

# Conclusiones

*“Las reglas de la moralidad no son la conclusión de nuestra razón.”*

— David Hume

La presente investigación ha logrado desarrollar satisfactoriamente un índice de precios de la vivienda en alquiler con los objetivos pretendidos: alto nivel de detalle temporal y geográfico, fiel reflejo del mercado mediante una alta consistencia con las fuentes oficiales y el uso de fuentes abiertas con métodos de aprendizaje estadístico.

Durante el transcurso de la misma, y para la consecución del objetivo, se han abordado de manera sistemática un conjunto de análisis que abarcan desde tener una perspectiva general del sector, hasta la definición concreta de los diferentes procesos metodológicos. A lo largo del proceso se han obtenido importantes conocimientos.

**El mercado español de la vivienda presenta un comportamiento diferencial al resto de los países de la Unión Europea. En nuestro país, existe una proporción mayoritaria de hogares en propiedad desde finales del siglo pasado. Sin embargo, tras la última gran crisis inmobiliaria se ha comenzado a revertir esta situación, motivada, no solamente, por un cambio actitudinal en las preferencias entre comprar o alquilar, sino también por la restricción crediticia de los primeros años de la década de 2010, la cual ha abocado a los grupos socioeconómicos menos favorecidos a tomar el alquiler como única alternativa. En paralelo, la presión de la demanda ante una oferta apenas creciente, en las áreas más densamente pobladas, ha dado lugar a un incremento progresivo de los precios del alquiler.**

1. Un factor importante en la escasez de oferta de vivienda fue consecuencia de la limitada construcción de vivienda nueva durante casi una década. Esto tiene su justificación en la explosión de la burbuja inmobiliaria, unida a un crecimiento de la población de más de dos millones de personas

- (principalmente en las áreas metropolitanas). La restricción crediticia hipotecaria desde finales de los 2000 hasta mediados de la década del 2010, impulsó de forma notable la demanda de vivienda en alquiler. Ambos factores han forzado un nuevo equilibrio entre oferta y demanda, que ha producido un incremento sostenido de los precios desde el año 2015.
2. España es un mercado con una oferta de alquiler generalmente inelástica, particularmente en las grandes áreas metropolitanas. Dichas zonas, cuya demanda ha sido creciente en la última década, experimentaron un notable incremento de precios, alimentado por una oferta estática o decreciente, producto de la escasez de vivienda nueva y social. Lo cual ha producido grandes diferencias en los mercados en función de la geografía, siendo las capitales más pobladas las que han sufrido mayores subidas de precios.
  3. El fuerte incremento de los precios afectó principalmente a los colectivos menos favorecidos, como los jóvenes o familias con menos recursos, por su casi imposibilidad de acceso al crédito. Lo cual es especialmente problemático en las grandes áreas metropolitanas del país, donde se concentra una gran demanda de vivienda junto con una oferta limitada. Dicha situación han contribuido a la degradación de las condiciones de vida y expulsión hacia zonas más económicas de los colectivos anteriores.
  4. La escasez de vivienda social está vinculada a una regulación del mercado de alquiler en España que no han favorecido la construcción pública de este tipo de vivienda. Además, anteriormente y durante un periodo prolongado (previo al estallido de la última burbuja) se había favorecido el régimen en propiedad. Los cambios de los últimos 10 años se han centrado en actuar sobre otros aspectos, como la implantación de incentivos fiscales para arrendadores y arrendatarios, las políticas de limitación de precios en ciertas zonas, y el control de la actualización anual de las rentas a un máximo por área. Aunque estas medidas animan a los propietarios privados a poner en el mercado vivienda social, esta es insuficiente en número e incentivos.
  5. Desde el punto de vista del fomento de la oferta, la normativa española ha asignado un rol principal al mercado privado como principal proveedor de vivienda en alquiler, tanto como promotor de vivienda social nueva como arrendador de los colectivos menos favorecidos. Con el fin de aliviar la carga sobre los propietarios individuales y fomentar la oferta, se otorgan incentivos fiscales a los propietarios y desincentivos o penalizaciones para los propietarios de viviendas desocupadas. Aún así, esta medida se ha mostrado insuficiente en impacto real y debería acompañarse una importante inversión en promoción de vivienda pública social.
  6. El aspecto anterior es, en parte, consecuencia de una normativa que

favorece los alquileres plurianuales y de la inestabilidad regulatoria (cambios frecuentes de las políticas de vivienda en los ámbitos nacionales y regionales). La cual produce, consecuentemente, que un porcentaje de los propietarios decidan dejar este mercado, realimentando la escasez en la oferta. Por tanto, la normativa debería proteger la situación legal de los mismos, para evitar este efecto pernicioso.

