

# Tesis Doctoral

**Metodología para la construcción Índices de Precios de la  
Vivienda robustos, segmentados y a escala basados en modelos  
de valoración hedónica**

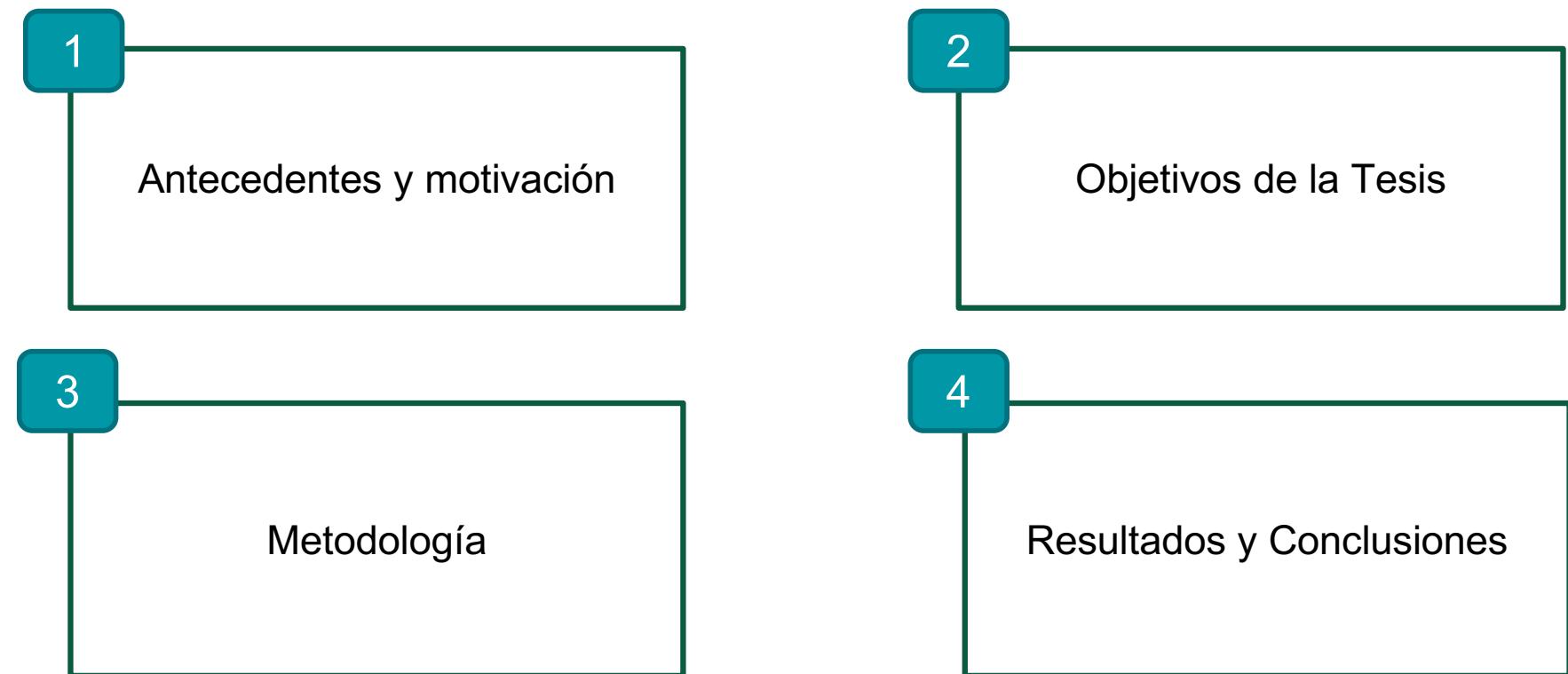
**AUTOR: David Rey Blanco**

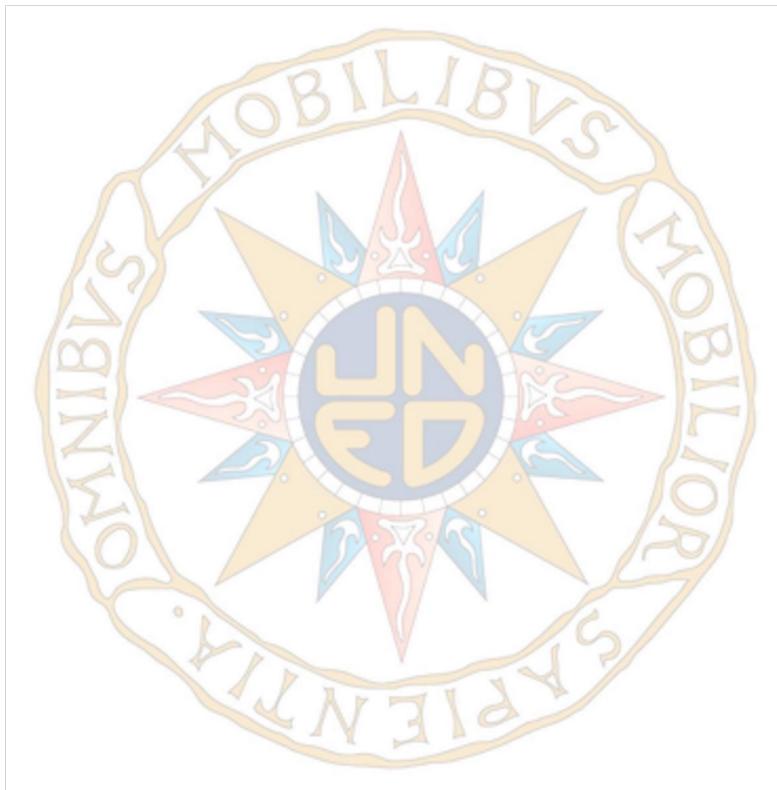
PROGRAMA DE DOCTORADO EN ECONOMÍA Y EMPRESA

DIRECTOR: Dr. Julio González Arias

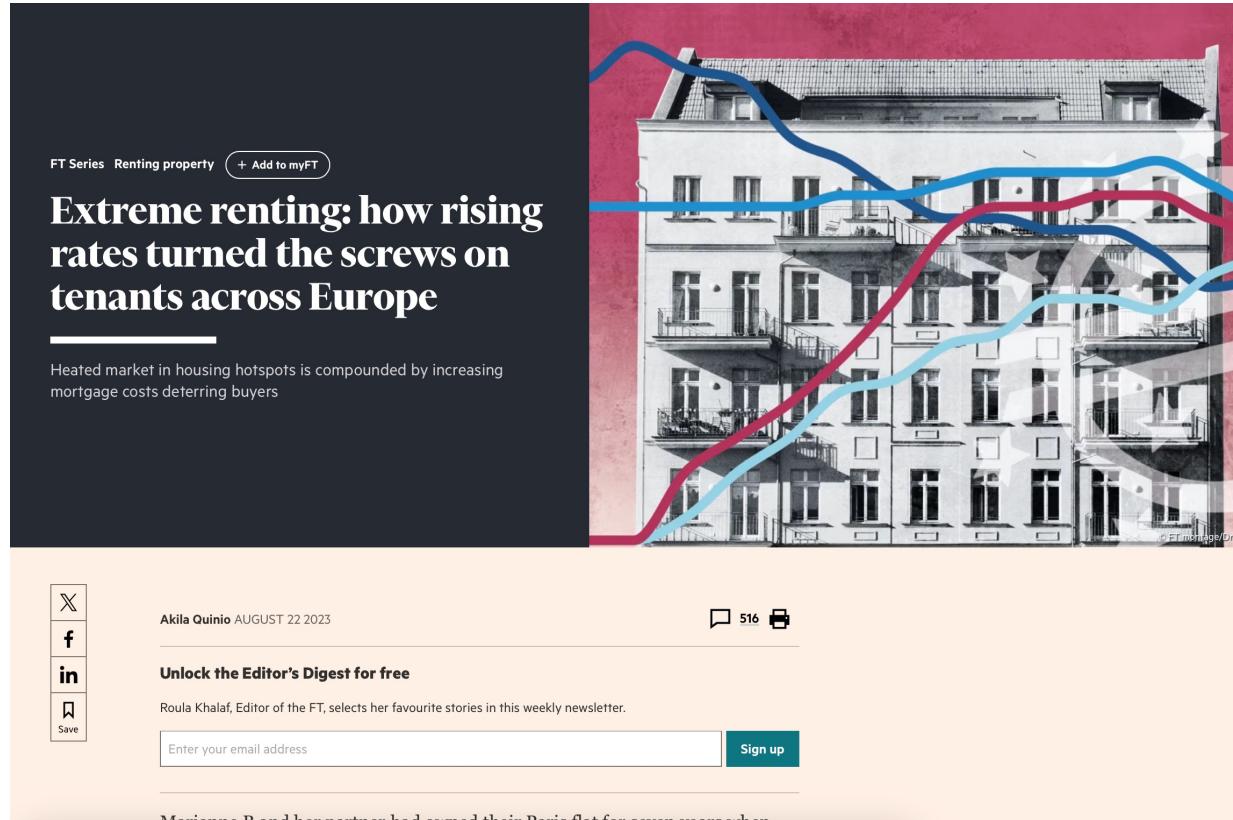


# Agenda





# Introducción



Fuente: Financial Times

## Se ha intentado solucionar con datos alternativos

- La disponibilidad de nuevas fuentes amplia la visión de la estadística oficial (Biancotti, 2020).
- Eurostat (2014, 2022) y los bancos centrales han mostrado su interés en los precios de internet (Loberto et al. 2018; Bricogne et al. 2023; Banco de España, 2021), pero aún no existe ninguna agencia con un índice de este tipo, quizá por la disparidad entre países (Diewert, 2009).
- Existe una alta correlación entre las series de precios de portales y de transacción (Ardila 2016; Chapelle et al., 2022; Kokot 2015; Loberto et at. 2018; Rey-Blanco et al. 2023a).
- Los IPV construidos sobre datos de portales versan el mercado de compraventa y tienen un escaso control de los sesgos (Desormeaux y Piguillem, 2003; Bricogne et al. 2023; Chauvet et al. 2013; Wang y Wu 2020).
- El acceso al microdato de portales ha propiciado nuevos modelos de precios, p.e. los basados en machine learning (Antipov y Pokryshevskaya, 2012; Baldominos, 2018; Alfaro et al. 2020; Rico y Taltavull, 2021).
- La especificación del precio como modelo lineal tiene importantes limitaciones, data la evidencia empírica que muestra la no linealidad de la relación entre precio y características de la vivienda (Des Rories y Thériault, 2006).
- En España, el MITMA (2020) publicó reciente un índice y, basado en él, el INE (2021) el IPVA (experimental), ambos muy agregados y con años de retraso.

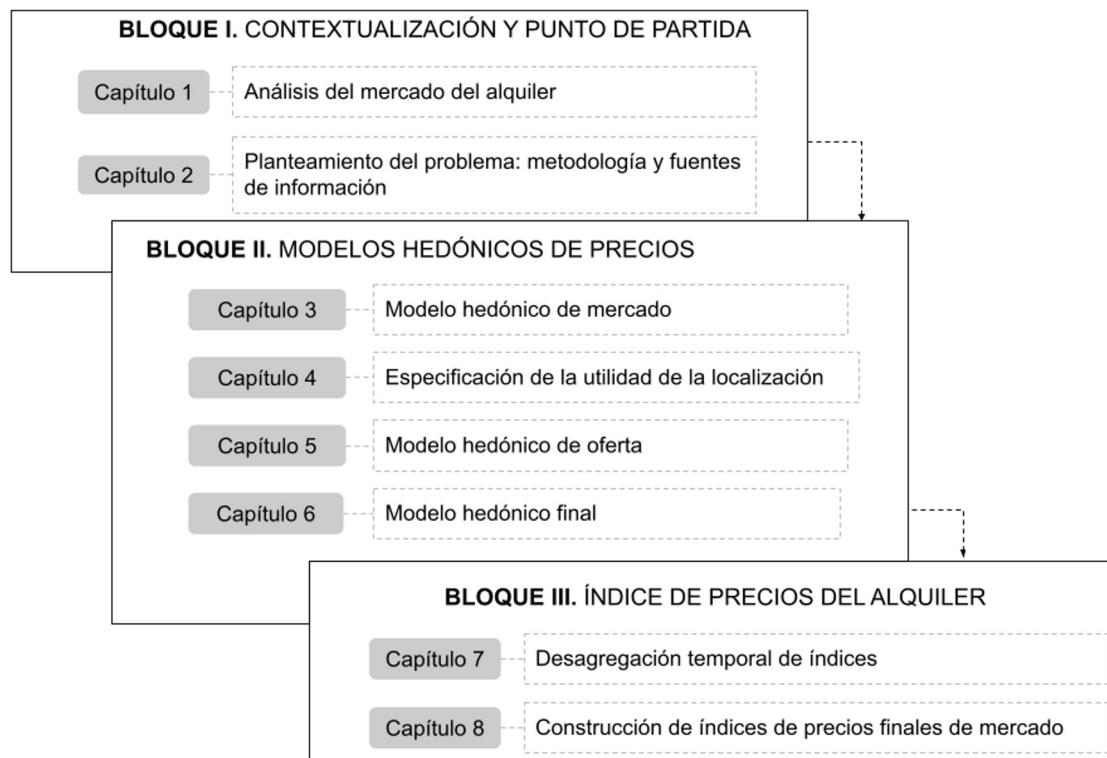
## Nuestra metodología ofrece un avance en el estado del arte

- Presenta una innovación metodológica al ser el primer índice de precios de transacción desarrollado sobre datos de portales inmobiliarios y estadística oficial.
- Incorpora mecanismos para el control de sesgos poblacionales.
- El índice es consistente con las magnitudes recogidas en la estadística pública.
- Ofrece una visión dual de la evolución del precio de mercado (transacción) y el de oferta (Rey-Blanco et al. 2023a).
- Proporciona una fuente de análisis sobre información casi en tiempo real, con un nivel de desglose de la información muy alto.
- Adicionalmente, se ha contribuido en el desarrollo de nuevas técnicas de modelado hedónico, orientadas a la mejora de la precisión y la interpretabilidad de la influencia de la zona (Rey et al., 2023b, Rey et al. 2024).
- Plantea una metodología multidisciplinar que integra técnicas econométricas, estadísticas y de las ciencias de la computación.
- Aparte del aporte académico, aporta una herramienta fácilmente industrializable.

# Objetivos del la metodología del índice

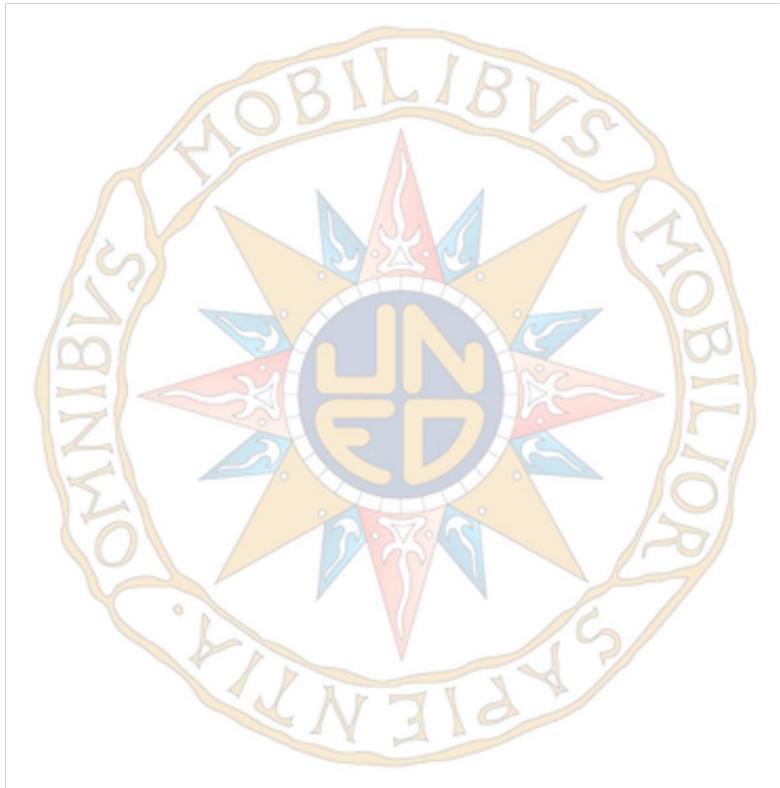
PRINCIPALES	Fiel reflejo del mercado del alquiler	Alto nivel de desagregación funcional	Alto nivel de desagregación temporal	Uso de fuentes y técnicas alternativas
SECUNDARIOS	Alineamiento con estadísticas oficiales	Segmentación geográfica	Frecuencia mensual	Uso de fuentes abiertas
	Definición del colectivo	Segmentación multidimensional características	Recencia del indicador	Uso de métodos de aprendizaje automático
	Ponderaciones poblacionales	Desglose funcional de las contribuciones		

# Estructura de contenidos de la Tesis



Plano de los precios

Plano temporal



# Metodología

## Catálogo de fuentes de información

Anuncios de  
Idealista

**Conjunto de anuncios publicados** en el portal Idealista para la Comunidad de Madrid con frecuencia mensual entre enero de 2011 a diciembre de 2019

Censo

EPF

IPV MITMA

**Estadística pública.** Representan datos observados de mercado y comportamiento de la demanda (microdatos y agregados por sección censal).

