

# Tesis Doctoral

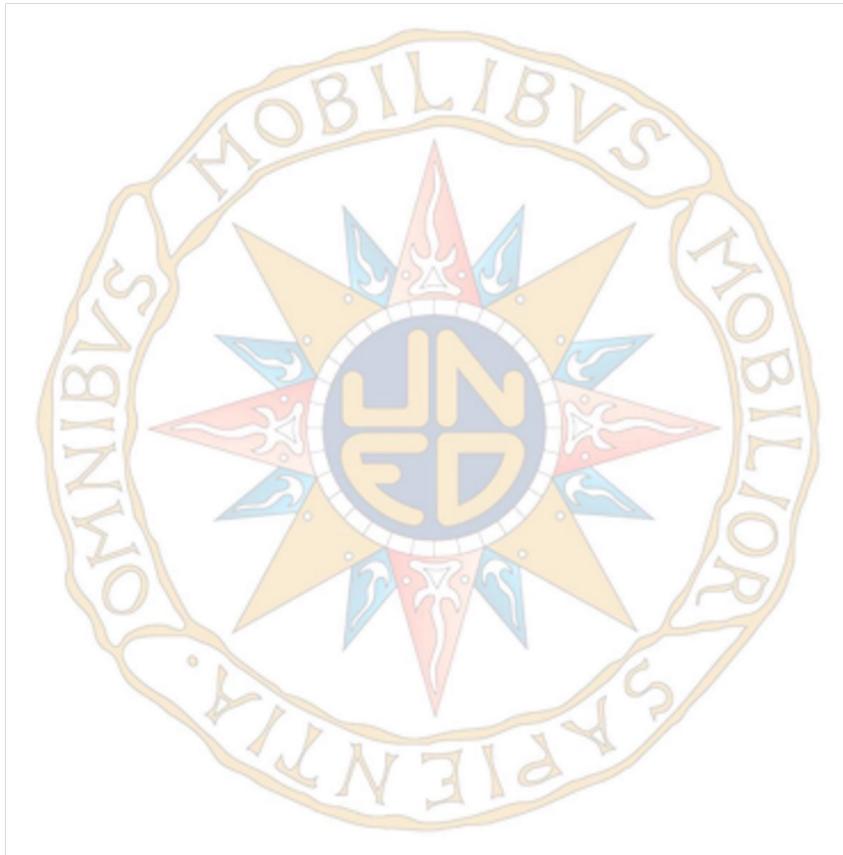
**Metodología para la construcción Índices de Precios de la  
Vivienda robustos, segmentados y a escala basados en modelos  
de valoración hedónica**

**AUTOR: David Rey Blanco**



**PROGRAMA DE DOCTORADO EN ECONOMÍA Y EMPRESA**

**DIRECTOR: Dr. Julio González Arias**



# Introducción

FT Series Renting property + Add to myFT

## Extreme renting: how rising rates turned the screws on tenants across Europe

Heated market in housing hotspots is compounded by increasing mortgage costs deterring buyers

Akila Quinoio AUGUST 22 2023

516

**Unlock the Editor's Digest for free**

Roula Khalaf, Editor of the FT, selects her favourite stories in this weekly newsletter.

Enter your email address  Sign up

Marianne B and her partner had owned their Paris flat for seven years when

Fuente: Financial Times

## Indices de precios usando métodos alternativas 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 (Behr, 2023; 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.
- La información con retraso puede desequilibrar el mercado (Anenberg y Laufer, 2017; Hill et al. 2023).

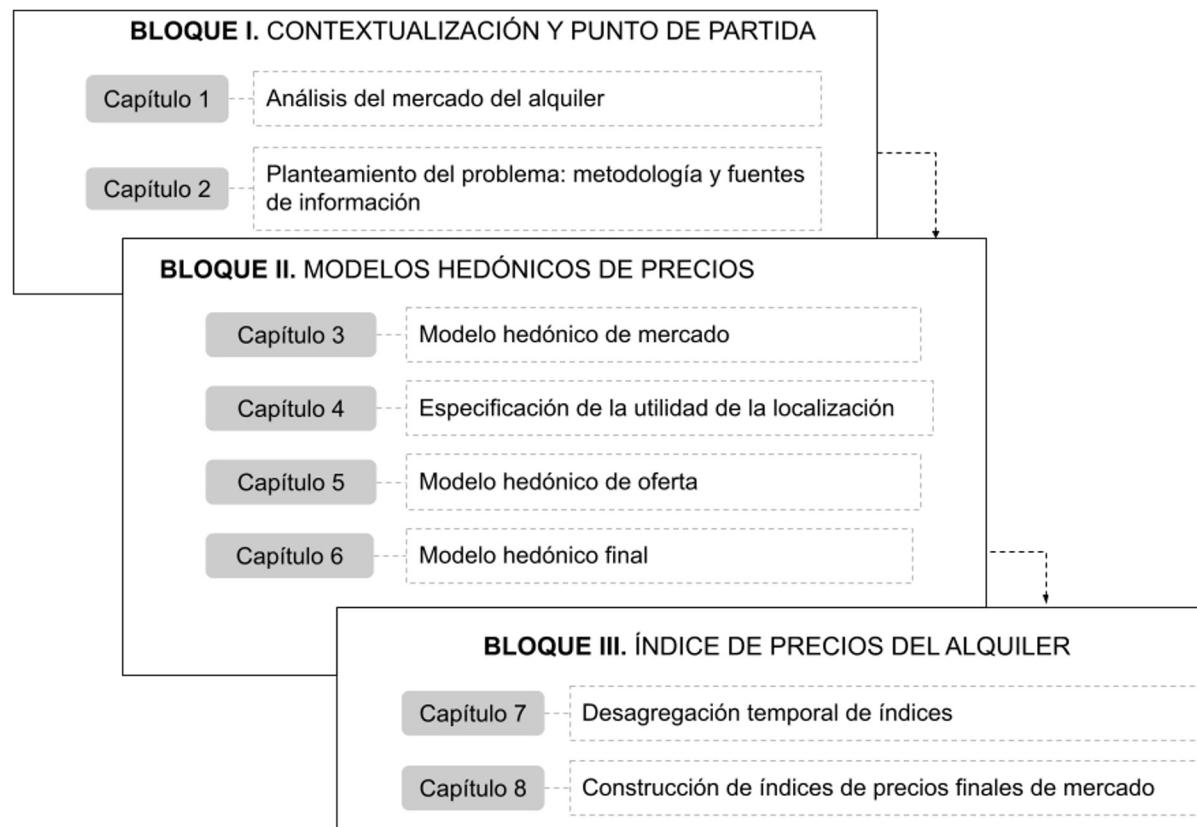
## Nuestra metodología ofrece un avance en el estado del arte

- 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 económicas, estadísticas y de las ciencias de la computación.
- Aparte del aporte académico (8 artículos aceptados / publicados), aporta una herramienta fácilmente industrializable.

# Objetivos del la metodología del índice

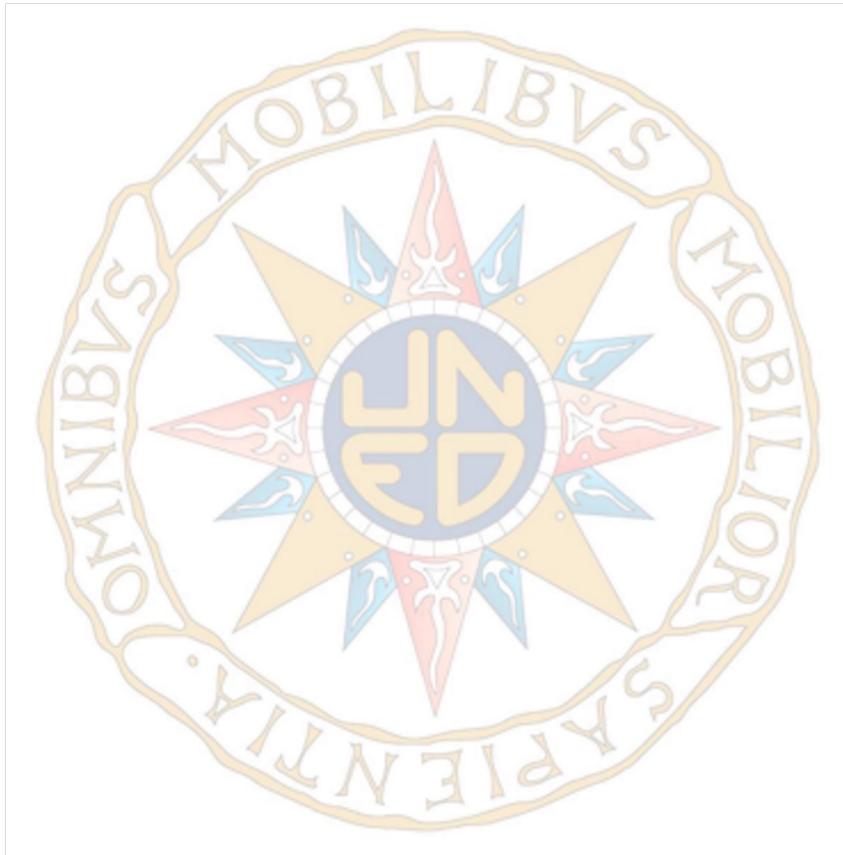
PRINCIPALES	Fiel reflejo del mercado del alquiler	Alto nivel de desagregación funcional	Alto nivel de desagregación temporal	Uso de fuentes y técnicas alternativas
SECUNDARIOS	Alineamiento con estadísticas oficiales	Segmentación geográfica	Frecuencia mensual	Uso de fuentes abiertas
	Definición del colectivo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

