Betting on the House

Subjective Expectations and Market Choices

Nicolas Bottan
Cornell University

Ricardo Perez-Truglia*
University of California, Los Angeles

Abstract

House price expectations play a central role in macroeconomics and finance. However, there is little *direct* evidence on how these expectations affect market choices. We provide the first experimental evidence based on a large-scale, high-stakes field experiment in the United States. We provided information by mail to 57,910 homeowners who recently listed their homes on the market. Collectively, these homes were worth $34 billion dollars. We randomized the information contained in the mailing to create non-deceptive, exogenous variation in the subjects' home price expectations. We then used rich administrative data to measure the effects of these information shocks on the subjectâs market choices. We find that, consistent with economic theory, higher home price expectations caused the subjects to delay selling their homes. These effects are statistically highly significant, economically large in magnitude, and robust to a number of sharp checks. Our results indicate that market choices are highly elastic to expectations: a 1 percentage point increase in home price expectations reduced the probability of selling within six months by 2.45 percentage points. Moreover, we provide evidence that this behavioral elasticity would be even higher if it were not for the presence of optimization frictions.

JEL Classification: C81, C93, D83, D84, R31.
Keywords: expectations, experiment, housing market, information.

*This Draft: June 17, 2020. First Draft: February 28, 2020. Bottan: Cornell University, Policy Analysis and Management, 2301 MVR Hall Ithaca, NY 14853, nicolas.bottan@cornell.edu. Perez-Truglia: University of California, Los Angeles, Anderson School of Management, Office C515, 110 Westwood Plaza, Los Angeles CA 90403, ricardo.truglia@anderson.ucla.edu. We are thankful for excellent comments from Emi Nakamura, Jon Steinsson, Bradley Nathan and seminar discussants at UC Berkeley. We thank funding from the UCLA's Ziman Center for Real Estate's Rosalinde and Arthur Gilbert Program in Real Estate, Finance and Urban Economics and UCLA Anderson's Behavioral Lab. Zhihao Han provided excellent research assistance. This project was reviewed and approved in advance by the Institutional Review Boards at University of California Los Angeles (#18-001496 and #19-000945) and Cornell University (#1811008440). The experiments were pre-registered in the AEA RCT Registry (#0003663).

1 Introduction

Consumer expectations play a central role in modern macroeconomics and finance, and they are of special interest to policymakers (Bernanke, 2007). Homeowners' expectations about the future growth in home prices, also known as home price expectations, are especially important. Because homes account for a large fraction of households' assets, home price expectations can have major welfare and policy implications. Moreover, home price expectations played a central role in the accounts of the U.S. housing crisis of the late 2000s (Shiller, 2005; Glaeser and Nathanson, 2015; Gennaioli and Shleifer, 2018; Kaplan et al., 2019).

According to economic theory, home price expectations should be a key input for homeownersâ decision-making in the real estate market. In particular, homeowners should be less willing to sell their homes when they have more optimistic home price expectations, compared with more pessimistic home price expectations. Despite their central role, there is little *direct* evidence on whether home price expectations have a causal effect on the decision to sell a property. In this paper, we fill this gap in the literature using a large-scale, high-stakes, pre-registered natural field experiment.

The relationship between home price expectations and market choices is plagued with challenges to causal identification. For example, consider the time-series evidence that average home price expectations co-move with home prices. The direction of causality is difficult to determine. Do home prices increase because of the increase in expectations, or do expectations rise in response to higher home prices? To make matters worse, other potential omitted-variable biases could go in either direction.[1] Without experimental data, quantifying the causal effect of home price expectations on market choices is challenging.

In the ideal experiment, we would take a sample of homeowners who are considering selling their properties. Before they start receiving offers, we would flip a coin to randomize their expectations. For example, if the coin falls heads up, the homeowner would be persuaded that median home prices will appreciate by 1% over the next year. If it falls tails up, the homeowners would be persuaded that median home prices will appreciate by 10% over the next year. Six months after the randomization, we would measure which owners sold their homes. According to economic models, the homeowners who were randomly assigned to the 1% home price expectation should be more likely than their counterparts to sell their properties within that time horizon. Moreover, the magnitude of the difference in behavior between those two conditions indicates how elastic homeowners' behavior is to their expectations, that is, how much less likely the decision to sale is for each percentage point (pp) difference in home price expectations.

[1]For instance, survey data are often subject to substantial measurement error, which can lead to attenuation bias and thus an under-estimation of the causal effect of expectations on behavior.

We designed a field experiment that closely mimics this ideal experiment. We mailed letters to homeowners in the United States who had recently listed their houses on the market. These letters included information on the current level and evolution of median home prices for comparable homes (i.e., homes in the same ZIP Code and with the same number of bedrooms). Some information in the letters was randomized to create non-deceptive, exogenous shocks to the subjectsâ home price expectations. We then used publicly available administrative records to check when homeowners sold their homes. We used these data to measure whether the exogenous shocks to home price expectations induced by our letters affected the subjectsâ subsequent market choices.

The shocks to home price expectations were based on an information-provision experiment. All letters included information about the current median home price. The letters could differ on two features, which were cross-randomized. First, we randomized whether the letter included or did not include additional information on the *evolution* of home prices, hereinafter referred to as the *disclosure-randomization*. Second, we randomized the *source* used for the information on the price evolution, hereinafter referred to as *source-randomization*. We used five sources that have significant effects on home price expectations, according to prior survey experiments (Armona et al., 2019; Fuster et al., 2018): the average price change over the past year, the average price change over the past two years, or one of three forecasts about the price change in the next year (according to three alternative statistical models).[2] To illustrate the exogenous shocks induced by the source-randomization, consider a subject selling a 2-bedroom home in ZIP Code 33308. That homeowner could be randomly allocated to one of the following five signals of home price growth: an annual growth rate of 1.2% over the past one year; an annual growth rate of 3.6% over the past two years; an annual growth forecast of 2.6% (according to statistical model 1); an annual growth forecast of 4.1% (model 2); or an annual growth forecast of 3.5% (model 3). Relative to receiving the first signal (1.2%), receiving the second signal (3.6%) should result in more optimistic home price expectations. Moreover, we can quantify the intensity of the information shock. Relative to receiving the first signal, receiving the second signal should amount to an information shock of 2.4 pp (= 3.6 − 1.2); likewise, relative to the first signal, receiving the third, fourth, or fifth signals should amount to information shocks of 1.4, 2.9, or 2.3 pp, respectively. The exogenous shocks induced by the disclosure-randomization operate in a similar fashion as the source-randomization, except that they exploit heterogeneity in signals within information sources (i.e., across markets) rather than heterogeneity across information sources.

We implemented the field experiment with a sample of individuals who had recently listed

[2] All letters, regardless of the source, were based on real data that homeowners could access from publicly available sources.

their homes for sale. We identified this subject pool using publicly available information from a major online listing website. Using unique identifiers for the property, we merged those records with rich administrative data from county assessors. These public records include detailed information about the property and its owners, such as their full names and mailing addresses. We then used that contact information to mail a letter to the owners of the listed properties. Using the same unique identifiers, we then used public records to track whether and when each property was sold over the next 6 months. In June 2019, we mailed the letters to 57,910 unique homeowners in 36 counties across seven U.S. states. The homes were collectively valued at $34 billion dollars.

The field experiment was designed to measure how the information shocks contained in the letters affect subsequent market choices. To aid in the interpretation of the main findings from the field experiment, we also designed a supplemental survey experiment to measure the effects of the information shocks on home price expectations. This supplemental survey experiment exposed 1,400 additional subjects to the exact same information treatments used in the field experiment. After the information provision experiment we elicited the subjects' home price expectations using standard survey methods. The supplemental survey was part of the same randomized control trial pre-registration as the field experiment, and it was also conducted on the same month as the field experiment.

The data from the supplementary survey experiment confirmed that our information shocks had the expected effects on home price expectations, with the expected sign and with a significant magnitude. A 1 pp information shock increased the one-year-ahead home price expectations by 0.205 pp (p-value=0.001). The degree to which individuals incorporated the information provided in the experiment is comparable in magnitude with the findings from survey experiments on home price expectations (Armona, Fuster, and Zafar, 2019; Fuster, Perez-Truglia, Wiederholt, and Zafar, 2018) and other macroeconomic expectations (Cavallo et al., 2017).

Most important, the results from the field experiment confirm that the information shocks affected actual, high-stakes market choices and in the direction predicted by economic theory. A larger information shock (i.e., making expectations more optimistic) reduced the speed at which the properties were sold. This effect was highly statistically significant and large in magnitude: a 1 pp higher information shock caused a 0.330 pp drop in the probability that the property was sold within 12 weeks (p-value=0.001), implying a behavioral elasticity of -0.33.

The results from the field experiment are robust to a number of sharp checks. We use an event study analysis to exploit variations in the timing of when subjects received and read our letters. First, we estimate the effects on the outcomes right before the letters were delivered. Because the letters had not reached the subjects yet, they should have no effect on the sales outcome. As expected, the effects of the information shocks were precisely estimated around zero in the pre-treatment period. Moreover, we exploit the fact that not all letters were delivered and read at the same time but instead were gradually opened over a period of seven weeks. We show that, as expected, the effects of the letters intensified during that period and stabilized thereafter. Moreover, the event-study analysis shows that the effects of our information shocks were highly persistent. For instance, behavioral elasticity estimated at 28 weeks post-treatment (-0.325, p-value=0.002) was close to the previously reported elasticity estimated at 12 weeks post-treatment (-0.330, p-value=0.001).

For an additional falsification test, we estimated placebo regressions that are identical to the baseline specification, except using dependent variables as pre-treatment characteristics, such as the number of days the property had been listed prior to our experiment or the original listing price. Because those outcomes were determined prior to the letter delivery, it should be impossible for the randomization of the information shocks contained in the letter to affect them. As expected, we find placebo effects that were close to zero, statistically insignificant, and precisely estimated. We provide a number of additional robustness checks. For example, we use binned scatterplots to show that our main effects were linear and not driven by outliers. We show that our results are consistent if, instead of combining the disclosure-randomization and source-randomization variation, we look at those two sources of experimental variation separately.

The reported elasticities (e.g., -0.330 elasticity at 24 weeks post-treatment) constitute an intention-to-treat effect. We provide estimates of the treatment effect on the treated, addressing two main forms of non-compliance. First, conditional on subjectsâ reading the letter, the information shock introduced in the letter may not affect their home price expectations. For example, the subjects may not update their expectationsbecause they have confidence in their prior beliefs. We used the results from the supplemental survey experiment as a measure of the rate of pass-through from information shocks to expectations. Second, some subjects may not have read the letter (e.g., letters may have been lost in the mail, discarded without opening, or read too late, such as after selling the property). Using data from our own mailing campaign and statistics from the U.S. Postal Office, we estimate that around 35.1% of subjects were affected by this form of non-compliance.

After correcting for the two sources of non-compliance, the estimated elasticity between expectations and the probability of selling in the next six months is -2.45 ($=\frac{-0.328}{0.205 \cdot 0.649}$). This result indicates that subjects were highly elastic to their expectations: increasing home price expectations by 1 pp caused a reduction in the sales probability of 2.45 pp. When combined with the high variation in home price expectations across individuals and over time, our experimental estimates suggest that home price expectations are a major driver of market choices. For example, according to our supplemental survey, the standard deviation in home price expectations was 5.39 pp. An increase in expectations of this magnitude (5.39 pp) would reduce the sales probability by 13.21 pp ($= 2.45 \cdot 5.39$).

Finally, we provide evidence that some subjects would like to react more strongly to changes in home price expectations but cannot do so due to optimization frictions. We study the (pre-registered) heterogeneity in the effects of information shocks between properties that were owner-occupied (78.4% of subjects) and non-owner-occupied (21.6% of subjects). We argue that, relative to the subjects who occupied the properties, subjects who do not live on their properties face fewer optimization frictions. To start with, subjects living on their properties need to move out of the property after selling. Moreover, they may need to move out by a deadline due to work or school. Moreover, the subjects moving out of the property may plan to buy another home in the same or nearby neighborhoods, or may have already bought one. As a result, whether they expect home values to appreciate in the neighborhood or not may not be a relevant factor in the decision to sell, because they will be exposed to the same appreciation regardless of whether and when they sell their properties. On the contrary, owners who are not occupying the property own them as investments and thus do not face the same constraints. They do not need to move out of the property or stay in the same neighborhood, and they can do whatever they want with the money after selling the property, such as buying real estate in other neighborhoods or investing in mutual funds.

We find that the effects of information shocks are qualitatively consistent for subjects who are occupants and those who are not occupants, as they have the same sign and are statistically significant. Consistent with the above conjecture, however, we find that the effects are quantitatively quite different: the effects are almost three times as large for subjects who do not live in the home as they are for subjects who do. For example, at 28 weeks after the start of the letter delivery, the effect of information shocks is -0.637 (p-value=0.003) for the non-occupant-owner and -0.225 (p-value=0.066) for occupant-owner, and the difference between these two coefficients is statistically significant (p-value=0.095) and persistent. This large difference suggests that, in the absence of frictions, the elasticity of behavior to expectations would be even larger.[3]

This study relates and contributes to various strands of literature. Most important, this study relates to literature on the role of subjective expectations on the housing market. Some non-experimental studies link survey expectations to decisions such as whether to buy or rent or the size of the mortgage (Bailey, Cao, Kuchler, and Stroebel, 2018; Bailey, Davila, Kuchler,

[3] For evidence on the role of optimization frictions in other contexts, see for example Giglio et al. (2019).

and Stroebel, 2018).[4] To our knowledge, the only experimental work on this topic is based on survey experiments and laboratory games (Armona, Fuster, and Zafar, 2019).