Open Street  
Map

Catastro

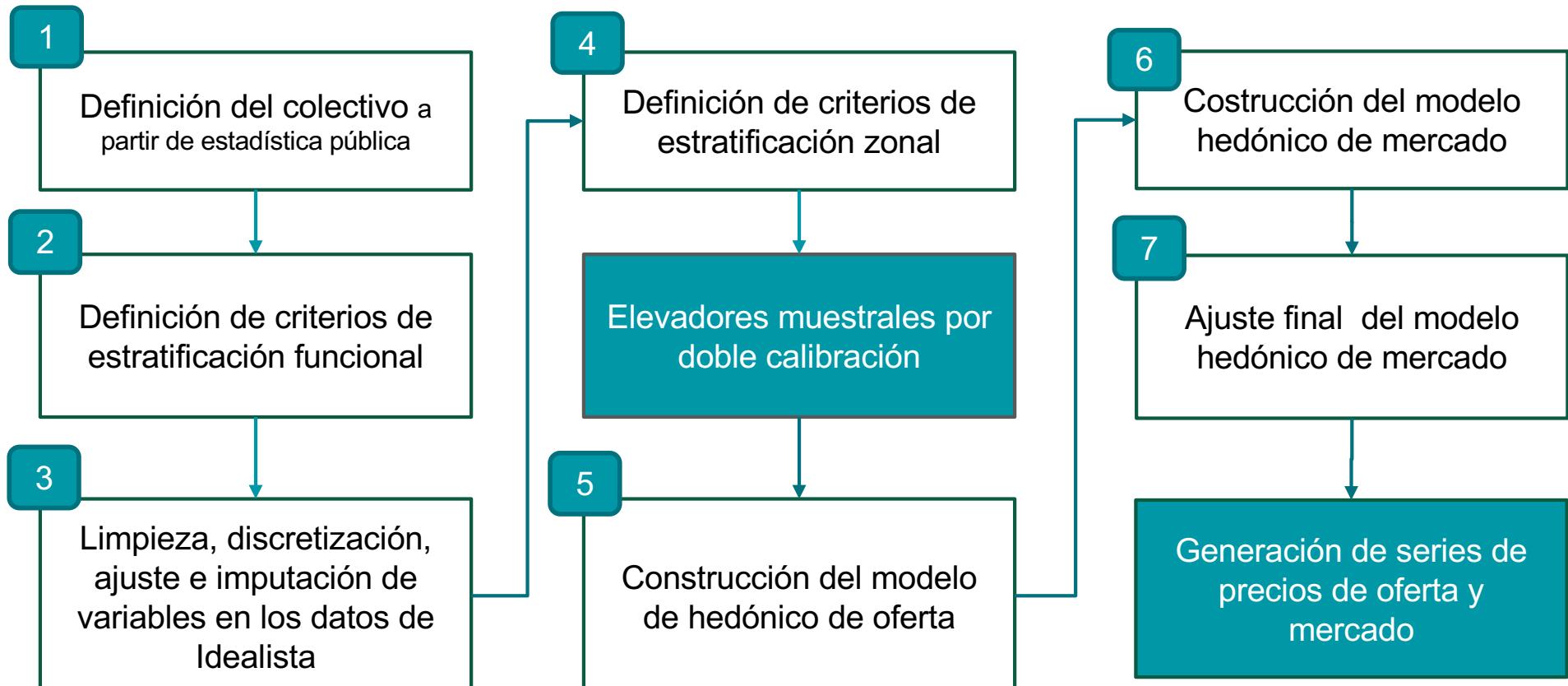
Otros

Otros datos de apoyo

# Metodología: plano de los precios



## Fases elaboración del modelo de precios

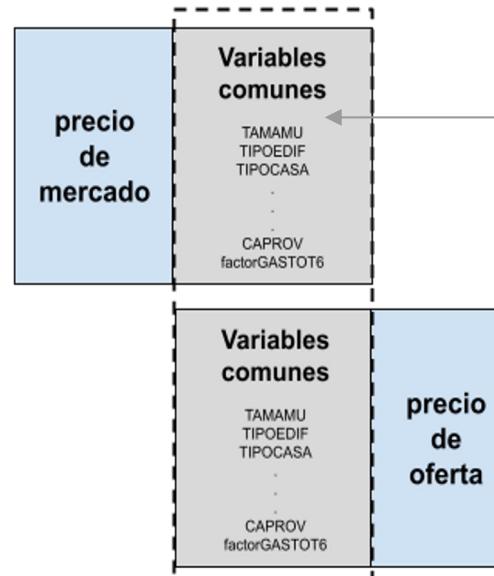


## Correspondencia de magnitudes entre poblaciones

Para llevar la información en el ámbito de la oferta al mercado de alquiler se deben relacionar ambas fuentes. El principio metodológico para hacerlo se basa en modelos de **correspondencia estadística**: calibración (pesos) y regresión hedónica (precios).

### Colectivo destino

Microdatos y datos agregados  
de la estadística oficial



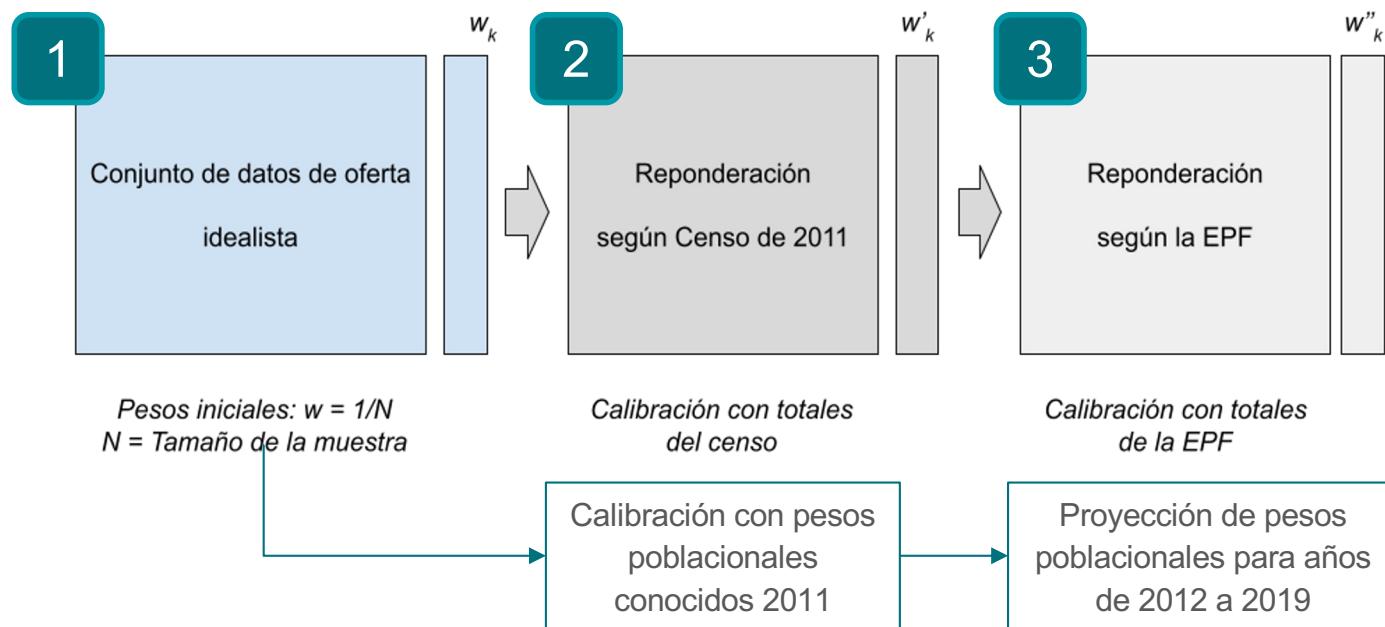
Se usan modelos que relaciona ambos  
ámbitos de información a través de  
variables auxiliares

### Muestra de origen

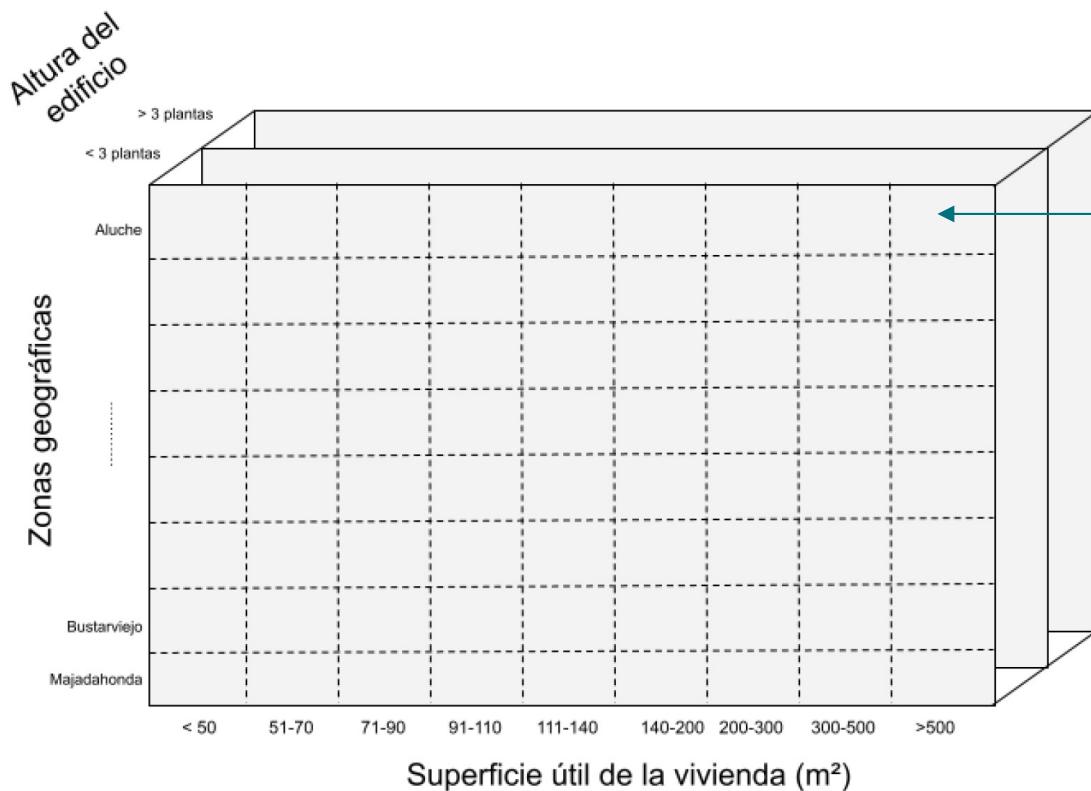
Microdatos y datos agregados  
de la oferta (Idealista)

## Cálculo de elevadores muestrales

Se estiman los pesos de cada observación de la muestra a través de una **doble reponderación mediante calibración** (Deville y Särdnal (1992) sobre la estructura censal proyectada.



## Reponderación mediante calibrado



Las medidas agregadas de cada estrato deben coincidir en mercado y oferta: **número de viviendas y total de metros construidos.**

# Estratificación (1/2): atributos de tipo funcional

**Tabla 3.1.** Resumen de variables de cada rejilla de trabajo

Variable	Descripción	Niveles distintos	Censo	EPF
ANOCONSC	Año construcción	4	X	X
ASCENSOR	Ascensor	2	X	
GARAJE	Garaje	2	X	
LOCATION	Zona	178	X	
NHAB	Número Habitaciones	4	X	X
PLANTAS	Plantas	5	X	X
SUT	Superficie Útil	4	X	X
CAPROV	Código de provincia	1		X
DENSI	Densidad de población	3		X
FACTORGASTOT6	Factor de gasto	3		X
INTERINPSP	Ingresos netos	3		X
TAMAU	Población del municipio	5		X
TIPOCASA	Tipo de vivienda	3		X
TIPOEDIF	Tipo de edificio	4		X
ZONARES	Tipo de zona	3		X

Fuente: elaboración propia

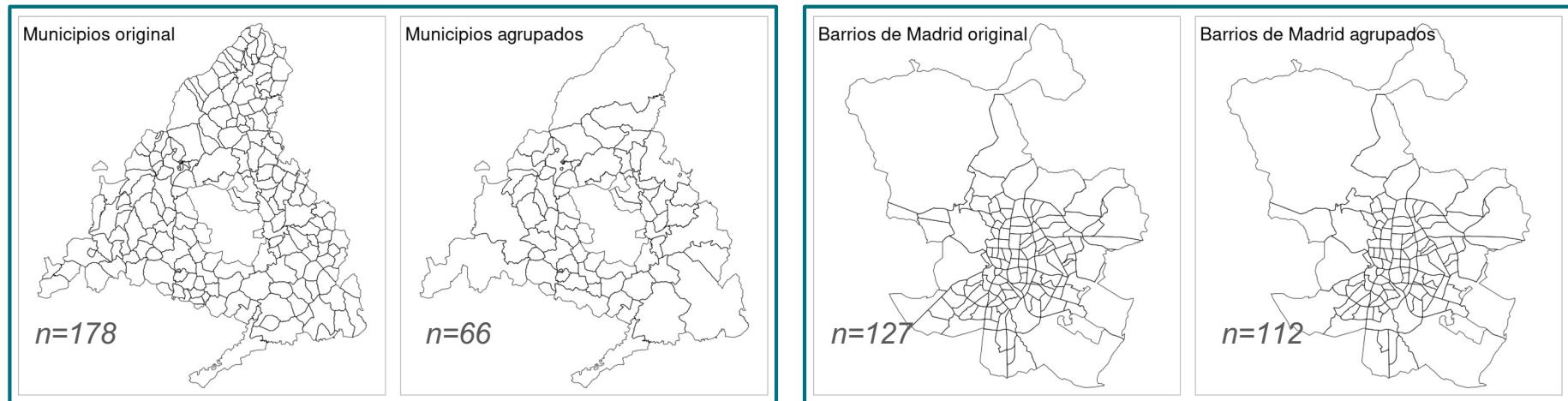
Homogeneización de variables en dos conjuntos

Ha sido necesario de construir modelos para la imputación de variables auxiliares comunes

**Nota:** la selección de las variables más adecuadas tiene más impacto que la técnica utilizada (Adamek, 1994)

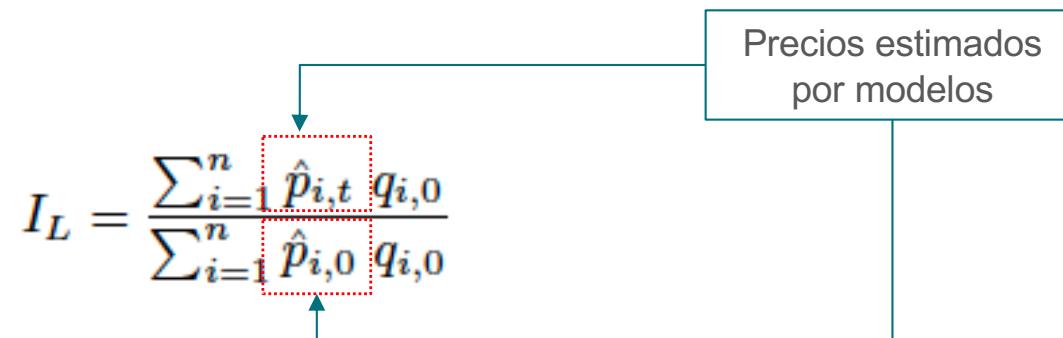
## Estratificación (2/2): atributos de tipo zonal

Se segmenta la geografía en 177 zonas de la Comunidad de Madrid usando un atributo de código de zona. Para la capital se parte de los barrios (128) y para el resto, municipios (66). Se fusionan las áreas cuando no cuentan con el suficiente nivel de soporte en los distintos estratos funcionales (en base a similitud de características inmobiliarias).



## Índice de precios hedónicos

El principio metodológico seguido en nuestro índice será el de **tipo hedónico de imputación doble**, que aporta ventajas ante la omisión de variables (Eurostat 2014; Hill 2013). Esta modalidad usa precios estimados por modelos en lugar del dato original (sometido a sesgos). El **modelado hedónico incorpora la naturaleza heterogénea de la vivienda** controlando los efectos sobre el precio de sus distintas características, es común en muchos IPV europeos (INE, 2016; Eurostat. 2014) .


$$I_L = \frac{\sum_{i=1}^n \hat{p}_{i,t} q_{i,0}}{\sum_{i=1}^n \hat{p}_{i,0} q_{i,0}}$$

**Expresión:** Índice elemental de Laspeyres basado en imputación doble

## Precio de la vivienda como modelo hedónico

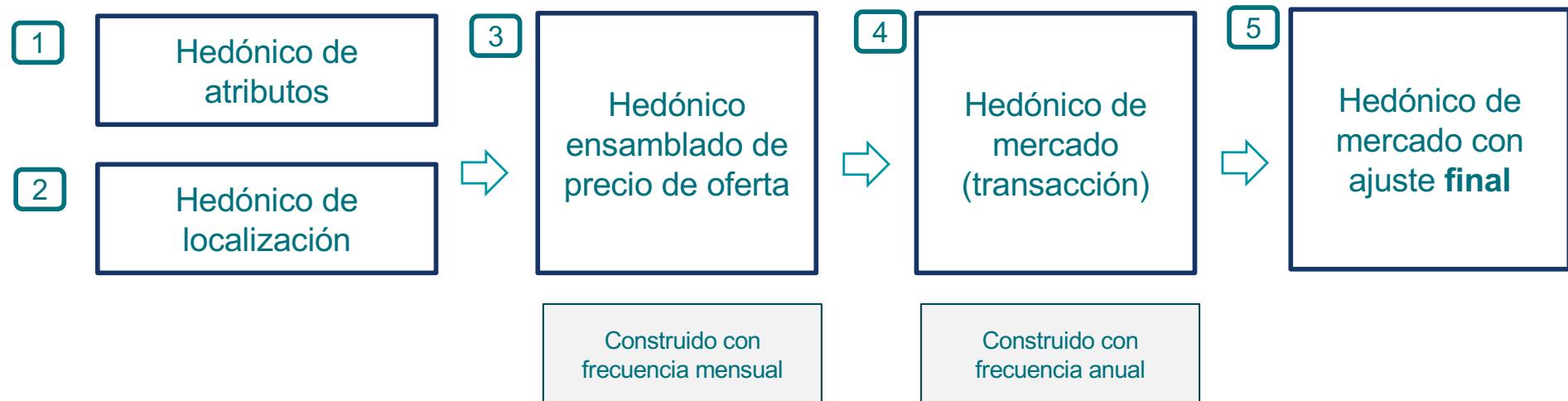
La vivienda es un activo difícil de valorar, al ser inusual en tres aspectos: heterogeneidad, durabilidad e inmovilidad (Kiel y Zabel, 2008). Desde hace décadas el modelo hedónico Rosen (1974) ha sido la herramienta común para estimar la formación de los precios en base a las contribuciones de los factores que describen un bien inmobiliario, y es de uso común en la industria y la academia.

$$\begin{array}{c}
 \text{Precio de la vivienda} \longrightarrow | \quad p^t = \beta_0 + \sum \delta \cdot D_{nk}^t + \sum \beta \cdot S_{nk}^t + \sum \gamma \cdot L_{nk}^t + \sum \mu^t \cdot M_{nk} + \epsilon \quad [1] \\
 \downarrow \qquad \qquad \qquad \downarrow \\
 \text{Tiempo} \qquad \qquad \qquad \text{Localización} \\
 \downarrow \qquad \qquad \qquad \downarrow \\
 \text{Estructura y Constructivas} \qquad \qquad \qquad \text{Dinámicas de mercado}
 \end{array}$$

donde  $D_{nk}$  son variables ficticias dicotómicas que representan el tiempo; atributos de estructura  $S_{nk}^t$ , representado por variables continuas; variables dicotómicas asociadas a la ubicación  $L$ ; y las características de mercado inmobiliario  $M_{nk}$  (Piazzesi et al., 2015).

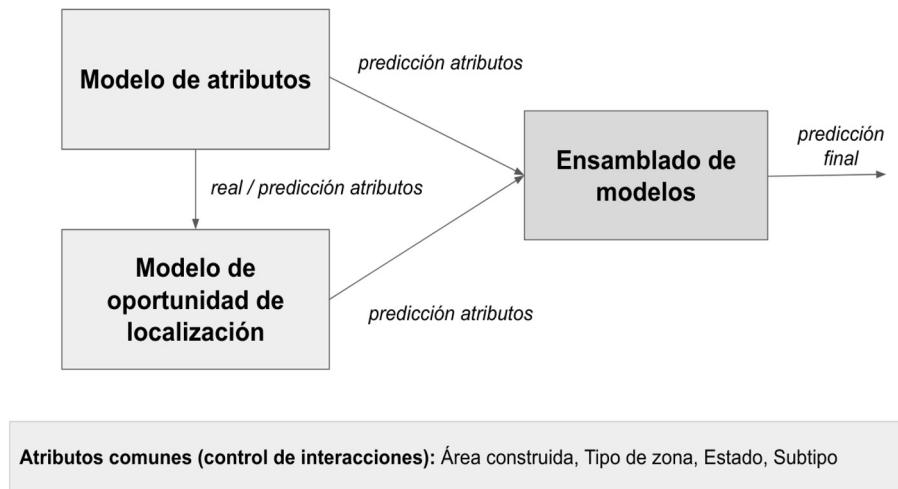
## Modelos de precio hedónicos desarrollados

El modelado hedónico se articula como una **cadena de modelos** que imputa a cada observación de la muestra el precio de mercado. Se estima toda la cadena para **cada año** de la serie. Se usan dos técnicas: Random Forests, por su precisión, bajo sesgo y capacidad de manejar gran cantidad de variables (Alfaro et al., 2020; Antipov y Pokryshevskaya, 2012; Truong et al. 2020) y GAM que aporta un equilibrio entre parsimonia y rendimiento en el ajuste (Hastie y Tibshirani, 2017).



