INSTITUTO NACIONAL DE ESTADISTICA

Index on Housing Prices. Base 2025

Methodology

June 2026

Contents

I Introduction

The Index on Housing Prices (IHP) was implemented in 2008 with the aim of complementing the information provided by the existing statistics on the real estate market in Spain, using an approach based on the quarterly evolution of the prices recorded in the deed of sale of the dwellings.

Until the publication of the IHP, the main statistic produced by public administrations was what is currently called the Appraised Housing Value, which is produced by the General Directorate of Architecture, Housing and Land, under the Ministry of Public Works. The aim of this operation is to offer quarterly estimates on the average value of homes per square metre. Its main source is appraisal companies, so the published information refers to the homes that have been appraised within the reference period, without having to account for the changes occurring in the type of homes that make up the sample, nor their quality, from one period to another, an aspect that the IHP emphasises.

As for information from the private sector, it came from real estate portals and appraisal companies, which also offered information on housing prices, specifically, asking prices and appraisal values. However, none of them were based on actual purchase prices.

This lack of information on purchase prices, together with the lack of treatment of changes in the sample of homes acquired in different periods, were the main reasons that prompted Instituto Nacional de Estadística (INE) to consider carrying out a statistic that covered both aspects. Thus, the IHP uses purchase prices and employs a mixed adjustment method that combines stratification and hedonic regression, which was a novel use of statistical methods that set it apart from the rest of the existing official statistics.

Likewise, within the European Union, interest in measuring the evolution of housing prices dates back to 2002, when a Pilot Study Group on homeownership was created within the framework of the Harmonised Index of Consumer Prices (HICP). Spain has been part of this group since its inception. The work carried out there culminated in the publication of the IHP in Spain. In this sense, the IHP has been designed according to the guidelines and recommendations of the European Statistical Office (Eurostat), which gives this indicator the possibility of establishing international comparisons.

Therefore, the implementation of the IHP resolved the information needs in the domestic and international spheres, insofar as it is essential to have a harmonised indicator that allows for international comparisons.

To carry it out, a work plan was designed with the objective of implementing an Index on Housing Prices (IHP), taking into account the work conducted within the Eurostat Study Group.

In addition, in 2005 the INE created an Interministerial and Bank of Spain Working Group, with representatives from the Presidency of the Government, the Bank of Spain and the Ministries of Economy and Finance, Justice and Housing. In this Working Group, all the elements that should make up an indicator on the evolution of housing prices were presented for analysis, with the necessary characteristics that allow for the required international comparison.

Eurostat began publishing the harmonised housing price index in December 2010 as an experimental index. Subsequently, following the approval of Regulation (EU) No

93/2013, the official publication of the Housing Price Index (HPI) began. The only differences with the IHP are that in the HPI the prices of new homes include VAT, as set out in the Regulation, and that since 2016 the reference period for the weightings is two calendar years in the IPH and one year in the harmonised one.

This methodology describes the procedure followed in the years prior to the implementation of the IHP and describes the most relevant characteristics of the indicator, the objectives and areas of research, the source of information and calculation methodology used in its preparation, and its dissemination. The methodological section, the most extensive, summarises the production process of the index and explains, in detail, the regression model used in the price estimation.

II Background

The creation of the Study Group on Homeownership coordinated by Eurostat served to promote the development of a new operation aimed at measuring the evolution of housing prices in Spain. In fact, the work carried out within this Group can be considered as the preliminary work of the project followed to obtain the IHP.

In the project's **first phase** (between 2002 and 2004), this Group —composed of Germany, Spain, Finland, Poland and the United Kingdom— was tasked with studying the situation of the real estate market in each country, locating the main sources of information available and proposing a statistical calculation procedure that would allow for comparisons of the evolution of new housing prices among EU countries.

During this phase, a pilot test of new housing prices was carried out in Spain, whose main objectives were to verify the field collection and test out a questionnaire model in order to finally analyse the differences between this system of information collection and obtaining it through appraisal companies.

To this end, the field test was limited to the provinces of Madrid and Segovia, where new housing developments were visited to collect the sale price and characteristics of all types of housing available in the development.

Preliminary studies were also conducted using appraisal values, where the average values of appraisals were taken from by the largest appraisal companies in Spain across eight provinces, distinguishing by the size of the dwelling.

The conclusions obtained from these initial tests served as a basis to begin work on the design of the future index. From these, the basic ideas for the development of the index methodology were conceived, such as the ideal calculation frequency, the cost/efficiency ratio of each method and other relevant aspects, such as how changes in housing quality would be treated.

Then, in the **second phase**, seven other countries joined the Study Group (which was developed in 2006 and 2007): Cyprus, Slovenia, France, Greece, Italy, the Netherlands and Slovakia. The objective of this phase was to obtain data on the evolution of housing prices, both new and second-hand, in each of these countries. The results had to be based on comparable calculation methodologies, which would allow for the assembly of an index of owner-occupied housing prices with the greatest possible degree of harmonisation. This is the main reason why the project to implement the IHP took the general guidelines set by Eurostat for the harmonised indicator as a reference.

In addition, each country had to undertake the task of calculating an index of costs associated with the dwelling purchase, studying and selecting for this purpose the typical items of that expenditure.

During this period, the INE analysed all available sources to obtain the information necessary for calculating the IHP, on one hand taking into account the availability, timeliness and content of the information, and on the other, the relevant Eurostat standards, which required, and still require, that the prices used to calculate the indices are prices actually paid for the sales carried out in the reference period of the index.

Each of the available sources of information on the housing market offers a different perspective on the buying and selling process. In addition to the two sources already mentioned (appraisals and real estate developments) there are other agents that

provide information about housing: deeds, mortgage loans, the Land Registry and real estate agencies.

After a thorough study of these sources and after weighing the usefulness and suitability of each for the established objectives, it was considered that the values from the deeds provide the most appropriate information for monitoring the prices of owned housing under the comparability and homogeneity parameters set out by Eurostat.

There is a variety of reasons why the other available sources were ruled out for use in this project. In the case of appraisal values, the value of the property estimated by the companies does not necessarily have to be the same as the transaction price; likewise, not all appraised properties are sold, nor does it consider properties which are paid for in cash, without a mortgage loan.

Mortgage loans, on the other hand, were ruled out because their amount may not match the transaction price, and this difference between the effective price and the loan value varies in each case.

Land Registry data was also disregarded because the registry value does not provide an adequate measure of housing prices for the purposes set by the IHP.

Finally, although collecting prices through a survey aimed at real estate developers could provide useful and complete information for the project, it would only cover new housing, which would require using other sources to obtain prices for second-hand housing. The complexity and cost of such an operation made it infeasible.

Of all these sources, the one finally selected for calculating the IHP was the notary register, which contains, among other data, the official prices of all sales that have taken place in Spain and correspond to the value given on the property deed.

Furthermore, the use of administrative records allows for the availability of information on the entire population that makes up the scope of study or research, which favours the accuracy of the results and reduces costs when compared to other statistics that use sampling techniques for field data collection.

The section set aside for information sources details the most notable characteristics of the notary database. In 2007, this database underwent a profound restructuring and expansion of content to respond to the information demands of the Spanish Tax Agency. The IHP began to be published in October 2008, once the new database was consolidated; the indices published up to 2016 were based on the year 2007, the first year available with the new restructuring; in 2026 a second restructuring takes place, so that the indices published up to 2026 were based on the year 2015.

In parallel with the study of the sources, the second phase determined the ideal calculation method for the housing price index, which would solve the problem of fluctuations in the quality and composition of the homes sample used each quarter to calculate the index. This method requires the use of regression models, as will be seen in detail in the section dedicated to the calculation methodology.