# 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

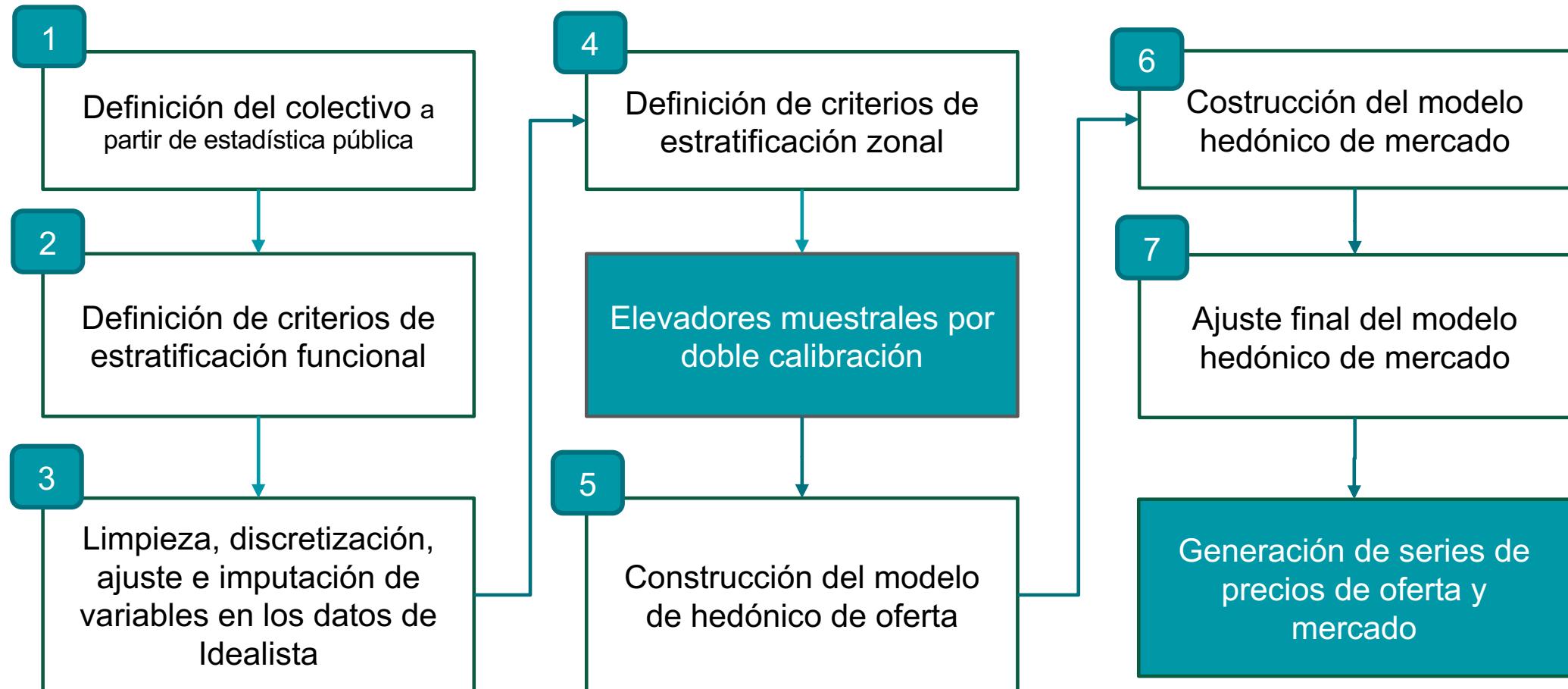
Otros

Otros datos de apoyo

# Metodología: plano de los precios



## Fases elaboración del modelo de precios

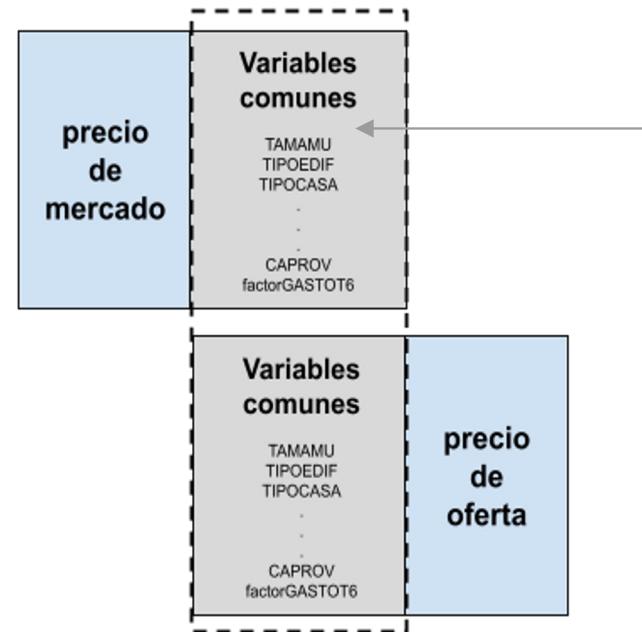


# Correspondencia de magnitudes entre poblaciones

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



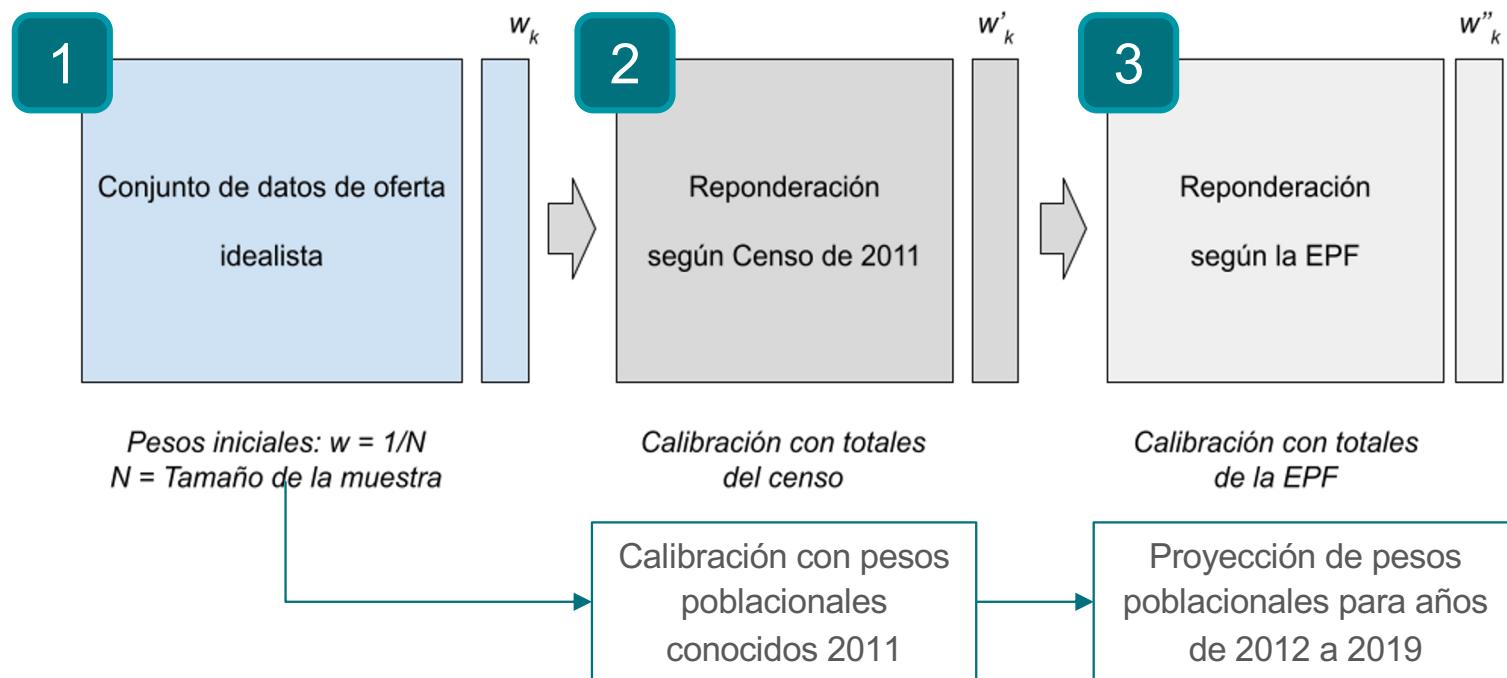
Se usan modelos que relaciona ambos  
ámbitos de información a través de  
variables auxiliares

