

Rental Housing Price Index (RHPI)

Technical Project



June 2023

Contents

1	Introduction	3
2	Objective	4
3	Scope of the research	5
4	Information sources	5
5	Calculation methodology	7
6	Dissemination	13

Rental Housing Price Index (RHPI)

1 Introduction

Spain is one of the European Union (EU) countries where the highest proportion of the population lives in the home that they own. That is why a large part of the efforts of public and private statistics have traditionally focused on understanding the sales market, to the detriment of the rental market.

In this regard, the *Instituto Nacional de Estadística* (INE), the National Statistics Institute, implemented the Housing Price Index (HPI) in 2008, with the aim of establishing a measurement of the evolution of housing transaction prices, based on international standards that allow for comparisons between EU countries.

However, with regard to the rental housing market, in 2020 there was no specific operation in the State Administration Statistical System that allowed for information on the evolution of rental prices. This is why the INE began to study the different alternatives to design and produce an indicator on the evolution of rental housing prices.

In order to make the most of existing administrative records, the production of this indicator began with information from the State Tax Administration Agency (AEAT), based on Form 100 of the annual Personal Income Tax (IRPF) return. It is important to note that the AEAT database does not contain information on housing in Comunidad Foral de Navarra or País Vasco, and its periodicity limits the analysis to a single data point per year, which prevents carrying out current market studies. Therefore, this statistic is published as experimental statistics and cannot be incorporated, at least for the moment, into the set of official statistics that make up the National Statistical Plan.

This source contains sufficient information to undertake the production of statistics on the evolution of rental prices. The basic information is the amount received by the reporting unit, which is dwelling's tenant, but it has other variables which are equally useful to meet this objective, which are complemented with land registry data. In this way, the exact location of the dwellings and their physical characteristics, such as their surface area, type, or age, are available.

In 2022, the RHPI was published for the first time as an experimental statistic, providing data from 2011 to 2020. Since this was a first foray into the housing rental sector, and only dwellings rented in consecutive years were taken into account in the calculation of the indices. In other words, for the calculation of rental price variations in a given year, the properties involved were those that had been rented both that year and the previous year¹. Depending on the year, common housing accounts for between 83 and 90% of the total declared rentals.

From the second publication onwards, referring to the year 2021, all dwellings for which the reporting parties have received income in the reference year are taken into account; therefore, the results include homes that were already rented in 2020 and that maintain the same contract, those that were rented in 2020 but that have undergone at least one change of contract in 2021, and a part that were incorporated into the index for the first time, as had not been listed as rented in 2020. For the latter, it was necessary to develop a specific methodology that would allow for an estimated evolution of their prices, despite not having their value in the previous year.

¹ Also included were dwellings rented in year a, not rented in a-1 but in a-2.

In short, the data from the RHPI since 2021 encompasses the entire rental market, which results in more accurate and representative results.

Another important new feature of the RHPI is the publication of the evolution of prices distinguishing between existing contracts and new contracts, which allows an analysis of both markets separately. To do so, the variable included in the AEAT databases since 2019 referring to the date of the rental contract has been used.

It is important to note that the rented homes considered *new contracts* are those whose contract was signed in 2021. For a proper interpretation of the results, it should be noted that, since the declared income refers to the full year¹, it may refer to different contracts (and/or tenants). Therefore, in cases where the dwelling has been rented with two different contracts in the same year, the annual price used for the calculation of the index includes both, so the calculated variation only reflects a part of the effect of the change of contract.

2 Objective

The Rental Housing Price Index (RHPI) aims to measure the annual evolution of rental prices for dwellings used as a primary household residence. The source of information used for its calculation is the database provided by the AEAT, which comes from model 100 of the IRPF declarations. Each year, the information declared for tax purposes is collected on real estate properties for which rental income from housing for habitual use has been declared.

The published indices allow for annual price variations at different levels of geographical disaggregation and according to different characteristics of the rented dwellings. Thus, comparisons can be made of the evolution of rental prices in autonomous communities, provinces, and some municipalities and districts, and according to different characteristics of rental housing such as surface area or type of building. Also, as a new feature, the updated publication from 2023 offers results by the age of the lease term.

