Between Scylla and Charybdis: Florida House Price Expectations amid Market Cooling and Hurricane Risk*

Samir Huseynov

Auburn University

July 2026

Abstract

Florida homeowners entered 2025 between a cooling housing market and looming hurricane risks. During the 2025 Atlantic hurricane season, we recruit property owners across the state and elicit incentivized prior and posterior probability beliefs about future house price growth in their ZIP codes. Homeowners show optimistic priors and assign around 60% probability to a price increase, although nearly all respondent-ZIPs later experience a price decline. Between the prior and posterior elicitations, we provide information treatments highlighting either market cooldown (Economic) or increased hurricane frequency (Natural). Relative to Baseline, the Economic and Natural treatments reduce the probability assigned to a price increase by 11 and 7 percentage points, respectively. Information treatments also shift probability mass and improve forecast accuracy relative to different market benchmarks and realized prices. However, these corrections do not translate into detectable changes in selling intentions, insurance premium tolerance, or modeled insurance decisions. After the hurricane season, our follow-up study no longer detects treatment effects. Minimal communication can immediately improve homeowner expectations, but does not guarantee durable learning.

JEL codes: C93, D83, D84, G50, Q54, R31.
Keywords: house-price expectations, information experiments, belief persistence, climate risk, probabilistic elicitation.

*Huseynov: Department of Agricultural Economics and Rural Sociology, Auburn University (e-mail: szh0158@auburn.edu). The experiment was registered in the AEA RCT Registry as AEARCTR-0016346.

1 Introduction

For U.S. homeowners, the home is both a place to live and a major balance-sheet asset. Climate risks and extreme weather events, therefore, increasingly matter for household wealth. Especially in Atlantic coastal markets, storms and hurricanes can affect housing through physical damage, insurance premiums, maintenance costs, disclosure rules, and resale values (Hsiang et al., 2017; Wing et al., 2022; Boustan et al., 2020). A growing literature shows that these risks are capitalized into property values through realized damage, floodplain maps and disclosures, insurance costs, and heterogeneous climate beliefs, although market capitalization can be incomplete or delayed.[1]

Transaction data do not fully explain this slow and incomplete market adjustment, because they are observed only when homeowners and buyers agree to trade. If incumbent owners are more optimistic than potential buyers about future house prices or resale values, they may set reservation prices above the market-clearing level, leading to weaker sales volume or longer time on the market (Genesove and Mayer, 2001; Armona et al., 2019; Keys and Mulder, 2020; Bottan and Perez-Truglia, 2025). However, transaction volume and listing durations can also reflect search frictions, credit restrictions, and bargaining dynamics. Transaction data alone are therefore not well equipped to identify the role of behavioral factors in the pass-through of climate risk to house prices. Survey-based experiments can provide insight into how homeowners form, update, and retain house-price beliefs, complementing transaction data (Manski, 2004). By employing incentivized experimental protocols, this paper examines how incumbent homeowners form house-price expectations in a climate-exposed market, whether simple risk-relevant information can correct those beliefs, and whether any correction persists long enough to matter for subsequent market decisions.

Several factors make Florida a unique setting to study these questions. The 2024

[1]For instance, see Ortega and Taspinar (2018); Bernstein et al. (2019); Baldauf et al. (2020); Gibson and Mullins (2020); Hino and Burke (2021); Bakkensen and Barrage (2022) and Gourevitch et al. (2023).

hurricane season delivered three landfalls in Florida, damaging local economies and private property.[2] The season also intensified pressure on the state's property-insurance market.[3] These pressures exposed broader fragility in Florida's insurance system, including private-market retrenchment and public concern about the long-run solvency of Citizens, the state-backed insurer of last resort.[4]

Yet these developments do not imply that homeowners necessarily incorporated escalating natural risk into their price expectations. Previous work shows that housing expectations are consequential for housing and mortgage decisions, but are often extrapolative and shaped by recent local price trends (Attanasio et al., 2011; Case et al., 2012; Glaeser and Nathanson, 2017; Kuchler and Zafar, 2019; Armona et al., 2019; Bailey et al., 2019; Kuchler et al., 2023; Huseynov, 2026). In Florida, the pandemic-era housing boom and associated price surge may therefore have served as a convincing historical precedent for homeowners to expect another strong housing-market year, even against a backdrop of hurricanes, insurance stress, and cooling market fundamentals (Gupta et al., 2022; Mondragon and Wieland, 2022; Brueckner et al., 2023; Howard et al., 2023).

Entering 2025, the case for a cooldown in the Florida housing market was already evident. Redfin explicitly warned that climate risk, insurance costs, HOA fees, and property taxes could push down prices or slow growth in coastal Florida, predicting buyer's markets, and cooling in much of the South and Southeast (Fairweather and Zhao, 2024;

[2]Overall U.S. damages are estimated at roughly \$2.5 billion from Debby, \$78.7 billion from Helene, and \$34.3 billion from Milton; Florida bore much of Milton's cost (National Hurricane Center, 2026a).

[3]The Insurance Information Institute reports that Florida's statewide average premium rose from \$3,040 in 2022 to \$3,340 in 2023, although reported increases in 2024 were much smaller (Insurance Information Institute, 2024). Some homeowners were priced out of the insurance market, and reportedly around 15% – 20% of Florida homeowners became uninsured (Miami Herald Editorial Board, 2024). Recent work also confirms that state-specific regulations can drastically change local home insurance markets (Oh et al., 2026).

[4]The U.S. Senate Budget Committee opened an inquiry into Citizens' long-run solvency and possible federal-bailout risk, citing Florida's growing exposure to climate-related property losses and Citizens' expanding market share (U.S. Senate Committee on the Budget, 2023, 2024). Florida law allows a Citizens deficit to trigger a surcharge of up to 15 percent on Citizens policyholders (Florida Senate, 2024; Citizens Property Insurance Corporation, 2023). The Florida Hurricane Catastrophe Fund's May 2025 report estimated that it had the capacity to cover claims up to \$17.77 billion, and \$17.0 billion of this capacity was statutory obligation (Florida Hurricane Catastrophe Fund, 2025).

Brissey, 2024). Florida homeowners were therefore between Scylla (cooling housing market) and Charybdis (increasing hurricane and flood risk). This tension makes 2025 a useful environment to study to what extent homeowners' price expectations incorporate either of these factors, whether information can correct those beliefs, and whether any belief corrections persist.

Using an online survey experiment, our main study recruits Florida homeowners ($N = 648$) during the 2025 Atlantic hurricane season and elicits their house price expectations. Each respondent predicts the year-over-year price growth for their own ZIP code in January 2026. Our incentivized belief-elicitation protocol recovers the full probability distribution of these expectations, as respondents allocate 100 probability points across eleven ordered potential house price change bins under a binarized quadratic scoring rule.

We then randomize respondents to *Baseline*, *Baseline+Natural*, or *Baseline+Economic* information treatments. The Baseline condition presents general facts about the Florida housing market, noting that price movements are shaped by the interaction of market forces and natural events. Importantly, this arm does not suggest any directional effect of either force. The *Natural* arm adds two sentences to the Baseline text, communicating the increased frequency of hurricanes and elevated flood risk in Florida. In the same vein, the *Economic* arm adds two sentences noting a predicted cooldown in the state's housing market. We re-elicit house-price expectations after respondents read their randomly assigned information condition. The design therefore yields *prior* and *posterior* beliefs with respect to information.

We also conduct a December follow-up study recruiting from the same sample ($N = 337$) after the end of hurricane season and elicit revised price expectations for the same target month. This final elicitation allows us to measure post-hurricane-season posterior beliefs and to test whether the information treatments remain detectable in price beliefs.

We find that prior house-price expectations are significantly optimistic. On average, homeowners assign 61.2 probability points to a price increase in their own ZIP code.

However, among the 642 respondent-ZIP outcomes, 622 experienced a price decline in January 2026 relative to the same month of the previous year.

For each participant, we construct price-prediction benchmarks using Zillow market data available by the date of study participation.[5] Participants' prior beliefs are considerably more optimistic than both the benchmark predictions and the realized market outcomes. Both comparisons indicate that respondents entered the experiment with substantial optimism.

Our analyses consistently show that recent trailing ZIP-level house-price appreciation is a primary predictor of prior optimism, consistent with the formation of extrapolative expectations. We also detect some associations between recent weather events and prior expectations, but these effects vanish once we control for the full set of potential predictors. Homeowner optimism is therefore systematically related to recent market momentum.

We detect correction in posterior beliefs toward our benchmarks, indicating that simple information treatments can help homeowners update price expectations in the right direction. Relative to the Baseline condition, the effects on the probability assigned to a price increase are -11.10 and -6.73 percentage points for the Economic and Natural information treatments, respectively. A direct equality test rejects equal probability responses across the two treatments. The Economic message, therefore, produces a larger immediate response than the Natural message, although the experiment does not isolate which difference in their content produces that gap.

Treatments also affect the belief distribution, as probability mass shifts out of every price-increase bin and accumulates in the no-change and price-decline bins. Posterior expectations also move toward the forecast benchmarks constructed from available market data, and improve in ranked probability scores relative to realized prices after the information treatments. These effects are economically meaningful because they move

[5] We use Zillow's May 2026 record. But it is possible that the current dataset contains revisions to historical observations. We therefore treat the benchmarks as date-truncated rather than strictly real-time.

the full reported belief distribution. However, posterior expectations still remain far from the benchmarks because prior miscalibration is substantial.

Despite substantial belief corrections, we do not detect corresponding effects on reported property-selling intentions. The information treatments also do not detectably shift reported home-insurance premium tolerance or modeled insurance decisions. These findings are consistent with recent findings in the literature and invite further scrutiny of the relationship between market beliefs and actions (Chopra et al., 2025).[6]

We conduct a December follow-up after the hurricane season and recruit 337 respondents from the previous waves. Because the follow-up includes only a sub-sample of the main-study respondents, comparing December posteriors across study arms could be affected by self-selection bias. One concern is that respondents who return to the follow-up may differ in unobserved ways from those who do not. To reduce this concern, we estimate December treatment effects using prior-adjusted specifications that control for main-study priors. The prior-adjusted treatment effects are -0.18 points for the Economic treatment and -1.66 points for the Natural treatment, respectively. These findings suggest that the immediate information-treatment effects in the main study are no longer detectable in the follow-up sample.

Because the follow-up sample is smaller and the main-study treatment effects move in the same direction, we also estimate a pooled-treatment effect. The pooled December effect is -0.96 points. For the Economic treatment, the Anderson–Rubin confidence set rejects full persistence, but we cannot reject zero persistence. For the Natural treatment, the immediate effect is weak, and the confidence set includes both zero and one. We therefore conclude that the full Economic-treatment effect did not persist, while the Natural-treatment evidence does not distinguish between no-persistence and full persis-

[6] An important contrast with our results was documented by Fairweather et al. (2024), showing that personalized flood-risk information can change market search, property bidding, and equilibrium prices. Their evidence suggests that tailored information is more effective at influencing market actions through belief changes.

tence.

We also ask respondents to report their insurance-premium ceiling in the main study. The stated annual premium tolerance is high and right-skewed. In particular, 82.9% state a premium level at or above the 90th percentile of the 2024 national insurance-premium distribution reported by Keys and Mulder (2024). We interpret this finding as a stated home-insurance premium surplus rather than as revealed demand.

We then construct a structural model to ask whether more accurate information about hurricane-damage risk can change respondents' insurance choices. The model first calculates whether each respondent would prefer insurance without additional information. It then exposes the respondent to one of two hypothetical signals: a warning signal that raises the posterior probability of damage, or a reassurance signal that lowers it. We find that 42.4–43.8% of the sample would change their insurance choice across signal realizations, while 56.2–57.6% would always buy or always decline. The latter group has zero calibrated information value. These findings suggest that the same information can change decisions for households near the insurance threshold while leaving the choice unchanged for the rest. Put differently, policy effectiveness depends on reaching households near a decision boundary.

