# Online Appendix (For Online Publication Only)

**Bottan and Perez-Truglia, "Betting on the House"**
**April 19, 2024**

# A Data Construction

## A.1 Data Sources

In this study we rely on two main data sources: online property listings and Assessor's Office data. In Figure A.2 we present screenshots to illustrate the original sources of data and how these were combined. First, we obtain the list of all active listings in a given real estate market from an online listing provider (e.g., Figure A.2.a). We are able to collect detailed information about the characteristics of each property, such as number of bedrooms and bathrooms, listing price, address, among other characteristics. In particular, we are interested in the unique parcel identifier (APN) – typically shown along with additional public information for the property (Figure A.2.c). We then use these unique identifiers to obtain additional information on the property and, importantly, their owners from the corresponding County Assessor's Office data (Figure A.2.d). Depending on the County, these data could be directly downloaded from the website, or data was shared through a request, or had to be scraped from the website.

## A.2 Implementation of the Field Experiment

This section describes the implementation of the field experiment. First we describe how the target populations (counties) were selected. Second, we define the eligibility criteria used to generate our experimental sample of properties.

Two main factors determined the geographic coverage of our mailing experiment. The first, and by far the most important, was access to the Assessor's data. Even though these data are publicly available, access varies greatly across counties (even within the same state). We manually reviewed accessibility to these data for the most populous counties in the United States. From this review, we identified 6 counties in different states (Los Angeles, CA; Maricaopa, AZ; Clark, NV; Cuyahoga, OH; King, WA; Harris, TX), and all counties in the State of Florida (the Florida Department of Revenue collects and publishes yearly secured property tax rolls for the entire state).

The second factor was the geographic coverage of the listing website. Even though coverage in large counties is not a problem, we had to drop some of the smaller counties in Florida because the listing website did not cover those markets. This left us a sample of 36 counties in 7 states.

The next step is to define the eligibility criteria for listings to our study. First, and most important, we want to focus on *individuals* who own and are selling their property. Therefore, we exclude all properties that are owned by business, banks, trusts, etc.[40] Second, we restrict to properties classified as residential by the County Assessor. Third, we restrict to individuals who own a single property in the county.

Figure A.3.a. shows the number and locations of listed properties for which we sent owners a letter. Figure A.3.b. shows the location where the owners lived at the time we sent out the letters (i.e., their mailing address according to Assessor's data). Most subjects (66.9%) live in the property that they listed.

## A.3 The Statistical Forecasts

The experimental variation in the paper comes from five different treatments including three different forecasts. In this section we provide a detailed description of the statistical models used to generate the three forecasts.

**Forecast-1**: The first statistical model uses five lags of the annual growth rate of home prices in the same ZIP code. Specifically, let $P_{z,y}$ be the home price in zip code $z$ in year $y$. $\Delta P_{z,y}$ is the annual percentage point change in the home price level for zip code $z$ in year $y$. The forecast-1 model is specified as follows:

$$\Delta P_{z,y} = \sum_{j=1}^{5} \Delta P_{z,y-j} + \epsilon_{z,y} \tag{A.1}$$

**Forecast-2**: The second statistical model uses five lags of the annual growth rate of home prices in the same ZIP code, plus five lags of the average annual growth rate of home prices in the same state. Let $P_{z,y}$ be the home price level in zip code $z$ in year $y$, and $P_{s,y}$ be the home price level in state $s$ where zip code $z$ is located. $\Delta P_{z,y}$ and $\Delta P_{s,y}$ are the annual percentage point changes in the home price level in year $y$ for zip code $z$ and state $s$ respectively.

$$\Delta P_{z,y} = \sum_{j=1}^{5} \Delta P_{z,y-j} + \sum_{j=1}^{5} \Delta P_{s,y-j} + \epsilon_{z,y} \tag{A.2}$$

[40] These were identified through regular expressions on the owner name field. For example, if the field included "LLC," "CORP," "estate," etc. We employed a very conservative approach when designing the list for classifying owners. Names in the final data set were manually checked to verify that owners were indeed individuals.

**Forecast-3**: The third model uses three lags of the annual growth rate of home prices in the same ZIP code, three lags of the average annual growth rate of home prices in the same city, and three lags of the city-level employment rate. Similar to the set-ups of the previous two models, let $P_{z,y}$ be the home price level in zip code $z$ in year $y$, and $P_{c,y}$ be the home price level in city $c$ where zip code $z$ is located. $\Delta P_{z,y}$ and $\Delta P_{c,y}$ are the annual percentage point changes in the home price level in the zip code and city respectively. $L_{c,y}$ is the employment rate in city $c$ in year $y$.

$$\Delta P_{z,y} = \sum_{j=1}^{3} \Delta P_{z,y-j} + \sum_{j=1}^{3} \Delta P_{c,y-j} + \sum_{j=1}^{3} L_{c,y-j} + \epsilon_{z,y} \tag{A.3}$$

In subsection A.5 below we provide statistics showing that on average these three models perform similar, and also are comparable to past information and to an external benchmark (Zillow forecasts).

## A.4 Further Descriptive Statistics

In this section we provide general descriptive statistics on the subject pool for the field experiment. We contrast our sample to a representative sample of homeowner household heads using data from the American Community Survey. Figure A.3 shows the geographic distribution of the listed properties (in panel (a)) and the location of homeowners' mailing addresses (in panel (b)) from the field experiment. A substantial proportion of the properties listed are non-owner occupied homes, as owners are located in different states and counties beyond those listings we include in our sample.

We present descriptive statistics for property and individual characteristics for our subject pools and representative samples in Table A.1. Statistics for field experiment subjects are presented in column (1). Property characteristics were obtained from online listings. Data on individual characteristics were obtained by merging data on homeowners and addresses with complementary data provided by a private vendor. This complementary data is based on voter lists and similar sources of data. The vendor (Aristotle International) was able to match most of the sample based on their full names and addresses, however there is missing data depending on the variable used.[41] To compare our sample with a representative sample of U.S. respondents, in column (3) we use data from the 2013-2017 ACS (Ruggles et al., 2024), where we restrict the sample to household heads, who are homeowners and reside

[41] For example, the share of missing data for gender was 25.6%, age 30.2%, education 66.3%, ethnicity 22.4%, income 22.4%.

in one of the counties where the field experiment properties are located. Note that even though we try to restrict the ACS to the same counties, not all counties in our experiment are sampled in these years in the ACS, so this is ultimately a subset of counties from the field experiment. Additionally, the variables definitions for the statistics in our field experiment and those in the ACS are not necessarily the same.[42] For instance, home value represents the Zillow Home Value Index for properties in the ZIP code for our field experiment, while it is the self-reported home value for the ACS. An additional caveat to consider when comparing samples is that homeowners in the experimental sample might not live in the same county as the property.

Despite these caveats, we find that the properties in field experiment sample are fairly representative of the counties where they are located. Homes in the field experiment are valued at $518,000 on average based on Zillow Home Value Index in the subject's ZIP code, while the self-reported property value in the ACS is $451,000. Listed properties in the field experiment on average have 3.24 bedrooms, compared to 3.10 in the ACS. The largest difference we observe is in the age of the property, where experiment properties are 10 years newer on average. This could partially explain the larger home values observed in the field experiment.

Homeowners in our field experiment population differs slightly in terms of individual characteristics. For instance, on average homeowners in our sample are 58.7 years old, while those in the ACS are 56.9. Homeowners in the field experiment are less likely to be female (32.9% vs 47.1%). Household incomes on average is $128,000 compared to $110,000 in the ACS. Again, note that these comparisons should be taken with a grain of salt because of different data definitions, and the large share of missing data on individual characteristics for our field experiment sample.

## A.5 Additional Details about the Information Sources

Based on data from recent years, all of these information sources were somewhat informative, in terms of their ex-ante predictive power. Figure A.5 presents a comparison of forecast model performance. Specifically, we generated an out-of-sample prediction of home price growth for the years 2015-2020 for each of the five treatments we use as information sources. We then compare the out-of-sample prediction to the actual home price growth in that year, and calculate the Root Mean Squared Error (RMSE). As a benchmark, we also show the RMSE for Zillow's official forecast. This figure shows that, on average, all of the information sources

[42] Because the ACS employs multiple years, in order to make self-reported home values comparable with current listing prices, we deflated reported values using the CPI and local property value growth based on the Zillow home price index.

we use are roughly equally informative. Moreover, the information sources we provided do about as well as Zillow's official forecasts.

Next, we provide additional information on the variation in signals. Even though the signals from different sources are correlated with each other, they are not perfectly so. The results are presented in Table A.2. Columns (1) and (2) present the means and standard deviations for each of the information treatments. On average, Past-2 is the most optimistic (5.2%) while Forecast-1 is the most pessimistic (2.6%). Columns (3) to (7) show the correlations between the different information treatments. The correlations range from 0.894 (Past-1 and Forecast-1) to 0.295 (Past-2 and Forecast-2). Even though the signals from different sources are highly correlated, there is substantial heterogeneity in signals, as shown in Figure A.4. Panels (a) to (c) show substantial heterogeneity within information sources but across markets. In panels (d) to (f) we show the heterogeneity across information sources. For example, panel (d) plots Past-1 versus Past-2, and the size of each dot represents the number of observations for that pair. Note that even though the slope coefficient of the regression is 0.852, the $R^2$ suggests that 66% of the variation in Past-2 is explained by variation in Past-1 signals; meaning that 34% of the variation remains unexplained between signals. We find similar levels of heterogeneity between signals in panels (e) and (f).

## A.6 Further Information about the AMT Sample

This section presents more details about the sample from the online survey experiment conducted on U.S. respondents recruited through AMT. We followed several best practices for recruiting participants in online surveys and experiments. We restricted the task only to participants who reside in the United States and with a HIT rating greater than 80%. When accepting the task, participants were re-directed to a Qualtrics survey. The median time to complete the survey was 5 minutes 21 seconds. Furthermore, we included an attention check question in the survey that was passed by 99.15% of respondents. We also asked respondents about the clarity of the survey and 92.17% of respondents answered the survey was "easy to understand", while only 0.64% thought it was "difficult to understand". After successful completion of the survey, participants were given a code to redeem their payment of $0.75 for completing the task.

The survey was distributed between June 20 and June 24 of 2019, approximately during the time same the first letters from the field experiment were being delivered. We obtained 2,251 responses. Subjects were assigned to each of the six treatment groups using equal probabilities. We took a number of steps to guarantee the integrity and quality of responses. First, we dropped respondents who entered an "invalid" ZIP code (670 respondents).[43] Second, we further dropped respondents who did not complete the entire survey (16 respondents). Third, we dropped respondents for whom the survey instrument did not record a treatment status, likely because of javascript incompatibility with the respondent's browser (23 respondents).

Finally, we dropped respondents that were assigned information treatments that were extreme outliers (35 respondents) and respondents who had extreme outliers for prior beliefs (less than -15% or greater than 25%; 103 respondents). These extreme beliefs may be the product of typos or lack of attention. As the prior belief was reported before the treatments, dropping these extreme prior beliefs should not contaminate the experimental analysis. Typos may also be present in posterior beliefs, but dropping individuals based on post-treatment outcomes could contaminate the experimental analysis. Instead, following Fuster et al. (2022), we winsorize the post-treatment outcomes using the same extreme values presented above (-15% and 25%). In any case, we use graphical analysis whenever possible to certify that the results are not driven by outliers. This yielded our final sample of 1,404 respondents.

Descriptive statistics for the AMT sample are provided in Table A.1. Using median home values for their corresponding zip code, the average Zillow Home Value Index in the respondent's ZIP code for a respondent in our AMT survey was $263,000 – about half that of our field experiment, and less than the U.S. average based on the ACS (column (4)). However, in relation to individual characteristics, our AMT sample more closely resembles the general sample of U.S. homeowner household heads . Though the sample is younger on average (39.2 vs 55.8 years), about half are female, and 80.2% are white (vs 83.5%), 9.05% black (vs 7.99%), and 4.4% are Hispanic (vs 9.2%).

The randomization was successful as shown in Table A.3. The different information treatment arms are well balanced across various covariates, where we do not find any significant differences using an F-test where under the null all averages are equal and the alternative hypothesis is at least one is different (column (8) presents the p-values for the test). On average, respondent's prior home price expectations were 3.877%, with a standard deviation of 0.144. The online survey included a brief follow-up module, where subjects were invited to participate through AMT a month after the baseline survey. We received 995 responses to the follow-up from July 23 to August 6, for a response rate of 71%.

[43] This is either because the ZIP code field was missing, or one that did not match our data set (needed to provide information in the survey experiment).

## A.7 Estimating the Reading Rate

Our main estimates capture an intent-to-treat effect of the information shock on the probability that the home is sold. Besides people not fully updating their beliefs from the information provided, there are many reasons why subjects may not have read the letters sent to them. In this section we provide details for how we estimate letter reading rates.

Our estimate of the reading rate is comprised of three components. The first component is estimate that 5% of standard mail is not delivered by the US Postal Service. The second component is the share of letters delivered too late (i.e., after the property had already been sold), our estimate is described below. In the weeks prior to the start of letter delivery, 0.59% of the properties were already sold. Because that happened before the letters were read, that 0.59% of subjects could only receive the letters when it is too late. During the first week after the letter delivery, 0.98% of the properties were sold. However, only 84.1% of the subjects had received the letters by the end of that week. Thus, we estimate that roughly 0.82% ($= 0.0098 \cdot 0.841$) received the letter too late during the first week. We can continue with this argument for each of the following weeks. We estimate that roughly 7.7% of the letters were read after the property had already been sold. In practice, 7.7% is likely a lower bound, because we are defining too late when the sale is registered in the Multiple Listing Service, but in reality it may be already too late before then. That is, once the contingent offer is accepted, while the buyer can later walk out, it is too late for the seller. Thus, the fraction of letters that arrived too late may be effectively somewhat larger than 7.7%. Thus, if anything, our reading rate is conservative. If it were even smaller under the alternative assumptions, it would lead to an even larger scale-up factor.

The third component is the share of letters that are delivered but not read by the recipient (e.g., thrown away unopened). This is the component for which there is the most uncertainty. Our baseline estimate is based on the 2018 sample (the latest available at the time of the analysis) of the Household Diary Survey. As reported in Figure 5.3 of Mazzone and Rehman (2019), 26% of the advertising mail is not read by the respondents. This estimate has been stable in the previous two rounds of the survey study (25% in 2016 and 24% in 2017). However, there are still some factors why this rate may be misleading.

