

Índice de precios de vivienda de segunda mano de la Comunidad de Madrid (IPVCM). Base 2022

NOTA METODOLÓGICA

Actualización: Diciembre 2025

1. **Presentación estadística**
 2. **Definiciones**
 3. **Metodología de procesamiento y cálculo de indicadores**
 - 3.1. **Recogida de ficheros de la fuente de datos**
 - 3.2. **Depuración y estratificación**
 - 3.3. **Modelo de ajuste de precios para cada tipología de vivienda**
 - 3.4. **Ponderación de cada tipología de vivienda**
 - 3.5. **Cálculo de índices de precios**
 4. **Difusión de resultados**
- ANEXO. Descripción del modelo de regresión**

1. **Presentación estadística**

El Índice de precios de vivienda de segunda mano de la Comunidad de Madrid (IPVCM) se crea en el año 2025 con la finalidad de aportar nueva información sobre la evolución de precios de compraventa de la vivienda usada de la Comunidad. Su propósito principal es ofrecer datos a un nivel territorial lo más detallado posible, completando de este modo la información proporcionada por el Índice de Precios de Vivienda (IPV) del Instituto Nacional de Estadística (INE), que ofrece datos sobre la materia hasta el nivel de comunidad autónoma.

Este nuevo índice de precios aporta información para todos los municipios de la Comunidad con más de 30.000 habitantes y para los 21 distritos del municipio de Madrid. Además, se incluyen resultados correspondientes a las 11 zonas estadísticas en que se divide la Comunidad.

El índice se publica por trimestres, comenzando las series temporales en el primer trimestre del año 2022. Junto con el índice se publican tasas de variación anual y trimestral.

Para elaborar este índice de precios se aprovechan fuentes de registros administrativos propios de la Administración de la Comunidad de Madrid, recopilados por la Dirección General de Tributos, donde se encuentra el precio de escrituración de las viviendas que han sido objeto de compraventa a lo largo del trimestre, así como información catastral acerca de las características de estas viviendas.

Con los datos provistos por esta fuente es posible estudiar los precios de vivienda de segunda mano, pero no así los de vivienda nueva, ya que no se dispone de datos completos de este tipo de viviendas.

En cuanto a la metodología de cálculo del Índice de precios de vivienda de segunda mano en la Comunidad de Madrid (IPVCM), esta se fundamenta en gran medida en la [metodología del Índice de Precios de Vivienda del Instituto Nacional de Estadística](#), la cual se encuentra armonizada a nivel europeo, como parte de la estadística sobre precios de vivienda que difunde EUROSTAT (House Price Index, HPI).

Se utiliza un índice de precios de tipo Laspeyres encadenado, que es ampliamente empleado en el cálculo de índices de precios, como por ejemplo el Índice de precios de consumo (IPC). Una característica destacada de los índices de Laspeyres es su capacidad para incorporar modificaciones metodológicas de forma periódica, en este caso, una vez al año coincidiendo con el primer trimestre. Esta flexibilidad permite introducir ajustes cuando resulta necesario, por ejemplo, en el modelo de estratificación.

Las viviendas de la muestra de cada trimestre se clasifican en distintas tipologías de vivienda obtenidas mediante estratificación según sus características. A partir de un modelo de regresión hedónica log-lineal se estiman los precios correspondientes a tipología de vivienda y, posteriormente, se calculan los índices de Laspeyres encadenados a partir de los precios estimados. Este modelo de ajuste por regresión resulta esencial para evitar que las variaciones en la composición de la muestra disponible en cada trimestre influyan en el cálculo del índice.

En relación con la comparabilidad con otras fuentes estadísticas, es necesario tener en cuenta que los precios que se utilizan para el cálculo del IPVCM son precios de escrituración de la compraventa antes de impuestos, lo cual permite cierta comparabilidad con los resultados del Índice de precios de vivienda del Instituto Nacional de Estadística con otras comunidades autónomas o con el total de España. Es importante tener en cuenta este aspecto, ya que otras estadísticas de precios de vivienda disponibles emplean valores de tasación o precios de oferta procedentes de portales inmobiliarios, que no son directamente comparables con los precios de escrituración.

Esta operación estadística Índice de precios de vivienda de segunda mano de la Comunidad de Madrid (IPVCM) por el momento tiene el carácter de experimental, estando pendiente de su inclusión en el Plan de Estadística de la Comunidad.

2. Definiciones

Año de antigüedad: Año de construcción del edificio donde se encuentra la vivienda, o bien el año que se realizó la última rehabilitación total del edificio. En cualquiera de los casos, este dato se toma de Catastro.

Índice de precios año base: Medida que permite comparar los precios respecto a un año específico, denominado “año base”. Se le asigna el valor 100 a la media de los cuatro trimestres del año base, y a partir de ahí se le calculan los índices de cada trimestre, en proporción al valor medio obtenido para el año base.

Índice de Laspeyres encadenado: Es una metodología estadística de uso común para calcular índices de precios. Los índices de Laspeyres se caracterizan por utilizar volúmenes de producto de periodos anteriores para ponderar los distintos productos de la cesta de precios con la que se construye el índice (para el IPVCM, los productos de la cesta de precios son tipologías de vivienda). Los índices encadenados se caracterizan por utilizar el último periodo de cada año para enlazar con los índices del año anterior.

La metodología de cálculo del IPVCM se basa en índices de Laspeyres encadenados, la cual se describe detalladamente en el punto 3.6 de este documento.

Media móvil ponderada: Es un procedimiento estadístico que se utiliza para suavizar series temporales. El valor suavizado se calcula haciendo la media del periodo actual y de varios valores de periodos anteriores, pero aplicando un mayor peso a los valores más recientes y menor a los más antiguos.