It should be noted that the calculation of rental price levels is not within the scope of the project, as this is already handled by the State Reference System for Rental Housing Prices, which provides median and percentile values of rental income at a very detailed geographical level.

Finally, although the main objective of the RHPI is to measure the evolution of rental housing prices, information is also provided on the distribution of housing rental expenditure by households (natural persons) for different levels of geographical and functional disaggregation. This information is included in the weightings that represent the structure of household spending on rental housing, according to location, characteristics of the rented dwellings and the age of the contract.

¹ When the property has been rented for less than a year, the Tax Studies and Statistics Service of the AEAT annualises the rental income taking into account the number of days it has been rented.

3 Scope of the research

POPULATION SCOPE

The population scope of the operation consists of real estate properties that include some premises used as housing, for which rental income from a main residence has been registered in form 100 of the IRPF (Personal Income Tax). Those who have generated income through kinship are excluded.

GEOGRAPHICAL SCOPE

The geographical scope, as this is an exploitation of tax data sources, consists of the common fiscal territory, that is, the entire national territory except the Comunidad Foral de Navarra and País Vasco. A goal for the future is to try to incorporate the information corresponding to these two autonomous communities. Therefore, they exclude:

- real estate owned by taxpayers whose tax address is within the common tax territory, but whose real estate is located outside of it (regional fiscal territory, foreign territory or unknown location); and
- real estate located in the common territory but declared by owners who are tax residents in the above regions or abroad.

TEMPORAL SCOPE

The reference period is annual.

The first publication, in February 2022, released the results for the years between 2011 and 2020.

The updated publication, published in 2023, offers results from 2021 that incorporate more homes into the calculation and a greater breakdown in the published tables.

In the future, it is planned that the annual results for the previous two years will be published in the first quarter of each year.

4 Information sources

The main source of information comes from Personal Income Tax (IRPF) returns relating to leases of habitual residences. This information is supplemented with land registry data on the physical characteristics of the properties, such as surface area, type or age.

The exploitation of both sources is carried out by the Tax Studies and Statistics Service of the Spanish Tax Agency and one of its results constitutes the State System of Reference Indices for Rental Housing Prices.

The criteria used by the AEAT to determine the homes that are part of the calculation of this indicator, according to its methodology¹, are as follows:

Urban leased properties that are not related to the economic activities conducted by the taxpayer, and whose use is not exclusively the declarant's main residence, with the following criteria:

The owners of the property have declared rental income.

2. Properties intended for habitual residence (Reduction for rental of properties intended for housing included on form 100).

3. Leases in which there is no family relationship between the lessor and the lessee (according to data collected in form 100).

4. In order to avoid considering the case of properties with numerous dwellings without horizontal division (where it is unknown to which dwellings and areas to assign the declared rents), those properties whose constructed area is three times greater than the area of the registered property in which they are located have been excluded.

5. The property data regarding its ownership and generated income must meet certain validations, so that it is not considered erroneous, suspicious or highly atypical.

6. Properties without a geographical assignment based on the land registry reference are excluded.

The information compiled under these criteria for the State Reference System for Rental Housing Prices is the one that the AEAT provides to the INE. Therefore, for the calculation of the RHPI, the properties selected as rented dwellings are those which have some premises used as housing, according to the construction typology of the largest surface area within the property and that have been rented for habitual residence. Furthermore, the property data regarding its ownership and generated income must meet certain validations, so that it is not considered erroneous, suspicious or highly atypical.

In these validations, the INE has established its own limits on the permitted surfaces (the same as those used in the HPI). Rental prices per square meter have also been limited, and housing types corresponding to homes and economic activities or services have been eliminated. Finally, properties whose rental price has been reduced by more than half or has increased by more than twice from one year to the next are excluded. This last criterion may, in some cases, lead to the exclusion of dwellings whose evolution has been reflected in the records; however, it is a way to eliminate possible errors and exclude circumstances such as non-payments, which do not imply a decrease in the rental price.

