

Bover, Olympia

## Article

# Measuring expectations from household survey: New results on subjective probabilities of future house prices

SERIEs - Journal of the Spanish Economic Association

## Provided in Cooperation with:

Spanish Economic Association

*Suggested Citation:* Bover, Olympia (2015) : Measuring expectations from household survey: New results on subjective probabilities of future house prices, SERIEs - Journal of the Spanish Economic Association, ISSN 1869-4195, Springer, Heidelberg, Vol. 6, Iss. 4, pp. 361-405, <https://doi.org/10.1007/s13209-015-0136-z>

This Version is available at:

<https://hdl.handle.net/10419/158545>

## Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

## Terms of use:

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<http://creativecommons.org/licenses/by/4.0/>

# Measuring expectations from household surveys: new results on subjective probabilities of future house prices

Olympia Bover<sup>1</sup>

Received: 30 October 2015 / Accepted: 11 November 2015 / Published online: 26 November 2015  
© The Author(s) 2015. This article is published with open access at Springerlink.com

**Abstract** I analyze new data on subjective probabilistic expectations on house prices collected in the Spanish Survey of Household Finances. Households are asked to distribute ten points among five different scenarios for the change in the price of their homes over the next 12 months. This paper is the first empirical study to document the beliefs of a representative sample of households about the future value of their homes. It also reviews the methodology of expectation measurement and recent work on household subjective probabilities. I model individual subjective probability densities using splines, construct quantiles from those densities, and analyze how the heterogeneity in the individual distributions relates to differences in housing and household characteristics. An important result of the paper is that women are more optimistic about the evolution of house prices than men. Location at the postal code level accounts for a large fraction of the variation in the subjective distributions across households. Finally, I provide some results on how subjective expectations matter for predicting spending behavior. Housing investment and car purchases are negatively associated with pessimistic expectations about future house price changes and with uncertainty about those expectations.

**Keywords** Household subjective probabilistic expectations · House price expectations · Gender bias · Consumption · Portfolio decisions

**JEL Classification** C81 · D84 · D12 · D14 · R21

---

This paper was written as a lecture for presentation as the Presidential Address to the Congress of the Spanish Economic Association in Vigo, December 2012. I am grateful to an anonymous referee and the Editor for their helpful comments.

---

✉ Olympia Bover  
bover@bde.es

<sup>1</sup> Banco de España, Madrid, Spain

## 1 Introduction

This lecture is concerned with household subjective expectations. Its central theme is the analysis of new data on subjective probabilistic expectations on house prices collected in the Spanish Survey of Household Finances (EFF). As a front-end, I first provide a review of the methodology of expectation measurement and of some recent work that use household subjective probabilities. Finally, as a back-end I provide some results on how subjective expectations matter for predicting consumption behavior.

Despite widespread agreement on the fundamental role of expectations in explaining behavior, direct measurement of individual expectations is a relatively recent activity. The standard practice in the economics of the last century was to infer the individuals' decision process from their observed choices. Following this revealed preference analysis, both preferences and the uncertainty about the future are identified from data on choices and market outcomes alone. Such strategy requires strong assumptions. For example, assuming individuals have rational expectations as well as knowledge of the model may be needed despite that this has often not been credible. In his seminal paper Manski (2004) strongly advocated for collecting self reported expectation data and using those jointly with observed choice data. The hope is this would improve economists' credibility and ability to predict behavior. But are household expectations collected through surveys trustworthy? Do subjective household survey expectations really improve the ability to predict behavior? To help put these questions in context, I begin by reviewing basic concepts of the methodology of expectation measurement as well as recent work on the elicitation and use of household subjective expectations.

The EFF is a representative survey of the Spanish population that contains detailed information on household assets, debts, income and consumption. Data have been collected every three years since 2002. Starting in 2011, the EFF introduced a new question to elicit household house price probabilistic expectations. Households were asked to distribute ten points among five different scenarios concerning the price change of their homes over the next 12 months. In this way respondents provide information not only about point expectations but also about the probabilities they assign to different future outcomes.

One motivation for introducing this question in the EFF is the importance of real estate assets in the wealth of Spanish households (80 % of the value of household assets) all along the wealth distribution (88 % for the bottom quartile and 67.5 % for the top decile). Aside from a high proportion of owner occupier households (83 %), 36 % of Spanish households hold some other real estate property.

It is also a timely question due to the housing market collapse that shattered house price expectations after 2007 in Spain. The number of households buying housing dropped dramatically from an overall annual average rate of 2.3 % between 2002 and 2005 to 1.1 % in 2011. According to the data I analyze in this paper, in 2011 over 23 % of households expected a large drop (of over 6 %) in the future price of their homes. Moreover, among households expecting such large drops, the fraction who bought a car was half the fraction in the total population (4.5 instead of 9.4 %).

This paper is one of the first empirical studies to document the beliefs of households about the future value of their homes, and the first one that uses a representative

sample of households. Questions on probabilistic house price expectations have only recently been introduced in household surveys, as detailed in Sect. 3. Niu and van Soest (2014) have independently obtained results that are complementary to ours using newly collected house price expectations data from the Rand American Life Panel.

I start by analyzing patterns of the answers provided by the EFF2011 respondents to the house price probabilistic expectation question to assess the coherency of responses. These include bunching, number of intervals used, and their association with the extent of non-response. Next I model individual probability densities and analyze how the heterogeneity in the individual distributions relates to differences in housing properties and in the characteristics of households.

An important result of the paper is that women are more optimistic about the evolution of house prices than men. Being a woman is associated with a positive shift in the median and the quartiles of the subjective distributions. I further examined potential differences in asset valuations by gender by considering self-assessed values of other assets reported in the EFF. I find that women tend to provide higher estimates for the value of their home compared to men but lower ones when it comes to value their financial assets.

Location at the postal code level accounts for a large fraction of the variation in the subjective distributions across households. Importantly, in the absence of postal code fixed effects the estimated effects of demographics on house price expectations would be biased. For example, the result on gender would not be found. Moreover, the location effects that emerge from the subjective probability data are meaningful and respond to economic fundamentals. In particular, estimated location fixed effects respond to past local house prices and unemployment rates.

Finally, I study whether reported household expectations predict household expenditure decisions. This is of substantive interest to understand household behavior and also a further step in the validation of the house price expectation responses. I exploit the availability in the EFF of information about purchases of secondary housing, cars, other big ticket items, and food. These data allow me to uncover some novel findings about correlations of house price expectations and their uncertainty with those purchases and expenditures. I find that housing investment and car purchases are negatively associated with pessimistic expectations about future house price changes and with uncertainty about those expectations. Moreover, these effects depend on household wealth. Specifically, the negative effects of holding very pessimistic house price expectations on secondary housing purchases are more pronounced at the top of the wealth distribution than at the median, while the opposite is true for car purchases.

The paper is structured as follows. In Sect. 2 the work on elicitation and use of household expectations is reviewed. I discuss the specificities in implementing expectation questions in household surveys and the validation of such questions. I also discuss some specific uses of subjective expectations, work on expectation formation, and some enlightening experiments conducted within expectation surveys. Section 3 contains the analysis of the house price expectations data in the EFF. First I describe the formulation of the question and I examine the quality of the responses. Next I estimate a probability density for each respondent, which I use to document the extent of heterogeneity in beliefs. Based on these individual densities I compute various

quantiles and measures of dispersion, and study their association with respondent and house characteristics. Finally, Sect. 4 reports the results on the relation between house price expectations and expenditure decisions. I present predictive results for the probabilities of purchasing secondary housing, an automobile, and other big ticket items.

## 2 The quantification of human uncertainty from social surveys

### 2.1 Preliminaries

After years of distrust, the measurement of individual expectations is becoming a very active topic in economics, both for research and for immediate policy use. Since the 1990s an increasing number of household surveys have been collecting data on subjective probabilistic expectations.<sup>1</sup> Expectation questions may be about future outcomes concerning the individual (e.g. own income, health, death, job security, home value, pension benefits, bequests) or about future aggregate conditions (e.g. inflation, house prices, stock market).

There are two important distinctions when considering asking expectations questions. First, whether the question is about eliciting point expectations as, for example, asking for the expected number of children, or about eliciting probabilistic expectations. A probabilistic counterpart to the previous example would be to ask about the probability of having no children, of having one child, of having two children, etc.

The second important distinction when considering eliciting expectations is whether the answer we seek is qualitative or quantitative. Qualitative questions to measure expectations have been used for some time. An example of qualitative question is as follows:

“Thinking about the next 12 months how likely do you think it is that you will lose your job? Possible answers: very likely, fairly likely, not too likely, not at all likely”.

An alternative probabilistic question on the same subject is:

“Using a scale from 0 to 100 what is the percent chance that you lose your job in the next 12 months?”.

This type of probabilistic questions are usually preceded by some explanations and examples about the meaning of probabilities (e.g. using examples about the probability of rain) and/or accompanied by some visual aid (e.g. a ruler).

Two limitations of verbal expressions of expectations (of the type “very likely”, “fairly likely”, “not too likely”) are that different respondents may interpret them differently and that they convey limited information about respondents’ expectations. In fact, [Dominitz and Manski \(1997, 2004\)](#) blame the early use of verbal expectations for the economists’ distrust of expectations data. In particular, they cite a controversy in the 1950s and 1960s about the usefulness of elicited verbal assessments of expected consumer finances in the Federal Reserve Board Survey of Consumer Finances con-

<sup>1</sup> Some of the most prominent are the US Health and Retirement Survey (HRS) and its UK counterpart the English Longitudinal Study of Ageing (ELSA), the US Survey of Economic Expectations, the American Life Panel (ALP), some Household Wealth Surveys (in particular the Italian SHIW, the Dutch VSB Panel, and the Spanish EFF).

ducted by the University of Michigan Survey Research Center. The debate had George Katona<sup>2</sup> as the leading proponent of qualitative attitudinal questions vs. Thomas Juster who did not find them useful in predicting behavior.<sup>3</sup> This debate would have left economists suspicious of any expectation data for a while.

The advantages of asking probabilistic expectations are that numeric answers are comparable across persons and over time, algebra may be used to examine consistency, and they allow respondents to express uncertainty or risk.

Measuring probabilistic expectations about future continuous outcomes entails obtaining each respondent's subjective probability distribution. An early example is the following question about earnings uncertainty included in the 1989 Survey of Household Income and Wealth (Banca d'Italia):

"We are interested in knowing your opinion about labor earnings or pensions 12 months from now. Suppose now that you have 100 points to be distributed between these intervals (a table is shown to the person interviewed). Are there intervals which you definitely exclude? Assign zero points to these intervals. How many points do you assign to each of the remaining intervals?"

A different formulation with the same objective could be

"How likely do you think it is that your income in the coming year will be higher than \_\_\_ (A/B/C) Rupees?"

as adopted in [Attanasio and Augsburg \(2012\)](#), where A, B, and C are different income thresholds. The information is elicited in the form of a probability density in the first case and of a cumulative distribution in the second.

Despite some potential added difficulty for the respondent in answering questions in a probabilistic form, most of the evidence shows that respondents are willing to answer probabilistic questions and that their responses are generally sensible and internally consistent. This is so when the questions concern well defined events that relate to respondents' lives (see for example evidence cited in [Manski 2004](#); [van der Klaauw et al. 2008](#)).

Recently probabilistic expectations data have also been collected in developing countries (see [Attanasio 2009](#); [Attanasio and Augsburg 2012](#)) where getting sensible answers to such questions has also proved feasible. Some controversy however remains related to [Tversky and Kahneman \(1974\)](#) randomized experiments, which reveal that individuals often use heuristic methods rather than Bayes theorem.

Studies on decision making under ambiguity take probability expectations one step further. Ambiguity arises when individuals do not hold a single subjective distribution but may hold a set of them. In the case of binary events this would translate into

<sup>2</sup> See [Katona \(1957\)](#).

<sup>3</sup> See [Juster \(1964, 1966\)](#) proposed eliciting probabilistic expectations by linking verbal expressions with numerical probabilities. His formulation of a purchase probability question regarding automobiles and other household appliances reads as follows (as reported in [Manski 2004](#)):

*Taking everything into account, what are the prospects that some member of your family will buy a \_\_\_ sometime during the next \_\_\_ months, between now and \_\_\_?*  
*Certainly, Practically Certain (99 in 100); Almost Sure (9 in 10); Very Probably (8 in 10); Probably (7 in 10); Good Possibility (6 in 10); Fairly Good Possibility (5 in 10); Fair Possibility (4 in 10); Some Possibility (3 in 10); Slight Possibility (2 in 10); Very Slight Possibility (1 in 10); No Chance, Almost No Chance (1 in 100).*

allowing answers in intervals of probabilities instead of only point probabilities (for an extended explanation see [Manski 2004](#)). [Manski \(2004\)](#) provides the following example in the case of binary events: “What do you think is the percent chance that event A will occur? Please respond with a particular value or a range of values, as you see fit”. He comments that this formulation enables respondents to express uncertainty or ambiguity. For example, complete ignorance may be expressed by reporting “0–100 %”, bounded ambiguity by reporting “30–70 %”, uncertainty by reporting “60 %,” or certainty by reporting “100 %”.

## 2.2 Elicitation methodology

Asking for uncertainty requires a process of elicitation. It is not like asking for age. Hence elicitation methods matter to what gets elicited. Understanding this is important but does not necessarily render the request for elicitation meaningless.

*Wording* A substantial amount of work has been produced to try to minimize bias and systematic error by refining the way information is elicited. This is relevant since even apparently minimal differences in wording may produce different interpretations of the question.

A salient example is the experiment conducted by the Federal Reserve Bank of New York, as part of their Household Inflation Expectations Project, on the effects of alternative wordings for eliciting inflation expectations. One conclusion is that reported expectations were higher when the question asked was about expectations of “prices in general” (as in the long standing Michigan Survey question) than when the formulation was in terms of “inflation” expectations (see for example [Bruine de Bruin et al. 2011b, 2012](#)). These authors report that question about “prices in general” and “prices you pay” focus respondents more on personal price experience and since these may be driven by prices of different goods over time the answers may be less comparable than the ones prompted by an “inflation” formulation.