Meanwhile, within Eurostat, the work plan of the Study Group continued, in which all EU countries, plus Norway and Iceland, ended up participating. The objectives were broadened and the quality of the indices improved, both in their coverage and in their calculation methodology, until achieving the degree of comparability and harmonisation between countries appropriate for it to begin to be disseminated. Thus, in 2011

Eurostat began to publish, on an experimental basis, a housing price index, initially for all housing, and later for new and second-hand homes. Independent data were provided for each country, as well as aggregated data for both the Monetary Union and the European Union.

Two years later, following the approval of European Commission Regulation No. 93/2013 of 23 February 2013, official publication of the Housing Price Index (HPI) began, a harmonised housing price index which differs slightly from the IHP published by the INE. The main difference relates to the prices of new housing which, as required by European regulations, include VAT in the harmonised indicator. Furthermore, since 2016, the reference period for the weightings differs in both indicators. While the IHP uses the home sales occurring during the two calendar years prior to the index reference year, the HPI uses those occurring in the last calendar year, as specified in the Regulation.

III Objectives

The IHP is a quarterly indicator whose main objective is to measure the **evolution in home sales on the open market**, both new and second-hand, over time. It is, therefore, an indicator designed solely to establish comparisons over time.

The measurement of price levels is not within its scope. Therefore, spatial comparisons of price levels cannot be established, although comparisons in price evolution can be made.

According to the criteria on coverage of the Harmonised Index of Consumer Prices (HICP), social housing is excluded from the calculation of the IHP because it is not accessible to all potential buyers and is not governed by the usual market mechanisms. These are homes whose type, dimensions and prices are government regulated, contingent upon buyers being able to take advantage of certain economic and tax benefits, who in turn must meet certain established conditions regarding property ownership, family income, etc.

Furthermore, within the scope of the production of European Union harmonised statistics, this index aims to serve as a point of comparison between the EU Member States, with regard to the evolution of housing prices. In this sense, it has been conceived under the same concepts and methodology used in the production of the EU HICPs. However, as mentioned in the previous section, there are some conceptual differences between the European indicator and the IHP; the most relevant being the inclusion of VAT in the prices of new homes in the HPI.

IV Investigation Scope

Statistical units

Given that the objective of the IHP is to measure the evolution of home sales in the open market in Spain, the *unit of analysis* is private housing. The *basic statistical unit* on which the information is collected is housing; however, during the calculation process, housing typologies are established, which allows for the calculation of the elementary indices. The housing typology, therefore, is a *derived statistical unit*, which includes homes with similar physical and location characteristics.

The *reporting unit* is the individual who provides the transfer data to the notary. All notarial information is centralised, and the Notarial Certification Agency (ANCERT for its Spanish initials) is responsible for supplying the data to various users, including the INE.

The *concept* to which the index refers is the home price.

Population scope

The reference population or stratum of the IHP includes the entire population (natural persons), both residents and non-residents in Spain, who have acquired a home in the reference period. Purchases made by legal entities or financial entities are not part of the IHP population scope.

Geographical scope

The geographical scope of the research is the entire national territory.

Temporal scope

The IHP is published on a quarterly basis.

V Information Sources

The information used to calculate the IHP comes from the General Council of Notaries, with whom the INE signed a collaboration agreement aimed at facilitating the use of notary data for statistical purposes. Accordingly, the General Council of Notaries, through the Notarial Certification Agency (ANCERT), provides the data that constitutes the main source of information for this indicator.

This source of information is the one that best suits the objectives of the IHP. The main characteristics that make it the ideal source are the following:

Availability

The agreement signed by the General Council of Notaries and the INE allows for the availability of the information necessary for calculating the IHP with the appropriate frequency and deadlines for this statistical operation.

Periodicity

ANCERT provides monthly, in electronic format, information on real estate transfers that have taken place in Spain during the previous four months. In each submission, the previously provided files are updated, new observations are incorporated, and modifications are made to those previously submitted.

Coverage

The quarterly calculation of the IHP uses data files received approximately one and a half months after the end of the reference quarter of the index. In this way, more than 90% of the total transactions carried out in the reference quarter of the index are included, which more than satisfies the needs of the indicator. This includes all sales of homes in the open market made by individuals in the Spain throughout the quarter, both new and second-hand homes, and regardless of the payment method, in cash or with financing.

Timeliness

The prices that must be included in the calculation of the index are those that refer to the sales of homes actually carried out in the reference quarter. Therefore, neither the offer prices nor the appraisal values are included. This fulfils the timeliness criterion which requires that the transactions carried out are included in the calculation of the index in the same quarter in which they take place.

Content

The notary database, in addition to the property purchase price, contains valuable additional information that allows, on the one hand, adaptation to the index coverage in terms of the type of transfer, type of property and buyer; and on the other hand, there is a good quality adjustment by having detailed information on the characteristics of the property, including its size and location, the most relevant aspects in determining the price.

It also includes the land registry reference of the property, from which the registered information can be accessed and used; currently, it is used in the data cleaning phase of the survey.

VI Variables used

The following variables are included in the data files regularly sent by ANCERT, grouped according to different aspects of interest:

– Time-based variable:

  - Authorisation date. Indicates the date when the property transfer takes place.

– Operation code:

  Transactions corresponding to the following legal acts are received:

  - Real estate sales.
  - Allocation of housing cooperative to its members.
  - Allocation to a member in the real estate development community.

– Variables related to housing prices:

  - Transaction price.
  - Item value.

  In acts with multiple items, that is, when several items are transferred in the same legal act (for example, a house and a parking space) and all of them are included in the *transaction price*, the value of each one is recorded in the *item value* field.

– Housing location variables:

  - Autonomous Community.
  - Province.
  - Municipality.
  - Postal code.
  - Type, name and number of the road.
  - Duplicate, block, staircase, floor and door.

– Variables related to housing characteristics:

  - Property type. Indicates the type of urban property: house, parking space, storage room, etc.
  - Land registry reference of the property, or reason for the lack thereof or why it could not be obtained, if applicable.
  - Variable that indicates whether the housing is market price or social housing.
  - Type of housing, distinguishing between flat and single-family home.
  - Variables that indicate whether parking and storage are included in the price.
  - Variable that indicates whether the property is new or second-hand.
  - A property is considered new when it is the first transfer in the deed of sale, normally made by the developer or builder in favour of the first purchaser; in the rest of the transfers, that is, when there is more than one transfer in the public deed, the property is considered second-hand.
  - Built area in square meters (m2).

– Buyer-related variables:
  - Type of person. Indicates whether the buyer is a natural person or a legal entity.
  - Country, province and municipality of residence of the purchaser.

All the variables mentioned above intervene, directly or indirectly, in the IHP production process, except those relating to the buyer's residence, since the IHP includes all homes acquired by individuals within Spanish territory, regardless of their nationality or place of residence.

As mentioned earlier, the location of the property, along with its size, are the most relevant elements in explaining the price. The database contains the exact address of the property; however, in order to use this information in the regression model used to estimate prices, it is necessary to group the provinces, municipalities and postal codes and thus have a number of categories that is not too high to avoid over-parametrization of the model. To this end, other sources of information have been used, which have made it possible to create new variables, classifications of the different geographical levels, according to one or more variables that are related to the price of housing. These variables are as follows:

– **Cluster of provinces**. Grouping of the 52 provinces into 6 groups by applying a cluster analysis, based on the average annual appraisal value of housing per province, published by the Ministry of Public Works.

– **Municipality size**. Classification of municipalities distinguishing between provincial capitals, large non-capital municipalities (over 50,000 inhabitants), medium (from 10,000 to 50,000 inhabitants) and small (under 10,000 inhabitants), using the latest available population data from the Continuous Register of Inhabitants (INE).