## Hedónico de precios oferta

El alto número de variables puede afectar a la representatividad de las variables (Zhou, 2021). Para evitarlo se desglosa el modelo en los aspectos hedónicos más comunes: características y localización (Hill, 2013), y se fusiona ensamblando las partes. Se logra **una mayor precisión promedio** en cada estrato (en torno una mejora de 5 puntos porcentuales en  $R^2$ ).



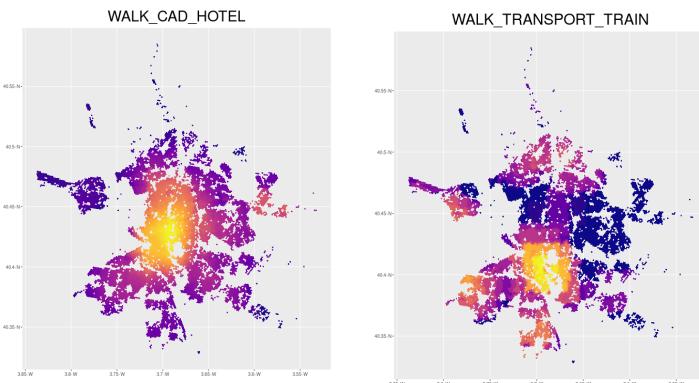
**Tabla 5.18.** Ajuste espacial promedio por clase de modelo y tipo

Capital	Tipo	Ensamblado		Único	
		$R^2$	RMSE	$R^2$	RMSE
No	Plurifamiliar	86,0%	1.29	79,9%	1.57
	Unifamiliar	80,6%	0.89	71,5%	1.22
Sí	Plurifamiliar	89,3%	0.67	87,1%	0.76
	Unifamiliar	91,3%	0.48	88,8%	0.59
Todos	Plurifamiliar	87,2%	1.06	82,5%	1.27
	Unifamiliar	85,9%	0.71	80,1%	0.94

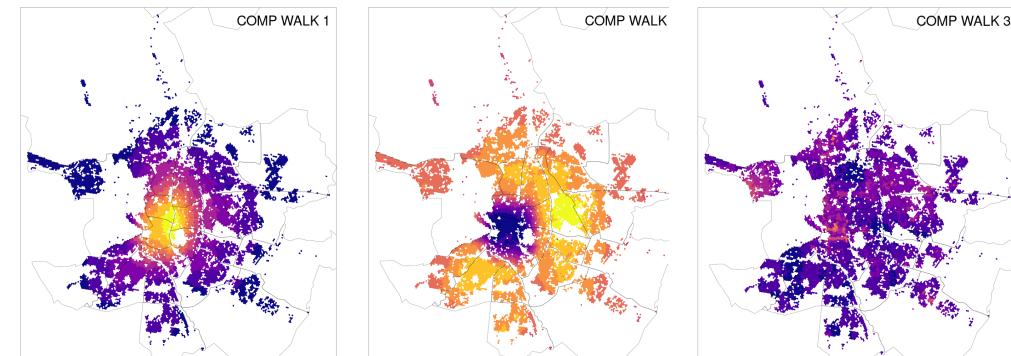
Fuente: elaboración propia

No hay un método canónico para incorar el factor (Hill y Sholz, 2018). Nuestro modelo de localización se basa en índices de accesibilidad que sintetizan las ventajas de una ubicación (**utilidad la ubicación**) en los modelos hedónicos (interpretables y eficientes). Cuatro índices son suficientes para representar adecuadamente la accesibilidad (Rey-Blanco et al., 2023b).

**Figura a:** Índices de accesibilidad originales.



**Figura b:** Índices sintéticos (calculados con PCA)



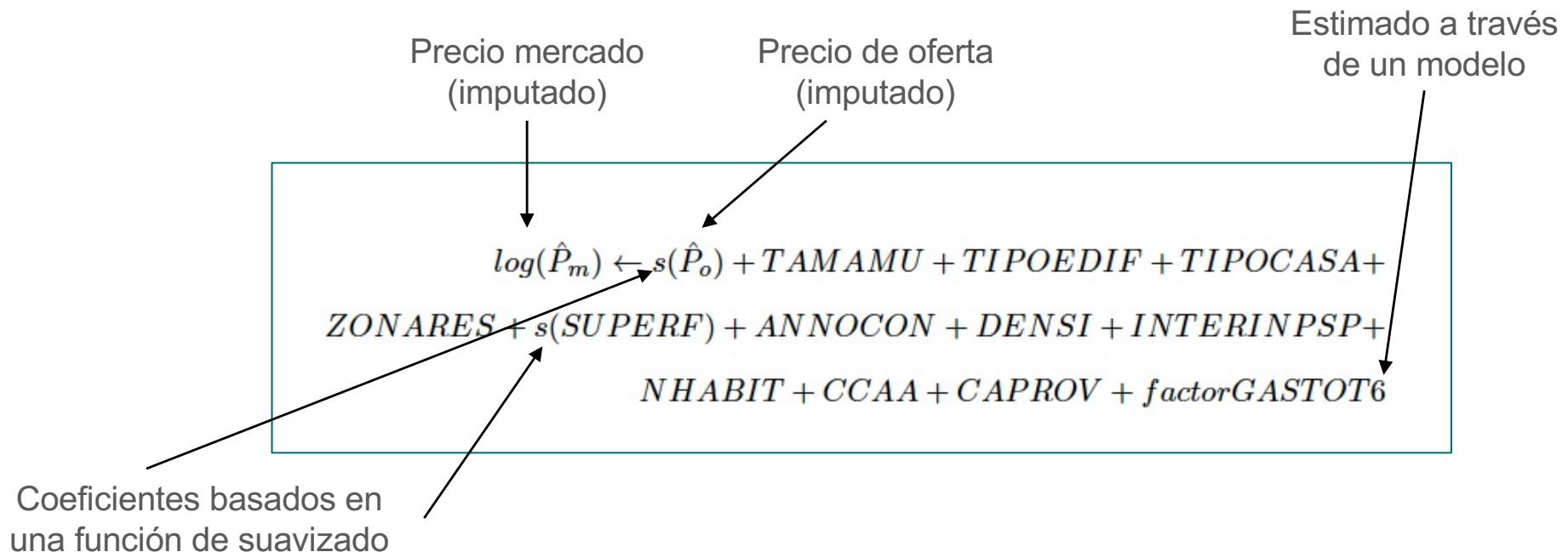
Equivalentes a usar datos geográficos

Su uso aplicado es muy eficiente

Permiten interpretar los efectos de ubicación en el precio

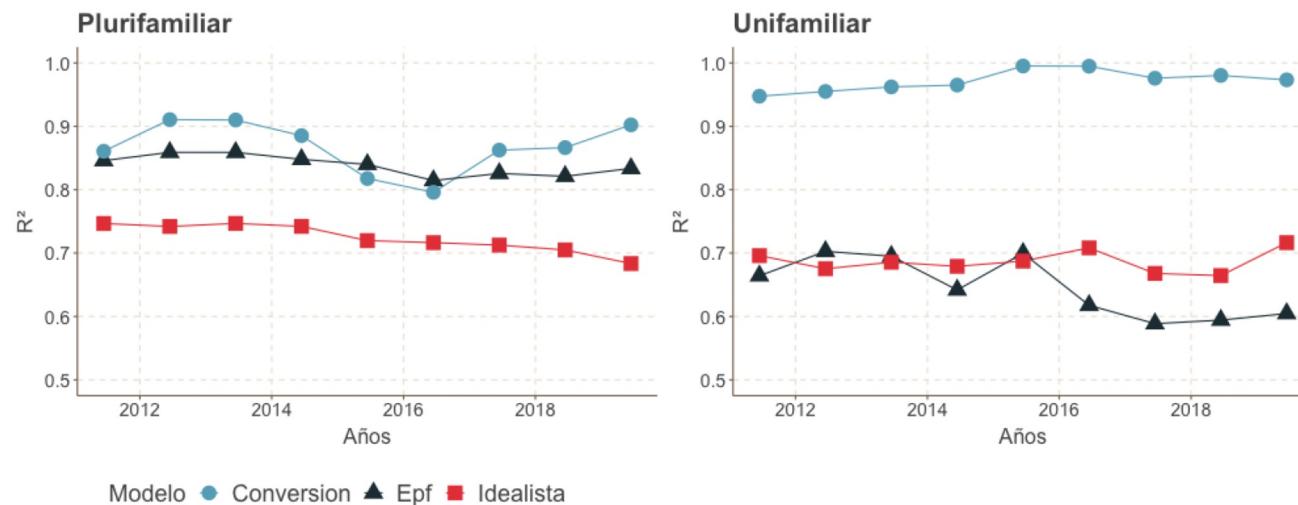
## Hedónico de precio mercado (forma funcional)

Se pretende estimar el precio de mercado (transacción) conocido a través de la EPF, a partir de un modelo hedónico que incorpora el precio de oferta y controla por características de la vivienda y de la zona.



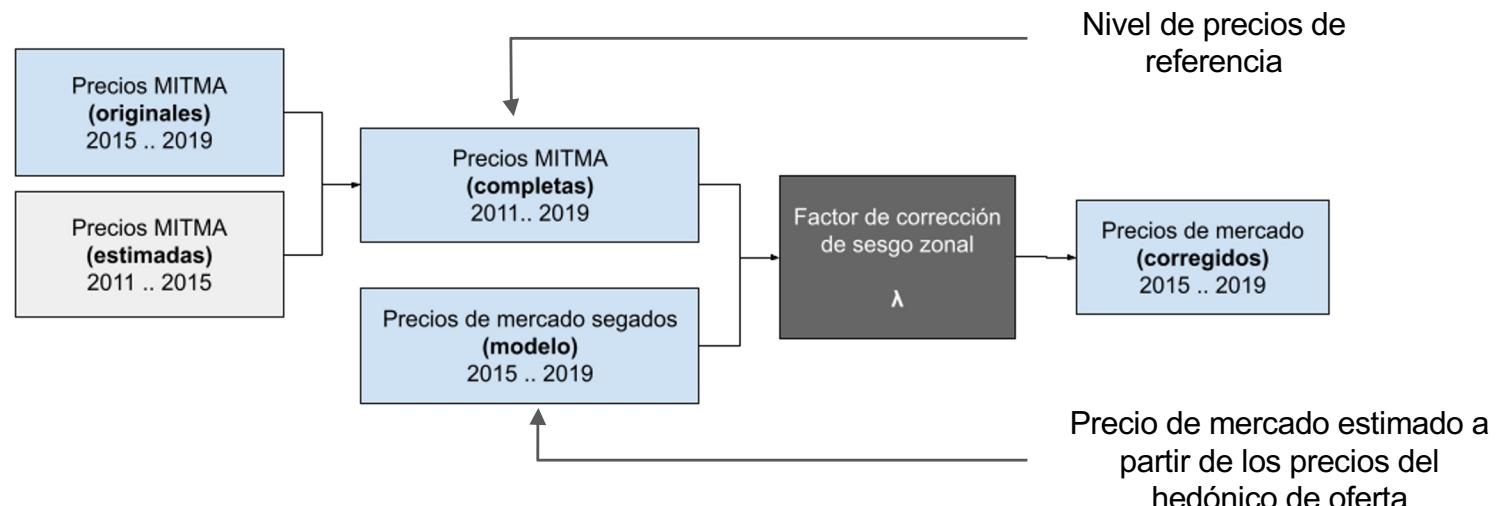
## Calidad del ajuste del modelo de mercado

El modelo logra un ajuste entre el 0,8 y 0,9 en vivienda plurifamiliar, y superior al 0,95 para vivienda unifamiliar, que puede considerarse excelente (Rico y Taltavull, 2021; Ho, 2021). Mantiene la distribución conjunta de los precios la distancia Hellinger entre los precios de los distintos estratos es de 0,042 (<0,05), lo que indica un buen ajuste (Leucescu y Agafitei, 2013).



## Eliminación del sesgo zonal (hedónico final)

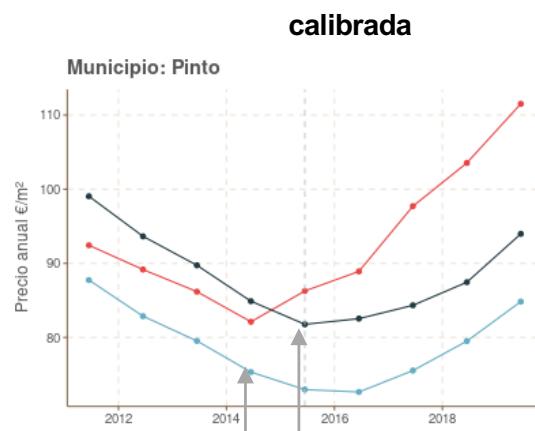
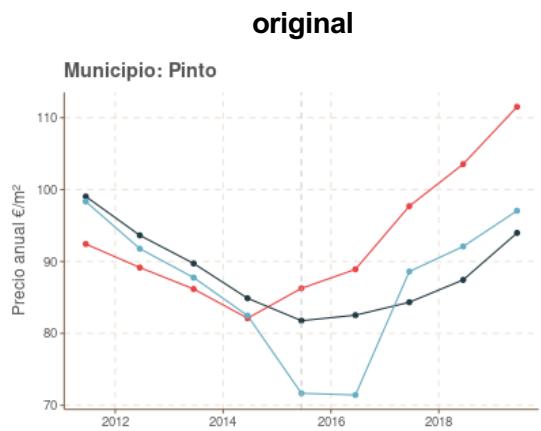
Las limitaciones en términos de desglose y número de registros de la EPF introduce falta de respuesta en los precios que no se debe ignorar (Fuller, 2011; Lohr, 2019). Para solventarlo se realiza un **proceso final de calibración basada en ratios** (factores de corrección de sesgo zonal  $\lambda$ ), calculados en base a la relación observada entre los precios de mercado publicados por el IPVA del MITMA, los precios de mercado de la EPF y los precios de oferta.



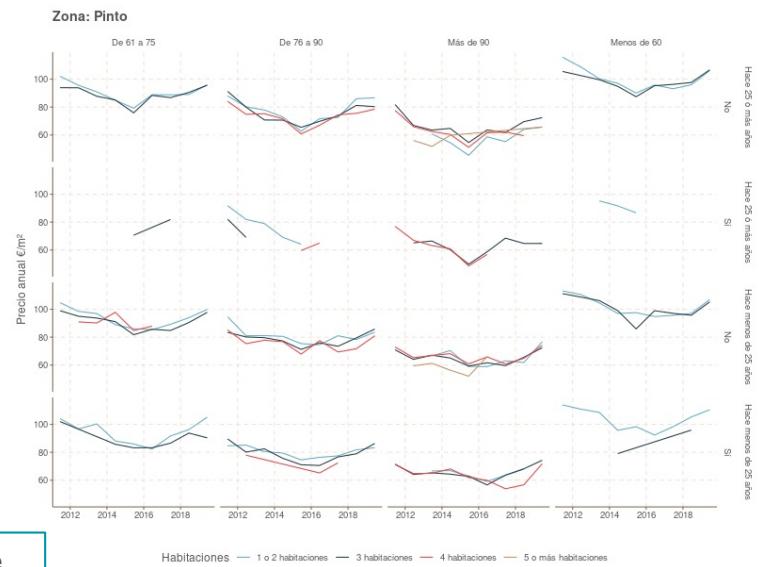
# Coherencia estructural del modelo hedónico final

De forma visual se puede observar como la coherencia de la serie agregada de un área, como del desglose funcional, como vemos para el caso de Pinto. Además menciona los parámetros generales de la distribución original de precios (media, desviación típica y cuartiles de precios).

**desglose superficie y habitaciones**

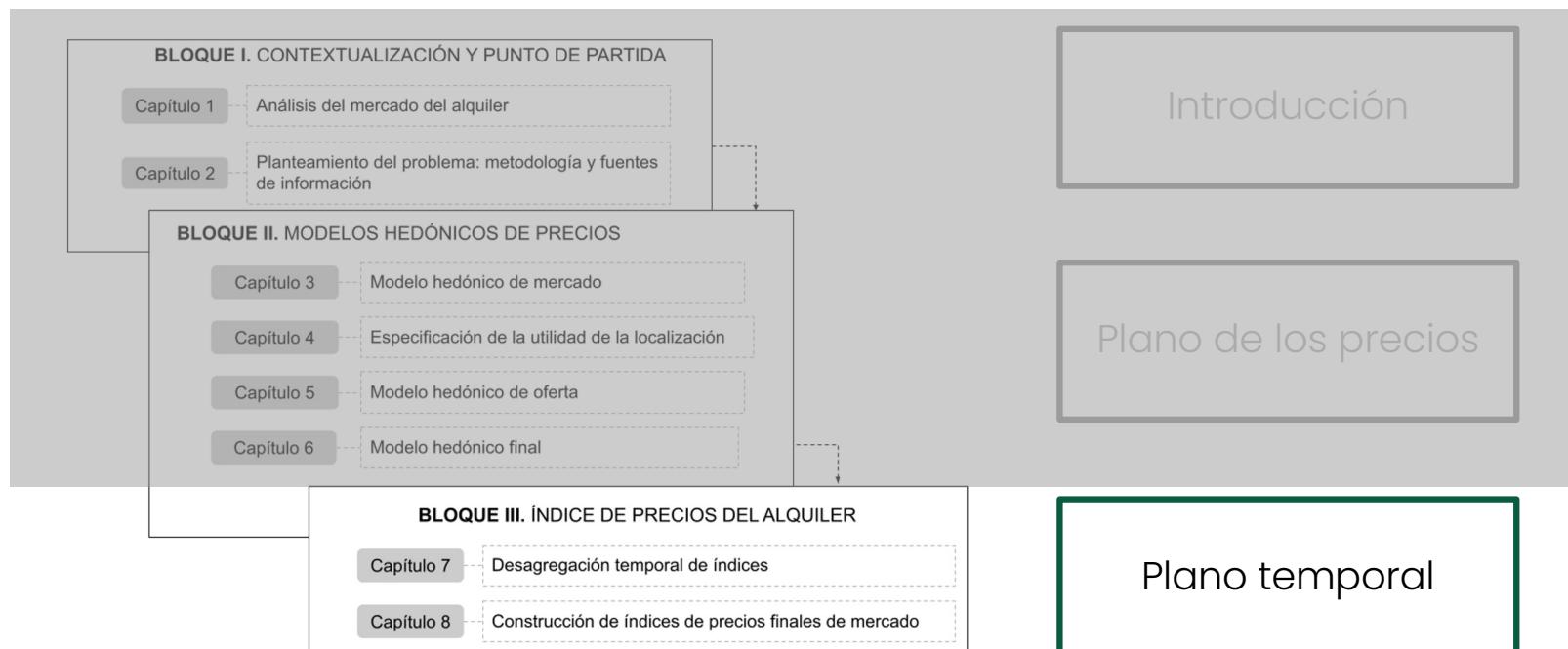


Serie de precios  
■ Mercado EPF  
■ Mercado mitma  
■ Oferta

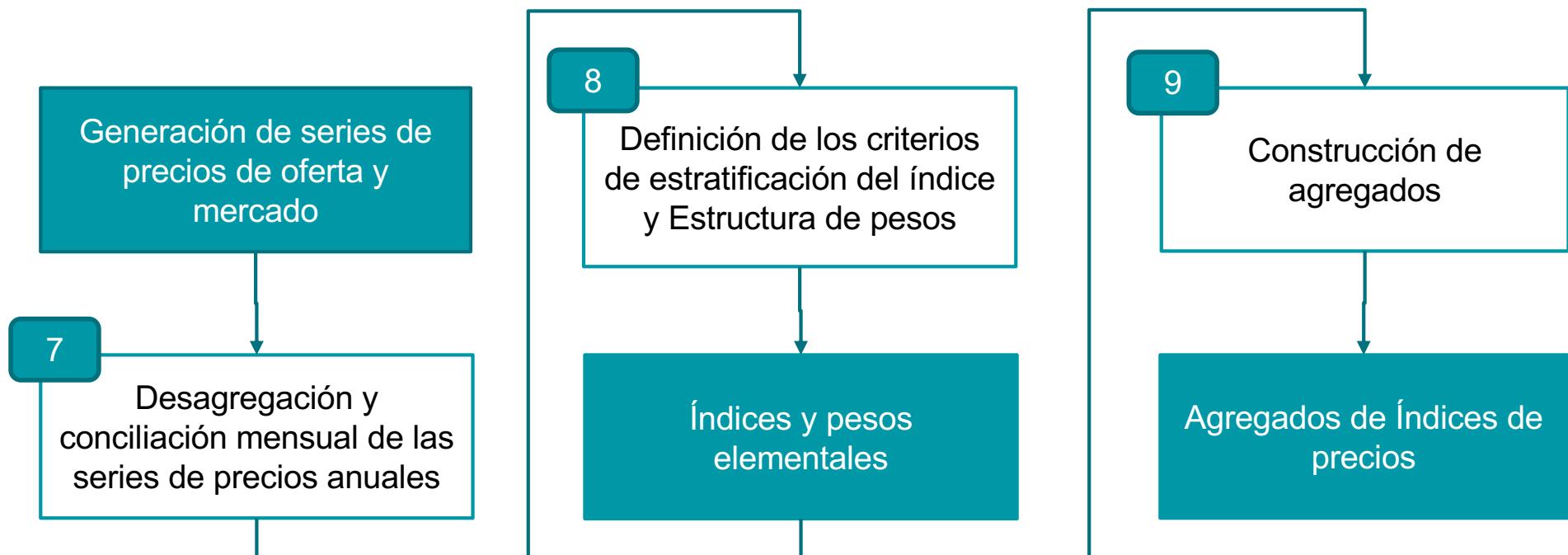


La serie corregida mantiene la escala de precios general del modelo de mercado porque las muestras de MITMA y INE/EPF no son las mismas