7. Para evitar el fuerte incremento de las rentas en los núcleos urbanos más densamente poblados de los últimos años (Madrid, Barcelona, Málaga, por ejemplo), se han articulado varias acciones orientadas a controlar el precio máximo en las zonas cuyos precios están más tensionados (la relación precio medio del alquiler / ingresos es muy alta). A la vista de los resultados, estas actuaciones no han tenido el efecto esperado porque, entre otras cuestiones, desincentivan a parte de los propietarios para mantener sus inmuebles en mercado, acrecentando la tensión sobre la oferta.
8. La principal crítica a las políticas restrictivas de precios, denominadas de segunda generación, es que introducen distorsiones entre las relaciones de los mercados de compra y alquiler, ignora aspectos fundamentales en la formación de los precios y, en el largo plazo, desincentiva las inversiones del propietario en mantener correctamente la propiedad. Por otra parte, tanto en el mercado nacional, como en los países del entorno, ninguna de las normativas de este tipo ofrecen evidencias claras sobre su eficacia.
9. Dentro de la administración no existe consenso respecto a quién debe liderar el desarrollo de la política de vivienda, planteando la duda de si estas deben ser nacionales, o si las comunidades autónomas o los ayuntamientos deberían ser los últimos responsables de decidir las normas a implementar. Parece evidente que las políticas deben articularse desde las entidades locales dentro de un marco general de política de la vivienda estatal o regional. De forma que se ataje el problema secular de la ausencia de vivienda social, a través de actuar sobre la oferta mediante un plan ambicioso de desarrollo de vivienda de esta naturaleza.
10. En paralelo a los puntos anteriores, en términos microeconómicos, la asimetría de información entre los oferentes y los demandantes de vivienda conduce a una escasez artificial, producto de la falta de confianza en el arrendatario, que se traduce en una prima de riesgo ante la posibilidad de impagos. Por lo cual el desarrollo de herramientas que ofrezcan más solvencia del inquilino y la creación de financieros que ofrezcan garantías ante los propietarios, para las personas más desfavorecidas, ayudarían a reducir las primas de riesgo en las cuotas.

**La contabilidad nacional es la referencia estadística y se acepta**

**comúnmente como fuente de las principales actividades de la economía. Sin embargo, presenta una serie de inconvenientes. En primer lugar, ofrece información muy agregada y con un retraso temporal importante para el sector inmobiliario. Por otra parte, el uso generalizado de Internet y aplicaciones móviles para la búsqueda de la vivienda ha permitido que dichas plataformas tecnológicas, sean una fuente de información relevante en el estudio de los fenómenos inmobiliarios. Sin embargo, el uso de este tipo de datos en la construcción de índices de precios de la vivienda plantea retos metodológicos producto de sus sesgos y la desigual calidad de la información.**

11. Al contrario de lo que sucede en el sector inmobiliario de compraventa a nivel nacional, la administración no dispone de herramientas de información sobre los precios del alquiler adecuados para el estudio exhaustivo de las dinámicas del mercado. El último intento de la administración central al respecto fue la creación, en 2020, de un índice de precios del alquiler, dirigido por el Ministerio de Transportes y elaborado en colaboración con otros organismos públicos. A pesar de este esfuerzo, dicho índice no satisfacía completamente todos los requisitos deseables, como el de contar con datos recientes, detallados y desglosados geográficamente.
12. Aunque existen fuentes de información privadas con un alto grado de actualización y desglose (portales inmobiliarios), estas no contienen precios finales de rentas, sino datos de oferta. Además, solo recopilan el producto anunciado en sus plataformas, lo que impide obtener una visión íntegra del mercado, que analizadas de forma aislada conforman una imagen sesgada de los mercados inmobiliarios.
13. Al contrario que en el uso de las fuentes estadísticas tradicionales, los datos de portales inmobiliarios adolecen de problemas estructurales de calidad de la información (duplicados, valores extremos, por ejemplo). Además, la diversidad de fuentes, magnitudes y métodos de elaboración exigen la homogenización y consolidación de las mismas.
14. Los datos registrados en los portales de Internet representan una perspectiva sesgada de la oferta y provienen de registros introducidos por usuarios. Por tanto, es necesaria la aplicación de procesos de tratamiento previo de la información tales como: eliminar valores atípicos, corrección errores, imputación de valores ausentes y eliminación de registros duplicados.
15. Se requiere información muy detallada en todas las fuentes utilizadas, particularmente en lo relativo al aspecto geográfico, ya que la heterogeneidad espacial asociada al precio no siempre puede explicarse mediante los atributos de la vivienda o tipologías de zona.

16. Los cambios metodológicos en las fuentes oficiales a lo largo del tiempo influyen de manera significativa en la estabilidad temporal de las medidas. En el presente caso, ha sido necesario aplicar un proceso de suavizado exponencial en el dato de la EPF para solventar este problema.
17. El empleo de múltiples fuentes con grandes volúmenes de datos (decenas o cientos de millones de registros) requiere la aplicación tecnología de procesos de datos eficiente y a gran escala, a través de plataformas tecnológicas en “la nube”.

**Los métodos de correspondencia estadística (*statistical matching*) son eficaces para replicar la estructura poblacional original de la muestra a través de los elevadores muestrales y mitigar el efecto de los sesgos de no respuesta (indisponibilidad de información para uno o varios estratos de la muestra), de composición, variables omitidas, de infrarrepresentación y sobrerrepresentación, habituales en los datos procedentes de portales inmobiliarios de Internet. La correspondencia estadística se aplica tanto en el muestreo como en la función que traduce los precios de oferta en precios de mercado.**