En el IPVCM se utiliza esta técnica para suavizar series de municipios o distritos donde la serie del índice presenta alta volatilidad, ocurriendo esto normalmente debido a que la muestra de compraventas disponible es relativamente pequeña. De este modo se amortiguan las oscilaciones y se facilita analizar la tendencia de la serie.

La fórmula concreta de suavizado que se utiliza se describe como parte del punto 3.6 de este documento.

Ponderaciones: Importancia de cada tipología de vivienda o categoría (municipio, distrito, etc.) sobre el total, que se utiliza para ponderar lo que debe representar cada tipología de vivienda en el cálculo del índice.

Para calcular estas ponderaciones se toma la muestra de compraventas acumulada en los dos años naturales anteriores al de referencia, donde el precio de cada vivienda se valora a partir de su superficie y el precio por metro cuadrado estimado por el modelo de regresión en el último trimestre del año anterior al de referencia.

Se hace notar que las ponderaciones que se aplican son las mismas para los cuatro trimestres de cada año. El cálculo de las ponderaciones se describe detalladamente en el punto 3.5 de este documento.

Precio de la vivienda: Precio de escrituración de la compraventa, sin incluir impuestos.

Superficie homogénea de la vivienda: Superficie construida de la vivienda, calculada después de aplicar un coeficiente entre 0 y 1 a la superficie de cada una de las unidades constructivas que conforman el inmueble catastral. Las unidades constructivas que estén clasificadas para uso de vivienda tienen siempre un coeficiente igual a 1.

Esta aplicación de coeficientes se lleva a cabo para que las zonas del inmueble catastral no destinadas a vivienda (trastero, garaje, piscina, zonas ajardinadas, zonas comunes, etc.) distorsionen lo mínimo posible el precio por metro cuadrado de compraventa.

Tipología constructiva del edificio: Tipo de construcción del edificio. Para el IPVCM se distingue entre pisos y viviendas unifamiliares.

Tipología de vivienda: Cada uno de los tipos de vivienda que se pueden obtener combinando los posibles valores de las variables sobre características de las viviendas. Para calcular el índice de precios, previamente cada vivienda se clasifica dentro de una tipología de vivienda concreta.

Variación anual: Porcentaje de cambio en el valor del índice respecto al mismo trimestre del año anterior.

Variación trimestral: Porcentaje de cambio en el valor del índice respecto al trimestre anterior.

Valor unitario de la vivienda: Precio en euros por metro cuadrado de la vivienda, resultante de dividir el precio de escrituración, antes de impuestos, entre la superficie homogénea de la vivienda.

Vivienda: Lugar cerrado y cubierto construido para ser habitado por personas.

A los efectos de esta estadística, se consideran viviendas los inmuebles catastrales para los que la unidad constructiva de mayor superficie se corresponde con uso para vivienda, pudiendo las restantes unidades constructivas estar destinadas a otros usos (trastero, garaje, piscina, zonas ajardinadas, zonas comunes, etc.)

Vivienda de segunda mano: Vivienda objeto de segunda y sucesivas transmisiones en la escritura de compraventa.

A los efectos de esta estadística, se consideran viviendas de segunda mano aquellas en las que el año de la fecha de la escritura de compraventa es al menos dos años posterior al año de antigüedad de la construcción (por ejemplo, para compraventas de 2025 se toman como viviendas de segunda mano las que tengan año de antigüedad en 2023 o anterior).

3. Metodología de procesamiento y cálculo de indicadores

3.1. Recogida de ficheros de la fuente de datos

El Índice de precios de vivienda de segunda mano de la Comunidad de Madrid (IPVCM) se basa en la explotación de registros administrativos que proceden de la Dirección General de Tributos de la Comunidad de Madrid.

Los registros contenidos en los ficheros se corresponden con las **compraventas del pleno dominio** de inmuebles catastrales que se han producido a lo largo del trimestre de referencia. Para que se incluya una compraventa en el fichero, de las unidades constructivas que componen el inmueble, la de mayor superficie debe corresponderse con uso para vivienda, pudiendo las restantes estar destinadas a otros usos: trastero, garaje, piscina, zonas ajardinadas, zonas comunes, etc.

Cada una de estas compraventas se asocian a una única referencia catastral. Los datos de características de vivienda se toman de la información que consta en Catastro.

Solo se incluyen las compraventas cuyo precio por metro cuadrado supera un determinado umbral para la valoración a efectos impositivos determinado por la Dirección General de Tributos, que depende de la ubicación geográfica de la vivienda. Las compraventas que quedan por debajo de dicho umbral no se consideran para esta operación estadística ya que se considera que no se vende a precio de mercado (compraventas entre familiares, edificaciones muy precarias o para demolición, etc.)

Los campos incluidos en los ficheros recibidos desde la Dirección General de Tributos son los siguientes:

- **Identificador único** del registro de compraventa:
Este campo permite distinguir de forma única cada registro, de modo que puedan detectarse registros duplicados o ya recibidos en entregas anteriores.
- **Fecha** de la operación:
Se refiere a la fecha en que se ha llevado a cabo la escrituración de la compraventa.
- Localización de la vivienda:
 - **Municipio**
 - **Distrito** (solo para el municipio de Madrid)
- Datos relativos al inmueble catastral:
 - **Superficie homogénea**: Se trata de la superficie construida de la vivienda calculada después de aplicar un coeficiente entre 0 y 1 a la superficie de cada una de las unidades constructivas que conforman el inmueble catastral. Las unidades constructivas que estén clasificadas para uso de vivienda tienen siempre un coeficiente igual a 1.
 - **Valor unitario**: Precio expresado en euros por metro cuadrado, resultado de dividir el precio de compraventa del inmueble, sin incluir impuestos, entre su superficie homogénea.

- Datos relativos al edificio donde se encuentra la vivienda:
 - **Tipología constructiva:** Se refiere al tipo de construcción tal y como conste en Catastro. Se distinguen tres posibles valores en estos ficheros: piso, vivienda unifamiliar y vivienda en hilera.
 - **Año de construcción** del edificio
 - **Año de antigüedad** del edificio: Año de la última reforma integral del edificio. En la gran mayoría de casos no hay reforma integral y se asigna el año de construcción.