– **Tourist municipality**. A tourist municipality is considered as such if it concentrates a high number of overnight stays in tourist accommodations and/or the proportion of secondary residences versus primary residences is high. These are classified into one of the first three categories:
  - Sun and beach tourism
  - Rural, inland or nature tourism
  - Cultural, urban or business tourism
  - Other (non-tourist)

  This is done using information on the number of annual overnight stays in each type of tourist establishment, provided by the INE surveys of Hotel Occupancy, Occupancy in Tourist Apartments and Occupancy in Rural Tourism accommodations, as well as data on main and secondary housing from the last housing census, the population of the municipality and its location (coast or inland).

– **Type of environment**. Classification of postal codes into 14 categories, based on information from the latest housing census (2011) and the price per square meter per postal code, in a previous annual period.

Every year, the variables that classify the provinces, municipalities and postal codes are updated with the latest information available from the sources used in their preparation, either keeping the number of categories and/or content or not. For example, the variable *tourist municipality* initially only had two categories, tourist or

non-tourist, based solely on the criterion of high overnight stays; in the early years, the provinces were grouped according to the average mortgage amount; and the limit for distinguishing medium-sized from large municipalities was 100,000 instead of 50,000 inhabitants.

The use of chained indices allows for changes to be introduced annually without their effect being noticeable in the rates of change of the indices.

VII Processing of information

The files provided by the General Council of Notaries must be adapted to the technical requirements necessary for calculating the IHP. To achieve this, a process has been designed that guarantees the internal consistency of the data and excludes extreme values.

The process involves the following phases:

Initial phase

In the initial phase of the process, the three monthly files comprising the quarter are joined, all received variables are formatted, and an initial filter is applied to adapt the records to the different areas of the indicator; for this purpose, sales made by individuals and relating exclusively to open market housing are selected. The resulting file must include only homes, so multiple transfers, when different real estate assets such as parking spaces and storage rooms are transferred together in the same act, must undergo a specific treatment that breaks down the values of each.

In this first phase, significantly high and low values are also eliminated, both in terms of surface area and housing price. Subsequently, the outliers will be objectively detected from the regression model, taking into account, as a whole, all the characteristics of the dwellings collected in the model.

The dwellings excluded due to having an atypical value in the surface area variable will be analysed in the subsequent phase. If the values are erroneous and the corrected ones are available, they will be added back into the study.

Filtering and imputation phase

The second phase of the process focuses on the imputation and filtering of values. It consists of detecting inconsistent values in geographical variables (for example, a postal code that does not belong to the recorded municipality). These errors are detected automatically and resolved by using external sources.

Although notarial databases are usually complete, in some cases it is necessary to impute the value of variables such as surface area or type of environment. In the case of surface area, the land registry information on the dwellings is used to verify and correct, in case of error, the registered value.

Regarding the type of environment, as already described in the previous section, it involves a classification of postal codes; however, not all of them are classified. Therefore, an imputation procedure has been designed based on the observation of average quarterly prices: when an unclassified postal code is observed, the type of environment of that postal code will be imputed which, being within the same municipality, has the average price per square meter most similar to the price per square meter of the observation, in the corresponding quarter.

Expansion phase

Finally, although some derived variables are obtained throughout the process, which occurs, for example, with the price per square meter (price divided by surface area) —necessary in the imputation phase— most are incorporated into the data file at the end. This is the case for many of the explanatory variables in the regression model.

As mentioned above, the location of the property, along with its size, are the most determining factors in explaining its price. The database contains the exact address of

the property; however, in order to use it in the regression model, it is necessary to summarise this information into a few variables with a small number of values each, thus avoiding over-parametrisation of the model. This is why the provinces, municipalities and postal codes have been grouped together, taking into account information from other sources on a certain variable closely related to the price of housing; in the previous section it has been explained how these variables have been created. Also, for the same purpose, 10 surface intervals have been established, and the values of the variable "floor," which reflects the floor height within the building, have been grouped into six categories.

The resulting file is fully adapted to the IHP coverage and prepared to obtain, based on a regression model described in the following section, the estimated prices that will be included in the index calculation formula.

VIII Calculation methodology

The IHP calculation system is based on the combination of two basic elements that reflect the characteristics of the real estate market, and which are essential in the calculation of price indices: housing prices, which represent the confluence of supply and demand in the market, and the weightings, or relative importance of each type of housing according to the purchase value.

The combination of these two elements to obtain the IHP is done through the **formula for the chained Laspeyres index**, the same one used in the calculation of the CPI/HICP.

In addition to considering the two elements mentioned above, another relevant aspect of any price index is the adjustment for changes in the quality of the goods whose prices are tracked over time. When the observed prices correspond to housing, this aspect is of utmost importance. In this case, it is not possible to observe the price of the same home every quarter; in fact, the composition of the sample of homes used to calculate the index is different each quarter, since it is made up of the homes sold in that period. Therefore, if prices are not adjusted for changes in the composition of the sample and the quality of the homes, the estimate of their evolution would not be representative of trends in the real estate market.

One possible solution is to group homes with similar characteristics into strata. In this way, the average price of each stratum is more representative, given its homogeneity. Logically, for a more accurate estimation of changing prices, it is advisable that there be a small number of strata, since the more delimited the housing typology is, the more efficient the adjustment for change in quality and composition will be.

On the other hand, to obtain representative average prices for each stratum with traditional estimators, it is necessary to have a minimum number of observations per stratum each quarter. This requirement would force a lower level of detail of the stratum, reducing the number of characteristics that define it. As a result, homes belonging to the same socioeconomic stratum may not be as homogeneous as would be desirable. Therefore, the IHP uses a mixed method that combines stratification techniques with hedonics, which allows prices to be estimated for each stratum regardless of the number of homes belonging to it during the quarter. In this way, there is a greater number of typologies considered and level of detail in their definition, which significantly improves the adjustment.

Hedonic models are commonly used in the calculation of price indices to control for quality variations in the products that make up the indices. These models aim to explain the value of an asset based on each of its attributes or characteristics. This makes it possible to determine how this value will change when the quantity varies in each of its attributes, and consequently, to predict prices.

The IHP calculation process is presented in further detail below, focusing on the aspects mentioned and, in particular, on the regression model used.

1 Prices

As mentioned earlier, prices per square meter are one of the basic elements in calculating this indicator. However, given the heterogeneity of dwellings, a process must be applied to these prices that guarantees their comparability; therefore, the prices that are ultimately used in the calculation of the IHP are those obtained for each stratum or type of housing after the application of the estimation process

The data received is processed as described in the previous section, in order to obtain housing records that are adapted to the IHP coverage, with complete, refined information. The regression model is applied to this final housing file to obtain the estimated model coefficients, which collect the implicit prices of the housing characteristics. Through them, prices are estimated for each type of housing, which are involved in the calculation of basic indices.

Although prices are estimated quarterly using a sample of home sales occurring within the quarter, the set of housing types remains fixed throughout the year. Each combination of values from the 11 variables included in the regression model (which will be seen in the next section), found in any of the transactions carried out during the reference period of the weightings, constitutes a housing typology. From 2016 onwards, the weighting reference period is the two previous years. In 2017, around 49,000 different housing types were built, which were determined by observing the physical characteristics and location of the homes sold during the years 2015 and 2016.

Prices for each type are estimated quarterly based on information provided by the quarterly sample of sales, regardless of the number of quarterly transactions for each type. Herein lies the main advantage of the method used, hedonic stratification. The model is able to obtain an estimated price for all housing types; this is done by multiplying the vector of characteristics that defines each housing type by the vector of parameters, which varies quarterly and collects the implicit prices of the characteristics.