We contribute to this literature by measuring the effects of home price expectations on market behavior with nearly ideal experimental data. Rather than using survey data to measure behavior, which is subject to many criticisms, such as experimenter demand effects, we combined the information experiments with rich behavioral data from administrative records to measure the effects on real, high-stakes market behavior. This is a naturally occurring context, as a large fraction of Americans face decisions about selling a home.[5] This also is a high-stakes context, as the decision accounts for a large fraction of the net worth of the decision maker. Indeed, buying or selling a home is arguably one of the biggest decisions that Americans make, financially and otherwise (Brooks, 2017). Our large-scale experiment involving nearly 60,000 subjects allows us to provide precise estimates and sharp falsification tests, such as providing an event-study analysis of the information provision experiment. Last, in addition to providing a qualitative test of the causal effect of expectations on behavior, our field experiment also allows for a meaningful quantification of the magnitude of the elasticity between expectations and market choices.

Our study also relates to a broad and growing literature using information-provision experiments to study subjective macroeconomic expectations. These survey experiments have been implemented along various dimensions of macroeconomic expectations, including home price expectations (Armona et al., 2019; Fuster et al., 2018), inflation expectations (Armantier et al., 2016; Cavallo et al., 2017; Coibion et al., 2018), and GDP growth (Roth and Wohlfart, 2019). These studies typically provide a random subset of respondents with a piece of information and measure the corresponding effects on their subsequent survey responses, including their posterior beliefs, attitudes, or even small-stakes laboratory choices (Armantier et al., 2015; Armona et al., 2019). We contribute to this literature by assessing the effects of macroeconomic expectations on *actual* behavior in a high-stakes and naturally occurring context.

Additionally, we intend to make a methodological contribution. Our field experiment can be implemented to study not only home price expectations but countless questions from macroeconomics, urban economics, finance, real estate economics, and behavioral economics. Our experimental framework does not have high barriers to entry, as we rely on data sources that are publicly available and easily accessible. The experiment can be implemented in just a few weeks for less than $0.25 per subject, and it is potentially scalable to up to a million

[4] For example, Bailey et al. (2018) presents evidence that individuals are more likely to transition from renting to owning after geographically distant friends experience large recent home price increases.

[5] According to the U.S. Census data for the fourth quarter of 2019, the homeownership rate was 65.1%. According to data from the National Association of Realtors, 5.34 million homes were sold in 2019.

subjects per experiment.

2 Research Design

2.1 Hypothesis

We seek to test a simple prediction from economic models: higher home price expectations should increase the reservation value of the seller and thus, on average, it should take longer for those properties to be sold. This basic prediction from an asset pricing model requires minimal assumptions. For illustration purposes, we provide a simple version of this model in Appendix A. The intuition behind this model is straightforward. The individual holds a real asset (i.e., the home) and can sell it at any time. If the individual expects the price of the home to appreciate, the decision to sell involves a simple trade-off between the expected appreciation and the risk from price volatility. An increase in the expected appreciation makes it optimal to wait for a more attractive offer. Thus, on average, it takes longer for the owner to get rid of the asset.

2.2 Econometric Model

In this section, we discuss the econometric model around which we designed the field experiment.

In the field experiment, we could randomize whether subjects received a signal about the future annual growth rate of home prices. Assume we randomized the provision of a fixed signal that takes the value $\theta\%$. The effect on the subsequent home price expectations and market choices depends on the prior beliefs of the individual. For subjects whose prior home price expectations were below $\theta\%$, we expect the signal to cause them to update their home price expectations upward (and thus take longer to sell their homes). On the contrary, subjects whose prior home price expectations exceeded $\theta\%$ should update their home price expectations downward (and put their homes up for sale homes faster). If the prior expectations were exactly $\theta\%$, then recipients should not update home price expectations (and should not change their market choices).

The challenge in our field experiment is that it is infeasible to measure the prior beliefs of tens of thousands of homeowners. Thus, we used a type of information-provision experiment that does not rely on information about prior beliefs (e.g., see Bottan and Perez-Truglia (2017), Bergolo et al. (2017) and Perez-Truglia and Troiano (2018)). We propose a simple design based on two sources of randomization: disclosure-randomization and source-randomization.

Let Y_i^{post} be the outcome of interest. In the field experiment, Y_i^{post} indicates if the homeowner sold the property by a given date; in the survey data, Y_i^{post} corresponds to the home price expectations. The subscript i indexes individuals, and the superscript *post* refers to an outcome that was measured in the post-treatment period.

Let E_i^j be a signal about the future growth of home prices, where the subscript i notes that this signal pertains to individual i and the superscript $j = 1, ..., J$ corresponds to the information source. For example, $j = 1$ could be the annual price change over the past one year, and $j = 2$ could be the annual price change over the past two years. Each individual is randomly assigned to one information source, denoted by j_i^*. Moreover, we randomized whether the signal produced by that information source ($E_i^{j_i^*}$) is disclosed to the subject. Let D_i be an indicator variable that equals one if the chosen signal is disclosed to the subject and zero otherwise.

The regression of interest is as follows:

$$Y_i^{post} = \nu_0 + \nu_1 \cdot E_i^{j_i^*} \cdot D_i + \nu_2 \cdot D_i + \sum_j \beta_j \cdot E_i^j + \varepsilon_i \tag{1}$$

We used this baseline specification in the analysis. For the sake of brevity, this equation controls linearly for E_i^j. In practice, we controlled for it flexibly by including, in addition to the linear terms, sets of decile dummies.

The main variable of interest is $E_i^{j_i^*} \cdot D_i$, which we call the *information shock*. Its corresponding coefficient, ν_1, measures the effect of the information shocks on the outcome of interest. The orthogonality assumption, required for causal inference, is $E[E_i^{j_i^*} \cdot D_i \cdot \varepsilon_i] = 0$. Given that both $E_i^{j_i^*}$ and D_i are randomly assigned, the orthogonality assumption should be satisfied.

The causal identification exploits two distinct sources of exogenous variation: disclosure-randomization and source-randomization. To illustrate the source-randomization, we focused on the special case in which the information is disclosed to everyone (i.e., $D_i = 1$ for every i) and thus the disclosure-randomization is irrelevant. Equation (1) then becomes the following:

$$Y_i^{post} = \nu_0 + \nu_1 \cdot E_i^{j_i^*} + \sum_j \beta_j \cdot E_i^j + \varepsilon_i \tag{2}$$

Each individual could be randomly assigned to one of many signals, some more optimistic than others. This specification asks, what is the effect of having been assigned to the signal that was randomly chosen, after controlling for each signal that the subject could have been assigned? Without loss of generality, we can normalize $E_i^{j_i^*}$ by differencing it out with respect to the first information source ($j = 1$). Define $\beta_1' = \beta_1 + \nu_1$ and $\beta_j' = \beta_j$ for $j = 2, ..., J$. We then re-express equation (2) as follows:

$$Y_i^{post} = \nu_0 + \nu_1 \cdot [E_i^{j_i^*} - E_i^1] + \sum_j \beta_j' \cdot E_i^j + \varepsilon_i \tag{3}$$

To illustrate the information shock, $[E_i^{j_i^*} - E_i^1]$, consider a subject selling a 2-bedroom home in ZIP Code 33308. That homeowner could be randomly presented with one of the following five signals: an annual growth rate of 1.2% over the past one year ($j = 1$); an annual growth rate of 3.6% over the past two years ($j = 2$); an annual growth forecast of 2.6% according to statistical model 1 ($j = 3$); an annual growth forecast of 4.1% according to model 2 ($j = 4$); or an annual growth forecast of 3.5% according to model 3 ($j = 5$). The variable $[E_i^{j_i^*} - E_i^1]$ equals zero if $j_i^* = 1$ and 2.4 if $j_i^* = 2$. That is, relative to receiving the first signal, receiving the second signal means an information shock of 2.4 pp (= 3.6 − 1.2). The information shock equals 1.4, 2.9, and 2.3 pp when $j_i^* = 3$, $j_i^* = 4$, and $j_i^* = 5$, respectively. Note that we need variation in signals across information sources for identification in this case.

The exogenous shocks induced by disclosure-randomization operate similarly to the source-randomization, except that they exploit heterogeneity in signals within information sources (i.e., across markets) rather than heterogeneity across information sources. To illustrate the disclosure-randomization, consider the case with a single signal (i.e., $J = 1$) and thus the source-randomization is irrelevant. In that case, equation (1) becomes the following:

$$Y_i^{post} = \nu_0 + \nu_1 \cdot E_i^1 \cdot D_i + \nu_2 \cdot D_i + \beta \cdot E_i^1 + \varepsilon_i \tag{4}$$

The parameter β measures the relationship between the signal (E_i^1) and the Y_i^{post} for individuals who are not shown the signal. The idea is that if individuals react to this signal, this relationship becomes even stronger among those who were shown the signal. The parameter ν_1 measures precisely that: how much stronger that relationship is for individuals who were shown the signal ($D_i = 1$), relative to individuals who were not shown the signal ($D_i = 0$). Note that this approach relies on heterogeneity in signals across subjects. If, for example, all subjects lived in the same neighborhood (i.e., the signal took the same value for every i), there would be no variation to identify ν_1.

As stated in the pre-registration, our baseline specification pools the two sources of random variation to maximize statistical power. We also estimated the parameter of interest using the two sources of variation separately to assess whether the results were consistent across the two sources of exogenous variation. To isolate the source-randomization, we re-estimated the baseline model, restricted to all the i's with $D_i = 1$. To isolate the disclosure-randomization, we estimated the baseline specification from equation (1) but included an additional control variable, $E_i^{j_i^*}$, which eliminated the source-randomization variation. Conditional on disclosing information, the effect of being assigned to one source or another was absorbed by $E_i^{j_i^*}$. Thus, the only remaining variation identifying ν_1 is that the signal is shown only to some individuals (i.e., the disclosure-randomization).

2.3 Mailing Design

All subjects in the field experiment were sent a letter by mail. Figure 1 shows a sample of the envelope. We took a number of measures to communicate that the letter, though unsolicited, came from a legitimate source. The top-left corner of the envelope included the logo for the University of California at Los Angeles (UCLA) and a note about the research study. The top-right corner of the envelope included non-profit organization postage.

Appendix C includes a sample letter (for a fictitious subject). Figure 2 shows the boilerplate content, including the front page (Figure 2.a) and back page (Figure 2.b). The boilerplate content comprised the official UCLA logo in the header, contact information in the footer, a physical correspondence address, and a URL of the study's website with additional information (general study information without specifying any hypotheses and contact information for the researchers and institutional review board). A copy of this website was hosted on UCLA's official website (see Appendix D).

Figure 2 shows the placeholders (marked as *«Information»* and *«Information Details»*) for the two pieces of information that differed across treatment groups. The *«Information»* section included a table with information about home prices. Figure 3 shows a sample table for each of the six treatment groups (discussed below). The emph«Information Details» section included methodological notes for the table, such as data sources and statistical models used. Appendix Figure B.1 shows the six corresponding methodological notes.

All letters contained information on the current median home value of similar properties. For example, subjects who listed a 3-bedroom home in ZIP Code 90210 received a letter indicating Zillow's estimated median home value for 3-bedroom homes in ZIP Code 90210. We used the same property types as those used by online real estate market platforms: 1-bedroom, 2-bedroom, 3-bedroom, 4-bedroom, and 5+ bedroom.[6]

In addition to information about the current median home values, the table could include information on to the evolution of median home values, which individuals could use in forming their home price expectations. Homeowners were randomized into one of six treatment groups. These treatment groups differ in whether the table includes additional information

[6]In a small minority of properties, the number of bedrooms was not available and thus we used a separate category ("all homes") that aggregated all previous categories. Also, Zillow does not produce estimates of median home values for some combinations of ZIP codes and bedrooms, in which case we used the estimates for "all homes" instead.

on the price evolution and the source of said information:

- **Baseline:** no additional information on price evolution
- **Past-1:** price change over the past year
- **Past-2:** price change over the past two years
- **Forecast-1:** price change forecast over the next year using statistical model 1
- **Forecast-2:** price change forecast over the next year using statistical model 2
- **Forecast-3:** price change forecast over the next year using statistical model 3

We choose two types of information sources for price changes, recent price changes and statistical forecasts, which have been proven to have significant effects on home expectations in survey experiments. For example, Fuster et al. (2018) show that, upon being shown one of these types of information, subjects update their expectations in the expected direction. Fuster et al. (2018) also show that most households are willing to pay positive amounts for these information sources, which suggests that they find them both useful. Beyond survey experiments, additional theoretical and empirical evidence suggests that these information sources may be relevant when forming expectations. According to the backward-looking expectations model, individuals form beliefs by looking at recent price changes (Case and Shiller, 1989; Shiller, 2005). According to models of rational expectation, households form expectations based on professional forecasts (Carroll, 2003).

The price changes over the past one year (Past-1) and past two years (Past-2) corresponded to the raw Zillow data. The three statistical forecasts (Forecast-1, Forecast-2, and Forecast-3) were based on the same Zillow data. All three forecasts were estimated using year/ZIP Code-level data on the Zillow Home Value Index for 1997â2019. All three models are autoregressive, but they differed in the set of explanatory variables chosen. These differences in specification yielded slightly different forecasts. The first model used five lags of the dependent variable. The second model used five lags of the dependent variable plus five lags of the state-level average of the dependent variable. The third model used three lags of the dependent variable, three lags of the city-level average of the dependent variable, and three lags of the city-level employment rate. Appendix B.3 presents more details about the information sources, including a comparison of the out-of-sample predictive power over recent years. All five information sources were informative to a reasonably similar degree: all had similar predictive power and were comparable to the predictive power of Zillow's official forecasts.