18. Los sesgos muestrales comunes en los datos de oferta (obtenidos de Idealista) incluyen el sesgo de no respuesta, el sesgo de composición, así como el sesgo de infra y sobrerrepresentación. Todos estos sesgos se deben a factores exógenos al mercado y distorsionan la percepción sobre el mismo. Algunos de ellos proceden, por ejemplo, de la cuota de mercado del portal, una mayor orientación hacia las agencias en detrimento de los particulares, o al nivel de adopción del servicio de los agentes inmobiliarios en una zona.
19. El proceso de calibración de poblaciones permite ajustar de forma apropiada los pesos poblacionales y enlazar fuentes diversas en una estratificación compleja, sin embargo requiere de variables auxiliares que vinculen los conjuntos de datos a conectar estadísticamente.
20. Cuando no se dispone alguna de las características en los dos conjuntos es posible estimarla mediante modelos. En nuestro caso, el modelo utilizado ha dependido de la naturaleza de la variable: año de construcción (a través de un proceso geográfico de imputación del año de construcción catastral), la superficie útil de la vivienda, o los ingresos familiares (a través de un modelo de tipo *Random Forests*).
21. El procedimiento de calibración en múltiples etapas resulta ventajoso para realizar la correspondencia cuando las variables comunes no están totalmente disponibles en todos los registros, y no es posible imputar la información a través de modelos. Para el presente caso, las magnitudes de

precios de oferta y reales de alquiler se vinculan mediante la combinación de calibración y modelos de regresión hedónica.

22. Sin embargo, la calibración presenta varias limitaciones que afectan a su eficacia, como se ha podido constatar en nuestro caso de aplicación:

- a. La estabilidad de las precios para un segmento heterogéneo, escasamente representado y con pesos irregulares a lo largo del tiempo, como sucede con las viviendas unifamiliares, es generalmente baja. Incluso en el caso en el que el nivel de ajuste en los modelos hedónicos es alto, la dispersión es mucho mayor comparada con estratos más representativos.
- b. Ante viviendas más complejas, como sucede con las viviendas unifamiliares, y considerando un menor tamaño muestral, la irregularidad en los resultados de los índices de precios de viviendas puede deberse a la omisión o la imprecisión de variables explicativas clave. Por ejemplo, la limitación máxima de superficie útil en la EFF a  $m^2$  dificulta la especificación de los modelos de valoración de viviendas con superficies amplias.
- c. La incorporación de zonas exactas resulta esencial para el correcto desempeño de los modelos, dada la presencia de heterogeneidad espacial en los precios. Cuando la información no está disponible en la fuente, como es el caso de la base de datos de la EPF, se requiere un nuevo proceso de reponderación basado en fuentes detalladas, como se realizó con los datos de precio del MITMA.
- d. Para abordar el problema espacial, se plantean diferentes alternativas. La primera consiste en efectuar un ajuste de sesgo posterior utilizando una fuente auxiliar (como en el caso del MITMA), que puede realizarse por medio de ratios o mediante un modelo lineal o de árboles. La segunda opción sería reemplazar la EPF por una encuesta que abarque todas las áreas de estudio y que permita construir el modelo de conversión considerando la ubicación del inmueble.
- e. La limitada granularidad de los microdatos de la EPF introduce una incertidumbre importante en el modelo de correspondencia de precios de oferta y mercado (correspondencia estadística de los precios). Este fenómeno conduce a errores en la estimación de los mismos en algunos estratos, los cuales requieren un ajuste posterior.
- f. Generalizando, aunque la correspondencia estadística aporta una solución que, de forma general, permite replicar la estructura muestral y la relación entre magnitudes de precios de oferta y reales, presenta desafíos en términos de representatividad, precisión y disponibilidad de variables y datos espaciales.

**La construcción de herramientas de supervisión y control de los mercados inmobiliarios plantea retos como consecuencia de la naturaleza altamente heterogénea de la vivienda. Lo cual implica que la construcción de índices de precios de la vivienda sea una disciplina compleja y que involucra numerosos aspectos (geografía, características constructivas o estado de conservación, por ejemplo). Para abordar la cuestión, la metodología se plantea en dos planos: el de los precios, donde se estudian los factores que los determinan; y el temporal, que se estudia la evolución temporal de los valores.**

23. En décadas recientes, los organismos encargados de establecer mecanismos macroprudenciales han exigido a las agencias estadísticas nacionales un control sobre los índices de la vivienda, lo cual ha derivado en la homogeneización de estas técnicas estadísticas. Aún así, no existe un consenso metodológico respecto a las variables a considerar para construir los modelos, ya que las realidades constructivas y urbanísticas de cada región determinan la idoneidad de las características a emplear.
24. Entre de las técnicas más idóneas y, por tanto, recomendadas por las agencias estadísticas, se encuentran aquellas basadas en modelos hedónicos, y que permiten la descomposición de los factores fundamentales del precio. Sin embargo, dada su complejidad de cálculo, un gran número de los índices de precios desarrollados por las agencias estadísticas estatales emplean modelos de ajuste mixto de medianas.
25. Los modelos hedónicos paramétricos basados en regresiones lineales poseen la ventaja de ser altamente interpretables, pero conllevan numerosas dificultades en su especificación funcional. Como alternativa a los métodos lineales, recientemente se han popularizado métodos hedónicos que emplean aprendizaje estadístico que resuelven, en gran medida, las debilidades de los primeros, pero a costa de una menor interpretabilidad en su funcionamiento.
26. Los métodos de aprendizaje estadístico estiman los coeficientes de cada una de las covariables del modelo, lo que conduce a que los índices de precios hedónicos, contruidos sobre los modelos, sean necesariamente los denominados de doble imputación. Esta aproximación aporta la ventaja de ser más eficaz ante el sesgo de omisión de variables, que es frecuente en este los modelos de valoración de la vivienda, al no disponer de toda la información del bien inmobiliario (principalmente por la complejidad y naturaleza única del bien).