The elementary indices for each type are calculated using the prices estimated by the model, which, together with their weighting, are used to calculate the aggregate indices.

2 Regression model

The regression model used to calculate the IHP is a semi-logarithmic model, frequently used in this field, where the dependent variable is the Napierian logarithm of the price per square meter of housing.

The explanatory variables, which include the physical characteristics and location of the dwelling, are all categorical, that is, they take a finite number of values.

The table below includes the twelve main effects or explanatory variables of the model, with their corresponding values or categories. Except for the floor variable, which began to be used in 2010, and the European variable, which began to be used in 2026, all have been part of the model since publication began.

Explanatory variables of the regression model used in the IHP

Variable	*Values*	*Categories or values*
New/Second hand	2	A home is considered new when it is the first transfer of ownership.
Type of dwelling	2	Flat or single-family
Garage	2	Yes or No
Storage room	2	Yes or No
Cooperative	2	Yes or No
European	2	Tourist provinces where the buyer is European.
Area	10	<40 m2; [40, 60); [60, 75); [75, 90); [90, 105); [105, 120); [120, 150); [150, 180); [180, 240); ≥240 m2
Floor	6	Basement, ground floor, first floor, second floor, remaining floors and attic apartments.
Classification of provinces	6	Using information on average appraisal value per province, 6 groups of provinces are established.
Size of municipality	4	Capital cities, non-capital municipalities with over 50,000 inhabitants, between 10,000 and 50,000 inhabitants, and with fewer than 10,000 inhabitants.
Tourist municipality	4	Tourist municipalities are considered to be those that concentrate a high number of overnight stays in tourist accommodations and/or a high percentage of second homes. These are classified into one of the first three categories: • sun and beach tourism • rural, nature or inland tourism • cultural, urban or business tourism • other (non-tourist)
Type of environment	14	Grouping of postal codes into 14 types of environments based on census information and the average price per square meter in the previous year.

The first six variables are dichotomous and are obtained directly from the notaries' data file; the last six have been created by grouping the values of some of the received variables. To this end, in some cases, it has been necessary to resort to information from other sources, as described in detail in the section on data processing, where the last four variables have been defined.

In addition, the model includes the most significant dual interactions between these main effects. There are three criteria followed for the selection of the interactions: it must be significant; its contribution to the explanatory power of the model must be as high as possible; and the number of quarterly observations for each combination or pair of possible values of the interaction must be greater than 30. Each interaction adds restrictions to the model; specifically, the average price observed and the price estimated by the model must coincide at each intersection, hence the need to require a minimum number of quarterly observations in the third criterion.

The number of model interactions has remained at nine; however, some may have varied from year to year. As a result, the number of model parameters has also fluctuated annually, between 120 and 150. This way, the prices of thousands of different types are estimated by means of the parameters estimated each quarter, which again highlights the main advantage of the method used.

Both the main effects and interactions may change annually, as the model is reviewed each year. This review consists of:

– Updating the location-based variables with the latest available information from the sources used in their preparation. Thus, from one year to the next the content (and even the number) of categories of the variables may vary: *cluster of provinces*, *size of municipality*, *tourist municipality* and *type of environment*.

– Adding potential new explanatory variables originating from the notaries' database, or created from supplementary information, as occurred in 2010 when the *floor* variable was incorporated, after analysing and grouping the values recorded in that field in the database, or in 2026 when the European variable is incorporated.

– Reviewing the model interactions. To do this, compliance with the established criteria is observed using data from the last four quarters.

Initially, the regression model is applied to the final data, obtained after processing the information. Next, outliers are selected and removed from the residuals of the initial model. It is therefore an objective method that jointly takes into account the price and the values of the 12 explanatory variables of the model. Finally, the final model, which is used to estimate the prices included in the calculation of the index, is weighted in order to correct the heteroscedasticity and also to assign a weight (less than one) to those observations where the value of a certain variable has been imputed. The specification of the model with all the formulation and technical information necessary for price estimation is described in Annex II.

3 Weightings

The weighting structure allows us to establish the importance or weight that each stratum or type of housing has in relation to all the others, based on the expenditure made on the purchase of each type of housing in relation to the total expenditure on housing purchases during the reference period. It is therefore a flow variable —the transactions made— and not a stock variable such as, for example, the number of owned homes in Spain.

The source of information used to obtain the weightings is the same as that used to obtain the prices, since the notarial data allows us to know the types of transactions carried out over time, both in number and value.

Due to the ever-changing nature of the real estate market, it is advisable to frequently update the weighting structure so that it represents the state of the market as faithfully as possible. The chained Laspeyres formula used by the IHP allows the weightings to be updated every year.

As with all indices designed from a chained index scheme, any changes introduced from year to year have some effect on the index's annual rates of change. However, this drawback is offset by the indicator's constant adaptation to market changes. In this ongoing adaptation, in addition to seeking to be current, a certain stability to the weighting structure must also be ensured. In this sense, the more years involved in calculating the weights, the less they will fluctuate and the smaller their effect on annual price variations. Likewise, the variety of housing types will also be greater, which will improve the adjustment in terms of quality and composition .

Initially, and until 2013, three years' worth of transaction data was used to obtain the IHP weighting structure. With the entry into force of the European regulation for the harmonised housing price index, which established the use of a single year in the calculation of weights, as in the HICP, the IHP also began to use a single year to calculate the annual weighting structure. However, since 2016, two years of information have been used in the calculation of weights in order to keep the indicator current and provide it with a consistent structure. Thus, the weightings in force in 2017 were obtained with data on transactions performed during the years 2015 and 2016.

The IHP uses hybrid weights, a term which means when quantities from one period are valued at the prices of another period. The formula for calculating the weighting of a housing type or stratum *e*, in the year *a*, is the following:

$$W_e^a = \frac{Q_e^{(a-1,a-2)} \times \hat{P}_e^{\,4,\,a-1}}{\sum_{\forall e} Q_e^{(a-1,a-2)} \times \hat{P}_e^{\,4,\,a-1}} \qquad a \geq 2016$$

where both prices and quantities refer to the same unit, the square meter of housing. Where:

$Q_e^{(a-1,a-2)}$ represents the average annual amount of square meters of housing belonging to the stratum *e*, sold in the weighting reference period *(a-1, a-2)*, and

$\hat{P}_e^{4,a-1}$ is the price per square meter estimated by the regression model for the stratum *e* in the 4th quarter of the previous year.

Thus, the annual weighting of each housing type represents the expenditure made in the previous two years on the purchase of homes of that type against the total number of homes, valued at prices of the fourth quarter of the previous year.

The reason for using estimated prices instead of collected prices is that in the fourth quarter of the year (*a-1*), possibly not all types of homes have been sold and, therefore, information on observed prices for all strata is not available. Furthermore, according to the chained Laspeyres formula, the prices involved in the calculation of the elementary indices must be the same as those used in the calculation of weights; in both cases, the IHP uses the prices estimated by the model.

The weighting of any aggregate *A*, whether functional or geographical, is obtained as the sum of the weights of the strata that comprise said aggregate:

$$W_A^a = \sum_{e \in A} W_e^a$$

4 Calculation of indices

The general formula used for calculating the IHP is a chained Laspeyres index, analogous to that used in the CPI/HICP. In the case of the IHP, since it is a quarterly indicator, the period used for linking is the fourth quarter of each year and not the month of December.

The use of chained indices allows for the annual updating of weights, as well as the possibility of making methodological changes (such as revising the regression model or including new housing types), unlike what happens with a fixed-base Laspeyres index, in which both the weights and the methodology remain fixed throughout the validity of the base period.

A chained index defines three reference periods:

– **Index reference period or base period.** Period for which the average of the indices becomes equal to 100. Normally this consists of an annual period. In the IHP, since 2017, the base year is 2015, and all published indices use that period as a reference.