Por tanto, la fuente de datos dispone de información sobre características principales que determinan el precio de compraventa de las viviendas, no obstante, existen otros factores determinantes de los que se carece de información, como pueden ser: la ubicación concreta dentro del municipio o distrito (cercanía a medios de transporte, vías principales, servicios, etc., o de igual forma, cercanía a zonas ruidosas, zonas conflictivas, etc.), el estado de conservación de la vivienda y del edificio en el que se ubica la vivienda, la altura dentro del edificio, la orientación y distribución de las estancias de la vivienda, circunstancias determinantes que hayan podido afectar a la negociación del precio, etc.

3.2. Depuración y estratificación

En primer lugar, se revisa si el formato del fichero es el correcto, se fusionan los ficheros recibidos y después se revisan todos los registros campo a campo, buscando información anómala, que puede consistir en un valor en blanco, un valor con formato incorrecto o un valor no esperado (fuera de rango). Los errores encontrados se examinan para poder consultar con el órgano remitente de los ficheros si se considera necesario. Los registros con información anómala por lo general se suprimen, con excepción de aquellos que puedan corregirse a través de imputación de la información faltante o errónea.

Después se lleva a cabo un proceso de filtrado de valores extremos o no representativos, que podrían distorsionar los resultados del modelo de cálculo.

En la fase de estratificación, se definen una serie de distintas tipologías de viviendas, de modo que cada vivienda de la muestra se clasifica dentro de una de estas tipologías, atendiendo a sus características.

Para ello, es necesario hacer varias transformaciones a las variables descriptivas de las viviendas:

- Para la tipología constructiva, la categoría “viviendas en hilera” se integra dentro de la categoría “viviendas unifamiliares”, ya que en la práctica se dan casos en que es difícil distinguir entre ambas categorías, por lo que pueden considerarse como la misma.
- Las variables superficie y antigüedad son numéricas en la fuente original, por lo que es preciso transformarlas en variables categóricas para poder estratificar, escogiendo una serie de intervalos que conformen varias categorías de superficie y de antigüedad. Adicionalmente, como en el caso de la superficie se observa una distribución muy distinta según la tipología constructiva, se van a considerar unos intervalos de superficie para pisos y otros diferentes para viviendas unifamiliares.

- En relación con la localización de la vivienda (municipio y distrito), es necesario tener en cuenta que el número de compraventas en los municipios de pequeño tamaño es insuficiente como para considerar cada uno de ellos por separado, por lo que es necesario diseñar una estratificación en la que haya categorías territoriales formadas por grupos de municipios pequeños, donde en la medida de lo posible los municipios de la misma categoría presenten características similares en cuanto a distribución de precios. Además, como se requiere poder calcular índices por zonas estadísticas, es necesario que los municipios de cada categoría territorial pertenezcan todos a la misma zona estadística.

A continuación, se define el modelo de estratificación escogido para los años 2022-2025:

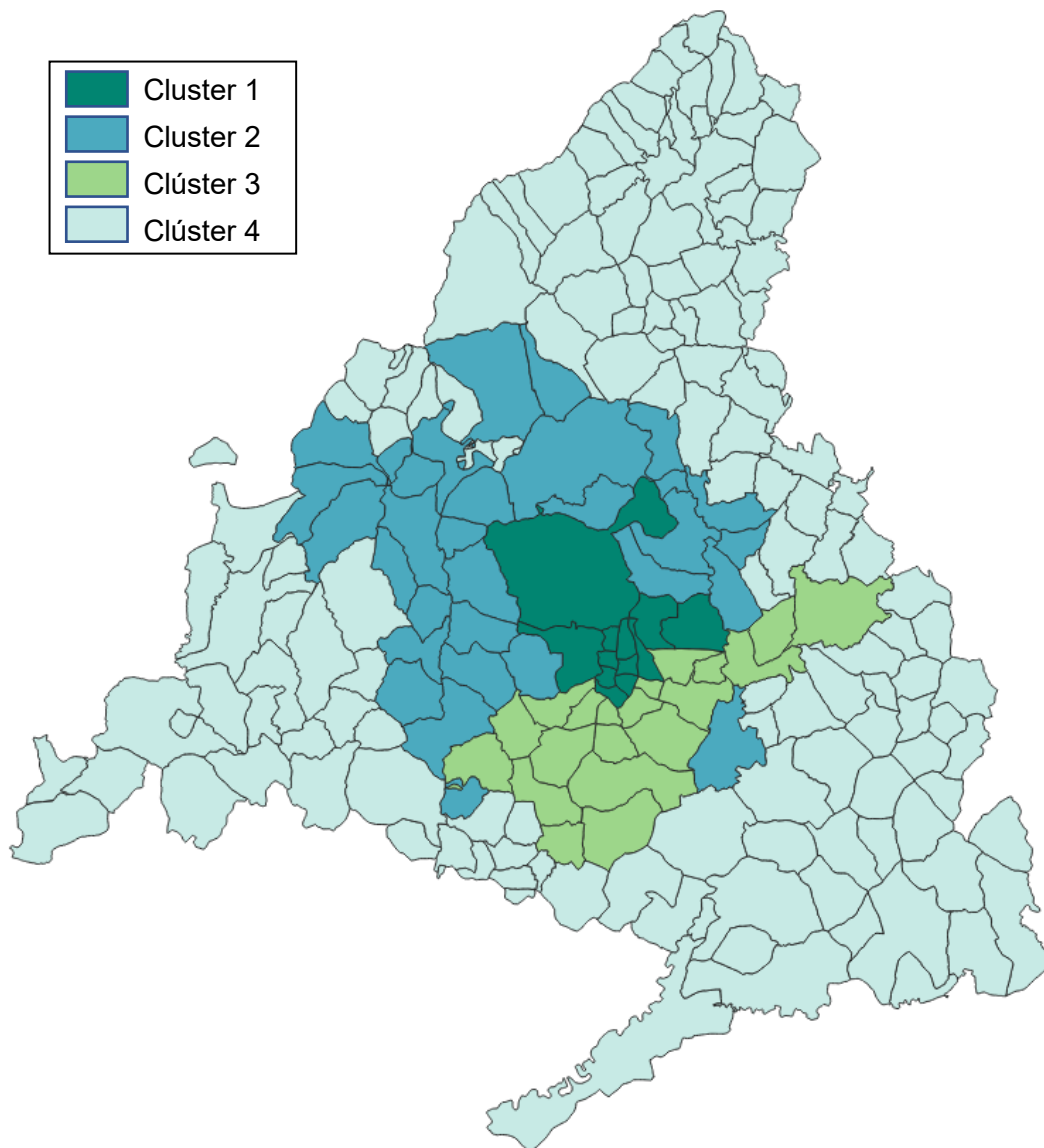
Variable	Categorías
Territorio (71 categorías)	- Cada uno de los 21 distritos de Madrid capital - 32 municipios (de tamaño grande y mediano) - 18 agrupaciones (de municipios de tamaño pequeño)
Año de antigüedad (10 categorías)	- Hasta 1900 - De 1901 a 1930 - De 1931 a 1950 - De 1951 a 1960 - De 1961 a 1970 - De 1971 a 1980 - De 1981 a 1990 - De 1991 a 2000 - De 2001 a 2010 - De 2011 y siguientes
Tipología constructiva y superficie homogénea (15 categorías)	- Pisos de menos de 40 m² - Pisos de 40 m² a menos de 60 m² - Pisos de 60 m² a menos de 80 m² - Pisos de 80 m² a menos de 100 m² - Pisos de 100 m² a menos de 120 m² - Pisos de 120 m² a menos de 140 m² - Pisos de 140 m² a menos de 160 m² - Pisos de 160 m² a menos de 200 m² - Pisos de 200 m² a menos de 240 m² - Pisos de 240 m² en adelante - Viviendas unifamiliares de menos de 120 m² - Viviendas unifamiliares de 120 m² a menos de 160 m² - Viviendas unifamiliares de 160 m² a menos de 200 m² - Viviendas unifamiliares de 200 m² a menos de 240 m² - Viviendas unifamiliares de 240 m² en adelante