Our paper makes three contributions. First, it elicits homeowners' full house-price belief distributions in a geography with salient climate and extreme-weather risks, and on the verge of a market cooldown. These behavioral data complement studies using transaction-level data and reveal belief patterns that prices alone leave latent. Our randomized information treatments show that minimal communication can immediately improve price expectations. However, the structural insurance exercise shows that shifting beliefs may still be insufficient to induce decision changes for a substantial share of the sample.

Second, we go beyond typical information-treatment studies and investigate whether communication lingers in price expectations. In this dimension, our study speaks to information experiments showing that households revise their beliefs when presented with relevant facts (Coibion et al., 2022; Haaland et al., 2023; Fuster and Zafar, 2023; Binder et al., 2026), as well as studies that find that belief updates can exhibit partial persistence at later horizons (Cavallo et al., 2017; Armona et al., 2019). In our follow-up, we reject full persistence but cannot reject zero persistence for the Economic treatment. Consistent with recent developments in the literature, we caution future studies and policymakers to pay close attention to the persistence of information campaigns (Burro et al., 2026).

Third, our paper offers an extensive set of measurements and disciplined analyses for information-treatment studies. We pair persistence ratios with weak-identification-robust test inversion and address attrition concerns with selection bounds, reweighting, and imputation. These analyses provide a blueprint for researchers and practitioners interested in studying latent belief factors that shape market decisions.

Taken together, our results suggest that policy communication can shift homeowners' price beliefs, but such belief changes do not guarantee durable learning or changes in market behavior. If belief corrections disappear before owners make market decisions, even accurate information may have limited market effects. Effective policy, therefore, requires more than making looming risks salient once. Information campaigns are more likely to be effective if they are repeated, tailored to the decision time, and paired with tools that enable households to act.

2 Conceptual Framework

In this section, we outline crucial measures and the concepts behind them. We first discuss the concept of belief persistence and how we construct behavioral measures to identify it. Then we discuss the insurance framework, which is the core of our calibration exercises.

2.1 Belief Persistence

Although information can immediately change beliefs, its impact on action depends on how persistent the change is. Even when beliefs retreat to their initial state, it is important to examine whether this is consistent with forgetting or with other factors, such as temporary reporting error. Our framework provides a way to study these possibilities without claiming that they can be fully separated.

Let Y_{it} denote respondent i's reported probability that home prices in their ZIP code will increase. We observe the prior at $t = 0$, the immediate posterior at $t = 1$, and the December report at $t = 2$. Treatment assignment is denoted by $a \in \{B, E, N\}$, where B is the Baseline arm, E the Economic treatment, and N the Natural treatment. For the immediate elicitation, we define

$$\Delta Y_{i1}(a) = Y_{i1}(a) - Y_{i0}.$$

The full-sample immediate effect of treatment $g \in \{E, N\}$ is

$$\delta_{1g}^{F} = E[\Delta Y_{i1}(g) - \Delta Y_{i1}(B)]\,. \tag{1}$$

Our random treatment assignments identify this effect in the main study. Measuring persistence requires a separate comparison because December beliefs are observed only among respondents who complete the follow-up. Let $R_i = 1$ indicate follow-up response. Within that sample, define

$$\delta_{1g}^{R} = E[\Delta Y_{i1}(g) - \Delta Y_{i1}(B) \mid R_i = 1]\,, \tag{2}$$

$$\delta_{2g}^{R} = E[Y_{i2}(g) - Y_{i2}(B) \mid R_i = 1]\,. \tag{3}$$

The first expression re-estimates the immediate effect among the same respondents who participate in the December follow-up.

The main study random treatment assignments identify δ_{1g}^F without depending on subjects' future participation in the follow-up. However, identifying lingering treatment effects in December requires additional assumptions about follow-up participation. Section 4.5 therefore examines response balance and reports selection bounds, weighting, and multiple-imputation results to assess how selective response may affect the December estimates.

We measure persistence as

$$\theta_g = \frac{\delta_{2g}^R}{\delta_{1g}^R}. \tag{4}$$

A value of one means that the December treatment contrast equals the treatment contrast of the same follow-up sample in the main study. Using the same respondents places the two effects on a common sample, but the immediate effect can still be imprecisely estimated. A value of zero means that no treatment–Baseline difference remains in the follow-up study. Values between zero and one describe partial persistence.

Since the main-study treatment effect of the follow-up sample is in the denominator of the persistence ratio, its imprecision makes the ratio unstable. Therefore, employing delta-method standard errors is problematic in our case. So, we apply the transformation of

$$\delta_{2g}^R - \theta\delta_{1g}^R = 0.$$

We invert this test to construct an Anderson–Rubin confidence set, which remains valid even when the main-study treatment effect is imprecisely estimated.

We also define belief persistence for a pooled treatment effect that combines the Economic and Natural treatments. This approach compares the average effect of assignment to either treatment with the Baseline arm. Let δ_{tP}^R denote the pooled coefficient at horizon

t, and define

$$\theta_P = \frac{\delta^R_{2P}}{\delta^R_{1P}}.$$

The pooled persistence ratio, therefore, measures how much of the average effect of assignment to either treatment persists at the December follow-up.

Belief has economic value when it persists until decision time. Suppose an action is taken at t^*, and its expected payoff depends on the latent belief μ_{it^*}. Then, the treatment effect is economically relevant if it still remains at t^*. Thus, we go beyond simply identifying whether treatment effects on beliefs vanish in December and examine potential reasons for this. We write the observed belief as

$$Y_{it} = \mu_{it} + e_{it}, \tag{5}$$

where μ_{it} is the underlying latent belief relevant for decisions and e_{it} is transitory reporting error. Suppose that a treatment reduces μ_{it}. At the same time, the main study posterior also contains a negative realization of e_{i1}. Notice that the follow-up posterior will move upward when that reporting error reverts back to its mean value, even though the underlying latent belief stays the same. Therefore, an observed posterior rebound in the December follow-up may combine both belief and mechanical revisions.

We use the Baseline arm to estimate how much of an immediate belief update ordinarily remains in December and then apply its implied reversion rate to the main-study treatment effects. Because the Baseline arm contains no appended treatment message, this relationship provides a benchmark for ordinary movement between main and follow-up elicitations.[7]

Finally, treatment messages could affect beliefs by shifting probability mass coherently toward small price-decline bins, or by rearranging mass in ways that keep the mean value

[7] We assume that the primary force behind the Baseline movement is a reporting error in the main study. This interpretation also requires this transitory reporting error to behave similarly across all three arms.

constant. We therefore estimate treatment effects on all eleven elicited price change bins and verify that the bin-level effects satisfy the adding-up restriction.

2.2 Insurance Decisions and the Value of a Signal

Our experiment did not randomize insurance prices, offer respondents an insurance contract, or observe whether they purchased coverage. However, our setting allows us to examine when natural-hazard information affects insurance decisions. Specifically, we use a stylized insurance problem to ask when information about hurricane damage could change insurance choice. Our calibration exercise places the elicited beliefs on a decision scale, without needing a full-scale insurance demand analysis.

Let d_i denote respondent i's reported probability that their home will be damaged by hurricanes or related natural disasters by the end of 2025. Our model uses this response directly as the probability of damage over its decision horizon.[8]

Respondents also reported the percentage of their home's value that "could be lost in the next hurricane or related natural disasters" during 2025. Let l_i denote this potential loss as a share of home value H_i.[9]

We do not observe respondents' total wealth. To put the insurance choice on a common scale, we model wealth as

$$W_i = \lambda H_i,$$

where H_i is home value and λ is the ratio of total wealth to home value. We do not estimate λ. Instead, we vary it across the calibration values and ask whether the decision margin is sensitive to that assumption.

Let c_i denote the annual full-insurance premium as a share of home value. In the main

[8]Because our study was fielded during the hurricane season, our elicited damage probabilities cover less than a full year, while the benchmark premium is annual. In this sense, our reported extensive margin is conservative relative to an otherwise comparable annualized-hazard exercise.

[9]Our survey question does not explicitly condition on the home being damaged. For the insurance exercise, our modeling assumption is to treat l_i as the loss share in the damage state.

calibration, H_i is the midpoint of respondent i's reported home-value interval. We use the national benchmark as the insurance premium amount.

In our setting, the homeowner chooses between no insurance and full insurance. Full insurance requires paying $c_i H_i$ and ensures the property's value is preserved after the hurricane season.[10] After dividing utility by the common factor $\sqrt{H_i}$, the utility from full insurance is

$$U_i^I = \sqrt{\lambda - c_i}.$$

Without insurance, the expected utility at damage probability d_i is

$$U_i^N(d_i) = (1 - d_i)\sqrt{\lambda} + d_i\sqrt{\lambda - l_i}.$$

Under the square-root utility, full insurance is preferred when

$$\sqrt{\lambda - c_i} \geq (1 - d_i)\sqrt{\lambda} + d_i\sqrt{\lambda - l_i}. \tag{6}$$

Solving this condition for the largest premium the owner would accept gives the model-implied reservation-premium share

$$c_i^* = \lambda - \left[(1 - d_i)\sqrt{\lambda} + d_i\sqrt{\lambda - l_i}\right]^2. \tag{7}$$

One should not conflate the model-implied cutoff with the premium tolerance subjects report in the survey. Let w_i denote the percentage of home value that respondent i stated they would be willing to pay each year as an insurance premium. Comparing w_i with c_i^* reveals whether the stated response is broadly consistent with the model. Therefore, we treat the recovery of c_i^* as a model-coherence diagnostic test.

We next introduce information. The model gives the owner one of two possible real-

[10] To make the exercise simple, we do not account for deductibles, coverage limits, exclusions, partial insurance, or other features of typical real-life insurance contracts.

izations of a hypothetical damage-risk signal. A *warning* indicates that damage is more likely, while a *reassurance* indicates that it is less likely.[11]

Let s_i denote signal sensitivity, the probability of receiving a warning when damage occurs, and let t_i denote specificity, the probability of receiving reassurance when damage does not occur. The unconditional probabilities of warning and reassurance are

$$q_{iW} = s_i d_i + (1 - t_i)(1 - d_i), \tag{8}$$

$$q_{iR} = (1 - s_i) d_i + t_i (1 - d_i). \tag{9}$$

After each realization, Bayes' rule gives

$$d_{iW} = \frac{s_i d_i}{q_{iW}}, \qquad d_{iR} = \frac{(1 - s_i) d_i}{q_{iR}}, \tag{10}$$

with the usual limiting convention when a signal has zero probability.

Without additional information, the owner chooses insurance once using the reported damage probability:

$$V_i^0 = \max\left\{U_i^I, U_i^N(d_i)\right\}.$$

With information, the owner can reconsider the choice after observing the signal:

$$V_i^1 = q_{iW} \max\left\{U_i^I, U_i^N(d_{iW})\right\} + q_{iR} \max\left\{U_i^I, U_i^N(d_{iR})\right\}.$$

The signal is actionable when the optimal insurance decision differs between warning and reassurance. In that case, the owner can insure after the warning and remain uninsured after the reassurance. If the same action is optimal after both realizations of the signal, the information cannot improve this binary decision. Therefore, by the law of iterated expectations, $V_i^1 = V_i^0$, and the value of information is exactly zero.

[11] Note that these are hypothetical signals used in the calibration and should not be conflated with the Economic and Natural treatments randomized in our experiment.

The main calibration sets sensitivity and specificity equal to 0.85. Note that this is an imposed symmetric benchmark, rather than an estimate of forecasting performance. However, our survey separately asks respondents how accurate they believe weather forecasting services are in their ZIP code. This measure does not reveal sensitivity and specificity separately. In the robustness checks, we treat reported accuracy a_i only as the constraint

$$a_i = s_i d_i + t_i(1 - d_i)$$

and consider the feasible sensitivity–specificity pairs that yield the same accuracy rate. Put differently, we assess a set of sensitivity–specificity pairs that coincide for each reported accuracy belief. This distinction allows us to examine signal pairs with the same overall reported accuracy that differ in their impact on insurance decisions.