– **Weighting reference period**. This is the period that refers to the data used in the calculation of weights.

  Each year the IHP weightings are calculated using the latest available information on the number of home sales carried out in a previous period; these amounts are valued at prices from the fourth quarter of the previous year. The period used to obtain the amounts has varied; from the three years used at the beginning of the publication, it was reduced to a single year in 2013, and since 2016, the two previous years have been used. Thus, and in 2017, the reference period for the weightings is that constituted by the years 2015 and 2016.

– **Price reference period.** The price reference period is the period against which current prices are compared, in other words, the period chosen for the calculation of the elementary indices. This refers to the fourth quarter of the year immediately preceding the current one.

The formula for calculating the elementary indices and the aggregate indices is given below, as well as the general system for calculating the chained indices.

ELEMENTARY INDICES

An elementary aggregate is the component with the lowest level of aggregation for which indices are obtained and whose calculation does not involve weightings; the indices of these aggregates are called elemental indices. In the IHP, the elementary aggregate is the stratum that includes the same type of housing.

The elementary index of the stratum *e* is obtained as the quotient of the price estimated by the model for the homes pertaining to that stratum in the current period, and the price estimated in the fourth quarter of the previous year:

$$_{(4,\,a-1)}I_e^{q,a} = \frac{\hat{P}_e^{q,a}}{\hat{P}_e^{4,\,a-1}} \times 100$$

where,

$\hat{P}_e^{q,a}$ the estimated price per square meter for homes in the stratum *e*, in the quarter *q* of the year *a*, and

$\hat{P}_e^{4,a-1}$ the estimated price per square meter for homes in stratum *e*, in the 4th quarter of the year *a-1*. This price estimate was made using the same regression model[1] used in estimating the price of the numerator.

AGGREGATE INDICES FOR THE FOURTH QUARTER

The index of an aggregation A, whether functional or geographical, is calculated from the elementary indices of the strata pertaining to said aggregation and their corresponding weights, according to the following expression:

$$_{(4,\,a-1)}I_A^{q,a} = \sum_{c \in A} W_e^a \times {}_{(4,\,a-1)}I_e^{q,a}$$

where,

W_e^a the weighting of stratum *e*, as a fraction, valid during the year *a*[2], and

$_{(4,\,a-1)}I_e^{q,a}$ the elementary index of the stratum *e*, in the quarter *q* of year *a*.

The above formula can be equivalently expressed as a quotient of average prices weighted by quantities, the same in numerator and denominator.

$$_{(4,\,a-1)}I_A^{q,a} = \frac{\sum_{e \in A} Q_e \times \hat{P}_e^{q,a}}{\sum_{e \in A} Q_e \times \hat{P}_e^{4,\,a-1}} \times 100$$

[1] The regression model is reviewed annually, so the prices for the fourth quarter of each year *a* must be estimated in two different ways. On the one hand; the model in force in year *a* will be used to calculate the numerator of the elementary indices for the fourth quarter of year *a*. On the other hand, with the revised model, in force in the year following *a+1*, the denominators of the elementary indices for the four quarters of the year *a+1* will be calculated.

[2] Depending on the year, the weightings are obtained with information on sales made in the three previous years (up to 2012), in the previous year (from 2013 to 2015) or in the two previous years (from 2016 onwards).

This is a typical expression in Laspeyres indices, where current prices (numerator) are compared with those of the fourth quarter of the previous year (denominator), thus keeping quantities constant.

Based on the indices for the fourth quarter, the quarterly and cumulative impacts throughout the year are calculated.

INDICES IN BASE 2025

The base 2025 indices are those that are published and are obtained by linking the indices referring to the fourth quarter of the previous year, according to the following expression:

$$ {}_{25}I_A^{q,a} = {}_{25}I_A^{4,(a-1)} \times \left(\frac{{}_{4,(a-1)}I_A^{q,a}}{100} \right) = $$

$$ = {}_{25}I_A^{4,25} \times \left(\frac{{}_{(4,25)}I_A^{4,26}}{100} \right) \times \ldots \times \left(\frac{{}_{(4,a-2)}I_A^{(4,a-1)}}{100} \right) \times \left(\frac{{}_{(4,a-1)}I_A^{q,a}}{100} \right) \qquad a \geq 2026 $$

Based on the 2025 base indices, the quarterly, cumulative (or year-to-date) and annual variation rates are obtained. The first two rates can also be calculated from the indices referring to the 4th quarter.

SERIES LINKING

In the IHP for base 2025, only the index reference period or base period has been changed, from the year 2015 to the year 2025. A rescaling coefficient has been calculated for each series published in base 2015, which converts the indices published in base year 2015, from the first quarter of 2007 to the fourth quarter of 2025, into indices in base year 2025.

This coefficient is that which makes the simple arithmetic average of the indices published in base 2015 for the year 2025 equal to 100.

$$ \frac{1}{4} \sum_{q=1}^{4} {}_{15}I^{q25} \times C_{re-escala} = 100 \quad \Leftrightarrow \quad C_{re-escala} = \frac{100}{\frac{1}{4} \sum_{q=1}^{4} {}_{15}I^{q25}} $$

5 Calculation of variation rates

QUARTERLY VARIATION RATE

The quarterly variation rate of an index is calculated as the ratio of the index of the current quarter and the index of the previous quarter, both in base 2025, according to the following formula:

$$ \Delta^{qa/(q-1)a} = \left(\frac{{}_{25}I^{qa}}{{}_{25}I^{(q-1)a}} - 1 \right) \times 100 $$

where:

$\Delta^{qa/(q-1)a}$ the quarterly rate of price change in the quarter *q* of the year *a*, as a percentage,

${}_{25}I^{qa}$ the index for quarter *q* for the year *a*, in base 2025, and

${}_{25}I^{(q-1)a}$the index for quarter *q-1* for the year *a*. in base 2025.

CUMULATIVE VARIATION RATE

The cumulative variation rate, or year-to-date, is calculated as the ratio between the index of the current quarter and the index of the fourth quarter of the previous year, both in base 2025:

$$\Delta^{qa/4(a-1)} = \left(\frac{{}_{25}I^{qa}}{{}_{25}I^{4(a-1)}} - 1 \right) \times 100$$

where:

$\Delta^{qa/4(a-1)}$ the cumulative rate of price change in quarter *q* of year *a*, as a percentage

${}_{25}I^{qa}$ the index, in base year 2025, for quarter *q* of year *a*, and

${}_{25}I^{4(a-1)}$ the index in base year 2025, for the fourth quarter of year *a-1*.

ANNUAL VARIATION RATE

The annual variation rate is calculated as the ratio between the published indices of the current quarter and the same quarter of the previous year, both in base year 2025:

$$\Delta^{qa/q(a-1)} = \left(\frac{{}_{25}I^{qa}}{{}_{25}I^{q(a-1)}} - 1 \right) \times 100$$

where:

$\Delta^{qa/q(a-1)}$ the annual rate of price change in the quarter *q* of year *a*, as a percentage,

${}_{25}I^{qa}$ the index, in base year 2025, for quarter *q* of year *a*, and

${}_{25}I^{q(a-1)}$ the index, in base year 2025, for quarter *q* of year *a-1*.

6 Calculation of impacts

QUARTERLY IMPACTS

The impact of the quarterly variation of a stratum or set of dwelling strata on the general index is defined as the portion of the quarterly variation of the general index that

corresponds to that stratum or set of strata. Therefore, the sum of the quarterly impacts from all dwelling strata in the IHP is equal to the quarterly variation of the general index.

In other words, the impact that the quarterly price variation of a stratum or set of strata has on the quarterly variation of the general index is the variation that the latter would have undergone if all the prices of all the other strata had not varied in that quarter.