On the one hand, there are some reasons to believe that the share of people ignoring our letter was lower than 26%. This 26% is based on all mass marketing mailings, which typically include mailers such as shopping catalogs and credit card offers. It is thus plausible that our letters may have picked up the attention of the recipient more that the type of promotion material that they receive regularly. Second, one of the reasons why unsolicited mail is thrown away is presumably because the sender is not trusted by the recipient. Indeed, the 2014 round of the Household Diary Survey provides some evidence in this respect. Table A3-

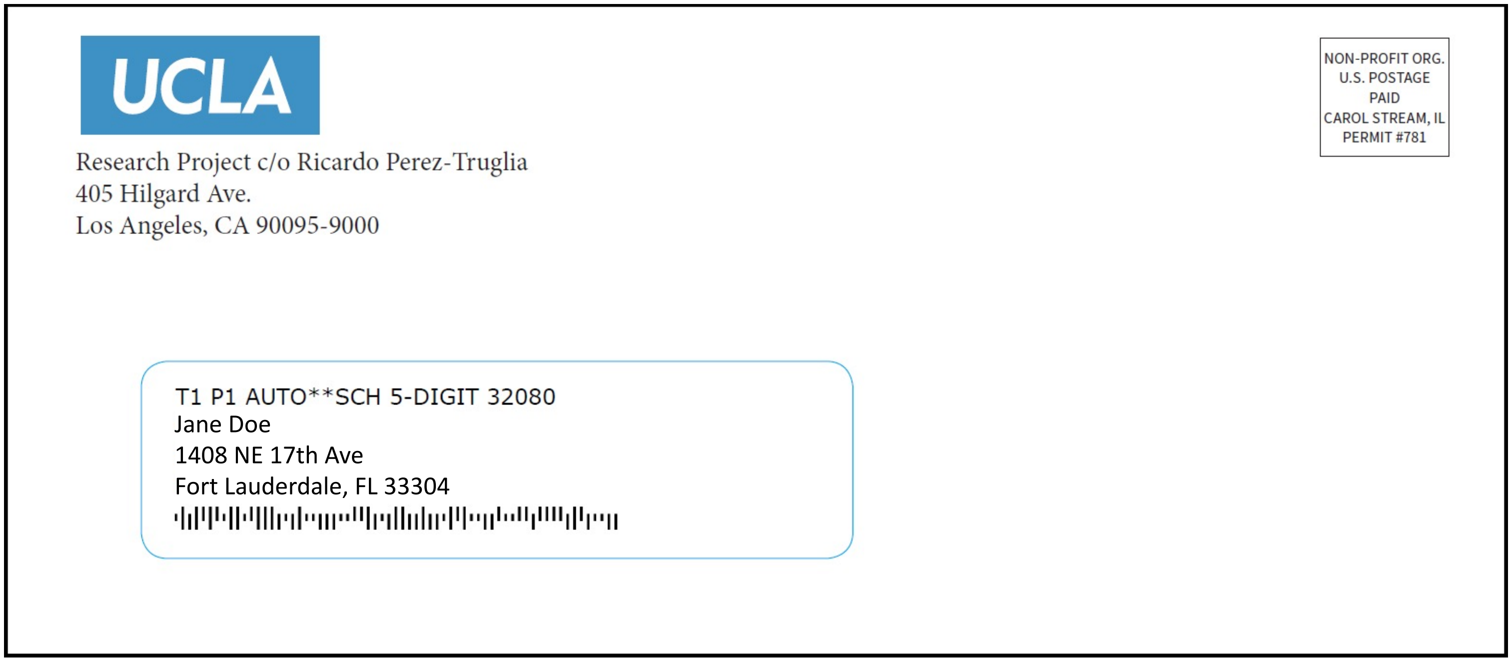
17 of Mazzone and Rehman (2019) shows that even though 27.5% of mailings were discarded for the whole sample, this fraction drops to 17.8% when the sender is an organization that the recipient had been a customer of in the past. Similarly, Table A3-47b of Mazzone and Rehman (2019) shows that the attention paid to the letters is larger for more trustworthy senders like the government or a charity. We took several measures to make our letter look trustworthy, such as addressing the envelope to the recipient using their full name, as well as putting in the outside of the envelope the official UCLA logo and a mention to the research study.

On the other hand, there is one main factor suggesting that the share of people ignoring our letter was actually higher than 26%. In Mazzone and Rehman (2019), the 74% probability of paying attention to the mailing is further broken down in a 49% probability of reading it plus a 23% probability of "scanning" it. If we want to be pessimistic about the reading rate, we could count the people who scanned the letter as not having paid enough attention, thus elevating the estimate of 26% to 49% (=26%+23%). However, there are two reasons why we believe this would be misleading. First, the main reason why the Household Diary Survey includes the alternative "scan" is because a big part of the mail advertising pieces include massive reading material such as catalogs with hundreds of pages and credit card offers with pages of small print. As a result, nobody would ever be expected to read the whole thing. Second, we believe that the fact that some people look at the information but only superficially is already being reflected in the estimates of the effect of the information shocks on the survey expectations. In the online survey experiment, the fact that the pass-through was low is probably to a great extent because some respondents were not paying a lot of attention. As a result, it would be misleading to correct twice for the same non-compliance.

It is also worth mentioning a couple other studies using estimates of reading rates: Perez-Truglia and Cruces (2017) uses an estimate of 21.5% and Gerber et al. (2020) uses an estimate of 50%. Due to the design of our mailing and the contextual information, we are confident the reading rate is much higher in our sample than in those other studies.[44] However, if anything, assuming these alternative estimates would lead to an even larger scale-up factor.

[44] Those studies were conducted under very different circumstances. For example, the mailer in Perez-Truglia and Cruces (2017) was a folded letter instead of a proper envelope, and sent to campaign contributors during a political campaign, while they receive a lot of non-solicited mailing related to the election.

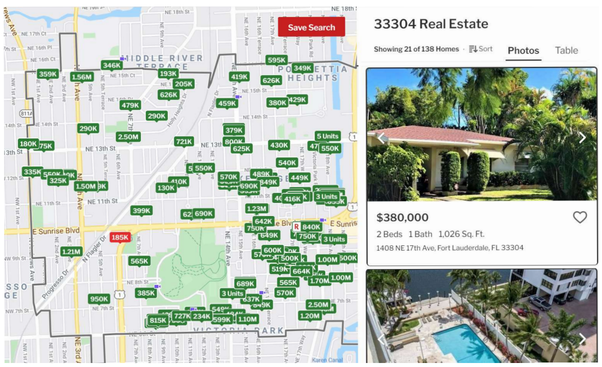
Figure A.1: Sample Envelope



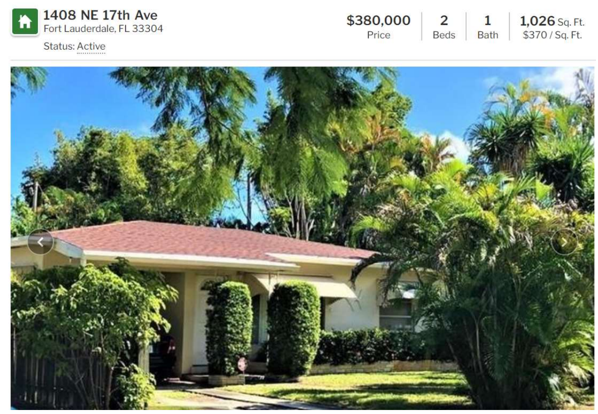

Notes: Screenshot of the outside of the envelope used in the field experiment.

Figure A.2: Screenshots from Sample Data Sources

**a.** Listings Map




**b.** Sample Listing




**c.** Parcel Identifier

**Public Facts for 1408 Northeast 17th Ave**

**Home Facts**

| | | | |
|---|---|---|---|
| Beds | — | Lot Size | 6,750 Sq. Ft. |
| Baths | — | Style | Single Family Residential |
| Finished Sq. Ft. | 1,026 | Year Built | 1951 |
| Unfinished Sq. Ft. | — | Year Renovated | 1968 |
| Total Sq. Ft. | 1,026 | County | Broward County |
| Stories | 1 | APN | 494234019540 |

Home facts updated by county records on Apr 4, 2020.

**d.** Assessor's Office Data

**ASSESSOR'S OFFICE**

Online Services | Dispute Assessment | Assessment Roll Search

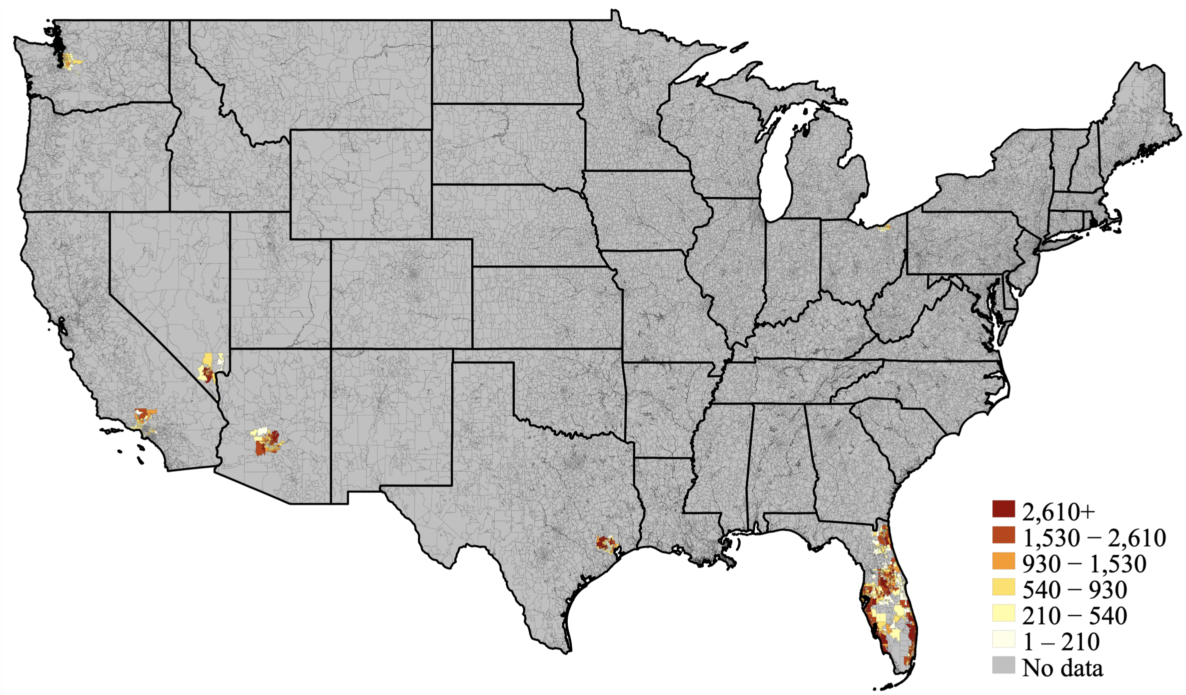
**Assessment Roll Search**

| | |
|---|---|
| Parcel # | 494234019540 |
| Address | 1408 NE 17th Ave, Fort Lauderdale, FL 33104 |
| Owner/s | Jane Doe |
| Mailing Address | 1408 NE 17th Ave, Fort Lauderdale, FL 33104 |
| School District | |
| Status | Active |
| Zoning Code | RE4 |
| Total size | 1,026 |
| Assessed Value | $350,000 |

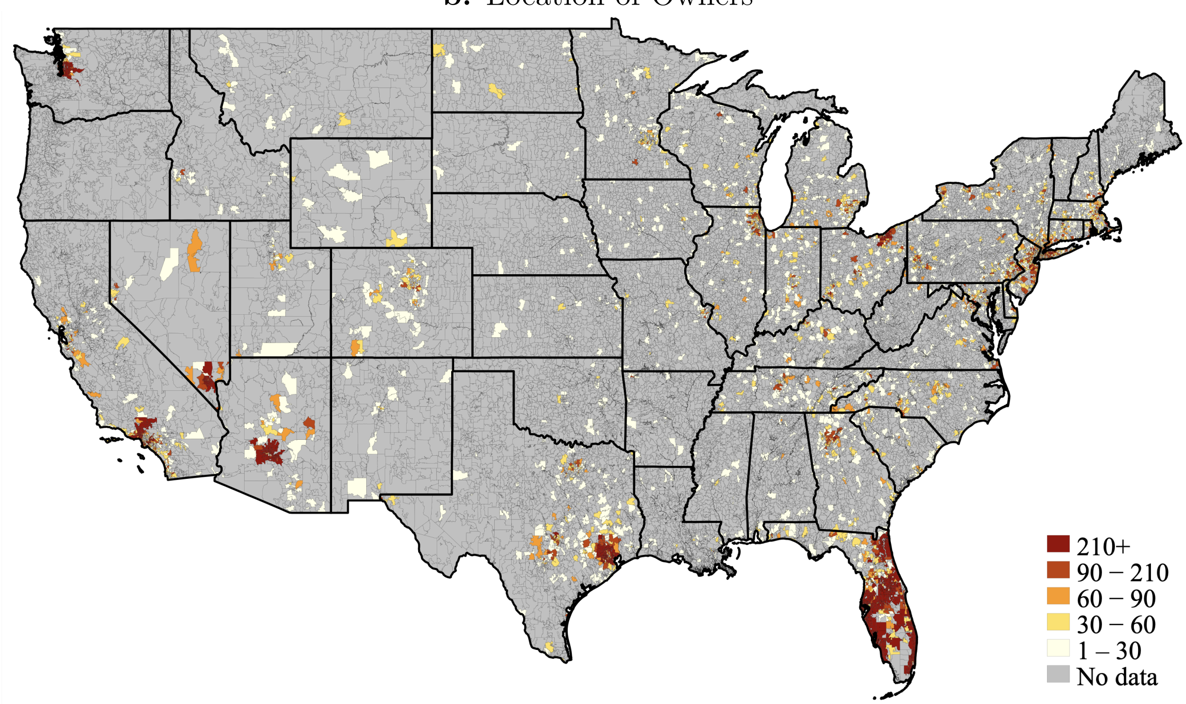
Notes: Panel (a) presents a screenshot of a sample real estate listings map that potential buyers would see on an online marketplace where we scraped the listings data from. Panel (b) presents a screenshot of a sample listing the potential buyers could view. Panel (c) shows a screenshot of the information available for a property on the online marketplace, including the Assessor's Parcel Number (APN), which we used to match the listings data with county assessors' data. Panel (d) shows a screenshot of the information available for each property in the county assessors' database when we search for it using the APN. This screenshot shows a manual search for a single record for illustrative purposes only, as we downloaded county assessors' data in bulk.