Como se puede observar, la variable territorial es la que tiene la estratificación más fina, por ser la más importante en la determinación del precio de las viviendas. Combinando entre sí las categorías de las tres variables, se podrían llegar a obtener hasta 10.500 posibles tipologías de vivienda.

Sin embargo, debido a la heterogeneidad del mercado inmobiliario de la Comunidad de Madrid, se han agrupado los municipios en varias zonas, que denominaremos conglomerados o clústeres, de modo que cada trimestre se aplica un modelo de ajuste de precios distinto para cada uno de estos clústeres.

Se han definido 4 clústeres para el periodo 2022-2025:

- Clúster 1: Precios altos; densidad de población alta.
(Madrid capital: Almendra Central y distritos en norte y oeste)
- Clúster 2: Precios altos; densidad de población media y baja.
(Parte de los municipios en área metropolitana norte y oeste y Sierra Central, junto con Arroyomolinos y Rivas-Vaciamadrid)
- Clúster 3: Precios medios; densidad de población alta.
(Madrid capital: distritos en sur y este; Municipios más poblados en área metropolitana este y sur)
- Clúster 4: Precios medios y bajos; densidad de población media y baja.
(Resto de la Comunidad de Madrid)



Al aplicar un modelo por separado para cada clúster, es necesario reajustar la estratificación de las variables, ya que para algunas categorías la disponibilidad de muestra es insuficiente. Por ejemplo, las viviendas con año de antigüedad anterior a 1900 se acumulan sobre todo en el clúster 1, siendo muy minoritarias en el resto de clústeres, donde es necesario agrupar esta categoría junto con otras con menos antigüedad para lograr suficiente muestra para el modelo. De igual modo, la disponibilidad de registros de vivienda unifamiliar en Madrid capital es muy escasa, por lo que es necesario unir todos los tramos de superficie de vivienda unifamiliar en una única categoría para el clúster 1.

En el caso de la variable territorial se garantiza que cada uno de los estratos de territorio tenga al menos aproximadamente 50 registros por trimestre. Para las restantes variables, cada categoría debe contar con al menos aproximadamente 100 registros por trimestre, excepto en los intervalos en los extremos, que son los más dispares con el resto, donde se rebaja la exigencia a aproximadamente 70 registros al trimestre.

Una vez hecho el ajuste necesario en la estratificación cada conglomerado, se reduce el número de tipologías de vivienda a aproximadamente 5.000, de las que realmente están representadas en los datos entre 2.500 a 2.800 tipologías de vivienda distintas, según trimestre.

El número de distintos valores por variable en cada clúster es el siguiente:

	Clúster 1	Clúster 2	Clúster 3	Clúster 4
Territorio	12	20	20	19
Antigüedad	10	8	6	6
Superficie y tipología constructiva	11	9	11	9
TOTAL	33	37	37	34

3.3. Modelo de ajuste de precios para cada tipología de vivienda

En este punto, para cada tipología de viviendas disponemos de varios precios de compraventa, por lo que la siguiente fase consiste en llevar a cabo la estimación de un único precio para el trimestre para cada tipología de vivienda. Esta estimación se utilizará más adelante en la fórmula de cálculo de los índices de precios.