More generally, the wording used in eliciting subjective probabilities has to convey the concept of probability in a manner the respondent understands, so that he is able to express his probabilistic beliefs. In developed countries the usual wording is “percent chance” or “how likely”, while in developing countries respondents are often given a number of beans or balls they are asked to distribute.<sup>4</sup> [Delavande et al. \(2011\)](#) compare distributing balls across bins to the percent chance approach. In their Indian setting beans generate usable answers for almost all respondents while a percent chance formulation produced a significant fraction of inconsistent answers.<sup>5</sup> A practical consideration is the number of beans respondents are given to distribute. Greater accuracy may be expected the larger this number is but with too many beans eventually proving difficult to handle by the respondent.

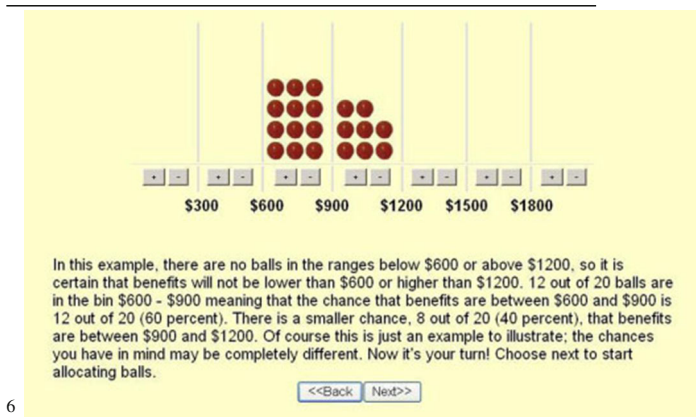
<sup>4</sup> But see [Delavande and Rohwedder \(2011\)](#) who ask Internet respondents in the US to allocate 20 balls across seven bins to express their beliefs about their future Social Security benefits.

<sup>5</sup> Along the same lines, [Manski \(2004\)](#) reports evidence that respondents perform much better when statistics are presented in the form of natural frequencies (e.g. 30 out of 10,000 cases) rather than in the form of objective probabilities (0.3 % of cases).



Visual aids are often employed to help respondents. In particular, a ruler may be used to explain the percent chance scale from 0 to 100 %. Visual aids have also proven useful in internet administered surveys in the US (see [Delavande and Rohwedder 2011](#)).<sup>6</sup> Often, time is also spent in providing examples about probability statements (for example, the probability of rain tomorrow) to try and make sure respondents understand probabilistic statements.<sup>7</sup>

*Eliciting subjective distributions: range of variation* Various elements need to be specified when formulating questions to obtain subjective distributions. The first consideration is to establish the range of variation of the outcome of interest. This may be obtained by asking the respondent to report the maximum and minimum possible outcome in a couple of preliminary questions. Alternatively the support may be chosen by the developer of the questionnaire and to be the same for all respondents.<sup>8</sup> The first option is now routinely used when the outcome is household or individual specific (e.g. own income) because it decreases the natural focus of the respondent on central tendencies and avoids that pre-established reference values influence his answers (also known as anchoring problem).<sup>9</sup> Predetermined ranges are predominant when eliciting expectations about aggregate outcomes (e.g. inflation). Once the range of variation is established it is divided in intervals (not necessarily equally wide) and corresponding cut-off points are determined. Presenting a large number of intervals may subsequently



6

<sup>7</sup> In the Health and Retirement Survey for example the explanations given are as follows: “Next we would like to ask your opinion about how likely you think various events might be. When I ask a question I’d like for you to give me a number from 0 to 100, where “0” means that you think there is absolutely no chance, and “100” means that you think the event is absolutely sure to happen. For example, no one can ever be sure about tomorrow’s weather, but if you think that rain is very unlikely tomorrow, you might say that there is a 10 % chance of rain. If you think there is a very good chance that it will rain tomorrow, you might say that there is an 80 % chance of rain”.

<sup>8</sup> [Dominitz and Manski \(1997\)](#) warn against interpreting the answers on minimum and maximum outcomes as absolute minimum and maximum possible outcomes and recommend using these only to help determine the range as opposed to fully determine it. Their suggestion would help overcome the problem discussed in [Delavande et al. \(2011\)](#) that self-reported ranges often produce less rounded interval bounds than would be the case with predetermined support. Non-rounded intervals are likely to be harder to think about for the respondent.

<sup>9</sup> See [Delavande et al. \(2011\)](#) for an attempt to compare the sensitivity of the results to differences in the specification of support.



allow for more precise statistics but be more cognitively demanding on the respondent. More intervals may be needed for individual outcomes with predetermined supports than with self-anchored ones to allow for individual heterogeneity in outcomes.<sup>10</sup>

*Eliciting subjective distributions: cdf vs. pdf* A third consideration when devising subjective distribution questions is whether to elicit the information in the form of a probability density (pdf) or a cumulative distribution (cdf). With a pdf format the respondent is faced with assessing the probabilities that the outcome lies in each interval (e.g. the 1989 SHIW question cited earlier) while with a cdf format he has to assess the probabilities that the outcome does not exceed the sequence of thresholds (e.g. as in [Attanasio and Augsburg 2012](#); also the question cited in the introduction).

Most studies have been eliciting cdfs although lately an increasing number of questions are being framed as pdfs (for examples of pdf questioning see [Arrondel et al. 2011](#), the New York Federal Reserve inflation question in [Bruine de Bruin et al. 2011b](#); [Delavande et al. 2011](#)). [Morgan and Henrion \(1990\)](#) cite experimental evidence reporting that individuals find it easier to deal with pdfs that allow an easier visualization of certain properties of the distribution like location and symmetry. Traditionally, the larger probabilities involved in cdfs was thought to help respondents.

An alternative to eliciting probabilities in the form of cdfs or pdfs is to ask for quantiles of the distribution, for example, the respondent is prompted to provide a value  $X$  such that there is a 25 % chance of her income being less than  $X$ . Early on both [Morgan and Henrion \(1990\)](#) and [Dominitz and Manski \(1997\)](#) rejected eliciting quantiles citing evidence that probabilities assessed in this way match less well empirical frequencies.

*Last but not least, knowledge about the subject matter* There are two basic considerations for successfully eliciting probabilistic expectations. The respondent should have knowledge about the event or outcome to be assessed as well as some skills in expressing beliefs in probabilistic form.<sup>11</sup> Although the later condition may often seem difficult to satisfy, there have been advances in learning forms of elicitation that may be easier for the respondent as we have discussed above. However, lack of knowledge about the subject matter may prove more difficult to overcome. This may be the case, for example, when trying to elicit stock market return expectations from low income and low education households. For many people mutual fund returns are not part of their lives and hence they lack knowledge of the subject matter which is a necessary condition for individuals to be able to express meaningful beliefs about it. Subjects in general know a lot about themselves but much less about aggregate circumstances.

<sup>10</sup> [Delavande et al. \(2011\)](#) use 20 intervals with predetermined support and four with a self-anchored one when eliciting expectations about the respondent's expected fish catch. [Attanasio and Augsburg \(2012\)](#) work with four intervals and self-anchored support when eliciting the cdf of expected individual income. Both studies were done in India. [Hurd et al. \(2011\)](#) and [van der Klaauw et al. \(2008\)](#) elicit expectations about aggregate variables (Dutch stock returns and US inflation, respectively) and define eight intervals with predetermined support.

<sup>11</sup> See [Delavande et al. \(2011\)](#) for examples of supporting evidence.

### 2.3 Validation diagnostics

*Response rates* Individuals are willing to answer probabilistic expectation questions. Response rates in many cases are high (e.g. 97 % in [Attanasio and Augsburg 2012](#), 99 % in [Bruine de Bruin et al. 2011a](#), 79–87 % in [Hurd et al. 2011](#)) and higher than for actual or historical outcomes in the same surveys. But non-response varies substantially with the matter being elicited. For example, in the 2006 HRS non-response was 4 % for the expected survival probability question but 24 % for the expected gain in the stock market.<sup>12</sup>

*Coherence* However, a major concern has been whether the answers obtained could really be interpreted as the respondent's subjective beliefs about uncertain outcomes. Therefore, in all studies some time is spent analyzing coherence of the responses in various ways. In the first place, checks to verify compliance with basic probability laws are usually reported. Authors working with cdf formulation type questions report a varying degree of monotonicity violations. In some cases high compliance is achieved with the help of a programmed automatic prompting in case of violation. [Dominitz and Manski \(1997\)](#) report around 10 % of monotonicity violations before the prompt and 5 % afterwards while [Attanasio and Augsburg \(2012\)](#) report 1 % without the help of such prompting. Automatic warnings for additivity violation (i.e. if probabilities or beans do not sum up to the required amount) in pdf questions are also useful.<sup>13</sup> [Bruine de Bruin et al. \(2011a\)](#) report other checks to support the validity of responses like the fraction of respondents who put positive probability mass in more than one bin (96.4 %) or the low fraction who put positive probability mass in non-contiguous bins (1.3 %) although some people may have bimodal beliefs.

*Correlations and predictive power* Correlations with other survey variables may sometimes provide information about the soundness of expectation answers. [Attanasio and Augsburg \(2012\)](#) make use of the standard preliminary question about the likelihood of rain. This question is often carried out to convey the idea of probability to respondents to further check the expected income distribution data they obtain from households in rural India. They find a significant correlation between the answers to the likelihood of rain and expected income for households whose main income is derived from agriculture and no significant correlation for those that do not. More routinely, assessing how answers to subjective probabilities vary with socio-demographic characteristics of the respondent (i.e. compliance with prior beliefs about correlates of expectations), is often seen as part of the validation of the data.

Predictive power is a desirable feature for the credibility of elicited expectations. However, beliefs may be inaccurate but nevertheless be the relevant measure behind observed behaviour. In many different surveys individual expectations about stock market gains have been found to be substantially lower than what observed past (and future) averages would justify. Additionally, young educated males are found to systematically hold more optimistic expectations about the stock market than other groups

<sup>12</sup> As expected, non-response is lower for stockholders (11 %) than for those not owning stocks (29 %).

<sup>13</sup> To some extent the need for prompts is a reflection of the limitations of the device used in implementing the question. For example, a prompt would not be necessary if the respondents were actually given ten balls to distribute using a mechanical or an electronic device.

(see [Hurd 2009](#), for this and other examples). Moreover, beliefs about stock market gains correlate with ownership of stocks.

**Rounding** Rounding of responses to the nearest 5 % is often reported although at the tails respondents may round to the nearest 1 % (see for example [Dominitz and Manski 1997](#); [Hudomiet et al. 2011](#); [Attanasio and Augsburg 2012](#)). Rounding may be influenced to some extent by the design of the visual aid attached to the question, for example, marks on a ruler.

**Epistemic uncertainty (ignorance about probabilities)** More importance has been given to the bunching of responses at 50 % for the expected probability of a binary event (e.g. the percent chance of a positive stock market return or the probability for a 70 years old person to live to at least the age of 80). Psychologists have reported that a 50 % reply may disguise a “don’t know” answer and reflect epistemic uncertainty, that is, the tendency to choose towards the middle of a scale when the respondent is not able to provide an answer or does not understand the question. Alternatively, such answers could reflect a genuine belief that the event is equally likely to occur or not to occur (see [Fischhoff and Bruine de Bruin 1999](#), for an early paper on the subject).

In order to disentangle responses that reflect a genuine probability belief from those reflecting epistemic uncertainty some studies have included a follow up question in the case of a 50 % answer. In 2006 the HRS added such an epistemic follow up question to some of the probability questions, which revealed that, for example, the fraction of 50 % answers to the survival probability question being simply ignorance (i.e. being unsure about the chances) was as high as 60 %. The HRS formulation of the follow up question for the percent chance of an increase in the value of mutual fund shares was: “Do you think that it is about equally likely that these mutual fund shares will increase in worth as it is that they will decrease in worth by this time next year or are you just unsure about the chance?”.

In contrast, [Dominitz and Manski \(2007\)](#) provide some evidence that such answers could reflect a genuine belief that the event is equally likely to occur or not to occur. In particular, they show that persons answering 50 % to the 2004 HRS question about their perceived percent chance of a positive stock return hold more stocks than persons with lower expected probabilities but less than persons with higher expected probabilities. They infer therefore that such answers reflect a higher perceived chance of a positive stock return than less than 50 % answers but lower perceived chance of a positive stock return than more than 50 % answers.

**Heaping** Heaping at 0 and 100 reported but this is usually less problematic than at 50 %. A high number of 0 and 100 responses probably reflects absence of precise beliefs and therefore some uncertainty. However, they convey the information that the chances of the event occurring are thought to be extremely low or extremely high. In any case focal answers at 0, 50, 100 reflect less precisely known probabilities than non-focal ones. [Lillard and Willis \(2001\)](#) find that the tendency to give focal answers is associated with lower cognitive ability. [Hurd et al. \(2011\)](#) find in their data a fraction of “50 %-respondents” lower than in many other surveys and attribute this to the fact that Dutch CentER Panel members are experienced survey respondents.

In the context of eliciting expected distributions of continuous variables (either cdf or pdf formulation) too many answers of 0 % (100 %) chance of the outcome to be

higher than the lowest (highest) threshold may sometimes indicate that the chosen range is not adequate.

*Addressing Kahneman’s critique* One critique to collecting subjective probabilistic expectations is that respondents would not apply much effort and hence would not provide thoughtful answers. In Kahneman’s dual system terminology, respondents will tend to use intuition (system 1) and not reasoning (system 2). [Gouret and Hollard \(2011\)](#) take this criticism seriously and try to separate the fraction of respondents that do provide valuable information about expected mutual fund return distribution. To achieve this they construct a coherency measure and show that only for the most coherent individuals there is a significant monotonic relationship between expected returns and perceived risk. They find that their measure of coherency correlates with education and income.