**El modelo hedónico es un mecanismo eficaz que permite descomponer el precio de la vivienda en base a sus determinantes fundamentales. En este**

**caso, dado el gran número de factores involucrado, la descomposición del proceso en distintos aspectos (funcional, geográfico y tipología) y su consolidación en un modelo ensamblado, ofrece un buen balance entre complejidad del modelo y precisión. El uso de Random Forests ensamblados como método de regresión permite solucionar algunos de los retos de la modelización hedónica tradicional (multicolinealidad, heterocedastidad, no linealidades, entre otros), ofreciendo buenos niveles de reducción de la varianza y sesgo.**

27. No existe un conjunto canónico de variables a utilizar, ya que la idoneidad está marcada por el tipo de vivienda y zona geográfica. No obstante y para evitar este inconveniente, se han tomado las categorías de variables más utilizadas en la academia y la industria, como son: elementos constructivos, morfología del edificio y de la vivienda, equipamiento, acceso a servicios cercanos, características de la zona y aspectos del mercado inmobiliario cercano.
28. *Random Forests* permite tratar una gran cantidad de variables sin incurrir en los problemas de la regresión lineal generalizada, controlando eficazmente los fenómenos de heterocedasticidad, multicolinealidad, las no linealidades y las interacciones complejas entre variables. Además, los algoritmos de árboles ofrecen muy buenos resultados en términos de ajuste con muy bajo sesgo. Dos ventajas adicionales de este método es su baja tendencia al sobreajuste y al hecho de ofrecer un mayor número de estadísticos en el proceso de entrenamiento (a través de las métricas sobre la muestra OOB).
29. Este modelo es menos sensible a las distorsiones introducidas por valores extremos que la regresión ordinaria. No obstante, tiende a excluir los valores menos usuales, y por tanto, la estimación de los mismos es poco fiable. En nuestro caso particular, se observa como el modelo ofrece peores resultados en precios por metro cuadrado anormalmente altos, mientras que no son problemáticos los casos extremos inferiormente.
30. El análisis experimental sugiere que la especialización de los modelos ofrece mejores resultados en escenarios en los que interviene un gran número de variables. La combinación de todos estos modelos en un ensamblaje final permite lograr un ajuste más equilibrado en todos los estratos, como se evidencia por un aumento en el coeficiente de determinación ( $R^2$ ) en todas las zonas.
31. El ensamblado permite separar la capacidad de capturar la varianza desde distintos ángulos (por los modelos individuales) y corregir el sesgo de cada uno de ellos en el proceso de combinación del resultado final. Un número excesivo de modelos a ensamblar también podría resultar problemático



por tanto se decide especializar en dos módulos: atributos de la vivienda y localización. Para lograr la unión eficaz los módulos se comparten atributos zonales (tipo de zona) y morfológicos (superficie y tipo de construcción).

32. Los modelos contruidos mediante ensamblado no están exentos de problemas. Algunos de los inconvenientes encontrados en su aplicación son su complejidad, la generación de un gran número de modelos y su capacidad reducida para ser interpretados y analizados en detalle.
33. Dado que los niveles de precios por metro cuadrado, las características y el nivel de soporte muestral de las viviendas unifamiliares y las plurifamiliares son muy diferentes, el uso de un modelo por cada tipo aporta un mecanismo de control adicional en la estimación del valor de la vivienda.
34. La selección del precio por superficie como variable objetivo reduce los errores de estimación en los valores muy altos. Aunque, desde el punto de vista inmobiliario, puede producir problemas de especificación de los modelos en las viviendas unifamiliares. En dichas propiedades, la superficie de la parcela es un elemento determinante en el valor final, pero no lo es en el precio la vivienda por unidad de superficie construida.

**La vivienda es un bien inusual en tres aspectos: heterogeneidad, durabilidad e inmovilidad. Este último factor señala a la localización como un elemento clave en la construcción de modelos de precios hedónicos. No obstante, los principales retos al introducir la ubicación en este tipo de métodos son: la inexistencia de una forma canónica de especificarla, la complejidad derivada de la heterogeneidad espacial y la dificultad de cálculo de los modelos de regresión geográfica.**

35. No se puede desligar el análisis espacial del modelado hedónico. El adecuado funcionamiento de los modelos hedónicos está condicionado por los efectos en el precio a lo largo del espacio, ya que las relaciones entre los predictores y el precio pueden variar en función de la zona (heterogeneidad espacial). Por tanto, un buen predictor en una zona puede no ser tan efectivo en otra, lo que invalida ciertas condiciones fundamentales de los modelos de regresión si no se incorporan variables que especifiquen correctamente la interacción con la geografía. Asimismo, es importante mencionar que muchos de los determinantes de los precios vinculados a la localización son desconocidos (por ejemplo, cuestiones actitudinales como las preferencias de los inquilinos por ciertas zonas que son muy difíciles de incorporar como atributos de los modelos).
36. Para incorporar de forma eficaz la influencia geográfica en los modelos manteniendo el principio de valoración hedónica, es posible especificarla

sobre el nivel de utilidad asociado a la localización. Esta magnitud se define como el grado de acceso a distintos servicios, medidos en tiempo de desplazamiento. Desde un punto de vista numérico se expresan a través de unos índices de accesibilidad en función de los elementos de posible interés del inquilino y que se encuentran cerca del inmueble.