## Muestra de origen

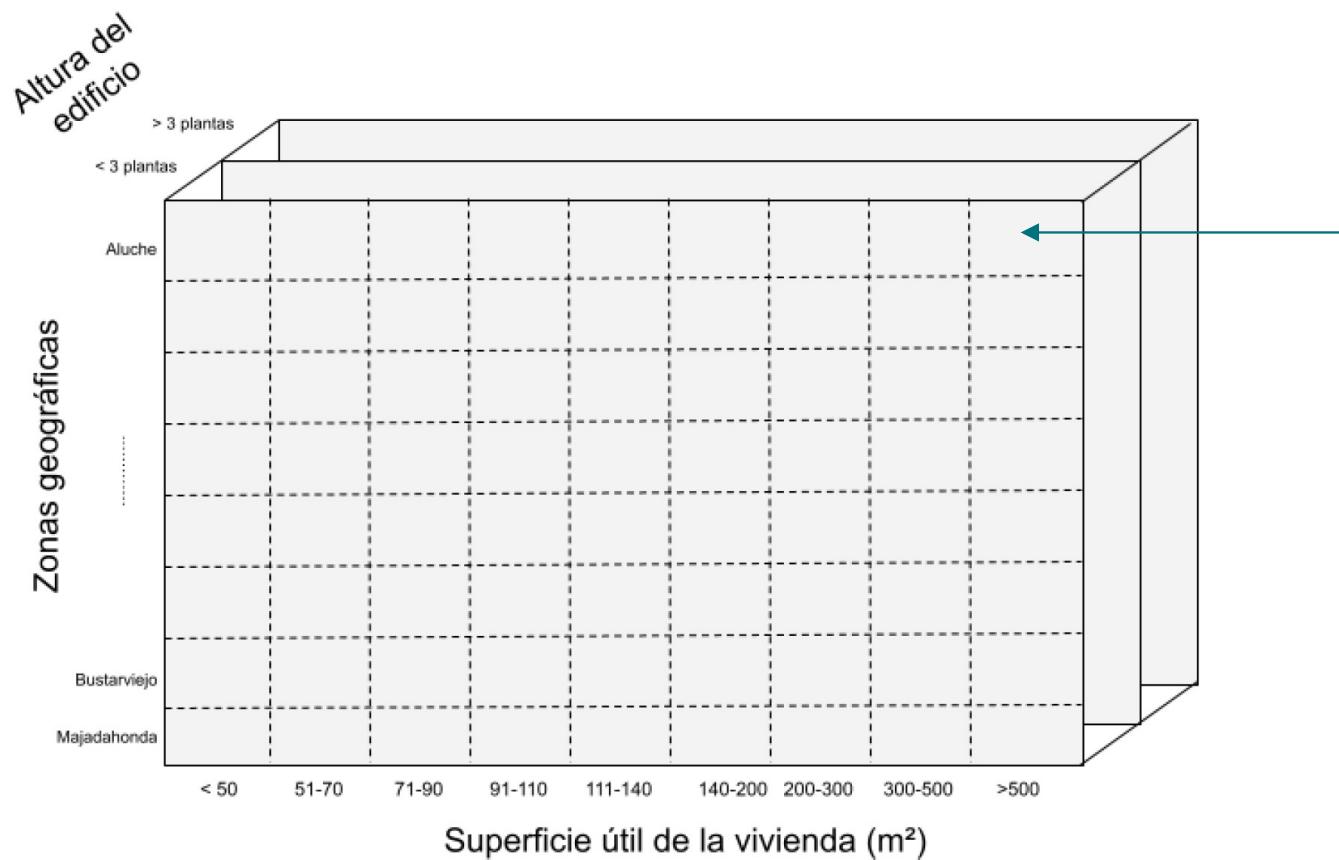
Microdatos y datos agregados  
de la oferta (Idealista)

## 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

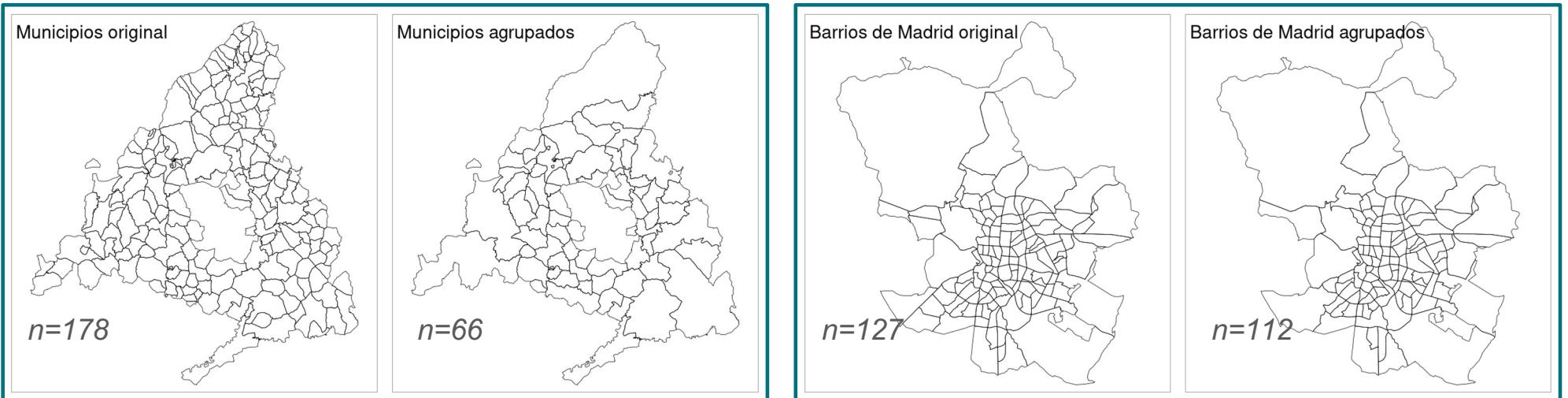
Homogeneización de variables en dos conjuntos

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)

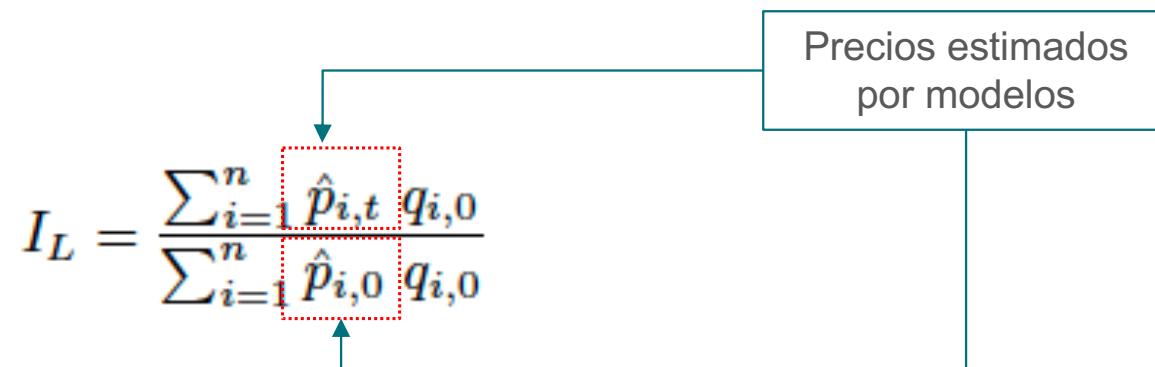
## Estratificación (2/2): atributos de tipo zonal

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 y sociodemográficas).



# 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). 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). Esta aproximación ha demostrado su superioridad ante otros métodos (Hill, R., & Trojanek, R. 2020).


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

**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 explicar los precios de la vivienda en base a la contribución de los factores que describen un bien inmobiliario, particularmente la influencia de la ubicación (Diewert y Shimizu, 2021).

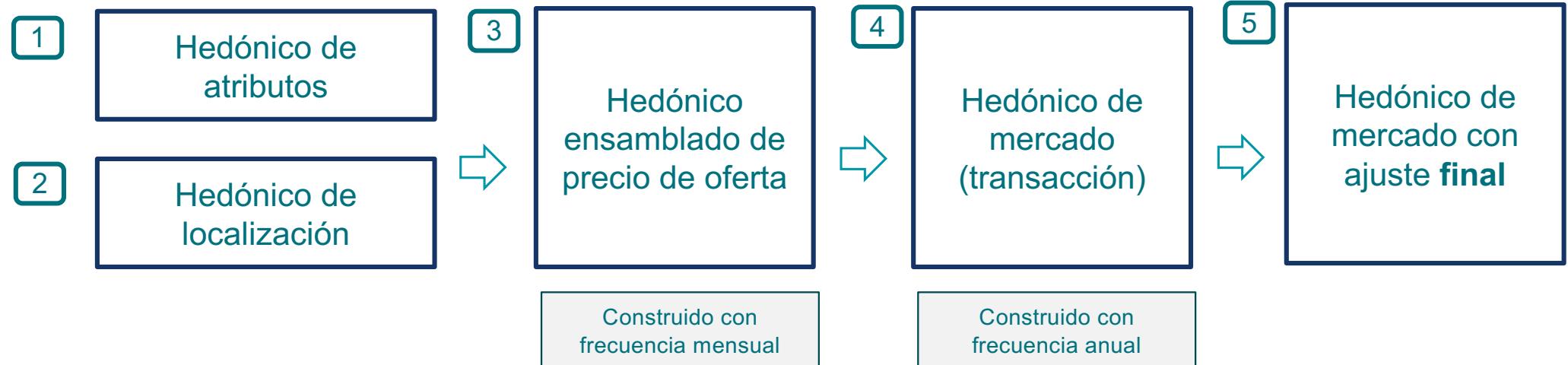
21).