In the first publication of the experimental statistics, only the dwellings that had been rented each year and the immediately preceding year were taken into account². Since the dwellings were the same in both years, no quality adjustments were necessary. From 2021 onwards, however, all homes for which rental income was declared in that year have been included, regardless of whether they were rented in the previous year or two years prior; this has been made possible by the development of a regression model,

¹ State System of Reference Indices for Rental Housing Prices, statistical exploitation of tax sources. Methodology

² Also included were dwellings rented in year a, not rented in a-1 but in a-2.

which allows the prices of rented homes to be estimated by taking into account their location and physical characteristics. The specification of the model and its use within the index calculation formula is detailed in the following section.

5 Calculation methodology

The RHPI calculation system is based on the combination of two basic elements: the prices of rental housing, and the weightings, or relative importance of each dwelling type according to the expenditure that households make on renting that type with respect to the total expenditure on rent. Both the location and the characteristics of the dwellings are taken into account to establish the types of housing whose prices will be tracked over time.

These two elements to obtain the RHPI is combined using the **formula for the chained Paasches index**.

It is important to point out that, unlike the Housing Price Index (HPI), where the sample of home sales changes completely every quarter, in the RHPI, most of the homes rented in a year were also rented the previous year, so it is not necessary to apply any kind of quality adjustment to them. However, in order to be able to use the total number of rented homes in the current year, regardless of whether or not the data from the previous year is available, an econometric model had to be developed that could estimate rental prices in each period in those dwellings where the rental value in the previous year is not available. This model, similar to that used in the HPI, takes into account the location of the dwelling, its physical characteristics such as surface area, type or year of construction, as well as the age of the lease agreement.

Due to the complexity involved in this type of methodological implementation, the first publication of the experimental statistics only used the dwellings that had been rented both in the current year and the previous year, regardless of whether they had changed tenants during that period. From 2021 onwards, the RHPI calculation includes all rented homes in the current year that are part of the study area. These can be divided into two groups: homes where the rental value is available for both the current year and the previous year (or two years prior), and the rest of the rented homes where the rental value for the previous year is not available. While the first block will consider the observed/declared rental prices, the second block will use the prices estimated by the regression model. However, the formulas for calculating the indices will use both the declared prices and those estimated by the model.

In total, in 2021 more than 2.1 million rented dwelling were used in the calculation of the RHPI, around 10% of which were able to be incorporated into the calculation thanks to the use of the regression model.

5.1 PRICES

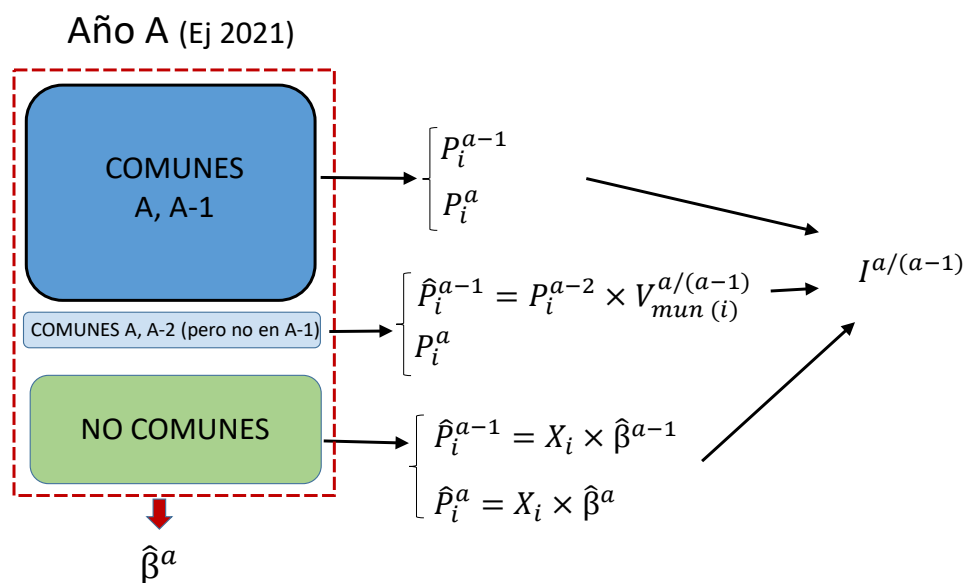
The rental prices used in the RHPI calculation are those obtained by the Tax Studies and Statistics Service of the Spanish Tax Agency from the rental income declared. Each year, properties that meet the established criteria, described in section 4, are selected.