Figure A.3: Map of Listings' and Owners' Location

**a.** Location of Listings

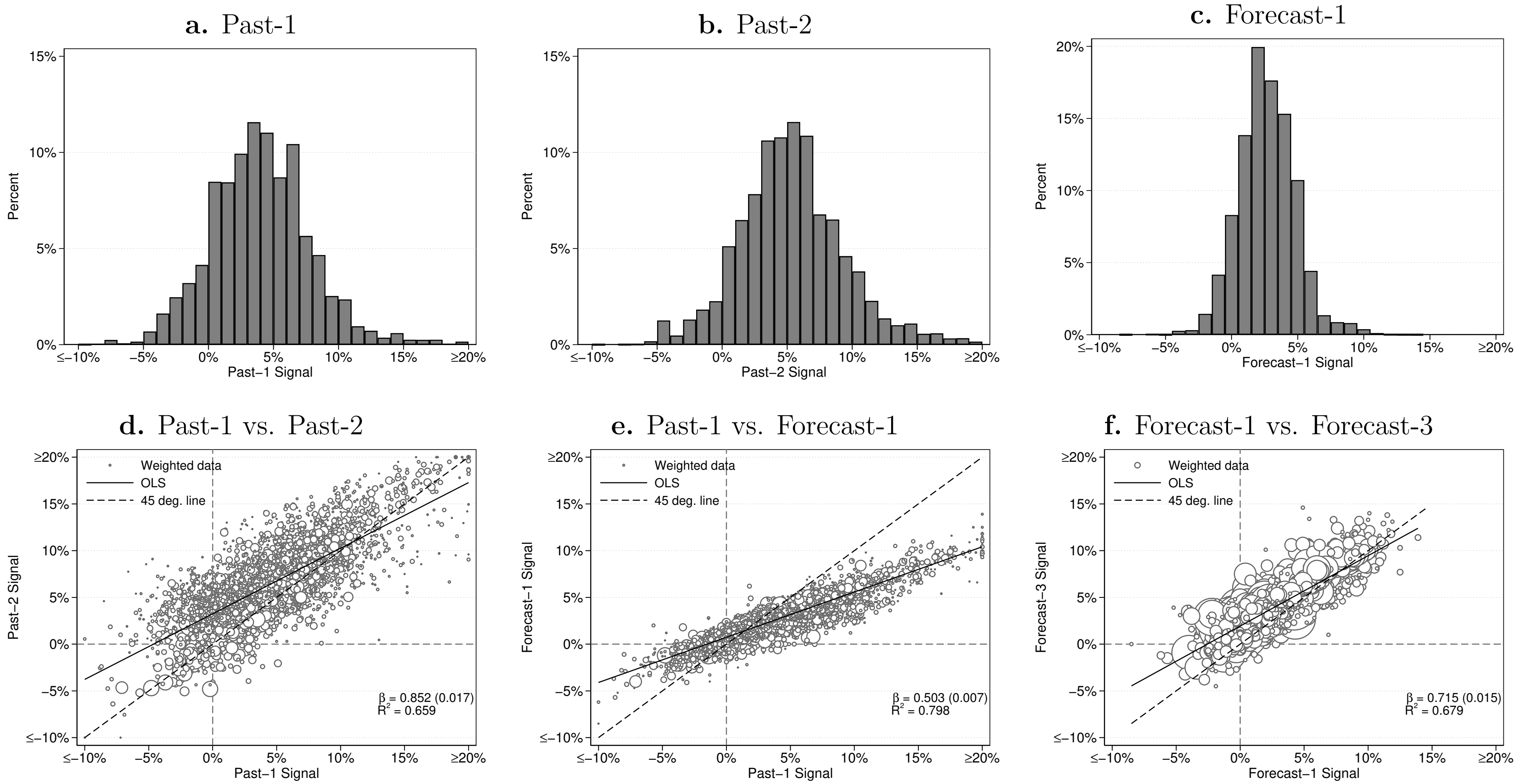



**b.** Location of Owners



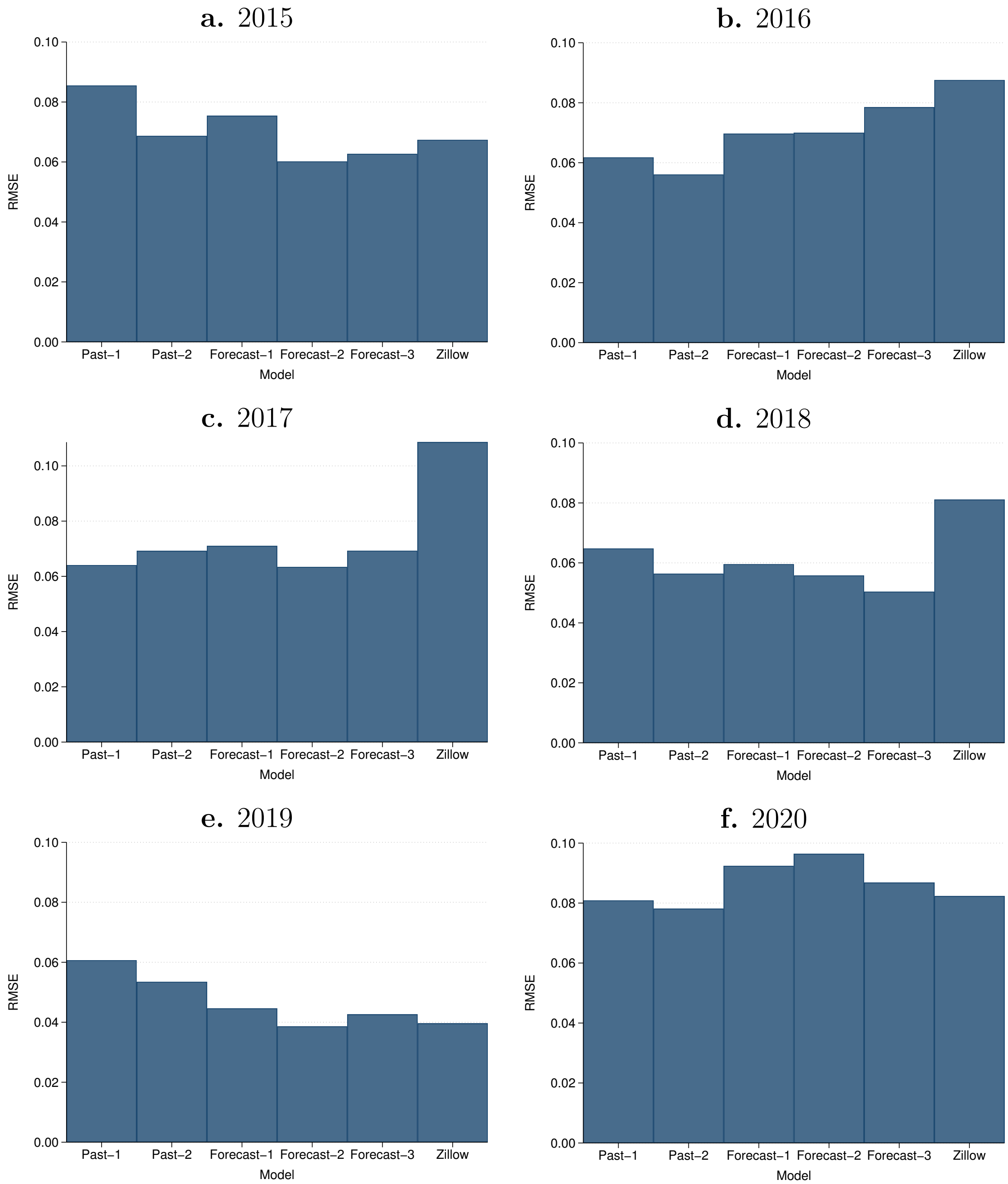

Notes: ZIP Code level maps. Panel (a) shows the locations of the listed properties that were for sale during the experiment. Panel (b) shows the locations of the owners per their mailing addresses.

Figure A.4: Heterogeneity in Signals Within and Between Subjects: Additional Details




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). Panels (b) and (c) show similar distributions for the Past-2 treatment (i.e., the annual growth rate over the past two years) and the Forecast-1 treatment (i.e., the price change forecast over the next year using statistical model 1) respectively. In panels (a), (b), and (c), the bins have a width of 1 pp and are truncated at -10% and +20%. Panel (d) is a scatterplot showing the relationship between the signals that the subjects would have received if they had been assigned to the Past-1 treatment versus the Past-2 treatment. Panel (e) is a similar scatterplot showing the relationship between the signals for the Past-1 treatment versus the Forecast-1 treatment. Panel (f) is a similar scatterplot showing the relationship between the signals for the Forecast-1 treatment versus the Forecast-3 treatment (i.e., the price change forecast over the next year using statistical model 3). In panels (d), (e), and (f), the size of the circles are proportional to the number of observations, and the signals are truncated at -10% and +20%.

Figure A.5: Ex-Ante Predictive Power for Each Information Source




Notes: Each figure represents a comparison of the performance of the sources used to forecast home price growth in a particular year. E.g., panel f shows the results for 2020 home price forecasts, where data prior to April 2019 is used to forecast the home price growth between April 2019 and April 2020. That forecast is then compared to the actual home price growth between April 2019 and April 2020, and the Root Mean Squared Error (RMSE) of the model is plotted. The forecast sources correspond to the sources used in the information-provision experiment. *Past-1* is the annual home price growth rate over the past year. *Past-2* is the annual home price growth rate over the past two years. *Forecast-1*, *Forecast-2*, and *Forecast-3* correspond to the statistical models presented in A.3. *Zillow* is Zillow's own home price forecast. The RMSE is averaged over the forecasts for all zip codes that we provided information for in the field experiment. 90% confidence intervals of RMSEs are small, and thus not shown.

Figure A.6: Sample Methodological Notes

## **a.** Baseline

*<u>Methodological Notes</u>: the median house price for 2019 corresponds to the Zillow Home Value Index: A smoothed, seasonally adjusted measure of the median estimated home value across a given region and housing type. This data is published and made publicly available online by Zillow Research.*

## **b.** Past 1-year

*<u>Methodological Notes</u>: the median house price for 2018-2019 corresponds to the Zillow Home Value Index: A smoothed, seasonally adjusted measure of the median estimated home value across a given region and housing type. This data is published and made publicly available online by Zillow Research.*

## **c.** Past 2-years

*<u>Methodological Notes</u>: the median house price for 2017-2019 corresponds to the Zillow Home Value Index: A smoothed, seasonally adjusted measure of the median estimated home value across a given region and housing type. This data is published and made publicly available online by Zillow Research.*

## **d.** Forecast-1

*<u>Methodological Notes</u>: the median house price for 2019 corresponds to the Zillow Home Value Index: A smoothed, seasonally adjusted measure of the median estimated home value across a given region and housing type. This data is published and made publicly available online by Zillow Research. The median house prices for 2020 correspond to our own forecasts, based on our own statistical models - as such, these forecasts are subject to error. To compute these forecasts, we fit a simple statistical model and then use the estimated parameters to make forecasts. The model uses the lagged values of the annual growth rate of housing prices as explanatory variables. These forecasts are based on ZIP Code-level data on the Zillow Home Value Index for the period 1997-2019.*

## **e.** Forecast-2

*<u>Methodological Notes</u>: the median house price for 2019 corresponds to the Zillow Home Value Index: A smoothed, seasonally adjusted measure of the median estimated home value across a given region and housing type. This data is published and made publicly available online by Zillow Research. The median house prices for 2020 correspond to our own forecasts, based on our own statistical models - as such, these forecasts are subject to error. To compute these forecasts, we fit a simple statistical model and then use the estimated parameters to make forecasts. The model uses the lagged values of the annual growth rate of housing prices (in the same ZIP Code and in the state as a whole) as explanatory variables. These forecasts are based on ZIP Code-level data on the Zillow Home Value Index for the period 1997-2019.*

## **f.** Forecast-3

*<u>Methodological Notes</u>: the median house price for 2019 corresponds to the Zillow Home Value Index: A smoothed, seasonally adjusted measure of the median estimated home value across a given region and housing type. This data is published and made publicly available online by Zillow Research. The median house prices for 2020 correspond to our own forecasts, based on our own statistical models - as such, these forecasts are subject to error. To compute these forecasts, we fit a simple statistical model and then use the estimated parameters to make forecasts. The model uses the lagged values of the annual growth rate of housing prices (in the same ZIP Code and in the city as a whole) and employment as explanatory variables. These forecasts are based on ZIP Code-level data on the Zillow Home Value Index for the period 1997-2019.*

<u>Notes:</u> Each panel of Figure 2 corresponds to the hypothetical table that a given individual would receive under the different scenarios. Each panel of this figure shows the corresponding methodological notes. The notes are placed in the middle of the second page of the letter, in the location of the placeholder *«Information Notes»* from Figure 1.

Table A.1: Descriptive Statistics: Subject Pool Vs. Representative Samples

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | | | Benchmark (ACS) | |
| | Field Exper. | Survey Exper. | Exper. Counties | All States |
| Property Characteristics | | | | |
| Home Value ($1000s) | 518.160 | 262.939 | 451.071 | 340.609 |
| | (3.284) | (4.786) | (0.921) | (0.219) |
| No. Bedrooms | 3.256 | 2.734 | 3.096 | 3.156 |
| | (0.005) | (0.027) | (0.001) | (0.000) |
| Year Built | 1989.920 | — | 1979.141 | 1975.924 |
| | (0.084) | (—) | (0.031) | (0.011) |
| Individual Characteristics | | | | |
| Age | 58.675 | 39.202 | 56.922 | 55.844 |
| | (0.080) | (0.329) | (0.024) | (0.008) |
| % Female | 31.890 | 54.915 | 47.050 | 47.335 |
| | (0.219) | (1.328) | (0.075) | (0.024) |
| % White | 62.872 | 80.271 | 77.710 | 83.517 |
| | (0.210) | (1.062) | (0.063) | (0.018) |
| % African-American | 3.203 | 9.046 | 8.404 | 7.987 |
| | (0.076) | (0.766) | (0.042) | (0.013) |
| % Hispanic | 13.556 | 4.416 | 17.906 | 9.218 |
| | (0.149) | (0.548) | (0.058) | (0.014) |
| % Homeowner | 100 | 56.838 | 100 | 100 |
| | (—) | (1.322) | (—) | (—) |
| Income ($1000s) | 128.059 | — | 109.559 | 104.448 |
| | (0.336) | (—) | (0.171) | (0.050) |
| % Higher Ed | 40.689 | — | 64.502 | 61.138 |
| | (0.276) | (—) | (0.072) | (0.024) |
| Observations | 57,910 | 1,404 | 24,476,728 | 75,292,592 |

Notes: Column (1) reports average characteristics of the subjects in the field experiment. Column (2) corresponds to the AMT survey respondents. Columns (3) and (4) provide benchmarks for the average characteristics based on samples of homeowner household heads from the American Community Survey (ACS) employing population weights. Column (3) restricts the sample to respondents in the counties where the field experiment was conducted, while column (4) includes all the United States. *Home Value* corresponds to the Zillow Home Value Index in the subject's ZIP code as of January of 2019 (columns (1) and (2)) or the self-reported value of the housing unit (columns (3) and (4)). In column (1), the individual characteristics are based on data provided by a private vendor.

Table A.2: Descriptive Statistics: Treatment Means, Standard Deviations and Correlations

| | (1) Mean | (2) S.D. | Correlation (3) Past-1 | (4) Past-2 | (5) Forecast-1 | (6) Forecast-2 | (7) Forecast-3 |
|---|---|---|---|---|---|---|---|
| Past-1 | 3.936 | 3.836 | 1.000 | 0.812 | 0.894 | 0.431 | 0.839 |
| Past-2 | 5.250 | 4.026 | 0.812 | 1.000 | 0.735 | 0.295 | 0.674 |
| Forecast-1 | 2.599 | 2.150 | 0.894 | 0.735 | 1.000 | 0.604 | 0.823 |
| Forecast-2 | 3.328 | 1.180 | 0.431 | 0.295 | 0.604 | 1.000 | 0.569 |
| Forecast-3 | 4.092 | 1.866 | 0.839 | 0.674 | 0.823 | 0.569 | 1.000 |

Notes: This table presents the means, standard deviations, and correlations between the treatments. *Past-1* represents the annual home price growth rate over the past year. *Past-2* represents the annual home price growth rate over the past two years. *Forecast-1*, *Forecast-2*, and *Forecast-3* correspond to the statistical models presented in A.3.

Table A.3: AMT Survey: Randomization Balance Test

| | | By Treatment Group | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| | All | Baseline | Past-1 | Past-2 | Forecast-1 | Forecast-2 | Forecast-3 | P-value |
| Property Characteristics | | | | | | | | |
| Home Value ($1000s) | 262.939 | 270.400 | 264.658 | 256.627 | 271.064 | 250.612 | 264.080 | 0.775 |
| | (4.786) | (11.745) | (12.385) | (11.431) | (11.080) | (11.085) | (12.569) | |
| No. Bedrooms | 2.734 | 2.740 | 2.697 | 2.657 | 2.774 | 2.711 | 2.819 | 0.530 |
| | (0.027) | (0.063) | (0.066) | (0.062) | (0.069) | (0.066) | (0.064) | |
| Prior home expectations (%) | 3.877 | 3.951 | 3.923 | 4.023 | 3.788 | 3.692 | 3.883 | 0.990 |
| | (0.144) | (0.339) | (0.356) | (0.363) | (0.344) | (0.363) | (0.353) | |
| Individual Characteristics | | | | | | | | |
| % Owner | 56.838 | 55.319 | 53.680 | 59.657 | 56.170 | 53.879 | 62.185 | 0.343 |
| | (1.322) | (3.250) | (3.288) | (3.221) | (3.244) | (3.280) | (3.150) | |
| Age | 39.202 | 39.130 | 39.041 | 39.768 | 39.551 | 38.987 | 38.737 | 0.957 |
| | (0.329) | (0.810) | (0.770) | (0.862) | (0.820) | (0.748) | (0.818) | |
| % Female | 54.915 | 55.319 | 52.381 | 53.648 | 52.340 | 58.190 | 57.563 | 0.688 |
| | (1.328) | (3.250) | (3.293) | (3.274) | (3.265) | (3.245) | (3.210) | |
| % White | 80.271 | 77.872 | 79.221 | 82.403 | 79.574 | 80.172 | 82.353 | 0.787 |
| | (1.062) | (2.714) | (2.675) | (2.500) | (2.636) | (2.623) | (2.476) | |
| % African-American | 9.046 | 9.787 | 6.494 | 8.584 | 9.787 | 10.776 | 8.824 | 0.626 |
| | (0.766) | (1.942) | (1.625) | (1.839) | (1.942) | (2.040) | (1.842) | |
| % Hispanic | 4.416 | 3.404 | 6.494 | 4.292 | 3.830 | 5.172 | 3.361 | 0.625 |
| | (0.548) | (1.185) | (1.625) | (1.331) | (1.255) | (1.457) | (1.171) | |
| Observations | 1,404 | 235 | 231 | 233 | 235 | 232 | 238 | |

Notes: Average characteristics for the 1,404 AMT survey respondents, 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. *Home Value* is the 2019 Zillow Home Value Index. *No. Bedrooms* is the property's number of bedrooms. *Prior home expectations* is the prior home price expectations elicited in the AMT survey (i.e., before the information-provision stage).