Para realizar esta estimación, se ha escogido la técnica de la regresión hedónica, que también utiliza el Instituto Nacional de Estadística para su Índice de Precios de Viviendas. En el modelo de regresión hedónica, se estima el valor que tiene cada una de las características que se han utilizado para construir los estratos de tipología de viviendas. De este modo, el precio para cada tipología se calcula finalmente teniendo en cuenta el valor estimado para cada una de sus características.

En el modelo de regresión que se aplica es log-lineal, es decir, la variable dependiente no es el precio en sí, sino su logaritmo neperiano, lo cual es bastante típico en modelización de precios, ya que ajusta mejor que un modelo puramente lineal. De ese modo, el valor que se da a cada categoría no suma, sino que multiplica el “precio base” por un valor determinado.

Se aplica un modelo de regresión distinto por cada trimestre y clúster, con un número de parámetros que varía entre 31 y 35, según la cantidad de categorías de cada variable para cada clúster. Cada una de las variables genera una cantidad de parámetros en el modelo igual a su número de posibles valores menos 1. Por tanto, el total de parámetros es la suma de los parámetros generados por cada una de las variables, menos el número de variables, al que hay que añadir un término más, llamado término constante o intercepto.

El modelo de regresión se explica detalladamente en el anexo al final del documento.

3.4. Ponderación de cada tipología de vivienda

Además del precio de cada tipología de vivienda, se necesita también ponderar la importancia relativa de cada tipología de vivienda respecto al total de viviendas de la muestra de compraventas.

Esto es necesario debido a que la composición de la muestra de viviendas disponible es diferente para cada trimestre, por tanto, se necesita tener fijada una referencia suficientemente estable de la composición de viviendas del mercado inmobiliario para que las comparaciones a lo largo del tiempo sean representativas. Una vez que se fija la ponderación al principio de cada año, esta se mantiene durante todos los trimestres del año y se revisa únicamente en el primer trimestre de cada año.

Para obtener las ponderaciones, se calcula la valoración de la suma total del precio de las viviendas de cada tipología en los dos años completos anteriores al año al que corresponde el trimestre de referencia de los datos. Estos precios se valoran según la estimación del modelo de regresión para el 4º trimestre del año anterior.

La fórmula de cálculo para las ponderaciones de cada estrato tipología de viviendas es la siguiente:

$$W_{e,a} = \frac{S_{e,(a-2,a-1)} \cdot \hat{P}_{e,(4,a-1)}}{\sum_{i \in E} (S_{i,(a-2,a-1)} \cdot \hat{P}_{i,(4,a-1)})}$$

donde:

$W_{e,a}$ es el peso de la tipología de vivienda (estrato) e durante el año a .

$S_{e,(a-2,a-1)}$ es la superficie total de las viviendas que corresponden a la tipología de vivienda e a lo largo de los años $a-2$ y $a-1$.

$\hat{P}_{e,(t,a-1)}$ es el precio estimado según el modelo de regresión para la tipología de vivienda e en el trimestre t del año $a-1$.

$E = \{e_1, \dots, e_N\}$ es el conjunto de todas las tipologías de viviendas.

Para los años 2022 y 2023 ha sido necesario variar ligeramente la fórmula de cálculo, ya que no se dispone de datos de años anteriores. La valoración para 2022 se hace a partir del primer trimestre disponible, que es el primer trimestre de 2022.

Las fórmulas de cálculo de ponderaciones son las siguientes para esos años 2022 y 2023:

$$W_{e,2022} = \frac{S_{e,(2022,2023)} \cdot \hat{P}_{e,(1,2022)}}{\sum_{i \in E} (S_{i,(2022,2023)} \cdot \hat{P}_{i,(1,2022)})}$$

$$W_{e, 2023} = \frac{S_{e, (2022, 2023)} \cdot \hat{P}_{e, (4, 2022)}}{\sum_{i \in E} (S_{i, (2022, 2023)} \cdot \hat{P}_{i, (4, 2022)})}$$

Una vez calculadas las ponderaciones de cada tipología de vivienda, se pueden calcular fácilmente las ponderaciones de cualquier agregado $A \subset E$, formado a partir de la unión de cualesquiera tipologías. Se obtiene como suma de las ponderaciones de las tipologías de vivienda que comprende dicho agregado.

$$W_{A, a} = \sum_{i \in A} W_{i, a}$$

Es fácil ver que se cumple que, para cualquier año a , la suma de las ponderaciones de todas las tipologías es 1, es decir, $W_{E, a} = 1$.

3.5. Cálculo de índices de precios

Para calcular los índices de precios se utiliza la fórmula de Laspeyres encadenado, la cual es una fórmula de uso común en el cálculo de índices de precios. Por ejemplo, se utiliza por parte del Instituto Nacional de Estadística (INE) en el cálculo del Índice de Precios de Vivienda (IPV) y el Índice de Precios de Consumo (IPC), entre otros.

Una característica muy importante de los índices Laspeyres encadenados es que permite hacer modificaciones metodológicas una vez cada año, ya sea la revisión del modelo de regresión o la modificación de las tipologías de vivienda. El trimestre de encadenamiento que se ha escogido es el cuarto trimestre del año, por lo que las modificaciones y actualizaciones metodológicas pueden hacerse en el primer trimestre de cada año.

Cálculo de índices elementales

Se denominan índices elementales aquellos que se refieren a una única tipología de vivienda, y en cuyo cálculo no intervienen las ponderaciones. A partir de los índices elementales se podrán calcular índices agregados de varias tipologías de vivienda, en donde sí tendrán que utilizarse las ponderaciones.

El índice elemental en un trimestre para una tipología de vivienda concreta se obtiene como el cociente del precio estimado por el modelo en el trimestre, dividido entre el precio estimado por el modelo para el cuarto trimestre del año anterior, multiplicado por 100. La fórmula de cálculo es la siguiente:

$$I_{e, (t, a)} = 100 \cdot \frac{\hat{P}_{e, (t, a)}}{\hat{P}_{e, (4, a-1)}} \quad , a > 2022$$

donde:

$I_{e, (t, a)}$ es el índice elemental para la tipología e en el trimestre t del año a .