Finally, we express the value of information in monetary terms. Let κ_i be the additional wealth, as a share of home value, that would make the owner without information as well off as the owner who receives the signal. We add $\kappa_i H_i$ to wealth in every state and allow the no-information insurance choice to be made again:

$$V_i^0(\kappa_i) = \max\left\{\sqrt{\lambda + \kappa_i - c_i}, (1 - d_i)\sqrt{\lambda + \kappa_i} + d_i\sqrt{\lambda + \kappa_i - l_i}\right\}.$$

The monetary-equivalent value solves

$$V_i^0(\kappa_i) = V_i^1.$$

Thus, κ_i is the wealth equivalent of access to the hypothetical damage-risk signal under square-root utility, full insurance, modeled wealth $W_i = \lambda H_i$, the specified premium, and the specified signal structure. Under these assumptions, the exercise shows for whom information can change an insurance decision and how large the resulting decision value is.

3 Experimental Design and Data

3.1 Study protocols, timeline, and sample

Our study borrows the primary building blocks from Huseynov (2026). We build the study in a Qualtrics survey module and recruit subjects through Prolific.com, an online crowdsourcing platform that provides access to homeowners.[12]

Prolific collects data on key individual characteristics before participants join the platform and makes these characteristics available as filters. We publish the study only to Florida residents who answered "I own the property I live in" to the platform's property-ownership question.[13] We also restrict the sample to participants with approval ratings of at least 80% in previous studies. This approval rate measures the share of participants' past study completions deemed appropriate by researchers.

Main Study: The main study is fielded in early July and closes on November 30, covering the Atlantic hurricane season.[14] Subjects spend around 10 minutes completing the study and are compensated at approximately a $22.00 hourly rate.

The study begins with an information page explaining general procedures and compensation. Subjects can proceed only after recording consent using the provided radio button. We also use Qualtrics's built-in bot-screening CAPTCHA function to screen out non-human entries.

The first survey block elicits key property and home-financing characteristics, including the respondent's home ZIP code. We then explain that subjects can earn a $2.00 bonus

[12]We obtained approval from the Auburn University IRB under application $N23 - 662$ EX 2401. We pre-registered the study under $AEARCTR - 0016346$ in the AEA RCT Registry. Screenshots of key experimental procedures can be accessed here: `https://samirhuseynov.com/research/fielded_survey_instrument.pdf`

[13]The platform regularly repeats surveys, allowing users to update responses to filtering questions. Previous work documents the reliability of Prolific filters (Exley and Nielsen, 2024).

[14]We paused the data collection in August and September. This decision was made based on low study take-up rates towards the end of July. We expected the active user base to increase in the second half of the hurricane season. Thus, we resumed our data collection in October. We conduct several robustness tests of our primary findings, in which the month effect is modeled as a wave fixed effect.

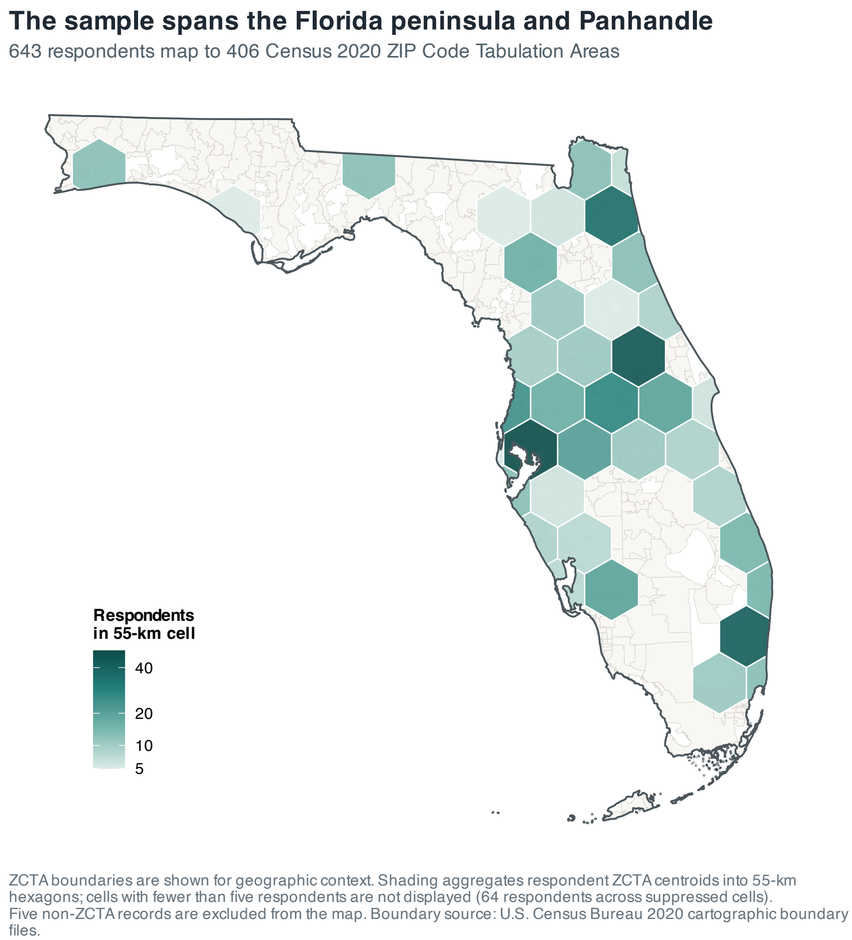

Figure 1: Florida sample geography. *Notes:* The map depicts 643 respondents across 406 Census 2020 ZIP Code Tabulation Areas. Five reported ZIP codes are valid but do not match any Census ZCTA. Hence, we exclude them. To protect respondent locations, we place ZCTA centroids into 55-kilometer hexagons and show only cells containing at least five respondents. This rule suppresses 64 respondents. State and county boundaries come from the U.S. Census Bureau (U.S. Census Bureau, 2021).

based on the accuracy of their reported price expectations.[15] We describe how the price-expectation task works and how the bonus is calculated.

Subjects are told that they will provide expectations for the January 2026 year-over-year average house-price growth in their ZIP code. We explicitly define this as the January 2026 average price level relative to January 2025. Subjects also see an example showing

[15] The accuracy bonus is meaningful relative to the platform's average hourly earnings. Previous work shows that elicited behavior is largely insensitive to the size of a monetary stake once a salient bonus structure is implemented (Camerer and Hogarth, 1999; Enke et al., 2023).

the price-change bins and the bonus calculation.

Specifically, we present 11 potential house-price-change bins, ordered in the data from the largest increase to the largest decline: {At least 4.00%}, {3.00% to 3.99%}, {2.00% to 2.99%}, {1.00% to 1.99%}, {0.01% to 0.99%}, {0.00%}, {−0.01% to −0.99%}, {−1.00% to −1.99%}, {−2.00% to −2.99%}, {−3.00% to −3.99%}, and {At most −4.00%}. Homeowners are endowed with 100 tokens each representing one probability point, and allocate these tokens across intervals to report their probabilistic beliefs about the year-over-year ZIP code house-price growth in January 2026.

We incentivize price expectations using a binarized quadratic scoring rule (Hossain and Okui, 2013):

$$\pi_{\text{lottery}} = 1 - \left(\mathbb{1}\{\text{event}\} - p_{\text{belief}}\right)^2 .$$

After January 2026, the experimenter randomly selects one price interval and determines whether the actual house-price change in the homeowner's ZIP code reported by Zillow falls into that interval. The event indicator equals one if the selected interval contains the realized price change and zero otherwise. The belief assigned to the selected interval is

$$p_{\text{belief}} = \frac{\text{Number of tokens in the selected interval}}{100}.$$

Each homeowner's bonus payoff is then determined by a lottery that grants a π_{lottery} chance of earning \$2.00.

Participants are informed that bonus rewards will be realized after January 2026, once market data for each ZIP code are available. Subjects know that they will provide beliefs twice and that the bonus will be based on a randomly selected belief report. It is therefore in their best interest to truthfully report expectations in both tasks.

Following Huseynov (2026), we also provided subjects with an online calculator that allowed them to allocate probability points and view their eventual chances of winning the bonus award across different realizations of the actual price growth. Subjects proceed to the actual price-belief elicitation only after correctly answering a quiz about the incentive structure.

Our main study recruits 648 Florida homeowners: 441 respondents in July 2025, 161 in October, and 46 in November. Our sample includes respondents from 411 ZIP code clusters and 57 counties. We show their ZIP code in Figure 1. Our subject pool represents both coastal and inland regions, allowing us to observe how differential market and natural conditions affect both price expectation priors and belief-updating.

Figure 2 depicts the response dates against the finalized 2025 Atlantic storm records. During our data collection, no named storm made landfall in Florida, although Imelda generated a tropical-storm watch for part of the state on September 27–28.

Follow-up Study: The follow-up began after the season closed and continued through the last day of December. We recruited only subjects from the main study, as Prolific allows researchers to publish study requests targeting samples from previous studies. In total, we received 341 completed follow-up submissions. One Prolific ID appeared in three submissions. We retained the latest complete submission by this subject. This left 339 distinct follow-up identifiers. Two subjects' Prolific IDs could not be matched to the main study data, leaving us with a linked sample of 337 respondents.

In the follow-up study, we reminded participants of the incentive structure and belief-elicitation protocol. We also provided the calculator to refresh their memory of how bonus payments worked and repeated the quiz from the main study. After re-eliciting price beliefs, we allowed subjects to choose the follow-up report to be binding for the bonus calculation.

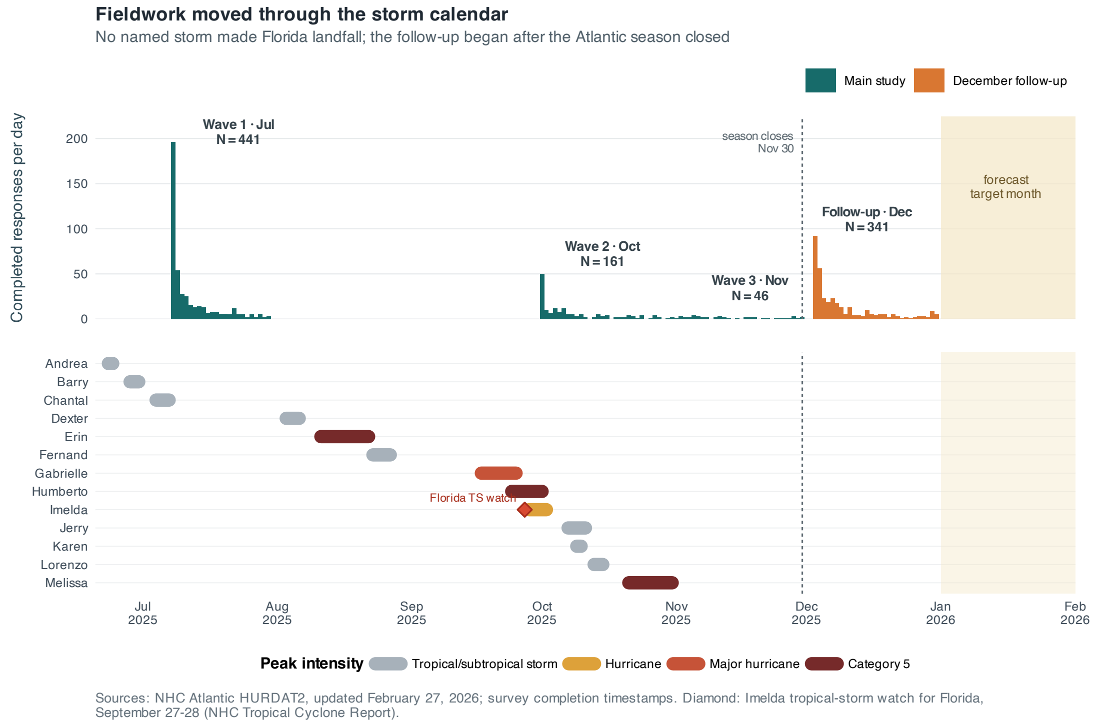

Figure 2: Data collection and the 2025 Atlantic storm calendar. *Notes:* Bars show the number of completed responses across study days. Storm lanes reflect the finalized National Hurricane Center HURDAT2 dates, and colors indicate peak intensity. The 2025 record contains 13 named storms, five hurricanes, four major hurricanes, and three Category 5 hurricanes (National Hurricane Center, 2026b). The diamond marks the Florida tropical-storm watch associated with Imelda on September 27–28 (Reinhart, 2026). No named storm made landfall in Florida during our fieldwork. The shaded January 2026 interval is the month for which respondents predicted house-price growth.