37. La definición de variables de accesibilidad es compleja desde un punto de vista metodológico y de cálculo. Por este motivo, en muchos casos, la literatura aporta soluciones parciales (uso de distancias euclidianas en lugar de tiempos de transporte) y arbitrarias (factores seleccionados por criterio del investigador sin mayor justificación). Estas aproximaciones parciales pueden no incorporar completamente los factores que influyen sobre el precio.
38. La síntesis de la accesibilidad en pocas variables (existe una gran variedad de aspectos), y que sean significativas para el problema a resolver no es un proceso trivial. Nuestra propuesta de índices gravitatorios basados sobre isodistancias en tiempos de desplazamiento, y seleccionadas de forma automática por su capacidad de reducción de la autocorrelación espacial, demuestra ser un método de especificación general, eficiente y sencillo.
39. El uso de modelos de aprendizaje estadístico de árboles complejos y la accesibilidad permite que los propios modelos incorporen la diversidad de comportamientos a lo largo del espacio. Además, el amplio rango de dimensiones de accesibilidad construidas, permite incorporar la influencia desde un gran número de aspectos. Se demuestra, estadísticamente, que estas variables de accesibilidad reducen la dependencia espacial de los residuos de los modelos y, en el caso de *Random Forests*, prácticamente los eliminan.
40. Dado que las características de utilidad pueden estar correlacionadas entre si es conveniente aplicar un proceso de factorización (por ejemplo, las medidas de utilidad de servicios de restauración y turismo suelen ser bastante similares en los centros de ciudades turísticas).
41. Cuando se desarrollan variables de localización sintéticas (factorizadas), a menudo resulta difícil estimar su grado de contribución. En nuestro caso, el método basado en factorización permite ofrecer una semántica al campo por dos razones: 1) el proceso de construcción selecciona las variables que potencialmente más contribuirían a explicar el precio, y 2) porque el proceso de análisis de componentes principales describe las contribuciones en los aspectos de la variable original y el predictor sintético.
42. A pesar de que el método ofrece un buen rendimiento predictivo, aún quedan ciertas cuestiones pendientes que deben ser abordadas en investigaciones

futuras. Tales como la extensión del método para que sea multivariante, y la integración de otras técnicas de econometría espacial en el proceso de selección heurística de medidas de utilidad.

**El modelo hedónico final, que estima el precio de mercado a través de correspondencia estadística con el precio de oferta, requiere incorporar información zonal a través de fuentes auxiliares. Como la unión se realiza anualmente y el número de modelos a desarrollar es muy amplio, podrían surgir inconsistencias de los modelos a lo largo del tiempo. El método utilizado demuestra funcionar adecuadamente en este último aspecto, ya que los modelos para un año dado predicen con bastante exactitud los precios para el año siguiente.**

43. Al igual que en los modelos de regresión convencionales, aquellos segmentos con menor representación exhiben un mayor grado de error en la predicción, como ocurre con las viviendas unifamiliares.
44. Los modelos ensamblados de árboles de regresión tienen una gran capacidad para capturar la varianza y producir modelos insesgados, pero presentan cierta rigidez estructural en términos de generalización. Lo anterior se deriva de que se asigna a nodo cada hoja del árbol el comportamiento promedio de una serie de observaciones, lo que dificulta la generalización de comportamientos en espacios de datos dispersos, escasos y altamente heterogéneos (como es frecuente en datos inmobiliarios). Alternativamente, se podrían proponer métodos más parsimoniosos que capturen la interacción espacial de forma general, y emplear esta estimación como entrada otro modelo (de tipo: bayesiano multinivel, basado en aprendizaje profundo, GAM, red neuronal o espacio-temporal).
45. A pesar del gran número de fuentes utilizadas, la información disponible es limitada en algunos estratos, particularmente en vivienda unifamiliar; en consecuencia, la exactitud del modelo podría mejorarse con la introducción de atributos específicos que representan sus singularidades, como fotografías del inmueble, textos de anuncios o incluso con imagen satélite de la finca. Alguno de ellos ya están presentes en los portales inmobiliarios, pero requieren un proceso avanzado de la información (visión por computador o minería de textos, por ejemplo).
46. El planteamiento de un modelo hedónico para cada año (impuesto por la restricción de datos en el proceso de correspondencia estadística) no garantiza la continuidad temporal de los valores estimados. Lo cual se refleja en variaciones abruptas de los precios estimados entre el mes de diciembre y el enero del año siguiente (también por la tendencia del árbol

a sobrerrepresentar el comportamiento anual medio, en detrimento de las diferencias mensuales). Para solucionar lo anterior, se podría plantear modelo único para todos los años, que a su vez presentaría los siguientes inconvenientes: 1) al incluir datos de un nuevo periodo, se podrían re-estimar datos de periodos anteriores en base a datos futuros, 2) el volumen de registros es muy elevado, y por tanto, se requerirían elementos más rigurosos para tratar la información.