Precio de la vivienda	$p^t = \beta_0 + \sum \delta \cdot D_{nk}^t + \sum \beta \cdot S_{nk}^t + \sum \gamma \cdot L_{nk}^t + \sum \mu^t M_{nk} + \epsilon \quad [1]$	Localización
	<b>Tiempo</b> <b>Estructura y Constructivas</b>	<b>Dinámicas de mercado</b>

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

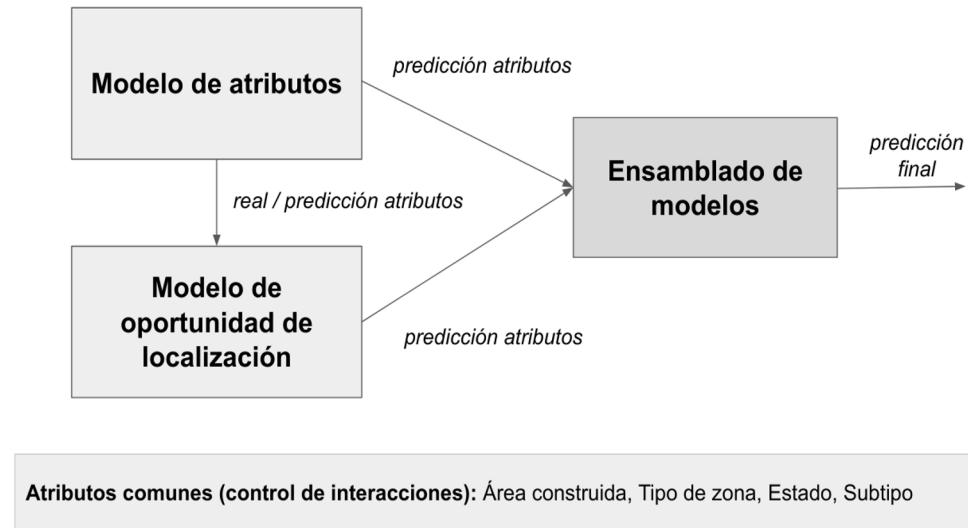
## Modelos de precio hedónicos desarrollados

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  $R^2$ ).



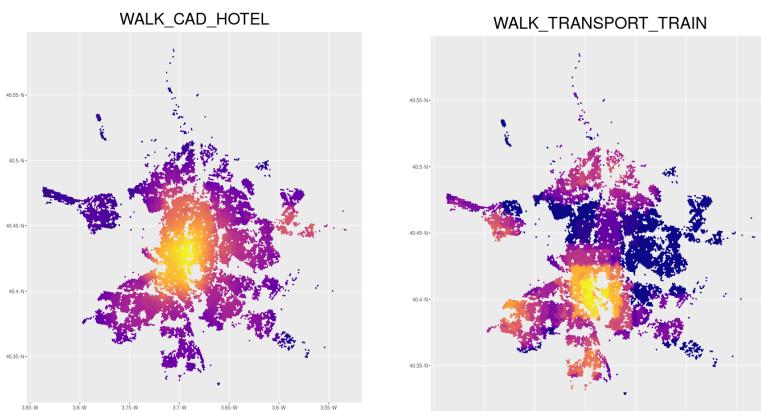
**Tabla 5.18.** Ajuste espacial promedio por clase de modelo y tipo

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

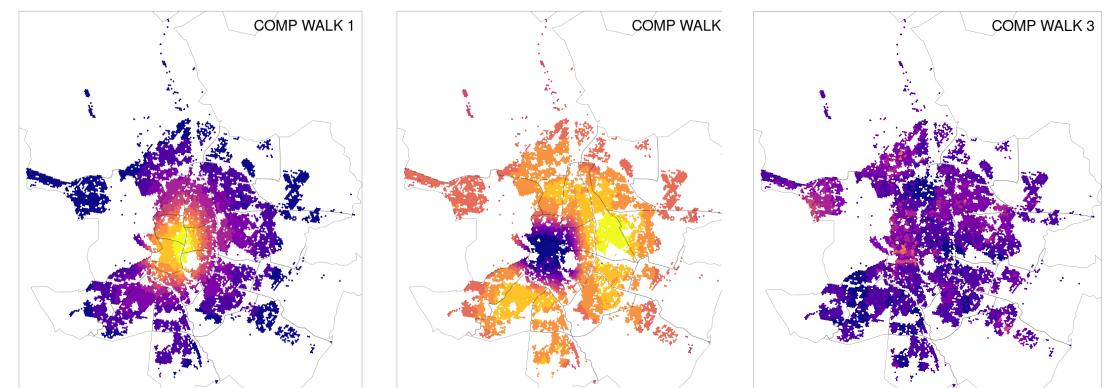
Fuente: elaboración propia

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)



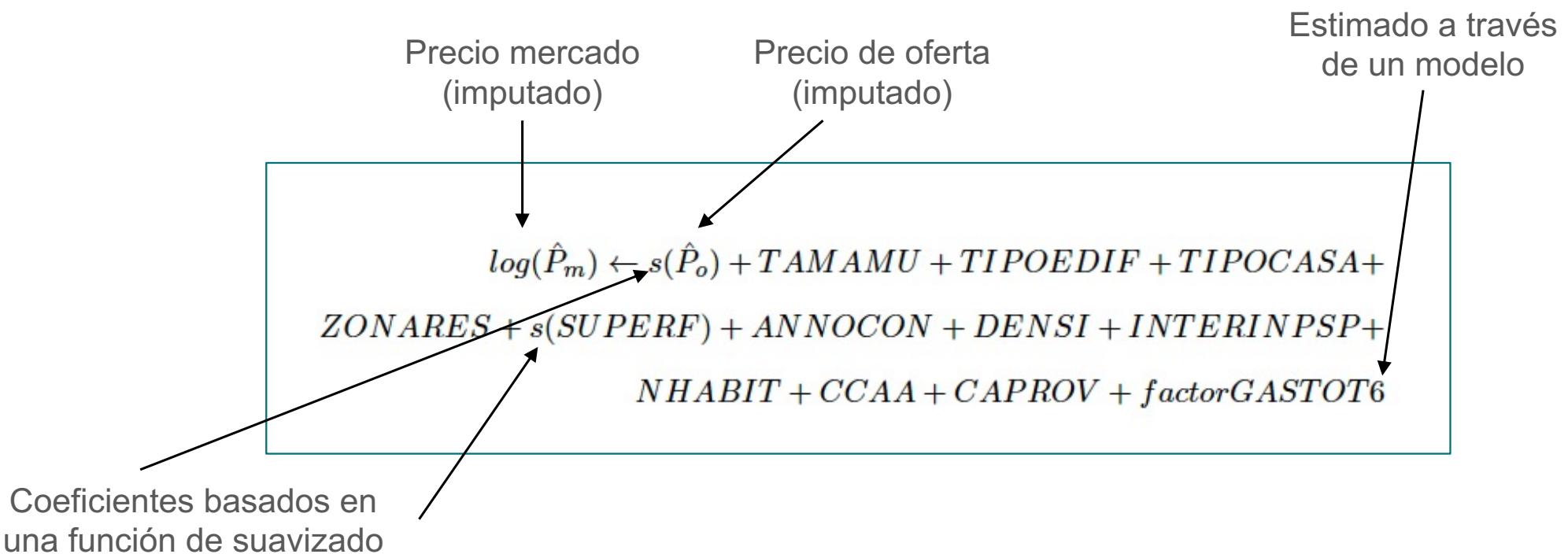
Equivalentes a usar datos geográficos

Su uso aplicado es muy eficiente

Permiten interpretar los efectos de ubicación en el precio

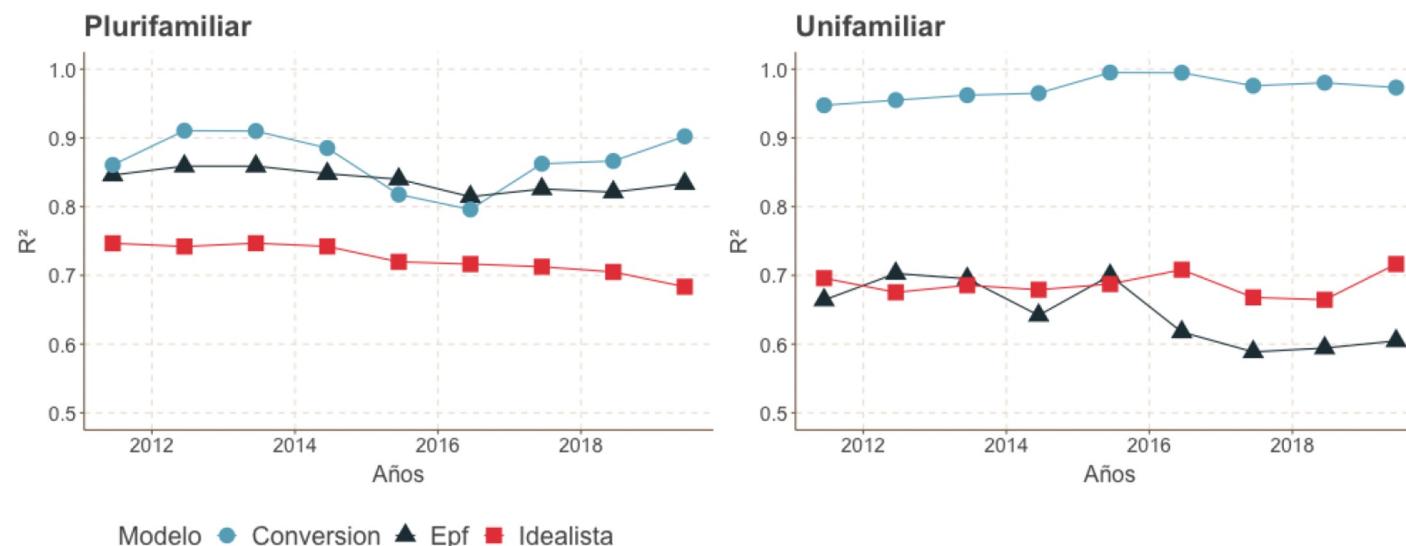
## Hedónico de precio mercado (forma funcional)

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.