This set is separated into two groups: common dwellings, which appear in the file for the current year as well as the previous year (or two years ago), therefore meeting the criteria in both years; and non-common dwellings, which are in the current year but not in the two previous years. Until 2020, only the homes in the first block were used.

Within the first group, for the homes in the file with the selected rents in the current year that are not in the file of the previous year but were two years ago, the price of the previous year is estimated by imputing the annual variation of the municipality to which the home belongs for the previous year. Thus, the observed biennial variation of $a-2$ to a is divided into two sections, allocating the annual variation of $a-2$ to $a-1$ the one published the previous year for the municipality to which the dwelling belongs and the remaining one, until accumulating the observed biennial variation, is assigned from $a-1$ to a .

In the second block, with non-common housing, it is necessary to estimate the prices, at least from the previous year (since they are not available); for this, a regression model is used which is described in the following section. These estimated prices will be used in the same way as the observed/declared prices of the common housing block, in the formula for calculating the indices.

The following diagram represents the subsets of rented dwellings that are used in the calculation of the HPI from 2021 onwards, along with their prices.



It is worth noting that for the estimation of the annual coefficients that will be used in the price estimation of the second block, both in the current year and in the previous year, all the homes rented in each year are taken into account.

5.2 REGRESSION MODEL

The regression model developed for the RHPI is used exclusively for estimating rental prices for homes where no value is available in either of the previous two years (non-common). This 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 the dwelling.

The explanatory variables include the physical and location characteristics of the dwelling, as well as the age of the contract¹. All of them are categorical, except the one relating to the surface area of the dwelling, which means they take a finite number of values. The table below includes the ten explanatory variables of the model, with their corresponding values or categories.

Explanatory variables of the regression model used in the RHPI in 2021

Variable	Values	Categories or values
CPRO	48	Province code, all except Navarra and the provinces of País Vasco
Size of municipality	14	Variable that takes into account the population of the municipality (with limits of one million inhabitants, 500,000, 300,000, 200,000, 100,000, 50,000, 20,000, 10,000 and 1,000 inhabitants) and whether or not it is a provincial capital
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: <ul style="list-style-type: none"> • sun and beach tourism • rural, nature or inland tourism • cultural, urban or business tourism • other (non-tourist)
CLDIS_Income	10	Grouping of census tracts into 10 classes according to their income level per consumption unit: <€30,000, (27,500, 30,000], (25,000, 27,500], (22,500, 25,000], (20,000, 22,500], (17,500, 20,000], (15,000, 17,500], (10,000, 12,500], <=10,000
ANT_Contract	11	Lease agreement year: prior to 1970, 1970-79, 1980-89, 1990-99, 2000-09, 2010-2015, 2016, 2017, 2018, 2019, 2020-21
ANT_Dwelling	7	Year of construction of the dwelling: prior to 1970, 1970-79, 1980-89, 1990-99, 2000-09, 2010 onwards, NC (not known)
Building type	6	Building typology: <ul style="list-style-type: none"> residential block of flats in an open block, residential block of flats in a closed block, rural residential dwelling, single-family residential, detached house single-family residential, row houses or closed block Don't know/No reply
Garage	2	Yes or No
Floor	6	Ground floor, first floor, second floor, third floor, fourth floor, fifth through tenth floor, the floor above the tenth floor, NC
Ln_Surface area		Numeric variable, Napierian logarithm of the surface area of the house

¹ The contract date is available from 2019 onwards. If there are multiple holders/declarants and the contract year is different for each holder, the most recent one is used.