In contrast, the results in [Zafar \(2011\)](#), analyzing a panel dataset of Northwestern University undergraduates that contains subjective expectations about major specific outcomes, support the hypothesis that students exert sufficient mental effort when reporting their beliefs.

However, in some cases, the problem may not lay in not exerting enough mental effort but in the wording of survey questions making it easy for some respondents to express their probability beliefs.

## 2.4 Some uses of subjective probability questions

An important motivation for introducing expectation questions in household surveys is to help explain household choices. Another still undeveloped use of individual responses is the construction of statistics like, for example, statistics about inequality in expected survival probabilities.<sup>14</sup>

Although there are already important studies that make use of subjective probabilities to explain economic behavior, a large proportion of the literature to date has focused on assessing the properties of the elicited information and establishing its validity. Further to the basic validation checks described previously, this literature has analyzed variation in subjective probabilities across individuals and their predictive power on outcomes.

To illustrate research work that uses subjective expectations survey data, I will briefly review findings regarding three questions: survival probability, probability of positive stock return, and expected inflation distribution.<sup>15</sup>

*Survival probability* The expected probability of survival to age 75 was introduced early on in the 1992 HRS.<sup>16</sup> Data from the first wave did show that the average survival probability was very similar to the 1990 survival rate from life tables. Once a second

<sup>14</sup> It would be interesting for example to see if heterogeneity in household expected survival probabilities is very different to heterogeneity in realized mortality.

<sup>15</sup> See [Manski \(2004\)](#) and [Hurd \(2009\)](#) for more detailed reviews on uses of expectation questions.

<sup>16</sup> Other subjective probability questions introduced in the 1992 HRS wave dealt with expectations about retirement age, health limitations, inflation, health care expenditures, unemployment, housing prices, Social Security benefits, giving financial help, and economic depression. A question about the expected probability of a positive stock return was added in 2002.

wave was available in 1994 subjective survival probabilities elicited in 1992 were proved to be a good predictor of mortality for the period between the two waves. This has been also true in the European SHARE (see [Winter 2008](#)). Moreover, after few years, it was established that elicited survival probabilities and actual mortality data correlate with variables like education, wealth, income etc. in a similar way. In general, as [Hurd \(2009\)](#) points out, subjective probabilities have “predictive power” when individuals have considerable private information about the subject matter. Indeed, predictive power in itself may not be as interesting as indirectly getting insight about private information.

Some work has also been done on using expected survival probability to explain economic behaviour. For example, [Hurd et al. \(1998\)](#), using the survey of the Asset and Health Dynamics among the Oldest Old (AHEAD), find that the probability of saving correlates in a significant and substantial way with individual subjective beliefs about their own mortality risk but not, when jointly included, with life-table probabilities. Using the HRS, [Hurd et al. \(2004\)](#) study whether individuals who expect to be long-lived claim Social Security benefits later than those expecting to be more short-lived. Although they find effects in the expected direction, their size is modest in general but increases with education. Finally, [Gan et al. \(2004\)](#) compare the ability of expected survival probability in predicting out of sample wealth with life-tables using a life-cycle model of consumption.

*Expectations about stock market return* Subjective expectations about stock market returns have proven to be useful in helping resolve the stock holding puzzle. Under the traditional assumption of rational and homogeneous expectations, observed low rates of stockholding would be attributed to high risk aversion. However, elicited data show that subjective stock return expectations are very heterogeneous and that this heterogeneity helps explain participation in the stock market (while there is no evidence of a risk aversion effect).<sup>17</sup> Individuals having more optimistic beliefs about returns are more likely to hold stocks. This effect was first found in [Dominitz and Manski \(2007\)](#) and has been confirmed by other authors in various contexts ([Hurd et al. 2011](#); [Hudomiet et al. 2011](#); [Arrondel et al. 2011](#)). Importantly, those heterogeneous beliefs seem to present systematic biases. Individuals are found to be more pessimistic about rates of return than the historical performance of the stock market (see evidence in [Hurd et al. 2011](#) for the Netherlands and [Kézdi and Willis 2008](#), for the US) and men are consistently found to be more optimistic than women. Observed heterogeneity in stock market expectations raises an important question about how beliefs are formed and what are the reasons behind such systematic differences given that information about stock prices is public and there is no private information.

*Inflation expectations* Household expected inflation is assumed to feed into realized prices if households take inflation into account when deciding about their purchase of large durables, saving instruments, wage negotiations, etc. Given this role of inflation expectations in the monetary transmission mechanism it is widely agreed that in order to control inflation it is important to learn about people’s beliefs concerning future inflation.

<sup>17</sup> Uncertainty about those expectations is also found to be heterogeneous when data about expected distributions are available.

For a long time many household surveys have asked point forecasts of expected inflation (e.g. the Michigan Panel, the Bank of England/NOP Inflation Attitudes Survey) but without eliciting related uncertainty.<sup>18,19</sup> For example, the Bank of England/NOP survey question is the following:

“How much would you expect prices in the shops generally to change over the next 12 months?”.

In 2007 the Federal Reserve Bank of New York (FRBNY) began to develop a survey to measure and analyse consumers’ inflation expectations.<sup>20</sup> In this survey, carried out every six weeks approximately, the full expected distribution is elicited asking respondents about the percent chance of inflation in the next 12 months being in eight separate intervals. After instructions, the wording of the question is as follows:

“What do you think is the percent chance that, during the next 12 months, the following things will happen? Prices in general will:

go up by 12 % or more \_\_\_\_\_ percent chance  
 go up by 8 to 12 % \_\_\_\_\_ percent chance  
 go up by 4 to 8 % \_\_\_\_\_ percent chance  
 go up by 2 to 4 % \_\_\_\_\_ percent chance  
 go up by 0 to 2 % \_\_\_\_\_ percent chance  
 go down by 0 to 2 % \_\_\_\_\_ percent chance  
 go down by 2 to 4 % \_\_\_\_\_ percent chance  
 go down by 4 % or more \_\_\_\_\_ percent chance  
 (100 % Total)”

Armentier et al. (2013) present various validation diagnostics for this question. For their experimental panel survey, non-response rate is less than half a percentage point, the proportion with positive probability in more than one bin is 89.4 % and the proportion with positive probability in non-contiguous bins is 1.6 %.

There is considerable heterogeneity across respondents in median forecasts which are higher for respondents who are women, less educated, poorer, single, or older. When conditioning for all demographics only education remains significant but when further controlling for financial literacy the effect of education is reduced.

Moreover, as we will see in detail below in Sect. 2.6, the authors find coherency between individual inflation expectations and financial choices. Related with the findings on the effect of education and literacy, these data reveal the inability of some groups of the population to form sensible expectations. The results are also indicative of the economic effects expectations of poor quality may have.

<sup>18</sup> An exception is the Bank of Italy Survey of Household Income and Wealth who elicited the expected inflation distribution in their 1989 and 1991 waves.

<sup>19</sup> There are also indirect ways to infer inflation expectations from the term structure of interest rates or from financial instruments but with some strong modelling assumptions.

<sup>20</sup> Until 2012 the survey was conducted over the internet with RAND’s American Life Panel.

Uncertainty about future inflation is positively related to mean and median expected inflation. Moreover, using the panel dimension of the survey, respondents who are more uncertain are found to make larger revisions to their expectations in the next survey (see [Bruine de Bruin et al. 2011a](#); [van der Klaauw et al. 2008](#)).

## 2.5 Expectation formation

The availability of data on individual subjective expectations has prompted renewed interest in analyzing their determinants and the amount of information households use when forming those expectations.

*Testing for rational expectations* There has been work with individual expectations data testing models of the way expectations are formed and in particular testing for rational expectations. When considering expectations over variables for which the individual has substantial private information (e.g. educational attainment, mortality risk) and in some cases are under his control up to some extent (e.g. retirement age) the rational expectations hypothesis cannot be rejected.<sup>21</sup> [Benítez-Silva et al. \(2008\)](#) test for rational expectations in the formation of retirement and longevity expectations using the Health and Retirement Study (1992–2002) and of educational attainment expectations using the National Longitudinal Survey of Youth (1979–2000). In their framework this amounts to testing that differences in expectations in successive periods cannot be forecast.<sup>22</sup> Using instrumental variables for measurement error and accounting for sample selection the authors cannot reject the rational expectations hypothesis.

Following a similar methodology [Das and Donkers \(1999\)](#) analyze the answers about expected income growth in the Netherlands' Socio-Economic Panel but they reject the hypothesis that these expectations are rational and find instead that households are excessively pessimistic about their future income growth. However, the force of the evidence is limited by the fact that expectations in that survey are elicited in a more qualitative way than in the HRS or the NLSY. In particular the set of possible answers are: “strong decrease”, “decrease”, “no change”, “increase”, “strong increase”.

House price change is a relevant variable for the macroeconomy that has been elicited in a few household surveys. The question may refer to house prices at the national level or at a more disaggregate level (area, own house) for which households may have more information. [Case et al. \(2012\)](#) test rationality of area house price expectations by regressing future house price change on the expected change. One-year price expectations are found to under-react to information while 10-year expectations seem likely to have been over-reacting although this longer term rationality is still difficult to assess with the authors survey data for the 2003–2012 period.

*Expectations about macro variables* A recent literature on this topic has been focusing on the study of individual expectations (or “sentiment”) about macroeco-

<sup>21</sup> For a detailed exposition of using survey expectation data for testing models of expectation formation see [Pesaran and Weale \(2006\)](#).

<sup>22</sup> However, a model free test may not be easy to perform.



conomic variables where there is public information but no individual information (e.g. inflation, house prices, stock returns). In those cases expectations are found to be systematically biased and the literature has unveiled heterogeneity in various dimensions.<sup>23</sup> Men, individuals who are young, highly educated, with high income are more optimistic and believe inflation will rise at a slower pace (Bruine de Bruin et al. 2010). However, these systematic biases in people's expectations are not constant over time (Souleles 2004). Similar findings are obtained by looking at expected stock returns (Dominitz and Manski 2007): there is variation in the empirical distributions over time and men report higher expected returns than women (and the young higher than the old).

A relevant question is therefore what could explain these demographic differences in expectations. Regarding inflation we have learned (see for example Bruine de Bruin et al. 2010) that inflation expectations are higher among respondents who thought relatively more about how to cover expenses and about specific prices, and among those with low financial literacy. Perceptions of past inflation are a major determinant of inflation expectations (see Blanchflower and MacCoille 2009, using UK data) but this is less so for individuals with high education. Cavallo et al. (2014) find that an individual's expectations are influenced both by inflation statistics and supermarket prices albeit more by the latter that are less costly to understand. Another finding regarding heterogeneity and biases in household inflation expectation is that individuals report biased beliefs on inflation in part because they use their price memories or other private information rather than inflation statistics. Moreover, this would mean that observed heterogeneity in household expectations reflects heterogeneity in individual beliefs rather than measurement error.

*Differences between consumers and professional forecasters* There have also been some results about patterns in individual expectations over time abstracting from the cross-sectional dimension of the data. Carroll (2003) finds that differences between professional forecasters and consumers narrow when inflation is more significant, probably due to increased coverage of the matter in the media and increased household interest who would improve their expectations when inflation matters. An alternative sticky-information model explanation (in Mankiw et al. 2003), by which economic agents do not update their information continuously because of the cost of collecting and processing the information, does not explain the positive association found between the level of inflation and the extent of the disagreement between consumers and professional forecasters.

## 2.6 Expectation experiments

*Do individuals act on their inflation beliefs?* To validate elicitation of inflation expectations data one would like to have evidence that reported beliefs on future inflation help explain financial decisions. This is especially relevant in a low inflation environment. Indeed, it may be argued that consumers may not act on their inflation beliefs

<sup>23</sup> There are older well known applications of the idea that individual agents may have incomplete aggregate information (Phelps 1970; Lucas 1963).

because the impact of future inflation is not sufficiently salient or because they may suffer from money illusion.

In an innovative paper [Armantier et al. \(2013\)](#) compare the behavior of consumers in a financially incentivized investment experiment with the beliefs they self-report in an inflation expectation survey. More precisely, respondents are first asked about their inflation beliefs as usually elicited in the FRBNY Survey. Several questions later they are asked to choose among different investment options in which the payoffs depend on future inflation. In particular, for each of the ten available choices, they are presented with two options: one where the payoff depends on inflation over the next 12 months and another where the payoff is fixed. The idea is to look at how reported expectations in the survey correlate with their decisions in the investment experiment.

The experiment was incentivized. Two participants randomly chosen would be paid one year later according to the investment choices they made in the experiment (which in turn were influenced by their inflation expectations).

An important characteristic of the design of this experiment is that when respondents reported their inflation expectations they were not aware of the experiment in which payoffs depend on future inflation.

Data on numeracy and financial literacy as well as a self-reported measure of risk tolerance are also collected as part of the survey.

The conclusion is that on average there is a high correspondence between reported beliefs and behavior in the experiment, and the substantial amount of heterogeneity across respondents can largely be explained by the respondent's self-reported risk tolerance. Moreover, when considering changes in beliefs over time for the same respondent, the adjustment in experimental behavior is mostly consistent with expected utility theory. Finally but importantly, individuals whose behavior is difficult to rationalize tend to obtain low scores on numeracy and financial literacy questions and are less educated.

*Revising expectations* Research that analyzes revisions to expectations in association with interim events or information may provide clues about how people form their expectations (as first advocated by [Manski 2004](#)). [Armantier et al. \(2013\)](#) carry out an information experiment embedded in one of the regular New York Federal Reserve Bank Surveys along those lines. They first elicit expectations for future inflation, then randomly provide a subset of respondents with information relevant to inflation (either past-year average food price inflation or professional economists' median forecast of the year ahead inflation), and finally expectations are re-elicited from all respondents. The findings are that respondents do revise their inflation expectations in response to information and that they do so in a meaningful way. In particular revisions are in the direction of the information provided and proportional to the prior perception gap and to the uncertainty of initial expectations. Moreover, updating behavior is heterogeneous with women updating more substantially than men and individuals with low education, low income, low financial literacy being more responsive to information treatment than their counterparts. These are the demographic groups who initially had the higher perception gaps and the more uncertain expectations. This leads the authors to advocate for a potential role for policies that incorporate public information campaigns.