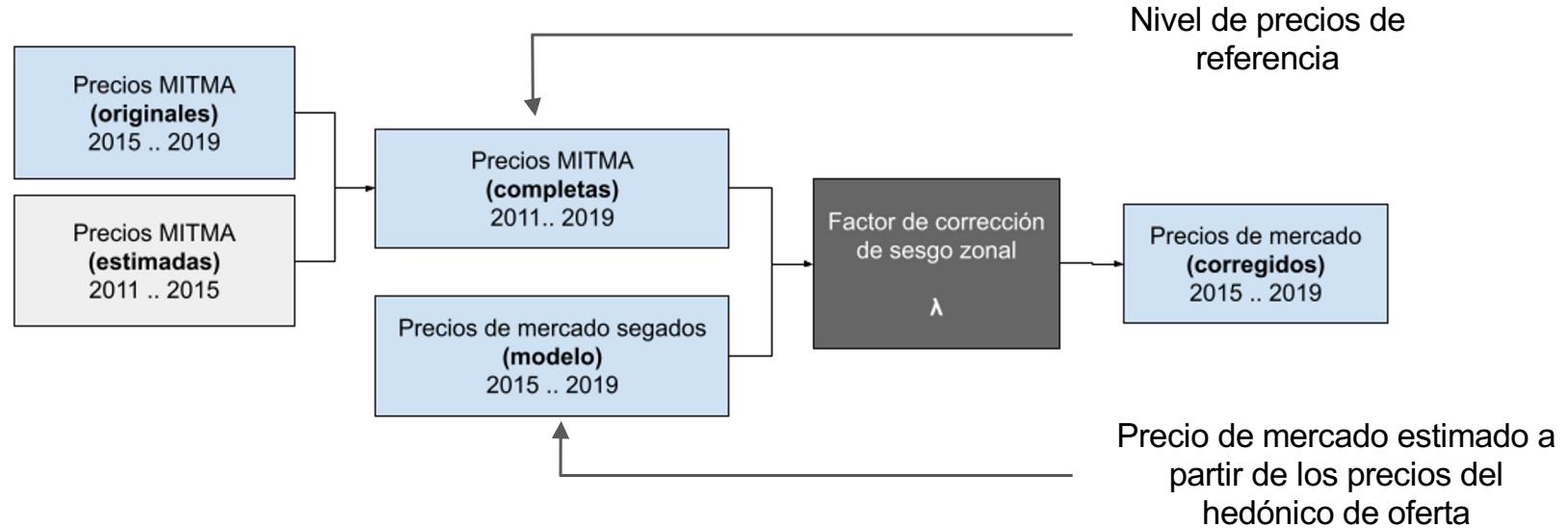
## Calidad del ajuste del modelo de mercado

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).



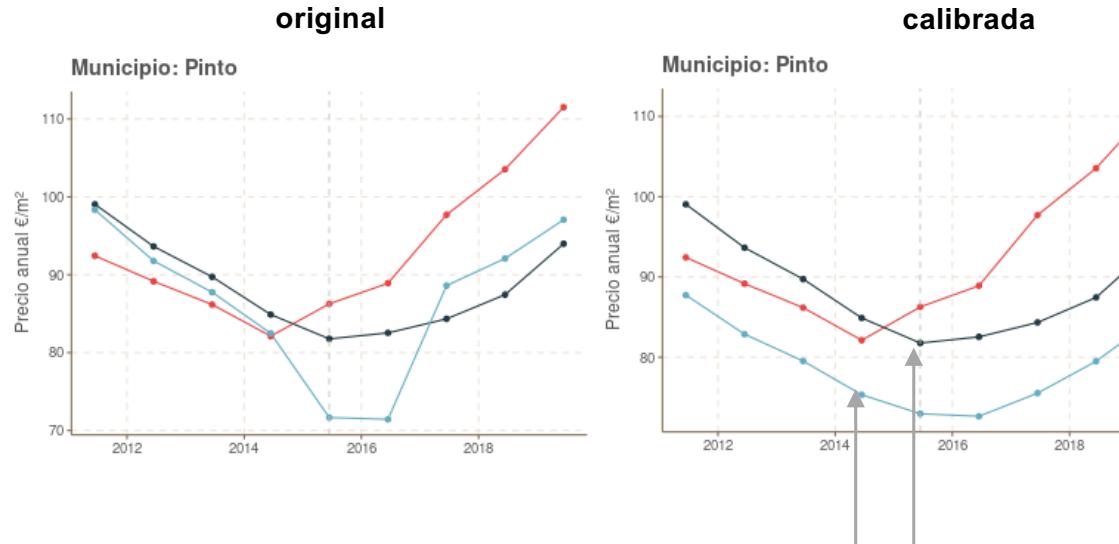
## Eliminación del sesgo zonal (hedónico final)

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 estratificada basada en ratios** (Lohr, 2019), que llamaremos 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.



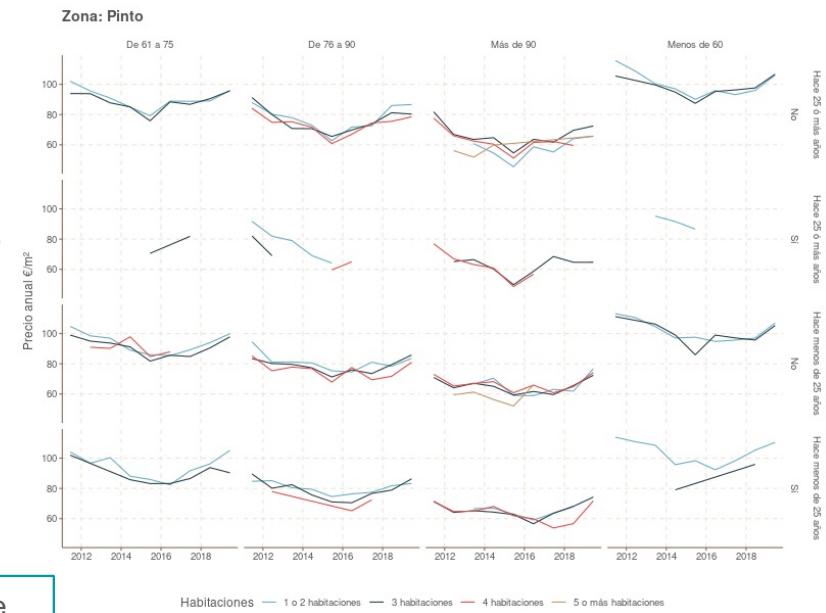
# Coherencia estructural del modelo hedónico final

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 mantiene los parámetros generales de la distribución original de precios (media, desviación típica y cuartiles de precios).