3.2 Treatments

After the prior elicitation, respondents were assigned by computer at the individual level to one of three information treatment screens. All screens began with the same text:

> "Florida has the second-largest residential real estate market in the United States, with a total market value close to \$3.62 trillion. Complex interaction of economic and natural factors eventually determines the movements in Florida house prices."

The *Baseline* condition ($N = 226$) only contained the above text. The *Baseline+Economic* study arm ($N = 212$) additionally provided: "Housing market is shifting to a healthier balance of buyers and sellers in Florida. This, in turn, is leading to a market cooldown." In turn, the *Baseline+Natural* treatment ($N = 210$) instead read: "Steady increases in ocean temperatures lead to more frequent hurricanes in Florida. This, in turn, is driving up flood risk as well." We enforced that subjects spend at least 20 seconds on the information screen.

Our identification is based on each information message's (Economic and Natural) total effect relative to the Baseline text. We should note that Economic and Natural information texts differ in subject matter and potential mechanisms that might affect house prices. Accordingly, the economic-minus-natural contrast is informative about the relative performance of two communication packages.

The Baseline condition provides common information without suggesting any directional price movements. For this condition, the average magnitude of the belief updating about the probability of a price increase is 2.57 points when we compare posteriors and prior beliefs. This suggests that posterior-prior contrasts for each study arm are not the treatment effect. We therefore identify the incremental effect of each appended treatment message relative to the Baseline condition.

3.3 Other controls

We use several data sources to construct ZIP code-level controls. We derive historical market price growth data from Zillow. We derive weather records from Ag Data, and also county-level 2024 flood insurance claims and associated variables from the FEMA National Flood Insurance Program (NFIP).[16]

[16] Ag Data was derived from https://www.aaronsmithagecon.com/us-county-weather, and FEMA data ZIP files were frozen on July 1, 2025, and downloaded via API access. Missing demographic values are not recoded as observed categories. Five reported ZIP codes are not represented by Census 2020 ZCTAs. They remain in randomized survey-outcome analyses, but we exclude them from the geography figure and location-dependent summaries.

3.4 Balance and inference

For each study condition, the Online Appendix reports the means for selected key variables. We also test if the main study waves (July, October, and November) display any covariate imbalance. A within-wave permutation omnibus test does not reject joint balance (statistic 0.552, $p = 0.162$, 9,999 permutations), and study arms do not differ across waves (χ^2 $p = 0.989$). Age is the largest individual imbalance: the Natural treatment arm is 4.6 years younger than the Baseline arm, with a joint test across study arms of $p = 0.002$. In our primary analyses, controlling for age and other variables does not affect our conclusions. Our primary treatment analyses also report within-wave randomization inference, a null-imposed Rademacher wild cluster bootstrap, and Romano–Wolf step-down adjustment, each with 9,999 draws (Young, 2019; Cameron et al., 2008; Roodman et al., 2019; Romano and Wolf, 2005).

4 Results

4.1 Optimistic Priors

Respondents assigned an average of 61.2 probability points to a price increase in their prior expectations and reported a mean expected growth rate of 0.8 percent. Over the January 2025–January 2026 window, the corresponding respondent-ZIP observations instead declined in 622 of 642 cases based on Zillow's market data, with a mean growth rate of -4.6%. Therefore, we document optimistic prior beliefs, both in sign and magnitude, which ex-post the market data substantially refute.

The priors provide descriptive insights into how different individual- and ZIP-code-level characteristics are associated with price expectations. The Online Appendix examines whether weather and NFIP history, ZIP market history, property characteristics, and personal characteristics, both individually and jointly, affect priors. We find that a one-standard-deviation increase in trailing twelve-month ZIP price appreciation predicts 8.0 additional probability points for price increase beliefs. This estimate falls to 5.6 points when we control for all controls. We also find that recent property purchasers are more optimistic, that owners of recently built properties are less optimistic, and that self-reported property value is positively associated with optimism. Of these three property associations, only purchase recency survives after statistical adjustments for multiple hypothesis testing in the full specification.

However, these associations have limited predictive content. On the common sample of 598 respondents, all controls explain $R^2 = 0.133$ (adjusted $R^2 = 0.063$), and the full model's ZIP-grouped cross-validated R^2 is -0.050. After adjusting for multiple hypothesis testing, the individual weather and NFIP associations do not remain statistically significant.

We construct several control variables for rainfall amount. A one-standard-deviation increase in June 2025 rainfall is associated with -5.33 points less prior increase mass and -4.65 points after adding price-history and property controls. With the addition of county fixed effects, our estimate remains negative (-4.80, SE 3.35) but still insignificant ($p = 0.153$). Moreover, our constructed 7–365-day rainfall lag controls do not predict stated damage probability or expected loss after we adjust for multiple hypothesis testing. We therefore treat rainfall-related analyses as suggestive and descriptive evidence that recent local experience may enter priors.

4.2 Forecast Benchmarks from Publication-Lag-Truncated Market History

To place respondents' priors on a market scale, we compare them with four simple statistical forecasts. We construct a random walk, trailing twelve-month momentum, ZIP-level AR(1), and their equal-weight mixture analyses. For each respondent, we truncate the

Zillow history at least 45 days before the survey date. No observation dated after that cutoff is included in our forecasts.[17]

Owner priors have a Root Mean Square Error (RMSE) of 6.18 percentage points and a mean error of +5.38 points. Raw RMSE ranges from 1.69 for momentum to 3.39 for the random walk. Our comparison is potentially uneven because probabilistic price expectation means are bounded by the elicitation support, while benchmark point forecasts are not. To address this potential concern, in the Online Appendix, we censor all point forecasts to $[-4.5, 4.5]$ and map each forecast to an elicitation bin. Even after the adjustment, benchmarks' censored RMSEs range from 2.56 to 3.53, compared to 6.18 for owners' priors, and their deterministic point-bin ranked probability scores range from 0.74 to 1.82, compared with 5.07 for owner priors. Our findings are robust to different truncation thresholds in the Zillow data.

Figure 3 summarizes the raw error comparison and lag sensitivity. Our benchmarks are deliberately simple, as their role is to show to what extent our treatments shift probabilistic beliefs toward forecasts based on market observations.

4.3 Belief Updating

4.3.1 Average treatment effects

How much did the two treatments change respondents' beliefs? Table 1 reports two complementary specifications. The first uses the change in beliefs between the prior and the immediate posterior. The second estimates treatment effects on the posterior while controlling for the respondent's prior. The two specifications produce very similar results.

[17] This timing rule does not, however, necessarily recreate the exact Zillow market information that was available in 2025. We build our analyses on the May 2026 publication of the Zillow data, and Zillow can revise values assigned to earlier months. Our analysis is therefore a publication-lag-truncated comparison using current-vintage history.

We calculate monthly price changes using consecutive observations in the Zillow series. In four cases, however, Zillow does not report consecutive observations. Our primary analysis retains these observations, while a sensitivity analysis excludes them. The results remain similar when we require every price change to be measured across consecutive calendar months.

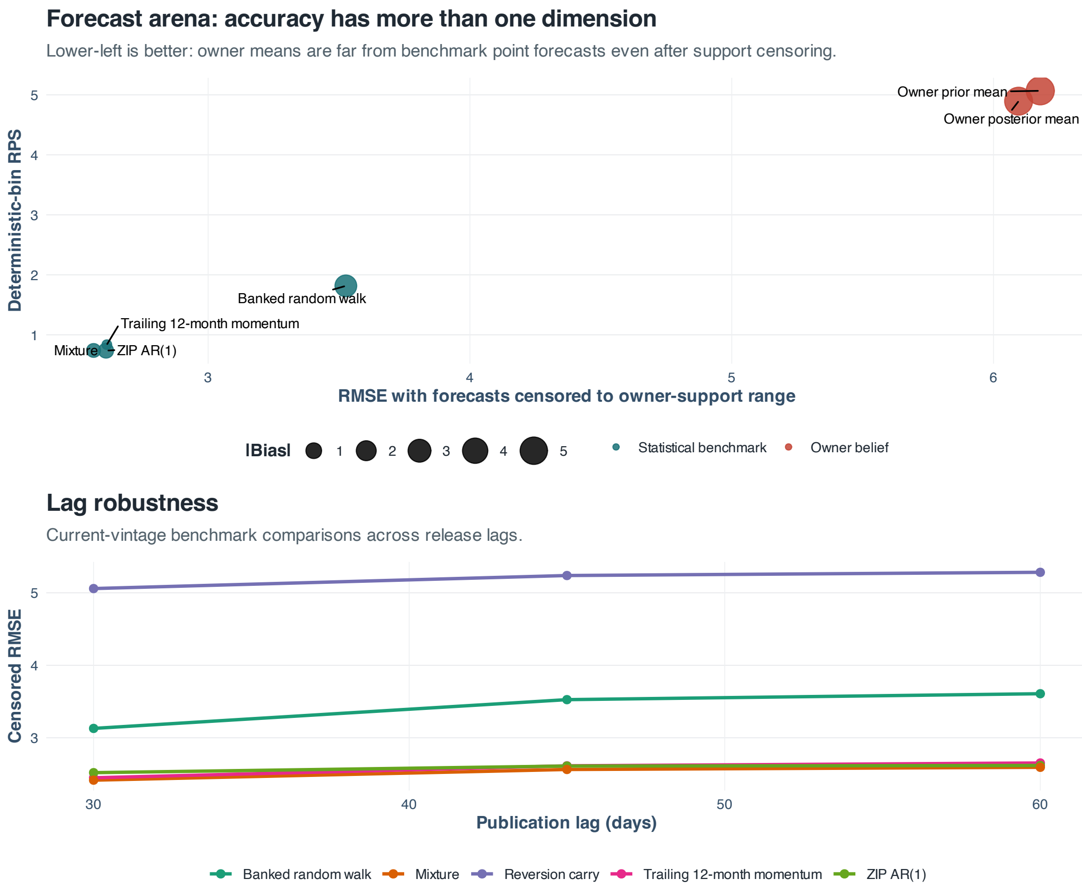

Figure 3: Owners and publication-lag-truncated statistical benchmarks. *Notes:* The upper panel compares homeowner beliefs with four statistical forecasts using two accuracy measures. Lower values indicate that forecasts are closer to the realized January 2025–January 2026 own-ZIP price growth. The lower panel shows how the benchmark Root Mean Square Error (RMSE) changes across different publication lags. The reversion-carry forecast applies the latest monthly price change observed at the publication cutoff through the survey month and uses the ZIP's long-run average monthly growth for the remaining months. The analysis includes 642 respondents due to Zillow's data coverage. We use the May 2026 Zillow market data.