# B Robustness Checks

## B.1 Estimating Duration Models

In this section we further explore the robustness of our main results in our field experiment by employing duration models. One benefit to using duration models in our context is that they can deal with truncation of our variable of interest. As this section shows, the results are both qualitatively and quantitatively similar to those presented in the paper.

We first provide graphical evidence in Figure B.1, where we show results of a proportional hazard models using different functional forms: Cox, Exponential, Wiebull and Lognormal. The figures compare properties that did not receive an information shock in blue (i.e., information shock equals 0%) to properties that were exposed to an information shock of 10% in red. Notably, the functional form chosen does not significantly affect the pattern of results, where the probability that a property remains on the market for properties exposed to an information shock of 10% is higher than receiving no information shock almost throughout the entire period – intensifying during the first 100 days and then remaining stable thereafter.

We formalize the graphical evidence in a regression table presented in Table B.1 using the same functional forms. The first row of the table presents the raw coefficient on information shock. The second row presents the marginal effect of increasing the information shock by 1 pp on the number of days to sell a property. For example, the estimates from column (2) suggest that a 10 pp increase in the information shock increases the number of days to sell a property by 18.09 days (about 10.9% of the mean number of days to sell).

## B.2 Effects on Listing Status

This section explores the effect of the information shocks on property's listing status to gain further insight into the mechanism driving our main effects on whether and when a property sells. In particular, we observe whether the property listing is sold, pending (i.e. sale offer has been accepted, but is pending the final close), active, or delisted each week. In Figure B.3 shows the share of properties that fall into each of those categories each week. Our main effects on whether a property sells can be the result of homeowners updating their reservation price and presumably holding onto the property waiting for a better offer. In a more extreme case, homeowners may decide to withdraw the listing altogether (i.e., delist).

Figure B.2 presents the event-study analysis that evaluates the impact of the information shock on each of the four secondary outcomes. Figure B.2.a is identical to the top half of Figure 6.a, which is the event-study analysis of the information intervention on the probability of selling the property. The information shock has no statistically significant effect on the probability of a property being pending or delisted (Figure B.2.b and Figure B.2.d respectively). However, Figure B.2.c shows a positive effect of the information shock on the probability of a property listing being active, which mirrors the negative effect on the probability of the property being sold in Figure B.2.a. This suggests that subjects reacted to the information shock primarily by keeping the properties on the market (as opposed to taking them off the market), presumably waiting for higher bids.

## B.3 Effects on Listing Price and Sales Price

In section 5.3 we present the effects of the information shock on average listing price changes and document that higher information shocks lead a higher average listing price. In this section we present a semi-parametric version of this analysis where we identify which margin is driving the average listing price change. For instance, the information shock could increase the likelihood that listing prices are raised. Alternatively, a higher information shock could decrease the probability of a small or large reduction in listing price. Indeed, on average listing prices decrease over time, increases are rare.

Figure B.4 shows the evolution of listing price over time for the properties in our sample. The majority of the listings left their listing price unchanged or decreased it over the course of the experiment. Figure B.6 shows the impact of the information shock on the probability of changing listing price. There is a small positive effect on the probability of a price increase. However, the proportion of properties that increased their listing price over time is very small overall (Figure B.4). The information shock also has a negative effect on the probability of a large (>5%) price decrease in the first half of the experiment. This suggests that the information shock caused owners to not decrease their listing price as sharply. The listing price variable used in this figure is not updated to reflect the final selling price if that selling price is different to the listing price. We also performed an alternative robustness check where the listing price variable is updated to incorporate the selling price, which produces very similar results, and is hence not reported here.

We can also estimate the effect of information shocks on sales prices. As mentioned in Section 5.3 (and as anticipated in the pre-registration), studying the effect on sales prices entails an identification challenge: by construction, the sale price is only observed if the property is sold, which was not the case for 42% of the properties (i.e., those that were not sold within 28 weeks of the start of the mail delivery). As a result, and despite the random assignment, estimating the effects on the censored variable is subject to potentially severe selection bias. Moreover, the effect on the sales price depends not only on the ex-ante intention of the seller (i.e., the reservation price), but most importantly on what happens ex-post in the market. Assume, for example, that many people decided to wait longer because they expected prices to go up, but then a crisis hit and the home prices plummet. As a result, the higher expectations will result in lower sales prices ex post. However, that will be mostly the product of the crisis, not the expectations.

To estimate the effect of the information treatment on sales prices, we need to make an additional assumption regarding the unsold properties. We impute sales prices for unsold properties using the corresponding Redfin estimate for each property. This is motivated by the fact that, among sold properties, the final sales price is very highly correlated with the Redfin estimate (r-squared=0.982) as shown in Figure B.5. There are 3,916 listings (6.7%) that are unsold and do not have an estimate available. For these we impute using the final listing price recorded in our data. Finally, we winzorize sales prices by +−25% of the original listing price (1,135 observations, or 1.9% of the sample).[45] Results are presented in Table B.2, where column (1) presents as a benchmark our main estimates of the effect of the information shock on the probability that the property is sold by week 29 after letter delivery. Column (2) presents the effect of the information shock on the logarithm of sales price. The positive coefficient (0.056) is directionally consistent with the effects documented on the effects on listing prices. Indeed, the effects on the sales price seem to be quantitatively stronger than the effects on listing price: e.g., a 10% higher information shock increases the average listing price by 0.12 pp (p-value=0.096). However, this result has to be taken with a grain of salt: the coefficient is imprecisely estimated and thus statistically insignificant at conventional levels (p-value=0.568).

## B.4 Next Purchase and Mail Forwarding Data

In this section we explore the extent to which owners purchase another property in the same neighborhood. We take each subject in the experiment and use the county assessor's data to figure out if they bought another property in the same county where the listed property was located. Because there are no unique identifiers in the county assessor's data, such as social security numbers, we need to use fuzzy matching methods based on full names, which is extremely time consuming. We focus this part of the analysis on the counties from the state of Florida, which constitute a majority (68%) of the subjects in the field experiment and for which the data is particularly easier to work with. However, we suspect the results should look similar for the other counties.

The results are presented in Table B.3. We find that the majority of subjects did not buy new properties in the same county that their listing was located in. For example, out of all the subjects who sold their properties within 28 weeks of the start of the letter delivery, only 8%

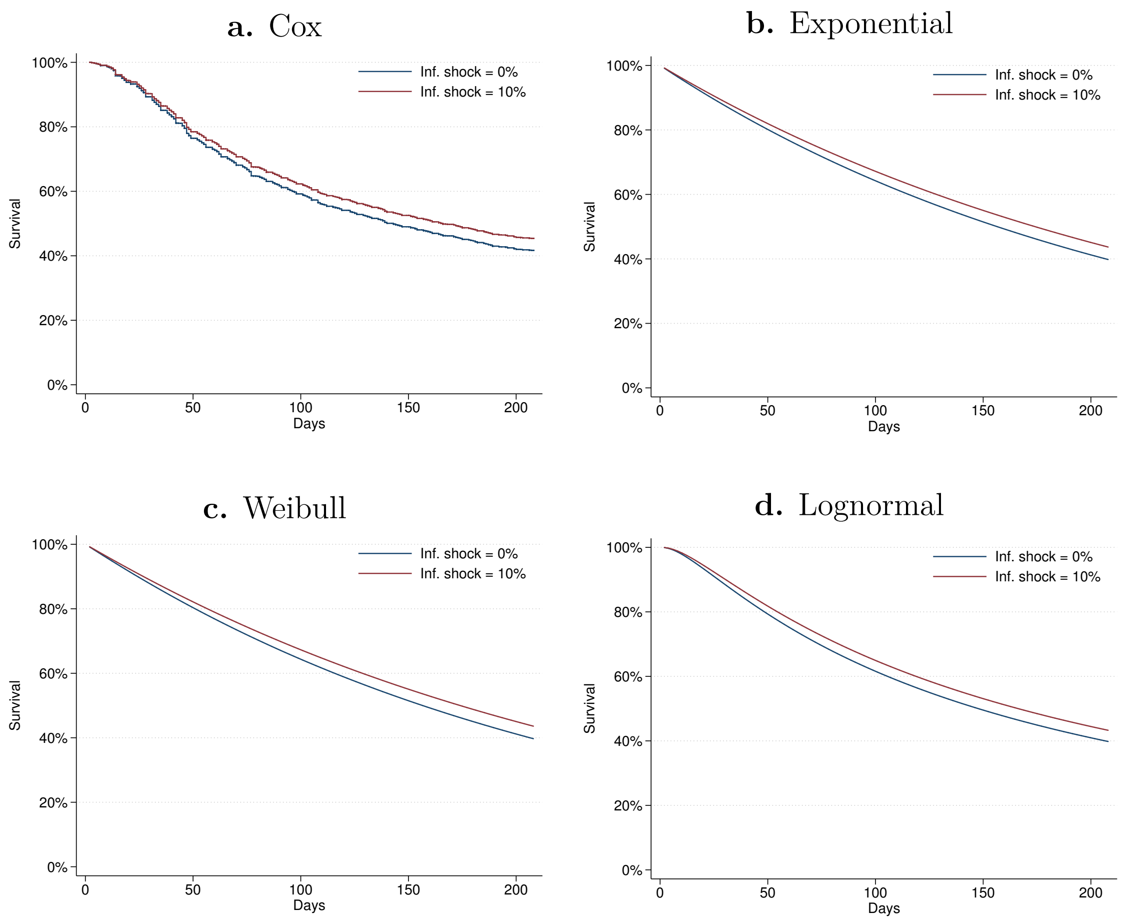
[45] Results are not sensitive to width of this window.

of them bought another property in the county over the same timeframe. We should take this estimate with a grain of salt. For example, the true share may bit somewhat higher because some of these subjects may buy a property a bit later. However, this estimate provides a rough sense of the order of magnitude. Indeed, the above estimate is roughly consistent in order of magnitude the estimate from an independent source: according to Zillow (2019), only 15% of homeowners bought their next home in the same neighborhood.

## B.5 Geographic Proximity

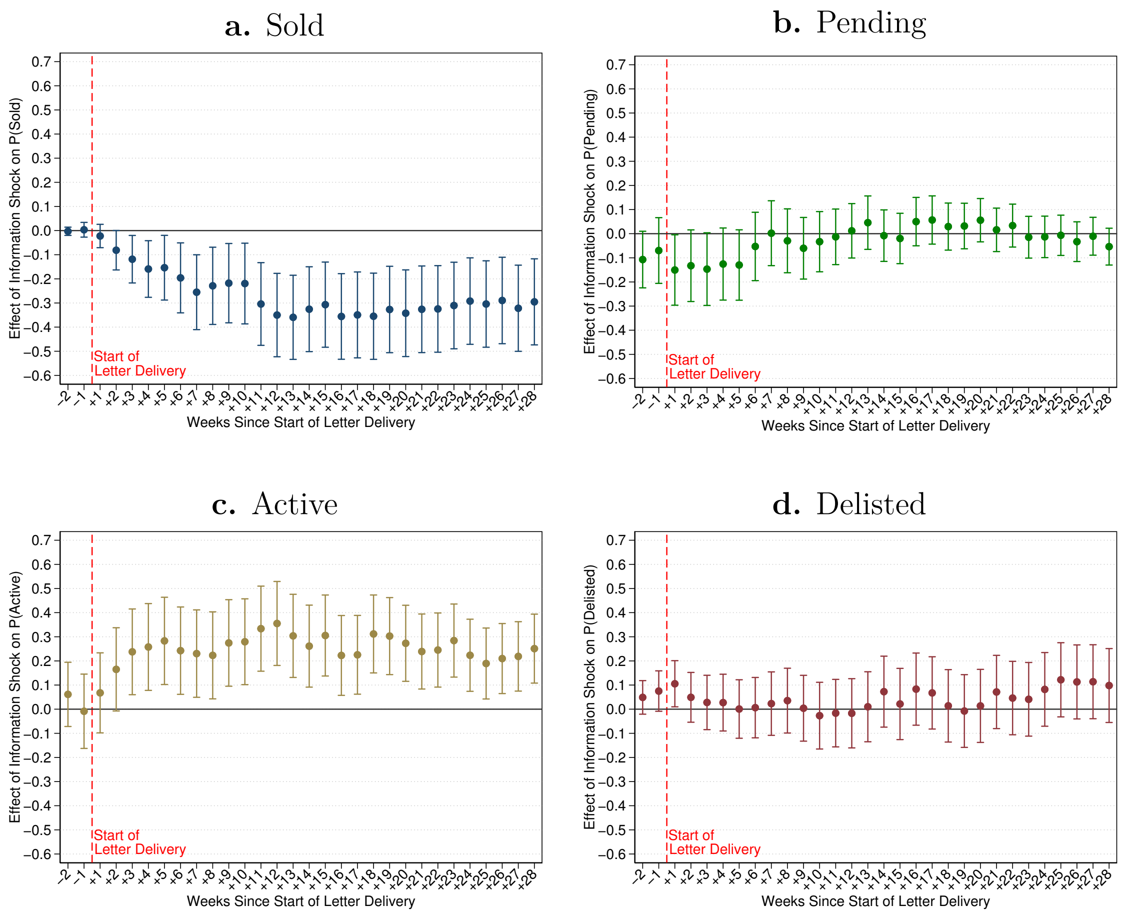
Figure B.7 shows the distribution of distances from each subject to their nearest fellow subject. This distance is computed using mailing addresses so that we can observe the distances between subjects rather than the distances between properties, as non-occupant owners do not live in their listed properties, and we are interested in the likelihood of subjects interacting to share information. The majority of subjects in the experiment live within 1 mile of another subject. However, with the exception of large high-rise buildings with multiple apartments up for sale, this distance is almost never zero - properties close to each other may be located on nearby streets. Moreover, it is highly unlikely that the subjects randomly know each other. Furthermore, the type of information shared is tailored not only to the ZIP Code, but also to the number of bedrooms of the subject's property, which renders sharing less straightforward. This figure supports our view that the potential attenuation bias from spillovers in the information treatments is likely negligible in our experiment.

Figure B.1: Survival Models



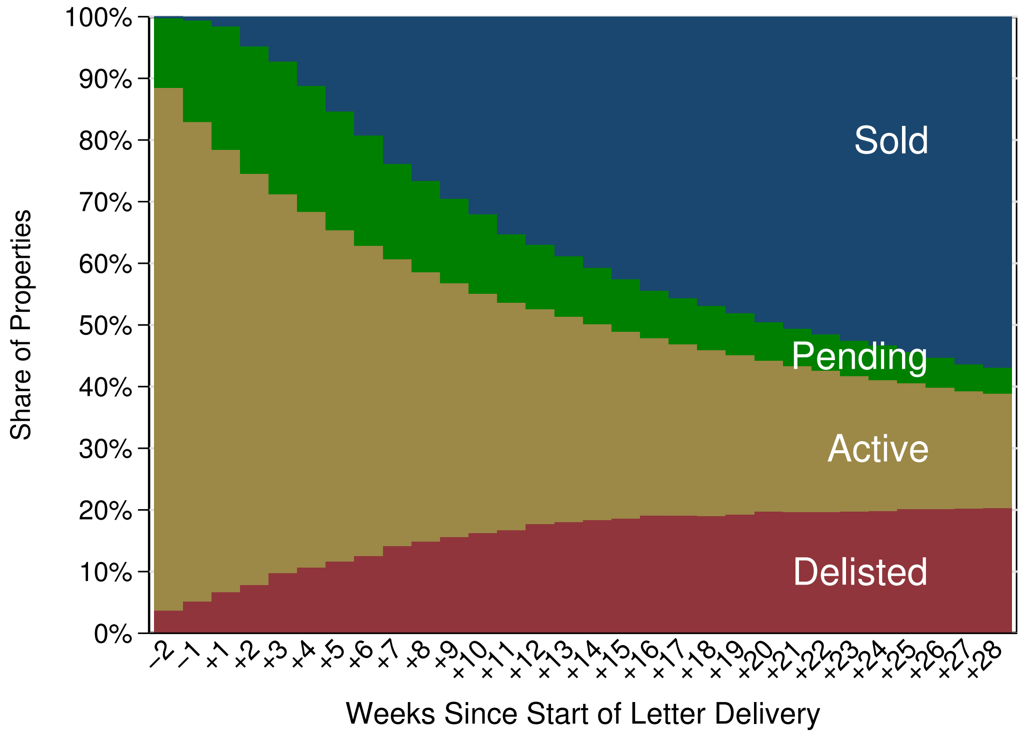

Notes: Each panel shows the results of a proportional hazards model with a particular functional form. In each panel, *Survival* represents the property remaining on the market (ie. not selling the property). The blue line shows the results for participants who did not experience an information shock (or experienced an information shock of 0%), while the red line shows the results for subjects who received an information shock of 10%. The x-axis indicates the days since the start of letter delivery.