47. Las fuentes oficiales usadas para la calibración de precios (EPF o censo) no cuentan con datos zonales exactos por tanto la correspondencia estadística en precios ha requerido el uso de una fuente de precios a nivel de zona exacta (datos de precios de alquiler de MITMA). Esto indica que las estadísticas agregadas por criterios funcionales o demográficos (zona urbana media, por ejemplo) combinan realidades de precios muy diferentes y su significancia es relativa.
48. La ausencia del control de zona exacta, descrita en el punto anterior, genera importantes problemas en la consistencia temporal de mercado en las series de precios (cambios de tendencia inesperados en las series), principalmente debidos a efectos de composición.
49. La correspondencia entre los precios vincula de manera muy eficiente los valores de oferta con sus equivalentes de alquiler, a tenor de su alto coeficiente de correlación a lo largo del tiempo (particularmente en vivienda plurifamiliar). Además, se observa una fuerte correlación cruzada entre las magnitudes del mercado y la oferta, con un retraso temporal, con el primero respondiendo a los cambios de la oferta.

**La construcción del aspecto temporal requiere el desarrollo del modelo para todos los periodos de la muestra. Sin embargo, es habitual que los datos disponibles no se encuentren en la frecuencia temporal más desagregada y, en particular, en esta investigación buena parte del dato público disponible es anual. Un proceso automático que estima el método de desagregación para cada zona, basado en la verosimilitud de una serie estructuralmente óptima, resulta eficiente para generar series de alta frecuencia de mayor calidad.**

50. La construcción de un índice de muy alta frecuencia (mensual, semanal o incluso diaria) que tenga en cuenta fuentes públicas requerirá generalmente de un proceso de integración temporal, principalmente porque muchas de las últimas se entregan anualmente o con retraso.
51. Se aprecian dos fenómenos en la integración de múltiples frecuencias de datos: 1) fuertes discontinuidades entre los meses finales e iniciales del

siguiente año; y 2) inconsistencia entre la media de las series mensuales generadas al aplicar los modelos directamente sobre el dato de oferta. Históricamente, estas cuestiones se solventan a través de diversos métodos de desagregación temporal que, por otra parte, introduce la dificultad de decidir qué métodos utilizar.

52. La desagregación temporal en un contexto de gran número de series y alta heterogeneidad entre ellas (por ser mercados inmobiliarios muy distintos), requiere un tratamiento particular de cada serie y es automatizable:
  - a. Existe un vasto número de métodos a aplicar, siendo el más apropiado aquel que depende principalmente de la naturaleza de la serie y de la calidad de las series originales y auxiliares.
  - b. Un enfoque de selección de parámetros y método basado en la calidad de las series generadas, como en nuestro caso con el método propuesto de máxima verosimilitud, ofrece resultados sólidos, pero el resultado también está fuertemente influenciado por la calidad de las series indicadoras utilizadas para estimar la serie final de alta frecuencia.
  - c. El criterio de verosimilitud ofrece una consistencia con el criterio experto, al sustentarse sobre los mismos criterios que toma un especialista para seleccionar un método sobre otro, logrando un mayor nivel de credibilidad del método de decisión.
  - d. Los métodos de selección multicriterio de verosimilitud resultan fácilmente interpretables y pueden mejorarse mediante la adición de nuevos criterios al proceso de selección. Particularmente, en nuestro caso, se plantean tres áreas de mejora futuras: la introducción de nuevos criterios, la desagregación multivariante y el uso de funciones de densidad no paramétricas para la estimación de la verosimilitud.
  - e. Sin embargo, el modelado hedónico determina en parte la calidad del proceso de desagregación, puesto que las series indicadoras se estiman a través de precios hedónicos de la oferta. A este respecto, es clave la forma en la que se especifica el aspecto temporal en los modelos y, particularmente, en los modelos ensamblado con alta dimensionalidad. En la presente investigación, se observa que al haberse incorporado la componente tiempo mediante variables dicotómicas *dummy* por mes, las mismas entran en competencia con otras variables altamente explicativas, como son la zona, la superficie o el estado de conservación, lo cual puede dar lugar una infravaloración de las contribuciones de las variables *dummy* de tiempo.

**Un índice de tipo encadenado superlativo basado en doble imputación hedónica permite mitigar sesgos producidas por variables omitidas.**

**El método de encadenamiento permite realizar ajustes dinámicos para tener en cuenta los cambios en la composición de la muestra a lo largo del tiempo. Este enfoque común en los mercados inmobiliarios, ya que estos están influenciados por múltiples factores de carácter dinámico, tales como los demográficos, financieros, urbanísticos, regulatorios o comportamentales de los agentes.**