# Metodología: plano temporal



## Fases elaboración del modelo de temporal



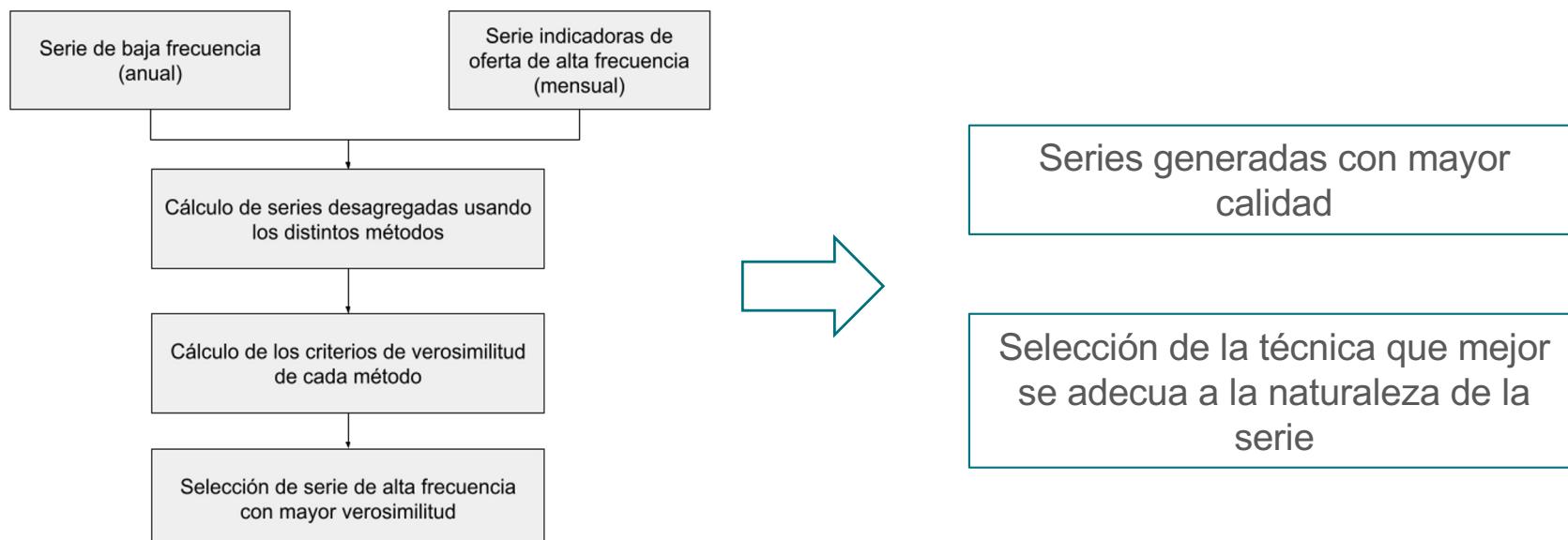
## Necesidad de desagregación temporal y reconciliación

El uso de información mensual y anual genera dos tipos de problemas en las series: (1) inconsistencia entre datos de mercado anuales y mensuales; (2) diferencias abruptas entre final y principio de año y variaciones mensuales extremas.



## Proceso de desagregación temporal

Gran variedad de series y métodos de desagregación y reconciliación (Eurostat, 2015), no hay un método óptimo y no es factible la elaboración manual. Usamos un método que **maximiza la verosimilitud** de encontrar el método **que mejor funciona** para cada serie, resolviendo los problemas habituales de la desagregación temporal, robustez del método y operatividad (Quilis, 2018).



## Estratificación del índice

La estratificación usa dos criterios de desagregación: el funcional y el zonal, combinables entre ambos.

**Tabla 8.1.** Variables usadas para la estratificación funcional

Variable	Descripción	Valores
SUT	Superficie útil	4 niveles
NHAB	Número de habitaciones	4 niveles
ANOCONSC	Año de construcción	4 niveles
TIPOEDIF	Tipo de edificio	4 niveles
TIPOCASA	Tipo de vivienda	3 niveles
ZONARES	Tipo de zona residencial	3 niveles: renta baja, media y alta
INTERINPSP	Intervalo de ingresos mensuales netos totales	3 niveles: ingresos bajos, medios y altos
FACTORGASTOT6	Gastos familiares	3 niveles, gasto bajo-medio, medio alto, alto
DENSI	Densidad de población del municipio	3 niveles
TAMAU	Tamaño del municipio en población	5 niveles

Estratificación zonal



## Definición del índice: ponderaciones

Se utiliza un índice Fisher encadenado, por su conveniencia como índice superlativo (Eurostat, 2014; Hill 2010), por sus propiedades económicas (Balk, 1995; Diewert, 1976) y menor propensión a sesgos (Graf, 2020). El dinamismo del mercado inmobiliario, cuyas condiciones cambian de forma continua, se usa un índice encadenado. Como el IPV de INE utiliza ponderaciones híbridas con los dos períodos anteriores, como recomienda Eurostat (2014).

$$W_e^a = \frac{\sqrt{Q_e^{(a-1,a-2)} \times \hat{P}_e^{a-1} \times \hat{P}_e^{(a-1,a-2)} \times Q_e^{a-1}}}{\sum_{\forall e} \sqrt{Q_e^{(a-1,a-2)} \times \hat{P}_e^{a-1} \times \hat{P}_e^{(a-1,a-2)} \times Q_e^{a-1}}}$$

donde  $Q_{(a-1,a-2),e}$  es la cantidad media, en metros cuadrados de superficie útil, en el estrato  $e$  para los dos años anteriores a  $a$ . Mientras que,  $P_{a-1,e}$  es el precio medio en €/m útiles del estrato en el año anterior. De forma análoga,  $P_{(a-1,a-2),e}$  es la media del precio de los dos años anteriores para  $e$ , y  $Q_{a-1,e}$  los metros útiles medios del año anterior.

## Índice elemental

El índice es de tipo hedónico basado en doble imputación, cuyos precios base y del periodo son los valores estimados por los modelos hedónicos (Eurostat, 2014). El uso de los pesos calibrados y el modelo hedónico permite controlar los cambios de calidad, asegurando las condiciones de homogeneidad, aprobadas por el grupo de trabajo de Eurostat aplicadas en la mayoría de los países de la UE.

Precio estimado por el modelo hedónico para el estrato  $e$  y año  $a$

$$I_e^a = \frac{\hat{P}_e^a}{\hat{P}_e^{a-1}} \times 100$$

Índice para el estrato  $e$  y el año  $a$

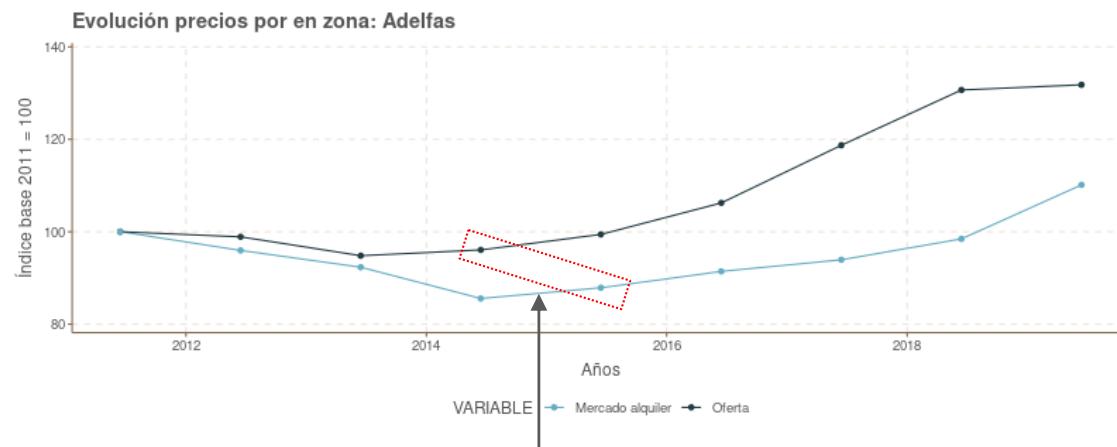
Pesos de los estratos  $e$  del agregado  $A$

$$I_A^a = \sum_{e \in A} W_e^a \times I_e^a$$

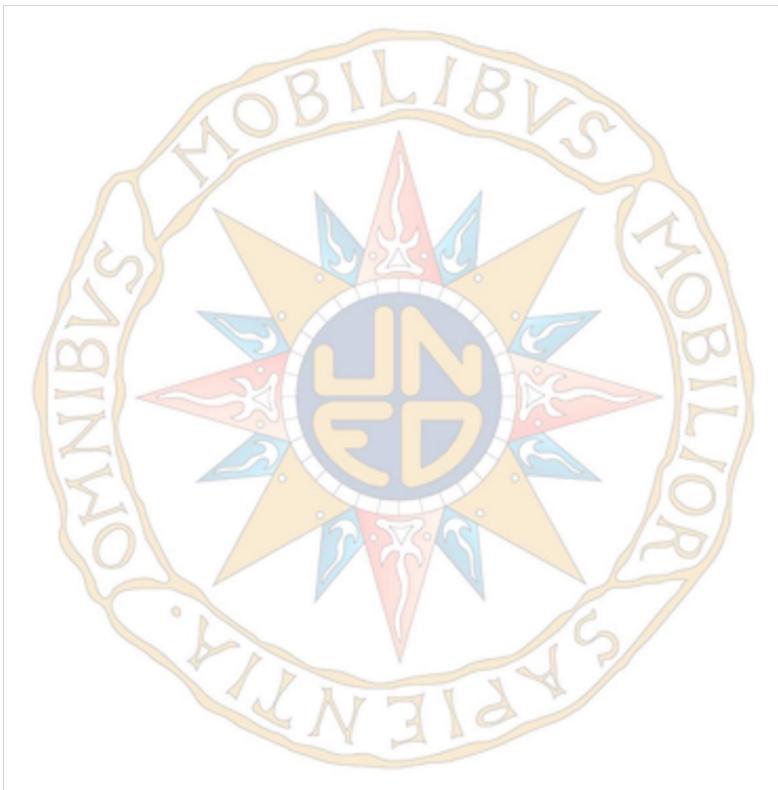
Índice para el agregado  $A$

## Construcción de las series de índices

Utilizando los índices elementales y los pesos se pueden construir las series para los agregados que se determinen.



Se observa como el precio de mercado y de alquiler están co-integradas, con la oferta temporalmente adelantada



# **Resultados y Conclusiones**

## Estabilidad de los índices de precios

El índice es estructuralmente estable, según indica la distancia Hellinger de las cantidades entre años contiguos de media < 0.05 (Leucescu y Agafitei, 2013). Se observa baja variabilidad en los estratos con suficiente muestra (hasta el percentil 99 de tamaño).

**Tabla 8.7.** Distancia Hellinger cantidades del índice entre pares de años (A y B), escalados a tanto por 1000

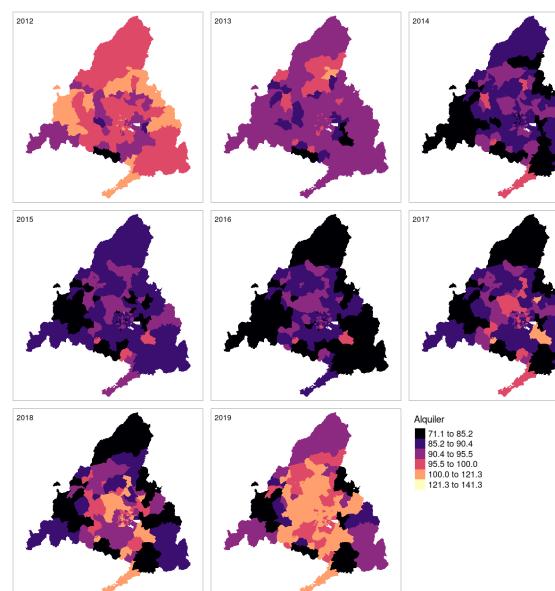
Año A	Año B							
	2012	2013	2014	2015	2016	2017	2018	2019
2011	0,06	0,07	0,08	0,07	0,07	0,10	0,13	0,14
2012		0,01	0,02	0,01	0,02	0,05	0,07	0,08
2013			0,01	0,01	0,01	0,03	0,06	0,07
2014				0,01	0,01	0,03	0,05	0,06
2015					0,01	0,04	0,07	0,07
2016						0,04	0,06	0,07
2017							0,03	0,04
2018								0,01

Fuente: elaboración propia

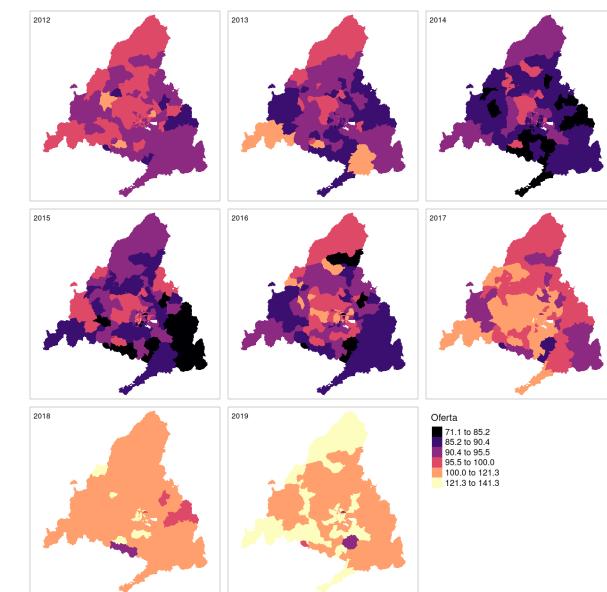
## Ventajas de un índice de precios dual: oferta y alquiler

La generación "dual" del índice, para el mercado de alquiler y de oferta permite un análisis pormenorizado de los dos fenómenos. El estudio de la evolución de ambos permite estimar los desequilibrios entre oferta y demanda, tiempos de comercialización o poder de negociación de los agentes (Galesi et al. 2020).

**Figura a.** Índice de precios de alquiler por zona

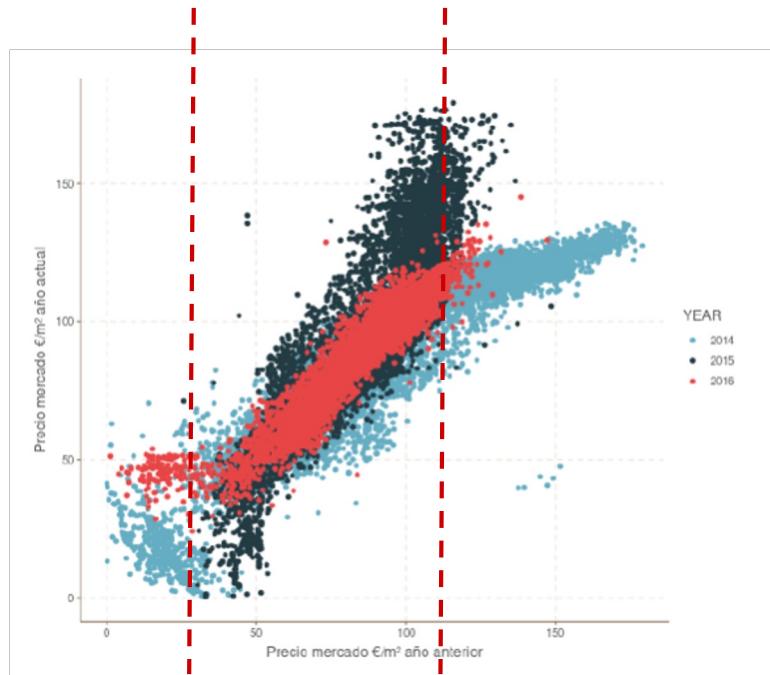


**Figura b.** Índice de precios de oferta por zona



## El modelo es válido varios años

Relación de precios de mercado estimados por el modelo del año y estimados para el año anterior.



Los valores de mercado estimados con modelos de años consecutivos, entre 2014 y 2016, guardan un **coeficiente de correlación que varía entre el 0,86 y 0,94** (mayor aún en la zona central de los precios).

Esto es particularmente interesante porque en los años 2014 y 2015 se produce un cambio de tendencia en los precios de alquiler.

## Aplicación interactiva y datos abiertos

Aplicación interactiva para probar los resultados, se incluye el acceso a los recursos, datos gráficos y de contenidos de la tesis, así como el de los artículos asociados.