The formula for the quarterly impact of a specific stratum (or set of strata), in quarter *q* of year *a*, is as follows:

$$R_e^{qa/(q-1)a} = \frac{{}_{4(a-1)}I_e^{qa} - {}_{4(a-1)}I_e^{(q-1)a}}{{}_{4(a-1)}I_G^{(q-1)a}} \times W_e^a \times 100$$

where:

${}_{4(a-1)}I_e^{qa}$ is the index, referenced to the fourth quarter of year *a-1*, of stratum *e* in quarter *q* of year *a*,

${}_{4(a-1)}I_e^{(q-1)a}$ is the index, referenced to the fourth quarter of year *a-1*, of stratum *e* in quarter *q-1* of year *a*,

${}_{4(a-1)}I_G^{(q-1)a}$ is the general index, referenced to the fourth quarter of year *a-1*, in quarter *q-1* of year *a*, and

W_e^a is the current weighting in the year *a* of stratum *e*, as a percentage.

As can be seen, the impacts are calculated from the indices referring to the fourth quarter of the previous year (unpublished indices). An alternative expression of the above formula is as follows:

$$R_e^{qa/(q-1)a} = \frac{{}_{4(a-1)}I_e^{qa} - {}_{4(a-1)}I_e^{(q-1)a}}{{}_{4(a-1)}I_G^{(q-1)a}} \times W_e^a \times 100 =$$

$$= \frac{{}_{4(a-1)}I_e^{qa} - {}_{4(a-1)}I_e^{(q-1)a}}{{}_{4(a-1)}I_G^{(q-1)a}} \times \frac{{}_{4(a-1)}I_e^{(q-1)a}}{{}_{4(a-1)}I_e^{(q-1)a}} \times W_e^a \times 100 =$$

$$= \Delta_e^{qa/(q-1)a} \times \frac{{}_{4(a-1)}I_e^{(q-1)a}}{{}_{4(a-1)}I_G^{(q-1)a}} \times W_e^a$$

Therefore, the quarterly impact of a specific stratum e is the product of its quarterly variation rate as a percentage, $\Delta_e^{qa/(q-1)a}$, its weighting in parts per unit, W_e^a, and the

ratio between the stratum index and the general index from the previous quarter, $\frac{{}_{4(a-1)}I_e^{(q-1)a}}{{}_{4(a-1)}I_G^{(q-1)a}}$.

As mentioned above, the sum of the quarterly impacts of all the strata that make up the set of housing typologies of the IHP is equal to the quarterly variation of the general index, as shown below.

$$\sum_e R_e^{qa/(q-1)a} = \sum_e \frac{{}_{4(a-1)}I_e^{qa} - {}_{4(a-1)}I_e^{(q-1)a}}{{}_{4(a-1)}I_G^{(q-1)a}} \times W_e^a \times 100 =$$

$$= \left(\frac{\sum_e {}_{4(a-1)}I_e^{qa} \times W_e^a - \sum_e {}_{4(a-1)}I_e^{(q-1)a} \times W_e^a}{{}_{4(a-1)}I_G^{(q-1)a}} \right) \times 100 =$$

$$= \frac{{}_{4(a-1)}I_G^{qa} - {}_{4(a-1)}I_G^{(q-1)a}}{{}_{4(a-1)}I_G^{(q-1)a}} \times 100 = \Delta_G^{qa/(q-1)a}$$

CUMULATIVE IMPACTS

The cumulative or year-to-date impact of a stratum or set of strata on the general index represents the cumulative variation that the general index would experience if the rest of the strata had not undergone any price variation thus far that year; or in other words, it is the part of the cumulative variation due to said stratum or set of strata.

The formula for the cumulative or year-to-date impact of a given stratum *e* (or a specific aggregation) in quarter *q* of year *a*, is the following:

$$R_e^{qa/4(a-1)} = \frac{{}_{4(a-1)}I_e^{qa} - {}_{4(a-1)}I_e^{q(a-1)}}{{}_{4(a-1)}I_G^{q(a-1)}} \times W_e^a \times 100 =$$

$$= \frac{{}_{4(a-1)}I_e^{qa} - 100}{100} \times W_e^a \times 100 = \Delta_e^{qa/4(a-1)} \times W_e^a$$

where:

$\Delta_e^{qa/4(a-1)}$ is the cumulative rate of change of the stratum *e*, in the quarter *q* of the year *a*, as a percentage, and

W_e^a the weighting of stratum *e* in force as of year *a*, as a percentage.

Therefore, the year-to-date impact is the product of the cumulative rate of change (as a percentage) and the weighting (as a percentage). This means the impact of a stratum or group of strata on the cumulative variation of the general index will be greater when its cumulative variation and its weighting are greater.

In this case, the sum of the cumulative impacts of all strata is equal to the year-to-date change in the overall index:

$$\sum_i R_e^{qa/4(a-1)} = \sum_e \left({}_{4(a-1)}I_e^{qa} - 100 \right) \times W_e^a =$$

$$= \sum_e {}_{4(a-1)}I_e^{qa} \times W_e^a - 100 \sum_e W_e^a = {}_{4(a-1)}I_G^{qa} - 100 =$$

$$= \frac{{}_{4(a-1)}I_G^{qa} - 100}{100} \times 100 = \Delta_G^{qa/4(a-1)}$$

IX Dissemination

The IHP is published quarterly, approximately 70 days after the end of the index's reference period, according to the INE's publication schedule. In the fourth quarter of each year, the INE publishes the calendar with the exact dates for the publication of statistical operations for the following year.

Each quarter, the INE publishes the IHP press release, which highlights the most significant price variations and presents the main results.

INEbase is the system used by the INE to store and disseminate all the statistical information it produces online, in electronic format. The INE website provides access to the online IHP database, which offers information on price indices, variation rates, and weightings.

Data is published for the country as a whole, the 17 autonomous communities and the Autonomous Cities of Ceuta and Melilla, which allows comparisons to be made between the evolution of prices in the different regions.

Regarding functional disaggregation, information is provided for new and second-hand homes, at the national level. Responding to user demand for more detailed information, since 2010, the breakdown by new and second-hand homes is also available by autonomous community.

The following data is published quarterly:

– indices in base 2025;
– 2025-2007 series in base 2015;
– 2016-2007 series in base 2007;
– quarterly variation rates;
– year-to-date variation rates;
– annual variation rates;
– quarterly impacts (for new and second-hand homes);
– year-to-date impacts (for new and second-hand homes).

In addition to the quarterly results, each year the annual averages of the indices and the variations of the annual averages are provided for each published series. The annual weighting structure is also made available to users.

The IHP data are definitive from the first time they are published, and are therefore not subject to revision.

The standardised methodological report, accessible from the INE website, contains the survey metadata, which helps for a better understanding and interpretation of the results.