The Economic treatment reduces the probability assigned to an increase in home prices by -11.10 percentage points (SE 1.99) in the change-score specification and by -11.45 points (SE 1.90) in the prior-adjusted specification. The corresponding effects of the Natural treatment are -6.73 (SE 1.75) and -7.29 points (SE 1.77). For both treatments and both specifications, the adjusted p-values are below 0.001, when we conduct randomization inference, the wild cluster bootstrap, and Romano–Wolf family adjustment for

Table 1: Immediate treatment effects on beliefs

	Effect	*p*-values				ZIP
	(CR1S SE)	RI	WCB	RW	N	clusters
Panel A. Probability assigned to a price increase (pp)						
Economic, change score	-11.10^{***} (1.99)	< 0.001	< 0.001	< 0.001	648	411
Natural, change score	-6.73^{***} (1.75)	< 0.001	< 0.001	< 0.001	648	411
Economic, prior-adjusted	-11.45^{***} (1.90)	< 0.001	< 0.001	< 0.001	648	411
Natural, prior-adjusted	-7.29^{***} (1.77)	< 0.001	< 0.001	< 0.001	648	411
Panel B. Treatment-effect equality: Economic − Natural						
Change score	-4.37^{**} (1.85)		CR1S $p = 0.018$		648	411
Prior-adjusted	-4.16^{**} (1.83)		CR1S $p = 0.024$		648	411
Panel C. Normalized entropy (dispersion; co-primary)						
Economic, change score	0.02 (0.95)	0.980	0.981	0.981	648	411
Natural, change score	−0.37 (0.88)	0.676	0.674	0.924	648	411
Economic, prior-adjusted	0.30 (0.94)	0.749	0.752	0.949	648	411
Natural, prior-adjusted	−0.14 (0.86)	0.867	0.867	0.981	648	411
Panel D. Baseline arm's own within-session update (descriptive)						
Increase probability (pp)	2.57 (1.33)		—		226	187
Entropy	0.32 (0.54)		—		226	187
Expected growth (pp)	0.17 (0.07)		—		226	187

Notes: Change-score models use the posterior minus the prior. Prior-adjusted models compare posteriors while controlling for the prior. All models include field-wave fixed effects and use CR1S standard errors clustered at the ZIP level. RI denotes within-wave randomization inference; WCB is the null-imposed Rademacher wild cluster bootstrap, and RW is the Romano–Wolf family adjustment (each approach uses 9,999 draws). Panel D reports the Baseline arm's own prior-to-posterior change. ($^{***}p < 0.01$, $^{**}p < 0.05$)

multiple hypothesis testing.

The Economic treatment's effect exceeds that of the Natural treatment by 4.37 percentage points ($p = 0.018$) and 4.16 percentage points ($p = 0.024$) in the change-score and prior-adjusted specification, respectively. Both treatments also shift the probability mass, but neither significantly changes the belief distribution's dispersion or entropy.

Interestingly, the posterior in the Baseline arm also moves with respect to its priors. We detect that the probability assigned to the price increase bins increases by 2.57 percentage points, even though the Baseline arm contains no directional message. This movement justifies our approach of estimating both treatment effects relative to the Base-

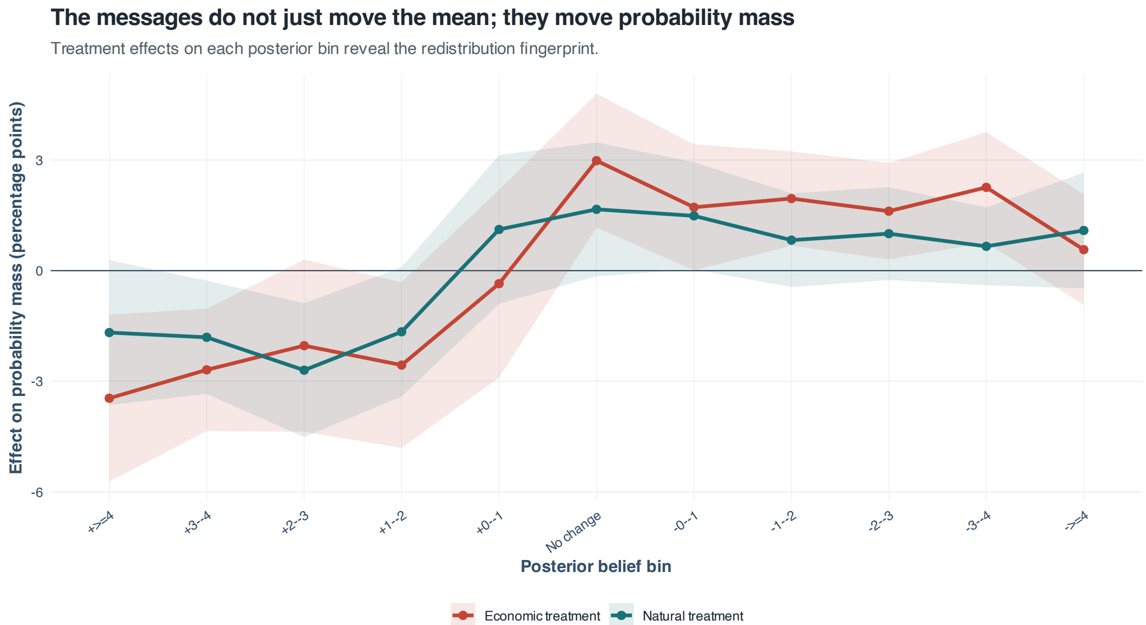

Figure 4: Treatment effects in the elicited distribution. *Notes:* Each line shows how a treatment changes the probability assigned to a price-growth bin relative to the Baseline arm. Negative values remove probability from a bin, while positive values add probability. Shaded areas are 95% confidence intervals based on ZIP-clustered CR1S standard errors. The models include all 648 respondents, 411 ZIP clusters, and field-wave fixed effects. Joint permutation tests reject no reallocation across the ten independent dimensions for both treatments (Economic: 65.772; Natural: 36.096; both $p < 0.001$).

line arm.

4.3.2 Distributional effects

The average treatment effect does not show how respondents rearrange the rest of their probability distribution. Figure 4 reports treatment effects separately for all eleven elicited price growth bins. Both treatments remove probability mass from the five bins representing price increases and move it toward the no change and price decline bins. The Economic treatment alone removes 3.46 percentage points from the most positive bin.[18] Our joint permutation tests over the ten independent dimensions reject the null that neither message reallocates probability mass.

[18] Because the probabilities assigned to the eleven bins must sum to one, the bin-level effects must sum to zero. Our construction satisfies this adding-up restriction.

4.3.3 How should we interpret the response?

Can these treatment effects be interpreted as the weight respondents place on a common signal? The Online Appendix estimates

$$P_i^1 = \alpha + \beta P_i^0 + \delta T_i + \gamma(T_i \times P_i^0) + u_i,$$

along with OLS, boundary-restricted, fractional-logit, and quantile specifications. Under Gaussian learning from a common signal, receiving additional information should reduce the posterior's dependence on the prior. This requires $\gamma/\beta \leq 0$, in which case $-\gamma/\beta$ can be interpreted as the additional weight placed on the signal (Coibion et al., 2022).

In our case, the estimated γ/β ratios are 0.109 and 0.164 in models with wave fixed effects, and 0.099 and 0.173 after adding property characteristics, demographics, lagged market conditions, weather, and NFIP controls. Since all four estimated ratios are positive, we do not interpret them as signal weights or measures of information precision. Estimated positive ratios suggest that the qualitative information in treatment messages does not behave like a single numerical signal.

We also examine whether belief updating differs across respondents' hurricane exposure and their local housing and insurance environments. We interact both treatments with four rainfall measures, FEMA Individual-Assistance designation, NFIP market conditions, lagged ZIP-level price growth and volatility, home value, mortgage status, tenure, and prior uncertainty. None of the 26 interactions survives the family- or global-Holm adjustment, with all global-adjusted p-values equal to one. We therefore do not detect any systematic effect of these controls on the price-belief updating process. However, the corresponding 80-percent MDEs range from roughly 3.4 to 12.5 percentage points, suggesting that our results do not establish that updating is homogeneous.

Table 2: Treatment effects on realized accuracy and on movement toward publication-lag-truncated benchmarks

	Economic		Natural	
Metric	Effect (SE)	RW p	Effect (SE)	RW p
Panel A. Ex-post realized accuracy (scored on realized own-ZIP growth)				
Mass on exact realized bin (pp)	0.43 (0.68)	0.679	−0.60 (0.69)	0.679
Mass on realized sign (pp)	6.80*** (1.78)	< 0.001	3.64* (1.64)	0.091
RPS improvement	0.53*** (0.11)	< 0.001	0.27** (0.11)	0.043
Panel B. Movement toward publication-lag-truncated benchmarks				
Distance to AR(1), improvement	0.43*** (0.10)	< 0.001	0.28*** (0.09)	0.008
Distance to random walk, improvement	0.44*** (0.08)	< 0.001	0.25*** (0.08)	0.006
Mass on AR(1)-implied sign (pp)	5.72*** (1.81)	0.007	4.19** (1.61)	0.042

Notes: Panel A asks whether the treatments improve accuracy relative to realized January 2025–January 2026 own-ZIP price growth. Panel B asks whether they move beliefs toward the AR(1) and random-walk forecasts constructed after excluding observations dated within 45 days of participation. The historical values come from the May 2026 Zillow record. Positive effects indicate improvement. RPS denotes the ranked probability score over the eleven ordered bins. All models include 642 respondents, 405 ZIP clusters, and field-wave fixed effects; CR1S standard errors appear in parentheses. RW p is the Romano–Wolf adjustment across fourteen accuracy and movement hypotheses using 9,999 null-imposed wild-bootstrap draws. The family contains the twelve contrasts displayed here and the two treatment effects on the BQSR payoff shown in the Online Appendix's accuracy figure. ($^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$).

4.4 Forecast Value of the Immediate Correction

Our treatments move respondents' beliefs downward. But is this move an improvement in price expectations? We construct benchmarks to answer this question.

Panel A of Table 2 employs a backward-looking approach and scores the main study posterior against realized January 2025–January 2026 home-price growth in each respondent's ZIP code. Panel B instead asks whether respondents move toward the publication-lag-truncated forecasts described in Section 4.2.

Moreover, Panel A measures the probability assigned to the realized sign and exact bin, together with the ranked probability score for the full distribution. Panel B measures whether beliefs move closer to the AR(1) and random-walk forecasts and whether respondents place more probability on the sign predicted by the AR(1) model.

Our evidence regarding improvements in belief is clearest for the Economic treatment. This treatment increases the probability assigned to the downward direction, in which home prices eventually moved, by 6.80 percentage points, and improves the ranked probability score by 0.53.

The Natural treatment also moves beliefs in the same direction, although the evidence for the belief improvement is relatively weaker. The Natural treatment increases the probability mass on the realized direction by 3.64 points (adjusted $p = 0.091$) and improves the ranked probability score by 0.27 (adjusted $p = 0.043$).

Our second inquiry is whether treatments increased the amount of token allocation in the price-change bin that eventually included the realized market outcome. We find that neither treatment measurably increases the probability assigned to that exact bin. The treatments, therefore, move beliefs in the correct direction without bringing them to the precise realized outcome.

The Online Appendix decomposes the within-arm ranked-score changes. Most of the improvement comes from better calibration rather than a stronger ability to distinguish ZIP codes with different outcomes. We also use a log-odds updating model to describe the size of the average shifts. Under that model, subjects behave *as if* the Economic treatment had precision 0.756, and the Natural treatment had precision 0.575. These results are aligned with our primary findings.

4.5 Persistence in the December Follow-Up

4.5.1 Follow-up sample and belief paths

In the main study, our treatments improved house-price beliefs, moving them closer to both the realized market outcomes and the publication-lag-truncated benchmarks. In December, we contacted the main study respondents to document how much of that improvement remained in beliefs. Of the 648 main study respondents, 337 (52%) participated in the follow-up study. Since follow-up participation was determined by subjects' availability in December and their willingness to participate, the follow-up sample may be selective. Below, we conduct a series of analyses to examine this issue.

Follow-up study response rates were 49.6% in the Baseline arm, 50.0% in the Economic treatment, and 56.7% in the Natural treatment. Relative to the Baseline arm, the regression-adjusted response-rate differences were 0.4 percentage points for the Economic treatment ($p = 0.931$) and 7.1 percentage points for the Natural treatment ($p = 0.136$). So, we do not find a statistically detectable relationship between treatment assignment and follow-up response. However, with only about half of the original sample returning, we cannot simply assume that the follow-up sample preserves the main study's random treatment assignments.