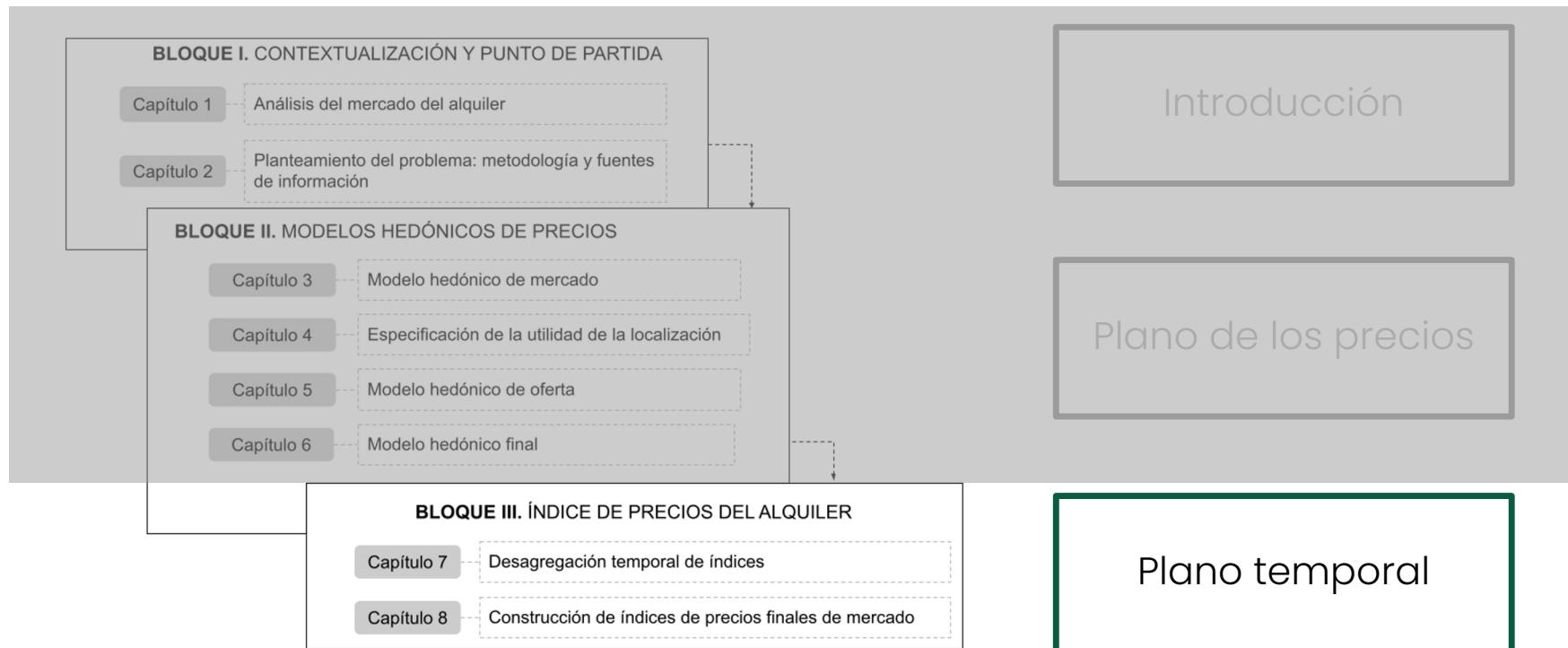


La serie corregida mantiene la escala de precios general del modelo de mercado porque las muestras de MITMA y INE/EPF no son las mismas

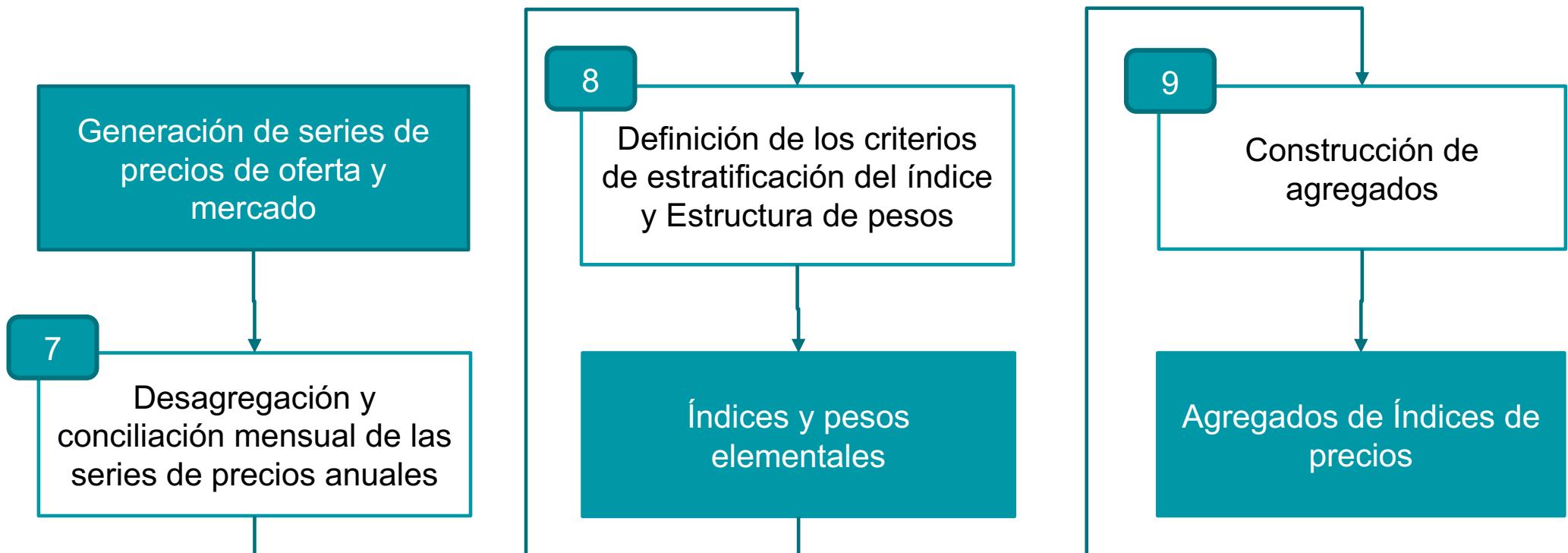
**desglose superficie y habitaciones**



# Metodología: plano temporal



## Fases elaboración del modelo de temporal



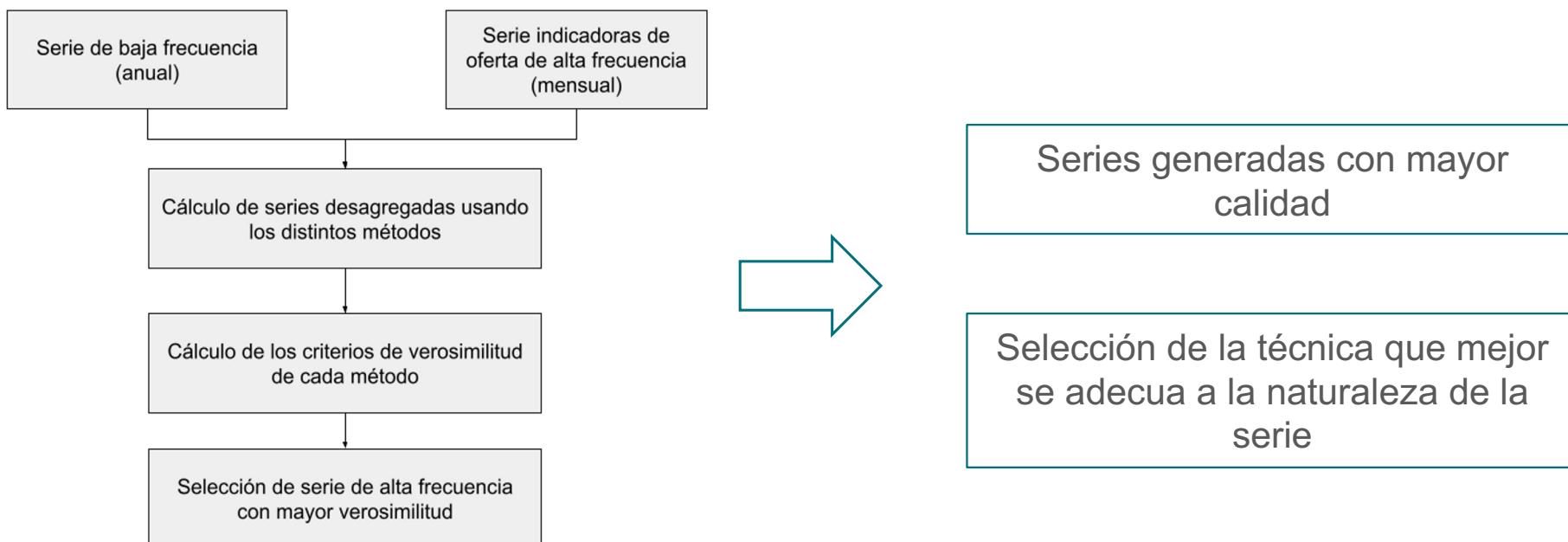
# Necesidad de desagregación temporal y reconciliación

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 periodos 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  $a$

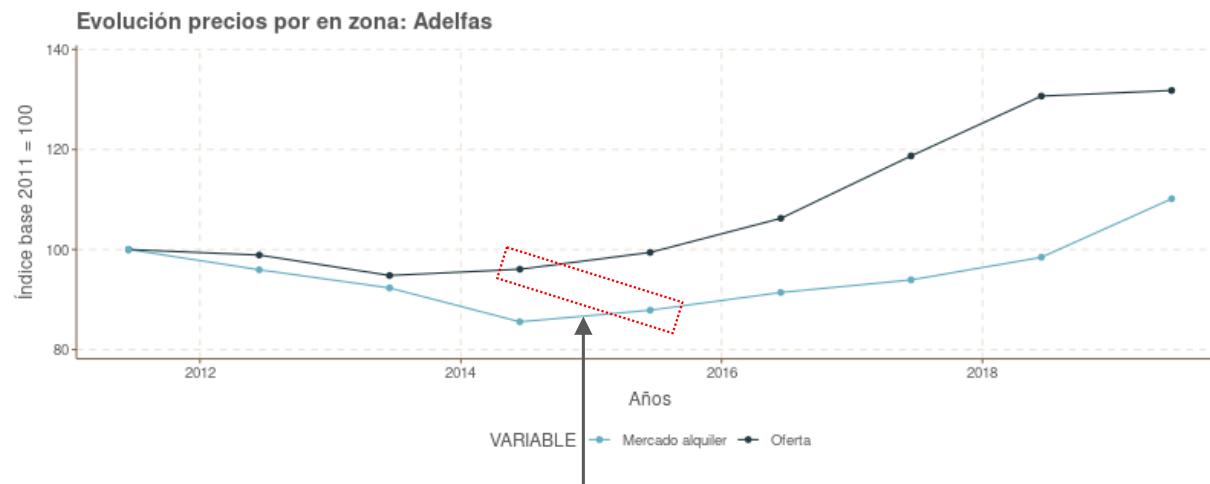
Pesos de los estratos  $e$  del agregado  $A$