Each panel of Figure 3 corresponds to the hypothetical table that a subject would receive, depending on the assigned treatment group. It shows real examples based on an individual selling a 2-bedroom home in ZIP Code 33308. Panel (a) shows the baseline letter, which includes the current median price level only. The following five panels add information on the price evolution: panel (b) shows an annual growth rate of 1.2% over the past year (Past-1 treatment); panel (c) shows an annual growth rate of 3.6% over the past two years (Past-2 treatment); panel (d) shows the annual growth of rate 2.6% projected by Model 1 (Forecast-1 treatment); panel (e) shows the annual growth rate of 4.1% projected by Model 2 (Forecast-2 treatment); and panel (f) shows the annual growth rate of 3.5% projected by Model 3 (Forecast-3 treatment). As discussed in Section 2, our identification strategy requires heterogeneity in signals across individuals and across information sources, which we show in Section 3 below.

As shown inside the blue box at the bottom of Figure 2.a, the letter includes a URL to an online survey. To verify that the respondents were legitimate subjects and to link survey responses at the individual level, we included a unique 5-letter survey code in the letter. Respondents had to enter the survey code in the first screen of the survey (see Appendix E) before they could answer any questions. The main goal for including the survey link was to provide a proxy for the dates when recipients opened the letters, as in Perez-Truglia and Cruces (2017) and Perez-Truglia and Troiano (2018).

2.4 Design of Supplemental Survey Experiment

The main hypothesis relates to the effect of the information provided in the letters on market behavior. We supplemented that evidence with a survey experiment. Specifically, we wanted to confirm that the information shocks affected home price expectations in the expected direction. Moreover, we wanted to quantify the effects of the information shocks on expectations to aid in the interpretation of the field experiment results.

Ideally, we would use the survey responses. However, based on experiences in previous projects, we anticipated an extremely low response rate and endogenous selection into the survey. For example, Perez-Truglia and Troiano (2018) conducted a mailing intervention in the context of tax compliance that included a link to an online survey, similar to ours. Only 0.2% of the subjects who received the letter responded to the survey. In anticipation of these challenges, we designed and implemented a supplemental survey experiment that randomized the same information included in the field experiment but was designed to be conducted on an auxiliary sample of respondents recruited from an online platform. The full survey instrument was included in the same pre-registration used for the field experiment and was conducted around the same date as the field experiment.

Appendix F includes the full survey instrument. The structure of the online survey was as follows:[7]

- **Step 1 (Elicit Property Details):** To provide randomized information relevant to the respondent, we asked respondents about their current residency, such as their number of bedrooms and 5-digit ZIP code.
- **Step 2 (Elicit Prior Belief):** Respondents were shown the current median home value (in May 2019) for a similar home (same number of bedrooms and ZIP code) and asked to provide the expected median value one year later (in May 2020).
- **Step 3 (Information-Provision Experiment):** All respondents were told that some survey participants will be randomly chosen to receive information about home prices. On the following screen, respondents find out the information selected for them. Respondents were assigned to one of the same six treatment groups from the field experiment described in Section 2.3.
- **Step 4 (Elicit Posterior Belief):** On the following screen, subjects were told that all participants can reassess their guess about future home prices, regardless of what they guessed initially or the information that they received from us. We re-elicited the same question about their expectations of the median price one year later, as well as an additional question on their expectations five years later. As a placebo outcome, respondents were asked about their stock market expectations (respondents were told the closing price of the Dow Jones on May 31st and asked what they expect the price to be one year later).

3 Data Sources and Implementation Details

3.1 Data Sources

To implement the mailing experiment, we combined two sources of data: data on active real estate listings and data for the property tax rolls from the county assessor. We used publicly available data on real estate listings from a major listing website. These data included rich information about the listed properties, such as address, listing price, property characteristics (number of bedrooms, bathrooms, size, days on market), and the assessor's unique parcel number (APN) for the property. We used the APN to match each listing to its corresponding

[7] The survey included some additional questions that could be useful for disentangling mechanisms and for heterogeneity analysisâ (see Appendix B.5 for more details).

record in the county assessor's tax rolls. Tax rolls contain rich information on the properties and owners. Most important for our experiment, the tax rolls include the names of the owners and their mailing addresses (for more details on the data sources, see Appendix B.1).

By law, tax rolls are publicly available for every county in the United States. However, accessibility to the tax rolls varies widely. Some counties post the data online. For example, raw data from many counties in Florida can be easily downloaded at any time using a file transfer protocol (FTP) address. Other counties, such as Alameda County in California, provide this information only in person and on a one-by-one basis. Many others, such as Los Angeles County, require filling out a short form and paying a fee to obtain a Compact Disc with the raw data.

For this field experiment, we selected a set of 36 counties to obtain a large enough subject pool and for which all the required information from the tax rolls (e.g., owner's name and mailing address) was easily accessible. These counties are distributed across seven states and include 30 counties in Florida, Los Angeles County in California, Maricopa County in Arizona, Clark County in Nevada, Cuyahoga County in Ohio, King County in Washington, and Harris County in Texas. In practice, many other U.S. counties likely would be feasible to include in this type of experiment.

3.2 Implementation of Mailing Experiment

On May 28, 2019, we obtained the information on the active real estate listings and the latest available version of the secured tax rolls. Of the 173,708 active listings scraped, around 164,298 included the APN. For these listings, we merged nearly all listings (164,176 out of 164,298) with the county assessor's data. As the number of individuals in this initial sample was substantially higher than the number of subjects needed for our experiment, we adopted a conservative approach and excluded individuals who were not ideal for the experiment. For example, we excluded non-residential properties and residential properties owned by businesses, because it was unclear whether our letter would be delivered to the person making selling decisions (e.g., the mailing address may correspond to the firm's lawyer). Similarly, we excluded individuals who recently moved, according to the latest mail-forwarding data from the U.S. Postal Services, and individuals who owned multiple properties in the same county (for more implementation details, see Appendix B.2).

After applying these filters, our pool of potential subjects consisted of 61,176 individuals. From those, we selected a random sample of 60,000 individuals to send a letter. After processing the data through the U.S. Postal Service, we excluded a minority (3.4%) of subjects whose mailing addresses were flagged as undeliverable or vacant (1,193) or whose updated tax rolls indicated that we sent the letter to a former rather than current owner (845).[8] The final subject pool comprised 57,910 individuals to receive letters. These individuals were randomly assigned to the following treatments: 20% to Baseline, 15% each to Past-1 and Past-2, and 16.6% each to Forecast-1, Forecast-2, and Forecast-3. All letters were mailed on June 10, 2019.

3.3 Descriptive Statistics and Randomization Balance

Table 1 shows some basic descriptive statistics about the sample. Column (1) presents the average characteristics for the whole subject pool of 57,910 recipients. The average property was on the market for 86 days, was listed for $575,000, had three bedrooms, 2.6 bathrooms, 2,300 sq. ft. of living space, and a lot size of 12,000 sq. ft. Additionally, columns (2) through (7) of Table 1 break down the average characteristics by each of the six treatment groups. The last column reports p-values for the null hypothesis that the average characteristics were equal across all six treatment groups. The results indicate successful random assignment and that the observable characteristics were balanced across treatment groups.

Appendix B.4 presents some descriptive statistics about the subjects based on an independent source of proprietary data: the average age was 58 years old, 31.9% were female, 68.8% were white, and their average annual household income of $128,000. We also show that our sample is representative of the universe of homeowners in the United States in terms of the observable characteristics of the owners and their properties. The main exception is home values, which were on average twice as large for the subject pool as for the country as a whole. We show that this difference arises mechanically, because the subject pool includes several counties with high property values (e.g., Los Angeles County).

3.4 Variation in Signals

As explained in Section 2, the identification strategy relied on variations in signals within and across information sources. In this section, we show that there was plenty of such variation.

Figure 4 presents the results. Figure 4.a shows the variation within information sources, which is relevant for the disclosure-randomization. This figure shows a histogram of the signal that subjects would have received had they been assigned to the Past-1 treatment. The results show plenty of variation. Subjects in the 10th percentile lived in areas where median home values declined by -0.7% in the previous 12 months, and subjects in the 90th percentile lived in areas where property values increased by 8.6%. The degree of heterogeneity is comparable to the other four sources (e.g., the standard deviation in signals across individuals is 3.8 pp

[8] The tax rolls were updated with a lag, thus 845 letters were sent to previous owners.

for Past-1, and the corresponding figure ranges from 1.2 to 4.0 for the other sources.) For more details, see Appendix B.6.

Figure 4.b presents the heterogeneity in signals across information sources, which is most relevant for the source-randomization. This scatterplot shows the relationship between the signals that the subjects would have received if they had been assigned to the Past-1 treatment (i.e., annual growth rate over the past one year) versus the Past-2 treatment (i.e., annual growth rate over the past two years). For example, this figure highlights a specific example: for 2-bedroom homes in ZIP Code 33308, the recipient would have been shown a price change of 1.2% if randomly assigned to the Past-1 treatment group and a price change of 3.5% if randomly assigned to the Past-2 treatment group. The two signals are highly correlated: on average, an extra 1% increase in the annual price change over the past one year is associated with an extra 0.659% increase in the annual price change over the past two years. This relationship is partly mechanical (the y-axis is an average that includes the x-axis) and partly due to the well-known momentum in home prices. In any case, the most important fact is that the relationship between these two potential signals is far from perfect: the $R^2 = 0.659$ is high but substantially below one. Moreover, in addition to the variation between the Past-1 and Past-2 treatments, significant variation occurs across other pairs of information sources (see Appendix B.6 for more details).

3.5 Letter Delivery

The letters were mailed on June 10, 2019. To make the experiment more affordable, we used non-profit postage. According to the U.S. Monitor Non-Profit Standard Mail Delivery Study, it takes non-profit mailings about 10 days to be delivered, with some letters arriving as much as a month after mailing (U.S. Monitor, 2014).[9] As such, some subjects received and read the letter a few days after that date, whereas others took weeks to receive and read the letter. Even after delivery, it may take days or even weeks for the typical subject to open and read the letter. Some subjects were travelling while the letter was delivered to their homes; some subjects received the letter right away but put it away and did not open it until weeks later; etcetera.

Following Perez-Truglia and Cruces (2017) and Perez-Truglia and Troiano (2018), we used the distribution of dates when the surveys included in the letter were completed as a proxy for when the letters were actually read; hereafter, we refer to these dates as the "read-receipt."[10] Figure 5.a presents the results. The first survey response was received on

[9]This delivery time is more than twice that of first-class mail, which is handled first, followed by presort standard and finally non-profit mail.

[10]Our proxy probably had some upward bias, because some people may have read the letter and waited a

June 15, thus marking the start of letter delivery. Indeed, this date coincided with the best guess provided by the mailing company and was based on the location of the shipping facility (Lombard, Illinois) and the location of the letter recipients. This figure suggests that the letters were opened gradually from the start of the letter delivery until eight weeks later. The median time from the start of the letter delivery until the read-receipt was approximately three weeks.

3.6 Outcome Variable: Home Sales

To measure the behavioral outcomes, we scraped the administrative data for the real estate listing website on a weekly basis, from two weeks before the start of letter delivery until 28 weeks after the start of letter delivery. We continue to scrape the data on a weekly basis, so in a future version we may be able to look at even longer time horizons.

Administrative records indicate whether the property was sold and on what date. Confirmation of property sale came from either the Multiple Listing Service (MLS) or the county assessor records.[11] This information is the source for our main outcome variable, the probability that a property is sold by a given date, as listed in the American Economic Association randomized controlled trial pre-registry.

Figure 5.b shows the evolution of the sales outcome for the subject pool. Note that the fraction of homes sold increases smoothly over time. By 12 weeks after the start of letter delivery, 38.8% of homes had been sold. By 20 weeks after the start of letter delivery, 50.6% of homes had been sold. By 28 weeks, the end of our panel data, 57.5% of the properties had been sold.

3.7 Implementation of the Supplemental Online Survey

For the supplemental online survey, we recruited subjects from Amazon's Mechanical Turk (AMT) online marketplace. We followed several best practices for recruiting participants in online surveys and experiments via AMT to ensure high-quality responses (Crump et al., 2013). We restricted participants to those who resided in the United States. We offered a $0.75 participation reward for a 5-minute survey and collected responses from June 21 to June 24.[12] The final AMT sample included 1,404 respondents, who were assigned to treatment groups using the same probabilities as in the main experiment (20% to the Baseline,

few days to respond to the survey. Another potential source of bias, which may be upwards or downwards, is that survey respondents could open the letters more or less slowly than survey non-respondents.

[11]If we had confirmation from both sources, we used the earliest date for which we had confirmation. Our records usually included both sources, and the two dates were normally just a few days apart.

[12]The survey included a short follow-up module. For details, see Appendix B.5.

15% each to Past-1 and Past-2, and 16.6% each to Forecast-1, Forecast-2, and Forecast-3). Appendix B.5 presents more details about the AMT sample. For example, we show that relative to the field experiment sample and the universe of U.S. homeowners, the online sample is younger, less wealthy, and lives in smaller and cheaper homes. Consistent with successful random assignment, the observable characteristics were balanced across treatment groups.

4 Main Results

4.1 Effects on Survey Expectations

Our main interest was to study the effects of information on market choices. We first analyzed the effects of information on home price expectations. This exercise confirmed that the information shocks affected expectations in the expected direction and would be useful to interpret the magnitude of the results from the field experiment.

Table 2 presents the main regression results. All regressions shown in this table use the exact same econometric specification corresponding to equation (1) from Section 2, with the key independent variable being the *Information Shock*: $E_i^{j_i^*} \cdot T_i$. The only difference between the different columns is that they can be based on different databases (the supplemental survey in columns (1)â(4) and the field experiment in columns (5)â(8)) and that each column uses a different dependent variable.

Columns (1) through (4) of Table 2 were estimated using data from the supplemental online survey. In column (1), the dependent variable is the posterior belief about the local home price one year ahead (H_{1y}^{post}). In other words, the coefficient on *Information Shock* measures how a 1 pp increase in the information shock affects the subsequent one-year-ahead expectation of the respondents. The coefficient on *Information Shock* from column (1) is positive (0.205) and statistically significant (p-value=0.001). A 1 pp increase in the information shock causes an increase in home price expectations of 0.205 pp. In other words, there is a 20.5% "pass-through" from the information shocks to the expectations. The fact that the coefficient on *Information Shock* is significantly greater than zero implies that the subjects found the information provided in the experiment relevant to form their home price expectations. The fact that the coefficient on *Information Shock* is significantly less than one suggests that the information provided to the subjects was not the only information they considered. Some subjects may have ignored the information given to them because they already knew it, because they did not trust the information source, or because they did not pay enough attention to the survey.