### 3 Subjective house price expectations in the Spanish Survey of Household Finances

#### 3.1 The EFF and its house price expectation question formulation

The Spanish Survey of Household Finances contains detailed information on household assets, debts, income and consumption and has now been conducted on five occasions (2002, 2005, 2008, 2011, and 2014).<sup>24</sup> The EFF was specially designed for the study of household wealth. While providing a representative picture of the structure of household assets and debt it incorporates an oversampling of wealthy households based on individual wealth tax files. In addition, there is an important panel component while the sample is being refreshed at each wave to maintain current population representativity. The sample size is around 6000 households, the exact number depending on the wave. Questions on assets, debts, consumption refer to the household as a whole while demographics and labour income information is available for each of its members. The person answering the survey is the one who is most knowledgeable about the household finances although very often help is provided from other members to answer individual specific information. The survey is administered by a computer assisted face to face interview.

Starting in the EFF2011 a new question to elicit household house price expectations was introduced. The motivation behind is the importance of real estate assets in household wealth (80 % of the value of household assets) all along the wealth distribution (88 % for the bottom quartile and 67.5 % for the top decile). Aside from a high proportion of owner occupier households (83 %), 36 % of Spanish households hold some other real estate property. Aggregate expectations about rates of return on housing have been found to be an important determinant of house purchase (see [Bover 2010](#)). Moreover, uncertainty about that return has also been found to play a role. Learning about household house price expectations at the individual level may be therefore useful in understanding portfolio composition as well as consumption behavior.

Other surveys eliciting subjective expectations about house prices are the HRS and ELSA targeted to the over 50 years of age households, the NYFRB internet survey, and the Asset Price and Expectations module in the ALP. The introduction of this question is in all cases very recent: 2011 in the ALP module and 2010 in the case of the HRS and the NYFRB survey. This paper is one of the first attempts to analyze answers to this type of questions.<sup>25</sup>

The person answering the 2011 EFF questionnaire was asked the following:<sup>26</sup>

*We are interested in knowing how you think the price of your home will evolve in the next 12 months: distribute ten points among the following five possibilities, assigning*

<sup>24</sup> Typically the fieldwork takes place during the last 3 months of the named year and the first 4 months of the next one with at least half of the interviews being conducted before the end of the named year.

<sup>25</sup> After writing and presenting the first version of this paper I learned of independent work in [Niu and van Soest \(2014\)](#).

<sup>26</sup> The original Spanish formulation is as follows:

*more points to the scenarios you think are more likely (assign 0 if a scenario looks impossible)*

*Large drop (more than 6 %)*

*Moderate drop (around 3 %)*

*Approximately stable*

*Moderate increase (around 3 %)*

*Large increase (more than 6 %)*

*Don't know*

*No answer*

Several comments are in order. The question refers to the price of the household main residence because of the belief that households have more information about their own house than about prices of houses in the area or nationwide. Moreover, answers provide information about unobservables and heterogeneity in the housing market even if people were to have plenty of information about aggregates. A sentiment about house prices nationwide could be inferred by aggregating from a representative sample like the EFF although these are of course different questions. The question was posed to all households and not only to home owners. When eliciting the subjective distribution numerical answer options are provided together with verbal descriptions. The number of intervals among which the probability mass is distributed is five and it was preferred to offer the respondent ten points to distribute as opposed to 100 because it is cognitively less demanding. For the same reason it was chosen to elicit the distribution using a density formulation rather than a cumulative distribution. Respondents are also handed out a sheet of paper containing the question and the response options on which they could draft their answers. Explanations are provided by the interviewer when needed. Finally, an automatic prompt would appear on the screen whenever the answers entered in the computer by the interviewer do not add up to ten. In such cases the household and the interviewer are asked to revise the answers.

The elicitation specificities in other surveys containing house price expectation questions are diverse. The HRS asks about own house price expectations (to owners only) using a cdf formulation with four cut-off points. The ALP module refers to house price in the area for renters and own home values for owners and has a pdf type of question with three intervals (two of them open ended). Finally the NYFRB survey asks about prices of a typical home in their zip code and follows their usual ten interval pdf formulation.

---

Footnote 26 continued

*“Estamos interesados en conocer cómo cree usted que evolucionará el valor de su vivienda en los próximos doce meses:*

*Reparta 10 puntos entre las cinco posibilidades siguientes, asignando más puntos a los escenarios que crea más probables (asigne cero puntos si alguno le parece imposible):*

*Caída grande (más de 6 %)*

*Caída moderada (en torno a 3 %)*

*Aproximadamente estable*

*Subida moderada (en torno a 3 %)*

*Subida grande (más de 6 %)*

*No sabe*

*No contesta”*

**Table 1** Don't know/no answer and bunching in the middle: demographic characteristics

Variables	(1) Number of respondents	DK/NA (%)		Bunching in the middle (%)	
		(2) Sample	(3) Population	(4) Sample	(5) Population
Women	2442	6.18	6.93	19.57	20.53
Men	3664	2.78	3.44	18.23	17.86
Primary educ.	2767	6.98	7.46	19.48	19.48
Secondary educ.	1466	2.32	2.47	18.62	17.99
University educ.	1851	1.40	1.67	17.83	19.47
Age under 35	279	2.87	2.26	16.85	20.48
Age 35 to 44	763	3.01	3.55	18.48	16.26
Age 45 to 54	1177	2.97	3.86	17.33	16.41
Age 55 to 64	1274	2.20	3.72	18.13	20.26
Age over 64	2613	6.08	8.85	20.02	21.97
Owner occupiers	5326	3.22	3.56	18.42	18.71
Non-owner occupiers	780	10.70	12.21	21.14	20.77
Total	6106	4.14	5.02	18.77	19.06

With the exception of the ALP, the previous surveys formulate their house price expectation question in terms of rates of change (as opposed to levels). In the EFF given that households provide a self-assessed current value for their home one could also derive the expected level of house price in 12 months time using the expected rate of change.

### 3.2 Item non-response

Only 4.1 % of households who participated in the EFF2011 did not answer the house price expectation question.<sup>27</sup> Table 1 (columns 2 and 3) provides some breakdown by demographic characteristics of the respondent. Sample shares are discussed in the text but the corresponding estimated shares for the population are also contained in Table 1 columns 3 and 5.

This percentage is higher for non-owner occupiers (10.7 %) than for owners of their main residence (3.2 %). In any case it compares favorably with the 2006 HRS response rates to an expected stock returns question, to which 24 % of households did not respond, suggesting how unfamiliar the stock market is for many households. Even among stockholders non response was 11 % (and 29 % for non-stock holders).

Men are more prone to answering the question than women (2.8 vs. 6.2 % non-response) and non-response rates decrease with education (7 % for individuals with up to primary education, 2.3 % for those with secondary education, and 1.4 % in case of holding a university degree). By age, the non-response of the over 64 stands out.

<sup>27</sup> Taking into account population weights the estimated percentage in the population is 5 %.

**Table 2** Observed answers and characteristics: multiple regressions

Variables	(1)	(2)	(3)	(4)	(5)
	dk/na	+ or – same	Points to increase ( $\times 10$ )	Points to drop ( $\times 10$ )	All points to one drop interval
Woman	0.025**	–0.010	0.959	–1.785	–0.032**
Secondary educ.	–0.010	0.019	–0.661	0.638	0.017
University educ.	–0.005	0.052	–1.938	–2.218	–0.028
Age 35–44	0.001	–0.009	–1.148	6.966	0.053
Age 45–54	0.043***	0.028	–0.694	–0.171	0.034
Age 55–64	0.034*	0.086*	–1.906	–3.583	0.042
Age over 64	0.061***	0.081*	–3.105	–3.703	0.036
Blue collar	0.005	0.026	–0.223	–2.910	–0.018
Self-employed	–0.008	0.021	–1.463	1.785	–0.015
Log (hh income)	–0.020**	0.002	–0.461	2.608*	0.013
Wealth percentiles					
25–50	–0.003	0.009	–0.512	–1.857	0.041
50–75	–0.024	0.041	0.034	–3.903	0.021
75–90	–0.005	0.017	0.265	–1.987	0.024
90–100	–0.010	0.002	–0.520	–0.845	–0.011
Bought main recently	–0.004	0.010	2.858**	–4.256	–0.008
Own other housing	–0.004	–0.014	–0.402	2.609	0.064***
Constant	0.233**	0.142	11.323	27.260	–0.050
Observations	5326	5326	5326	5326	5326

t-statistics in parentheses

All specifications include postal dummies and have been estimated taking into account population weights and the five imputed datasets

\*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

Table 2 (column 1) presents results from a multiple regression including income and wealth variables as well.

In the EFF I construct various measures to assess the amount of questions the household has provided an answer for. Among others, I calculate the percentage of monetary questions that have been answered with a point value (as opposed to an interval) as the ratio of exact answers to total questions posed to the households. The correlation of this precise information ratio with not having answered the house price expectations question is  $-0.10$  ( $-0.17$  with a t-ratio of 8.2 in a simple regression). Not answering the house price expectation question also correlates significantly with not having been able to provide an estimate of the current value of their home (0.10; 0.05 with a t-ratio of 7.4 in a simple regression).<sup>28</sup>

<sup>28</sup> Only homeowners are asked to provide an estimate of how much their house is worth.

### 3.3 Coherency analysis

*Bunching in the middle of the scale* The percentage of respondents placing all ten points in the middle-of-the-scale option is 18.8 %. For reference, in the 2006 HRS 23 % of respondents chose the middle of the scale to the question on survival probability to age 75 and 30 % chose it as a response to a question about the probability of stock market gains.<sup>29</sup>

There is certain heterogeneity by demographic groups (see Table 1, columns 4 and 5). Among home-owners 18.4 % chose this answer while the share among non home owners is 21.1 %. There is also some variation by education (varying from 19.5 % for respondents with no secondary education to 17.8 % in the case of University educated respondents). By gender there are some differences as well (18.2 % in the case of men, 19.6 % for women). Differences by age are less noticeable (ranging from 16.8 % among the under 34–20 % among the over 64). In a multiple regression (see Table 2, column 2) only being aged over 64 has a significant (positive) effect on bunching. All in all these are small differences across groups, which is suggestive of bunching driven by beliefs more than by ignorance, except may be for the older respondents.

The correlation between the constructed information ratio variable and choosing to put all ten points in the middle of the scale is not significant (0.004 and 0.01 with a t-ratio of 0.31 in a simple regression). Along the same lines, the correlation with not being able to provide a value of their home is not significant either (−0.002 and −0.002 with a t-ratio of 0.13 in a simple regression).

The effects of demographic variables do not work in the same direction as in the case of non-response and are much less significant in this case despite the sizeable number of such respondents (Table 2, column 2). This may indicate that there are different factors at work. Namely, while a fraction of individuals giving all ten points to the approximately no house price change option may do so because they are unable to express beliefs about the future path of house prices there are others who strongly believe (i.e. put all ten points) that the price of their house will experience no change over the next 12 months (see more details on epistemic uncertainty in Sect. 2.3). The absence of correlation with the information ratio and with not answering the current value of their house points in this direction as well. Unfortunately, I cannot separate the two types of answers because in the EFF the house price expectation question is not followed by one trying to disentangle ignorance from genuine belief of no change in house prices.

*Number of intervals used* 61 % of the respondents express uncertainty and put some probability mass in more than one interval while 28 % of all respondents use more than two intervals (see Table 3). Only 6.32 % use all five intervals.

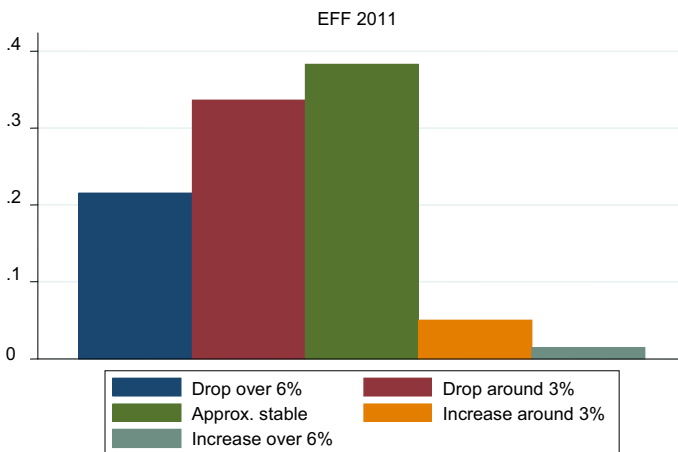
*Using non-adjacent intervals* There is a very small fraction of respondents (1.6 %) that assign non-zero probabilities to non-adjacent intervals.

<sup>29</sup> In the HRS survival probability question answering the middle of the scale corresponds to a 50 % chance answer.



**Table 3** Bins used (%)

No of bins used	Sample	Population
1	38.55	36.17
2	33.62	35.01
3	17.13	16.26
4	4.37	5.31
5	6.32	7.24
Using non-adjacent bins	1.59	1.66

**Fig. 1** Average probability distribution of future HP EFF 2011

### 3.4 Preliminary analysis

*Average histogram and most frequent answers* Figure 1 shows an average histogram showing the percentage probability mass in each of the five predefined intervals of the density function. The figure shows that respondents overwhelmingly put most of the probability mass in the expected drop-in-price region. Therefore, Spanish households at the end of 2011 were in general not expecting increases in the price of their homes over the next 12 months.<sup>30</sup> But importantly, around this average of distributions there is a large heterogeneity in individual subjective probability distributions. To provide more detail about the pattern of answers, Table 4 shows the most frequent answers up to 90 % of the cumulative sample distribution. The ten most frequent answers collectively account for 60 % of the sample.