$\hat{P}_{e,(t,a)}$ es el precio estimado según el modelo de regresión para la tipología e en el trimestre t del año a .

Los índices elementales del primer año de datos, el año 2022, se calculan respecto al primer trimestre del año 2022, ya que no hay disponibilidad de datos para el cuarto trimestre del año anterior, por lo que la fórmula de cálculo para ese año es:

$$I_{e,(t,2022)} = 100 \cdot \frac{\hat{P}_{e,(t,2022)}}{\hat{P}_{e,(1,2022)}}$$

Por tanto, para el caso particular de los índices elementales del primer trimestre del año 2022, son todos iguales a 100.

Cálculo de índices agregados

Para calcular el índice de una agregación A de tipologías de vivienda cualesquiera, ya sea una agregación a nivel geográfico o a partir de otra(s) variable(s), se utilizan los índices elementales y las ponderaciones para cada tipología de vivienda de la agregación. La fórmula de cálculo es:

$$I_{A,(t,a)} = \frac{\sum_{i \in A} (W_{i,a} \cdot I_{i,(t,a)})}{W_{A,a}}$$

donde $I_{A,(t,a)}$ es el índice para el agregado A en el trimestre t del año a .

A partir de esta fórmula, desarrollando tanto la ponderación como los índices elementales, y simplificando términos después, se puede encontrar otra expresión para los índices agregados, que depende de las superficies y de los precios valorados por el modelo, en lugar de depender de las ponderaciones y los índices elementales:

$$I_{A,(t,a)} = 100 \cdot \frac{\sum_{i \in A} (S_{i,(a-2,a-1)} \cdot \hat{P}_{i,(t,a)})}{\sum_{i \in A} (S_{i,(a-2,a-1)} \cdot \hat{P}_{i,(4,a-1)})}$$

Se observa en esta expresión que los índices agregados comparan los precios totales de las viviendas del agregado, según la valoración en el trimestre actual contra los precios de las mismas viviendas según la valoración en el 4º trimestre del año anterior.

Cálculo de índices año base 2022

Los índices año base 2022, que son los que finalmente se publican y a partir de los cuales se calculan las tasas de variación de los índices, se obtienen encadenando los índices referidos al cuarto trimestre, de un año con el siguiente.

El primer año de la serie es una excepción, ya que no hay año anterior con el que encadenar. En este caso lo que hace es llevar a cabo un ajuste para que la media del año base tenga valor igual a 100 para cualquier agregado.

$$I_{A, (t, 2022)}^{2022} = 100 \cdot \frac{I_{A, (t, 2022)}}{\frac{1}{4} \sum_{k=1}^4 I_{A, (k, 2022)}}$$

donde $I_{A, (t, a)}^{2022}$ es el índice base 2022 para el agregado A en el trimestre t del año a .

A partir del año 2023, el índice año base se calcula enlazando con el 4º trimestre del año anterior:

$$I_{A, (t, a)}^{2022} = I_{A, (t, a)} \cdot \frac{I_{A, (4, a-1)}^{2022}}{100}, \quad a > 2022$$

Desarrollando el encadenamiento, para cualquier trimestre del año 2023, los índices base 2022 quedarían así:

$$I_{A, (t, 2023)}^{2022} = I_{A, (t, 2023)} \cdot \frac{I_{A, (4, 2022)}^{2022}}{100}$$

Del mismo modo, para cualquier trimestre del año 2024:

$$\begin{aligned} I_{A, (t, 2024)}^{2022} &= I_{A, (t, 2024)} \cdot \frac{I_{A, (4, 2023)}^{2022}}{100} \\ &= I_{A, (t, 2024)} \cdot \frac{I_{A, (4, 2023)}^{2022}}{100} \cdot \frac{I_{A, (4, 2022)}^{2022}}{100} \end{aligned}$$

Para años sucesivos, la fórmula general una vez desarrollado el encadenamiento es:

$$I_{A, (t, a)}^{2022} = I_{A, (t, a)} \cdot \frac{I_{A, (4, a-1)}^{2022}}{100} \cdot \frac{I_{A, (4, a-2)}^{2022}}{100} \cdot \dots \cdot \frac{I_{A, (4, 2022)}^{2022}}{100}$$

Índices suavizados a partir de los índices año base (para algunos municipios y distritos)

Estudiando las series temporales que se obtienen a partir de los índices año base para municipios mayores de 30.000 habitantes y para los distritos del municipio de Madrid, se observa que para algunos municipios y distritos aparece una alta volatilidad en las variaciones trimestrales, especialmente cuando la muestra disponible es relativamente pequeña. Este hecho dificulta el uso de estas series temporales para análisis coyuntural.

Para estos casos se ha optado por utilizar un ajuste por suavizado utilizando la media móvil ponderada exponencialmente, utilizando únicamente los 2 periodos anteriores al dato, para mejorar la estabilidad de las series y sus tasas de variación y así facilitar el análisis a medio y largo plazo de las series.

El ajuste realizado es el siguiente: si tenemos una serie temporal de índices año base con n periodos $s_1, s_2, \dots, s_n$, la serie suavizada $s_1^*, s_2^*, \dots, s_n^*$, vendría definida por la expresión:

$$s_i^* = \frac{s_i + \alpha s_{i-1} + \alpha^2 s_{i-2}}{1 + \alpha + \alpha^2}$$

Son excepción los 2 primeros términos de la serie, en donde se suaviza con los términos anteriores disponibles, es decir:

$$s_1^* = s_1$$

$$s_2^* = \frac{s_2 + \alpha s_1}{1 + \alpha}$$

El valor de α se fija en función del tamaño de la muestra disponible y de la volatilidad de las variaciones trimestrales observada en los 3 años más recientes. Este α oscila entre 0,3 y 0,5.