Figure B.2: Effects on Behavior: Non-Sold Sub-Categories



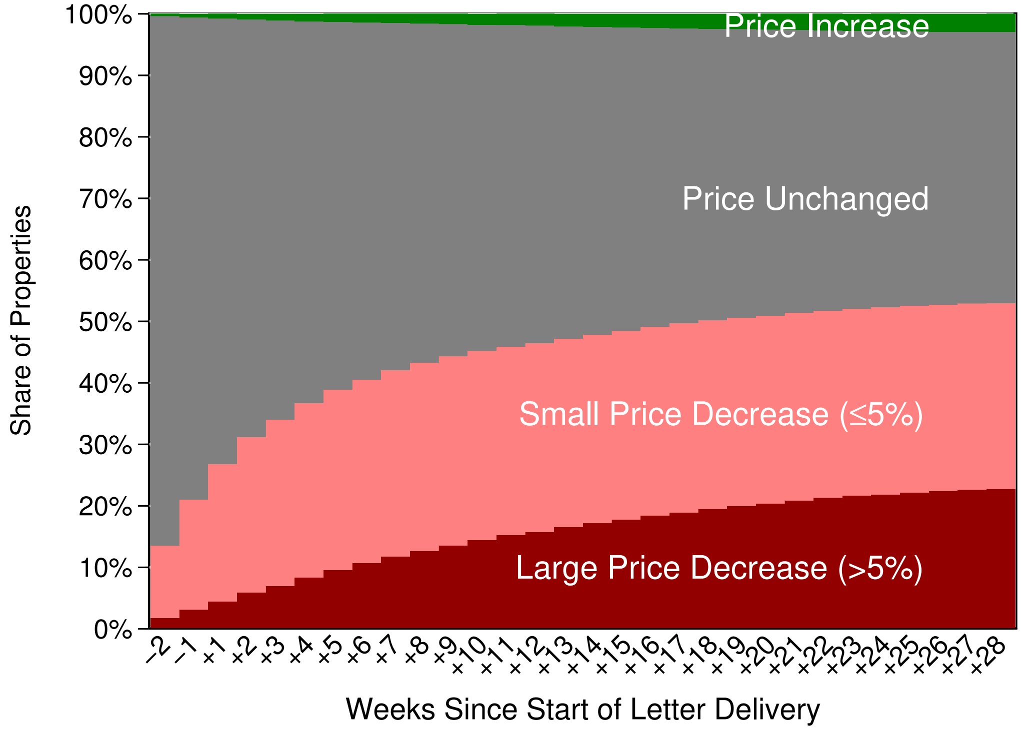

Notes: Each panel represents an event-study tracking the effect of the information shock of the experiment for properties falling into different outcome categories. Panel (a) presents the results for sold properties. Panel (b) presents the results for properties marked as pending (i.e., sale is close to a done deal but pending the final close). Panel (c) presents the results for listed properties. Panel (d) presents the results for delisted properties (i.e., the home has not been recorded as sold, but has been taken off the market). Within each panel, each coefficient corresponds to a separate regression based on 57,910 subjects from the field experiment. Every regression is based on the equation (4) from Section 2, and the coefficient being graphed corresponds to the coefficient on the key independent variable, *Information Shock* ($E_m^{j_i^*} \cdot D_i$). 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 is in a particular outcome category at a number of weeks after the start of the letter delivery and 0 otherwise. For example, the coefficient on *+12 weeks* in panel (a) 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 vertical red line indicates the estimated date when the first letter was delivered (June 15 2019). The 90% confidence intervals are based on heteroskedasticity-robust standard errors.

Figure B.3: Evolution of Sales Outcome: Listing Status



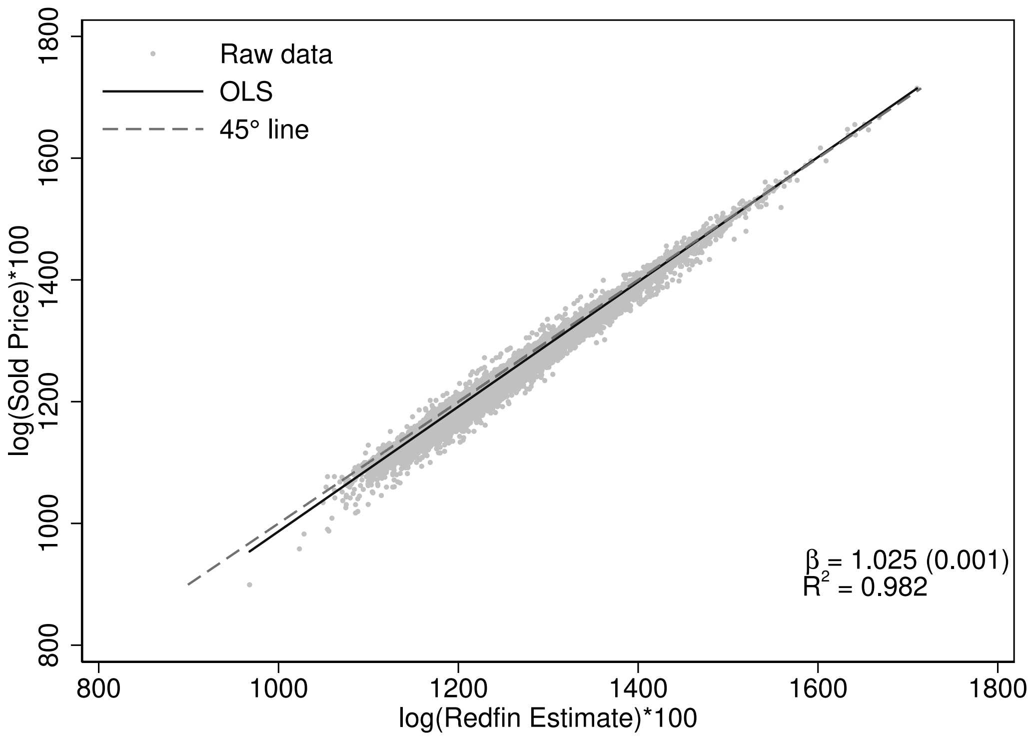

Notes: The x-axis represents the number of weeks elapsed since the start of letter delivery (June 15 2019). For each week, the stacked bars shows the proportion of properties in the subject pool (out of a total of 57,910) that fell in each of the four possible statuses: delisted, active, pending or sold. The bars labeled *Sold* represent the proportion of properties that have been sold. The bars labeled *Pending* represent the proportion of properties that were marked as sale pending (i.e., the offer has been accepted pending final close). The bars labeled *Active* represent the proportion of properties that were still active. The bars labeled *Delisted* represent the proportion of properties that were delisted (i.e., the listing had been taken down according to our scrapped data from the real estate website but no sale had been recorded).

Figure B.4: Evolution of Listing Prices



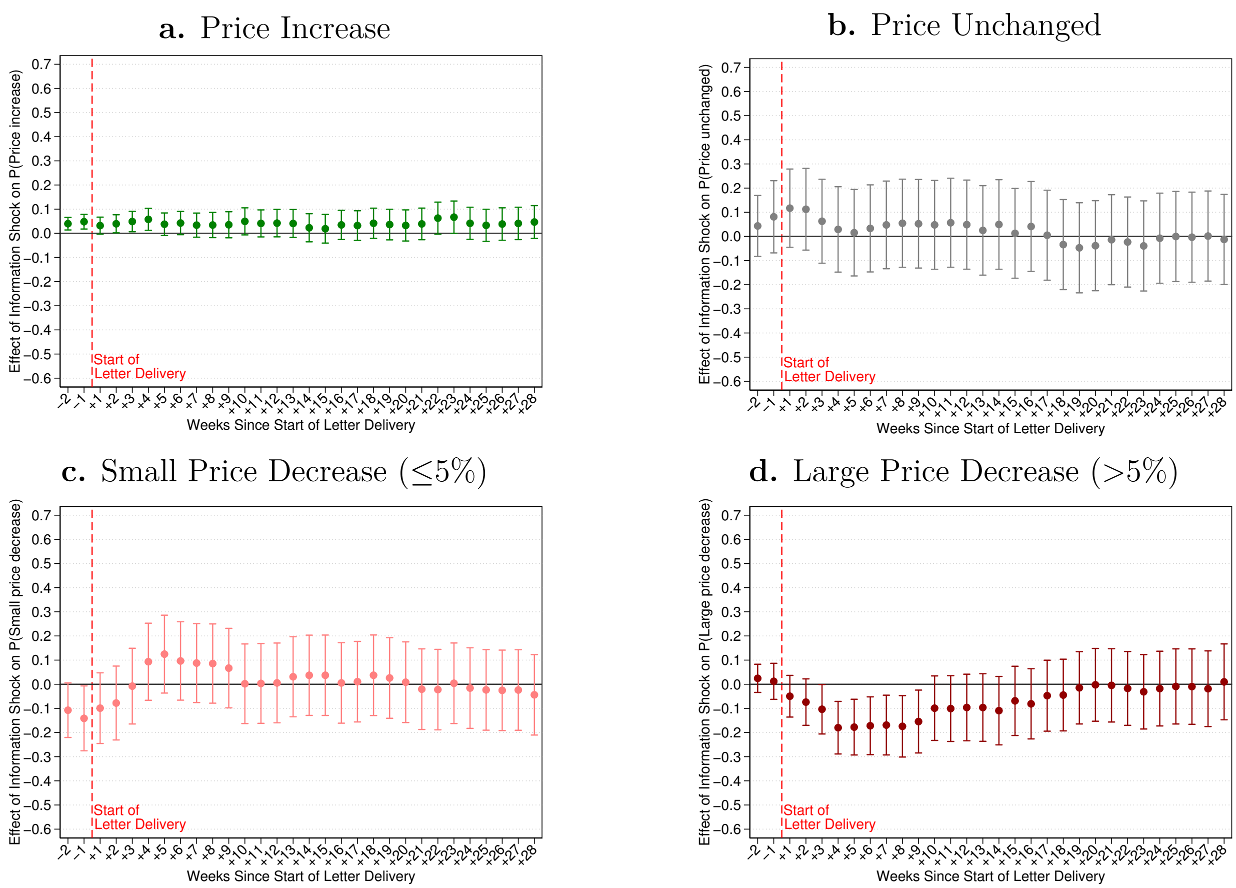

Notes: The x-axis represents the number of weeks elapsed since the start of letter delivery (June 15 2019). Each bar shows the proportion of properties (out of a total of 57,910 properties) that saw a particular change in listing price (relative to the original listing price). The bars labeled *Price Increase* represent the proportion of properties for which the listing price was higher than the original. The bars labeled *Price Unchanged* represent the proportion of properties for which the listing price was exactly the same as the original. The bars labeled *Small Price Decrease* represent the proportion of properties for which the listing price was up to 5% below the original. The bars labeled *Large Price Decrease* represent the proportion of properties for which the listing price was below the original by over 5%.

Figure B.5: Correlation between sold price and Redfin estimate



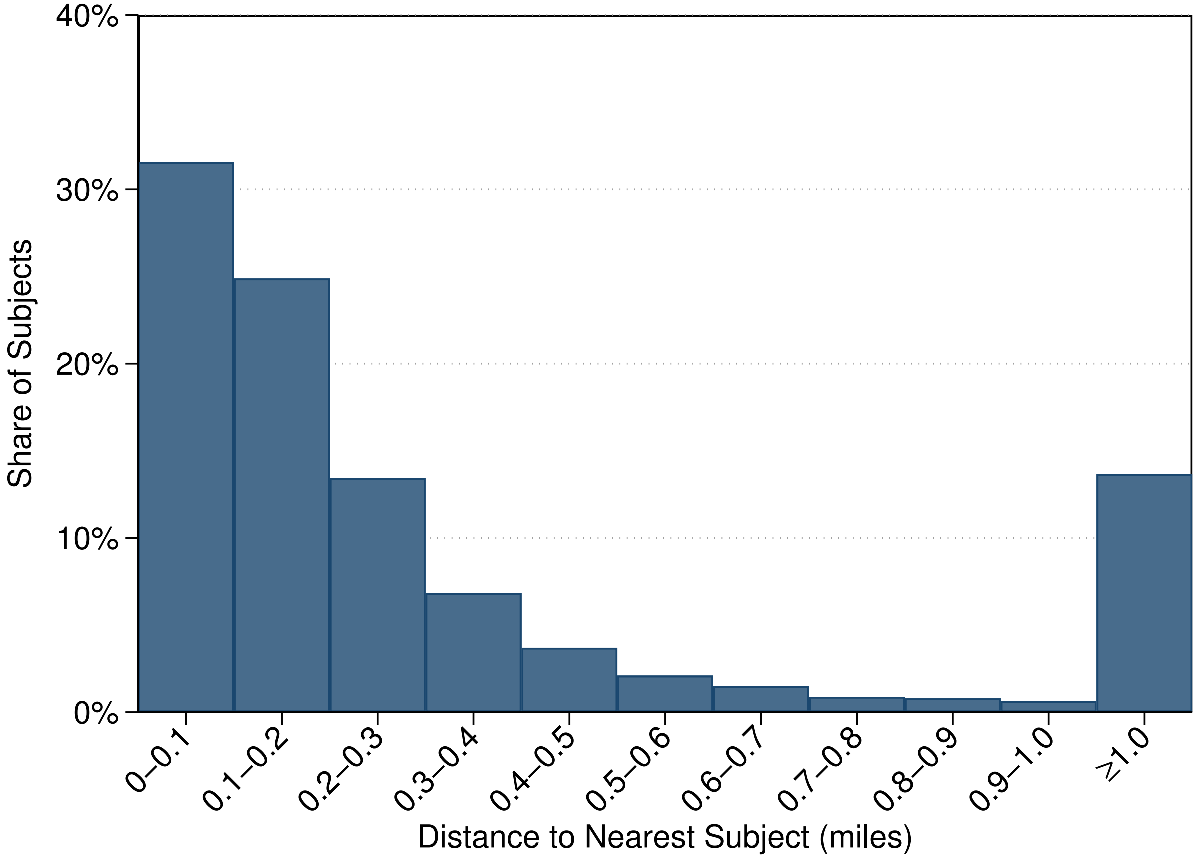

Notes: Sample of 32,949 properties sold during the study period. The x-axis presents 100 times the natural logarithm of the home's value according to the Redfin Estimate. The y-axis shows 100 times the natural logarithm of the actual sales price. The slope coefficient, heteroskedasticity-robust standard error (in parenthesis) and r-squared correspond to an OLS regression.