The adjusted R^2 of the model has been approximately 0.76, both in 2020 and 2019; therefore, 76% of the variability in rental prices is explained by the explanatory variables included in the model.

With the file of selected rental properties in the current year a , the regression coefficients $\hat{\beta}^a$ are estimated, which reflect the implicit prices of the characteristics of year a . In this way, all dwellings rented during the year that meet the established criteria are included in the parameter estimates, regardless of whether they have been rented in previous years. The same process is followed with the rentals selected the previous year, to obtain the coefficient vector: $\hat{\beta}^{a-1}$.

For dwellings rented this year that were not in the records of the previous two years, their prices will be estimated according to the value that the explanatory variables of the model have for each of those dwellings; thus, it will depend on their location, their physical characteristics as well as the age of the contract, all of which will be included in a vector of characteristics associated with each dwelling, x_i . The characteristic vector is combined with the parameter vector for each year, $\hat{\beta}^a$ y $\hat{\beta}^{a-1}$, to obtain estimates of rental prices for both the previous and current year:

$$\ln(\widehat{alq_m2}_{t,a}) = x_i \widehat{\beta}_a \text{ and } \ln(\widehat{alq_m2}_{t,a-1}) = x_i \widehat{\beta}_{a-1}$$

Although the declared rent for the current year is available, it is advisable to use estimated prices for both periods to avoid the effect that omitted variables may cause in the regression model.

5.3 CALCULATION OF INDICES

The general formula used for calculating the RHPI is an annually chained Paasche index. This refers to the relationship between the sum of rental prices in the current year a and the sum of the prices of the same rentals in the previous year $a-1$ where the sums are weighted by the amounts in the current year:

$${}_{a-1}I^a = \frac{\sum_e Q_e^a \times P_e^a}{\sum_e Q_e^a \times P_e^{a-1}} \times 100$$

It can equivalently be expressed as a weighted arithmetic mean of simple price indices, where the weights reflect the value of rents in the current period calculated at prices from the previous year.

$${}_{a-1}I^a = \sum_e W_e^a \times I_e^{a/a-1} \text{ con } W_e^a = \frac{Q_e^a \times P_e^{a-1}}{\sum_e Q_e^a \times P_e^{a-1}} \text{ e } I_e^{a/a-1} = \frac{P_e^a}{P_e^{a-1}} \times 100$$

where

W_e^a is the weighting of stratum e , as a fraction, valid during year a

The Paasche index, along with the Laspeyres and Fischer indices, are the most widely used to calculate housing price indices, both for buying and renting; the Laspeyres index is possibly the most commonly used due to the difficulties in obtaining data on the amounts or expenses in the current period, which is not the case in the RHPI.

1. Basic 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 or simple indices. In the RHPI, the elementary aggregate is the stratum, which is defined as a combination of variables whose values reflect both the physical characteristics of the rented homes as well as their location, and from 2021 onwards, the age of the lease agreement. The variables used to construct the strata are:

- Autonomous Community
- Province
- Municipality or group of municipalities¹
- District²
- Type of building (block of flats or single-family)³
- Housing size (with six intervals: ≤ 60 m², (60-75], (75-90], (90-120], (120-150] and >150 m²)
- Age of lease agreement: new agreement/existing agreement.

Each combination of values of the above variables found in any of the rented dwellings in the current year forms a stratum. Each year, average rental prices and rental spending weights are calculated for each stratum.

