INSTITUTO NACIONAL DE ESTADISTICA

Household Projections

Methodology

February 2023

1 Introduction

The demand for data on the evolution of the number, type and composition of households has gradually grown in recent years. Traditionally, the population and housing censuses - conducted every ten years - are the main sources of this type of information. Thus, the most recent 2011 census revealed not only the extraordinary population increase as a consequence of the massive arrival of foreigners over the past decade, but fundamentally, that experienced in the number of households and in their typology.

1.1 THE 2011 POPULATION AND HOUSING CENSUS

The 2011 census showed a 14.6% increase in Spain's population, and a 27.5% increase in the number of households, since the previous, 2001 census. This decade-long increase was accompanied by an increase in the number of dwellings (20.3%).

Spain's population grew by almost 6 million people between the two censuses, with the arrival of foreigners contributing to this extraordinary population increase. In those ten years, the foreign population increased by more than three and a half million, reaching the figure of 5.3 million; this represented 11.2% of the population.

The housing stock also increased between the two censuses. In the 2011 census, the total number of homes in Spain exceeded 25.2 million, having grown by 20.3%.

The most pronounced increase occurred among primary dwellings, whose number increased by nearly 4 million (27.5%), to reach 18.1 million.

It should be taken into account that for the purposes of the census, the terms "primary dwellings" and "households" are synonymous, since the group of individuals who habitually occupy the same dwelling is called a household. The number of households in Spain thus reached 18,083,692 in November 2011, a figure significantly higher than that estimated from surveys (for example, the Labour Force Survey estimated a total of 17.4 million households in the fourth quarter of 2011).

Comparison of dwellings, households and population in 1991, 2001 and 2011

	Census 1990-1991(*)	Census 2001(**)	Census 2011(**)	Absolute growth 1990-2001	Absolute growth 2001-2011	Relative growth 1990-2001	Relative growth 2001-2011
Dwellings	17.220.399	20.946.554	25.208.623	3.726.155	4.262.069	21,6%	20,3%
Households	11.852.075	14.187.169	18.083.692	2.335.094	3.896.523	19,7%	27,5%
Population	38.872.268	40.847.371	46.815.916	1.975.103	5.968.545	5,1%	14,6%

(*) Reference date: 1 March 1991

(**) Reference date: 1 November

In short, while the population increased by 14.6% over the course of a decade, the growth in number of households was much more intense (27.5%) and considerably more accentuated than that of the previous decade (19.7 %).

This larger increase in households than in population has been taking place since the previous decades: thus, between the 1970 and 1980 censuses, households grew by 22.14% while the population grew by 10.88%, and between 1980 and 1991, the increases were 13.37% and 2.98%, respectively.

In addition, there were significant changes in both household composition and size. The average number of household members has decreased considerably since 1970, from almost four people per household to 2.58 in 2011.

1.2 THE NEED FOR UP-TO-DATE DATA ON THE NUMBER OF HOUSEHOLDS

To the extent possible, the statistical system should reflect this new reality, and should do so without the need to wait ten years until a new census is carried out.

Once the results of the 2011 Population and Housing Census were obtained - which provided precise information on the number and composition of the households making up the population for November 2011 - the challenge of producing updated data on the number of households was undertaken. This represented an important advance in the quality of the information produced by the INE.

In the absence of other data, household surveys were frequently used to discover recent developments in the number of households. One of the problems frequently associated with household surveys, however, is the bias that can be introduced into results owing to a lack of response from certain types of households that are more difficult to locate (not at home) or who are unwilling to collaborate. This means that the description of the society and of the forms of coexistence provided by the survey, and even the number of households estimated, may be distorted.

Furthermore, as with the population, the recent evolution in the number of households is not the only item of interest. It is also necessary to show, in advance, household creation and dissolution dynamics. That is why the INE proposed the **Household Projection** operation.

This operation was created with the purpose of gradually incorporating as many additional sources of information and methodological improvements into the methodology as deemed necessary to better identify and measure household evolution and composition. These methodological improvements will be incorporated in forthcoming editions, and will be documented in order to inform researchers and users.