The baseline specification from equation (1) assumes that the relationship between the outcome of interest and the information shock was linear and symmetric around zero. Indeed, a simple Bayesian learning model predicts this type of relationship when the outcome is the posterior belief (i.e., column (1) of Table 2). This linear relationship has been found to fit the data almost perfectly in a variety of information-provision experiments, including home price expectations (Armantier et al., 2016; Cavallo et al., 2017; Bottan and Perez-Truglia, 2017; Fuster et al., 2018; Cullen and Perez-Truglia, 2018). We used binned scatterplots to assess whether the implicit functional form assumptions provided a good fit for the data. Figure 6.a shows the binned scatterplot version of the results from column (1) of Table 2. The results indicate that the linear specification was a reasonable approximation. This figure also suggests that the results were not driven by outliers.[13]

Columns (2) through (4) of Table 2 present the effects on other survey outcomes. The dependent variable in column (2) is identical to the dependent variable in column (1), except that it corresponds to the expectation for five-years-ahead instead of one-year-ahead. The coefficient on *Information Shock* is positive (0.167) and statistically significant (p-value=0.017). This result implies that when a subject receives an information shock, it propagates to the more immediate expectation (one-year-ahead) and to the longer-term expectations (five-years-ahead). The point estimate from column (1) is smaller in magnitude than the point estimate from column (2), suggesting that the information shocks affect more immediate expectations more strongly; however, that comparison must be considered in context, as the coefficients from column (1) and (2) are statistically indistinguishable from each other (p-value=0.560).

Columns (3) and (4) of Table 2 present some falsification tests. The dependent variables from columns (1) and (3) are identical, except that the outcome variable from column (1) is a posterior belief (i.e., elicited after the information-provision experiment) and the outcome variable from column (3) is a prior belief (i.e., the home price expectations elicited before the information-provision experiment). The information shock has not been administered to the subject yet, so it should have no effect on prior beliefs. As expected, the coefficient on *Information Shock* from column (3) is close to zero (-0.014), statistically insignificant (p-value=0.837), and precisely estimated. Moreover, we can reject the null hypothesis that the baseline coefficient from column (1) is equal to the falsification coefficient from column (3), with a p-value<0.001.[14]

[13] In Appendix B.7, we provide further results on how subjects update beliefs based on the information provided to them.

[14] This equality test between two coefficients is based on the same data but different regressions. To allow for a non-zero covariance between these two coefficients, we estimate a system of seemingly unrelated regressions. In the remainder of the paper, when comparing coefficients from the same data but different regressions, we always use this method.

Column (4) of Table 2 presents the other falsification test. The dependent variables in columns (1) and (4) are both posterior beliefs (i.e., elicited after the information-provision experiment). However, whereas the dependent variable from column (1) corresponds to the home price expectations, the dependent variable from column (4) corresponds to stock market expectations. Because the information shock is specific to local home prices, we do not expect individuals to extrapolate this information to the stock market expectations, and if they do, it suggests spurious motives behind the belief updating, such as numerical anchoring or experimenter-demand Cavallo et al. (2017). As expected, the coefficient on *Information Shock* from column (4) is close to zero (0.017), statistically insignificant (p-value=0.899), and precisely estimated.

4.2 Effects on Market Choices

Next, we turn to the effects of information shocks on market behavior. The main regression results are presented in columns (5) through (8) of Table 2, which are estimated with the data from the field experiment. Our main outcome of interest is whether the property was sold at a given post-treatment date. Note that our information shock cannot have an instantaneous effect on the sales outcome. Most likely, a few weeks should go by from when the letter is read until the information contained in the letter could influence the sales outcome. Typically, after reading the letter the seller will have to wait a couple of weeks to get some offers on the house. And even after the seller accepts an offer, a couple of additional weeks must go by until the sale shows up in the administrative records, due to the standard real estate closing process.

In column (5) of Table 2, the dependent variable equals 100 if the property was sold by 12 weeks after the start of letter delivery and zero otherwise. This is just a starting point. Below, we report estimates for all the possible time horizons. According to the read-receipt proxy from Section 3.5, virtually all subjects read our letter within eight weeks after the start of the letter delivery. As a result, when looking at the sales outcome at 12 weeks after the start of the letter delivery, most subjects had been "exposed" to the information for 4â11 weeks, allowing plenty of time for the information to affect sales outcomes. Moreover, around 37% of the properties were sold within this time horizon, allowing plenty of variation in this outcome to identify the effects of our information.

The coefficient on *Information Shock* from column (5) of Table 2 is negative (-0.330) and highly statistically significant (p-value=0.001). This negative sign is consistent with the prediction from economic theory (i.e., a positive shock to expectations should decrease the probability that the property is sold). This coefficient is also economically large. A 1 pp larger information shock causes a 0.330 pp drop in the probability that the property is sold within 12 weeks. This estimate implies a behavioral elasticity between the sales probability and the information shock of 0.33. Note that this elasticity corresponds to an intention-to-treat effect, because the information shocks do not fully materialize based on changes in expectations. In Section 4.4, we provide estimates of a more relevant elasticity between home price expectations and market choices.

As previously discussed, the baseline specification implicitly assumes a relationship between information shocks and sales outcomes that is linear and symmetric around zero. In practice, this may not be a good approximation: for example, individuals may find it easier to react to good news than bad news, or they may care about their expectations only when they are too optimistic or too pessimistic. To explore these possibilities, Figure 6.b presents the binned scatterplot version of the results from column (5) of Table 2. The results indicate that the baseline specification fits the data perfectly. Moreover, this binned scatterplot shows that outliers do not drive the results.

Column (6) of Table 2 is identical to column (5), except that the dependent variable indicates if the property was sold within 28 weeks, instead of within 12 weeks, after the start of letter delivery (i.e., the longest horizon we can assess with the currently available data). These results indicate if the effects of the information shocks were short-lived (e.g., some owners waited an additional few weeks) or persistent. The coefficient on *Information Shock* from column (6) is negative (-0.325) and statistically highly significant (p-value=0.002). Indeed, the coefficient for 28 weeks later (-0.325, from column (6)) is almost identical and statistically indistinguishable from the corresponding coefficient for 12 weeks later (-0.330, from column (5)). These results indicate that the effects of the information shocks were highly persistent and remained as strong at six months after the start of letter delivery as they were at three months after the start of the letter delivery.

Columns (7) and (8) of Table 2 present some falsification tests. Column (7) is identical to column (1), except that the horizon is one week pre-treatment instead of 12 weeks post-treatment. In other words, the dependent variable in column (7) equals 100 if the property was sold right before the start of letter delivery and zero otherwise. As the information shocks had not been administered to the subjects at that point, the information shocks should have no effect on those market choices. As expected, the coefficient on *Information Shock* from column (7) is close to zero (0.014), statistically insignificant (p-value=0.469), and precisely estimated. Indeed, we can reject the null hypothesis that the falsification coefficient (0.014, from column (4)) is equal to the baseline coefficient (-0.330, from column (1)), with a p-value<0.001.

We further assess whether the timing of effects relates to the timing of delivery of letters, as expected. To do so, we present an event-study analysis of the effects of our information intervention. The top half of Figure 7 presents the results. This figure reproduces the results from columns (5) and (6) of Table 2, corresponding to the horizons at 12 weeks after and 28 weeks after the start of letter delivery, as well as the corresponding results for every other time horizon, from two weeks before the start of letter delivery (when the administrative data was first downloaded to create the letters) until 28 weeks after the start of letter delivery (the last date on which we collected the administrative data). To facilitate the comparison of timing of read-receipts and the effects of the information, the bottom half of Figure 7 shows the evolution of read-receipts (i.e., a reproduction of Figure 5.a).

The evidence indicates that the timing of the experimental effects is largely consistent with the timing of letter delivery. First, the information should not have any effects prior to the start of the letter delivery. Indeed, Figure 7 shows effects of information that are close to zero and statistically insignificant for each of the two weeks prior to the start of letter delivery. We do not expect the information shocks to materialize immediately after the start of letter delivery for two reasons. First, the effects should build up over time as letters are opened. Second, even after all letters were read (around week 8, according to our read-receipt proxy), a few additional weeks must go by for the information to potentially affect sales outcomes. In other words, sellers must receive offers over the following weeks, and if they accept one, it takes another couple of weeks for the closing process. The timing of the effects of information shown in Figure 7 are as expected. The effects of the information started to build up at two weeks after the start of letter delivery and intensify during the period in which the letters were gradually opened, finally stabilizing a month after all letters were read.

Column (8) of Table 2 presents an additional falsification test. In Section 3.3, we show that the pre-treatment characteristics were balanced across the six treatment groups. However, given that our econometric model focused on treatment heterogeneity, this test was not the most relevant balance test. A more direct falsification test would consist of reproducing the same regression as in the baseline specification (column (5) of Table 2), using the pre-treatment characteristics as dependent variables. An example of this type of falsification test is provided in column (8) of Table 2. In that specification, the dependent variable is the (log) number of days that the property was listed prior to the start of our experiment. Because this outcome was determined before the letters were mailed, the information shocks should have no effect on it. As expected, the coefficient on *Information Shock* from column (8) is close to zero (0.001), statistically insignificant (p-value=0.755), and precisely estimated. We cannot compare the coefficient on *Information Shock* from column (8) to the corresponding coefficient from column (5), because the two dependent variables have different scales. To make that quantitative comparison possible, we normalized the coefficients. We constructed a standardized coefficient by multiplying it by 100 and then dividing it by the standard deviation of the corresponding dependent variable (reported at the bottom of Table 2). The standardized coefficients are -0.683 (p-value=0.001) on the sales probability at 12 weeks post-treatment and 0.068 (p-value=0.755) on the pre-treatment number of days listed. The difference between the two is highly statistically significant (p-value=0.023).

Figure 8 extends this falsification analysis to other pre-treatment characteristics. All coefficients in this figure are standardized as described above to make them comparable to each other. The two leftmost coefficients correspond to the two post-treatment outcomes shown in columns (5) and (6) of Table 2: whether the property was sold at 12 weeks after or 28 weeks after letter delivery. The six estimates to the right are based on six different pre-treatment characteristics: the (log) number of days the property was listed (as in column (8) of Table 2), the (log) initial listing price, the number of bedrooms, the number of bathrooms, the square footage of the building, and lot size square footage. Consistent with the results from column (8) of Table 2), the effects on the pre-treatment outcomes are close to zero, statistically insignificant, precisely estimated, and statistically different from the corresponding effects on the post-treatment outcomes.

4.3 Additional Robustness Checks

Table 3 presents some additional robustness checks. Column (1) is the baseline specification, identical to column (5) from Table 2. The specification from column (2) is identical to that of column (1), except that it includes additional control variables: the (log) number of the days the property was on the market prior to the experiment, the (log) initial listing price, a set of four indicator variables for the number of bedrooms, four indicator variables for the number of bathrooms, the (log) square footage built, the (log) lot size, and six indicator variables for the state where the property is located. Note that the R^2 increases substantially, from 0.034 in column (1) to 0.111 in column (2), meaning that the control variables have substantial explanatory power. Because the treatment is randomized, controlling for additional variables should not make a significant difference for the coefficient on *Information Shock*. As expected, the point estimate (-0.325, from column (2)) is almost identical to the baseline coefficient (-0.330, from column (1)), and the difference is statistically insignificant (p-value=0.990).

In the baseline specification, one of the control variables is an indicator variable that equals one if the information was disclosed to the subject. Because the letters disclosed information from different sources, we controlled for each type of disclosure separately. The specification from column (3) of Table 3 is identical to that of column (1), except that instead of controlling for one treatment indicator, it controls for a set of five treatment indicators (i.e., one indicator for each treatment, with Baseline being the omitted category). The results are almost identical under this alternative specification. The coefficient from this extended specification (-0.325, from column (3)) is almost identical to the baseline coefficient (-0.330, from column (1)), and the difference is statistically insignificant (p-value=0.999). Indeed, controlling for additional disclosure indicators does not add any explanatory power: the R^2 is identical between columns (1) and (3).

As discussed in Section 2, the baseline specification combines the two sources of exogenous variation: disclosure-randomization and source-randomization. Columns (4) and (5) of Table 3 provide estimates that rely on the two sources of variation separately to assess whether the results are similar across both sources of identification. Column (4) provides an estimate that exploits only the disclosure-randomization: as discussed in Section 2, this specification is identical to that of column (1), except for an additional control for the value of the signal chosen for the subject and without the interaction with the disclosure indicator. Column (5) provides an estimate that exploits only the source-randomization by dropping the observations corresponding to the baseline group. The results from Table 3 indicate that the results are robust across the two identification strategies. The coefficients are similar in the baseline specification (-0.330, from column (1)), in the specification that only uses the disclosure-randomization (-0.286, from column (4)), and in the specification that only uses the source-randomization (-0.330, from column (5)). These three coefficients are statistically significant on their own (p-values of 0.001, 0.049, and 0.005) and statistically indistinguishable from each other.