The elementary index of the stratum e is obtained as the quotient of the mean geometric price in that stratum in the current period, \bar{P}_e^a , and its mean geometric price in previous year \bar{P}_e^{a-1} :

$${}_{a-1}I_e^a = \frac{\bar{P}_e^a}{\bar{P}_e^{a-1}} \times 100 \quad \text{con} \quad \bar{P}_e^a = \sqrt[n_e^a]{\prod_{i=1}^{n_e^a} P_i^a} \quad \text{y} \quad \bar{P}_e^{a-1} = \sqrt[n_e^{a-1}]{\prod_{i=1}^{n_e^{a-1}} P_i^{a-1}}$$

The formula used for calculating average prices is the geometric mean, since it has two fundamental advantages:

¹ In each province, small municipalities with fewer than 10,000 inhabitants are grouped into a single category

² In provincial capital municipalities, the results are broken down by district.

³ Rentals that do not fall into either of the two building categories (block of flats or single-family) are excluded

- It gives equal importance to each dwelling involved in the calculation, therefore, it does not depend on levels, which occurs in the arithmetic mean, where the higher values have greater influence on the result of the average.
- It does not depend on the unit, so the evolution of the resulting index is the same whether the total rent in euros or the rent per square meter is used. This property would even allow for the unavailability of the dwelling surface area, which may be useful when the data of the missing autonomous communities can be incorporated, if they do not have such information.

2. Aggregate indices compared to the previous year

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

$${}_{a-1}I_A^a = \sum_{e \in A} W_e^a \times {}_{a-1}I_e^a$$

where:

W_e^a is the weighting of stratum e , as a fraction, valid during the year a , and

${}_{a-1}I_e^a$ is the elementary index of the stratum e , in the year a .

The weights for each stratum represent the relationship between the expenditure that households make on rent pertaining to that stratum and the total expenditure made on rent.

According to the Paasche formula used in the RHPI, the current weightings for each year are calculated using information on the dwelling rented in that current year a , with the prices in the previous year $a-1$, in each stratum:

$$W_e^a = \frac{N_e^a \times \bar{P}_e^{a-1}}{\sum_e N_e^a \times \bar{P}_e^{a-1}}$$

where:

N_e^a is the number of dwellings pertaining to stratum e leased in the year a

\bar{P}_e^a the mean geometric price in the year a of the dwellings rented in that year pertaining to stratum e .

3. Indices in base 2015

The indices published are the 2015 base indices, which are obtained by linking the indices referring to the previous year, according to the following expression:

$${}_{15}I_A^a = {}_{15}I_A^{a-1} \times \left(\frac{{}_{a-1}I_A^a}{100} \right) = {}_{15}I_A^{16} \times \left(\frac{{}_{16}I_A^{17}}{100} \right) \times \dots \times \left(\frac{{}_{a-2}I_A^{a-1}}{100} \right) \times \left(\frac{{}_{a-1}I_A^a}{100} \right) \quad a > 2016$$

Based on the 2015 base indices, the annual variation rates and rates for periods longer than one year are calculated.

6 Dissemination

Data is published for the entire country, by autonomous community (except Comunidad Foral de Navarra and País Vasco), province, municipality and municipal groupings with fewer than 10,000 inhabitants, and by district (in the provincial capital municipalities) and provided that a minimum number of rental dwellings is exceeded in each territory.

Regarding functional disaggregation, the information is broken down by building type (block of flats or single-family), by dwelling size (six intervals) and from 2021, by contract age. This itemisation is provided for the country as a whole, by autonomous community and by province.

The following data has been published for each year since 2011:

- indices in base 2015
- annual variation rates
- weightings

From the first publication, tables with the results for each year and maps are provided, one showing the evolution of rental prices since the base year, 2015, and another showing the latest annual rate.

In addition, the weightings of all aggregates are published for each index. The weights represent the relationship between the expenditure that households make on rent pertaining to that stratum and the total expenditure made on rent.

Publication schedule

The first publication, in February 2022, disseminated the annual results from 2011 to 2020, based on dwellings rented in two consecutive years where it was not necessary to make any quality adjustment, as they were the same dwellings.

In the update of the publication carried out in 2023, the data for 2021 have been published, which concerned the calculation of the total number of rented homes declared in that year which belong to the scope of study.

In the future, it is planned that the annual results for the previous two years will be published in the first quarter of each year.