2 Objectives

The Household Projection statistical operation was designed to provide updated information on household number and size, as well as to offer, every two years, a statistical simulation for the next 15 years, based on the demographic trends and social behaviours currently observed. In both cases, information is provided regarding the homes of people residing in Spain, its autonomous communities and its provinces. Furthermore, due to their construction, the operation results are fully consistent with the results of Population Projections, also released by the INE.

More specifically, the objectives of the Household Projection statistical operation are:

1. Obtain and make available information for the next 15-year period on the evolution of the number of households by size in Spain, its autonomous communities and its provinces. This data will be published **every two years** starting in 2014.

2. This objective will allow different social agents to plan social policies based on the expected knowledge regarding the evolution of forms of coexistence.
3. In addition, household **estimates** are constructed each quarter and each semester for use in calibrating household surveys. These are calculated using the population figures available at the time and are at the same level of geographic disaggregation as the projections.

These estimates will not be a specific publication; rather, their results will appear explicitly in all household surveys that require them for calibration. For example, the CHS uses the semi-annual estimates, while the EAPS, due to its urgency and immediacy, requires the results from the quarterly estimates.

This makes it possible to obtain updated data on households both quarterly and semi-annually at the province level, as well as distribution according to household size.

3 Information Sources

The following basic sources of information are used to prepare the Household Projection:

2011 Population and Housing Census

Provides the number of households on the census date, classified by household size and by province, as well as the population residing in family dwellings classified by the size of the household to which this population belongs, age group, sex and province of residence.

Continuous Register since 2012

The Continuous Register constructs a framework of postal addresses for primary dwellings to be created, using the addresses that appear in the person's individual records.

Although for the time being the information in the Register is not sufficiently accurate to determine the number of occupied dwellings (or households), it can be used to construct an framework of "occupied" postal addresses. This makes it possible to study **the evolution** of the number of households by size, comparing the data from the register at the beginning of each year with the previous year, since 2012. This variation in the number of households, classified by household size and by province, as well as the population classified according to the size of the household to which this population belongs, age group, sex and province of residence, is used to estimate the propensity variation of belonging to a certain type of household each year.

Population Figures, available from 2012 onwards

Contains the population series since January 2012 disaggregated by sex, age and province, but not by household size. This series is available for both the total population, and separately for the population residing in family dwellings for internal use. In addition to the figures published each semester (at least 6 months after the reference date), estimates are produced for internal use at the end of each quarter.

4 Concepts and definitions

Resident Population: the resident population in a given geographical area is the persons that, on the reference date, have their regular residence in said area.

The **Regular Residence**[1] is defined as the place where a person normally spends their daily rest periods, regardless of temporary absences for leisure purposes, holidays, visits to relatives and friends, business, medical treatment or religious pilgrimage. Nonetheless, it must be noted that only the following will be considered regular residents in a district:

- Those persons who, according to the previous definition, have regularly resided therein for a continuous period of at least twelve months.
- Those persons that, according to the above definition, had established their regular residence therein less than twelve months before, but intended to stay there for at least one year.

Family Dwelling: dwelling intended to be inhabited by one or more people who do not constitute a collective, regardless of the ties between them.

Collective Dwelling: dwelling intended to be inhabited by a collective, that is, by a group of people subject to a common authority or regime not based on family ties or coexistence. A collective dwelling may occupy only part of a building or - more frequently - the whole of it.

Household (according to residence criteria): a human group formed by one or several persons who habitually reside in a family dwelling all or most of the year; family ties do not necessarily exist among the members of the group.

Household Size: number of household members.

5. Basic Household Projection Methodology

The model chosen to carry out the household projections is based on the so-called propensity method. This method is based on the following article: “Bell, M., Cooper, J., et al (1995) Household and Family Forecasting Models. A review. Canberra. Departament of Housing and Regional Development”.