La serie resultante suavizada no se reescala nuevamente, por lo que el valor medio del año base puede ser ligeramente distinto de 100 en series suavizadas.

Cálculo de tasas de variación

La tasa de variación anual de un agregado A en un determinado trimestre t_2 del año a_2 mide el porcentaje de variación, ya sea crecimiento o decrecimiento, del índice año base en el trimestre en relación con un trimestre anterior, t_1 del año a_1 :

$$\Delta_{A,(t_2, a_2),(t_1, a_1)} = 100 \cdot \left(\frac{I_{A, (t_2, a_2)}^{2022}}{I_{A, (t_1, a_1)}^{2022}} - 1 \right)$$

Cuando se trata de la variación anual se compara con el mismo trimestre del año anterior, mientras que en el caso de la variación trimestral se compara con el trimestre inmediatamente anterior.

4. Difusión de resultados

A fecha de redacción de esta metodología, solamente se ha llevado a cabo la publicación inicial de este índice, que abarca desde el primer trimestre de 2022 hasta el segundo trimestre de 2025. Está prevista la difusión trimestral a lo largo del tercer mes posterior a la finalización de cada trimestre, en fecha por concretar.

La periodicidad de las series es trimestral para los índices propiamente dichos y las variaciones anual y trimestral, mientras que las ponderaciones tienen una periodicidad anual, ya que la ponderación no varía a lo largo del año.

Los índices, tasas de variación y ponderaciones se desagregan por:

- Municipio (solo mayores de 30.000 habitantes, dado que con población inferior no presentan suficiente muestra)
- Distrito del municipio de Madrid
- Zona estadística
- Tipología constructiva (pisos / viviendas unifamiliares)
- Intervalos de superficie construida (solo para pisos, debido a que en viviendas unifamiliares no existe muestra suficiente en determinadas zonas de la Comunidad)
- Intervalos de año de antigüedad de la vivienda

ANEXO. Descripción del modelo de regresión

El modelo de regresión se utiliza para estimar los precios por metro cuadrado para las diferentes tipologías de vivienda en cada trimestre del año, a partir de los datos de la muestra de compraventas de vivienda de las que se dispone.

Para cada registro de la muestra de viviendas, por un lado, tenemos el precio de venta por metro cuadrado (antes de impuestos) y, por otro lado, una serie de características clave que influyen en la determinación del precio de la vivienda y que encuadran a cada vivienda dentro de una determinada tipología de viviendas.

Estas características pertenecen a una serie de variables categóricas V_i (que en el modelo actual son 3 variables categóricas: territorio, año de antigüedad, tipología constructiva / superficie), cada una de las cuales con n_i categorías. Empezamos construyendo vectores de n_i elementos para cada una de estas variables categóricas en cada vivienda, de modo que tomando se le asigna un valor 0 a todos los elementos del vector, excepto al elemento que se corresponde con el número de categoría a la que corresponde la vivienda, que se le asigna el valor 1.

Debido a que el último elemento de estos vectores se puede deducir, ya que si una vivienda no está en las $n_i - 1$ primeras categorías entonces forzosamente estará en la última categoría, solo necesitamos los $n_i - 1$ primeros elementos de estos vectores. Además, es necesario eliminar este último elemento ya que si no tendríamos problemas posteriormente en los cálculos del modelo.

A continuación, para cada vivienda se unen secuencialmente los vectores de $n_i - 1$ de todas las variables, formando un vector con valores 0 y 1, que recogerá las características de cada vivienda. Además, para que pueda funcionar el modelo, es necesario añadir un elemento adicional a ese vector, usualmente colocado al principio del vector, que tomará valor 1 para todas las viviendas, que se denomina término constante (o término independiente o intercepto). Este vector se denominará vector de características de la vivienda, $x = (x_1, x_2, \dots, x_k)$.

El número de total de elementos de x , por tanto, es igual al número total de características posibles de todas distintas variables de características que intervienen en el modelo, menos el número de variables de características, más 1. Este es el número de parámetros del modelo.

Una vez construido el vector de características, pasamos al modelo de regresión en sí que se va a utilizar, que es log-lineal, es decir, que no depende del precio por metro cuadrado, sino de su logaritmo neperiano.

Para cada vivienda de la muestra (vivienda i -ésima) tendremos que:

$$\ln p_i = x_{i,1}\beta_1 + x_{i,2}\beta_2 + \dots + x_{i,k}\beta_k + u_i$$

donde:

$(x_{i,1}, x_{i,2}, \dots, x_{i,k})$ es el vector de características de la vivienda, construido como se ha descrito más arriba, que solo toma valores 0 o 1.

$(\beta_1, \beta_2, \dots, \beta_k)$ es el vector de parámetros del modelo, que se obtendrá como resultado del ajuste del modelo de regresión.

u_i es la componente aleatoria o término de error del modelo.

Llamaremos:

n al número total de viviendas de la muestra.
 k al número de parámetros del modelo.

El modelo que se ha visto anteriormente para cada vivienda se puede poner en forma matricial, de modo que tendremos el modelo completo con todas las viviendas en una única expresión:

$$Y = \ln P = X\beta + U$$

donde:

P es un vector de dimensión $(n \times 1)$ que contiene los precios de las viviendas.
 Y es un vector de dimensión $(n \times 1)$ que contiene los logaritmos neperianos de los precios de las viviendas.

Por tanto: $Y = \ln P$

X es la matriz de características, de dimensión $(n \times k)$, cuyas n filas contienen los vectores de características de cada vivienda de la muestra.

β es el vector de parámetros del modelo desconocidos inicialmente y que se obtendrá como resultado del modelo de regresión, de dimensión $(k \times 1)$.

U es el vector términos de error (o residuos) del modelo, de dimensión $(n \times 1)$, un término de error por cada vivienda.