Figure B.6: Effects on Behavior: Probability of Changing Listing Price




Notes: Each panel represents an event-study tracking the effect of the information shock of the experiment on the probability that the listing price falls into one of four categories: price increase, constant price, small price decrease and large price decrease. Panel (a) presents the results for properties whose price increased. Panel (b) presents the results for properties whose price was the same as the original price. Panel (c) presents the results for properties for which the listing price was up to 5% below the original. Panel (d) presents the results for properties that saw a price decrease of over 5% relative to the original listing price. Within each panel, each coefficient corresponds to a separate regression based on 57,910 subjects from the field experiment. Every regression is based on the equation (4) from Section 2, and the coefficient being graphed corresponds to the coefficient on the key independent variable, *Information Shock* ($E_m^{j_i^*} \cdot D_i$). 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 had a particular change in listing price at a number of weeks after the start of the letter delivery and 0 otherwise. For example, the coefficient on *+12 weeks* in panel (a) is based on a dependent variable that takes the value 100 if the property's price 12 weeks after the start of the letter delivery was higher than at the start of the experiment. The vertical red line indicates the estimated date when the first letter was delivered (June 15 2019). The 90% confidence intervals are based on heteroskedasticity-robust standard errors.

Figure B.7: Distance to Nearest Subject




Notes: This figure shows the distribution of distance to the nearest subject based on the mailing addresses of the subjects in the field experiment. The results shown in this graph exclude 3,451 out of 57,910 subjects whose exact geo-locations could not be obtained for a variety of reasons (e.g., they used a P.O. box). Distance to the nearest subject is computed using the latitudes and longitudes obtained by geo-location for the scraped data.

Table B.1: Survival Models

| | (1) Cox | (2) Exponential | (3) Weibull | (4) Lognormal |
|---|---|---|---|---|
| Raw Coefficient: | | | | |
| Information shock | 0.990*** | 0.989*** | 0.989*** | 0.012*** |
| | (0.0029) | (0.0030) | (0.0030) | (0.0032) |
| Marginal Effects: | | | | |
| Information shock | | 1.809*** | 1.794*** | 1.918*** |
| | | (0.5201) | (0.5160) | (0.5209) |
| Median Days to Sell | | 165.47 | 165.34 | 157.46 |
| Observations | 57,574 | 57,574 | 57,574 | 57,574 |

Notes: This table presents the results from the survival models. The first row presents the raw coefficient on the information shock, which can be interpreted as a proportional hazard ratio in columns (1)-(3) (i.e. for column (1), a 1 pp increase in the information shock decreases the likelihood of selling the property by 1 - 0.990 = 1.0 pp). In column (4), the coefficient can be interpreted as a non-exponentiated raw coefficient (i.e. a 1 pp increase in the information shock decreases the likelihood of selling by 1-$e^{-0.012}$= 1.2 pp). The second row presents the marginal effect of increasing the information shock by 1 pp on the number of days to sell the property (i.e. for column (2), a 1 pp increase in the information shock increases the number of days to sell the property by 1.809 days). The fourth row presents the median number of days to sell the property as predicted by each model. The fifth row shows the number of observations used in estimating the models.

Table B.2: Sales Price

| | (1) $S_{29w}$ | (2) $log(SalePrice)$ |
|---|---|---|
| Information Shock | -0.295*** | 0.056 |
| | (0.108) | (0.097) |
| Mean Outcome | 56.90 | 1282.02 |
| SD Outcome | 49.52 | 73.29 |
| $R^2$ | 0.154 | 0.699 |
| Observations | 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 are based on data from the field experiment. The dependent variable in column (1) is an indicator variable ($S_{+29w}$) that takes the value 100 if the property was sold at 29 weeks after the start of the letter delivery and 0 otherwise. The dependent variable in column (2) is the natural logarithm of sales price multiplied by 100. See details in Appendix B.3 for details about the construction of sales price and imputation of prices for properties that were not sold during the study period. Both estimates employ the pooled specification given by equation (4), with *Information Shock* referring to the key independent variable: $E_m^{j_i^*} \cdot D_i$.

Table B.3: Next Purchase

| | Shares (%) | | | No. of Households | | |
|---|---|---|---|---|---|---|
| | (1) All subjects | (2) Occupant owners | (3) Non-occupant owners | (4) All subjects | (5) Occupant owners | (6) Non-occupant owners |
| (a) Sold within 28 weeks: | | | | | | |
| Bought in same county | 8.29 | 11.04 | 3.59 | 1,752 | 1,471 | 281 |
| Did not buy in same county | 91.71 | 88.96 | 96.41 | 19,389 | 11,850 | 7,539 |
| Total | 100 | 100 | 100 | 21,141 | 13,321 | 7,820 |
| (b) Did not sell within 28 weeks: | | | | | | |
| Bought in same county | 2.56 | 2.82 | 2.03 | 473 | 348 | 125 |
| Did not buy in same county | 97.44 | 97.18 | 97.97 | 18,031 | 12,004 | 6,027 |
| Total | 100 | 100 | 100 | 18,504 | 12,352 | 6,152 |

Notes: This table presents a breakdown of new property purchases for the 39,645 subjects in the field experiment with property listings in Florida. The breakdown provides information on whether the subject bought a new property in the same county that their original listing was located in. Information on new property purchases of subjects is obtained from county assessors' data for Florida counties.

# C Heterogeneity Analysis

## C.1 Heterogeneity by Information Source

We compare the effects of two types of information sources in Figure C.1. This figure estimates the same model from equation (4) of Section 2, only that in two separate samples: on the one hand, using groups Baseline, Past-1 and Past-2; on the other hand, using groups Baseline, Forecast-1, Forecast-2 and Forecast-3. It is important to note that the pooled estimates presented in the main body of the paper are more than the sum of the two effects. The pooled estimate also exploits the comparison between the Past and Forecast groups. As a result, we significantly lose power when separating the sample into these different groups and estimate them separately.

The point estimate suggests that information about the past was more effective. However, the effects of the forecasts are imprecisely estimated, and as a result we cannot rule out that the effects are as large as the effects of the past information. For example, at 12 weeks after the start of the letter delivery, the coefficient on the past-information shock is -0.426 (p-value=0.002) and on the forecast-information shock is -0.196 (p-value=0.349). However, we cannot confidently reject the null hypothesis that these two effects are equal to each other (p-value=0.541). In other horizons (16 out of the 28 one-week horizons we have in the data) we can reject the null hypothesis that the two coefficients are the same, however, even in those cases we cannot rule out that the effects of forecasts are half as large as the effects of information about the past.

## C.2 Heterogeneity Analysis

The evidence presented indicates that, when deciding whether to sell a property, the average subject cares about home price expectations. Next, we explore whether the average effect masks meaningful heterogeneity across subjects. The heterogeneity analysis is presented in Figure C.2. The coefficient at the very top of the figure corresponds to the effect for the average subject (-0.350, from column (3) of Table 2). The other coefficients are divided in pairs, corresponding to the same regression specification but estimated in split samples.

We split the sample by various observable characteristics of the property, the seller or the market, that are summarized below. The first form of heterogeneity presented in Figure C.2 corresponds to the owner occupancy status of the property. This form of heterogeneity was pre-registered based on the widespread conjecture that non-owner-occupied properties have a disproportionate influence on speculation in the housing market (Nathanson and Zwick, 2018; Gao et al., 2020). We split the sample by whether the owner was living in the property at the time of letter delivery (66.99% of subjects) or not (33.01% of subjects).[46] It is possible that home price expectations are more relevant for homes that are more expensive as the stakes are higher. To shed light on this question, we split the sample by whether the listing price is above or below the median listing price in our sample ($359,900). Our letters provided information about the median home values. Recipients may trust the information more when their listing price is close to that median home value, because they may infer that the information is more applicable to their own property. Thus, we split the subjects by whether the initial listing price of the listed property is closer or farther away from the median home value in their market. It is possible that owners react differently depending on whether they started with a more aggressive or conservative listing price. We split the sample by whether the initial listing price is above or below the estimated market value for that same property according to the website Redfin.com. For a subset of subjects for which demographic data was available via a data vendor, we split the sample by gender, age and ethnicity.[47] Last, we consider a number of local housing market characteristics. We split the sample by listings in counties from Florida (68% of the subjects) versus the rest of the counties. To differentiate between more and less expensive areas, we split the ZIP Codes by whether median home values are above or below the median ZIP Code. And we also split ZIP Codes by a measure of seller market power (the Zillow Seller Power Index) as well as the market size (Zillow's Size Rank).

Figure C.2 shows that the effects of home price expectations are similar across the board. None of the pairwise differences in Figure C.2 are statistically significant at conventional levels. And almost all of the pairwise differences are close to zero.[48] The most notable difference are by owner occupancy status and age. We find that the effects of the information shocks are almost twice as strong for the non-occupant owners (-0.439, p-value=0.025) as for the occupant owners (-0.247, p-value=0.060). The difference between those two estimates (-0.439 and -0.247) must be taken with a grain of salt, however, as it is statistically insignificant (p-value=0.409).[49] We also find that the point estimates were larger (in absolute value) for individuals aged 59 or older (-0.404, p-value=0.032) than for individuals below the age of 59 (-0.058, p-value=0.781). This result must also be taken with a grain of salt, however, as

[46] We can split subjects in this way by comparing the owner's mailing address compared to the listed property address using county assessor's office records. Because some mailing address may take some time to be updated in the official records, we complemented those records with mail forwarding data from the United States Postal Service.

[47] The vendor was able to provide data for most, but not all, the homeowners in the subject pool. For more details about this complementary data see Appendix A.4.

[48] These results must be taken with a grain of salt, however: given the statistical precision, in some cases we cannot rule out moderate or even large differences.

[49] In Appendix C.3 we reproduce this heterogeneity analysis for other time horizons, and the results are broadly consistent.

the difference in coefficients (-0.404 and -0.058) is statistically insignificant (p-value=0.220). Indeed, the data on the seller's age is missing for a third of the sample, and as a result, the heterogeneity by age is less precisely estimated.

The differences in occupancy status and age are statistically insignificant, so we should not draw any strong conclusions. Moreover, even if those differences were statistically significant, we should still be cautious with our interpretation because we did not randomize the owner occupancy status or the age of the subjects. With those caveats in mind, we believe that the above evidence may be suggestive of selling frictions. Owners who are non-occupants and owners who are age 59 or older probably face less frictions when deciding whether to sell their home or when to sell it, and thus are better able to react when their expectations change. Occupant owners need to move out of the property after selling it, while non-occupant owners do not need to move out.[50] And relative to those aged 59 or above, owners aged below 59 years are more likely to experience changes in jobs, school or marital status that may put constraints on when the property can be sold.[51]

## C.3 Occupancy Status and Selling Frictions

In this section we provide more details on the heterogeneity of information shocks by occupancy status, which suggests the possibility of selling frictions in the home selling process for owner-occupied homes. Short of randomizing the selling frictions through an experiment, we cannot establish that this is the only mechanism at play. However, we do provide evidence against some alternative interpretations. A first potential concern is that the effects on non-occupant are stronger because the level of the dependent variable is higher for that group. However, we can quickly rule out that explanation, as the timing of sales outcomes looks quite similar between the two groups. For example, the probability of selling the property at 12 weeks post-treatment was 38.09 pp for non-occupant owners versus 36.45 pp for occupant owners.

Figure C.3 shows the share of properties sold over time during the experiment, distinguishing between occupant owners versus non-occupant owners. The timing of the sales is generally similar between the two groups. However, the share of properties sold is larger for non-occupant owners relative to occupant owners at any point in time during the experiment. This discrepancy is almost negligible at the beginning, but widens over time, which is consistent with the possibility of non-occupant owners facing fewer selling frictions.

[50] In principle, the large heterogeneity by occupancy status could be due to other factors. In Appendix C.3 we discuss alternative explanations and present some evidence suggesting they are less plausible.

[51] Our interpretation of selling frictions in the housing market is related to evidence of frictions on another important asset market: the market for stocks and bonds (Giglio et al., 2019).

Another potential concern could be that non-occupant owners may be more responsive because they are different in other characteristics. For example, if non-occupants are more educated they may find it easier to understand the information and incorporate it into their decision-making. However, we provide evidence against this interpretation. Columns (2) and (3) of Table C.1 breaks down the average characteristics of the subjects and their properties by occupancy status. Column (4), in turn, presents the p-value corresponding to the null hypothesis that the average characteristics are equal between occupant and non-occupant owners. Due to the high statistical power from our large sample sizes, almost all of the pairwise differences are highly statistically significant. However, those differences are economically small across the board. For instance, consider the differences in education that motivated this analysis. The share of college graduates is 40.3% among occupants and 41.4% among non-occupants. It is highly unlikely that this tiny difference in education attainment could explain the large differences in the effects of the information shocks that we observe between occupants and non-occupants.

Perhaps non-occupant owners react more to the information because they do not live near the property and are thus less informed about the evolution of home prices in the property's area. We can test this hypothesis directly by exploiting variation in the distance to the property among non-occupants. The results are presented in Figure C.4.b, which breaks down the event-study analysis the sample by non-occupant owners that live in the same ZIP Code as the property (44.08% of the non-occupant owners) versus the non-occupant owners that live in a different ZIP Code (the remaining 55.92%). According to this alternative channel, the effects of information should be concentrated in the non-occupant owners that live in a different ZIP Code. Instead, the results indicate that the effects of the information shocks are quite similar regardless of whether the non-occupants live in the same ZIP code as the property or not. For all 28 one-week horizons, the pairwise differences in coefficients are always close to zero and statistically insignificant.

Last, recall from Section 7 above that one potential source of attenuation bias is that some sellers are planning to buy again in the same area. A related concern could be that the estimates for non-occupants are weaker because they are more affected by this source of attenuation bias. More precisely, if after selling the property, the occupants are more likely than non-occupants to buy in the same area again, that could explain the difference in responses to the information. The share of subjects who, after selling the listed property, buy another property in the same county is 11% for occupants and 4% for non-occupants. Consider the most extreme form of attenuation bias, in which the effects of information are zero for property owners who buy in the same area. Even in this extreme case, we would expect the effects of information shocks to be 7% stronger ($= 11 - 4$) for non-occupants than for occupants. By comparison, we observe that the point estimate is 77% stronger ($= \frac{-0.439}{-0.247}$, at 12 weeks post-treatment) for non-occupants (though recall the difference is not statistically significant). That is, even in its most extreme form, the attenuation bias would likely not explain the potential differences in effects between non-occupants and occupants.

Because a majority of subjects are not buying again in the same county, a natural question is what are they doing instead? Are they staying in the same county but renting? Or are they moving to a different county? To address these questions, we use National Change of Address (NCOA) Data from the United States Postal Service (USPS). By merging the mail forwarding data to the data from the experiment, we can identify subjects' new mailing addresses. We are able to match 96% of subjects from our original sample by using unique identifiers provided in the datasets (i.e., full names and full addresses).