The method consists of calculating the propensity (probability) that, based on basic sex and age characteristics, persons of a certain population group will belong to a certain type of household. This propensity is measured by the quotient between the total number of persons belonging to the population group under consideration who reside in households of that type, and the total number of persons in the population group in question. That is, using the formula:

[1] Definition of regular residence established in *Regulation 763/2008 of the European Parliament and of the Council on Population and Housing Censuses* and in *Regulation 1260/2013 of the European Parliament and of the Council on European Demographic Statistics.*

$$PH_{h,s,x}(k) = \frac{P_{h,s,x}(k)}{P_{h,s,x}}$$

Being:

- $PH_{h,s,x}(k)$ the propensity that persons of sex s and age x in province h have of belonging to a type k household.
- $P_{h,s,x}(k)$ the population of persons of sex s and age x in province h in type k households.
- $P_{h,s,x}$ the resident population in family dwellings of persons of sex s and age x in province h.

These propensities are applied to the figures for the population permanently residing in family dwellings; i.e., the population residing in collective dwellings is excluded. With this, an estimate of the number of households of each type consistent with the population in question is obtained.

The household type will be defined based on size, using the following types of households:

1. Household Size 1
2. Household Size 2
3. Household Size 3
4. Household Size 4
5. Household greater than or equal to 5

Ten population groups have been established for the combinations of sex (man, woman) and age group (under 20 years, 20 to 39, 40 to 59, 60 to 79, over 79 years), bearing in mind that household composition and dissolution dynamics are mainly related to the differences in these variables. At the start, nationality was also considered, but it was finally discarded, since the foreign population is smaller, and is influenced by the processes of acquisition of Spanish nationality.

5.1 EVOLUTION OF PROPENSITIES IN THE CONTINUOUS REGISTER

Initially, the household data obtained from various surveys between 2008 and 2013 (LCS, EAPS, HBS, ICT-H) and even data obtained from the Register were analysed. It was observed that the decrease in average household size did not appear to be constant over time. The Continuous Household Survey was not included in this analysis since at the time only 2013 data was available. It was thus determined that taking the propensities derived from the 2011 Census results was not sufficient; rather, these should evolve over time. The information contained in the continuous Register has been used for this purpose, as described below.

Using the continuous Register on January 1st of each year, the number of people residing at each postal address is determined as an approximation of the number of households, and for each province, the propensities of belonging to each type of household are calculated for the different defined population groups. These figures provide a household

structure by size that is not entirely comparable to that of the censuses. The information used is thus the evolution over time of the propensity of belonging to these “households” by size, and not the absolute values.

For the same dates, the average size of households of five or more persons is calculated for each province, using the quotient between the population residing in households greater than or equal to five, and the number of households of said size.

The annual evolution of propensities and average sizes of households with five or more members is derived from the variation observed between January of one year and January of the following year.

5.2 ESTIMATED STARTING HOUSEHOLDS FOR THE PROJECTION AND SEMI-ANNUAL ESTIMATE

To prepare the population projections, the number of households by size and province as of January 1 of the starting year of the projection is firstly needed. This corresponds to the population provided by the latest provisional Population Figures, published in June of that same year.

Since these populations are published six months after the reference date, the Municipal Register data as of January 1 is always available. It is thus sufficient to start with the propensities and average sizes of households of 5 or more people from the 2011 Census and apply the annual evolution of both elements observed in the Register to gradually obtain the average propensities and sizes as of January 1, for every year after the census.

These propensities are applied to the Population Figures as of January 1 of each reference year to obtain households by size (those of the last year will be provisional, while the previous ones are definitive).

The surveys with less immediacy in the need for households for calibration, such as the EAPS, use this series. It is updated every six months together with the Population Figures publications, in June and December of each year, and produces provisional and definitive results, respectively. Estimates are also provided for the intermediate dates of the year, using the populations available at each moment.

5.3 HOUSEHOLD PROJECTION TO 15 YEARS

The household projection consists of simulating the number of households there would be by size and province for the years following a certain start year, on January 1 of each year, so long as the trends of recent years continue.