In the December survey, we also allowed respondents to replace both main study belief reports with their December price beliefs. We added this design feature to incentivize truthful responses. We document that 283 of 337 respondents (84.0 percent) chose to replace their earlier beliefs with the December report. Main study treatment assignment does not jointly predict replacement ($p = 0.947$), and replacers are not more accurate in December after controlling for their earlier score and field wave ($p = 0.858$). The high replacement rate suggests that respondents took the follow-up decision seriously.

The time between the main study and the December follow-up differs across field waves. We therefore measure persistence at the observed December horizon rather than estimating the causal effect of waiting an additional month.

Our first analysis examines the main-study treatment effects in our follow-up study sub-sample. This exercise allows us to reveal to what extent the potentially self-selected follow-up sample reacted to the main study treatments. Figure 5 shows the belief paths between the main and follow-up studies. In the Economic treatment, the average probability assigned to a price increase falls from 60.7 before treatment to 52.6 immediately afterward, then returns to 58.9 in December. In the Natural treatment, it moves from 60.4 to 57.3 and remains at 57.2. The Baseline arm shows a different path, rising from 66.0 to 67.7 within the initial survey and then falling to 61.0 in December. Therefore, the

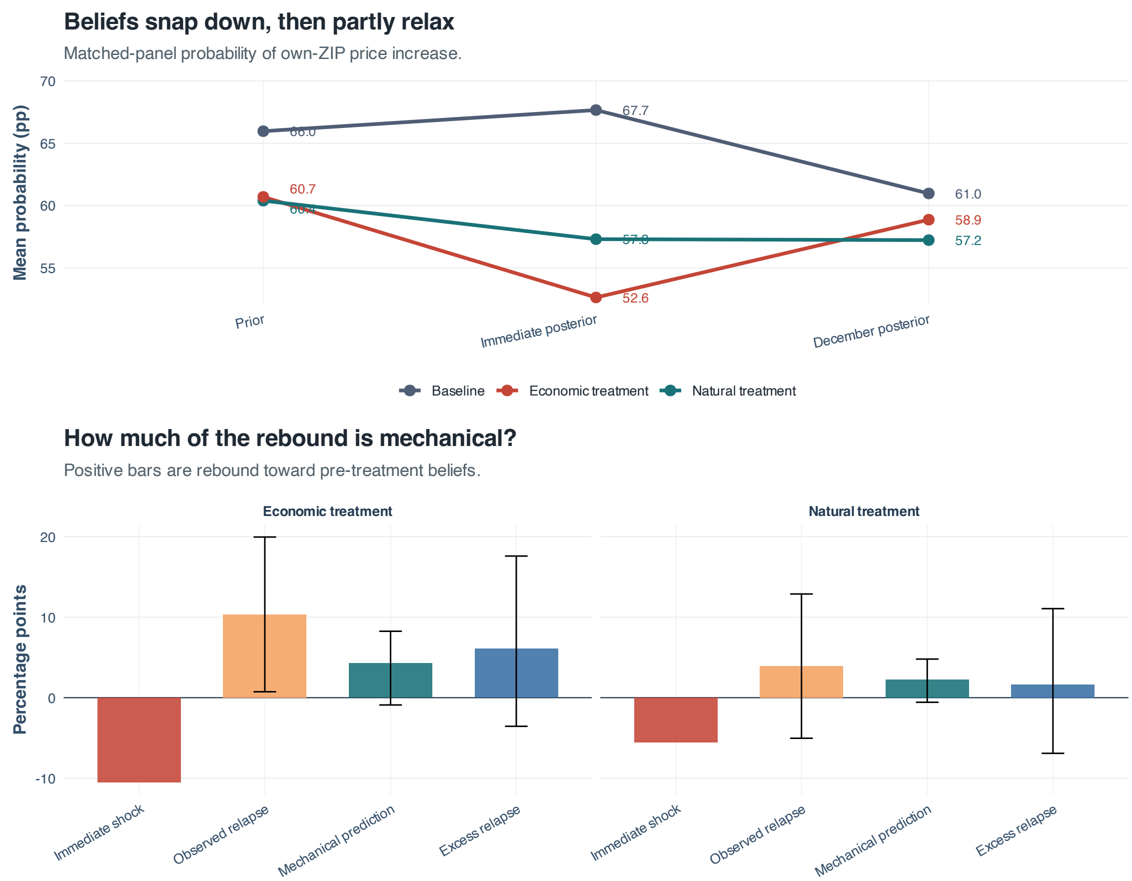

Figure 5: Belief paths and the reversion benchmark. *Notes:* The upper panel shows mean prior, immediate-posterior, and December beliefs for the 337 respondents who completed the follow-up. The lower panel begins with each treatment's immediate effect (labeled as *Immediate shock*) in this same sample and compares the observed prior-adjusted rebound with the amount predicted from the Baseline arm's retention pattern. The difference is labeled "excess relapse." The confidence intervals are based on ZIP-cluster bootstrap draws.

convergence of study arms can be attributed to changes in treated arms and also in the Baseline condition.

4.5.2 December contrasts and attrition

We first examine the convergence of study arms in follow-up while controlling for the main study priors. Table 3 reports prior-adjusted treatment effect estimates in the follow-up sample. The probability assigned to a price increase differs from the Baseline arm by -0.18 percentage points (SE 4.50) in the Economic treatment and by -1.66 points (SE

4.22) in the Natural treatment. We also do not detect treatment differences in entropy. Our conclusions do not change when we use wild cluster bootstrap and Romano–Wolf adjustments.

We also pool the two treatments and compare them with the Baseline arm. The pooled December effect is -0.96 percentage points (SE 3.81). The pooling improves the precision of our estimates, as the two-sided 80-percent MDE falls from 12.64 and 11.88 points in the separate treatment comparisons to 10.72 points.

The follow-up study is not equally informative about the two treatments. The December MDEs are 12.64 and 11.88 points, while the corresponding immediate effects in the follow-up sample are -10.52 and -5.58 points. In absolute value, each immediate effect is smaller than its December MDE. Thus, even if the Natural-treatment effect had remained unchanged, the December sample would not reliably detect an effect of that size. The Economic-treatment effect is closer to the detection boundary. For this reason, we do not interpret a failure to reject zero as evidence that no effect remained.

Our next exercise examines whether selective participation in the follow-up study could change our December results. We use Lee bounds and assume that being in either of the main study treatments can only increase or leave unchanged a respondent's probability of completing the follow-up compared to the Baseline condition. This assumption is consistent with the treatment messages making some respondents more engaged with the study and therefore more likely to return. However, we cannot verify this condition for each respondent, so we treat the bounds as a sensitivity analysis rather than as a correction for attrition. Since response rates are higher in the treatment arms than in Baseline, we pool respondents across field waves and remove enough treatment-arm observations to equalize response rates across the study arms. We then calculate the bounds by removing respondents with the highest and lowest price increase probabilities. These calculations provide the smallest and largest treatment effects consistent with our assumption about follow-up response.

Table 3: December treatment contrasts among follow-up respondents

	Effect (CR1S SE)	p-values RI	p-values WCB	p-values RW	80% MDE
Panel A. Prior-adjusted December contrasts					
Probability (pp): Economic	−0.18 (4.50)	0.969	0.968	0.998	12.64
Probability (pp): Natural	−1.66 (4.22)	0.690	0.697	0.970	11.88
Probability (pp): Any treatment (pooled)	−0.96 (3.81)	0.794	0.803	0.986	10.72
Entropy: Economic	2.52 (2.50)	0.319	0.318	0.746	7.02
Entropy: Natural	−0.11 (2.61)	0.966	0.963	0.998	7.35

Panel B. Selection bounds (randomized-population estimands)	Lee bounds	Manski bounds
Probability: Economic	[−2.47, −1.58] IM [−12.86, 9.08]	[−51.23, 49.22] IM [−56.85, 54.65]
Probability: Natural	[−9.88, 4.37] IM [−20.64, 14.91]	[−48.23, 45.55] IM [−53.81, 51.15]
Entropy: Economic	[5.71, 6.60] IM [−0.27, 13.12]	[−47.17, 53.27] IM [−52.19, 57.80]
Entropy: Natural	[−1.77, 8.05] IM [−8.44, 14.85]	[−44.58, 49.20] IM [−49.63, 53.95]

Panel C. Response and missingness sensitivity (probability, Economic / Natural)	
Response rates (Base / Econ / Nat)	49.6% / 50.0% / 56.7%
IPW, weights re-estimated per draw	−0.44 [−9.0, 7.8] / −1.71 [−9.9, 6.3]
Multiple imputation ($m = 50$)	+0.30 [−10.0, 10.6] / −1.45 [−11.5, 8.6]

Notes: Panel A compares each treatment with the Baseline arm among the 337 follow-up respondents, controlling for the corresponding prior and field wave. Standard errors are CR1S and clustered across 260 ZIP codes. RI denotes within-wave randomization inference, WCB the null-imposed wild cluster bootstrap, and RW the Romano–Wolf adjustment; each uses 9,999 draws. The MDE is the two-sided 80-percent minimum detectable effect. Panel B asks how selective follow-up could change the results. Lee bounds assume that treatment does not reduce response. The calculation pools respondents across field waves and fractionally trims only the higher-response treatment arms until their response rates match the Baseline arm. Manski bounds use only the logical range of the outcome. IM rows are 95-percent Imbens–Manski intervals based on 1,999 ZIP-cluster bootstrap draws. Panel C reports inverse-probability weighting and 50-imputation estimates; brackets are 95-percent intervals.

Table 3 reports this range together with Imbens–Manski confidence intervals that account for sampling uncertainty by resampling ZIP clusters. The confidence intervals are wide and include both no-December-effect and effects consistent with substantial persistence. Therefore, after accounting for possible selection bias in the follow-up study, the magnitude of the main study treatment effect remaining in December becomes noisy.

Note that Manski's bounds are limited when the follow-up response rate is only 52 percent. We also conduct inverse probability weighting and multiple imputation, yielding estimates that are close to the unweighted results.

4.5.3 Persistence ratios

We next place the December treatment estimates on the same scale as the main-study effects. Specifically, we divide each treatment's December effect by its own immediate main-study effect in the follow-up sample. A persistence ratio of zero indicates no remaining treatment effect in December, while a ratio of one indicates full persistence.

Since the persistence ratio divides the December effect by the immediate main-study effect, even a small noise in the immediate effect can lead to a large change in the ratio. Employing the Anderson–Rubin approach allows us to circumvent this issue and yields a confidence set for persistence, even when the immediate effect is imprecisely estimated.

When we pool the two treatments, the immediate effect is -7.91 percentage points, with the corresponding F-statistic of 12.76. The estimated persistence ratio for the pooled treatment effect is 0.121, while the Anderson–Rubin 95-percent confidence set is $[-1.03, 1.22]$. Since the Anderson–Rubin set contains both zero and one, the pooled treatment effect persistence analysis cannot distinguish no-persistence from full persistence at the five-percent level ($p(\theta = 1) = 0.099$).

The immediate effect of the Economic treatment in the follow-up sample is -10.52 percentage points ($F = 15.54$), and the estimated persistence ratio is 0.017. The Anderson–Rubin confidence set is $[-1.02, 0.91]$ and excludes one, allowing us to reject full persistence ($p(\theta = 1) = 0.036$). However, the confidence set includes zero, and we cannot reject no-persistence ($p(\theta = 0) = 0.969$). Therefore, we conclude that the entire Economic-treatment effect did not persist through December, but we cannot exclude the possibility that some effects persisted in the follow-up.