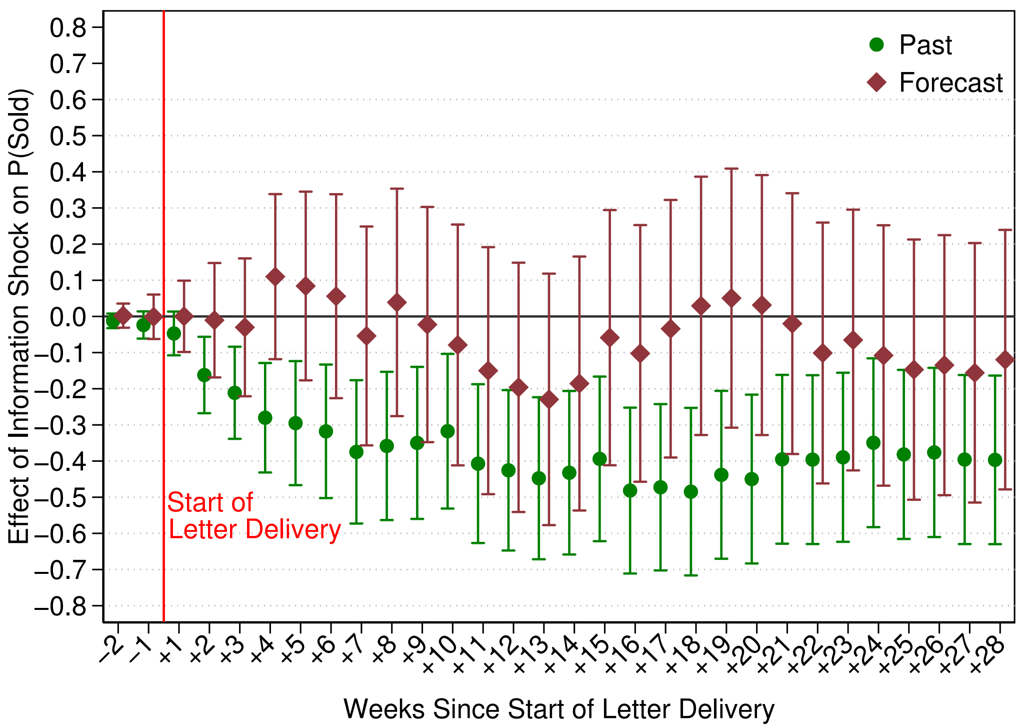
We compare the subjects' new mailing addresses to the address of their property listing, and report the results in Table C.2.a for subjects who sold their property within 28 weeks and in Table C.2.b for subjects who did not sell their property within 28 weeks below.

As expected, occupant owners are more likely to move after selling their property than non-occupant owners: in our sample, the share of subjects who move are 53% for occupant owners and 4% for non-occupant owners. The share of subjects who move may be underestimated, as some subjects may not have used mail forwarding services during their move, or chose their data to remain confidential, and thus do not show up in the NCOA records.

The majority of the subjects who move relocate to a different zip code but stay in the same state. Most stay in the same county, although that pattern is less pronounced. Given that most subjects move to a different geographical area after selling their property, it may be possible that the information shock provided in the experiment did not have as strong an effect on subjects' expectations about the home price of their new properties.

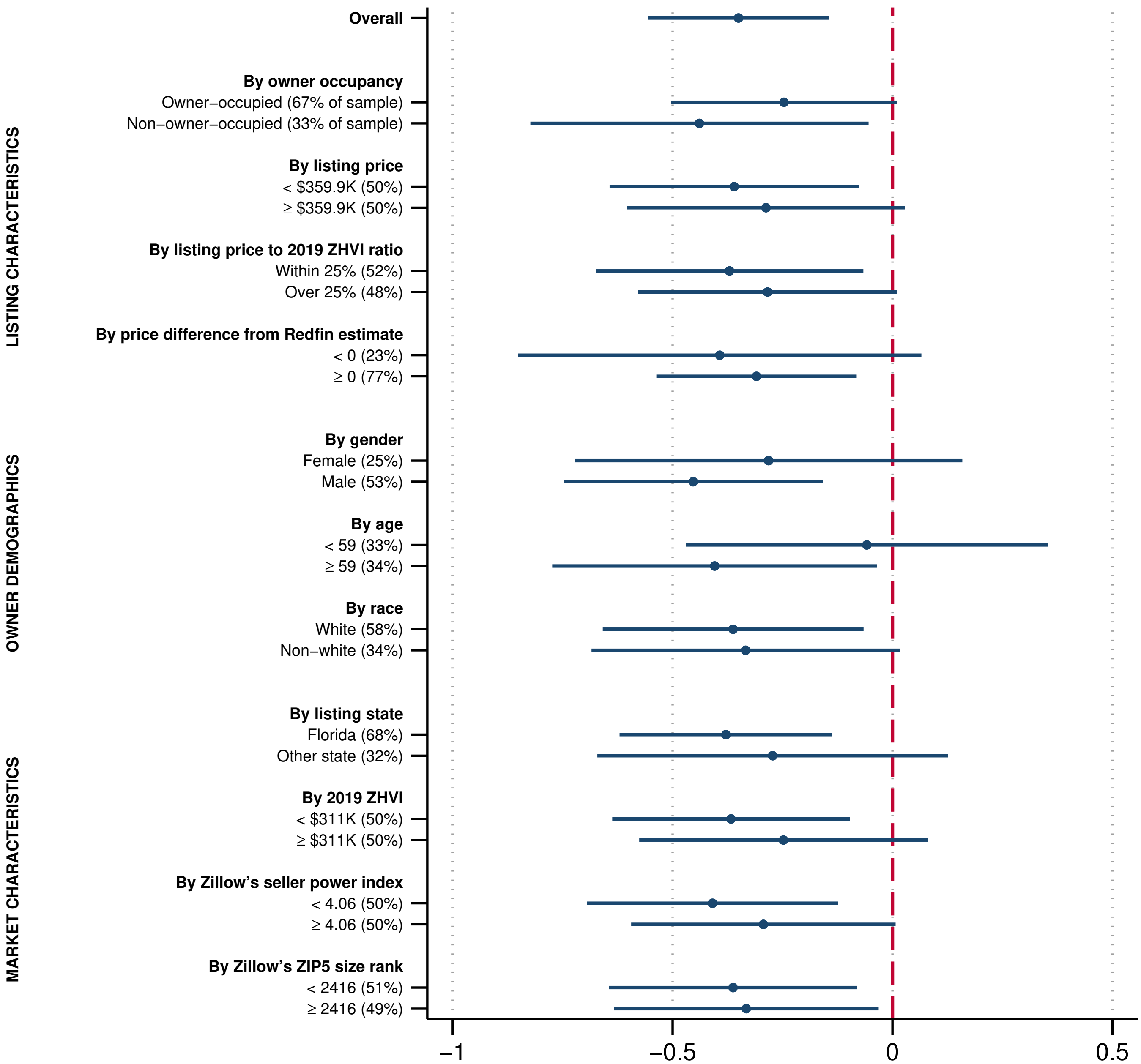
Subjects who were unable to sell their home within the 28 weeks of the experiment were significantly less likely to move in that time frame, with only 7% of occupant owners and 3% of non-occupant owners moving away. The pattern of new addresses is similar: while most subjects who move remain in the same state, the majority move to a different zip code, and around half move to a different county.

Figure C.1: Heterogeneity of Behavioral Effects by Past/Forecast Information Sources: Event-Study Analysis



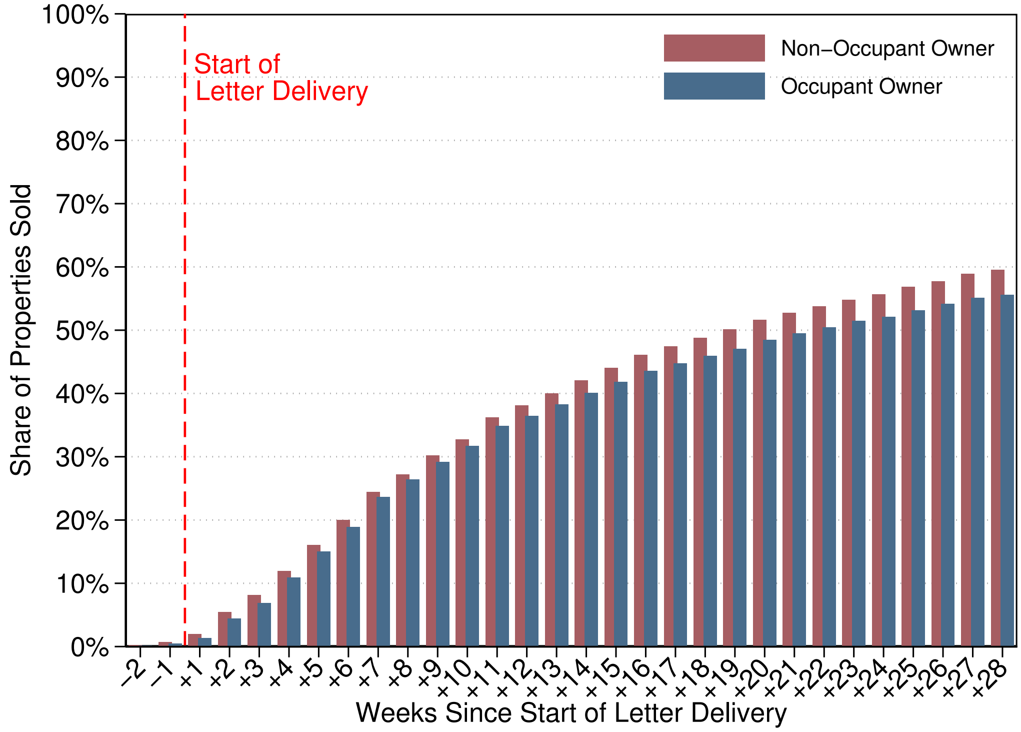

Notes: Each coefficient corresponds to a separate regression based on 57,910 subjects from the field experiment. Every regression corresponds to equation (4) from Section 2, and the coefficient being graphed corresponds to the coefficient on the key independent variable, *Information Shock* ($E_m^{j_i^*} \cdot D_i$). All regressions are identical except for two features: the dependent variable and the sample. Each of the green circles are based on a regression with the 28,828 subjects from the field experiment who received either the Baseline treatment, the Past-1 treatment (i.e., annual home price change over the past year) or the Past-2 treatment (ie., annual home price change over the past two years). Each of the red diamonds are based on a regression with the 40,569 subjects from the field experiment who received the Baseline, Forecast-1, Forecast-2, or Forecast-3 treatments. 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 vertical red line indicates the estimated date when the first letter was delivered (June 15 2019). The 90% confidence intervals are based on heteroskedasticity-robust standard errors.

Figure C.2: Heterogeneity Analysis




Notes: Each coefficient corresponds to a separate regression based on 57,910 subjects corresponding to equation (4) from Section 2. The coefficients correspond to the one on the key independent variable, *Information Shock* ($E_m^{j_i^*} \cdot D_i$). The dependent variable used is an indicator variable that takes the value 100 if the property has been sold at 12 weeks after the start of the letter delivery. All regressions are identical except for the sample restrictions, which are presented in the y-axis labels. For example, the coefficient for *Owner-occupied* represents the coefficient on *Information shock* when the regression is run for occupant owners only. The percentage totals in parentheses represent the proportion of the 57,910 subjects that each sample represents. The percentages do not sum to 100% for the demographic breakdowns due to missing values. The 90% confidence intervals are based on heteroskedasticity-robust standard errors.

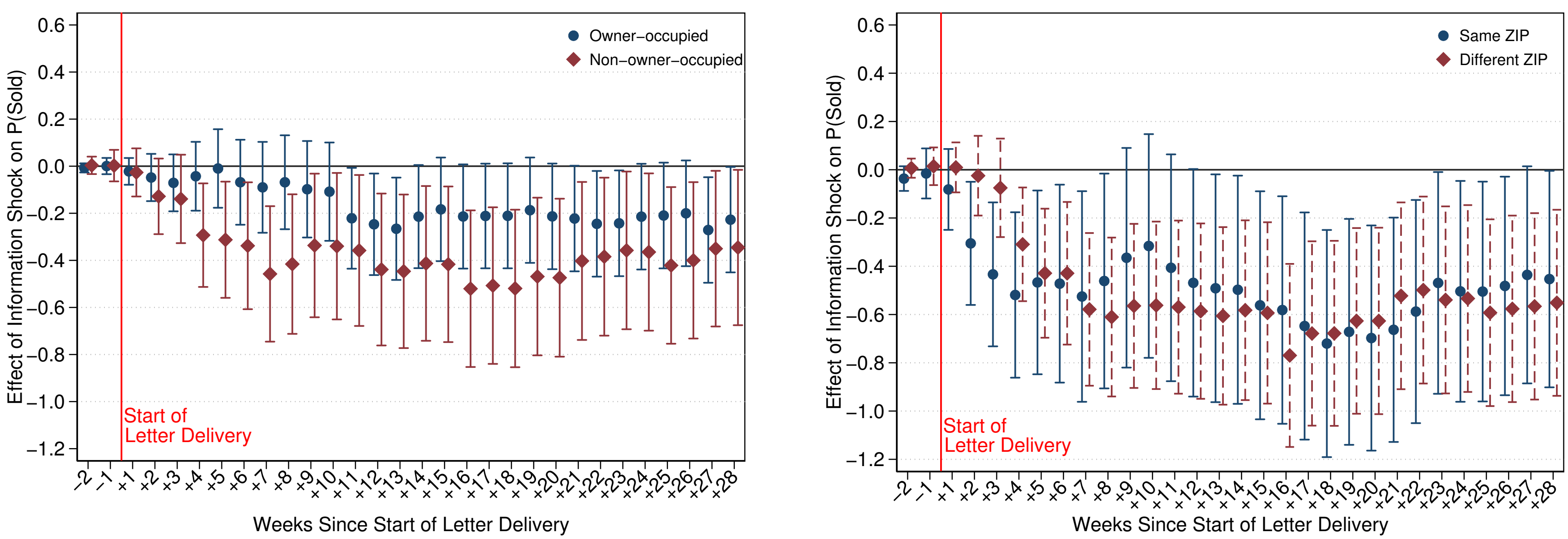
Figure C.3: Timing of Property Sales by Owner-Occupied Status




Notes: The graph shows the fraction of the properties in the subject pool that were sold at each point in time, according to the administrative records. The blue bars present the results for occupant owners, while the red bars present the results for non-occupant owners. The vertical red line indicates the estimated date when the first letter was delivered (June 15 2019).

Figure C.4: Heterogeneity of Behavioral Effects by Owner-Occupied Status: Event-Study Analysis

**a.** By Owner Occupancy **b.** Non-Occupants by Place of Residence




Notes: Panel (a) shows the heterogeneity of behavioral effects depending on whether the owner lives in the listed property or not. Each coefficient in the panel corresponds to a separate regression. Every regression corresponds to equation (4) from Section 2, and the coefficient being graphed corresponds to the coefficient on the key independent variable, *Information Shock* ($E_m^{j_i^*} \cdot D_i$). 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 38,795 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 19,115 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 vertical red line indicates the estimated date when the first letter was delivered (June 15 2019). The 90% confidence intervals are based on heteroskedasticity-robust standard errors. Panel (b) further breaks down the heterogeneity of behavioral effects for non-occupant owners, depending on whether they reside in the same zip code as the property they are selling (as determined by their mailing address). Each coefficient corresponds to a separate regression. The blue circles show the results for the 8,425 non-occupant owners who reside in the same zip code as their listed property. The red diamonds show the results for the 10,690 non-occupant owners who reside in a different zip code to their listed property. The design of the regression, dependent variable, and confidence intervals are identical to panel (a).

Table C.1: Descriptive Statistics about the Properties and their Owners

| | (1) All | (2) Occupant Owners | (3) Non-Occupant Owners | (4) P-value |
|---|---|---|---|---|
| Property Characteristics | | | | |
| Days Listed | 86.654 | 81.947 | 96.208 | <0.001 |
| | (0.477) | (0.576) | (0.843) | |
| List Price ($1,000s) | 574.756 | 619.176 | 484.604 | <0.001 |
| | (3.914) | (5.274) | (5.040) | |
| No. Bedrooms | 3.256 | 3.390 | 2.985 | <0.001 |
| | (0.005) | (0.005) | (0.008) | |
| No. Bathrooms | 2.608 | 2.693 | 2.435 | <0.001 |
| | (0.004) | (0.005) | (0.007) | |
| Sq. Ft. Built (1,000s) | 2.295 | 2.416 | 2.049 | <0.001 |
| | (0.005) | (0.007) | (0.008) | |
| Sq. Ft. Lot (1,000s) | 12.958 | 12.978 | 12.920 | 0.769 |
| | (0.089) | (0.103) | (0.168) | |
| Individual Characteristics | | | | |
| Age | 58.675 | 57.081 | 61.911 | <0.001 |
| | (0.080) | (0.097) | (0.138) | |
| % Female | 31.890 | 32.251 | 31.153 | 0.018 |
| | (0.219) | (0.268) | (0.379) | |
| % White | 62.872 | 61.603 | 65.424 | <0.001 |
| | (0.210) | (0.258) | (0.358) | |
| % African-American | 3.203 | 3.432 | 2.742 | <0.001 |
| | (0.076) | (0.097) | (0.123) | |
| % Hispanic | 13.556 | 15.695 | 9.257 | <0.001 |
| | (0.149) | (0.193) | (0.218) | |
| Income ($1,000s) | 128.059 | 126.752 | 130.806 | <0.001 |
| | (0.336) | (0.403) | (0.605) | |
| % Higher Ed | 40.689 | 40.257 | 41.419 | 0.042 |
| | (0.276) | (0.347) | (0.454) | |
| Observations | 57,910 | 38,795 | 19,115 | |

Notes: Average characteristics, with standard errors in parentheses. Column (1) corresponds to the entire sample of 57,910 subjects. Column (2) presents the characteristics for the 38,795 subjects who live in the property they are selling. Column (3) presents the characteristics for the 19,115 subjects who do not live in the property they are selling. Column (4) reports the p-value of the test of equal means across occupant owners and non-occupant owners. All 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. Individual characteristics based on data provided by a private vendor.