The first edition of household projections was carried out in 2014 and covered a period of 15 years, from 2014 to 2029. A new edition is published every two years, meaning that household projections were published in 2016, 2018 and 2020.

In order to carry out this projection - which will be calculated starting in 2022 and will have a projection horizon of 15 years, that is, until 2037, the calculation method is divided into several steps:

1. Projection of propensities and average sizes of households of five or more people throughout the projective period, by extrapolating the average annual evolution observed in the Municipal Register in the last three years (between 2019 and 2022).
2. Application of projected propensities to projected populations in the Population Projections operation.
3. Obtainment of the number of each size of household in each province for each year of the projection, and application of the projected average sizes of households of five or more people to calculate households of this size.

Thus, the basis for determining the household projection is the population projection, but hypotheses must also be made regarding the evolution of propensities and average sizes.

To project the propensities, an extrapolation is carried out using limited exponential-type formulas, which place a lower cap on decreasing propensities and an upper cap on increasing propensities. The same criterion, with small variations, is used to extrapolate the mean sizes of households of five or more people. These limits act as asymptotes, toward which average sizes and propensities tend over time.

The use of these formulas ensures that the average size of households of five or more persons could not be less than five, or trigger its growth, and prevents the decreasing propensities from becoming negative and the increasing propensities from reaching values that cause the sum of the propensities of each population group to be much higher than one (1); with one being the theoretical value of said sum.

The formulas used to **calculate the propensities** for each year are the following:

a) For the propensities that, according to the census reference framework for years t-3 and t-1 January, are **decreasing** (that is, if $PHP_{h,s,x}^{t-3} > PHP_{h,s,x}^{t}$), it is calculated:

$$PH_{h,s,x}^{t+n}(k) = 0{,}0001 + \left[\left(PH_{h,s,x}^{t}(k) - 0{,}0001\right) \cdot \left(\frac{PHP_{h,s,x}^{t}(k)}{PHP_{h,s,x}^{t-3}(k)}\right)^{\frac{n}{3}}\right]$$

Where:

- t is the start year for the projection, which in this case will be t = 2020 and will refer to January 1, as in all the years involved in the projections.
- n is the number of years elapsed since the reference point from which the projection starts (with values n = 1, 2, 3,….15).
- k is the type of household according to size, with values k = 1, 2, 3, 4, 5, with k = 5 referring to households with 5 or more people.
- $PH_{h,s,x}^{t+n}(k)$ is the propensity for a person of sex s and age group x in province h to belong to a household of size k on January 1 of year t+n.
- $PH_{h,s,x}^{t}(k)$ the propensity for a person of sex s and age group x in province h to belong to a household of size k in the starting year of the projection.
- 0.0001 is the lower cap set for decreasing propensities.

– $PHP_{h,s,x}^{t-3}(k)$, $PHP_{h,s,x}^{t}(k)$ are the propensities that a person of sex s and age group x in province h belongs to a household of size k on January 1, years t-3 and t respectively, according to the reference framework of the Continuous Municipal Register on said dates.

This formula makes any propensity calculated worth more than 0.0001, which is an arbitrary value chosen based on the values of the minimum propensities calculated from the 2011 census data. The lowest propensities seen are for those of belonging to size one households, by people in the first age group. Specifically, for the national total, the propensity for men is 0.000965 and that of women, 0.000931. By provinces, propensities oscillate close to that value, many that the highest is 0.00373, but always above 0.0001, except in six cases that are 0 (zero) absolute, and the lowest non-zero is 0.00017.