*Probability of a positive return* I calculate the respondent probability of a positive change in house prices as the sum of the number of points attributed to intervals 4 and 5 (i.e. to a moderate increase of around 3 % and a large increase of over 6 %). A

<sup>30</sup> Aggregate house prices had been falling in Spain since 2007.

**Table 4** Most frequent answers to probabilistic expectations of future house prices

	– > 6 %	– around 3 %	+ or – same	+ around 3 %	+ > 6 %	Frequency	Percent	Cumpercent
1.	0	0	10	0	0	1146	19.58	19.58
2.	0	10	0	0	0	583	9.96	29.54
3.	10	0	0	0	0	558	9.53	39.7
4.	0	5	5	0	0	382	6.53	45.60
5.	5	5	0	0	0	323	5.52	51.12
6.	8	2	0	0	0	125	2.14	53.25
7.	0	3	7	0	0	116	1.98	55.24
8.	0	2	8	0	0	111	1.90	57.13
9.	6	4	0	0	0	91	1.55	58.69
10.	0	8	2	0	0	79	1.35	60.4
11.	3	7	0	0	0	79	1.35	61.39
12.	2	8	0	0	0	73	1.25	62.63
13.	0	0	5	5	0	73	1.25	63.88
14.	7	3	0	0	0	70	1.20	65.8
15.	0	6	4	0	0	70	1.20	66.27
16.	0	4	6	0	0	65	1.11	67.38
17.	0	7	3	0	0	64	1.9	68.48
18.	4	6	0	0	0	64	1.9	69.57
19.	2	6	2	0	0	57	0.97	70.55
20.	5	3	2	0	0	56	0.96	71.50
21.	4	4	2	0	0	55	0.94	72.44
22.	0	0	0	10	0	54	0.92	73.36
23.	3	5	2	0	0	52	0.89	74.25
24.	2	3	5	0	0	52	0.89	75.14
25.	2	5	3	0	0	51	0.87	76.1
26.	6	3	1	0	0	47	0.80	76.82
27.	2	4	4	0	0	36	0.62	77.43
28.	0	0	8	2	0	35	0.60	78.3
29.	3	3	4	0	0	32	0.55	78.58
30.	2	2	6	0	0	31	0.53	79.10
31.	0	1	9	0	0	31	0.53	79.63
32.	0	2	6	2	0	28	0.48	80.11
33.	4	3	3	0	0	27	0.46	80.57
34.	5	4	1	0	0	26	0.44	81.2
35.	9	1	0	0	0	26	0.44	81.46
36.	2	7	1	0	0	23	0.39	81.86
37.	3	4	3	0	0	23	0.39	82.25
38.	6	2	2	0	0	22	0.38	82.62
39.	3	6	1	0	0	21	0.36	82.98
40.	1	2	7	0	0	21	0.36	83.34

**Table 4** continued

	– > 6%	– around 3%	+ or – same	+ around 3%	+ > 6%	frequency	percent	cumpercent
41.	0	0	7	3	0	20	0.34	83.68
42.	0	3	5	2	0	20	0.34	84.3
43.	1	7	2	0	0	20	0.34	84.37
44.	7	2	1	0	0	20	0.34	84.71
45.	0	3	4	3	0	18	0.31	85.2
46.	0	5	3	2	0	18	0.31	85.32
47.	4	5	1	0	0	15	0.31	85.63
48.	0	0	0	5	5	17	0.29	85.92
49.	0	3	6	1	0	16	0.27	86.20
50.	1	8	1	0	0	16	0.27	86.47
51.	1	2	5	2	0	16	0.27	86.74
52.	2	3	4	1	0	16	0.27	87.2
53.	8	1	1	0	0	14	0.24	87.25
54.	0	0	0	0	10	13	0.22	87.48
55.	3	2	5	0	0	13	0.22	87.70
56.	1	6	3	0	0	13	0.22	87.92
57.	0	9	1	0	0	13	0.22	88.14
58.	1	4	5	0	0	12	0.21	88.35
59.	1	3	5	1	0	12	0.21	88.55
60.	0	0	4	6	0	12	0.21	88.76
61.	0	1	8	1	0	12	0.21	88.96
62.	0	2	7	1	0	11	0.19	89.15
63.	2	4	3	1	0	11	0.19	89.34
64.	0	2	5	3	0	11	0.19	89.53
65.	0	0	5	3	2	11	0.19	89.71
66.	3	4	2	1	0	11	0.19	89.90
67.	3	3	3	1	0	10	0.17	90.7

fraction of 15.7 % of respondents put some probability mass to an increase in house price and 3 % (2.5 % of men, 4.1 % of women) believe this probability exceeds 50 %.

The demographic characteristics behind the likelihood attributed to an increase are analyzed by reporting linear regression results for the probability of a positive return (Table 2, column 3).<sup>31</sup> The positive effect of having bought the main residence recently stands out. Other noticeable effects are the negative effects of age and having a University degree although these are not precisely estimated.

*Probability of a negative return* The respondent's probability of a negative change in house prices is calculated as the sum of the number of points attributed to intervals 1 and 2 (i.e. to a moderate drop of around 3 % and a large drop of over 6 %). The results

<sup>31</sup> The sum of points is multiplied by 10 to provide results in percentage points.

(Table 2, column 4) show no significant association of such beliefs with household characteristics, except for a not very precise positive effect of household income. Negative house price expectations were therefore widespread across groups of the population at the end of 2011.

*No uncertainty* 32.7 % of respondents believe the price of their homes will drop for sure during 2012 (i.e. they distribute all points between intervals 1 and 2—large drop over 6 %, moderate drop around 3 %). Over half of them (57.2 %) attribute all ten points to one of the two price drop alternatives and hence answer without uncertainty. The results in the fifth column of Table 2 are an attempt to uncover demographic differences associated with these “no uncertainty” answers. The only significant difference between these no-uncertainty respondents and the rest of respondents expecting a drop is gender and owning other housing.<sup>32</sup> According to these results, women are less likely than men to give a 100 % probability to one of the two drop-in-price scenarios (and hence more likely than men to distribute the chances among the two alternatives). Additionally, households owning other housing aside from their main residence are more likely to believe in a drop with no uncertainty about its magnitude.

Analyzing answers without uncertainty in the expected positive domain is not undertaken because it is hampered by the small number of observations.

### 3.5 Fitting subjective house price distributions

*Calculating individual distributions* As seen above, subjects are asked to distribute ten points among five possible changes to the price of their homes over the next year. I use the subject responses to fit a saturated probability distribution for each respondent. This is useful because it facilitates the calculation of comparable measures of position, uncertainty, and quantiles for all individuals. Using a saturated distribution avoids placing restrictions on the form of the distribution relative to the information in the data.

I assume that the probability distributions have a pre-specified support and a pre-specified neighborhood around zero for the no-change category. Having specified end-points and an interval around zero, to get a full cdf I connect the observed points using straight lines so that the cdf is piece-wise linear and the density is flat within segments. This allows calculating all quantiles by linear extrapolation.

Figure 2 illustrates the estimation of the probability distribution for a respondent having distributed his ten points as follows: one point to a drop of more than 6 %, six points to a drop of around 3 %, one point to more or less the same, one point to an increase of around 3 % and one point to an increase larger than 6 %. The limits of the support are defined to be  $-15$  and  $+15$  % and the interval around zero for the non-change category to be between  $-1$  and  $+1$  %. To obtain the  $\tau$ -quantile  $q_{\tau i}$  for some  $\tau \in (z_{li}, z_{(l+1)i})$  we use:

<sup>32</sup> This analysis is conditioned on expecting a drop because I do not wish to mix determinants of certainty with determinants of expecting a rise. Given the macroeconomic scenario, respondents that are certain of a rise are few and probably with special characteristics. As for those putting all points to the “more or less the same” option we have already analyzed their characteristics above.

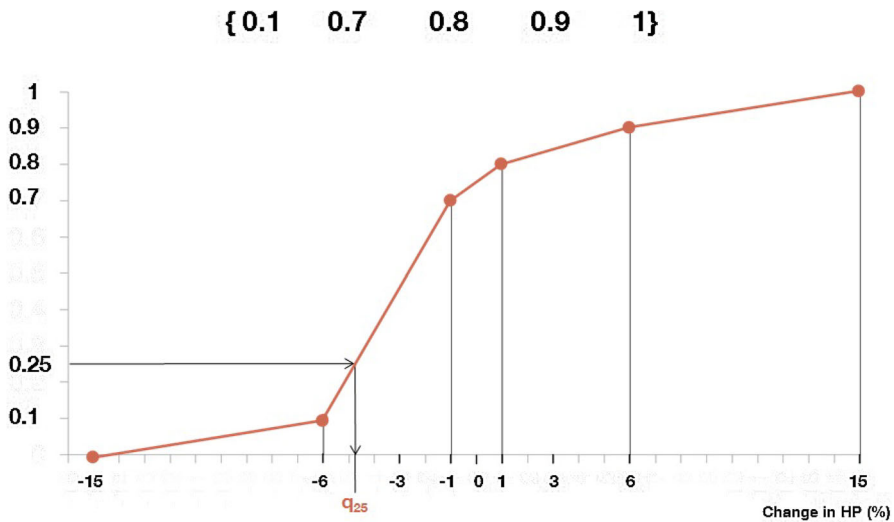


Fig. 2 Fitting individual saturated distributions

$$q_{\tau i} = q_{zli} + [(\tau - z_{li}) / (z_{(l+1)i} - z_{li})](q_{z(l+1)i} - q_{zli})$$

where the  $z_{li}$  are cumulative probabilities and  $q_{zli}$  the corresponding quantiles for  $l = 0, 1, \dots, 5$ , which are given by  $(-15, -6, -1, 1, 6, 15)$ .

*Quantile regressions from subjective quantile variables* Measured quantiles  $q_{\tau i}$  are to be interpreted as conditional quantiles given characteristics of the individual and the house, both observable and unobservable. To look at the variability in these distributions, I estimate least squares regressions of individual quantiles on measured characteristics and postal code dummies (that is within postal code quantile estimates). These quantile regressions are very different from ordinary quantile regressions where one fits a quantile model to data that are sample draws from the distribution. Here the left hand side variable consists of direct measures of the conditional quantiles.

*A factor model for unobserved heterogeneity in subjective quantiles* The quantile regression errors capture unobservable heterogeneity in the subjective probability distributions (except for functional form approximation errors). I estimate a random effects model for the errors of different quantiles to see to what extent a single factor captures the unobserved heterogeneity in the distributions.

Consider for example regressions for  $q_{.25i}, q_{.50i}, q_{.75i}$

$$q_{.25i} = x'_{i.25} \beta_{.25} + u_{.25i}$$

$$q_{.50i} = x'_{i.50} \beta_{.50} + u_{.50i}$$

$$q_{.75i} = x'_{i.75} \beta_{.75} + u_{.75i}$$

The factor model is:

$$u_{\tau i} = \eta_i + \varepsilon_{\tau i} \quad \tau = 0.25, 0.5, 0.75$$

**Table 5** Quantiles of subjective probability distributions of house prices (within postal code estimates)

Variables	(1)	(2)	(3)	(4)	(5)
	q10	q25	q50	q75	q90
Log (price/m2)	0.014	0.131	0.147	0.121	0.116
Age 45–64	0.608*	0.504*	0.379	0.150	−0.087
Age over 64	1.052***	0.728**	0.432	0.100	−0.241
Blue collar	0.676**	0.604**	0.523**	0.406**	0.364**
Self-employed	−0.563	−0.363	−0.327	−0.385	−0.412*
Secondary education	0.221	0.041	0.030	−0.025	−0.082
University education	0.595	0.506	0.382	0.268	0.137
Woman	0.228	0.367*	0.401**	0.323**	0.206
Own other housing	−0.141	−0.225	−0.310	−0.331*	−0.358**
Bought main residence recently	0.609	0.621*	0.578**	0.551**	0.508**
Log (household income)	−0.103	−0.103	−0.125	−0.125	−0.131
Wealth percentiles 25–50	0.455	0.322	0.223	0.091	−0.181
Wealth percentiles 50–75	1.054**	0.765*	0.520	0.328	−0.010
Wealth percentiles 75–90	0.758	0.638	0.555	0.389	0.049
Wealth percentiles 90–100	0.258	0.246	0.212	0.139	−0.102
Constant	−7.230**	−5.811**	−3.405*	−1.279	0.643
Observations	5023	5023	5023	5023	5023
Adjusted R2	0.366	0.345	0.353	0.382	0.400

All specifications include postal dummies and have been estimated taking into account population weights and the five imputed datasets

\*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

I estimate the variance of the common factor  $\eta_i$ , the variances of the random errors  $\varepsilon_{\tau i}$  and the factor loadings  $\delta_\tau$  subject to  $\delta_{0.5} = 1$  and the assumption that  $\eta_i$  and the  $\varepsilon_{\tau i}$  are mutually independent.

### 3.6 Relating heterogeneity in expectations to housing and household characteristics

*Individual density position measures and demographics* I examine the association between quantiles at various points of the estimated individual densities and demographics, within postal codes.<sup>33</sup> In particular I consider the individual median and the 10th, 25th, 75th, and 90th percentiles as distributional measures for each respondent. Multiple regression results for those variables on demographics may be found in Table 5.

The regression equations are of the form:

$$q_{\tau i} = X_i \beta_\tau + Z_i \gamma_\tau + u_{\tau i}$$

<sup>33</sup> There are 1094 postal codes in our data, 212 of them have only one household and 71 have ten or more.

where  $X_i$  is a vector of household characteristics such as age, education, gender, income and wealth. Moreover,  $Z_i$  is a vector of house characteristics, which includes postal code dummies, log (price/square meter) and in some cases also an indicator of age of the house.

The results include the household estimated price (per square meter) of their home. Interestingly, the self-assessed house price of a household is not a significant predictor of the expected evolution of the price of its home conditional on postal code dummies (and the rest of included controls).

We observe lower expected declines in the lower part of the distribution as age increases. This relates to the finding by [Malmendier and Nagel \(2013\)](#) that experience of older individuals draw on longer history of data when forming their expectations while expectations of younger ones are dominated by more recent data. In Spain in 2011 the house price drops experienced since 2007 came after decades of rising house prices.