53. El índice construido garantiza la consistencia con datos recogidos por fuentes oficiales a través de la calibración y la bondad del ajuste de los modelos hedónicos. Como el nivel de detalle de las fuentes es limitado, sería deseable, además, poder contrastar los resultados desglosados con otra información objetiva (por ejemplo paneles o encuestas de control).
54. El detallado desglose funcional y geográfico facilita un análisis profundo en la evolución de los precios. Como se ha observado, las diversas zonas exhiben distintas velocidades de recuperación en función del nivel de demanda particular. Por otra parte, el análisis conjunto de los índices de precios de oferta y alquiler constituye una excelente herramienta para evaluar dinámicas del mercado, como las tensiones en los precios.
55. La falta de soporte en ciertos estratos introduce ruido en los índices, como sucede en el caso de las viviendas unifamiliares. Este fenómeno aporta inestabilidad a las series desagregadas, cuyo efecto no se puede mitigar completamente mediante el uso de índices de precios encadenados. No obstante, dado el reducido peso de dichos estratos, los problemas no afloran significativamente en los índices más agrupados.
56. Los índices de precios demuestran un comportamiento similar, en general, al IPVA experimental del INE, aunque se detectan ligeras discrepancias en la escala, que se pueden atribuir a las diferencias en los métodos y la estructura poblacional de ambos enfoques.
57. La evolución de los precios de oferta representa un indicador adelantado de la coyuntura de los precios de alquiler. Además, el grado de correlación entre ambos precios es muy elevado, especialmente para los inmuebles de tipo plurifamiliar, lo que confirma la adecuación de esta fuente para construir índices de precios, una vez aplicadas las ponderaciones sobre un diseño muestral que permite corregir los sesgos poblacionales.
58. En el índice se pueden identificar las diferentes velocidades, en las fases de caída y recuperación, vinculadas a los diversos niveles de liquidez del mercado inmobiliario. Por ejemplo, las zonas de Madrid capital y coronas metropolitanas inmediatas son las primeras en caer y recuperarse, mientras que las zonas rurales o de menor densidad poblacional presentan una recuperación más lenta. Esta misma distinción se puede realizar en función

de la excentricidad de dichas caídas, siendo las caídas más pronunciadas en los mercados ilíquidos.

**En el análisis experimental sobre un índice de precios construido para la Comunidad de Madrid en el periodo 2011 - 2019, se observa que las áreas densamente pobladas y urbanas de mayor nivel adquisitivo, y las viviendas con características constructivas y de instalaciones más lujosas experimentaron un crecimiento más pronunciado en los precios de la vivienda. Estos resultados parecen confirmar la hipótesis inicial de que las áreas más pobladas y urbanas de lujo de la Comunidad experimentaron un comportamiento mejor comparado con las áreas rurales y menos exclusivas. Estas diferencias se pueden relacionar con el mayor o menor dinamismo de dichos mercados.**

59. El índice de precios generales muestra una tendencia decreciente en el alquiler desde el año base hasta 2016, seguido de un crecimiento sostenido hasta 2019. Las fases de crecimiento y decrecimiento varían ligeramente por zona, aunque en todas ellas se aprecian una fase de caída y una de recuperación. Estas desigualdades se manifiestan como una caída anticipada en la zona central y más poblada, siendo además la primera en recuperarse. En cambio, las zonas periféricas y rurales tuvieron una recuperación más lenta y valores inferiores en 2019 respecto al año base 2011.
60. El análisis zonal también revela que la oferta de viviendas actúa como un indicador adelantado del precio del mercado, aunque con diferentes escalas de valores, y este diferente nivel de intensidad está relacionado con una mayor liquidez de la zona. En este sentido, la capital, Madrid, es la primera zona en recuperarse y de las últimas en depreciarse debido a la alta demanda que presenta, independientemente del ciclo económico. Por el contrario, las zonas con menor poder adquisitivo tardan más tiempo en recuperarse y además, dado a su menor interés inmobiliario, no muestran caídas ni subidas de precios exageradas.
61. Las diferencias en el comportamiento de los precios entre la capital (Madrid) y el resto de la provincia, están motivadas por la mayor demanda del área urbana y, por otra, porque en las zonas de alto poder adquisitivo el precio se mantiene estable ante cambios en el ciclo económico o inmobiliario, por su mayor elasticidad en los precios ante las condiciones económicas.
62. El análisis basado en características constructivas revela que las viviendas más antiguas no resistieron la crisis tan bien como las viviendas más nuevas, aunque experimentaron crecimientos similares durante la fase de

- recuperación a partir del año 2016.
63. En cuanto a la segmentación por superficie útil, las viviendas de mayor tamaño mostraron un mejor comportamiento en comparación con las viviendas más pequeñas, que experimentaron caídas más pronunciadas durante el periodo 2011-2016.
  64. El análisis por tipo de vivienda y tipo de edificio muestra una mayor estabilidad y crecimiento en precios para viviendas unifamiliares y más exclusivas en el resto de la Comunidad, especialmente en áreas densamente pobladas y urbanizaciones metropolitanas de alta gama.
  65. Las áreas rurales exhibieron los descensos más pronunciados con tasas de crecimiento más bajas, en comparación con las zonas intermedias y densamente pobladas, lo cual puede atribuirse a la mayor presión de la demanda en las zonas urbanas.
  66. La presencia de ciertas instalaciones en la vivienda, como el aparcamiento y la piscina, mostró una diferencia significativa en la evolución de los precios en función de las mismas, especialmente en el caso del resto de la Comunidad. Por contra, otros equipamientos que intuitivamente se relacionan con una mayor calidad del inmueble, como el aire acondicionado, no mostraron diferencias significativas en la evolución de los precios.
  67. Se puede trazar una relación entre la desigualdad zonal y su composición socioeconómica. En municipios de alto poder adquisitivo, por ejemplo la zona noroeste del eje de la A6 (Pozuelo de Alarcón, Majadahoda y Las Rozas), se presenta una disminución inicial en los precios hasta 2013, con una meseta sin crecimiento y una vuelta al crecimiento a partir de 2018. En contraste, en zonas metropolitanas de poder adquisitivo medio y bajo, se percibe una caída progresiva de precios hasta 2015, seguida de una recuperación a partir de 2016. En este contexto, la transferencia de precios de oferta al alquiler es más lenta en comparación con municipios de mayores ingresos. Se percibe también una diferencia en la intensidad de las caídas y subidas según el poder adquisitivo, con transiciones más suaves en las zonas más acomodadas.
  68. Los precios de oferta y alquiler presentan comportamientos homogéneos a lo largo del tiempo en todas las áreas geográfica, aunque con discrepancias en algunas zonas puntuales como la Sierra Norte y Oeste, que experimentan caídas en oferta de menor intensidad que en alquiler.