Tablas

	PERIOD	CAPROV	VARIABLE	LOCATION	LOCATIONNAME	IDX_CHAIN	IDX_ACC	IDX_ACC_SIMPLE
198	2011-06-15 00:00:00	No	PRICE_RENT	M_28080	Majadahonda	1	1	1
199	2011-06-15 00:00:00	No	PRICE_ASKING	M_28080	Majadahonda	1	1	1
200	2011-06-15 00:00:00	No	PRICE_RENT	M_28080	Majadahonda	0.0172	0.0172	0.0172



Usa el QR para acceder a la app web

## Síntesis de conclusiones

1

Esta metodología presenta una novedad en la construcción de Índice de Precios de la Vivienda en alquiler usando datos alternativos, que soluciona gran parte de los retos de las aproximaciones tradicionales, y ofrece un alto grado de desagregación y actualizado

2

Al producirse por diseño dos índices de forma simultánea, oferta y mercado, permite el análisis del mercado de una forma completa

3

Se combina información inmobiliaria de portales inmobiliarios de forma coherente con la estadística pública oficial, eliminando los sesgos poblacionales. El diseño muestral trata eficazmente la heterogeneidad de la vivienda

4

Aproximación multidisciplinar en la que se exploran espacios comunes entre las ciencias de la computación y las ciencias económicas, con aplicabilidad académica e industrial

## Próximas líneas de investigación

1

Uso de métodos hedónicos  
**más precisos**

2

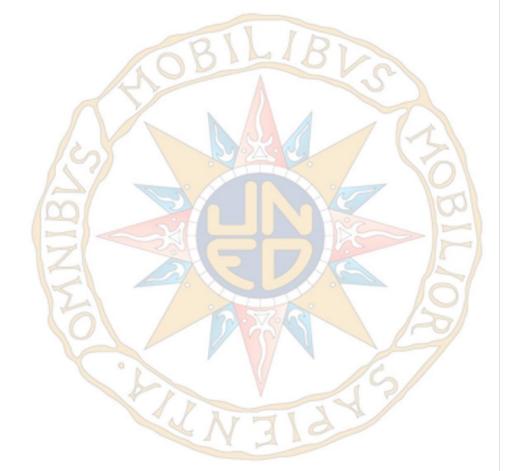
Solución de los **sesgos** de  
información ausente y zonal

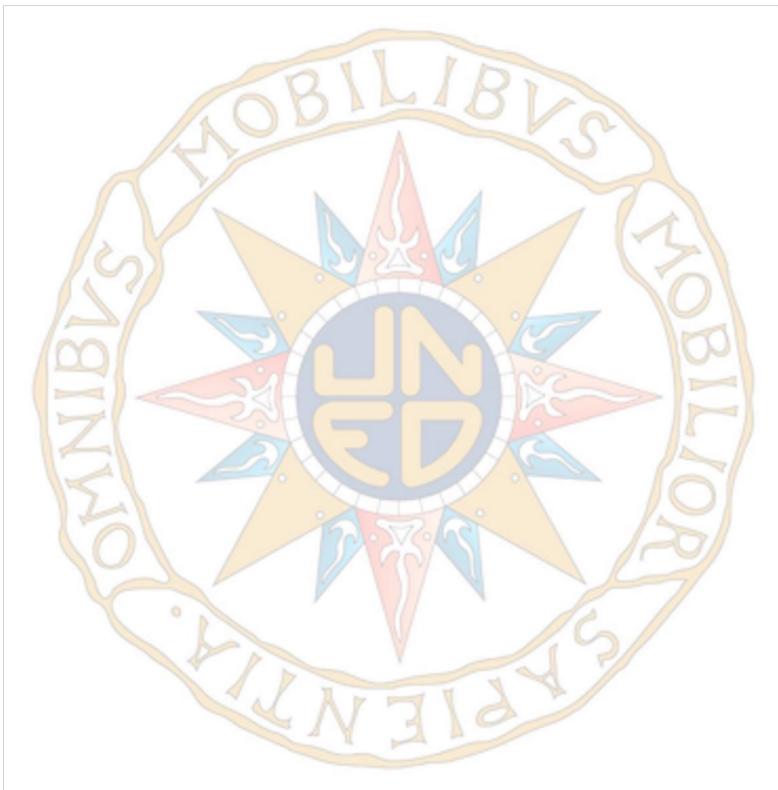
3

Incorporación de fuentes  
adicionales y evaluación en  
ámbito más amplio

- Aplicación de métodos de regresión basados en redes neuronales artificiales de aprendizaje profundo (*deep learning*). Particularmente en redes profundas de grafos (GNN), y las de tipo espacio-temporal.
- Contrastar resultados del ajuste de la reponderación de la EPF con el Censo de 2021
- Nuevas fuentes de datos que complementen aquellos segmentos que tienen menor soporte de información.
- Incorporar el aspecto geográfico detallado en la reponderación.
- Controlar anomalías en la estructura de las series de precios generadas.
- Mejoras en el proceso de desagregación temporal

¡ Muchas gracias !

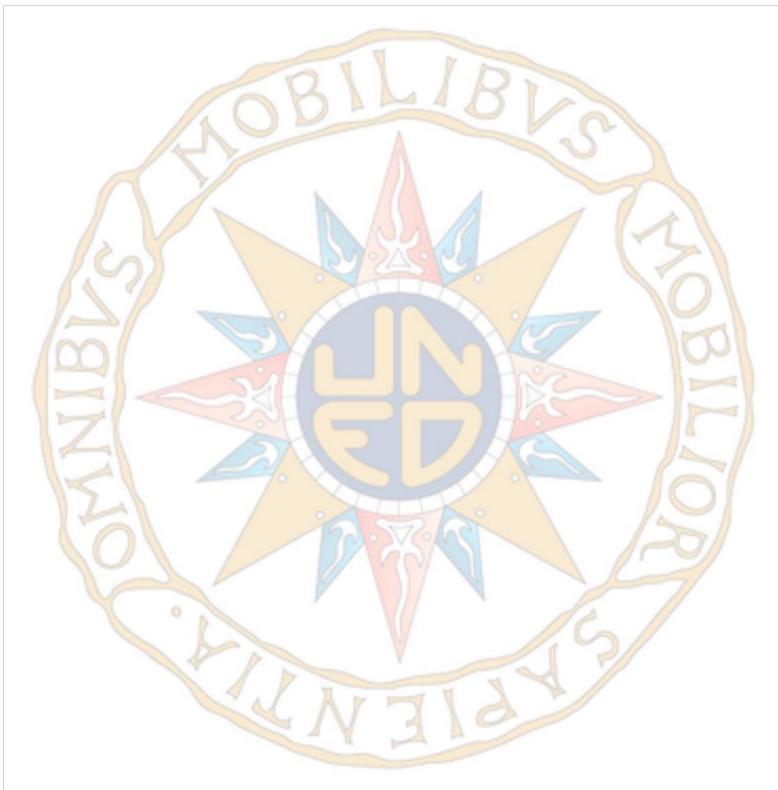




# Bibliografía

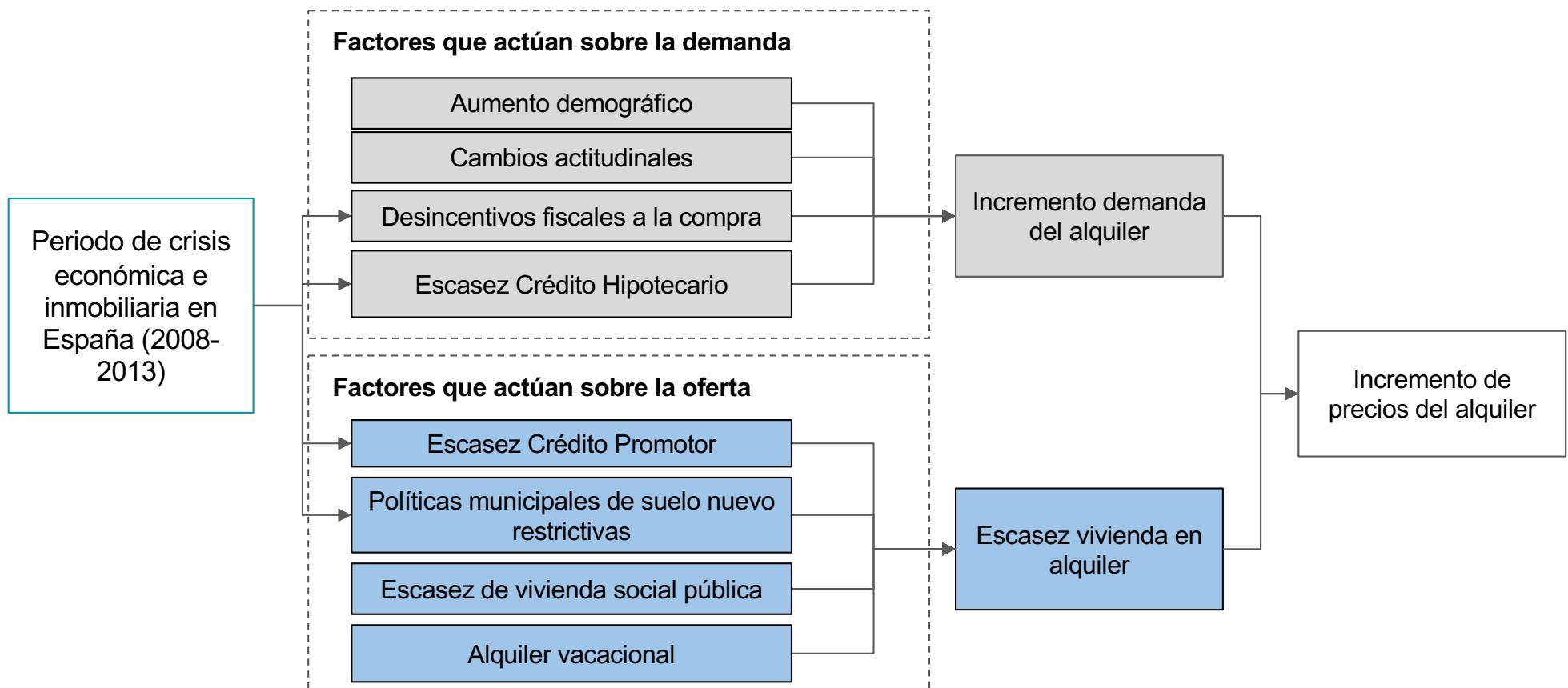
- Chen B. y Andrews S. (2008). *An empirical review of methods for temporal distribution and interpolation in the national accounts*. *Survey of current business* 88 (5): 31.
- Deville J.-C. y Särndal C.-E. (1992). *Calibration estimators in survey sampling*. *Journal of the American statistical Association* 87 (418): 376-382.
- Friedman J. y Weinberg D.H. (1981). *The demand for rental housing: Evidence from the housing allowance demand experiment*. *Journal of Urban Economics* 9 (3): 311-331.
- Kiel K.A. y Zabel J.E. (2008). *Location, location, location: The 3L Approach to house price determination*. *Journal of Housing Economics* 17 (2): 175-190
- Leucescu A. y Agafitei M. (2013). *Statistical matching: a model based approach for data integration*. *Eurostat methodologies and Working papers*. <https://doi.org/10.2785/44822>.
- Särndal C.-E. y Lundström S. (2008). *Assessing auxiliary vectors for control of nonresponse bias in the calibration estimator*. *Journal of Official Statistics* 24 (2): 167.

**COMPLETAR**

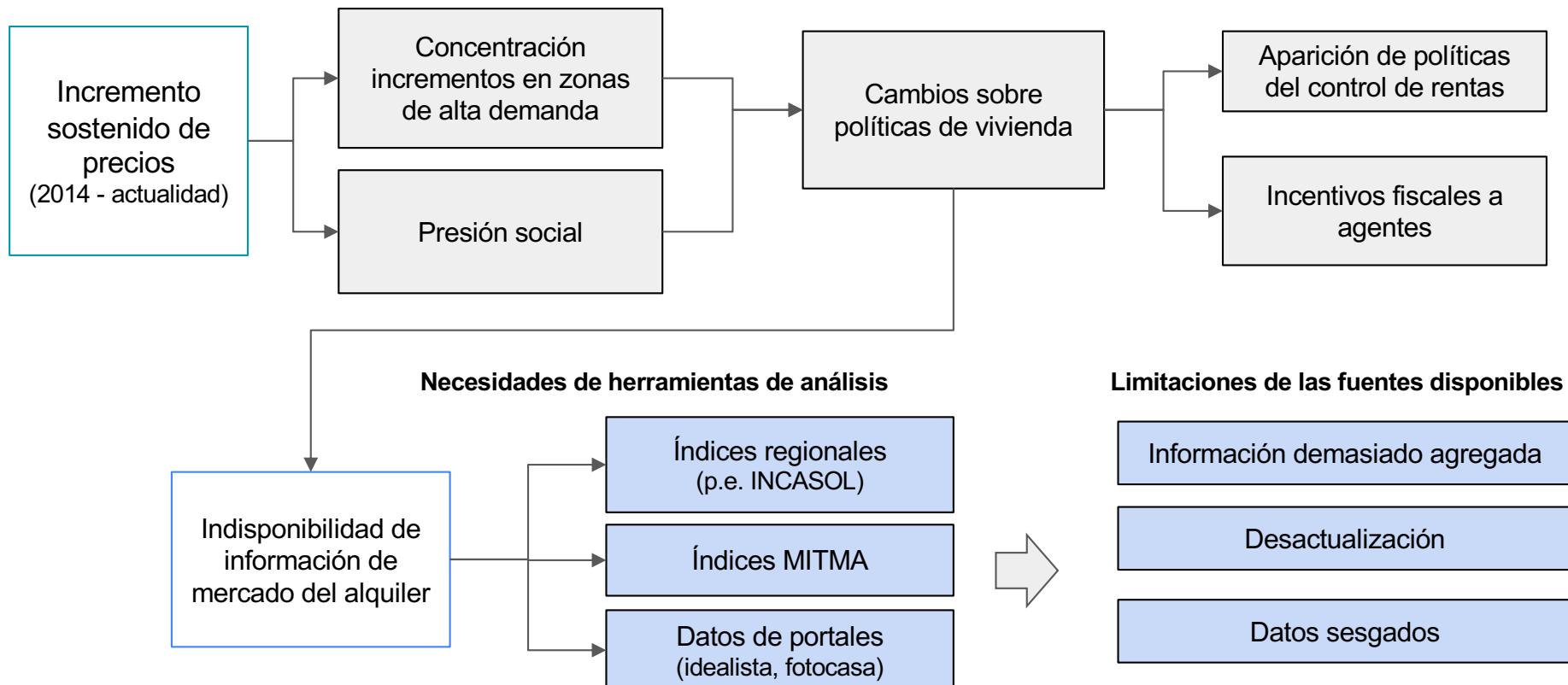


# Anexos

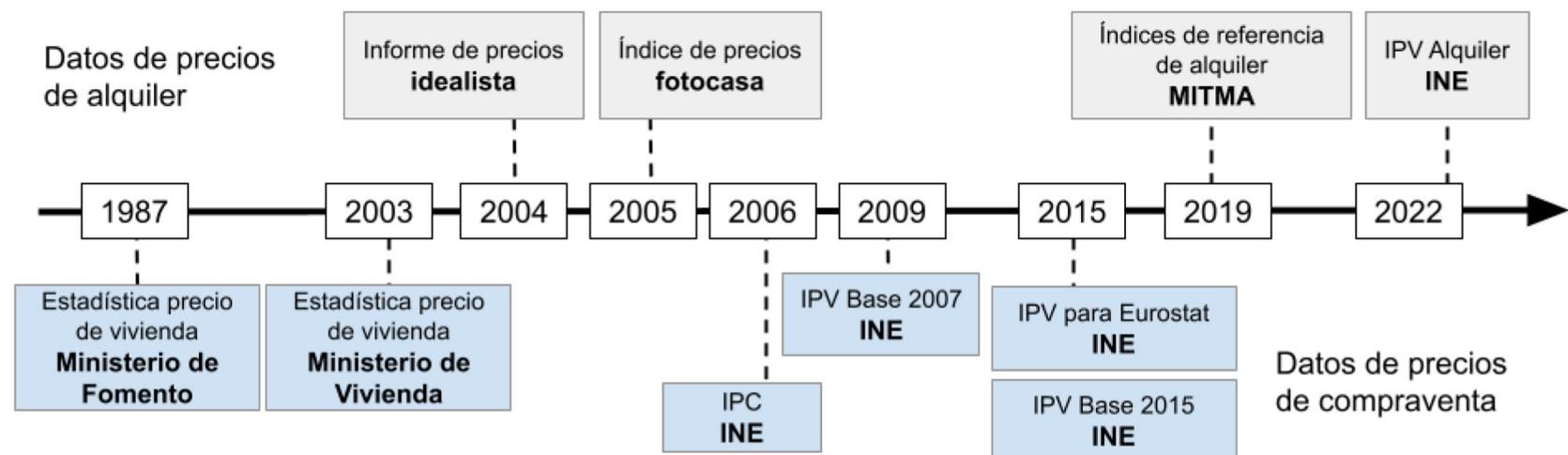
# Antecedentes: mercado inmobiliario post gran crisis - actualidad



# Motivación: visión parcial y con retraso de las dinámicas del mercado



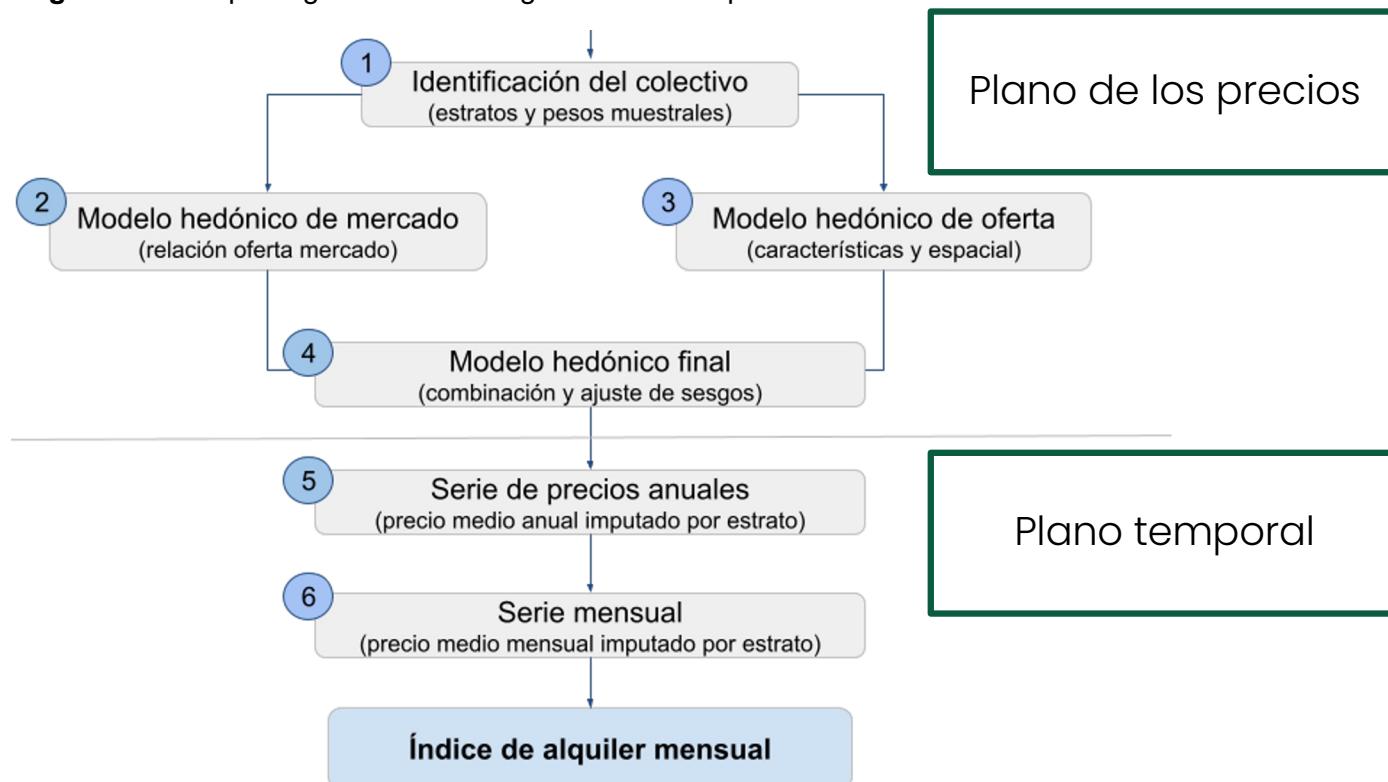
**Figura 2.10.** Evolución histórica de la publicación datos de precios de la vivienda en España



*Fuente:* elaboración propia.