Table C.2: New Mailing Address

| | Shares (%) | | | No. of Households | | |
|---|---|---|---|---|---|---|
| | (1) All subjects | (2) Occupant owners | (3) Non-occupant owners | (4) All subjects | (5) Occupant owners | (6) Non-occupant owners |
| (a) Sold within 28 weeks: | | | | | | |
| Didn't move | 63.98 | 47.18 | 95.79 | 21,080 | 10,172 | 10,908 |
| Moved | 36.02 | 52.82 | 4.21 | 11,869 | 11,390 | 479 |
| Same ZIP Code | 5.77 | 8.54 | .53 | 1,902 | 1,842 | 60 |
| Different ZIP Code | 30.25 | 44.28 | 3.68 | 9,967 | 9,548 | 419 |
| Same county | 20.16 | 29.9 | 1.69 | 6,641 | 6,448 | 193 |
| Different county | 15.87 | 22.92 | 2.51 | 5,228 | 4,942 | 286 |
| Same state | 28.49 | 42.13 | 2.66 | 9,388 | 9,085 | 303 |
| Different state | 7.53 | 10.69 | 1.55 | 2,481 | 2,305 | 176 |
| Total | 100 | 100 | 100 | 32,949 | 21,562 | 11,387 |
| (b) Did not sell within 28 weeks: | | | | | | |
| Didn't move | 94.12 | 92.85 | 96.95 | 23,493 | 16,001 | 7,492 |
| Moved | 5.88 | 7.15 | 3.05 | 1,468 | 1,232 | 236 |
| Same ZIP Code | 1.01 | 1.24 | .5 | 252 | 213 | 39 |
| Different ZIP Code | 4.87 | 5.91 | 2.55 | 1,216 | 1,019 | 197 |
| Same county | 2.78 | 3.59 | .97 | 694 | 619 | 75 |
| Different county | 3.1 | 3.56 | 2.08 | 774 | 613 | 161 |
| Same state | 4.13 | 5.29 | 1.57 | 1,032 | 911 | 121 |
| Different state | 1.75 | 1.86 | 1.49 | 436 | 321 | 115 |
| Total | 100 | 100 | 100 | 24,961 | 17,233 | 7,728 |

Notes: This table presents the breakdown of moving patterns for the 57,910 subjects from the field experiment who sold their property within 28 weeks, broken down by whether they sold their property within 28 weeks or not. Data on the relocations is obtained by comparing mailing addresses provided to NCOA. The categories under *Moved* represent the address of the owner's new property in relation to the address of the property they sold. For example, if a homeowner had a new mailing address on record with NCOA, and that new address was in the same ZIP Code as their property, their information will be captured under *Moved - Same ZIP Code*.

# D Sample of Full Letter

UCLA

Los Angeles, May 10th 2019

Dear Jane Doe,

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:

**Median Price**

**2-bedroom home in ZIP Code 33308**

**May 2019:** $343,000

Notes: for more details, see the notes in the back of this page.

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 **ABCDEF**

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

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* | *Nicolas Bottan* |
| Assistant Professor of Economics | Post-Doctoral Associate |
| University of California, Los Angeles | 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: the median price for 2019 corresponds to the Zillow Home Value Index: A smoothed, seasonally adjusted measure of the median estimated home value across a given region and housing type. This data is Published and made publicly available online by Zillow Research.*

T1 P1 AUTO**SCH 5-DIGIT 32080
Jane Doe
1408 NE 17th Ave
Fort Lauderdale, FL 33104

# E Housing Study Website

UCLA Anderson School of Management

APPLY | FOR COMPANIES | GIVE

## Information about Housing Study

You were randomly chosen to receive this letter, and we will not send you any more letters in the future.

This study was approved by the UCLA and Cornell Institutional Review Boards. Your participation is voluntary and greatly appreciated. The results of this study can provide valuable insights to homeowners across the country. You may withdraw from the study at any time.

Your privacy is very important to us. When information is transferred online there is a possibility that at may be viewed by a third party. To reduce the risk that an outside party could identify you or observe your responses, this survey employs Transport Layer Security (TLS) encryption for all transmitted data. As a result, we anticipate that your participation to this survey presents no greater risk than everyday use of the internet. Your responses will be used solely for research purposes and will be kept strictly confidential, shared only with the researchers named below.

This study is being conducted by Ricardo Perez-Truglia (Assistant Professor at UCLA) and Nicolas Bottan (Postdoctoral Fellow at Cornell). If you have any questions or concerns about this survey please contact us at ricardo.truglia@anderson.ucla.edu or nicolas.bottan@cornell.edu.

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.

Thank you for your attention,

The Research Team

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

Nicolas Bottan
Post-Doctoral Associate
Cornell University

*FOR VISITORS*
campus tour
maps & directions
master calendar
facility use

*FOR COMPANIES*
recruit a student
post a job
consulting teams
for GAP companies

*FOR THE NEWS MEDIA*
media relations
ucla anderson forecast
anderson in the news
faculty directory
faculty expertise

directory
site index
portal
library
UCLA
feedback
© UC Regents

# F Letter Survey Instrument

Thank you for your interest in completing our brief, 2-minute survey about household decision making in the US housing market!

Your participation is voluntary and is greatly appreciated. You may withdraw from the survey at any time. Your responses will be used solely for research purposes and will be kept strictly confidential, used only by the Principal Investigators.

Please enter the validation code included in the letter to begin:

The company Zillow measures median home prices, and makes it publicly-available through their website.

According to that data, in May of 2019, the median price for a **2-bedroom** home in **ZIP Code 33062** was **$336,000.**

In the following screens, we will ask you to guess how this price will change in the future.

In May of 2019, the median price for a 2-bedroom home in ZIP Code 33062 was $336,000.

What do you think this median price will be **1 year later (in May 2020)**?

How confident are you about this prediction?

| Not at all confident | Somewhat confident | Confident | Very confident |
|---|---|---|---|
| ○ | ○ | ○ | ○ |

In May of 2019, the median price for a 2-bedroom home in ZIP Code 33062 was $336,000.

What do you think this median price will be **5 years later (in May 2024)**?

How confident are you about this prediction?

| Not at all confident | Somewhat confident | Confident | Very confident |
|---|---|---|---|
| ○ | ○ | ○ | ○ |

Recent research on decision making shows that choices are affected by the context in which they are made. Differences in how people feel, in their previous knowledge and experience, and in their environment can influence the choices they make. To help us understand how people make decisions, we are interested in information about you, specifically whether you actually take the time to read the instructions; if you don't, some results may fail to tell us very much about decision making in the real world. To help us confirm that you have read these instructions, please ignore the question below about how you are feeling and instead check only the "none of the above" option. Thank you very much.

- ☐ Interested
- ☐ Hostile
- ☐ Nervous
- ☐ Distressed
- ☐ Enthusiastic
- ☐ Determined
- ☐ Excited
- ☐ Proud
- ☐ Attentive
- ☐ Upset
- ☐ Irritable
- ☐ Jittery
- ☐ Strong
- ☐ Alert
- ☐ Active
- ☐ Scared
- ☐ Inspired
- ☐ None of the above

In your opinion, were the questions included in this survey easy or difficult to understand?

- ○ Easy to understand
- ○ Neither easy nor difficult
- ○ Difficult to understand

# G AMT Survey Instrument

Hi. We are a group of academic researchers from Cornell and UCLA. This survey is part of a research study. Our goal is to understand how you form expectations about the future. It should take you approximately 7 minutes to complete this survey. You will obtain the HIT code after completing the survey.

Participation in this study does not involve any physical or mental risks different from those found in everyday life. Your participation in this study is purely voluntary, and you may withdraw and retract your responses at any time without penalty. If you have any questions about this study, you may contact us at nlb75@cornell.edu or ricardo.truglia@anderson.ucla.edu. You must be a U.S. resident aged 18 years or older in order to participate.

*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 OHRPP by phone: (310) 206-2040; by email: participants@research.ucla.edu or by mail: Box 951406, Los Angeles, CA 900950-1406.*

○ Yes, I would like to take part in this study and confirm that I am a U.S. RESIDENT, aged 18 or older. I understand I may not qualify to participate in the survey.

This survey is about perceptions and expectations of the housing market. For that, we need to know a little bit of information about yourself:

Do you currently live with your parents/legal guardians?

- ◯ Yes
- ◯ No

Do you (or your parents/legal guardians) rent or own your primary residence (the place where you usually live)?

- ◯ Rent
- ◯ Own

Please indicate the type of your current primary residence. Is your primary residence a:

- ◯ Single Family Home
- ◯ Apartment/Condo/Co-op
- ◯ Townhouse/Duplex
- ◯ Mobile/Manufactured home
- ◯ Other

How many bedrooms does your primary residence have?

- ◯ 1 Bedroom
- ◯ 2 Bedrooms
- ◯ 3 Bedrooms
- ◯ 4 Bedrooms
- ◯ 5+ Bedrooms

What is the (5-digit) ZIP code of your primary residence?

How many years have you lived in your primary residence?

The company Zillow measures median home prices, and makes it publicly available through their website.

According to that data, in May of 2019, the median price for a **2 bedroom** home in **ZIP code 33149** was **$781,000.**

In the next screens, we will ask you how you think that this median price will evolve in the future.

In May 2019, the median price for a 2 bedroom home in ZIP code 33149 was $781,000.

What do you think this median price will be **1 year later (in May 2020)**? (Note: please do not write in dollar signs, commas or decimal points)

How confident are you about this value?

| Not at all confident | Somewhat confident | Confident | Very confident |
|---|---|---|---|
| ○ | ○ | ○ | ○ |

Next, a group of individuals participating in this  survey will be randomly chosen to receive some information about home prices in your ZIP code.

Please continue to the next screen to find out if you will be selected to receive this information.

You have been selected to receive the following information. Please take a moment to review the information carefully.

This information is only shown once and you will not be able to come back to it.

| Median Price | | | |
|---|---|---|---|
| **2 bedroom home in ZIP code 33149** | | | |
| **May 2017:** | $869,000 | | |
| **May 2018:** | $828,000 | ↓ | -4.7% |
| **May 2019:** | $781,000 | ↓ | -5.7% |

Methodological Notes

We want to give you the opportunity to reassess your answer to some of the previous questions. This opportunity is given automatically to all survey participants, regardless of their responses or whether they received information from us.

In May 2019, the median price for a 2 bedroom home in ZIP code 33149 was $781,000.

What do you think this median price will be **1 year later (in May 2020)**? (Note: please do not write in dollar signs, commas or decimal points)

How confident are you about this prediction?

| Not at all confident | Somewhat confident | Confident | Very confident |
|---|---|---|---|
| ○ | ○ | ○ | ○ |

In May of 2019, the median price for a  2 bedroom home in ZIP code 33149 was $781,000.

What do you think this median price will be **5 years later (in May 2024)**? (Note: please do not write in dollar signs, commas or decimal points)

How confident are you about this prediction?

| Not at all confident | Somewhat confident | Confident | Very confident |
|---|---|---|---|
| ○ | ○ | ○ | ○ |

Now, we want to ask you about your perceptions of returns to the U.S. stock market. More specifically, the Dow Jones Industrial Average.

The closing price for the Dow Jones was **$24,815** on May 31st 2019.

What do you think this price will be **1 year from now (on May 31st 2020)**? (Note: please do not write in dollar signs, commas or decimal points)

How confident are you about this prediction?

| Not at all confident | Somewhat confident | Confident | Very confident |
|---|---|---|---|
| ○ | ○ | ○ | ○ |

In the next page, you will be asked to choose between two investments. Once the study is completed, we will randomly pick one survey participant. Twelve months from now, this lucky participant will be paid extra money according to the investment choice they made. So answer the question in the next page carefully, as it may affect your future earnings.

You have $200 to split between two types of investments:

**Investment Options**

**Stock Market**

The rate of return will be equal to the rate of return of the U.S. stock market (more specifically, the Dow Jones Industrial Average). If the stock market grows by X% between now and May 2020, you will earn the amount you invest in this option plus X% in interest.

**Housing Market**

The rate of return will be equal to the growth rate of the median price of 2 bedroom homes in ZIP code 33149. So if the median house price grows by X% between now and May 2020, you will earn the amount you invest in this option plus X% interest.

How would you split your funds between these two investments? (Note: your investment must sum up to $200)

| | |
|---|---|
| Stock Market | 0 |
| Housing Market | 0 |
| Total | 0 |

If you are chosen as the lucky participant, we will contact you through mTurk.

Generally speaking, do you think now is a good time or a bad time to buy a house in your ZIP code?

| Very Bad | Bad | Good | Very Good |
| --- | --- | --- | --- |
| ○ | ○ | ○ | ○ |

What about selling a house -- generally speaking, do you think now is a good time or a bad time to sell a house in your ZIP code?

| Very Bad | Bad | Good | Very Good |
| --- | --- | --- | --- |
| ○ | ○ | ○ | ○ |

Thinking about your home -- what is the percent chance that you will put your home up for sale in the next 12 months?

Over the last 12 months, how often have you consulted websites or other sources that give you information on the past, present or future home prices?

- Never
- Once a year
- A few times a year
- Once a month
- Once a week or more

We are almost done. We would like to ask you a few more questions about yourself before finishing the survey.

Please indicate your gender:

- ◯ Female
- ◯ Male

How old are you?

- ◯ 18 - 24
- ◯ 25 - 30
- ◯ 31 - 34
- ◯ 35 - 40
- ◯ 41 - 50
- ◯ 51 - 60
- ◯ 61 - 70
- ◯ 70 or older

Which of the following best describes your ethnicity?

- ◯ White
- ◯ Black or African American
- ◯ Asian or Native Hawaiian and other Pacific Islander
- ◯ American Indian or Alaska Native
- ◯ Hispanic or Latino Origin

Are you currently married or living with a partner (not including roommates)?

- ◯ Yes
- ◯ No

Do you have kids?

- ◯ Yes
- ◯ No

Recent research on decision making shows that choices are affected by the context in which they are made. Differences in how people feel, in their previous knowledge, experience, and in their environment can influence the choices they make. To help us understand how people make decisions, we are interested in information about you. Specifically, whether you actually take the time to read the instructions. If you don't, some results may fail to tell us very much about decision making in the real world. To help us confirm that you have read these instructions, please ignore the question about how you are feeling. Instead, only check the "none of the above" option. Thank you very much.

- ☐ Interested
- ☐ Hostile
- ☐ Nervous
- ☐ Distressed
- ☐ Enthusiastic
- ☐ Determined
- ☐ Excited
- ☐ Proud
- ☐ Attentive
- ☐ Upset
- ☐ Irritable
- ☐ Jittery
- ☐ Strong
- ☐ Alert
- ☐ Active
- ☐ Scared
- ☐ Inspired
- ☐ None of the above

In your opinion, were the questions included in this survey easy or difficult to understand?

- ◯ Easy to understand
- ◯ Neither easy nor difficult
- ◯ Difficult to understand

Feel free to share any comments with us below.