In column (5) of Table 3, we show the results when excluding the Baseline treatment group. In columns (6) through (10), we explore whether the results are sensitive to dropping one of the other five treatment groups. Relative to the baseline specification, the resulting coefficients are a bit less precisely estimated, because dropping one treatment group means throwing away between 15% and 20% of the subjects. However, most important, the results are not driven by any individual treatment group. The six coefficients (-0.330, -0.338, -0.257, -0.325, -0.382, and -0.320 from columns (5) through (10), respectively) are consistent in magnitude to the corresponding baseline coefficient (-0.330, from column (1)) and robust in terms of statistical significance.[15]

Due to space constraints, the results for the secondary outcomes are presented in the Appendix. In the previous analysis, the outcome variable is the probability that property is sold at a given point in time. An alternative outcome variable would be the number of days

[15] A related question is whether the information about the past (Past-1 and Past-2 treatments) was more or less compelling than the information about the forecasts (Forecast-1, Forecast-2, and Forecast-3 treatments). For example, if most subjects have backward-looking expectations, they may be more elastic to information about the past than to the forecasts (Case and Shiller, 1989; Shiller, 2005). In Appendix B.9, we provide some suggestive evidence that the information about the past was more effective than the forecasts.

elapsed from the start of letter delivery to the sale of the property. As we anticipated in the pre-registration, a problem with this outcome is that it is truncated: the 42% of properties in the sample that were not sold by the end of the sample window could have been sold one week or 1,000 weeks later. By looking at the probability of selling at a given point in time, we avoided the biases caused by truncation. In any case, as reported in Appendix B.8, the results are qualitatively and quantitatively robust if we estimate the effects using duration models.

The data that we scraped to construct the main outcome of interest contained other information that can be used to construct secondary outcomes. Appendix A presents the effects on these secondary outcomes. One of those outcomes, which is listed in the pre-registration, is sale price. As anticipated in the pre-registration, however, the challenge with this outcome is that it is censored: we did not observe the sale price for 42% of the sample, because those properties were not sold. In Appendix B.10, we show that the evidence is inconclusive, because the results were sensitive to different approaches used to deal with the bias caused by censoring this dependent variable.[16] In addition to sale price, our scraped data included other weekly outcomes related to the listings, such as whether the property remained listed on the website or the listing price changed. Appendix B.11 presents the results for these outcomes.

4.4 Elasticity of Behavior to Expectations

In this section, we discuss the magnitude of our findings. The main object of interest is how elastic the sellers' decisions to sell were in relation to their home price expectations. A good starting point is how elastic sellers' decisions were to the information shocks. The results presented in column (6) of Table 2) suggest that a 1 pp higher information shock decreased the probability of selling within 28 weeks after the start of letter delivery by 0.325 pp. This estimate can be interpreted as a behavioral elasticity of -0.325. However, the challenge is that this elasticity between behavior and information shocks is an intention-to-treat estimate, because part of the information shocks do not translate into changes in home price expectations. This imperfect compliance implies that the elasticity of -0.325 is just a lower bound. Thus, we propose an approach to scale up the intention-to-treat elasticity into treatment-on-the-treated elasticity.

First, conditional on reading the letter, the information shock introduced in the letter may not fully materialize in the reader's home price expectations. For example, a subject

[16] If in the future we collect administrative data for a time horizon longer than 28 weeks after the start of letter delivery, the degree of censoring may be less severe, because more homes would be sold in the longer horizon).

may not react to the information provided in the letter for several reasons (e.g., they have already seen the information, they do not trust the source, or they feel confident in their prior beliefs). In Section 4.1, we use the auxiliary survey to estimate the rate of pass-through from information shocks to expectations: a 1 pp higher information shock increased home price expectations by 0.205 pp. The results from this auxiliary sample should be taken in context, however, because the incentives to pay attention to information about home prices may be much lower than it is in the sample used for the field experiment. However, if we extrapolate the results from the auxiliary survey, these estimates imply that a 1 pp increase in home price expectations causes a decline of 1.59 pp $=(\frac{0.325}{0.205})$ in sales probability (i.e., a behavioral elasticity between expectations and sales probability of 1.59).

The second source of non-compliance is that some letters may not have been delivered, were delivered but not read, or were read after the property already sold. To correct for this type of non-compliance, as in Perez-Truglia and Cruces (2017), we estimated a reading rate (i.e., the share of recipients who read the letter in time). According to the U.S. Monitor Non-Profit Standard Mail Delivery Study, around 5% of standard non-profit mailers fail to be delivered (U.S. Monitor, 2014). Based on data from the U.S. Postal Service Household Diary Survey (Mazzone and Rehman, 2019), we estimate that, conditional on delivery, around 26% of our letters were not read by the recipient.[17] Based on the timing of survey responses and the timing of sales in the baseline group, we estimated that roughly 7.7% of the letters were not read until after the property was sold. Combining these three estimates leads to a reading rate of 64.9% ($= 0.95 \cdot 0.74 \cdot 0.923$). If anything, this estimate is likely conservative, as it could be a bit smaller under alternative assumptions and thus could lead to an even larger scale-up factor. See Appendix B.12 for more details. To account for this source of attenuation bias, the scaled-up behavioral elasticity between expectations and sales probability becomes 2.45 $(= \frac{1.59}{0.649})$.

This elasticity of -2.45 between market choices and home price expectations suggests that sellers were highly elastic to their expectations. Moreover, this estimate implies that the large variation in expectations typically seen in survey data across individuals and over time could influence market choices and equilibrium outcomes. For example, according to the prior belief measures in our supplemental survey, the standard deviation in home price expectations across individuals is 5.39 pp.[18] According to the elasticity of -2.45, an increase in home price expectations of this magnitude (5.39 pp) would reduce the probability that

[17]This 26% figure is based on the 2018 HDS Recruitment Sample and corresponds to the estimate of treatment of advertising mail reported in Figure 5.3 of (Mazzone and Rehman, 2019).

[18]This measure corresponds to the raw standard deviation. The results are similar (5.51 instead of 5.39) if we instead use the standard deviation within a given property type (i.e., same ZIP code and number of beds).

one of our subjects sells the property within 28 weeks by 13.21 pp ($= 2.45 \cdot 5.39$).

4.5 Role of Optimization Frictions

Because these estimates pertain to the population as a whole, they may mask meaningful heterogeneity. In particular, some subjects may want to react more strongly to changes in their home price expectations, but they may be unable to due to optimization frictions. To explore this hypothesis, we study the (pre-registered) heterogeneity of the effects of information by assessing whether the listed property is owner-occupied (78.4% of subjects) or non-owner-occupied (21.6% of subjects).

We argue that, relative to subjects who live on the listed property, those who do not face fewer optimization frictions. First, the occupants are likely to have less flexibility regarding *when* they can sell their properties, as the decision to sell is consequential for their daily lives: they have to move out after selling it, or the occupants may be selling for time-sensitive reasons, such as a new job or school, because they already bought another home, or because of marriage, divorce, or the birth of a child. For the non-occupants, the decisions on whether and when to sell usually are based purely on investment motives. Second, the owner-occupant is likely less flexible regarding *where* (or if) to buy another house, especially if they need to buy (or already bought) another home in the same or similar neighborhood.[19] As a result, whether these sellers expect home values to appreciate in their current neighborhood may not be relevant in their decisions to sell, because they will be exposed to the same neighborhood appreciation regardless of when or if they sell their current property. On the contrary, non-occupant-owners do not need to buy again in the same neighborhood and can instead buy real estate in other neighborhoods that they expect to appreciate more or make entirely different investments, such as in the stock market.

Occupant and non-occupant owners differ markedly in their optimization frictions but are otherwise similar in observable characteristics, such as age, ethnic composition, income, and education. Despite some differences, the timing of sales outcomes does not look dramatically different: the probability of selling the property at 28 weeks post-treatment was 52.89 pp for non-occupant owners versus 57.99 pp for occupant owners (for more details, see Appendix B.13).

Figure B.6 presents the heterogeneity results. This figure shows the event-study analysis of the effects of the information shocks, broken down by owner occupancy status. We find that the effects of information shocks are qualitatively similar for occupant owners (depicted

[19] For example, according to mail forwarding data from the U.S. Postal Services, a significant fraction of owner-occupants in our sample moved to another home in the same or adjacent neighborhoods (results presented in Appendix B.13).

in blue circles) and non-occupant owners (depicted in red diamonds): they point in the same direction and are both statistically significant. Consistent with the above conjecture, however, we find that the effects are quantitatively quite different: the effects are almost three times as strong for the non-occupant owners as for the occupant owners. For example, at 28 weeks after the start of letter delivery, the coefficient on the information shock is -0.637 (p-value=0.003) for the non-occupants and -0.225 (p-value=0.066) for the occupants, and the difference between these two coefficients is statistically significant (p-value=0.095). This large difference suggests that, in the absence of optimization frictions, the market choices of individuals would be substantially more elastic to their expectations. Indeed, this evidence on the role of optimization friction is consistent with suggestive evidence from other contexts such as the choice of how much equity to hold (Giglio et al., 2019). When we apply the same corrections for the previous pass-through rate (0.205) and reading rate (0.649), the coefficient for the non-occupant owners (-0.637) implies an elasticity between expectations and market choices of -4.79 pp $=(\frac{-0.637}{0.205 \cdot 0.649})$: that is, a 1 pp increase in home price expectations causes a decline of 4.79 pp in the probability of selling the home within 28 weeks. This estimate shows that, in the absence of optimization frictions, home price expectations can have an even larger effect on market choices. Additionally, this evidence supports the widespread view that non-owner-occupied properties have a disproportionate influence on speculation in the housing market (Gao et al., 2020).

Our favorite interpretation for the heterogeneity by occupant status is based on the two previously mentioned frictions. However, we cannot rule out alternative interpretations. Perhaps non-occupant owners are more responsive because they are different in other respects, such as financial sophistication, and thus more likely to understand the information and use it correctly.[20] However, the similarities in these two groups across observable characteristics, including income and education, constitute evidence against this alternative interpretation.

5 Conclusions

We provide the first experimental evidence on the causal effects of home price expectations on market choices. We conducted a large-scale, high-stakes, field experiment involving 57,910 U.S. individuals who recently listed their homes on the market. We sent letters to them with randomized information about their home prices to generate exogenous shocks to their home price expectations. We then used the administrative records to measure the effects of those information shocks on the recipients' decisions to sell their homes. Consistent with economic

[20] Also, perhaps non-occupant owners are more likely to incorporate the information shocks into their expectations because they are less informed about local home prices due to not living in the area.

theory, we found that higher home price expectations caused the owners to delay selling their homes. Moreover, the magnitude of this reaction indicates that market choices were highly elastic to expectations: a 1 pp increase in home price expectations reduced the probability of selling within six months by 2.45 pp.

Our estimates address an open question that is central to the literature on macroeconomics and finance: what are the causal effects of expectations on market choices? We believe our experimental framework can be adapted to study other questions in a range of fields, such as macroeconomics, urban economics, finance, real estate economics, and behavioral economics. The information provided in the mailings does not need to be related to home prices. The letters can provide any type of information that the researcher hypothesizes may be relevant to the decision to sell. Our experimental framework also has several advantages that could warrant its widespread adoption. For example, it provides causal inferences based on experimental variation. Also, the effects on behavior are measured using objective data from administrative records in a naturally occurring context and based on high-stakes choices. Our framework has practical advantages, too. Following our detailed instructions, the experiment can be implemented relatively quickly after a few weeks of preparation, and the results can be ready in a few months. The experiment is relatively cheap, costing less than $0.25 per subject. Last, the experiment can be implemented at massive scales, with potentially up to 1 million subjects at a time.[21]

References

Armantier, O., W. Bruine de Bruin, G. Topa, W. van der Klaauw, and B. Zafar (2015). Inflation Expectations and Behavior: Do Survey Respondents Act on their Beliefs? *International Economic Review 56*(10), 505–536.

Armantier, O., S. Nelson, G. Topa, W. van der Klaauw, and B. Zafar (2016). The Price Is Right: Updating Inflation Expectations in a Randomized Price Information Experiment. *Review of Economics and Statistics 98*(3), 503–523.

Armona, L., A. Fuster, and B. Zafar (2019). Home Price Expectations and Behaviour: Evidence from a Randomized Information Experiment. *Review of Economic Studies 86*(4), 1371–1410.

Bailey, M., R. Cao, T. Kuchler, and J. Stroebel (2018). The Economic Effects of Social Networks: Evidence from the Housing Market. *Journal of Political Economy 126*(6), 2224–2276.

[21]For example, Zillow Research (https://www.zillow.com/research/data/) shows 1,705,251 unique active listings in May 2019. County assessor records likely include data for most of these listings.

Bailey, M., E. Davila, T. Kuchler, and J. Stroebel (2018). House Price Beliefs And Mortgage Leverage Choice. *The Review of Economic Studies 86*(6), 2403–2452.

Bergolo, M. L., R. Ceni, G. Cruces, M. Giaccobasso, and R. Perez-Truglia (2017, 7). Tax Audits as Scarecrows: Evidence from a Large-Scale Field Experiment. *NBER Working Paper No. 23631*.

Bernanke, B. S. (2007). *Inflation Expectations and Inflation Forecasting.* Cambridge, MA: Speech at the Monetary Economics Workshop of the NBER Summer Institute.

Bottan, N. and R. Perez-Truglia (2017). Choosing Your Pond: Location Choices and Relative Income. *NBER Working Paper No. 23615*.

Brooks, D. (2017). The Home Buying Decision. *New York Times, January 6, 2017*.

Carroll, C. (2003). Macroeconomic Expectations of Households and Professional Forecasters. *Quarterly Journal of Economics 118*(1), 269–298.

Case, K. and R. Shiller (1989). The Efficiency of the Market for Single-Family Homes. *American Economic Review 79*(1), 125–137.

Cavallo, A., G. Cruces, and R. Perez-Truglia (2017). Inflation Expectations, Learning, and Supermarket Prices: Evidence from Survey Experiments. *American Economic Journal: Macroeconomics 9*(3), 1–35.

Coibion, O., Y. Gorodnichenko, and S. Kumar (2018). How Do Firms Form Their Expectations? New Survey Evidence. *American Economic Review 108*(9), 2671–2713.