Blue collar workers are associated with more optimistic expectations all over the distribution while for the self-employed there is a negative shift in the upper part of the distribution. Households in the middle-upper part of the wealth distribution have their expectation distribution shifted upwards (more pronounced in the lower part).

Interestingly, there is a positive effect for those households who bought their main residence recently (in the last 6 years). Moreover, this effect is quite uniform across the whole range of the distribution although more precise in the upper part of the distribution. Recent buyers may be more reluctant to accept a prospect of no house price increases as compared to non-recent buyers who have experienced sizeable house price returns. This effect may also reflect reverse causality, that is, buyers who expected higher house price changes than the rest were the ones who bought recently. [Table 12](#) presents results omitting this variable and the results are unchanged (columns 1 and 2). The same result is obtained if instead a variable reflecting that the house was built in the last 6 years is included. These could be taken as suggesting that the previous result does not seem to be driven by reverse causality.

The results on gender stand up. Being a woman produces a positive shift that is particularly noticeable at the median and at the top quartile. This is difficult to explain in terms of differences in information as one may do with occupation or age. It does not seem to be related to risk aversion either. Indeed, I have also included a measure of risk aversion available in the EFF but the results are unchanged (see [Table 12](#) columns 3 and 4).<sup>34</sup>

What these results say is that there is a difference by gender among the respondents to the survey (controlling for postal code and other covariates), who are meant to be the

<sup>34</sup> We classify as risk lover those individuals answering options 1 or 2 to the following question:

*Which of the following statements do you feel best describes your household in terms of the amount of financial risk you are willing to run when you make an investment?*

Take on a lot of risk in the expectation of obtaining a lot of profit 1

Take on a reasonable amount of risk in the expectation of obtaining an above-normal profit 2

Take on a medium level of risk in the expectation of obtaining an average profit 3

You are not willing to take on financial risk 4

Don't know 98

No answer 99



most knowledgeable about the household finances, as explained earlier. Whether these differences would still hold for randomly selected individuals cannot be answered on the current data.

To check the robustness of the gender result I estimated an [Abadie and Imbens \(2006\)](#) matching estimator of the gender average treatment effect which uses the control variables in a non-parametric way. This produces similar results both in magnitude and significance. The same result is also found estimating the gender average treatment effect by weighting on the propensity score. This is at odds with the generally accepted finding that women tend to be less optimistic than men (see for example [Balasuriya et al. 2010](#)).

To further assess potential differences in asset valuations by gender we regressed self-assessed values of different assets reported in the EFF on the same demographic and socio-economic characteristics. The results in [Table 13](#) show that women tend to provide higher estimates for the value of their home compared to men but lower ones when it comes to value their financial assets.

An open research question in economic psychology is to what extent people's price perceptions and expectations are mediated by psychological variables like emotions and attitudes (see for example [Ranyard et al. 2008](#)). One hypothesis for further research that could explain our results would be that women positive affective feelings for their home (and its value) are stronger than for men and that these preferences affect the judgment of men and women. For a detailed description and evidence see [Slovic et al. \(2001\)](#).

Are women more optimistic, or simply more realistic? A bold answer to this question can be based on the aggregate of counterfactual point predictions of house price changes across all households as if all were male respondents. Using the median as a point forecast measure, the estimation results inform us that the counterfactual female aggregate is 0.4 percentage points higher than the corresponding male aggregate. We can now look at the actual aggregate house price change between 2011 and 2012 to find out which one of the two genders was closer to the truth. The national house price change December 2011–December 2012 for second hand housing was around  $-10\%$ .<sup>35</sup> The counterfactual aggregate male and female point forecasts are  $-3$  and  $-2.6\%$ , respectively. Even if the position of the subjective probability distribution may be affected by framing, the distance between actual and predicted changes is sufficiently large to conclude that women were more optimistic rather than more realistic by comparison with men.

*Uncertainty and demographics* As a first measure of individual forecast uncertainty I consider the inter-quartile range. I also analyze the range given by the difference between the 90th and 10th percentiles. Heterogeneity in self-reported uncertainty is examined in [Table 6](#). A distinct effect on uncertainty in a multiple correlation context is age. Older people express less uncertain expectations. Households in the middle-upper part of the wealth distribution are also less uncertain about their expectations. In line with other authors (see for example [Bruine de Bruin et al. 2011a](#)) I also find

<sup>35</sup> It was  $-14\%$  according to the index from the National Statistics Office (INE) and  $-10\%$  according to the Ministry of Public Works and online search-sites based on asking price data.

**Table 6** Uncertainty in subjective probability distributions of house prices (within postal code estimates)

Variables	(1)	(2)
	q75–q25	q90–q10
Log (price/m2)	–0.010	0.102
Age 45–64	–0.354**	–0.694***
Age over 64	–0.628***	–1.293***
Blue collar	–0.198	–0.312
Self-employed	–0.023	0.151
Secondary education	–0.066	–0.303
University education	–0.237	–0.458
Woman	–0.044	–0.023
Own other housing	–0.107	–0.217
Bought main residence recently	–0.070	–0.101
Log (household income)	–0.022	–0.029
Wealth percentiles 25–50	–0.231	–0.636*
Wealth percentiles 50–75	–0.437**	–1.064***
Wealth percentiles 75–90	–0.249	–0.709*
Wealth percentiles 90–100	–0.107	–0.360
Constant	4.533***	7.873***
Observations	5023	5023
Adjusted R <sup>2</sup>	0.432	0.456

All specifications include postal dummies and have been estimated taking into account population weights and the five imputed datasets  
 \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

that differences in uncertainty across demographic groups are smaller than those in central tendency forecasts.

Are people with more certain expectations more accurate? Since older people have more certain expectations, we can answer the question with relation to age. This is relevant because age is the main observable associated with differences in the degree of certainty in expectations. It turns out that age does not have a significant effect on point-forecasts as measured by the subjective median. Therefore, there is no evidence of differences in predictive accuracy according to the degree of certainty as captured by age.

As another indicator of the potential association between accuracy and certainty I calculated the correlation between the median and the inter-quartile range of the individual subjective distributions. It turns out to be  $-0.4$ . Therefore, more certain individuals tend to predict lower falls in house prices. Given the actual declines described above, such negative correlation would suggest that more certain expectations are less accurate. This result is consistent with recent evidence in psychology that superforecasters are more uncertain about their forecasts (Tetlock and Gardner 2015).

*Robustness to alternative cutting points and to bunching* As explained above, the individual densities required specifying values for various cutting points in the probability density. We analyzed the determinants of robustness of the analysis of beliefs and their uncertainty to alternative values of the cutting points. Table 12 (columns 5 and 6) presents results obtained increasing the minimum and maximum values of the

**Table 7** Importance of the location of the house

	(1)	(2)
	On q50	On q75–q25
% Of explained variation due to postal code dummies	96.6	94.7
% Of postal code variation explained:		
By municipality dummies	63.7	75.4
By province dummies	29.2	29.3

The reference for these calculations are Table 5 (column 3) in the case of the first column and Table 6 (column 1) in the case of column 2

support (from  $\mp 15$  to  $\mp 20$ ). As we can see the results are qualitatively robust to these alternative ways of fitting the distribution. The size of the effects varies depending on the cutting point but relative effects as well as significance are maintained. The conclusions hold for other changes in these values and in the interval chosen around zero.

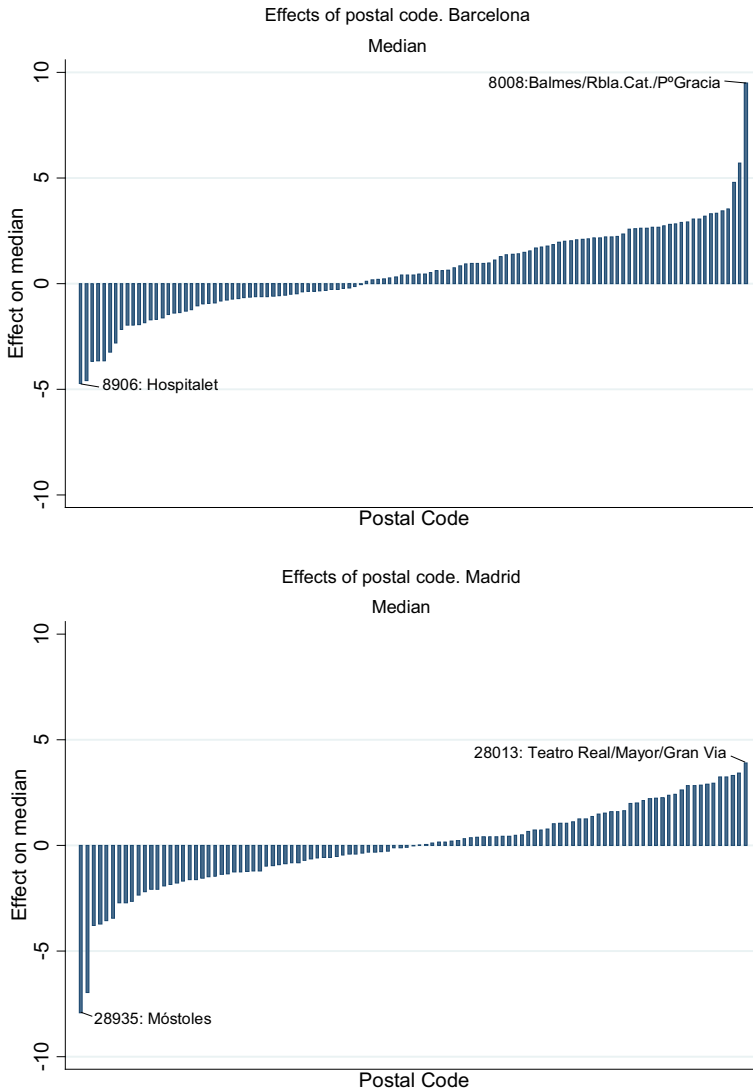
As a further robustness check I estimate the models in Tables 5 and 6 dropping those respondents who put all ten points in the “more or less the same” alternative. The results (not shown) are similar except for the various effects of age that mostly disappear. This is not surprising given the estimates presented in Table 2 column 2 about the factors influencing the probability of assigning all ten points to the middle interval.

*Importance of detailed location of the house* Table 7 highlights the central importance of the detailed location of the house and in particular of introducing postal code information. Location at the postal code level accounts for 97 % of the observed variation in the estimated median expectation and for 95 % of the variation in uncertainty across households (as measured by the inter-quartile range). More aggregate location information like municipality or province do not do such a good job, as one would expect. Municipality dummies account for 66 % of explained variation in the median (and 80 % in the inter-quartile range).

Table 14 (columns 1–3) presents some of the regressions reported in Tables 5 and 6 but without location information. This shows how misleading the estimated effects of other variables could be in the absence of location information. In particular, the gender effect would not be found. Municipality dummies produce results more similar to estimates that control for postal code dummies but still quite different (Table 14 columns 4–6). As expected, it is location at a very disaggregate level that matters for house prices.

*Relating expectations to local housing and labour markets* Inspection of the estimated postal code effects estimated in Tables 5 and 6 indicate that respondents expect the price of their home to grow more in areas where housing prices are already high. Figure 3 plots the estimated postal code effects for Barcelona and Madrid sorted in ascending order. The highest postal code effects in both cities correspond to sought-after areas. The opposite is true at the other end of the scale.

In Tables 8 and 9 estimated postal code fixed effects are regressed on housing and labour market variables, in particular rates of return on housing and unemployment



**Fig. 3** Estimated postal code effects for the two major cities:

rates at the province level. The results show that when forming expectations about the future price of their home respondents extrapolate the recent evolution of the province labour and housing markets. This is true both for the location of the distribution and for the measure of uncertainty. For example, an increase in the unemployment rate in the previous year of 1 percentage point leads to a decrease of 0.18 percentage point in expected median house price and to an increase of 0.1 point in uncertainty as measured by the inter-quartile range.

**Table 8** Postal code dummies on housing and labour market variables

Variables	(1) q10	(2) q25	(3) q50	(4) q75	(5) q90
Rate of return on housing 2010	0.005	0.044	0.078	0.098**	0.102*
Rate of return on housing 2009	0.150**	0.185***	0.174***	0.153***	0.123***
Change in unemployment rate 2010	−0.278***	−0.234***	−0.179***	−0.143**	−0.085
Constant	1.680**	2.095***	2.124***	2.034***	1.784***
Observations	1093	1093	1093	1093	1093
Adjusted R <sup>2</sup>	0.009	0.012	0.014	0.013	0.007

Housing and labour market variables are available at the province level

Postal code dummies estimated in Table 5

\*\*\* p < 0.01, \*\* p < 0.05, \* p < 0.1

**Table 9** Postal code dummies on housing and labour market variables <sup>1,2</sup>

Variables	(1) q75–q25	(2) q90*–q10
Change in rate of return on housing 2010	0.042*	0.029
Change in unemployment rate 2010	0.097**	0.114***
Constant	−0.210*	−3.307***
Observations	1093	1093
Adjusted R <sup>2</sup>	0.006	0.006

Housing and labour market variables are available at the province level

Postal code dummies estimated in Table 6

\*\*\* p < 0.01, \*\* p < 0.05, \* p < 0.1

*Quantile error structure* The first principal component of the (.1, .5, .9) quantile residuals explains 99 % of total variation in a model with postal code dummies, and 98 % with province dummies. When five residuals are used (.1, .25, .5, .75, .9) the variation captured by the first principal component is 91 % with postal code dummies and 89 % with province dummies. Estimation of the random effects model produces an estimated residual variance at the zero boundary (a Heywood case), which is not surprising given the high correlation among residuals. The estimated factor loadings for 0.25 and 0.75 in the three error specification are close to unity (0.94 and 0.95) with corresponding residual variances in the 0.10 range. Relative to those residuals the single common factor explains 97 % of total variation.

