad a partir del posicionamiento de teléfonos móviles.
- Instituto Nacional de Estadística (INE), España. (2022c). Cifras oficiales de población a 1 de enero de 2021.
- Kestens, Y., Thériault, M. & Des Rosiers, F. (2006). Heterogeneity in hedonic modelling of house prices: looking at buyers' household profiles. *Journal of Geographical Systems*, 8(1), 61-96. <http://dx.doi.org/10.1007/s10109-005-0011-8>
- Kishor, N. K. & Marfatia, H. A. (2017). The dynamic relationship between housing prices and the macroeconomy: Evidence from OECD countries. *The Journal of Real Estate Finance and Economics*, 54(2), 237-268. <https://doi.org/10.1007/s11146-015-9546-8>
- Kokot, S. & Bas, M. (2015). The comparative analysis of asking and traded price indices in different floor area subsegments of the residential property market. *Real Estate Management and Valuation*, 23(3), 14-25. <https://doi.org/10.1515/remav-2015-0021>
- Kolbe, J., Schulz, R., Wersing, M. & Werwatz, A. (2021). Real estate listings and their usefulness for hedonic regressions. *Empirical economics*, 61(6), 3239-3269. <https://doi.org/10.1007/s00181-020-01992-3>
- Lacerda, N. (2018). Mercado imobiliário de bens patrimoniais: um modelo interpretativo a partir do centro histórico do Recife (Brasil). *Revista EURE – Revista de Estudos Urbano Regionales*, 44(132), 113-132. <http://dx.doi.org/10.4067/s0250-71612018000200089>
- Leucescu, A. & Agafitei, M. (2013). Statistical matching: a model based approach for data integration. *Eurostat methodologies and Working papers*. <https://doi.org/10.2785/44822>
- Loberto, M., Luciani, A. & Pangallo, M. (2018). The potential of big housing data: an application to the Italian real-estate market. Temi di discussione (Working Papers), n° 1171. Banca d'Italia, Eurosystema. [https://www.bancaditalia.it/publicazioni/temi-discussione/2018/2018-1171/en\\_tema\\_1171.pdf?language\\_id=1](https://www.bancaditalia.it/publicazioni/temi-discussione/2018/2018-1171/en_tema_1171.pdf?language_id=1)
- McLaughlin, R. & Young, C. (2018). Data democratization and spatial heterogeneity in the housing market. En C. Herbert, J. Spader, J. Molinsky & S. Rieger, *A Shared Future: Fostering Communities of Inclusion in an Era of Inequality* (pp. 126-139). Harvard Joint Center for Housing Studies. [https://www.jchs.harvard.edu/sites/default/files/A\\_Shared\\_Future\\_Chapter\\_6\\_Data\\_Democratization.pdf](https://www.jchs.harvard.edu/sites/default/files/A_Shared_Future_Chapter_6_Data_Democratization.pdf)
- Merlo, A. & Ortalo-Magne, F. (2004). Bargaining over residential real estate: evidence from England. *Journal of Urban Economics*, 56(2), 192-216. <https://doi.org/10.1016/j.jue.2004.05.004>
- Merlo, A., Ortalo-Magné, F. & Rust, J. (2015). The home selling problem: Theory and evidence. *International Economic Review*, 56(2), 457-484. <https://doi.org/10.1111/iere.12111>
- Mikhed, V. & Zemčík, P. (2009). Do house prices reflect fundamentals? Aggregate and panel data evidence. *Journal of Housing Economics*, 18(2), 140-149. <https://doi.org/10.1016/j.jhe.2009.03.001>
- Ministerio de Transportes, Movilidad y Agenda Urbana (MITMA), España. (2020). *Sistema Estatal de Índices de Referencia del Precio del Alquiler de Vivienda*. <http://www.fomento.gob.es/be2/?nivel=2&orden=34000000>
- Montalvo, J. G. (2011). *De la quimera inmobiliaria al colapso financiero*. Antoni Bosch Editor.