A zero propensity for these population groups would mean that there are no people under the age of 20 living alone: this is rare today, but to say that there are no cases would seem very extreme. If propensities worth zero are projected, the proposed formula would assign the value zero to the entire series of propensities. In these cases, for the formula to work in the first projections based on the 2011 census data, it was necessary to replace the null propensities for another arbitrary value, with the chosen value being 0.0002.

b) For the propensities that, according to the census reference framework for years t-3 and t, are **increasing** (that is, if $PHP_{h,s,x}^{t-3} < PHP_{h,s,x}^{t}$), it is calculated:

$$PH_{h,s,x}^{t+n}(k) = LÍMITE - \left[\left(LÍMITE - PH_{h,s,x}^{t}(k)\right) \cdot \left(\frac{PHP_{h,s,x}^{t-3}(k)}{PHP_{h,s,x}^{t}(k)}\right)^{\frac{n}{3}}\right]$$

With all the formula elements defined as for the decreasing case, and in this case considering the CAP as the smallest of the following four values:

– $PH_{h,s,x}^{t}(k) + \mathrm{C}$, to prevent the propensities from increasing too much in absolute value.

– $PH_{h,s,x}^{t}(k) \cdot M$, to prevent the propensities from increasing too much in relative value. For example, the propensities of belonging to size one households of persons under 20 years of age are very close to 0, but the relative variation between the patterns can be very high, and transferring this evolution in the long term can provide values completely unrelated to those of the starting point.

– T, to avoid the propensities being in any case greater than a certain value.

– $PH_{h,s,x}^{t}(k) + \left[\left(\frac{PHP_{h,s,x}^{t}(k)}{PHP_{h,s,x}^{t-3}(k)}\right)^{\frac{1}{3}} - 1\right] \cdot S$, to avoid the exponential formula that calculates the propensities producing very large increases in the long term.

The values set for these formulas are C=0.12, M=2, T=0.7 and S=10, and they have been determined based on a series of tests that vary the possible values and check their operation.

After propensities for the projection period are calculated, they must be adjusted so that the sum of the propensities of all sizes for each population group continues to add up to

one - since, when projecting each size separately, it is common for them to stop adding up to one, although they do not deviate too much. This adjustment is made using a recursive procedure, which consists of taking the propensities within each sex and age group, and dividing each of them by the sum of all the propensities in that sex and age group. Should the growth of any these propensities change sign regarding the evolution of the Municipal Register, what remains to be assigned is left constant and is distributed among the propensities of other sizes. The process is repeated until an adjustment is reached that respects the growth sign in all sizes.

The following formulas are used to **calculate the average sizes** of households of five and more people for each year:

a) For the average sizes of households of five and more people, which, according to the census reference framework for years t-3 and January 1, are **decreasing** (that is, if $TMP_h^{t-3} > TMP_h^t$), the formulas used are:

$$TM_h^{t+n}(5) = 5{,}1 + \left[(TM_h^t(5) - 5{,}1) \cdot \left(\frac{TMP_h^t(5)}{TMP_h^{t-3}(5)} \right)^{\frac{12n}{3}} \right]$$

Where:

- $TM_h^{t+n}(5)$ is the average size of households with five or more members in province h in year t+n.
- $TM_h^t(5)$ is the average size of households with five or more members at the time of reference from which the projection begins, on January 1 of year t, for province h.
- $TMP_h^{t-3}(5)$, $TMP_h^t(5)$ are the average sizes of households with five or more members in province h on January 1 for the years t-3 and t respectively, according to the Municipal Register reference framework on those dates.
- is the minimum value established for the average size of households of five or more persons, in view of the average sizes calculated from the 2011 census data.

With this formula, the mean size cannot be less than 5.1; this value is an arbitrary and modifiable figure, but it seems reasonable that it should be greater than five, since five would be the extreme case in which there was no household with more than five members, which seems unlikely.

In this formula - unlike the formula used for propensities - the exponent for the power is multiplied by 12. In both cases, an attempt is made to achieve a reasonable evolution of these variables, so that the theoretical series of sizes and propensities have values close to those obtained at the series origin, and evolve similarly to that seen in the Registers for the years considered. Since the magnitude of the two items is different (the propensities less than 1, and the mean sizes greater than 5), the speed of approach to the asymptote marked by the cap is different: if we do not multiply by a constant in the case of mean sizes, the speed is very slow, and the imposition of the cap causes the curve to descend excessively (in the case of the upper cap; analogous to the lower cap). The value 12 was chosen empirically.