## 4 House price expectations and consumption decisions

*Expectations and decisions* One of the main purposes of collecting subjective expectation data is to help understand behavior. In this section I study whether house price expectations reported in the EFF predict household expenditure decisions. This is of

substantial interest in its own right and also a further step in the validation of the information collected.

There were large unexpected shocks to house price expectations in Spain after 2007. The percentage of households buying second housing decreased dramatically since the bursting of the housing bubble. In the three year period between the 2002 and the 2005 EFFs, 5.2 % of households bought a second house (an average of over 1.7 % a year) while this percentage was only 0.6 % for 2011. Also according to EFF data, 9.4 % of the Spanish households bought a car in 2011. However, among the households who are very pessimistic about the future price of their house (i.e. those assigning all ten points to the over 6 % drop scenario) only 4.5 % did so.<sup>36</sup> In this section I use information on expenditure outcomes on various items available in the EFF to see if house price expectations are predictive of purchase and expenditure decisions once a rich set of controls are taken into account.

*Expenditure and purchases in the EFF* In the EFF households provide information on whether they bought a car in the last 12 months and the price paid for those who did. The same information is collected about other big ticket items (furniture, washing machines etc) as a whole. Amounts spent on food at home and outside as well as on other non-durables are also collected.

The EFF provides detailed information on purchases of secondary housing (for households owning their main residence). Housing purchases are both consumption and investment decisions. Bover (2010) provides evidence that aggregate predicted returns on housing have a large positive effect on the hazard of purchasing a house. However, aggregate returns are probably masking different individual expectations concerning future house prices, both in terms of differences in household characteristics and in terms of differences in house specific attributes like location. I therefore explore if individual household expectations about house prices help predict the probability of purchasing a house and, in case of purchase, the amount spent on it.

A word about the timing of subjective expectations and expenditure outcomes. Ideally, the interest is in how expectations held at  $t$  about the future influence decisions at  $t$ . The expectation data correspond to beliefs held at the time of the interview, while the expenditure data refer to purchases during the last 12 months, which is a good timing approximation, specially for durables.

*Empirical model* First, probit estimates are presented for the probability of (1) buying secondary housing, (2) buying a car, and (3) buying other big ticket items (see Table 10). To analyze expenditure I present tobit estimates for the amounts spent on (1) other housing, (2) cars, and (3) other big ticket items, and multivariate regressions for the amount spent on (4) food and other non durables (see Table 11a). As is well-known, tobit estimates rely on the assumption that the same relationship holds both for the decision to purchase and for the amount spent. To check how restrictive this assumption is here the implied tobit estimates for the various purchase probabilities are also be provided (see Table 11b) and compared with the probit estimates.

<sup>36</sup> At the end of 2011 23.3 % of households expected a large decrease (over 6 %) in the price of their home over the next 12 months. Among those, 30.7 % expected this large drop without uncertainty that is 7.2 % of the population of households.

**Table 10** Effects of house price expectations on average probabilities of purchase (at various points of the wealth distribution; probit estimates)

	Other housing		Car purchase		Other big ticket items	
	0.57 (1)	(2)	9.42 (3)	(4)	41.99 (5)	(6)
	Full sample	Within postal codes	Full sample	Within postal codes	Full sample	Within postal codes
	Weights	No weights	Weights	No weights	Weights	No weights
Large certain drop in HP <sup>b</sup>						
At median net wealth	−0.826**	−4.06	−4.46**	−11.7***	2.80	−7.26**
At 80th percentile of net wealth	−1.24***	−4.31	−2.83	−11.3***	2.07	−6.46**
Uncertainty in HP <sup>c</sup>						
At median net wealth	−0.629*	−5.88**	2.67*	0.296	8.09***	0.461
At 80th percentile of net wealth	−0.803**	−8.27***	2.64	−0.334	12.5***	1.59
Income higher than current <sup>d</sup>	−0.145	11.8*	2.14	1.56	0.724	4.95**
Income lower than current	0.656	−4.43	1.77	1.83	−0.007	−1.91
Observations <sup>e</sup>	5019	381	5019	2158	5019	4189

\*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

<sup>a</sup> Control variables include: interactions of expectation dummies with log net wealth, gender, age brackets, number of persons in the household dummies, children dummy, couple dummy, labour status dummies for respondent and partner (if any), positive and negative income shocks dummies and in columns 1, 3, 5 and 7 municipality size dummies instead of postal code dummies

<sup>b</sup> Large certain drop = 1 if respondent assigns all ten points to “large drop (more than 6%)”

<sup>c</sup> Large certain drop = 0 if respondent assigns all ten points to any other of the 4 other options

<sup>d</sup> Uncertainty = 1 if respondent assigns points to more than one option

<sup>e</sup> Uncertainty = 0 if respondent assigns ten points to any of the 4 options different from “large drop (>6%)”

<sup>f</sup> Income higher (lower) than current = 1 if future expected income higher(lower) than current, 0 otherwise

<sup>g</sup> Sample sizes vary slightly across imputations. The numbers shown correspond to the smaller sample size

<sup>h</sup> Probabilities are shown in percentage terms



**Table 11** Effects of house price expectations on (a) expenditures and (b) average probabilities of purchase from tobit model (at various points of the wealth distribution)

Expectation variables <sup>a</sup>	Other Housing (Tobit)		Car purchase (Tobit)		Other big ticket items (Tobit)		Food and other non-durables (log)	
	(1)	(2) Within postal codes <sup>b</sup>	(3)	(4) Within postal codes	(5) Within postal codes	(6) Within postal codes	(7)	(8) Within postal codes
(a)								
Large certain drop in HP <sup>c</sup>								
At median net wealth	−876,501**	−811,159***	−11,580*	−12,975**	260.1	−705.2	0.0319	0.0402
At 80th percentile of net wealth	−676,699**	−483,728***	−4995	−5105	652.0	−667.8	0.0431	0.0566
Uncertainty in HP <sup>d</sup>								
At median net wealth	−229,984**	−507,576***	3761	1871	646.9	69.92	0.0255	0.0102
At 80th percentile of net wealth	−180,981**	−380,539***	4063	2793	1497**	626.2	0.0486*	0.0295
Income higher than current <sup>e</sup>	−47,607	50,864	2694	−843.1	−315.5	222.3	0.0769***	−0.106***
Income lower than current	155,664*	−143,444	2137	2104	−282.4	−669.6	−0.0140	0.00542
Observations <sup>f</sup>	5019	5019	5019	5019	5019	5019	5019	5019
Of which uncensored <sup>g</sup>	40	40	412	412	1959	1959	5019	5019

**Table 11** continued

Expectation variables <sup>a</sup>	Other Housing (Tobit)		Car purchase (Tobit)		Other big ticket items (Tobit)	
	(1)	(2) Within postal codes <sup>b</sup>	(3)	(4) Within postal codes	(5)	(6) Within postal codes
(b)						
Large certain drop in HP <sup>c</sup>						
At median net wealth	−0.771**	−0.753***	−5.64**	−5.95***	1.35	−3.55
At 80th percentile of net wealth	−1.22***	−0.935***	−3.05	−2.87	3.48	−3.45
Uncertainty in HP <sup>d</sup>						
At median net wealth	−0.535*	−0.600***	2.51	1.02	3.41	0.360
At 80th percentile of net wealth	−0.729**	−0.709***	2.96	1.66	8.18***	3.34
Income higher than current <sup>e</sup>	−0.124	0.0841	1.93	−0.477	−1.65	1.15
Income lower than current	0.570	−0.231	1.51	1.21	−1.48	−3.40
Observations <sup>f</sup>	5019	5019	5019	5019	5019	5019
Of which uncensored <sup>g</sup>	40	40	412	412	1959	1959

\*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

<sup>a</sup> Control variables include: interactions of expectation dummies with log net wealth, gender, age brackets, number of persons in the household dummies, children dummy, couple dummy, labour status dummies for respondent and partner (if any), positive and negative income shocks dummies and in columns 1, 3, 5 and 7 (in panel a) municipality dummies instead of postal code dummies

<sup>b</sup> Population weights are used in all columns (both in parameter estimation and in the computation of the effects)

<sup>c</sup> Large certain drop = 1 if respondent assigns all ten points to “large drop (more than 6 %)”

<sup>d</sup> Large certain drop = 0 if respondent assigns all ten points to any other of the four other options

<sup>e</sup> Uncertainty = 1 if respondent assigns points to more than one option

<sup>f</sup> Uncertainty = 0 if respondent assigns ten points to any of the 4 options different from “large drop ( $> 6\%$ )”

<sup>g</sup> Income higher (lower) than current = 1 if future expected income higher(lower) than current, 0 otherwise

<sup>h</sup> Sample sizes vary slightly across imputations. The numbers shown correspond to the smaller sample size

<sup>i</sup> Sample sizes vary slightly across imputations. The numbers shown correspond to the smaller number of uncensored observations

In the empirical models I include two variables measuring household beliefs about future house prices. To reflect the location of those expectations, a 0/1 dummy is defined taking the value 1 for people expecting a large certain drop with certainty (i.e. people who assign all ten points to the option “more than 6 %” drop). To capture uncertainty about the expectation location another 0/1 dummy is defined taking the value 1 for respondents assigning points to more than one option. Note that these indicators are constructed directly from the household responses, and not from fitted individual probability distributions.

A potential concern of reverse causality is that the uncertainty about the future price of the main residence may be reduced by investment in information associated with the purchase of other housing. However, the results in Table 6 indicate a lack of association between uncertainty and having bought the main residence recently, which suggests that endogenous reductions in uncertainty may not be very important.

Importantly, I am able to control for expectations about future household income and hence identify house price expectations net of income expectations. In the EFF2011 expectations about future household income are collected albeit in a qualitative way. Households are asked whether, compared to their current income, they believe their income in the future will be higher, lower, or approximately the same. Two indicators are constructed containing such that information.<sup>37</sup>

Additionally, the occurrence of positive or negative income shocks is controlled for by exploiting the information provided by a question in the EFF on whether current household income is higher than usual, lower than usual or as usual.

Other variables included in the estimated models are: log net household wealth and its interactions with the house price expectations dummies, respondent gender, age (six interval variables), number of persons in the household (six 0/1 dummies), couple dummy, children dummy, labour status dummies for respondent (four categories) and partner (if any).

Regarding location variables two sets of results are presented: controlling for municipality size (seven categories) or by postal code. However, probit estimates controlling for postal codes rely on a significantly reduced number of observations because of the requirement of observing households who buy and households who do not buy at postal code level (columns 2, 4, 6 in Table 10). Those results therefore cannot be taken as representative of the population of households. This is not the case for the tobit estimates in Table 11a, b.

**Results** The results in Table 10 (column 1) show that the most pessimistic households have a significantly lower probability of buying a house than the rest. The reduction in probability is of 0.8 percentage point at the median level of wealth but higher (1.24 pp) at the 80th wealth percentile. Uncertainty about the evolution of future house prices is also associated with reductions in the probability of buying a house. The magnitude of this reduction is 0.63 pp at the median and 0.8 at the 80th percentile. Larger effects at the top of the wealth distribution appear sensible as those are the households most prone to buying second housing.

<sup>37</sup> Starting from the 2014 wave the EFF includes a question on probabilistic expectations about future total household income.

Expecting a large drop in house prices is also associated with a 4.5 pp smaller probability of buying a car at the median level of wealth but not for wealthier households. However, uncertain expectations are positively correlated with the probability of buying a car and, mostly, with other big ticket items. These results could reflect some substitution effects.

Table 11a shows the estimates for the various expenditures. Again, the larger and most significant effects are the reduction in the amounts spent when buying second housing for households expecting a large drop in the price of their house or for those being uncertain about the evolution of the value of their home. For these households the amounts spent when buying a car are also significantly lower (−13,000 € at median wealth). These conclusions hold when postal code dummies are included.

Similarly to the results on purchase probabilities, there seem to be some evidence of some substitution effects for expenditures on other big ticket items and on food and other non-durables among wealthy households uncertain about future house prices. However, these results do not hold when controlling for postal code.

Finally, Table 11b reports the probabilities of purchase of the various items obtained from the tobit model. These are very much in line with those obtained with the probit model (see Table 10). Furthermore, in this case estimates of the probabilities controlling for postal code can be obtained for the whole sample and confirm the results in columns 1, 3, and 5 of Tables 10 and 11b.

## 5 Conclusions

Household predictive probability distributions play an important role in explaining behavior in many contexts. Not only the location of these distributions but also the uncertainty with which households hold their beliefs is relevant. It is therefore important to have at our disposal measures of such probabilistic expectations directly collected from households.

In this paper I have first reviewed recent work about expectations elicited in household surveys. In particular, I have described relevant implementation specificities as well as reviewed various uses of subjective probabilistic expectation questions.

Second, I have analyzed the answers to a question recently introduced in the EFF on probabilistic house price expectations. This analysis shows that asking such type of questions to Spanish households is feasible (as long as respondents are familiar with the subject matter), as shown by the high response rate and the results of a coherency analysis.

The results show significant heterogeneity in house price expectations across respondents. Heterogeneity is found to be significant both for the location of such expectations as well as for the amount of uncertainty around them. I find that women and blue collar workers are more optimistic about the evolution of house prices for 2012, and older respondents are more certain.

The results also provide valuable information about heterogeneity in the housing market. Location of the house at the postal code level is shown to explain most of the observed heterogeneity in expectations. Moreover, past returns to housing and unemployment rates are found to be strong determinants of the estimated effects of location.

Furthermore, the results show that in the absence of controlling for detailed location information about the house the estimated effects of demographic characteristics on house price expectations are biased and misleading.

I also exploit the availability of information about various durable and non-durable expenditures in the EFF and present some novel findings about the association between house price probabilistic expectations (location and uncertainty) and various durable expenditures. The results show that households holding pessimistic expectations have significantly lower probabilities of buying a house and of buying a car. Moreover, the amounts spent on those items by buyers are also smaller than in the absence of such negative expectations. However, I find no association between house price expectations and expenditure on other big ticket items, nor on food and other nondurable expenditure.