The same exercise for the Natural treatment reveals that the immediate effect in the follow-up sample is -5.58 percentage points, with an F-statistic of only 5.50. Its estimated persistence ratio is 0.297, with the Anderson–Rubin confidence set of $[-2.27, 3.19]$. Since the set contains both zero and one, we cannot distinguish between no persistence and full

Table 4: Persistence ratios with weak-identification-robust inference

	Economic	Natural	Any treatment
Immediate effect, matched panel (SE)	−10.52 (2.67)	−5.58 (2.38)	−7.91 (2.21)
First-stage F	15.54	5.50	12.76
First-stage strength	$F \geq 10$	Weak: $F < 10$	$F \geq 10$
December effect (SE)	−0.18 (4.50)	−1.66 (4.22)	−0.96 (3.81)
Persistence ratio $\hat{\theta}$	0.017	0.297	0.121
Anderson–Rubin 95% set	[−1.02, 0.91]	[−2.27, 3.19]	[−1.03, 1.22]
$p(\theta = 0)$	0.969	0.695	0.801
$p(\theta = 1)$	0.036	0.392	0.099

Notes: Both the immediate and December effects are estimated among the same 337 follow-up respondents, covering 260 ZIP clusters. The models control for the prior and the field-wave fixed effects. The persistence ratio $\hat{\theta}$ divides the December effect by the treatment's own immediate effect in this sample. A value of zero indicates no remaining contrast, while one indicates full persistence. Because the denominator can be imprecisely estimated, the Anderson–Rubin procedure tests each candidate ratio directly. The "Any treatment" column compares assignment to either treatment with the Baseline arm. All Anderson–Rubin sets use CR1S ZIP-clustered covariance.

persistence for the Natural treatment.

4.5.4 How much reversion is mechanical?

In this section, we scrutinize potential reasons behind the disappearance of the treatment effects. One reason could be that respondents may have changed their underlying price beliefs between the main and follow-up studies. However, part of the observed rebound may also arise from reporting error in the main study posterior. Put differently, our belief-elicitation procedure may add temporary noise to the reported posterior. If an unusually low posterior partly reflects such error, then the follow-up posterior will tend to move upward even when the underlying house-price belief remains unchanged. We refer to this movement as mechanical reversion.

We investigate mechanical reversion in several steps. Our first inquiry is to determine the relationship between the main study prior and the December posterior. Panel A of Table 5 examines whether this relationship enables us to predict December beliefs after accounting for the main-study posterior. We then predict the December posterior using both the main study priors and posteriors, and find that their estimated coefficients are

0.183 and 0.227, respectively. We cannot reject the equality of estimates ($p = 0.847$), suggesting that the prior continues to predict December beliefs even after conditioning on the main-study posterior. This result is consistent with respondents returning to their original belief anchor. However, our regression approach is not conclusive about whether this return reflects forgetting of the treatment information or the disappearance of the reporting error.

Panel B next uses the Baseline arm to estimate how much of an immediate belief update ordinarily remains in December. Because this arm does not include an appended treatment message, it provides a benchmark for movement between the main and follow-up elicitation. Among Baseline respondents, a one-point main-study update predicts that 0.594 points remain at follow-up (SE 0.213). Under the assumption that the same reporting process applies to all three arms, we use the remaining fraction, $1 - 0.594$, as the share predicted to reverse mechanically. Applying this benchmark to the immediate treatment effects predicts rebounds of 4.27 points for the Economic treatment and 2.26 points for the Natural treatment.

The observed follow-up rebounds are 10.34 points for the Economic treatment and 3.92 points for the Natural treatment. The differences between the observed and predicted rebounds are therefore 6.07 and 1.65 points, respectively. The bootstrap intervals for both differences include zero. Our finding does not imply that forgetting played no role. Rather, the estimates are too imprecise to determine how much of either rebound is mechanical and how much reflects a change in underlying beliefs.[19]

In the Online Appendix, we also examine whether the initial belief improvement disappears when beliefs are evaluated against realized house prices. The analysis includes 335 respondents.[20] The Economic treatment initially improves the ranked probability

[19]The Online Appendix estimates the relationship between the immediate and December updates in both directions. This exercise allows us to put the two estimates on the same scale. Results reveal that the conclusion is sensitive to which elicited update is treated as measured with error. Moreover, the confidence intervals widen further when we account for selective participation in the follow-up study.

[20]Two follow-up respondents are excluded because realized Zillow price growth is unavailable for their

Table 5: Anchor estimates and the mechanical-reversion benchmark

Panel A. Main study prior and posteriors equally predict December beliefs				
	Prior (SE)	Posterior (SE)	Equality p	N
Probability	0.183 (0.120)	0.227 (0.115)	0.847	337
Entropy	0.351 (0.109)	0.112 (0.102)	0.244	337
Panel B. Observed relapse vs. the mechanical benchmark (probability, pp)				
	Observed relapse (SE)	Predicted mechanical [95% CI]	Excess	Excess [95% CI]
Economic treatment	10.34 (4.90)	4.27 $[-0.91, 8.25]$	6.07	$[-3.56, 17.59]$
Natural treatment	3.92 (4.57)	2.26 $[-0.57, 4.79]$	1.65	$[-6.91, 11.06]$

Notes: Panel A predicts the December belief using both the main-study prior and immediate posterior, together with treatment indicators and field-wave fixed effects. Standard errors are CR1S and clustered across 260 ZIP codes. The equality test compares the prior and posterior coefficients using their full covariance. Panel B compares the observed prior-adjusted rebound with a mechanical benchmark from the Baseline arm. Among 112 Baseline respondents, a one-point immediate update predicts that $b = 0.594$ points remain in December (SE 0.213). We therefore use $1 - b$ as the share of each treatment's immediate effect predicted to reverse. "Excess" is the observed rebound minus this prediction. The Economic-treatment rebound has $p = 0.036$. Intervals come from 1,999 bootstrap draws that resample ZIP clusters and repeat the full calculation.

score by 0.394 (Holm-adjusted $p = 0.085$). The change from the immediate posterior to December is -0.614 (SE 0.333, with Holm-adjusted $p = 0.332$). In turn, the change from the prior to December is -0.220 (Holm-adjusted $p = 1.000$). The point estimates suggest that respondents also lost the main study accuracy gain in follow-up, but this evidence becomes inconclusive after adjustment for multiple testing.

Finally, respondents who revise their beliefs more immediately also tend to move back more by December. The estimated slopes are -0.41, -0.49, and -0.61 in the Baseline arm, Economic treatment, and Natural treatment, respectively. We cannot reject the equality of these slopes ($p = 0.774$). The pattern is consistent with proportional reversion across all three study arms. However, the immediate posterior enters the initial update and the subsequent change with opposite signs, which mechanically creates a negative relationship. We therefore do not interpret these slopes as evidence about memory.

Taken together, the December evidence differs across the two treatments. For the

ZIP codes.

Economic treatment, we reject full persistence but cannot determine why the effect disappeared. For the Natural treatment, the immediate effect in the follow-up sample is too imprecisely estimated to distinguish no persistence from full persistence.

4.6 Downstream Outcomes and Decision Relevance

Do the randomized messages also change respondents' intended behavior or their beliefs about hurricane damage and insurance? Our insurance survey block, consisting of four insurance questions, measures different objects. Respondents first reported the probability that their home would be damaged by hurricanes or related natural disasters by the end of 2025. We refer to this response as stated damage probability. A separate question asked, "What percentage of your home's value do you think could be lost in the next hurricane or related natural disasters?" We refer to this response as the potential loss share. In the structural exercise below, we interpret this response as the share of home value lost in the damage state. The third insurance question asked respondents for the percentage of home value they would be willing to pay each year in insurance premiums. We refer to this response as stated annual premium tolerance. The final item asked respondents to rate the accuracy of weather forecasting services in predicting hurricanes or related natural disasters in their ZIP code, where zero meant "no accuracy at all" and 100 meant "extremely accurate." We refer to this response as perceived forecast accuracy.

We also elicit homeowners' intention to list their property by the end of 2025 in both the main and follow-up studies. In the main study, we also asked respondents whether they thought it was a good time to sell their property, measuring their perceptions of market conditions.

Table 6 reports reduced-form treatment effects on selling intentions, perceived housing market conditions, four insurance-related responses, and December listing and selling intentions. We do not detect a treatment effect on these reported perceptions or intentions.

Table 6: Treatment effects on stated decision margins, with minimum detectable effects

Outcome	Arm	Effect (SE)	RI p	RW p	80% MDE	N
Immediate (full sample)						
Considering sale (0–7)	Econ	0.02 (0.04)	0.576	0.990	0.11	648
	Nat	−0.03 (0.04)	0.477	0.990	0.12	648
Good time to sell (−2–2)	Econ	0.01 (0.11)	0.919	0.917	0.30	642
	Nat	0.21 (0.11)	0.075	0.125	0.32	642
Damage probability (pp)	Econ	0.97 (1.91)	0.614	0.990	5.36	648
	Nat	2.79 (2.05)	0.170	0.809	5.75	648
Conditional loss share (pp)	Econ	−0.72 (2.72)	0.793	0.990	7.63	648
	Nat	0.17 (2.73)	0.948	0.990	7.66	648
Stated premium (% of value)	Econ	−0.64 (1.99)	0.742	0.990	5.59	648
	Nat	−1.81 (1.92)	0.353	0.966	5.39	648
Forecast trust (0–100)	Econ	−1.57 (2.40)	0.512	0.990	6.73	648
	Nat	−1.72 (2.39)	0.472	0.990	6.71	648
December follow-up (among respondents)						
Listed property (0/1)	Econ	−0.02 (0.03)	0.402	0.628	0.08	337
	Nat	−0.02 (0.03)	0.501	0.628	0.08	337
December selling prob. (0–7)	Econ	−0.02 (0.18)	0.920	0.991	0.50	321
	Nat	−0.01 (0.16)	0.944	0.991	0.46	321

Notes: Each effect compares a treatment with the Baseline arm and controls for the field wave. CR1S standard errors are clustered by ZIP. The immediate outcomes use the full sample except "Good time to sell," which excludes six "unsure" responses. December outcomes are observed only among follow-up respondents. RI denotes within-wave randomization inference and RW the Romano–Wolf adjustment within each sample-defined outcome family. The Online Appendix reports the family map and wild-bootstrap results. The MDE is the two-sided 80-percent minimum detectable effect. No effect survives family adjustment at the 10-percent level.

The belief changes may have been too small or too short-lived to cross decision thresholds.

4.6.1 Stated premium tolerance on an external scale

How large are respondents' stated premium tolerances when compared to premiums observed in the market? Panel A of Figure 6 plots the full distribution of this measure and also provides an external benchmark.[21] We borrow benchmarks from the national distribution of 2024 mortgage-escrow premiums reported by Keys and Mulder (2024).

Respondents report premium tolerances that are high relative to these benchmarks.

[21] This measure records the maximum premium share stated in the survey, not contract-specific insurance demand.