# Metodología: esquema general

**Figura 4.** Descripción general metodología de índice de precios residenciales.



Fuente: elaboración propia.

# Novedades de la metodología propuesta

nuevos datos

Mayor disponibilidad  
información de fuentes  
públicas

Datos catastrales

Datos desagregados de  
portales inmobiliarias

Información cartográfica  
pública

Otras fuentes públicas

nuevas técnicas  
de modelado

Técnicas  
complementarias a la  
econometría clásica y  
mayores capacidades  
de proceso

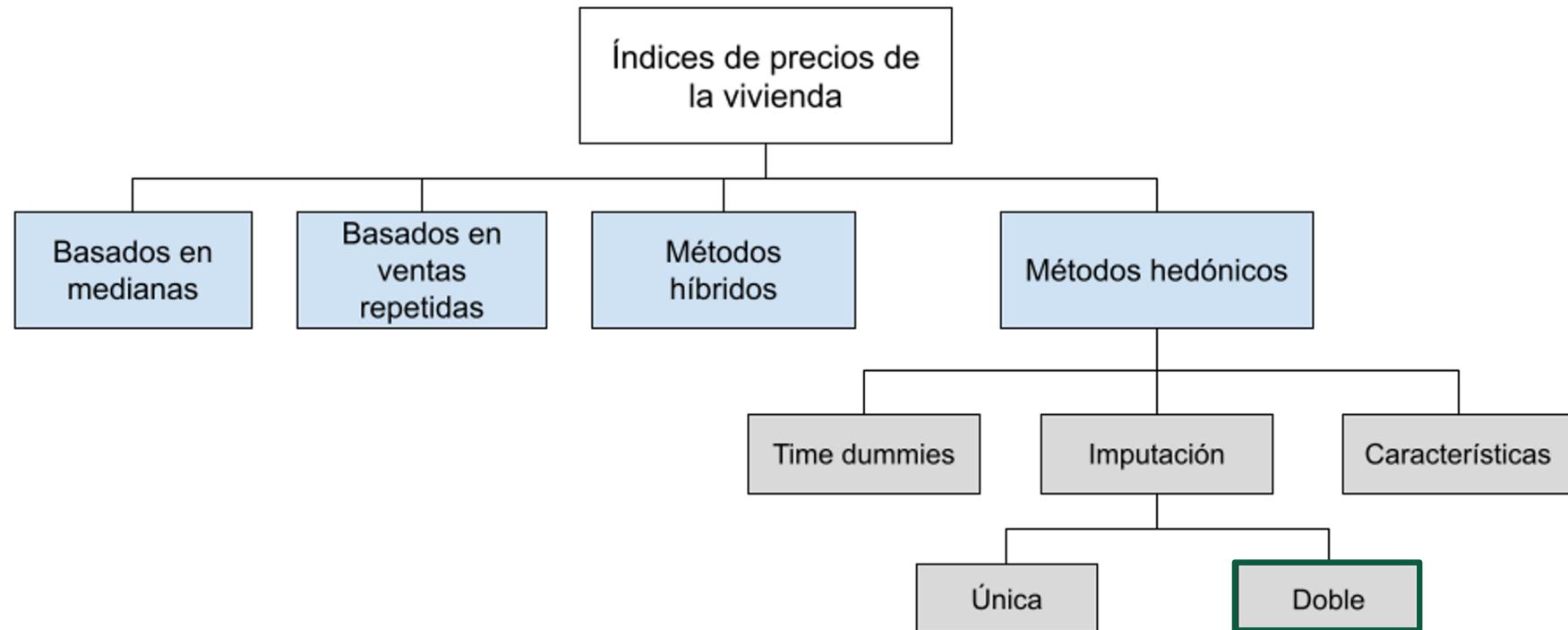
Aprendizaje estadístico /  
automático

Geoestadística

Big Data

# Técnicas de construcción de índices de precios de Vivienda

**Figura 5.** Taxonomía de metodologías para la construcción de Índices de Precios de la Vivienda.



Fuente: elaboración propia

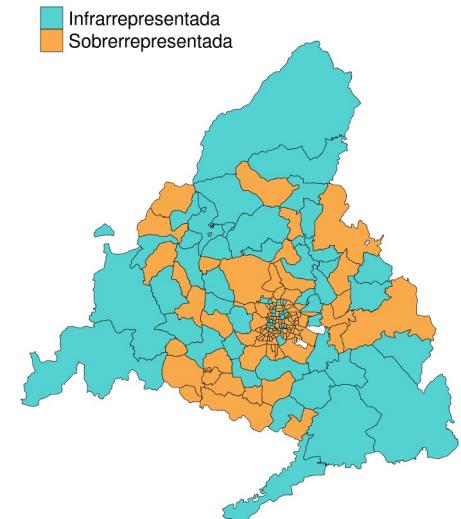
## Retos ante las fuentes de información

Las fuentes de información llevan asociados ciertos retos, en particular para el censo, la EPF e Idealista, que determinarán la estructura metodológica, en el caso de la falta de datos en Idealista se construyen varios modelos: imputación de año de construcción, superficie útil y estimación de gasto familiar.

Estadística pública:

- Información anual (EPF) o poco frecuente (Censo)
- El nivel de desglose zonal en EPF es muy limitado.
- El número de variables es muy reducido.
- Los cambios metodológicos pueden hacer los datos poco estables

**Sesgos de infra y sobrerepresentación.**



El dato de Idealista:

- El dato de oferta con sesgos de no respuesta e infra y sobre representación.
- Dato generado por los usuarios: subjetividad, errores, etc.
- No cuenta con algunas variables para realizar el cruce con EPF y Censo (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*).

## Resultados de la doble calibración

La calibración ha requerido aplicar un proceso iterativo para encontrar las bandas superior e inferior de los elevadores, se ha seguido un proceso iterativo para encontrarlas (Rao, 1996). Siendo el rango más amplio de 0,05, para la parte inferior, a 10 para la superior. Se observa que el rango se amplía, especialmente para los años 2018 y 2019, pero incluso en este caso no representan g-distancias extremas (D'Orazio et al., 2006). En términos zonales, no se aprecian diferencias importantes entre Madrid y el resto de la provincia.

**Tabla 3.6.** Bandas de calibración Censo y EPF

Año	Censo				EPF	
	Madrid		Resto CAM		Todas las zonas	
	Inferior	Superior	Inferior	Superior	Inferior	Superior
2011	0,28	4,50	0,27	6,67		
2012	0,24	6,00	0,24	6,95	0,46	1,90
2013	0,20	5,00	0,20	5,00	0,58	2,00
2014	0,25	4,00	0,20	3,00	0,28	3,00
2015	0,26	4,00	0,16	4,50	0,38	4,00
2016	0,18	3,00	0,18	3,25	0,38	3,00
2017	0,10	3,50	0,10	4,50	0,08	5,00
2018	0,05	10,00	0,05	10,00	0,10	5,50
2019	0,05	10,00	0,05	10,00	0,10	5,50

Fuente: elaboración propia

# Variabilidad de la composición geográfica de pesos de la EPF

La composición en estratos al discretizar por cada una de las variables, usando los criterios de la EPF (para poder hacer el cruce) da lugar a ciertas inestabilidad temporal en la presencia/ausencia de los distintos estratos, que se puede trasladar a una mayor variabilidad de los datos agregados a ese nivel

**Tabla 3.8.** Presencia de estratos, municipio de Tres Cantos (casa de tipo medio)

Densidad	Tipo	Edificio	Hab.	Zona	2011	2012	2013	2014	2015	2016	2017	2018	2019	
densa	Casa media	10 ó más	1 o 2	Urbana alta	X	X	X	X	X					
densa	Casa media	10 ó más	1 o 2	Urbana media	X	X	X	X	X					
densa	Casa media	menos de 10	1 o 2	Urbana alta	X	X	X	X	X					
densa	Casa media	menos de 10	1 o 2	Urbana media	X	X	X	X	X					
densa	Casa media	10 ó más	3	Urbana alta	X	X	X	X	X					
densa	Casa media	10 ó más	3	Urbana media	X	X	X	X	X					
densa	Casa media	menos de 10	3	Urbana alta	X	X	X	X	X					
densa	Casa media	menos de 10	3	Urbana media	X	X	X	X	X					
densa	Casa media	10 ó más	4	Urbana alta		X	X	X	X					
densa	Casa media	10 ó más	4	Urbana media	X	X	X	X	X					
densa	Casa media	menos de 10	4	Urbana alta	X	X	X	X	X					
densa	Casa media	menos de 10	4	Urbana media	X	X	X	X	X					
densa	Casa media	10 ó más	5 o más	Urbana media			X	X	X					
densa	Casa media	menos de 10	5 o más	Urbana alta				X						
densa	Casa media	menos de 10	5 o más	Urbana media		X	X							
intermedia	Casa media	10 ó más	1 o 2	Urbana alta				X	X	X				
intermedia	Casa media	10 ó más	1 o 2	Urbana media				X	X	X				
intermedia	Casa media	menos de 10	1 o 2	Urbana alta				X	X	X				
intermedia	Casa media	menos de 10	1 o 2	Urbana media				X	X	X				
intermedia	Casa media	10 ó más	3	Urbana alta				X	X	X				
intermedia	Casa media	10 ó más	3	Urbana media				X	X	X				
intermedia	Casa media	menos de 10	3	Urbana alta				X						
intermedia	Casa media	menos de 10	3	Urbana media				X	X	X				
intermedia	Casa media	10 ó más	4	Urbana media				X	X	X				
intermedia	Casa media	10 ó más	4	Urbana alta				X	X	X				
intermedia	Casa media	menos de 10	4	Urbana alta				X	X	X				
intermedia	Casa media	menos de 10	4	Urbana media				X	X	X				
intermedia	Casa media	10 ó más	5 o más	Urbana media				X	X	X				
intermedia	Casa media	menos de 10	5 o más	Urbana media				X						

Fuente: elaboración propia

## Problemas de inconsistencia temporal de la EPF

Los totales originales de la EPF son significativamente variables año a año para algunos estratos, para solventar el impacto en el modelo y bajo el supuesto de que estas variaciones no son verosímiles se realiza un proceso de suavizado exponencial sobre la serie temporal de totales de calibración.

**Tabla 3.17.** Totales originales para calibración EPF

Variable	2011	2012	2013	2014	2015	2016	2017	2018	2019
Total	428667	96%	95%	95%	99%	111%	113%	109%	122%
1 o 2 habitaciones	267959	106%	104%	106%	110%	120%	112%	118%	138%
3 habitaciones	126257	87%	78%	81%	84%	101%	115%	98%	99%
4 habitaciones	25964	89%	158%	82%	132%	144%	161%	131%	142%
5 o más habitaciones	8487	48%	39%	70%	39%	33%	40%	41%	68%
Menos de 60	174119	110%	93%	104%	108%	120%	105%	108%	122%
De 61 a 75	99464	77%	86%	97%	81%	89%	92%	93%	106%
De 76 a 90	86505	100%	93%	77%	117%	148%	155%	138%	174%
Más de 90	68579	98%	133%	95%	99%	109%	145%	119%	124%
Menos de 10 viviendas	81814	95%	114%	107%	99%	112%	130%	115%	113%

## Ajuste del modelo hedónico de oferta

El nivel de ajuste de los modelos hedónicos de oferta es muy alto, obteniendo valores más altos en el caso de la vivienda unifamiliar. Las cifras son relativamente estables en el tiempo con cierta degradación en el año 2019.

**Tabla 7a.** Coeficiente R<sup>2</sup> del modelo ensamblado de oferta.

Tipo	Modelo	2011	2012	2013	2014	2015	2016	2017	2018	2019
Unifamiliar	atributos	88.2%	79.7%	83.9%	88.5%	89.6%	89.1%	81.6%	78.7%	79.2%
	accesibilidad	96.9%	96.2%	96.7%	96.5%	95.0%	94.9%	90.9%	90.2%	91.7%
	ensamblado	91.1%	82.2%	86.7%	91.8%	92.2%	91.6%	83.9%	80.8%	81.5%
Plurifamiliar	atributos	68.8%	70.9%	77.4%	75.9%	72.6%	71.7%	69.2%	69.0%	64.6%
	accesibilidad	93.6%	92.5%	93.9%	92.2%	92.4%	91.2%	89.3%	93.4%	88.7%
	ensamblado	79.2%	81.8%	87.9%	87.3%	83.7%	83.9%	81.3%	80.5%	76.4%

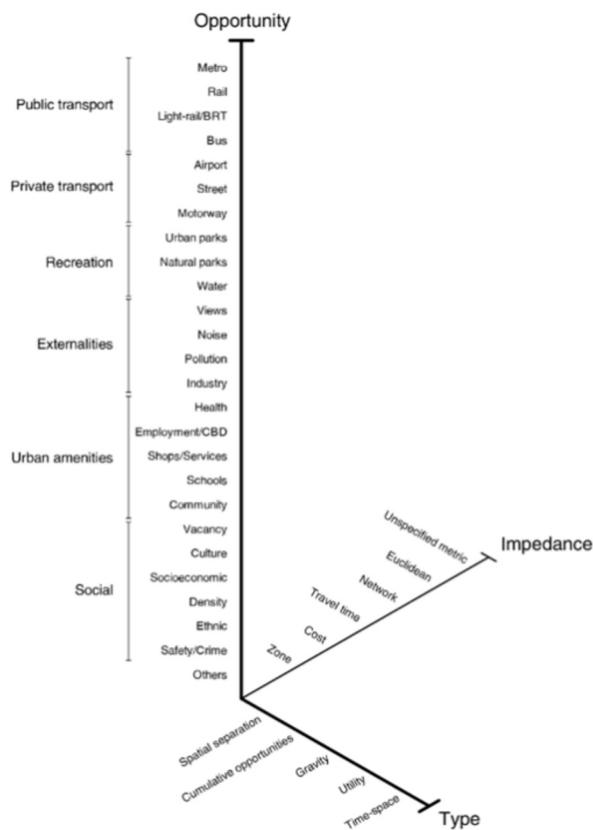
Fuente: elaboración propia

**Tabla 7b.** Error cuadrático medio (RMSE) y RMSE normalizado (en

Métrica	Tipo	2011	2012	2013	2014	2015	2016	2017	2018	2019
Absoluta	Plurifamiliar	28,13	24,59	18,84	20,04	25,30	29,18	34,43	34,54	38,77
	Unifamiliar	15,00	20,87	16,61	12,19	11,61	12,36	17,78	21,70	22,04
Normalizada	Plurifamiliar	17.0%	15.7%	12.8%	13.5%	16.3%	17.3%	17.7%	16.0%	17.3%
	Unifamiliar	14.3%	20.8%	17.9%	13.4%	12.6%	12.6%	17.3%	19.7%	19.2%

Fuente: elaboración propia

## Localización especificada a través de la accesibilidad



La localización como inductor del precio, calculado como la prima pagada por la utilidad marginal de la ubicación. Esta medida la estimamos como un **índice de accesibilidad u oportunidad**

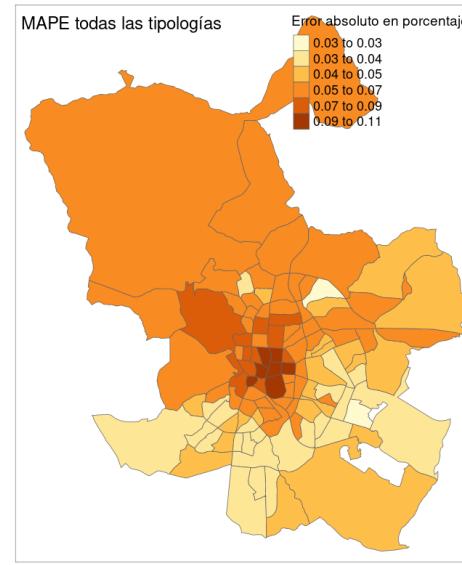
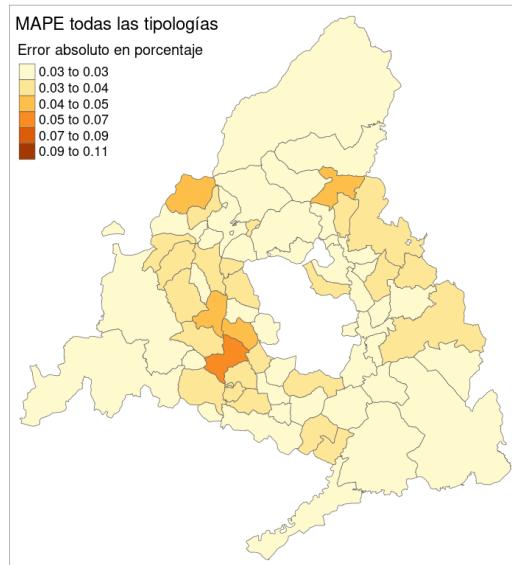
**Oportunidad:** elemento que aporta algún tipo de ventaja (**utilidad**) a un usuario de una vivienda

**Impedancia:** coste para desplazarse (acceder) de un lugar a un tipo de oportunidad

**Tipo:** forma de agregar las distintas oportunidades que están al alcance de una **impedancia** determinada

## Calidad del ajuste del modelo de oferta

El nivel de ajuste de los precios de oferta medio por zona de análisis es muy bajo: **por debajo del 5%** de error porcentual medio en los municipios de la Comunidad fuera de la capital. Para Madrid los errores también son bajos, con una concentración ligeramente mayor en el área del centro histórico.



## Localización especificada a través de la utilidad

La utilidad se puede especificar como una medida numérica de **accesibilidad** a través de la agregación de un conjunto de oportunidades que tiene alrededor, particularmente los de tipo **gravitatorio** se apoyan en la ley de Tobbler que asume una mayor importancia a los elementos más cercados. Se trata de una “media ponderada” espacial:

---

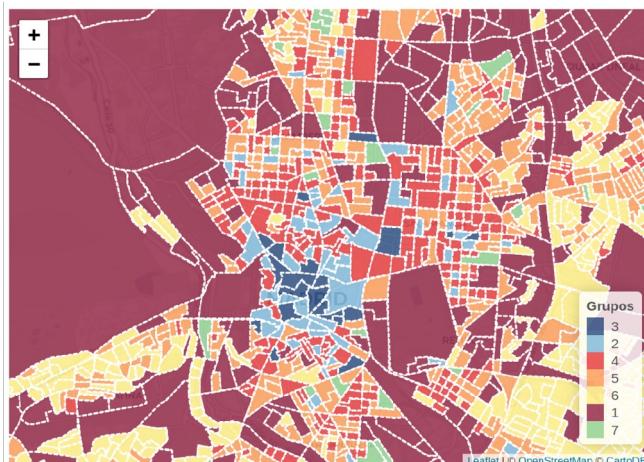
$$A_{nk}^{tm} = \sum_{i(m)=1}^I O_k(X, Y, d_{i(m)}) \cdot e^{-\beta_m \cdot d_{i(m)}}, m = \text{coche, a pie, } [2]$$

donde  $O_k(X, Y, d_i(m))$  representa un tipo de **oportunidad** a una distancia  $d$  desde un lugar ubicado en las coordenadas  $(X, Y)$ , dentro de los límites geográficos marcados por una serie de isócronas. La medida de distancia  $d_i(m)$  corresponde a los rangos de tiempos de viaje (**impedancia** en minutos). El parámetro  $\beta_m$  es la caída exponencial del índice para la impedancia aplicada, el valor de dicha variable se estima para cada oportunidad por si capacidad de explicar la dependencia espacial (medido numéricamente como el residuo de un modelo que no incorpora la interacción espacial).