Finally, greater uncertainty in house price expectations is associated with a lower probability of buying a secondary house (as well as with smaller amounts spent) but not with the purchase or the amount spent in other goods.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## 6 Appendix

See Tables 12, 13 and 14.

**Table 12** Various robustness checks

Variables	(1)	(2)	(3)		(4)		(5)	(6)
	No recent buyer variable		Risk aversion		Risk aversion		Wider support	
	q50	q75–q25	q50	q75–q25	q50	q75–q25	q50	q75–q25
Log (price/m2)	0.182	–0.014	0.145	–0.012	0.179	–0.045	0.179	–0.045
Age 45–64	0.325	–0.347**	0.326	–0.301*	0.424	–0.491**	0.424	–0.491**
Age over 64	0.378	0.621***	0.321	0.514***	0.493	0.792***	0.493	0.792***
Blue collar	0.530**	–0.199	0.554***	–0.227*	0.639***	–0.327*	0.639***	–0.327*
Self-employed	–0.340	–0.021	–0.362	0.007	–0.379	0.037	–0.379	0.037
Secondary education	0.050	–0.068	0.022	–0.055	0.015	–0.052	0.015	–0.052
University education	0.417	–0.242	0.353	–0.197	0.479	–0.369	0.479	–0.369
Woman	0.381**	–0.041	0.422**	–0.050	0.500**	–0.141	0.500**	–0.141
Own other housing	–0.295	–0.109	–0.273	–0.153	–0.410*	–0.054	–0.410*	–0.054
Bought main residence recently	–	–	0.501*	0.004	0.677**	–0.176	0.677**	–0.176
Risk lover	–	–	–0.189	0.676	–	–	–	–
Log (household income)	–0.113	–0.024	–0.133	–0.008	–0.146	–0.030	–0.146	–0.030
Wealth percentiles 25–50	0.129	–0.220	0.212	–0.205	0.261	–0.364	0.261	–0.364
Wealth percentiles 50–75	0.410	–0.423**	0.522	–0.430**	0.623*	–0.635**	0.623*	–0.635**
Wealth percentiles 75–90	0.430	–0.233	0.574	–0.265	0.720*	–0.457	0.720*	–0.457
Wealth percentiles 90–100	0.090	–0.092	0.252	–0.143	0.310	–0.167	0.310	–0.167
Constant	–3.392*	4.531***	–3.256	4.338***	–3.860*	5.801***	–3.860*	5.801***
Observations	5023	5023	5004	5004	5023	5023	5023	5023
Adjusted R <sup>2</sup>	0.351	0.432	0.354	0.437	0.359	0.407	0.359	0.407

\*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

All specifications include postal dummies and have been estimated taking into account population weights and the five imputed datasets

**Table 13** Self-assessed value of different assets (for those owning each asset)

Variables	(1) Log (home price/m2)	(2) Log (current accounts)	(3) Log (stocks)	(4) Log (pension funds)
Age 45–64	−0.149***	−0.082	0.281	0.565***
Age over 64	−0.211***	0.473***	0.009	0.296
Blue collar	0.008	−0.176**	−0.954***	−0.057
Self-employed	−0.151***	−0.023	0.552	0.164
Secondary education	0.017	−0.011	−0.494	0.043
University education	−0.008	0.588***	−0.393	0.108
Woman	0.054***	−0.266***	−0.611*	−0.256
Log (household income)	0.028*	0.534***	−0.010	0.187
Wealth percentiles 25–50	0.250***	0.669***	−0.030	0.235
Wealth percentiles 50–75	0.457***	1.081***	0.485	0.604*
Wealth percentiles 75–90	0.520***	1.403***	0.285	1.117***
Wealth percentiles 90–100	0.635***	1.461***	1.025*	1.866***
Constant	6.905***	1.776**	9.047***	5.707***
Observations	5190	5717	1631	1858
Adjusted R <sup>2</sup>	0.663	0.447	0.673	0.584

\*\*\* p &lt; 0.01, \*\* p &lt; 0.05, \* p &lt; 0.1

All specifications include postal dummies and have been estimated taking into account population weights and the five imputed datasets

**Table 14** Some quantile measures of subjective probability distributions of house prices: (1) no location information, (2) municipality dummies

Variables	(1) No location information		(3)		(4) Municipality dummies		(5)		(6)	
	q25	q50	q75–q25	q50	q25	q50	q25	q50	q75–q25	q75–q25
Log (price/m2)	–0.237	–0.074	0.256***		–0.142		–0.019		0.148	
Age 45–64	0.079	0.113	–0.061		0.281		0.272		–0.157	
Age over 64	0.398	0.114	–0.621***		0.645**		0.481*		–0.473***	
Blue collar	0.423	0.387*	–0.118		0.464*		0.412*		–0.188	
Self-employed	–0.395	–0.444	–0.171		–0.315		–0.217		0.032	
Secondary education	–0.295	–0.286	–0.009		–0.124		–0.102		–0.028	
University education	0.421	0.345	–0.199		0.454		0.396		–0.177	
Woman	0.294	0.226	–0.161		0.269		0.313*		–0.001	
Own other housing	–0.102	–0.236	–0.174		–0.123		–0.217		–0.116	
Bought main residence recently	0.878**	0.605*	–0.437**		0.753**		0.639**		–0.166	
Log (household income)	–0.071	–0.144	–0.108		–0.121		–0.132		–0.021	
Wealth percentiles 25–50	0.635	0.221	–0.650***		0.652*		0.316		–0.471**	
Wealth percentiles 50–75	1.024**	0.474	–0.833***		1.056***		0.583*		–0.642***	
Wealth percentiles 75–90	0.954**	0.425	–0.764***		0.891**		0.474		–0.531**	
Wealth percentiles 90–100	0.719	0.304	–0.537*		0.546		0.178		–0.396	
Constant	–3.939*	–1.508	4.314***		0.723		0.777		0.736	
Observations	5023	5023	5023		5023		5023		5023	
Adjusted R <sup>2</sup>	0.0116	0.00988	0.0218		0.208		0.209		0.318	

All specifications have been estimated taking into account population weights and the five imputed datasets. Columns 1 to 3 include no location dummies and columns 4 to 6 include municipality dummies

\*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$



## References

- Abadie A, Imbens G (2006) Large sample properties of matching estimators for average treatment effects. *Econometrica* 74:235–267
- Armantier O, Bruine de Bruin W, Potter S, Topa G, van der Klaauw W, Zafar B (2013) Measuring inflation expectations. *Annu Rev Econ* 5(1):273–301
- Arrondel L, Calvo-Pardo H, Tas D (2011) Subjective stock market expectations, information and stock market participation: evidence from France (unpublished manuscript)
- Attanasio O (2009) Expectations and perceptions in developing countries: their measurement and their use. *Am Econ Rev Papers Proc* 99:87–92
- Attanasio O, Augsburg B (2012) Subjective expectations and income process in rural India (unpublished manuscript)
- Balasuriya J, Muradoglu G, Ayton P (2010) Optimism and portfolio choice (unpublished manuscript)
- Benítez-Silva H, Dwyer DS, Gayle W, Muench TJ (2008) Expectations in micro data: rationality revisited. *Empir Econ* 34(2):381–416
- Blanchflower DG, MacCoille C (2009) The formation of inflation expectations: an empirical analysis for the UK, NBER Working Paper No 15388
- Bover O (2010) Housing purchases and the dynamics of housing wealth, CEPR Discussion Paper No 8128, Banco de España Working Paper 1036
- Bruine de Bruin W, van der Klaauw W, Downs J, Fischhoff B, Topa G, Armantier O (2010) Expectations of inflation: the role of demographic variables, expectation formation, and financial literacy. *J Consum Aff* 44(2):381–402
- Bruine de Bruin W, Manski CF, Topa G, Van der Klaauw W (2011a) Measuring consumer uncertainty about future inflation. *J Appl Econom* 26:454–478
- Bruine de Bruin W, van der Klaauw W, Topa G (2011b) Expectations of inflation: the biasing effect of thoughts about specific prices. *J Econ Psychol* 32:834–845
- Bruine de Bruin W, van der Klaauw W, Downs J, Fischhoff B, Topa G, Armantier O (2012) The effect of question wording on reported expectations and perceptions of inflation. *J Econ Psychol* 33:749–757
- Carroll CD (2003) Macroeconomic expectations of households and professional forecasters. *Q J Econ* 118(1):269–298
- Case KE, Shiller RJ, Thompson A (2012) What have they been thinking? Home buyer behavior in hot and cold markets, NBER Working Paper No 18400
- Cavallo A, Cruces G, Pérez-Truglia R (2014) Inflation expectations, learning and supermarket prices: evidence from field experiments, NBER WP 20576
- Das M, Donkers B (1999) How certain are Dutch households about future income? An empirical analysis. *Rev Income Wealth* 45(3):325–338
- Delavande A, Giné X, McKenzie D (2011) Eliciting probabilistic expectations with visual aids in developing countries: how sensitive are answers to variations in elicitation design? *J Appl Econom* 26:479–497
- Delavande A, Rohwedder S (2011) Individuals' uncertainty about future social security benefits and portfolio choice. *J Appl Econom* 26:498–519
- Dominitz J, Manski CF (1997) Using expectations data to study subjective income expectations. *J Am Stat Assoc* 92(439):855–867
- Dominitz J, Manski CF (2004) How should we measure consumer confidence? *J Econ Perspect* 18(2):51–66
- Dominitz J, Manski CF (2007) Expected equity returns and portfolio choice: evidence from the health and retirement study. *J Eur Econ Assoc* 5(2/3):369–379 (**Papers and Proceedings**)
- Fischhoff B, Bruine de Bruin W (1999) Fifty-fifty = 50 %? *J Behav Decis Making* 12:149–163
- Gan L, Gong G, Hurd M, McFadden D (2004) Subjective mortality risk and bequests, NBER Working Paper No. 10789
- Gouret F, Hollard G (2011) When Kahneman meets Manksi: using dual systems of reasoning to interpret subjective expectations of equity returns. *J Appl Econom* 26:371–392
- Hudomiet P, Kézdi G, Willis RJ (2011) Stock market crash and expectations of American households. *J Appl Econom* 26:393–415
- Hurd M (2009) Subjective probabilities in household surveys. *Annu Rev Econ* 1:543–562
- Hurd M, McFadden D, Gan L (1998) Subjective survival curves and life cycle saving behavior. In: Wise D (ed) *Inquiries in the economics of aging*. The University of Chicago Press, Chicago, pp 259–305
- Hurd M, Smith J, Zissimopoulos J (2004) The effects of subjective survival on retirement and social security claiming. *J Appl Econom* 19:761–775

- Hurd MD, van Rooij M, Winter J (2011) Stock market expectations of Dutch households. *J Appl Econom* 26:416–436
- Juster FT (1964) *Anticipations and purchases: an analysis of consumer behavior*. Princeton University Press, Princeton
- Juster FT (1966) Consumer Buying intentions and purchase probability: an experiment in survey design. *J Am Stat Assoc* 61:658–696
- Katona G (1957) Federal Reserve Board Committee Reports on Consumer expectations and savings statistics. *Rev Econ Stat* 39:40–46
- Kézdi G, Willis J (2008) Stock market expectations and portfolio choice of American households (unpublished manuscript)
- Lillard LA, Willis RJ (2001) Cognition and wealth: the importance of probabilistic thinking (unpublished manuscript)
- Lucas RE (1973) Some International evidence on output-inflation tradeoffs. *Am Econ Rev* 63:326–340
- Malmendier U, Nagel S (2013) *Learning from inflation experiences*, Mimeo
- Mankiw NG, Reis R, Wolfers J (2003) Disagreement about inflation expectations, NBER Working Paper No. 9796
- Manski CF (2004) Measuring expectations. *Econometrica* 72(5):1329–1376
- Morgan G, Henrion M (1990) *Uncertainty: a guide to dealing with uncertainty in quantitative risk and policy analysis*. Cambridge University Press, New York
- Niu G, van Soest A (2014) House price expectations, IZA DP No. 8536
- Pesaran H, Weale M (2006) Survey expectations. In: Elliott G, Granger CW, Timmermann A (eds) *Handbook of econometric forecasting*, vol 1, pp 715–776, North Holland
- Phelps ES (1970) Money-wage dynamics and labor-market equilibrium. In: Phelps ES (eds) *Microeconomic*. Norton, Foundations of Employment and Inflation Theory, New York
- Ranyard R, Del Missier F, Bonini N, Duxbury D, Summers B (2008) Perceptions and expectations of price changes and inflation: a review and conceptual work. *J Econ Psychol* 29:378–400
- Slovic P, Finucane M, Peters E, MacGregor D (2001) The affect heuristic. In: Gilovich T, Griffin D, Kahneman D (eds) *Heuristic and biases: the psychology of intuitive judgment*. Cambridge University Press, New York, pp 397–420
- Souleles NS (2004) Expectations, heterogeneous forecast errors and consumption: micro evidence from the Michigan Consumer Sentiment Survey. *J Money Credit Bank* 36:39–72
- Tetlock Ph, Gardner D (2015) *Superforecasting: the art and science of prediction*. Random House Books, London
- Tversky A, Kahneman D (1974) Judgement under uncertainty: heuristics and biases. *Science* 185:1124–113
- van der Klaauw W, Bruine de Bruin W, Topa G, Potter S, Bryan M (2008) Rethinking the measurement of household inflation expectations: preliminary findings. Federal Reserve Bank of New York Staff, Report no. 359
- Winter J (2008) Expectations and attitudes, Expectations and attitudes. In: Borsch-Supan A, Brügiavini A, Jürges H, Kapteyn A, Mackenbach J, Siegrist J, Weber G (eds) *First results from the survey of health, ageing and retirement in Europe 2004–2007*, pp 306–311. Mannheim Research Institute for the Economics of Aging, Mannheim
- Zafar B (2011) Can subjective expectations data be used in choice models? Evidence on cognitive biases. *J Appl Econom* 26:520–544