Crump, M. J. C., J. V. McDonnell, and T. M. Gureckis (2013). Evaluating Amazon's Mechanical Turk as a Tool for Experimental Behavioral Research. *PLOS ONE 8*(3), 1–18.

Cullen, Z. and R. Perez-Truglia (2018). How Much Does Your Boss Make? The Effects of Salary Comparisons. *NBER Working Paper No. 24841*.

Fuster, A., R. Perez-Truglia, M. Wiederholt, and B. Zafar (2018). Expectations with Endogenous Information Acquisition: An Experimental Investigation. *NBER Working Paper No. 24767*.

Gao, Z., M. Sockin, and W. Xiong (2020). Economic Consequences of Housing Speculation. *The Review of Financial Studies, forthcoming*.

Gennaioli, N. and A. Shleifer (2018). *A Crisis of Beliefs: Investor Psychology and Financial Fragility.* New Jersey: Princeton University Press.

Giglio, S., M. Maggiori, J. Stroebel, and S. Utkus (2019). Five Facts about Beliefs and Portfolios. *NBER Working Paper No. 25744*.

Glaeser, E. L. and C. G. Nathanson (2015). Housing Bubbles. In G. Duranton, J. V. Henderson, and W. C. Strange (Eds.), *Handbook of Regional and Urban Economics*, Volume 5, Chapter 11, pp. 701–751. Elsevier.

Kaplan, G., K. Mitman, and G. L. Violante (2019). The Housing Boom and Bust: Model Meets Evidence. *Journal of Political Economy, forthcoming*.

Mazzone, J. and S. Rehman (2019). The Household Diary Study Mail Use and Attitudes in FY 2018. *Retrieved March 28, 2020, from https://www.prc.gov/dockets/document/109368*.

Perez-Truglia, R. and G. Cruces (2017). Partisan interactions: Evidence from a field experiment in the United States. *Journal of Political Economy 125*(4), 1208–1243.

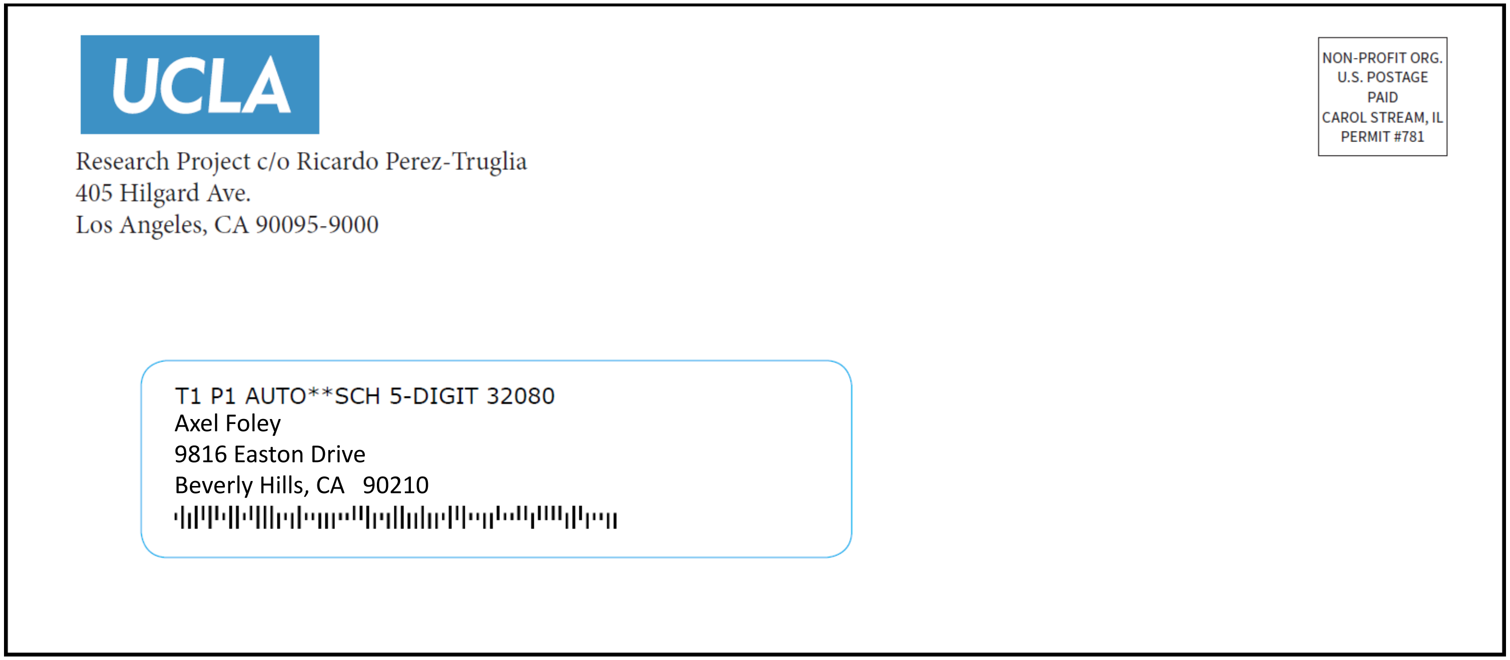
Perez-Truglia, R. and U. Troiano (2018). Shaming Tax Delinquents. *Journal of Public Economics 167*, 120–137.

Roth, C. and J. Wohlfart (2019). How Do Expectations About the Macroeconomy Affect Personal Expectations and Behavior? *Review of Economics and Statistics*.

Shiller, R. (2005). *Irrational Exuberance.* New Jersey: Princeton University Press.

U.S. Monitor (2014). 7 Myths of Direct Mailing. *Retrieved March 28, 2020, from https://www.targetmarketingmag.com/promo/7MythsofDM.pdf*.

Figure 1: Sample Envelope



Notes: Screenshot of the outside of the envelope used in the field experiment.

Figure 2: Sample Letter

a. First Page

UCLA

Los Angeles, May 31st 2019

Dear **Axel Foley**,

We are researchers at UCLA and we are reaching out to you as part of a research study about decision making of homeowners.

According to our records, you may be considering selling a property. We know these decisions can be difficult, so we want to share some information that we hope can be helpful:

<<INFORMATION>>

If you would like to help us with our study, we kindly ask you fill out the following 2-minute survey:

Visit **www.surveyhousing.com** and enter validation code

Participation is voluntary and responses are 100% confidential. The results of this study can provide valuable insights to homeowners across the country. Your participation in the survey is greatly appreciated.

110 Westwood Plaza, Suite C515
Los Angeles, CA 90095-1481

Website: http://www.anderson.ucla.edu/housingstudy

Please recycle

b. Second Page

Your household was randomly chosen to receive this letter. *We will not send you any more letters in the future.*

If you have any questions about the study, you can find contact information on our website: www.anderson.ucla.edu/housingstudy.

Thank you for your attention!

Ricardo Perez-Truglia
Assistant Professor of Economics
University of California, Los Angeles

Nicolas Bottan
Post-Doctoral Associate
Cornell University

If you have questions about your rights as a research subject, or you have concerns or suggestions and you want to talk to someone other than the researchers, you may contact the UCLA Office of the Human Research Protection Program by phone: (310) 206-2040; by email: participants@research.ucla.edu or by mail: Box 951406, Los Angeles, CA 90095-1406.

Methodological Notes:

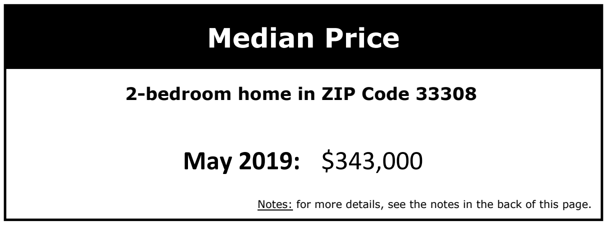
<<INFORMATION DETAILS>>

T1 P1 AUTO**SCH 5-DIGIT 32080
Axel Foley
9816 Easton Drive
Beverly Hills, CA 90210

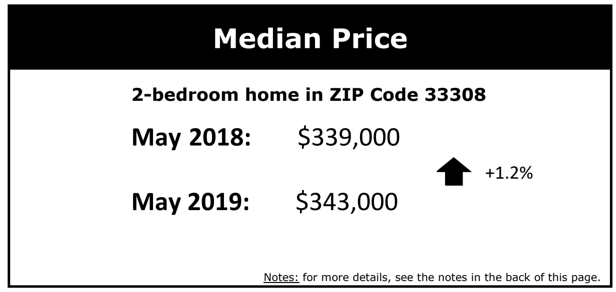
Notes: Screenshot of the letter used in the field experiment. The two placeholders (marked as *«Information»* and *«Information Details»*) indicate the placement of the two components that were randomly allocated. Their samples, by treatment group, are presented in Figures 3 and B.1 respectively. Appendix C shows a sample of the final product.

Figure 3: Sample Information Tables

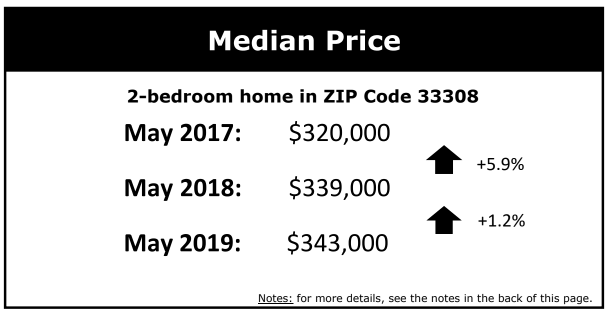
a. Baseline



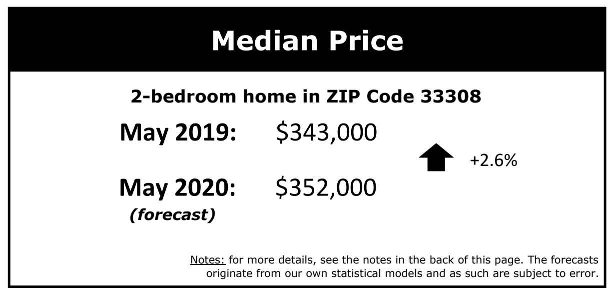
b. Past-1



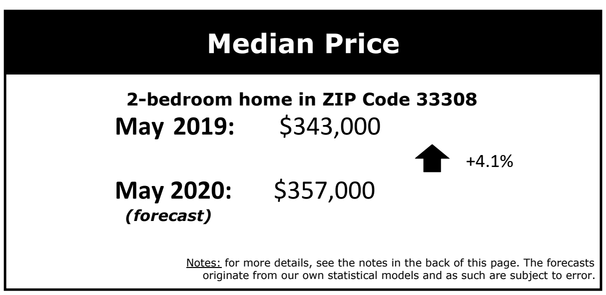
c. Past-2



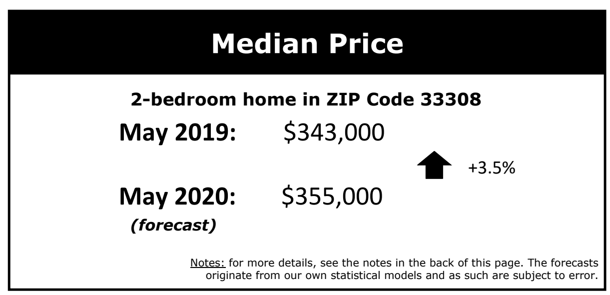
d. Forecast-1



e. Forecast-2



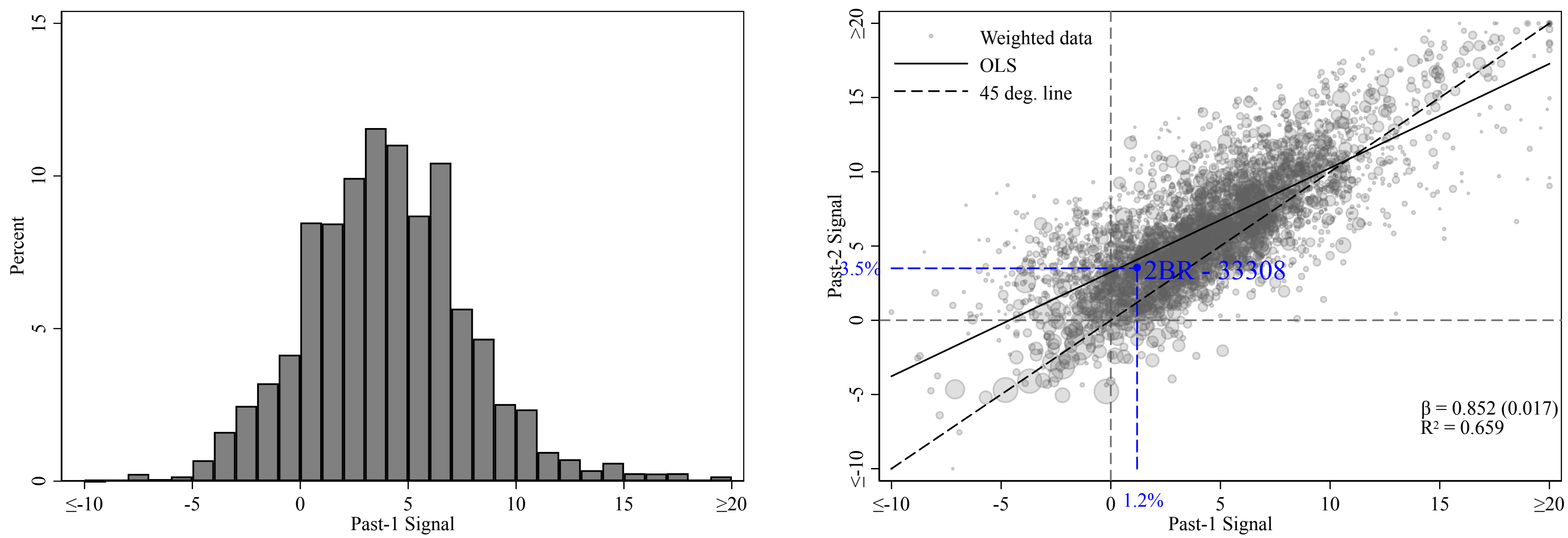
f. Forecast-3



Notes: Each panel corresponds to the hypothetical table that a given individual would receive under the different treatment groups. The table is then placed in the middle of the first page of the letter, in the location of the placeholder *«Information»* from Figure 2. See Figure B.1 for the methodological notes accompanying each table.