**El índice de precios construido ofrece una herramienta de análisis en profundidad del mercado inmobiliario de alquiler, por múltiples motivos: 1) alta desagregación temporal, zonal y funcional; 2) incorpora las perspectivas de oferta y mercado; y 3) mantiene la coherencia con las cifras generales del mercado publicadas por los organismos oficiales.**



69. La capacidad de segmentar la información de precios (tanto en valor absoluto como en número índice), permite el análisis en profundidad de los fenómenos inmobiliarios en un mercado complejo como es el de la vivienda desde tres aspectos en paralelo: zonal, funcional y temporal. Este potencial de análisis es esencial en el desarrollo de una política de la vivienda eficaz por parte de las distintas entidades nacionales, autonómicas y locales.
70. Esta metodología es replicable y extensible con datos locales y mucho más granulares, por tanto, se podrían desarrollar nuevos índices con mayor desagregación y con datos específicos para municipios o comarcas.
71. Por construcción, este índice mantiene la coherencia entre cifras generales publicadas por los organismos oficiales (INE o MITMA). Lo que permite su mejor combinación con fuentes de información públicas y fomenta su potencial adopción por las entidades públicas encargadas de la política de la vivienda.
72. El modelo que vincula el precio de oferta y el de mercado, permite analizar en profundidad, entre otras cuestiones, las relaciones funcionales entre oferta y demanda, el poder de negociación de propietarios y arrendatarios. En lo que se refiere al aspecto temporal de esta relación, el índice se ha demostrado su validez como predictor de los precios futuros.
73. El uso de información de portales permite disponer de los resultados del índice de precios actualizado y que sería fácilmente adaptable a un proceso casi en tiempo real (frecuencia diaria). Esto ofrece una perspectiva de control del mercado inmediato inusitada en el ámbito de la vivienda.
74. La complejidad de la presente metodología plantea varios desafíos operativos para su puesta en práctica. El número de modelos, variables, fuentes y estratos involucrados implican un coste de mantenimiento operativo alto, resolubles con la automatización de los procesos de carga, tratamiento de la información y ajuste de modelos.

Las próximas líneas de investigación asociadas a la presente Tesis doctoral se centrarán en tres aspectos: 1) el uso de métodos hedónicos más precisos; 2) la solución de los sesgos de información ausente y zonal; y 3) la mejora de los métodos de tratamiento de series temporales. Más particularmente, atendiendo a los siguientes puntos:

- a. Los recientes avances en el campo del aprendizaje profundo, y la limitada aplicación al campo de la valoración de la vivienda, permite plantear la aplicación de métodos de regresión basados en redes neuronales artificiales de aprendizaje profundo (*deep learning*). Más concretamente, se trabajará sobre las propuestas metodológicas del estado del arte en redes profundas

- de grafos, y las de tipo espacio-temporal.
- b. Analizar el impacto de incorporar o desarrollar nuevas fuentes de datos que complementen aquellos segmentos que tienen menor soporte de información, como las viviendas unifamiliares o los inmuebles singulares.
  - c. Replantear el proceso de reponderación, para que calibración incorpore los aspectos geográficos de los inmuebles con mayor detalle, de tal forma que se evite la aplicación de un modelo final de corrección de sesgos.
  - d. Incorporar procesos de identificación de anomalías en la estructura de las series de precios generadas, de forma que sea posible su corrección automática a través de mecanismos de imputación.
  - e. Plantear mejoras en el proceso de desagregación temporal y conciliación de series temporales con distinta frecuencia. Primeramente, con la inclusión de distintos aspectos zonales, temporales y funcionales en el proceso para asegurar la consistencia multinivel de las series; en segundo lugar, incorporar una aproximación multivariada para lograr resultados más robustos; y finalmente, incorporar funciones de densidad no paramétricas para estimar las funciones de verosimilitud en el algoritmo de selección del mejor método de descomposición.

En resumen, este trabajo contribuye al campo de la economía ofreciendo una serie de novedosos modelos de índices de precios hedónicos de la vivienda en alquiler, cuya innovación es aplicación conjunta de la estadística, econometría tradicional, el aprendizaje automático y las fuentes de datos de portales inmobiliarios.

Se han abordado eficazmente los retos en la incorporación del factor de la localización como determinantes de los precios; la solución de los problemas de coherencia y calidad de los datos inmobiliarios procedentes de portales; la especificación de una función de enlace entre precios de oferta y reales; y finalmente, el diseño temporal y poblacional de los resultados a través de un índice de precios altamente desagregado que guarda consistencia con los datos de la estadística oficial. Se espera que este trabajo sea una referencia para que permita sucesivas investigaciones para el avance del estado del arte en diferentes ámbitos como son la aplicación de técnicas no tradicionales en el modelado hedónico de la vivienda, o la incorporación de fuentes alternativas para complementar o desarrollar herramientas de información para la gestión pública.