b) For the average sizes of households of five and more people, which, according to the census reference framework for years t-3 and January 1, **are increasing** (that is, if $TMP_h^{t-3} < TMP_h^t$), the formulas used are:

$$TM_h^{t+n}(5) = LÍMITE + \left[\left(LÍMITE - TM_h^t(5)\right) \cdot \left(\frac{TMP_h^{t-3}(5)}{TMP_h^t(5)}\right)^{\frac{12n}{3}}\right]$$

With all the formula elements defined as for the decreasing case, and in this case considering the CAP as the smallest of the following four values:

- $TM_h^t(5) + C$, to prevent the average sizes from increasing too much in absolute value.
- $TM_h^t(5) \cdot M$, to prevent the average sizes from increasing too much in relative value. For example, the propensities of belonging to size one households of persons under 20 years of age are very close to 0, but the relative variation between the patterns can be very high, and transferring this evolution in the long term can provide values completely unrelated to those of the starting point.
- T, to avoid the average sizes being in any case greater than a certain value.

The values set for these formulas are C=0.52, M=1.09 and T=6.6, and they have been determined based on a series of tests based on known sources that vary the possible values and check their operation.

Finally, the number of households by size and province on January 1 of each year is obtained from the following formula, which shows all elements previously obtained: that is, propensities, the average sizes of households of five or more members, and the resident population figures in family dwellings, as follows:

$$H_h^{t+n}(k) = \frac{\sum_{s,x} PH_{h,s,x}^{t+n}(k) \cdot P_{h,s,x}^{t+n}}{TM_h^{t+n}(k)}$$

Being:

- $H_h^{t+n}(k)$ the number of households of size k in the province h on January 1 of year t+n, with t=2020 and n=1.2…15.
- $PH_{h,s,x}^{t+n}(k)$ the propensity that a person of sex s and age group x belongs to a household of size k in province h on January 1 of year t+n.
- $P_{h,s,x}^{t+n}$ the population of sex s and age x in province h on January 1 of year t+n that resides in family dwellings.
- $TM_h^{t+n}(k)$ the average size of households of size k in province h on January 1 of year t+n

5.4 QUARTERLY ESTIMATE OF HOUSEHOLDS

The quarterly estimate of households should be ready at the beginning of the month that follows the end of the reference quarter. It is thus necessary to extrapolate propensities a few months from January 1, which is the final date on which Municipal Register data is available. This extrapolation is carried out in the same way as projections for several years ahead. Some differences are thus applied in the previous formulas:

- T-3 is replaced by t-1; that is, the evolution of the propensities or average sizes for the last year of the Register is applied, since they are the current estimates.
- Consequently, in the quotients that reflect the Municipal Register's evolution, the exponent $\frac{1}{3}$ disappears; this had been to obtain the three-year average annual growth.
- The exponent n will now take values for a fraction of a year. For example, if you wish to extrapolate 6 months, it would be $n = \frac{6}{12}$.

Once these elements have been extrapolated, they are applied to the quarterly family dwelling Population Figures that are elaborated internally for survey calibration, particularly for the EAPS.

6 Calendar and Dissemination Plan

The 15-year Household Projection is published every two years, starting in 2014. The last publication took place in October 2022, and contains, in accordance with the Population Projections, information at the provincial level for the years 2022 through 2037.

The plan is to publish a single table, which would present the evolution of the number of households for Spain, its autonomous communities, and its provinces on January 1 of each year for the next 15 years, by household size (1, 2, 3, 4, and more than 4 members). Average household size is also published for the first time in this edition.

In principle, it is planned to publish a single table containing the evolution of the number of households for Spain, its Autonomous Communities and its provinces on 1 January of each year for the next 15 years, by household size (1, 2, 3, 3, 4, more than 4 members). This table also includes the average household size.

At the end of each quarter, and six months after each semester, estimates are made of the households needed for household survey calibration. As previously noted, these projections do not result in a specific publication of results, but the data itself is nonetheless made public: it appears in the EAPS publication and in other household surveys.