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

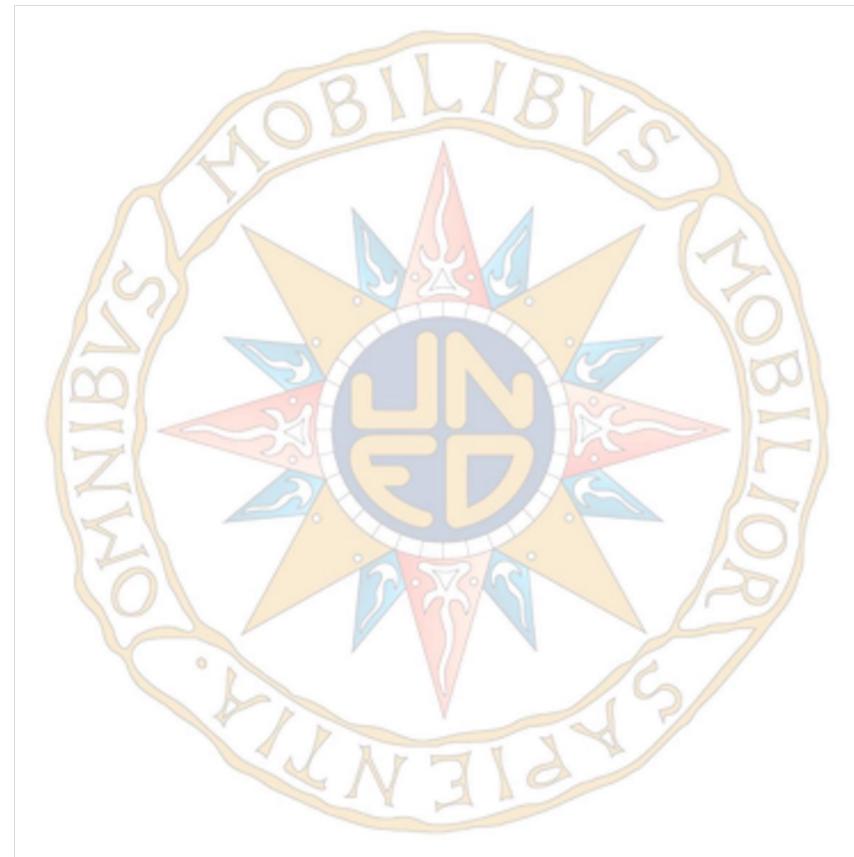
Índice para el agregado  $A$

## 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

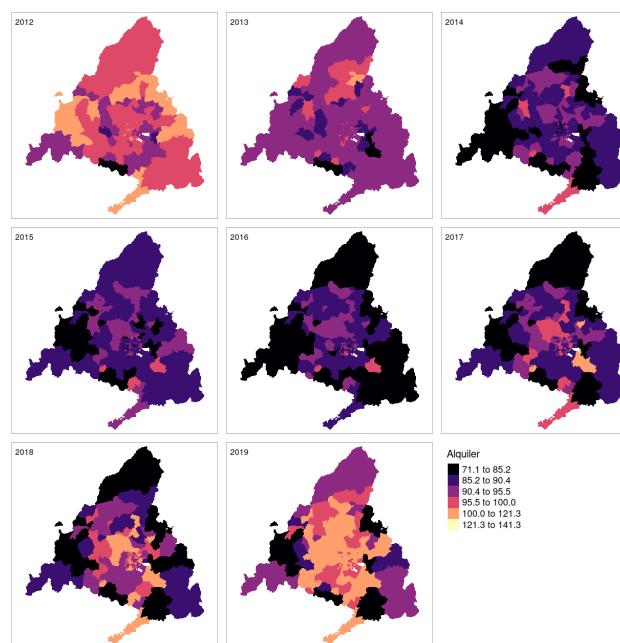
Año A	Año B								
	Año	2012	2013	2014	2015	2016	2017	2018	2019
2011	2012	0,06	0,07	0,08	0,07	0,07	0,10	0,13	0,14
2011	2013		0,01	0,02	0,01	0,02	0,05	0,07	0,08
2011	2014			0,01	0,01	0,01	0,03	0,06	0,07
2011	2015				0,01	0,04	0,07	0,07	
2011	2016					0,04	0,06	0,07	
2011	2017						0,03	0,04	
2011	2018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

Fuente: elaboración propia

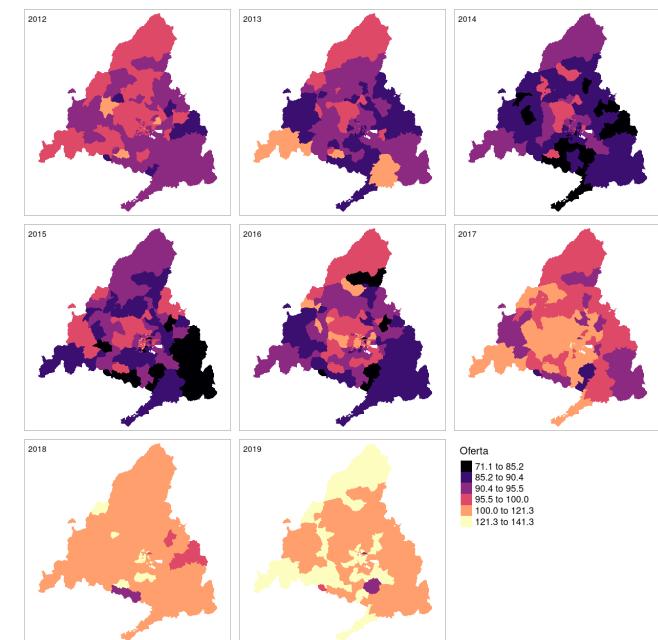
## Ventajas de un índice de precios dual: oferta y alquiler

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

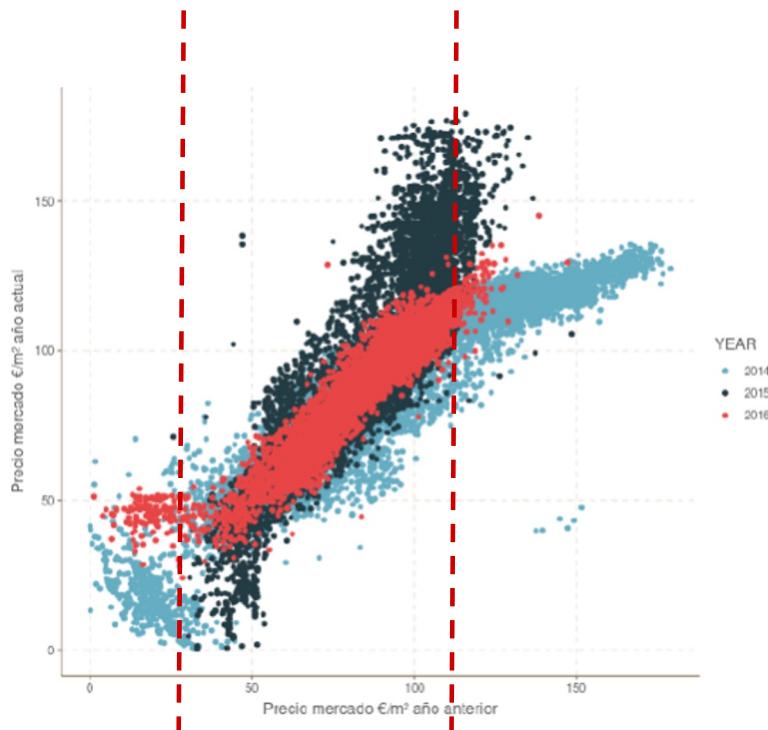


**Figura b.** Índice de precios de oferta por zona



# Capacidad de generalización para períodos futuros

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



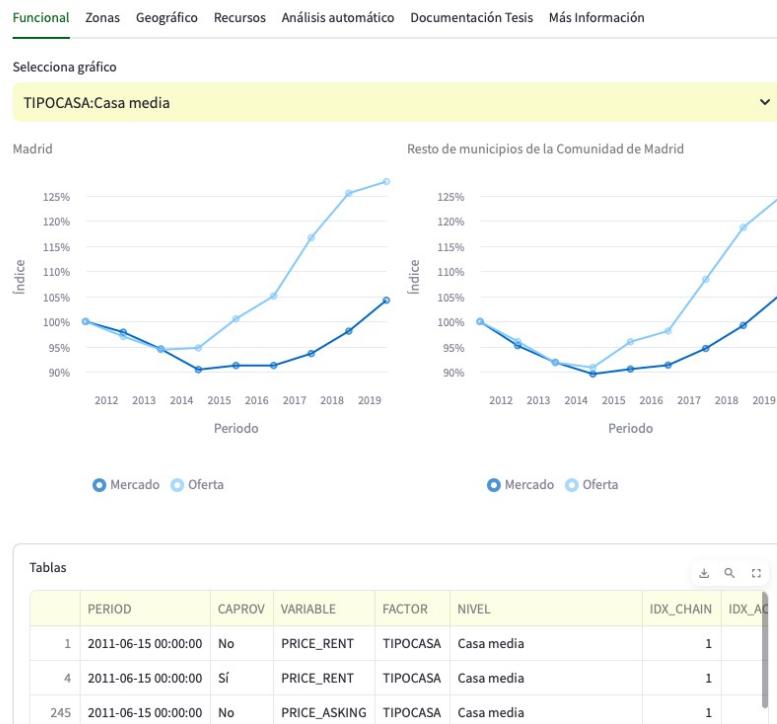
Los valores de mercado estimados con modelos de años consecutivos, entre 2014 y 2016, guardan un **coeficiente de correlación medio de 0,91** (mayor aún en la zona central de los precios).