Figure 4: Heterogeneity in Signals Within and Across Information Sources

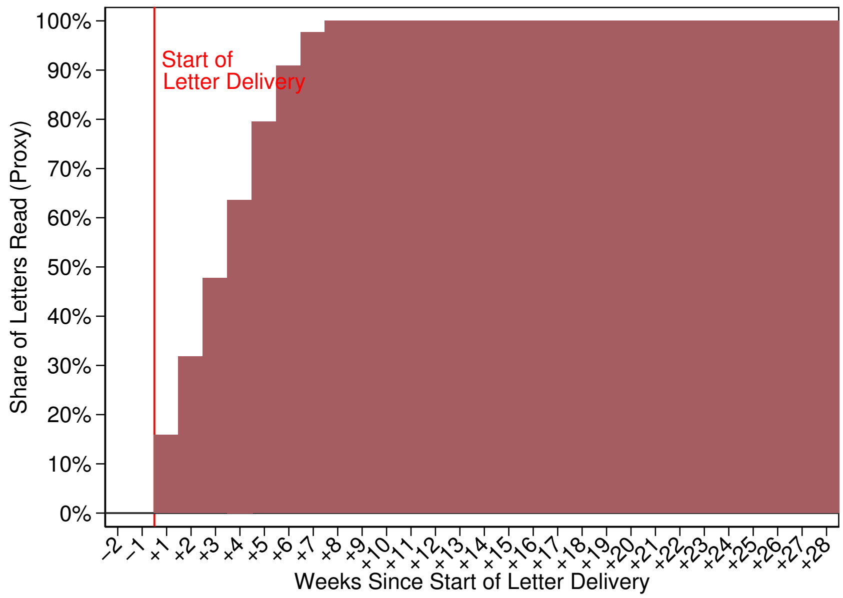
a. Past-1 **b.** Past-1 vs. Past-2



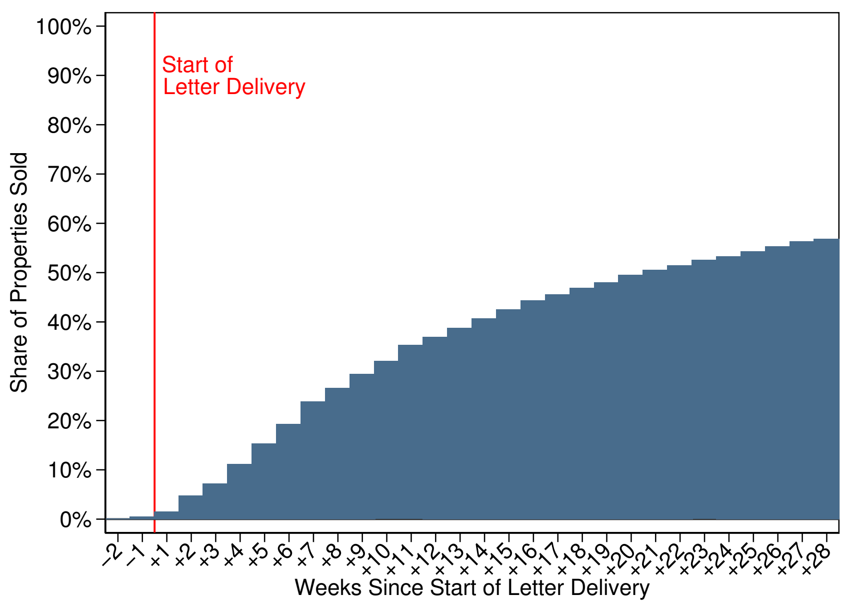
Notes: Panel (a) shows the distribution of the signals that the 57,910 subjects would have received if they had been assigned to the Past-1 treatment (i.e., the annual growth rate over the past year). The bins have a width of 1 pp and are truncated at -10% and +20%. Panel (b) is a scatterplot showing the relationship between the signals that the 57,910 subjects would have received if they had been assigned to the Past-1 treatment (i.e., the annual growth rate over the past year) versus the Past-2 treatment (i.e., the annual growth rate over the past two years). The size of the circles are proportional to the number of observations, and the signals are truncated at -10% and +20%.

Figure 5: Timing of Read-Receipt and Property Sales

a. Read-Receipt



b. Property Sales

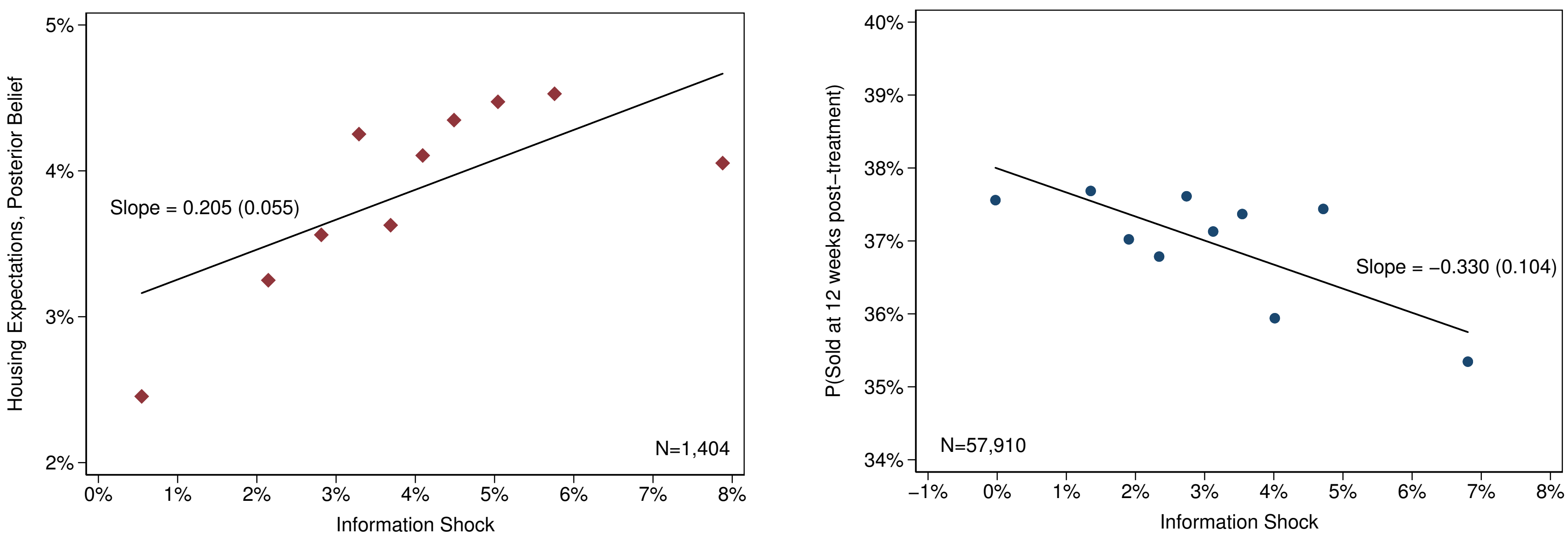


Notes: The red line indicates the estimated delivery date for the first letter (June 15 2019). Panel (a) shows the evolution of the responses to the online survey included in the letter. These dates constitute our read-receipt: i.e., our proxy for the dates when the letters were actually read. Panel (b) shows the fraction of the properties in the subject pool that were sold at each point in time, according to the administrative records.

Figure 6: Effects of Information Shocks on Expectations and Behavior: Binned Scatterplots

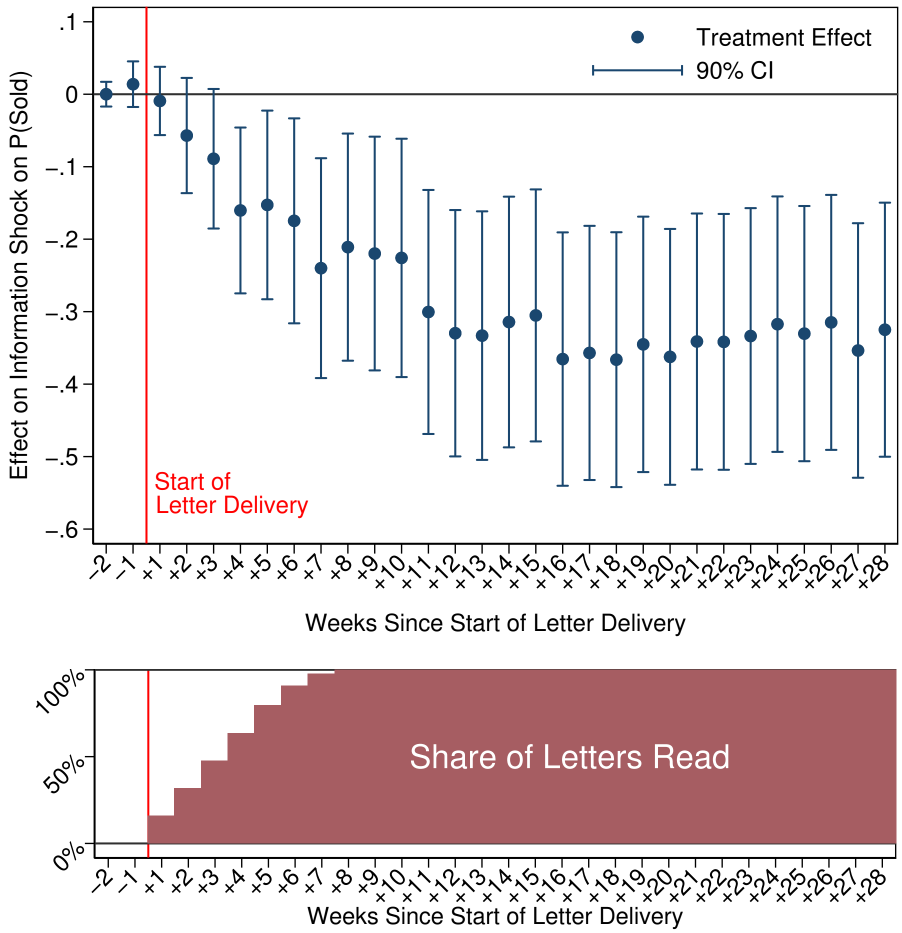
a. Effects on Expectations

b. Effects on Behavior



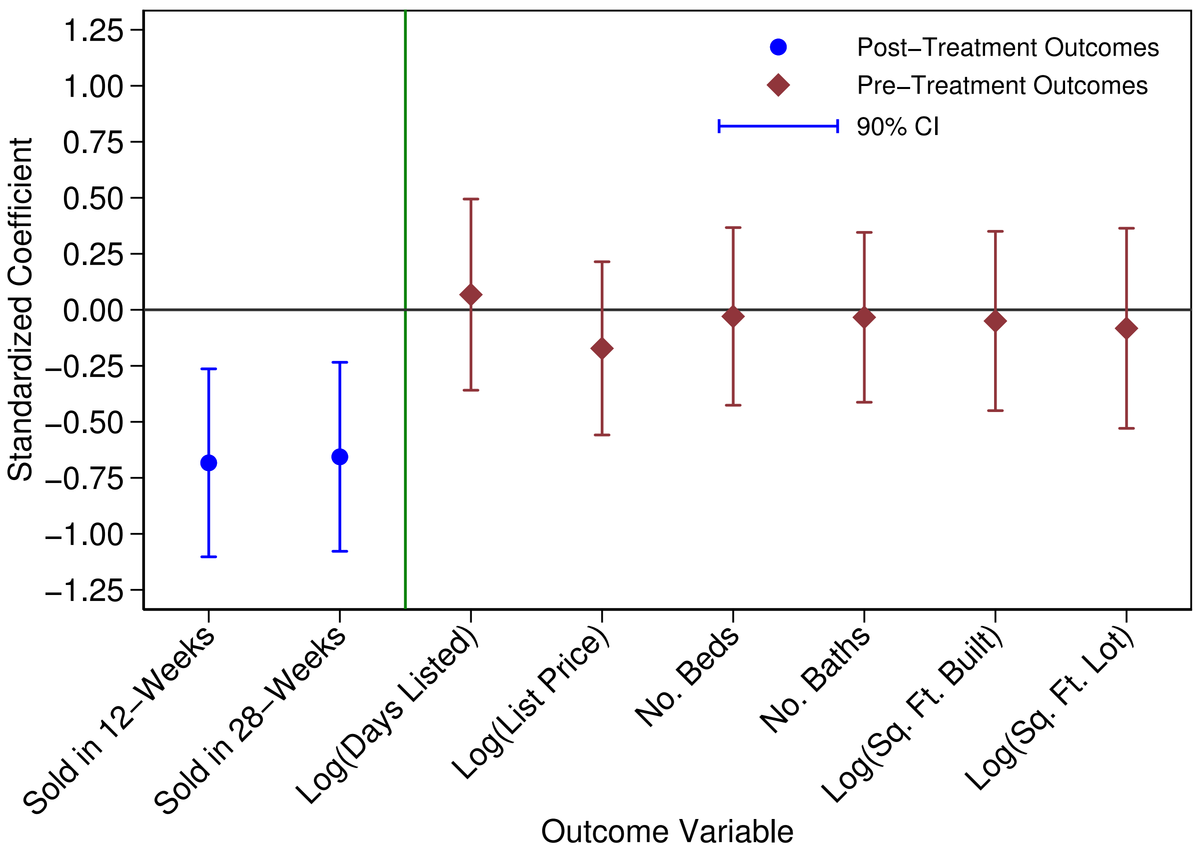
Notes: Panel (a) corresponds to a regression given by equation (1) from Section 2. This binned scatterplot focuses on the key independent variable, *Information Shock* $(E_i^{j_i^*} \cdot T_i)$. Results are based on 1,404 subjects from the AMT supplemental survey. The dependent variable is the posterior belief (i.e., elicited after the information-provision experiment) about the expected growth rate of the median home value over the following year. Panel (b) corresponds to a regression given by equation (1). This binned scatterplot focuses on the key independent variable, *Information Shock* $(E_i^{j_i^*} \cdot T_i)$. Results are based on 57,910 subjects from the field experiment. The dependent variable is an indicator variable taking the value 100 if the property was sold at 12 weeks after the start of the letter delivery and 0 otherwise. Each panel reports the slope with heteroskedasticity-robust standard errors in parentheses.