## Atributos comunes: Clusters de zona como variable

Al igual que el calibrado, para garantizar la estabilidad del ensamblado se ha introducido variables comunes: tipo, superficie y cluster de zona. Esta agrupación se calcula a nivel de sección censal mediante K Medoides (Schubert y Rousseeuw, 2019), en dos ámbitos Madrid y resto de la CAM. Para evitar distorsiones de escala y colinealidad se preprocesan las variables mediante un proceso de escalado multidimensional (SMACOF).

**Figura 7.** Clusters de zona ayuntamiento de Madrid.



Fuente: elaboración propia

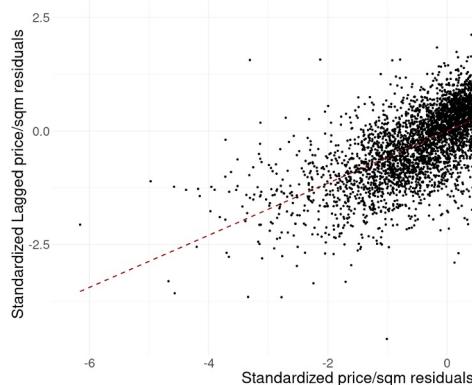
Código	Descripción	Secciones censales
1	Zona con la renta más alta. Muy baja densidad de viviendas, comercios y bares-restaurantes	471
2	Gran densidad de viviendas y lugares de ocio. Alta densidad de comercios y bares-restaurantes	66
3	Alta densidad de viviendas. Gran densidad de comercios, bares-restaurantes, hoteles y monumentos	21
4	Gran densidad de viviendas. Alta densidad de comercios y bares-restaurantes	313
5	Media densidad de viviendas, comercios y bares-restaurantes	587
6	Zona con la renta más baja. Baja densidad de viviendas, comercios y bares-restaurantes	897
7	Zona industrial. Baja densidad de viviendas. Media densidad de comercios y bares-restaurantes	51

## Capacidades de las medidas de accesibilidad

Las medidas de accesibilidad construidas permiten especificar de forma eficaz a la vista de los residuos del modelo. Para la ciudad de Madrid, estas medidas permiten ser sustituidas por las formas de especificación tradicional de la ubicación: location dummies, coordenadas, etc.

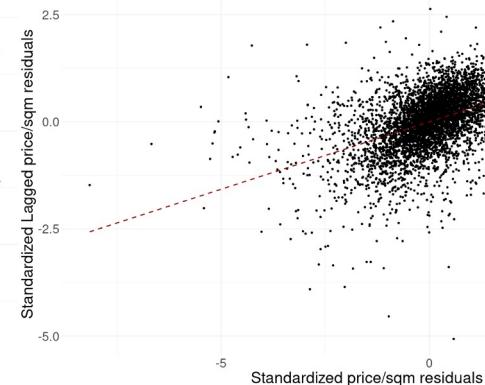
Modelo lineal sin accesibilidad

Moran's I Plot (OLS without location attributes): 0.55848  
Regression line approximate Moran's I



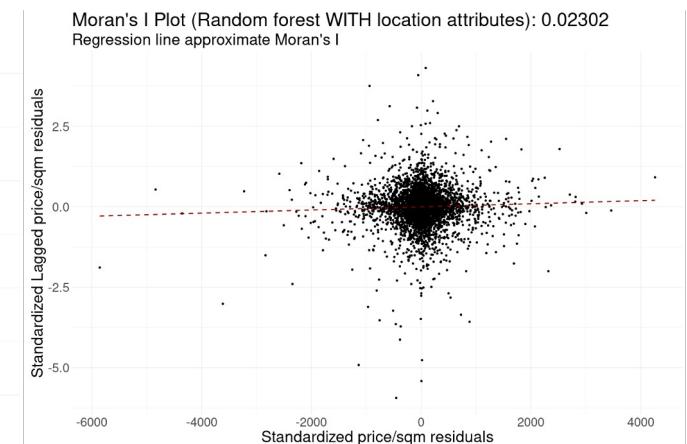
Modelo lineal con accesibilidad

Moran's I Plot (OLS WITH location attributes): 0.30235  
Regression line approximate Moran's I



Random Forests con accesibilidad

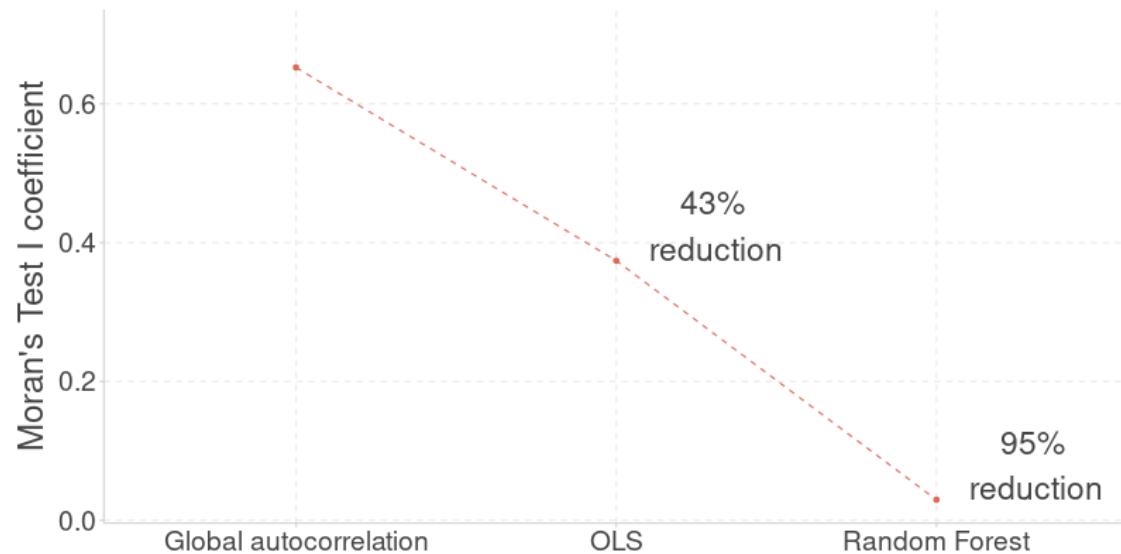
Moran's I Plot (Random forest WITH location attributes): 0.02302  
Regression line approximate Moran's I



**Mayor** autocorrelación espacial de los errores del modelo hedónico

**Menor** autocorrelación espacial de los errores del modelo hedónico

## Capacidades de las medidas de accesibilidad



# Interpretabilidad de las medidas de accesibilidad

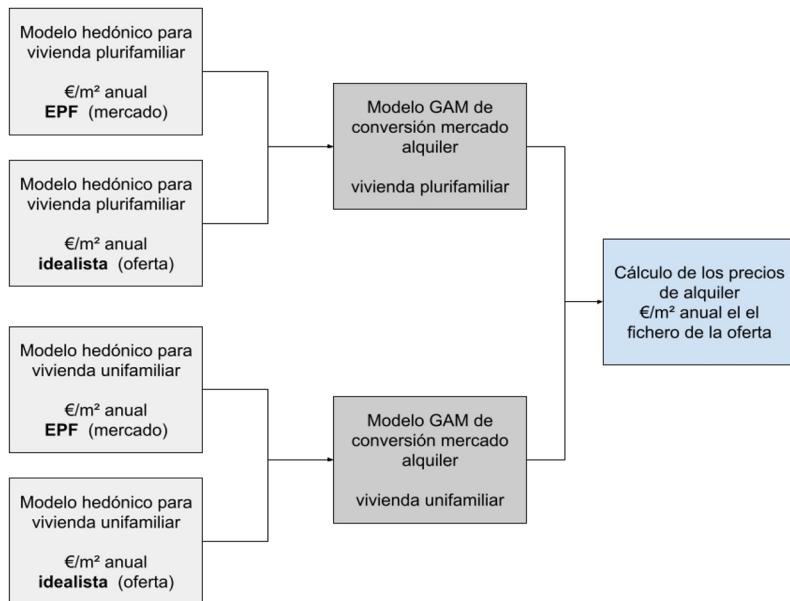
Componente	COMP	COMP	COMP	COMP
	WALK	WALK	WALK	WALK
	1	2	3	4
VACATIONAL	<b>0.93</b>	0.19		
TRANSPORT BUS	0.72	0.63		
TOURISM	<b>0.92</b>	0.23		
SHOP	<b>0.85</b>	0.44		
HOTEL	<b>0.95</b>	0.13		
FOOD	<b>0.93</b>	0.29		
CAD VENUES	<b>0.91</b>	0.35		
CAD RELIGION	0.77	0.59	0.12	
CAD PUBLIC	<b>0.88</b>	0.39	0.10	
CAD OFFICE	0.75	0.56	0.13	
CAD HOTEL	<b>0.83</b>	0.51		
CAD COMMERCE	0.70	0.68		
TRANSPORT METRO	0.54	<b>0.74</b>	0.15	
SPORT	0.18	<b>0.86</b>	0.10	
PARK	0.46	0.78		
HEALTH	0.64	0.70		
EDUCATION	0.52	<b>0.80</b>	0.16	
CAD SCHOMCO	0.57	0.77	0.15	
CAD RESIDENTIAL	0.58	0.76	0.13	
CAD INDUSTRY		<b>0.85</b>		
ROUTING	0.11	0.12	<b>0.98</b>	
COMPLEXITY				
CAD SPORT		0.17		<b>0.97</b>
TRANSPORT TRAIN	0.35	0.28		
ROUTING HIGHWAY	0.21	0.25		

Cargas de los componentes principales de modo de desplazamiento a pie

## Las variables de accesibilidad resultan interpretables:

- El primer componente principal COMP\_WAK\_1 se refiere áreas con un alto grado de servicios de ocio, áreas comerciales y bien conectadas con el transporte público y con una alta existencia de apartamentos de vacaciones.
- El segundo componente destaca el anillo exterior inmediato del centro de la ciudad, áreas urbanas residenciales prósperas desde un punto de vista urbano.
- El tercero no menos interpretable, y destaca áreas específicas de la ciudad que presentan comportamientos diferentes a los generales dentro de sus distritos. Por ejemplo, se identifican valores altos en calles importantes como la calle principal de Madrid (Gran Vía) y áreas sujetas a fenómenos de gentrificación en el centro de la ciudad.

# Esquema hedónico de precio mercado (oferta-transacción)



## Modelos hedónicos instrumentales

1) Se construye un hedónico para **imputar a cada observación** el precio de la transacción de alquiler usando el microdato de la EPF con los atributos comunes.

2) Se construye un hedónico para **imputar a cada observación** el precio del precio en oferta (Idealista) usando los mismos atributos que en 1).

Ambos modelos controlan por las variables del modelo de correspondencia de precios de mercado



## Modelo hedónico de mercado

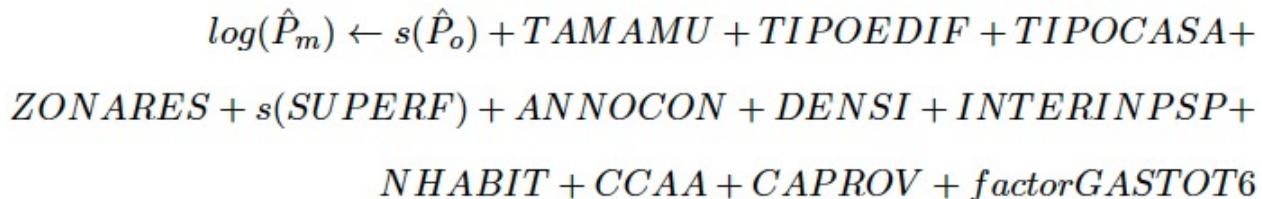
3) Se construye el **modelo hedónico de mercado** que estima el precio de alquiler de mercado (transacción) en base a los atributos comunes y el precio de oferta.

## Limitaciones del modelo de mercado

El modelo de mercado controla por un número limitado de variables de la vivienda, en particular lo que sucede es que distintos niveles de precio de oferta se corresponden a un mismo nivel de mercado (debemos tomar en cuenta que esta magnitud es una media). Podemos atribuirlo a dos cuestiones: (1) varias realidades de precios marcadas por factores omitidos no se están controlando y la estimación tiende a la media (a pesar de controlar por variables de ingresos, gastos o si se encuentran en la capital). (2) La segunda cuestión es que el nivel de soporte en la EPF es bajo lo que puede dar lugar a inestabilidad temporal en la estimación.

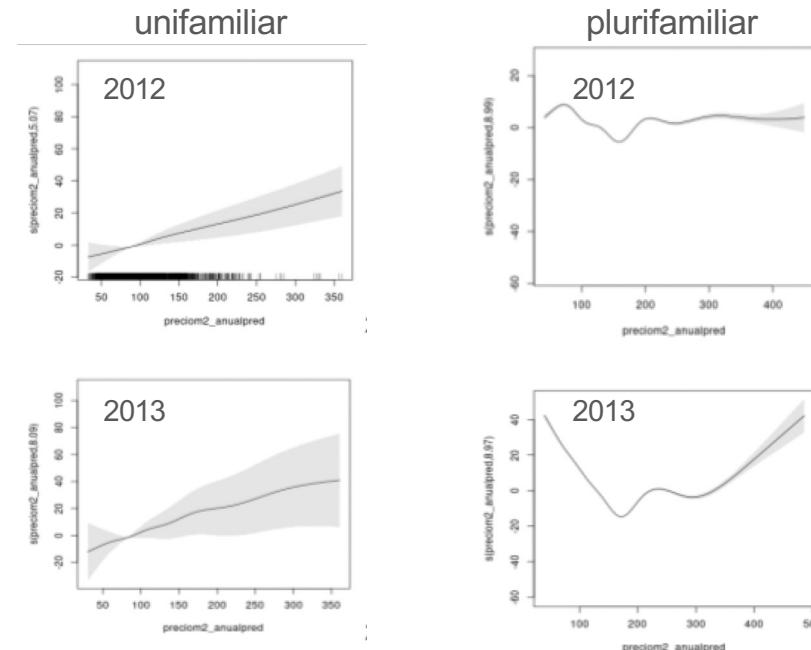
Precio mercado  
(imputado)

Precio de oferta  
(imputado)


$$\log(\hat{P}_m) \leftarrow s(\hat{P}_o) + TAMAMU + TIPOEDIF + TIPOCASA + ZONARES + s(SUPERF) + ANNOCON + DENSI + INTERINPSP + NHABIT + CCAA + CAPROV + factorGASTOT6$$

## Limitaciones del ajuste del modelo de mercado

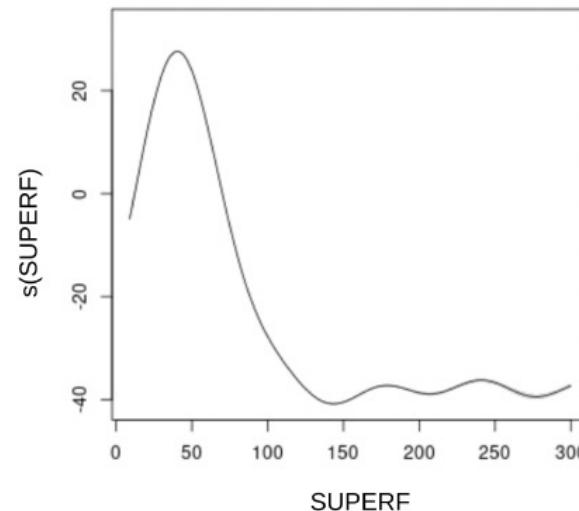
A pesar del nivel de ajuste el modelo tiene limitaciones. Por ejemplo, en vivienda unifamiliar la variabilidad de los errores del modelo es alta a partir de 300 m<sup>2</sup>, debido a que la superficie en la EPF está limitada a 300 metros cuadrados.



**Figura 7.** Función de suavizado (s) del modelo GAM sobre la covariante precio de oferta (unifamiliar versus plurifamiliar). Fuente: elaboración propia

## ¿Por qué usar un GAM?

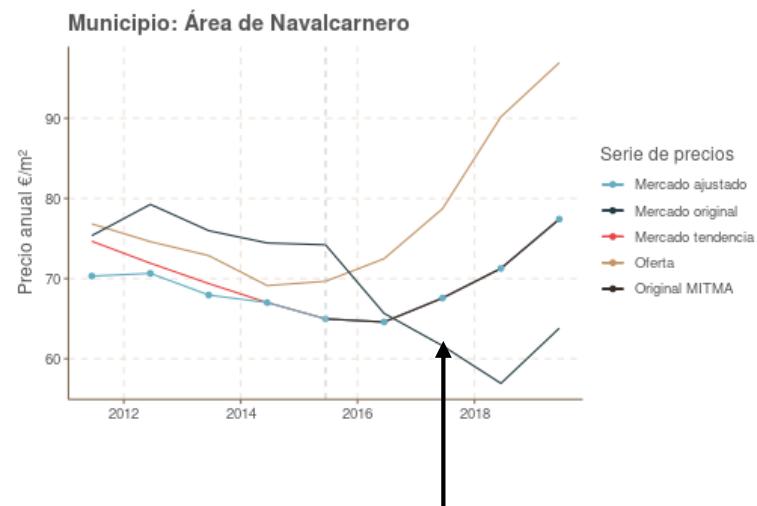
**Figura 3.12.** Valores de la función de suavizado ( $s$ ) para la variable superficie útil en modelo de correspondencia



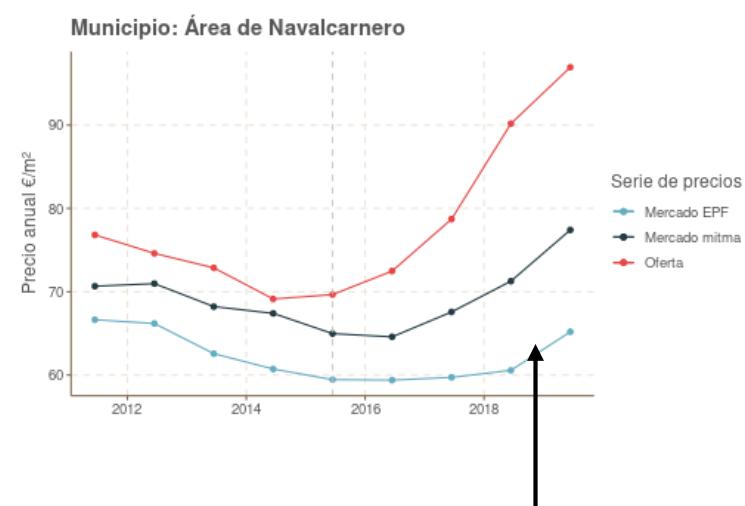
Se observa como para superficies pequeñas el coeficiente es positivo hasta llegar a 50 m. aproximadamente. Entre 50 m. y 80 m., la contribución es positiva pero decreciente, mientras que a partir de 80 m., la contribución al precio es negativa. Para valores superiores la relación es decreciente, hasta que los 140 m., donde los incrementos en la superficie no implican incrementos significativos en el precio.

## Eliminación del sesgo zonal (hedónico final)

El dato de mercado estimado a través del resultado del modelo de oferta incorpora de forma efectiva el precio del suelo, pero debido a que los datos de la EPF no disponen de detalle suficiente de la ubicación. Por consiguiente, los precios de mercado por zona ofrecen datos inconsistentes. Para corregir este problema utilizaremos una serie que ayude a calibrar la forma de las series en algunos niveles de agregación (usamos series de MITMA).