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.

## Aplicación Interativa - Índice de Precios de la Vivienda en Alquiler



Usa el QR para acceder a la app web

# Principales conclusiones

1

El índice de precios construido ofrece una herramienta de análisis en profundidad del mercado inmobiliario con: **alta desagregación** temporal, zonal y funcional y **mantiene la coherencia** con las cifras generales del mercado publicadas por los organismos oficiales.

2

**Los portales de internet son una fuente válida** para construir un índice de precios de la vivienda con actualización frecuente, convenientemente ajustados con fuentes de información de mercado. El nivel de detalle y cobertura de estas fuentes pueden afectar al nivel de precisión del modelo de correspondencia estadística.

3

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 y alta precisión.

4

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

# Próximas líneas de investigación

1

Ajustes en el modelo de mercado, mejora en casos de muestra escasa y coherencia temporal

2

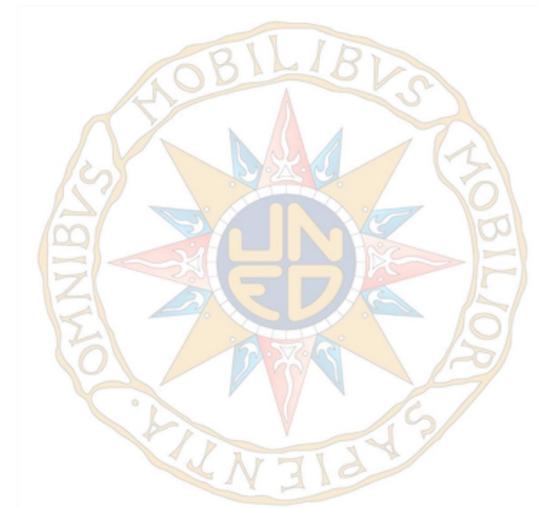
Incorporación de fuentes adicionales y evaluación en ámbito más amplio

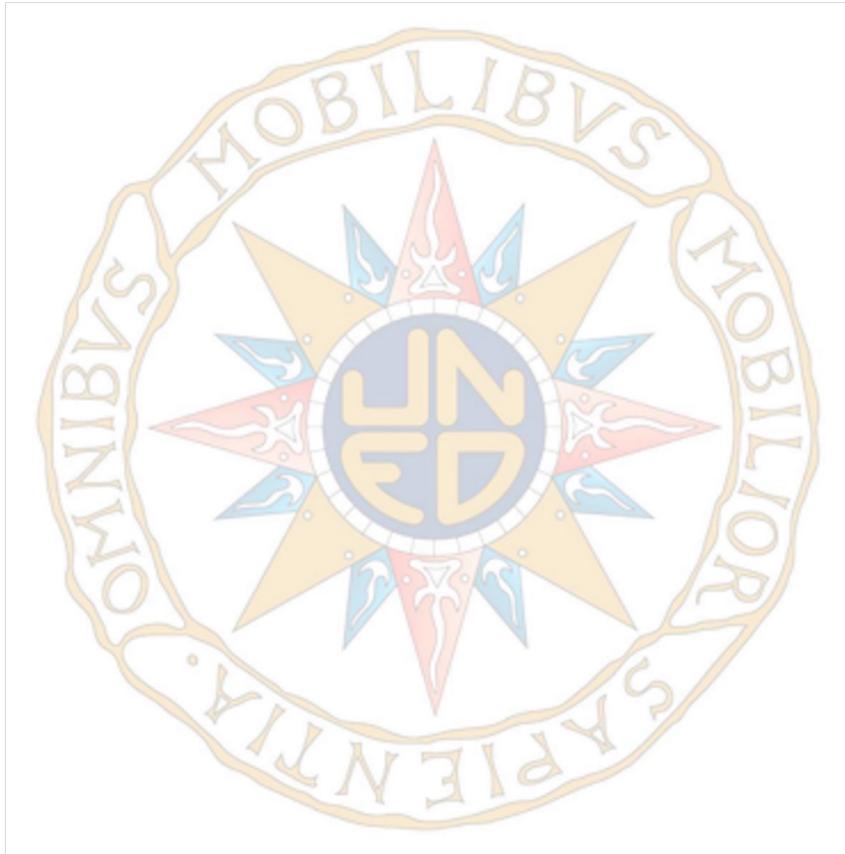
3

Uso de métodos hedónicos **alternativos** y profundizar en explicabilidad

- 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 y controlar anomalías en estratos con baja muestra
- Estudiar mecanismos que permitan mejorar la descriptibilidad del comportamiento de los hedónicos
- Mejoras en el proceso de desagregación temporal

**Muchas gracias por  
su atención**

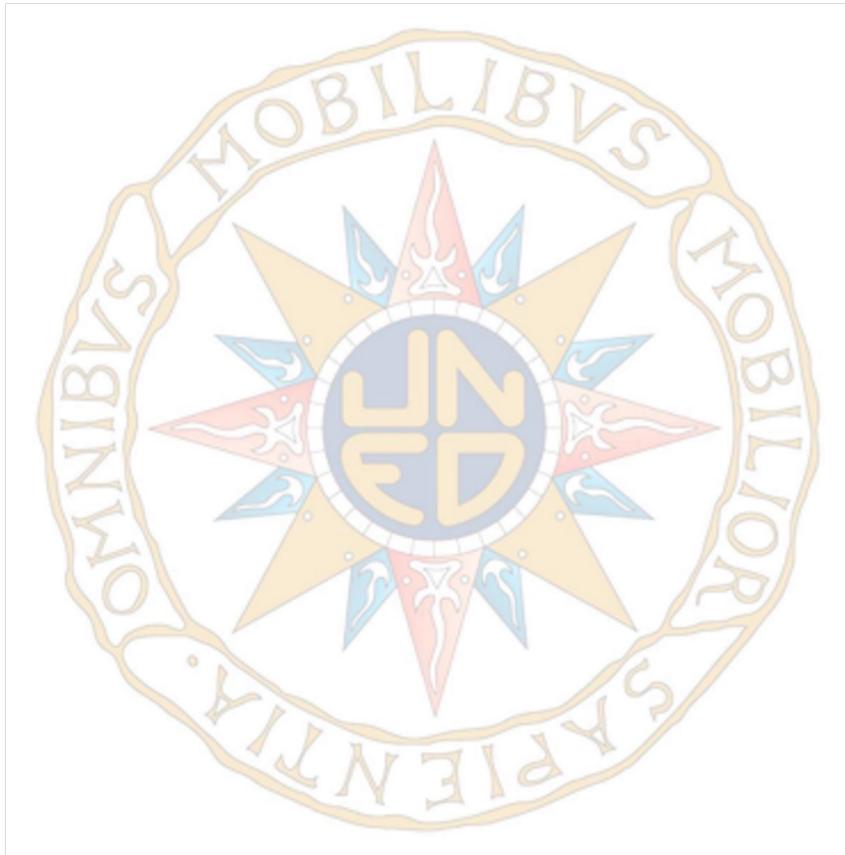




# Bibliografía

Las referencias bibliográficas se encuentran en el siguiente enlace:

<https://github.com/davidreyblanco/idx/blob/master/doc/chapters/refs.pdf>



# Anexos: artículos

# Artículos

1. Spatial Real Estate submarkets: [Using machine learning to identify spatial market segments. A reproducible study of major Spanish markets](#)
2. Rental Price Regulation impact in Catalonia : [Lessons from an Aborted Second-Generation Rent Control in Catalonia](#)
3. Time On Market and Price Spread relation in Real Estate Markets: [Regional Housing Market Conditions in Spain](#)
4. Use of web listing data as a proxy of transaction data: [Relación entre precios de alquiler en portales inmobiliarios y precios de mercado. Evidencias para la Comunidad de Madrid](#)
5. Accessibility applied to Hedonic House Price Models: [Improving hedonic housing price models by integrating optimal accessibility indices into regression and random forest analyses](#)
6. Time Series disaggregation: [Método de optimización en la selección de métodos de desagregación temporal](#)
7. Alternative House Price Index methodology: [Metodología para la estimación de índices de precios de la vivienda de alquiler](#)
8. Urban morphology and House Prices: [Impacto de la morfología urbana en la segmentación del mercado de la vivienda. Caso de estudio Madrid](#)
9. Urban morphology and Inequality: [Intra-Urban House Prices In Madrid \(Spain\) Following The Financial Crisis. An Exploration Of Spatial Inequality](#)
10. Time On Market in Real Estate Markets with survival models: [Metodología de estimación del tiempo de venta a través de modelos de supervivencia](#)