Figure 7: Effects of Information Shocks on Behavior: Event-Study Analysis



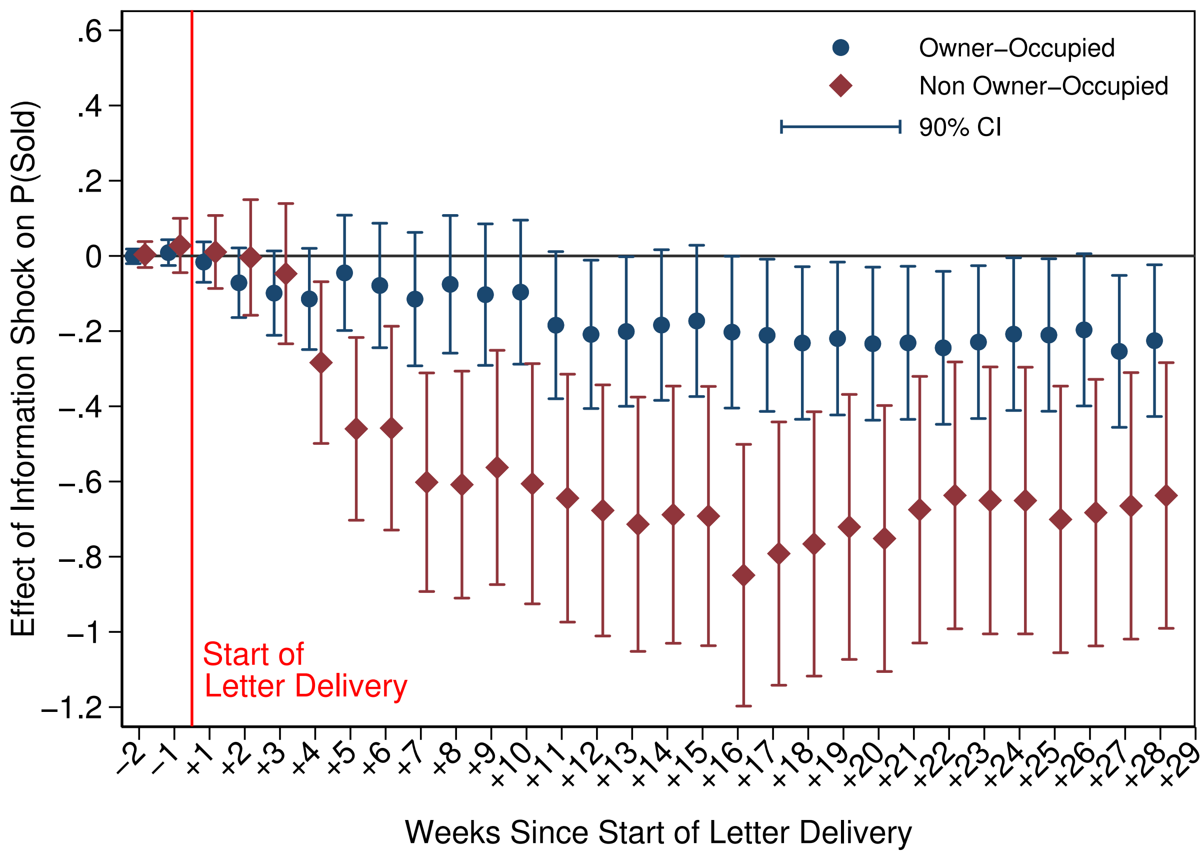
Notes: Each coefficient corresponds to a separate regression based on 57,910 subjects from the field experiment. Every regression corresponds to equation (1) from Section 2, and the coefficient being graphed corresponds to the coefficient on the key independent variable, *Information Shock* ($E_i^{j_i^*} \cdot T_i$). All regressions are identical except for the dependent variable. The x-axis indicates the dependent variable used, which is always an indicator variable that takes the value 100 if the property has been sold at a number of weeks after the start of the letter delivery and 0 otherwise. For example, the coefficient on *+12 weeks* is based on a dependent variable that takes the value 100 if the property was sold at 12 weeks after the start of the letter delivery. The red line indicates the estimated delivery date for the first letter (June 15 2019). The smaller figure at the bottom shows the proportion of the letters from the field experiment that had been read at every point in time according to our proxy (the responses to the online survey included in the letter). The 90% confidence intervals are based on heteroskedasticity-robust standard errors.

Figure 8: Effects of Information Shocks on Behavior: Placebo Outcomes



Notes: All the regressions shown in this table correspond to equation (1) from Section 2, based on data on the 57,910 subjects in the field experiment. Each coefficient corresponds to a separate but identical regressions, with the only difference being the dependent variables. We report the coefficient on the key independent variable, *Information Shock* ($E_i^{j_i^*} \cdot T_i$). All the coefficients have been normalized by multiplying it by 100 and then dividing it by the standard deviations of the respective dependent variable. Each dependent variable is listed in the x-axis. We use blue circles to denote the post-treatment outcomes (i.e., that were determined after the start of letter delivery): *Sold in 12-Weeks* is an indicator variable that takes the value 100 if the property was sold at 12 weeks after the start of the letter delivery; and *Sold in 28-Weeks* is an indicator variable that takes the value 100 if the property was sold at 28 weeks after the start of the letter delivery. We use red circles to denote the pre-treatment outcomes (i.e., that were determined before the start of letter delivery): *Log(Days Listed)* is the logarithm of the number of days that the property had been listed for before our experiment; *Log(Listing Price)* is the logarithm of the original listing price of the property; *No. Beds* is the property's number of bedrooms; *No. Baths* is the property's number of bathrooms; *Log(Sq. Ft. Built)* is the logarithm of the property's built area in square feet; *Log(Sq. Ft. Lot)* is logarithm of the property's lot size in square feet. The 90% confidence intervals are based on heteroskedasticity-robust standard errors.

Figure 9: Heterogeneity of Behavioral Effects by Owner-Occupied Status: Event-Study Analysis



Notes: Each coefficient corresponds to a separate regression. Every regression corresponds to equation (1) from Section 2, and the coefficient being graphed corresponds to the coefficient on the key independent variable, *Information Shock* ($E_i^{j_i^*} \cdot T_i$). All regressions are identical except for two features: the dependent variable and the sample. Each of the blue circles are based on a regression with the 45,405 subjects from a field experiment who were living on the property while the property was listed for sale. Each of the red diamonds are based on a regression with the 12,505 subjects from the field experiment who were not living on the property while the property was listed for sale. The x-axis indicates the dependent variable used, which is always an indicator variable that takes the value 100 if the property has been sold at a number of weeks after the start of the letter delivery and 0 otherwise. For example, the coefficient on *+12 weeks* is based on a dependent variable that takes the value 100 if the property was sold at 12 weeks after the start of the letter delivery. The red line indicates the estimated delivery date for the first letter (June 15 2019). The 90% confidence intervals are based on heteroskedasticity-robust standard errors.

Table 1: Descriptive Statistics and Randomization Balance Test

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		By Treatment Group						
	All	Baseline	Past-1	Past-2	Forecast-1	Forecast-2	Forecast-3	P-value
Days Listed	86.654	85.997	87.212	87.029	85.936	86.208	87.778	0.829
	(0.477)	(1.017)	(1.298)	(1.253)	(1.093)	(1.203)	(1.191)	
List Price ($1,000s)	574.756	575.774	586.787	559.192	574.306	586.835	565.092	0.303
	(3.914)	(8.517)	(11.445)	(9.516)	(8.147)	(10.937)	(9.081)	
No. Beds	3.256	3.249	3.259	3.243	3.254	3.269	3.264	0.582
	(0.005)	(0.010)	(0.012)	(0.012)	(0.011)	(0.011)	(0.011)	
No. baths	2.608	2.607	2.619	2.599	2.608	2.617	2.600	0.678
	(0.004)	(0.010)	(0.011)	(0.011)	(0.010)	(0.011)	(0.010)	
Sq. Ft. Built (1,000s)	2.295	2.292	2.304	2.292	2.287	2.308	2.288	0.820
	(0.005)	(0.012)	(0.014)	(0.013)	(0.012)	(0.013)	(0.013)	
Sq. Ft. Lot (1,000s)	12.958	12.992	12.666	12.934	13.223	13.163	12.730	0.389
	(0.089)	(0.199)	(0.222)	(0.230)	(0.219)	(0.221)	(0.213)	
	57,910	11,487	8,672	8,669	9,818	9,635	9,629	

Notes: Average characteristics on the 57,910 subjects in the field experiment, with standard errors reported in parentheses. Column (1) corresponds to the entire sample. Columns (2) through (7) correspond to each of the six treatment groups. Column (8) reports the p-value of the test of equal means across all six treatment groups. All the variables correspond to pre-treatment characteristics (i.e., that were determined before the start of letter delivery). *Days Listed* is the number of days that the property had been listed for before our experiment. *List Price* is the original listing price of the property. *No. Beds* is the property's number of bedrooms. *No. Baths* is the property's number of bathrooms. *Sq. Ft. Built* is the property's built area in square feet. *Sq. Ft. Lot* is the property's lot size in square feet.

Table 2: Main Regression Results

	Survey Data				Behavioral Data			
	(1) H_{1y}^{post}	(2) H_{5y}^{post}	(3) H_{1y}^{prior}	(4) M_{1y}^{post}	(5) S_{+12w}	(6) S_{+28w}	(7) S_{-1w}	(8) D_{pre}
Information Shock	0.205***	0.167**	-0.014	0.017	-0.330***	-0.325***	0.014	0.001
	(0.064)	(0.070)	(0.066)	(0.134)	(0.103)	(0.107)	(0.019)	(0.003)
Mean Outcome	3.86	2.31	3.88	3.58	36.99	56.90	0.58	3.81
Std. Dev. Outcome	4.42	4.36	5.39	9.05	48.28	49.52	7.61	1.28
Observations	1,404	1,404	1,404	1,404	57,910	57,910	57,910	57,910

Notes: Significant at *10%, **5%, ***1%. Heteroskedasticity-robust standard errors in parentheses. Each column corresponds to a different regression. All regressions correspond to equation (1) from Section 2, with *Information Shock* referring to the key independent variable: $E_i^{j_i^*} \cdot T_i$. The only difference between columns is that they use a different dependent variable. Columns (1) through (4) are based on data from the AMT supplemental survey. H_{1y}^{post} is the posterior belief (i.e., elicited after the information-provision experiment) about the expected growth rate of the median home value over the following year. H_{5y}^{post} is is the posterior belief about the annualized expected growth growth rate of the median home value over the following five years. H_{1y}^{prior} the prior belief (i.e., elicited before the information-provision experiment) about the expected growth rate of the median home value over the following year. M_{1y}^{post} is the posterior belief about the annualized expected growth growth rate of the stock market index over the following year. Columns (5) through (8) are based on data from the field experiment. S_{+12w} is an indicator variable that takes the value 100 if the property was sold at 12 weeks after the start of the letter delivery. S_{+28w} is an indicator variable that takes the value 100 if the property was sold at 28 weeks after the start of the letter delivery. S_{-1w} is an indicator variable that takes the value 100 if the property was sold at 1 week prior to the start of the letter delivery. And D_{pre} is the logarithm of the number of days that the property had been listed for before our experiment.

Table 3: Additional Robustness Checks

	Dep. Var: S_{+12w}									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Information Shock	-0.330***	-0.325***	-0.325***	-0.286**	-0.330***	-0.338***	-0.257*	-0.325***	-0.382***	-0.320***
	(0.103)	(0.100)	(0.111)	(0.146)	(0.119)	(0.118)	(0.137)	(0.110)	(0.111)	(0.106)
Additional Controls		Y								
Extended Dummies			Y							
Control for Feedback				Y						
Group Left Out					Baseline	Past-1	Past-2	Forecast-1	Forecast-2	Forecast-3
R^2	0.034	0.111	0.034	0.034	0.032	0.035	0.034	0.033	0.034	0.034
Observations	57,910	57,910	57,910	57,910	46,423	49,238	49,241	48,092	48,275	48,281

Notes: Significant at *10%, **5%, ***1%. Heteroskedasticity-robust standard errors in parentheses. Each column corresponds to a different regression. All regressions are based on data from the field experiment and using the same dependent variable: an indicator variable (S_{+12w}) that takes the value 100 if the property was sold at 12 weeks after the start of the letter delivery and 0 otherwise. Column (1) corresponds to equation (1) from Section 2, with *Information Shock* referring to the key independent variable: $E_i^{j_i^*} \cdot T_i$. Column (2) is identical to column (1) except that it includes some additional control variables: the logarithm of the days the property was on the market prior to the experiment, the logarithm of the initial listing price, four dummies for number of beds, four dummies for number of bedrooms, the logarithm of square footage built, the logarithm of lot size, and six state dummies. Column (3) is identical to column (1) except that instead of controlling for one treatment indicator, it controls for a set of five treatment indicators (i.e., one for each of the five treatments that are not Baseline). Column (4) is identical to column (1) except that it includes an additional control variable: the value of the signal chosen for the individual without the treatment interaction (i.e., $E_i^{j_i^*}$). Column (5) through (10) are identical to column (1) except that they exclude subjects for one treatment group at a time.