- Páez, A. (2021). Open spatial sciences: an introduction. *Journal of Geographical Systems*, 23, 467-476. <https://doi.org/10.1007/s10109-021-00364-4>
- Páez, A., Long, F. & Farber, S. (2008). Moving window approaches for hedonic price estimation: an empirical comparison of modelling techniques. *Urban Studies*, 45(8), 1565-1581. <http://dx.doi.org/10.1177/0042098008091491>
- Pangallo, M. & Loberto, M. (2018). Home is where the ad is: online interest proxies housing demand. *EPJ Data science*, 7(1), 47. <https://doi.org/10.1140/epjds/s13688-018-0176-2>
- Rey, D. (2023). Online asking and actual house price set - Region of Madrid. *Mendeley Data*, V1. <http://dx.doi.org/10.17632/r53fjrcwt3.1>
- Rey-Blanco, D., Arbués, P., López, F. A. & Páez, A. (2023). Using machine learning to identify spatial market segments. A reproducible study of major Spanish markets. *Environment and Planning B: Urban Analytics and City Science*, 0(0). <https://doi.org/10.1177/23998083231166952>
- Rico, J. R. & Taltavull, P. (2021). Machine learning with explainability or spatial hedonics tools? An analysis of the asking prices in the housing market in Alicante, Spain. *Expert Systems with Applications*, 171, 114590. <https://doi.org/10.1016/j.eswa.2021.114590>
- Rosen, S. (1974). Hedonic prices and implicit markets: product differentiation in pure competition. *Journal of Political Economy*, 82(1), 34-55. <https://doi.org/10.1086/260169>
- Rosenfeld, D. (2022). Using real-time indicators for economic decision-making in government. Lessons from the Covid-19 crisis in the UK. ODI Emerging analysis. [https://cdn.odi.org/media/documents/ODI\\_Emerging\\_analysis\\_Using\\_real-time\\_indicators\\_for\\_economic\\_decision-making\\_Q5O4ha9.pdf](https://cdn.odi.org/media/documents/ODI_Emerging_analysis_Using_real-time_indicators_for_economic_decision-making_Q5O4ha9.pdf)
- Schachter, A. & Besbris, M. (2017). Immigration and neighborhood change: Methodological possibilities for future research. *City & Community*, 16(3), 244-251. <https://doi.org/10.1111/cico.12242>
- Shimizu, C., Nishimura, K. & Watanabe, T. (2016). House prices at different stages of the buying/selling process. *Regional Science and Urban Economics*, 59, 37-53. <https://doi.org/10.1016/j.regsciurbeco.2016.04.001>
- Similarweb. (2022). *Cifras de tráfico web comparativo para España de los portales idealista.com, fotocasa.com y pisos.com*. <https://www.similarweb.com> (septiembre).
- Struijs, P. & Daas, P. (2014). Quality approaches to big data in official statistics. Texto presentado en la European Conference on Quality in Official Statistics (Q2014), Vienna, Austria. [http://www.pietdaas.nl/beta/pubs/pubs/Q2014\\_session\\_33\\_paper.pdf](http://www.pietdaas.nl/beta/pubs/pubs/Q2014_session_33_paper.pdf)
- Sun, D., Du, Y., Xu, W., Zuo, M. Y., Zhang, C. & Zhou, J. (2015). Combining online news articles and web search to predict the fluctuation of real estate market in big data context. *Pacific Asia Journal of the Association for Information Systems*, 6(4), 2. <https://doi.org/10.17705/1pais.06403>
- Trivelli, P. (2012). Análisis trimestral 30 años de mercado de suelo en el Gran Santiago. *Mercado de Suelo Urbano*, Boletín n° 121 (octubre). [https://www.researchgate.net/publication/347463738\\_ANALISIS\\_TRIMESTRAL\\_30\\_anos\\_boletin\\_numero\\_121\\_de\\_Mercado\\_de\\_suelo\\_Urbano](https://www.researchgate.net/publication/347463738_ANALISIS_TRIMESTRAL_30_anos_boletin_numero_121_de_Mercado_de_suelo_Urbano)

- United Nations Economic Commission for Europe (UNECE). (2015). *Using administrative and secondary sources for official statistics: A handbook of principles and practices*. [https://unstats.un.org/unsd/nationalaccount/workshops/2014/St\\_Lucia/R-SD9.pdf](https://unstats.un.org/unsd/nationalaccount/workshops/2014/St_Lucia/R-SD9.pdf)
- Wang, X., Li, K. & Wu, J. (2020). House price index based on online listing information: the case of China. *Journal of Housing Economics*, 50, 101715. <https://doi.org/10.1016/j.jhe.2020.101715>
- Wu, Y., Wei, Y. D. & Li, H. (2020). Analyzing spatial heterogeneity of housing prices using large datasets. *Applied Spatial Analysis and Policy*, 13(1), 223-256. <https://doi.org/10.1007/s12061-019-09301-x>
- Xu, X. & Zhang, Y. (2022). Cointegration between housing prices: evidence from one hundred Chinese cities. *Journal of Property Research*, 40(1), 53-75. <https://doi.org/10.1080/09599916.2022.2114926>