El estimador MCO (mínimos cuadrados ordinarios) de β , que llamaremos $\hat{\beta}$, que es el estimador que minimiza la suma de los cuadrados de los residuos, viene dado por la expresión matricial:

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

(Se utiliza la notación X^T para la matriz transpuesta de X)

Para que este estimador proporcione resultados adecuados es importante verificar que los residuos siguen una distribución normal de media 0, es decir, $E[U] = 0$. Bajo esta condición se sigue que $\hat{\beta}$ es un estimador insesgado de β , es decir, se cumple que $E[\hat{\beta}] = E[\beta]$.

La matriz de varianzas-covarianzas del estimador, que llamaremos V , es una matriz cuadrada de dimensión $(k \times k)$, que se obtiene como:

$$V = \text{Var}[\hat{\beta}] = \sigma^2 (X^T X)^{-1}$$

donde σ^2 es la varianza de los residuos.

El precio estimado para cada vivienda por el modelo de MCO finalmente se obtiene por la expresión:

$$\hat{P} = \exp(X\hat{\beta})$$

que es vector, de dimensión $(n \times 1)$.

Sin embargo, es necesario hacer correcciones a esta estimación por MCO para que resulte más precisa, las cuales se describen a continuación.

Corrección de la heterocedasticidad

Al aplicar el modelo de mínimos cuadrados ordinarios a los datos, los residuos presentan signos de heterocedasticidad, por lo que es recomendable realizar una transformación que haga que el modelo sea homocedástico.

En los modelos homocedásticos la varianza de los residuos es constante: $Var[U] = \sigma^2 I_{n \times n}$ donde $I_{n \times n}$ es la matriz identidad de dimensión $n \times n$. Esto implica que la varianza de los residuos es la misma, siempre σ^2 , independientemente del elemento de la muestra que se trate.

En los modelos heterocedásticos la varianza no es constante, por lo que cada elemento está sujeto a una varianza distinta $\{\sigma_1^2, \dots, \sigma_n^2\}$.

Se puede definir una matriz Ω :

$$\Omega = \left(\frac{\sigma^2}{\sigma_1^2}, \dots, \frac{\sigma^2}{\sigma_n^2} \right) \cdot I_{n \times n}$$

que es una matriz diagonal de dimensión $n \times n$, de modo que $Var[U] = \sigma^2 \Omega^{-1}$.

El modelo puede hacerse homocedástico premultiplicando por la matriz $\Omega^{1/2}$, es decir, que podríamos formar el siguiente modelo:

$$\Omega^{1/2} L = \Omega^{1/2} X \beta + \Omega^{1/2} U$$

del cual el estimador $\hat{\beta}$ que minimiza la suma ponderada de los cuadrados de los errores, llamado estimador MCP (mínimos cuadrados ponderados), pasa a tener esta expresión:

$$\hat{\beta} = (X^T \Omega X)^{-1} X^T \Omega Y$$

y su matriz de varianzas-covarianzas es:

$$V = Var[\hat{\beta}] = \sigma^2 (X^T \Omega X)^{-1}$$

La idea subyacente en relación con la introducción de esta matriz Ω es que los elementos de su diagonal constituyen un vector de pesos para cada elemento de la muestra de viviendas que participan en el modelo, de modo que las observaciones con varianzas grandes se les otorga un peso más pequeño en el modelo, al considerarse menos fiables que las observaciones con varianzas más pequeñas, a las que se da un peso mayor, al considerarse que estas son más fiables.

El problema radica en encontrar la matriz Ω , que en general no es tarea sencilla y normalmente hay que enfocar el problema en encontrar una aproximación a partir del análisis de las varianzas de los residuos que se obtienen del modelo de mínimos cuadrados ordinarios.

Corrección del sesgo que se produce al aplicar la función exponencial al resultado del modelo de mínimos cuadrados ponderados

Una vez corregida la heterocedasticidad, si tenemos el vector de características de una tipología de vivienda concreta $x_i = (x_{i,1}, x_{i,2}, \dots, x_{i,k})$ no habría más que aplicar la fórmula para obtener el precio estimado de dicha tipología, utilizando los resultados obtenidos para el estimador después de aplicar mínimos cuadrados ponderados $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k)$:

$$\hat{P}_i = \exp(x_i \hat{\beta}) = \exp(x_{i,1} \hat{\beta}_1 + x_{i,2} \hat{\beta}_2 + \dots + x_{i,k} \hat{\beta}_k)$$

El problema que el modelo de MCP nos devuelve logaritmos de precios y no los precios en sí, por lo que es necesario volver a la escala original aplicando la función exponencial, lo cual introduce un sesgo debido a que la función exponencial es no lineal. Concretamente, al ser una función convexa, la exponencial produce un valor por debajo del que debería.

Para corregir este sesgo de transformación se utiliza un estimador propuesto por El-Shaarawi y Viveros (1997), que tiene esta expresión:

$$\hat{P}_i = \exp\left(x_i \hat{\beta} - \frac{1}{2} x_i \hat{V} x_i' + \frac{1}{2} \hat{\phi} \hat{\sigma}^2\right)$$

donde $\hat{\sigma}^2$ se estima a partir de la media de los cuadrados residual, sin aplicar pesos y ajustada en función del número de parámetros del modelo:

$$\hat{\sigma}^2 = \frac{1}{n-k} \sum_{i=1}^n u_i^2$$

Para que funcione este estimador, se asume que los errores siguen una distribución normal. El parámetro $\hat{\phi}$ se calcula con la expresión:

$$\hat{\phi} = 1 - \frac{\hat{\sigma}^2}{2(n-k)} - \frac{\hat{\sigma}^4}{3(n-k)^2}$$

La matriz de varianzas-covarianzas se estima como: $\hat{V} = \hat{\sigma}^2 (X^T X)^{-1}$

Para cualquier consulta o aclaración sobre esta operación estadística, puede contactar con el Instituto Estadístico en el correo electrónico: iestadis@madrid.org