Annex I. Glossary of terms

- **ANCERT**. *Agencia Notarial de Certificación*, is Spain's notarial certification agency and association of Spanish notaries, created by the General Council of Notaries with the aim of modernising and placing the body of notaries of Spain at the technological forefront, as well as the different agencies in the notary community.
- **Cell**. Combination of the possible values of the variables or characteristics (main effects of the regression model) that define a specific type of housing.
- ***Consejo General del Notariado***. The General Council of Notaries, an entity that coordinates the notarial associations in Spain. It manages the database relating to real estate transactions (computerised index of notaries), which is used for the calculation of the IHP.
- **Housing cooperative**. It is the group of people who, fulfilling the general requirements of the cooperative (drafting of by-laws, registration in the *Registro de Sociedades Cooperativas* [Register of Cooperative Societies], formation of the bodies by which it is governed, accounting, etc.), come together to participate in a common project, carrying out all the necessary activities (search for land, search for a financial entity to finance the construction, commissioning the architect, drafting incorporation contracts, construction contract, housing allocation contracts, etc.) to obtain accommodations and/or premises and complementary facilities, for themselves or for the people who live with them.
- **Main effect**. Explanatory variable of the regression model.
- **Mortgage**. The right that the lender acquires against the borrower in case of non-payment of the latter's obligations, and which is exercised over the asset that appears as a guarantee or collateral. In the case of a mortgage loan for a home, the mortgaged property is usually the home that was purchased.
- **Interaction**. Explanatory variable of the regression model, obtained as a combination of other explanatory variables (main effects) of the model.
- **Hedonic regression model**. Hedonic pricing models analyse the price of an item based on its multiple characteristics, estimating the implicit price of each.
- **Flats**. These are dwellings that are part of a building of two or more floors or levels, all having a common access to them from the public road. Whenever there are private areas and common areas, there is a special form of co-ownership established as horizontal property.
- **Appraisal**. An appraisal is an estimate of the market value of an asset based on various determining factors; in the case of homes, these factors can include size, location, age, etc. Most home appraisals are commissioned by a bank for the purpose of granting a mortgage loan to purchase the property, and are typically carried out by appraisal companies.
- **Deed value**. To notarise something is to record a grant or an event in a public deed and in a legal manner.
- The deed value of a home is the value stated in the public deed of sale and is, therefore, the official price of the home.
- **Dwelling**/**Housing**. Any structurally separate and independent enclosure that, by the way it was built, rebuilt, transformed or adapted, is designed to be used by people and is part of a building.

- **Second-hand housing**. The classification of homes as new or second-hand is based on the order of the transfer. Thus, when there is more than one transfer in the public deed, the property is classified as second-hand.
- **Open-market housing**. That which is not considered social housing.
- **New home**. The classification of homes as new or second-hand is based on the order of the transfer. Thus, when it is the first transfer in the deed of sale, normally made by the developer or builder in favour of the first purchaser, the home is classified as new.
- **Social housing**. That which has received any type of subsidy for its construction, regardless of the agency that grants it, and where limitations of surface area and price are taken into account. Excluded from this definition are homes that have already exceeded the expiry date of said subsidy and those others that, although they have not exceeded it, appear with a net assets value defined in the Ministerial Order of Economy and Finance. These last two considerations give the dwelling the category of open-market housing.
- **Single-family home**. A dwelling located on an independent plot of land, and which serves as a residence for a single family.

Annex II. Regression model

Specification of the regression model

The regression model used to calculate the estimated price per square meter in the preparation of the IHP is specified below. For each quarter *q* it is assumed that the price per square meter, *P*, of home *i* belonging to cell *c*, is:

$$l_{i,c}^{q} = \ln P_{i,c}^{q} = \mathbf{x}_{c}^{'}\boldsymbol{\beta}^{q} + \varepsilon_{i,c}^{q} \qquad (1)$$

where:

$\mathbf{x}_{c}^{'}$ is a dimension vector *(1 x p)* whose elements are equal to either *0* or *1*, based on the characteristics that define cell *c*, as far as main effects and interactions are concerned,

β^{q} is a vector *p* of unknown parameters, of dimension (*p x 1*), and

$\varepsilon_{i,c}^{q}$ is the random component of the model, in quarter *q*.

The vector β^{q} defines the proportional effect on the expected price per square meter of housing of the *p* dichotomous variables included in $\mathbf{x}_{c}^{'}$, in quarter *q*. The *p* unknown parameters include the constant and the parameters of the dichotomous variables associated with the main effects and interactions of the model.

For each possible category *r* that has a main effect, the model includes (*r-1*) parameters. If the interaction has (*r x s*) possible combinations of values, *(r-1) x (s-1)* parameters enter the model. All in all, the 2008 model consists of 157 parameters.

The disturbances $\varepsilon_{i,c}^{q}$ verify:

$$E[\varepsilon_{i,c}^{q}] = 0, \;\; Var[\varepsilon_{i,c}^{q}] = \sigma_{q}^{2}, \;\; Cov[\varepsilon_{i,c}^{q}, \varepsilon_{j,d}^{q'}] = 0, \;\; \forall (q,i,c) \neq (q',j,d) \qquad (2)$$

Once the model, which will be in effect for one year, has been defined, the vector $\boldsymbol{\beta}^{q}$ must be estimated each quarter, with the information available. To do this, model (1) is formulated in matrix notation, as follows:

$$\mathbf{L}^{q} = \mathbf{X}^{q}\,\boldsymbol{\beta}^{q} + \boldsymbol{\varepsilon}^{q} \qquad (3)$$

where:

$\mathbf{L}^{q}$ is a vector of dimension (n^{q} *x 1*) that contains the n^{q} elements $l_{i,c}^{q}$ of quarter *q*. In other words, it contains as many rows as there were home sales in quarter *q* (n^{q}),

$\mathbf{X}^q$ is a dimension matrix ($n^q \times p$), whose elements are equal to either *0* or *1*. In this matrix, each row represents a dwelling and each column contains one of the *p* characteristics that define said housing, in quarter *q*,

β^q is a vector of dimension (*p x 1*) that contains the *p* unknown parameters of quarter *q*. It includes the constant and parameters of the dichotomous variables associated with the main effects and interactions of the model, and

$\boldsymbol{\varepsilon}^q$ is a vector of dimension (n^q *x 1*) that contains the n^q random disturbances of the model in quarter *q*. This disturbance vector verifies:

$$E[\boldsymbol{\varepsilon}^q] = \mathbf{0}, \quad Var[\boldsymbol{\varepsilon}^q] = \sigma_q^2 \, \mathbf{I}_{n^q \times n^q} \tag{4}$$

The OLS (ordinary least squares) [3] estimator of β^q is:

$$\hat{\boldsymbol{\beta}}^q = (\mathbf{X}'^q \mathbf{X}^q)^{-1} \mathbf{X}'^q \mathbf{L}^q \tag{5}$$

and its variance is:

$$Var[\hat{\boldsymbol{\beta}}^q] = \sigma_q^2 (\mathbf{X}'^q \mathbf{X}^q)^{-1} = \mathbf{V}^q \tag{6}$$

where the matrix V^q has dimension (*p x p*).

The parameter vector $\hat{\beta}^q$ varies according to the data for each quarter and is the fundamental element used to estimate the average price per cell.

Price estimate

To prepare the IHP, it is necessary to have the estimated average price in each quarter that corresponds to each cell. This estimated price is obtained from the price in formula (1); thus, the estimated price in cell *c*, in quarter *q* is as follows:

$$\hat{P}_c^q = \exp(\mathbf{x}_c' \hat{\boldsymbol{\beta}}^q) \tag{7}$$

[3] The derivation of these results can be found, for example, in the works of Peña (1993, 2002), Draper (1998) and Montgomery (2001)

The problem with this estimator, which has a simple expression, is that it has a high bias. To correct this bias, the estimator proposed by El-Shaarawi and Viveros (1997) is used:

$$\hat{P}_c^q = \exp\left\{ \mathbf{x}_c^{'}\hat{\boldsymbol{\beta}}^q - \frac{1}{2}\mathbf{x}_c^{'}\hat{\mathbf{V}}^q\mathbf{x}_c + \frac{1}{2}\hat{\varphi}^q\hat{\sigma}_q^2 \right\} \qquad (8)$$

where

$$\hat{\varphi}^q = 1 - \frac{\hat{\sigma}_q^2}{2(n^q - p)} - \frac{\hat{\sigma}_q^4}{3(n^q - p)^2} \qquad (9)$$

Estimator (8) substantially corrects the bias of estimator (7), assuming the normality of the errors $\varepsilon_{i,c}^q$.