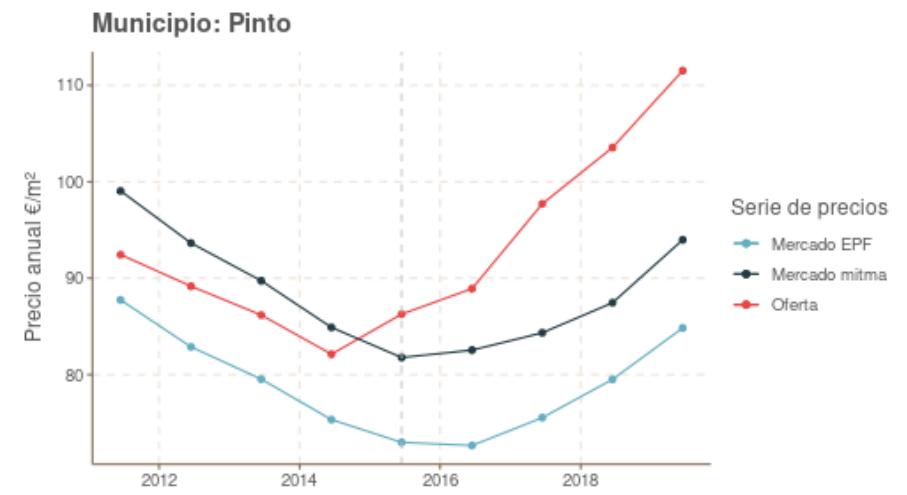
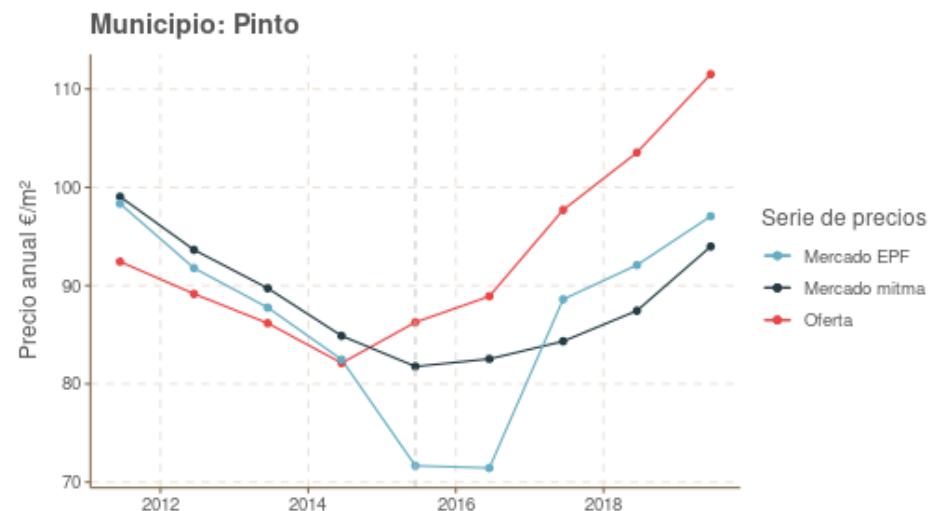


Anomalía en la serie original del modelo de mercado



Serie ajustada por el modelo final

## Coherencia estructural del modelo hedónico final



## Coherencia estructural del modelo hedónico final

El proceso de reajuste del modelo de mercado mantiene las métricas generales del modelo estructural, una vez corregidas las incoherencias con la serie de referencia del Ministerio.

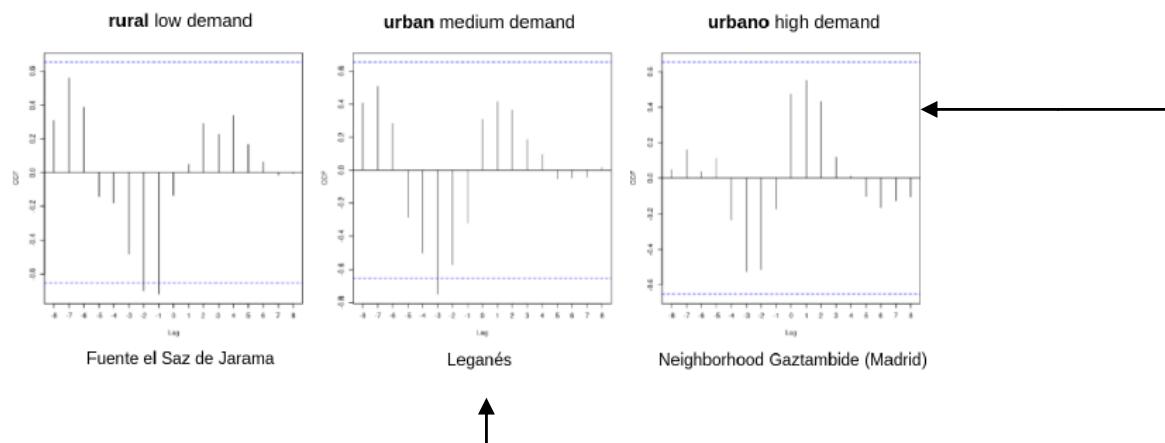
**Tabla 6.2.** Comparativa de agregados de series antes y después del ajuste zonal

Capital	Tipo	Paso	Media	Desv	Min	p25	p50	p75	Máx.
No	Plurifamiliar	Ajustado	80.73	4.44	75.59	76.91	79.97	83.54	88.94
		Original	81.43	4.33	76.97	78.16	79.96	84.72	89.41
	Unifamiliar	Ajustado	57.27	2.95	53.27	55.34	56.23	58.41	63.27
		Original	54.40	5.13	49.99	50.67	52.60	55.84	65.40
Sí	Plurifamiliar	Ajustado	102.08	5.13	96.46	97.01	100.99	105.03	110.90
		Original	102.02	5.90	92.58	98.03	104.70	105.84	109.04
	Unifamiliar	Ajustado	61.12	9.62	49.88	56.12	57.27	67.94	78.54
		Original	61.12	9.62	49.88	56.12	57.27	67.94	78.54

## Validez de uso de serie de oferta como indicadora

Se observa la alta correlación temporal cruzada entre las series de oferta y de alquiler, lo que demuestra la validez de la hipótesis de usar como serie de oferta como “indicadora” para realizar la desagregación. Además se observa que la naturaleza de cada zona condiciona el comportamiento de la serie.

**Figura 7.5.** Correlación cruzada de series mensuales de oferta y alquiler: Zonas  
Fuente del Saz, Barrio Gaztambide y Leganés



En general la correlación cruzada oferta – alquiler indica la correlación con *lag* entre ambas series

El precio del alquiler en zonas más activas responden más rápidamente ante cambios en los precios de oferta

## Calidad de las series desagregadas según MLE

La selección por verosimilitud es mejor que la selección arbitraria o aleatoria para todas las series (según se observa en valores absolutos y verosimilitud de cumplimiento de los criterios de calidad). Esto confirma la validez del método y confirma la necesidad de adaptar el método de desagregación en función de la serie.

Los modelos autorregresivos seleccionados logran altos niveles de ajuste, un  $R^2$  del 0,87 alquiler y 0,94 (oferta).

**Tabla 7.8.** Comparativa de probabilidades en series seleccionadas por máxima verosimilitud contra todos los métodos y una selección aleatoria

Método	Prob. VF	Prob. VM	Prob. POS	Prob. final
Máxima verosimilitud	37.65%	20.91%	100.00%	8.88%
Denton Cholette	32.38%	20.41%	100.00%	7.57%
Litterman Maxlog	30.01%	19.92%	100.00%	6.94%
Dynamic Maxlog	27.54%	15.54%	100.00%	6.45%
Aleatorio	24.94%	16.79%	100.00%	5.64%
Chow Lin Maxlog	23.40%	17.45%	100.00%	5.33%
Causey Tragger	11.36%	10.61%	100.00%	1.92%

Fuente: elaboración propia

Resumen de criterios: **VF**: variación frontera. **VM**: variación mensual, **POS**: positividad

## ¿Cómo se selecciona un método de desagregación?

Existen distintos métodos de desagregación temporal, lo ideal es que el método que se use maximice la calidad de las series generadas, por tanto el método debe maximizar la verosimilitud del resultado óptimo según 5 criterios de validación (Chen y Andrews, 2008):

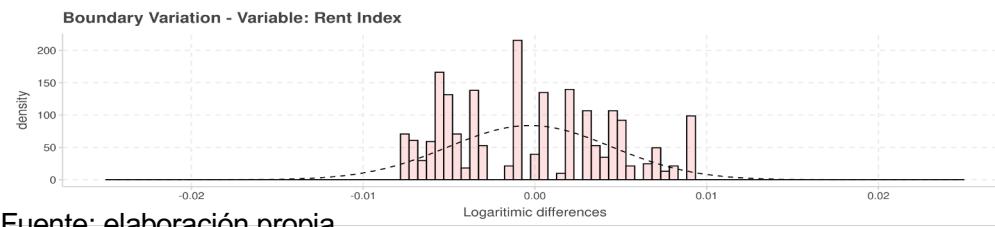
- **Variación frontera:** diferencia logarítmica del índice entre diciembre y enero.
- **Distribución de las diferencias mensuales** logarítmicas.
- **Ajuste de la media anual de la serie** generada con el valor anual (RMSPE).
- **Positividad de la serie** (en este caso no se deben permitir precios negativos).
- **Ausencia de variaciones extremas a principio y fin** de las series.

$$\log \hat{\theta}(\hat{H}) = \log \prod_{i=1}^n f(X_i|\theta) = \sum_{i=1}^n \log f(X_i|\theta)$$

La verosimilitud final se calcula combinando las probabilidades de cada criterio **m**, aproximando la distribución de valores de cada uno de los criterios como una distribución normal.

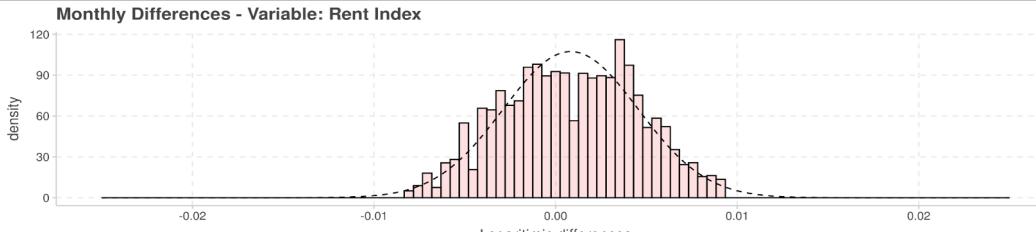
La verosimilitud se estima paramétricamente en base a la distribución “ideal” de valores basado en la estimación de los parámetros de las distribuciones de probabilidad de las diferencias logarítmicas, sin embargo las distribuciones de referencia no siguen fielmente una distribución normal. Aunque los criterios de selección funcionan de la forma esperada es posible que una aproximación no paramétrica para estimar la verosimilitud podría ofrecer un mejor funcionamiento del algoritmo.

**Figura 6.** Distribución de variación frontera (oferta)



Fuente: elaboración propia

**Figura 6.** Distribución de variación frontera (oferta)

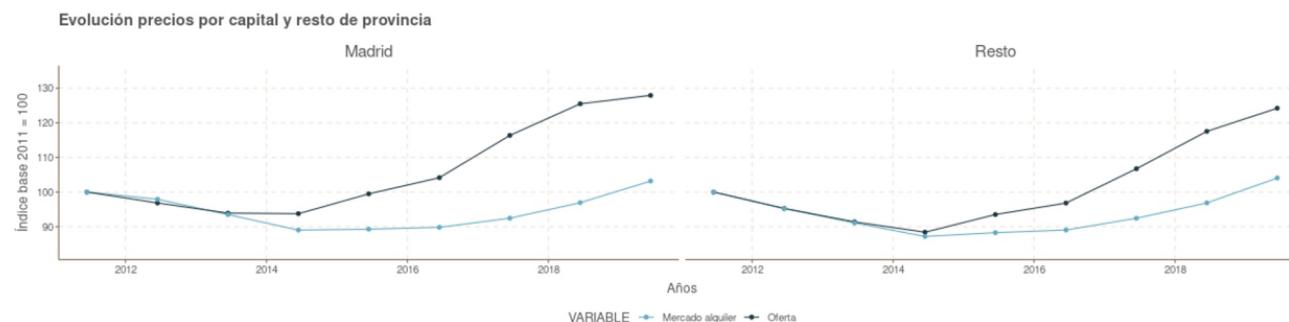


Fuente: elaboración propia

## Consistencia de los índices de oferta y alquiler

La ventaja de la metodología es que permite realizar un análisis conjunto de la evolución de mercado y de oferta

**Figura 8.4.** Comparativa evolución del precio de oferta y de alquiler, desglose por capital y resto de provincia

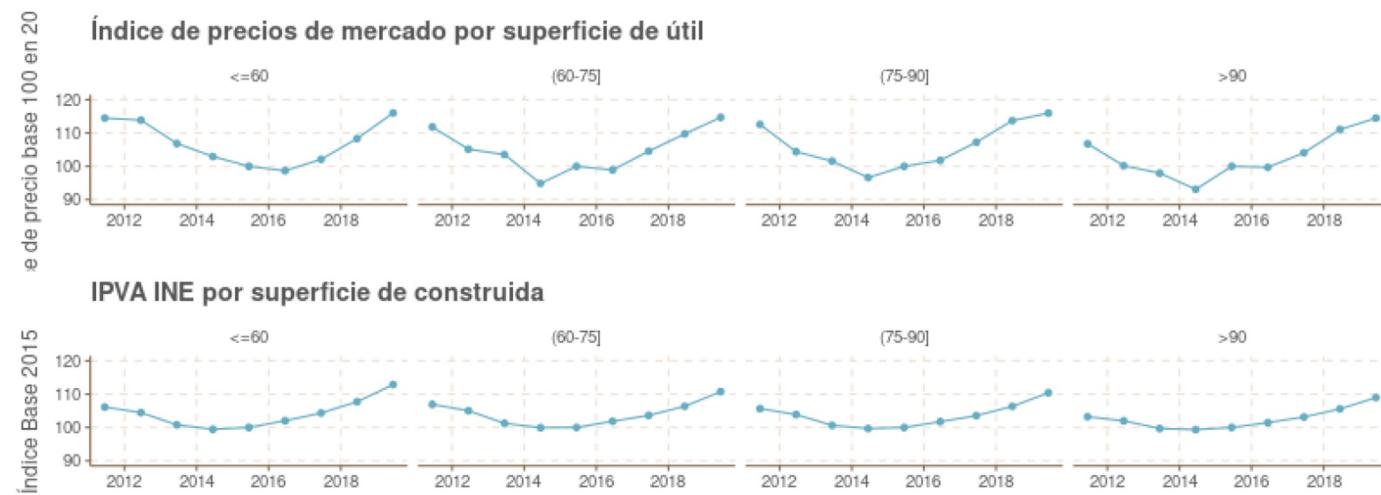


*Fuente:* elaboración propia.

## Comparativa con IPVA del INE

Nuestro índice (agregado) es coherente con el desarrollado por el índice experimental desarrollado por el ministerio de vivienda. Existen diferencias que atribuibles a la diferencia del colectivo de análisis en cada caso.

**Figura 8.38.** Comparativa indice de mercado por superficie útil e IPVA por superficie construida



*Fuente:* elaboración propia.

# Conclusiones principales

**Los portales de internet son una fuente adecuada** como variable para la estimación de los precios mercado. Sin embargo, adolecen de problemas de calidad de la información que requieren un tratamiento precio de la información.

La correspondencia estadística con calibración es **eficaz para replicar la estructura poblacional original** de la muestra a través de los elevadores muestrales y mitigar el efecto de los sesgos. Aunque está limita cuando existe una cobertura de datos parcial

El uso de **estadística oficial** en el modelado asegura la coherencia entre las magnitudes macro y micro. Aunque la estabilidad de los resultados está **expuesta a los cambios metodológicos de las primeras**.

**Los modelos hedónicos son eficaces** para recoger **contribuciones complejas** en el precio. La vivienda tiene características que requieren un tratamiento particular como la **heterogeneidad de espacial**

Los modelos basados en **aprendizaje estadístico** son eficaces para resolver los problemas habituales de los modelos paramétricos de regresión lineal, con una menor dificultad en la especificación.

El **ensamblado de un gran número de modelos** ofrece un alto nivel de **ajuste**, pero tiene como **inconvenientes** el cálculo de una medida general de ajuste y la gestión operativa su mantenimiento.

Se ha demostrado empíricamente que es posible especificar de forma eficiente la **contribución espacial** en función de variables que representan la **utilidad** asociada a la ubicación

La **coherencia temporal** se ve limitada cuando se desarrollan **modelos anuales**. Los mecanismos de conciliación y desagregación temporal solventan la cuestión, pero deberían incorporarse las restricciones espaciales en el diseño

Un **modelo dual oferta-mercado** con un alto nivel de detalle permite un análisis más exhaustivo del mercado e información extra: tensiones en oferta y demanda, poder de negociación o tiempo en mercado

# Conclusiones principales I

- Los portales de internet son una fuente adecuada como variable para la estimación de los precios mercado. Sin embargo, adolecen de problemas de calidad de la información que requieren un tratamiento precio de la información.
- La correspondencia estadística, a través de calibración, es eficaz 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, de composición, variables omitidas, de infrarrepresentación y sobrerepresentación, habituales en los datos procedentes de portales inmobiliarios de Internet.
- La calibración presenta limitaciones que afectan a su eficacia: estabilidad temporal, insuficiente muestra, representatividad del precio en viviendas complejas (unifamiliar), granularidad de la información de los atributos comunes (en particular la geográfica).
- El uso de estadística oficial en el modelado asegura la coherencia entre las magnitudes macro y micro. Aunque es necesario considerar el efecto de los cambios metodológicos en las fuentes oficiales, para asegurar la estabilidad temporal de las medidas.

## Conclusiones principales II

- Los modelos hedónicos son eficaces para recoger contribuciones complejas en el precio. La vivienda tiene una serie de características que requieren un tratamiento particular, especialmente el introducido por la heterogeneidad de espacial en los efectos
- Los modelos basados en aprendizaje estadístico son eficaces para resolver los problemas habituales de los modelos paramétricos de regresión lineal, con una menor dificultad en la especificación.
- El ensamblado de un gran número de modelos ofrece un alto nivel individual de ajuste y un mejor tratamiento para cada estrato, sin embargo, tiene como inconvenientes el cálculo de una medida general de ajuste y la gestión operativa su mantenimiento.
- Se ha demostrado empíricamente que es posible especificar de forma eficiente la contribución espacial en función de variables que representan la utilidad asociada a la ubicación

## Conclusiones principales III

- La coherencia temporal se ve limitada cuando se desarrollan modelos anuales. Los mecanismos de conciliación y desagregación temporal solventan la cuestión, pero deberían incorporarse las restricciones espaciales y temporales en el propio diseño de los modelos
- A pesar de los altos niveles de ajuste, el encadenamiento y la doble imputación hedónica no solventan los problemas de inestabilidad en los precios ante estratos con muy poco soporte.
- El modelado dual oferta y mercado con un alto nivel de detalle permite un análisis más exhaustivo del mercado. Ofreciendo información de las tensiones en oferta y demanda, así como la estimación de los poderes de negociación de los agentes.

## Resumen final

El modelo desarrollado aporta una metodología totalmente novedosa para construir un índice de precios del alquiler que soluciona gran parte de los retos de las aproximaciones tradicionales.

Por una parte, aplica nuevas fuentes de información, como los portales inmobiliarios de internet y el aprendizaje automático para conseguir una función más precisa para especificar las funciones de precio y enlace entre oferta y mercado. Adicionalmente, se ha profundizado en propuestas innovadoras en cuestiones de modelado espacial sobre el fundamento de la utilidad de la ubicación, y la desagregación temporal.

Con dicha aportación, se espera que la presente tesis pueda servir de referencia para futuras investigaciones en el avance del estado del arte en modelado hedónico de la vivienda con de técnicas no tradicionales.