To obtain the variance estimate that appears in the above expressions, the residuals $e_{i,c}^q$ are defined as the difference between the natural logarithms of the observed price and the estimated price, that is:

$$e_{i,c}^q = l_{i,c}^q - \mathbf{x}_c^{'}\hat{\boldsymbol{\beta}}^q \qquad (10)$$

The variance σ_q^2 is estimated using the residual mean squares:

$$\hat{\sigma}_q^2 = \frac{1}{n^q - p} \sum_{c,i}^{n^q} \left(e_{i,c}^q \right)^2 \qquad (11)$$

Heteroscedasticity correction

When applying the regression model to the data for each quarter, the residuals show signs of heteroscedasticity for one of the variables included in the model, as well as for the set of observations that have imputed values. Therefore, the model must be transformed to make it homoscedastic.

In heteroscedastic models, the variance of the residuals is not constant, since:

$$\text{var}[\boldsymbol{\varepsilon}^q] = \sigma_q^2 \left(\mathbf{W}^q \right)^{-1} \qquad (12)$$

where W^q is a diagonal matrix with dimension (n^q x n^q) and all its positive elements.

Given that:

$$\mathrm{var}\left(\left(\mathbf{W}^q\right)^{1/2}\boldsymbol{\varepsilon}^q\right)=\sigma_q^2\,\mathbf{I}_{n^q\times n^q} \qquad (13)$$

the model can be made homoscedastic by pre-multiplying by the matrix $\left(\mathrm{W}^q\right)^{1/2}$; that is:

$$\left(\mathbf{W}^q\right)^{1/2}\mathbf{L}^q=\left(\mathbf{W}^q\right)^{1/2}\mathbf{X}^q\boldsymbol{\beta}^q+\left(\mathbf{W}^q\right)^{1/2}\boldsymbol{\varepsilon}^q \qquad (14)$$

The estimator $\hat{\boldsymbol{\beta}}^q$ that minimises the weighted sum of the squares of the errors is as follows:

$$\hat{\boldsymbol{\beta}}^q=(\mathbf{X}^{q\prime}\mathbf{W}^q\mathbf{X}^q)^{-1}\mathbf{X}^{q\prime}\mathbf{W}^q\mathbf{L}^q \qquad (15)$$

and its variance is:

$$Var[\hat{\boldsymbol{\beta}}^q]=\sigma_q^2(\mathbf{X}^{q\prime}\mathbf{W}^q\mathbf{X}^q)^{-1}=\mathbf{V}^q \qquad (16)$$

The rationale for introducing the matrix W^q in the model assumes that if the variance of the data is different for the different categories of a variable, the observations belonging to the categories with lower variance are more reliable and should have more weight in the weighted sum of squares of the errors than those with higher variance (on average, the lower their variance, the less they will deviate from the mean value that should be estimated). Something similar happens with complete observations (without imputed values) which, in general, have less variance than those in which it has been necessary to impute a value.

The elements of matrix W^q are determined from the analysis of the model's heteroscedasticity. Thus, to correct this, in formula (8) for the estimated average price per cell, the new expressions $\hat{\boldsymbol{\beta}}^q$ and V^q must be used, and the residual variance of the corrected model will be obtained from the weighted residuals:

$$e_{i,c}^q=\sqrt{w_i^q}\;\left(l_{i,c}^q-\mathbf{x}_c^{\prime}\hat{\boldsymbol{\beta}}^q\right) \qquad (17)$$

where W_i^q is the element (*i,i*) of the matrix W^q.

ALLOCATION OF HETEROSCEDASTICITY WEIGHTS BY THE IMPUTATION OF VALUES

In the notaries' database, most of the variables that are directly or indirectly involved in the model are complete. However, when this is not the case, it is necessary to impute the values that are not reported.

Since the variability of the residuals in observations where the value of one of the model's explanatory variables has been imputed is greater than in the set of those are completed in the data file, the complete observations are assigned a weight equal to one in the regression, while those with imputed values are assigned a lower weight.

The mean squared error (MCE) is used to calculate these weights: for the set of observations that have an imputed value from a set of main effects *U*, the corresponding weight is obtained as a quotient of the mean squared error of the complete model, with all the main effects (MCE_T^q) and the mean squared error of the model that excludes the main effects and interactions associated with the set *U* of imputed variables (MCE_{T-U}^q). To calculate these terms, MCE_T^q and MCE_{T-U}^q , the complete set of observations *C* is used; in other words, all observations from the quarter that have an imputed value from any of the main effects of the model are excluded.

Since the complete model has a lower residual variance than the submodel that excludes one or more main effects (and their corresponding interactions), it is verified that:

$$0 \le \lambda_u^q = \frac{MCE_T^q}{MCE_{T-U}^q} \le 1 \qquad (18)$$

where:

MCE_T^q the mean squared error of the model that includes all main effects and interactions applied to the set *C* of observations without imputed values in quarter *q,* and

MCE_{T-U}^q the mean squared error of the model that excludes the main effects *U* in which some value has been imputed, applied to the set *C* of observations without imputed values in quarter *q.*

It is logical to assume that those observations that have been subjected to an imputation process have a larger error variance (or a smaller weight in the model adjustment). To take this into account, a heteroscedastic model of type (12) is considered, where the weights W_i are defined as follows:

– If observation *i-th* of quarter *q* has complete information, then $\mathrm{W}_i^{impu} = 1$.

– If observation *i-th* of quarter *q* is incomplete and lacks the data corresponding to the set of explanatory variables *U*, then $W_i^{impu} = \lambda_U$.

As many weights λ_U will be calculated as possible cases or combinations of imputed main effects that may have occurred in the quarter. In the simplest case, only the value of one main effect will be imputed in the model, and it will only be necessary to calculate one weight other than one.

HETEROSCEDASTICITY CORRECTION BETWEEN CATEGORIES

The analysis of the residuals from the previous weighted model may necessitate a final heteroscedasticity correction existing in some of the explanatory variables. To make this correction, the steps to follow are:

Where C_1, C_2,…,C_u the *U* possible values of the variable on which the heteroscedasticity will be corrected:

1. The previous weighted model is adjusted.
2. The residuals of the previous weighted model are calculated. $\hat{e}_i^q \;\; i = 1,...,n^q$.
3. The estimated variances of the residuals within each category are obtained:

$$S_r^2 = \frac{1}{n_r - 1} \sum_{i \in C_r} \left(\hat{e}_i - \bar{\hat{e}}_r \right)^2, \;\; n_r = card\left(C_r \right), \;\; \bar{\hat{e}}_r = \frac{1}{n_r} \sum_{i \in C_r} \hat{e}_i \qquad (19)$$

4. Thus defining

$$w_i^{cate} = \frac{\min\left[S_1^2,...,S_U^2 \right]}{S_1^2} \;\; \forall i \in C_1, \;\; ....., \;\; w_i^{cate} = \frac{\min\left[S_1^2,...,S_U^2 \right]}{S_U^2} \;\; \forall i \in C_U$$

The joint correction for heteroscedasticity is performed with a weighted model, defining the weight of each observation as the product of the two coefficients calculated in the previous section and in this one below:

$$w_i^{hete} = w_i^{impu} \times w_i^{cate}$$

where w_i^{impu} is the coefficient assigned to the observation *i-th*, taking into account the imputed values it has, and w_i^{cate} the coefficient assigned to the value or category of the variable that presents problems of heteroscedasticity, in the observation *i-th*.

The matrix W^{q} of the homoscedastic model (14), is a diagonal matrix, of dimension $n^{q} \times n^{q}$, where the elements of the main diagonal are the coefficients w_{i}^{hete}.