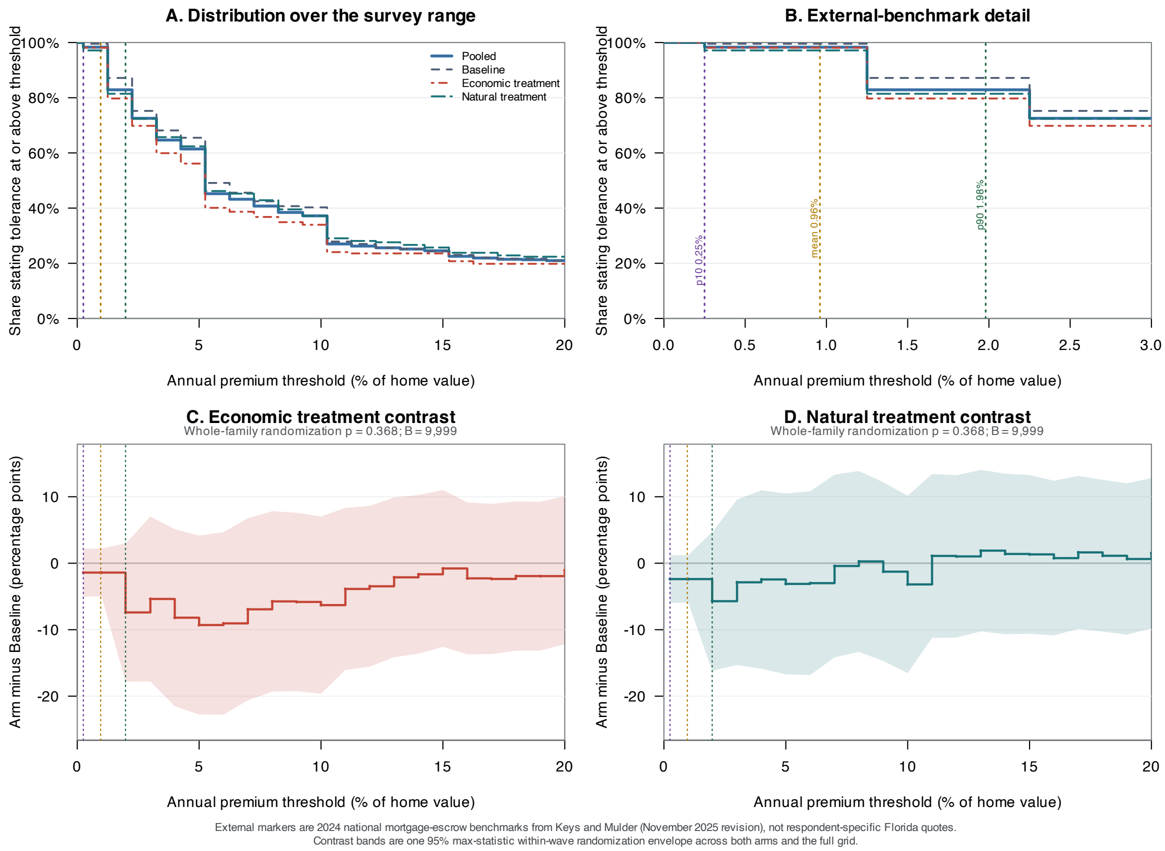

Figure 6: Stated premium tolerance and external market benchmarks. *Notes:* Panel A shows the share of respondents whose stated annual premium tolerance is at least each percentage of home value. Panel B enlarges the part of the distribution containing the national 2024 mortgage-escrow p10, mean, and p90 reported by Keys and Mulder (2024). These vertical lines provide external scale markers rather than respondent-specific Florida quotes. Panels C and D compare each treatment with the Baseline arm at every premium threshold. Shaded areas are simultaneous 95-percent intervals. A 9,999-draw randomization test across both treatment contrasts, and all thresholds gives $p = 0.368$.

At both the external p10 and mean markers, 98.3 percent of respondents report tolerance at or above the corresponding premium share.[22] More strikingly, 82.9 percent report tolerance at or above the external p90 premium share of 1.98 percent. As Panels C and D of Figure 6 show, the randomized treatment messages do not shift the distribution of stated tolerance. In the Online Appendix, we convert stated tolerances into USD using

[22]The shares are identical at the two markers because respondents reported their tolerance in whole percentage points. The external p10 and mean, at 0.25 and 0.96 percent, therefore fall within the same interval of the elicited response scale.

survey responses across home-value brackets and provide sensitivity checks.

This comparison provides a market scale for the survey responses, but it does not identify insurance demand.[23] We therefore conclude only that most respondents report a premium tolerance above even the p90 marker in the national benchmark distribution.

4.6.2 Can information change an insurance decision?

When could better information about hurricane damage change a homeowner's insurance decision? Figure 7 summarizes our primary structural assumptions and reports the resulting model-implied decision margins.

We first convert each respondent's reported home-value interval to its midpoint. This allows us to express the reported loss and insurance premium as a proportion of home value. As we do not observe respondent-specific insurance quotes, our calibrations utilize the 2024 national mortgage-escrow mean annual premium of $2,712.98 reported by Keys and Mulder (2024).

[23]The premium-tolerance question was not incentive-compatible and did not specify an insurer, deductible, coverage limit, or other contract details. The national benchmark also uses purchase price as its denominator, whereas respondents reported tolerance relative to home value.

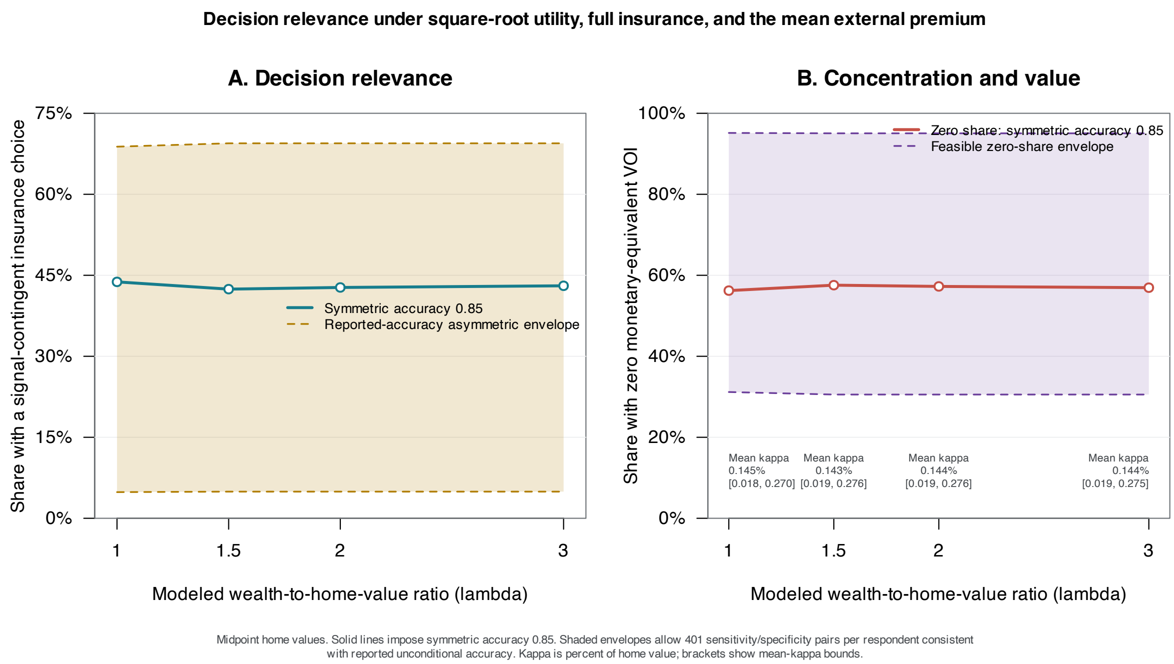

Figure 7: Insurance decision margins. *Notes:* Solid lines show the benchmark calibration under square-root utility, full insurance, midpoint home values, the external mean premium, and signal sensitivity and specificity of 0.85. The shaded areas keep each respondent's reported overall forecast accuracy fixed while allowing it to come from 401 different combinations of correct warnings when damage occurs and correct reassurances when it does not. These combinations produce different posterior damage probabilities and, therefore, different insurance choices. The shaded areas show the resulting range. The measure κ is the additional wealth, as a share of home value, that would make a respondent as well off without the signal as with it after reconsidering the insurance choice. All results are model-implied under the stated assumptions.

Our model compares full insurance with no insurance and evaluates each respondent's decision under two hypothetical realizations of a damage-risk signal. A *warning* raises the posterior probability of damage, whereas a *reassurance* lowers it. These hypothetical signals are distinct from the treatment messages randomized in our experiment. In our calibration, both signal sensitivity and specificity are 0.85.

This national mean gives every respondent a common premium scale, but it may be lower than the premiums faced by many Florida homeowners. Raising the premium does not change the modeled decision margin in the same direction for everyone. Depending on where a respondent lies relative to the insurance threshold, it can create or eliminate a choice that changes across the two signals. The Online Appendix therefore reports how the result varies across premium-to-loss ratios.

Under our modeling assumptions, 42.4–43.8 percent of respondents make different insurance choices under the warning and reassurance realizations. Put differently, insurance is optimal for these respondents after the warning but not after the reassurance.

We also examine whether the assumed relationship between home value and total wealth changes this decision margin. We vary modeled wealth from one to three times home value, $W_i = \lambda H_i$, and obtain nearly the same extensive-margin result.[24]

For the remaining 56.2–57.6 percent of respondents, the same insurance choice is optimal after the realization of either signal. Therefore, for this group of homeowners, information consequently has zero value in this binary decision problem. Although mean κ is 0.143–0.145 percent of home value, the median is zero because most respondents do not cross the insurance threshold.[25]

4.7 Robustness and Scope

In the appendix, we provide a series of robustness checks.[26] Main study treatment effects remain similar after excluding (i) respondents who allocate tokens only in multiples of ten, (ii) the fastest five percent of survey durations, and (iii) boundary beliefs. The Economic treatment ranges from -12.69 to -9.11 points, while the Natural treatment ranges from -8.45 to -6.29 points throughout these checks.

We also conduct leave-one-county-out estimates, post-stratification, post-double-selection, and rank tests. With alternative outcome, control, and timing definitions, we

[24] At $\lambda = 1$, we exclude 29 respondents whose reported conditional loss equals their entire modeled wealth. All respondents satisfy the model's domain conditions at the larger wealth ratios.

[25] Here, κ is the additional wealth, expressed relative to home value, that would make the optimized no-information decision as valuable as receiving the signal.

[26] We also provide an extensive supporting evidence archive here: `https://samirhuseynov.com/research/supplementary_results_archive.pdf`. This archive is organized as a ledger of our systematic analyses.

still document treatment effects.

As a further robustness check, we use double machine learning with prior beliefs, property characteristics, demographics, and field-wave indicators (Chernozhukov et al., 2018). This method uses Lasso to select the controls that help predict respondents' posterior beliefs and treatment assignment. We divide ZIP codes into five groups. The model is estimated using respondents from four groups and then used to generate predictions for respondents in the remaining group. We repeat this process until every group has been held out once. All respondents from the same ZIP code always remain in the same group. After making this data-driven adjustment, the estimated effects are −12.01 percentage points for the Economic treatment and −7.88 percentage points for the Natural treatment. These estimates are similar to our main results. Thus, our findings remain unchanged when we allow the data to determine which controls are most useful for the adjustment.

Our spatial inference reaches the same conclusion. Conley standard errors for the same-sample economic treatment estimate are 1.87 at 50 km and 1.85 at 100 km, compared with 1.91 under ZIP clustering. The corresponding Natural-treatment values are 1.75, 1.84, and 1.79. Neither respondent-level nor ZIP-mean residual Moran tests reject spatial randomness. Within-ZIP peer-assignment spillover tests are also null.

Only 52 percent of the original sample completed the December follow-up. In the Online Appendix, we examine how this attrition could affect the results. Weighting respondents by their estimated probability of return and imputing missing December outcomes continue to produce estimates close to zero. Lee bounds, which assume that treatment assignment did not reduce follow-up response, also place the Economic-treatment effect below its immediate correction in the main study, although the corresponding confidence intervals are wide. Rank- and sign-based tests do not detect a December effect. Manski bounds, which place almost no restrictions on missing outcomes, remain very wide. These exercises, therefore, do not eliminate attrition as a concern.

Our design also cannot completely rule out the experimenter demand effect. In the treatment conditions, we elicit posterior beliefs immediately after respondents read directional messages. Subjects may revise their beliefs to align with treatment expectations. Our incentive-compatible belief scoring rules alleviate this concern but cannot completely dismiss it.

5 Conclusion

We provide insights on how latent price beliefs are shaped and updated in a changing housing market squeezed between market cooling and hurricane risks. Our incentivized experimental protocols capture the latent component of price expectations that is absent from market data and complement secondary data studies by showing how different cognitive mechanisms can lead to distinct market beliefs.

An important contribution of this paper is that we not only capture optimism in house-price expectations but also provide several forward-looking benchmarks based on publication-lag-truncated market data and backward-looking benchmarks based on realized house prices. Both comparisons show that the homeowners are miscalibrated. This miscalibration matters, especially given that residential property constitutes an important portion of household wealth.

We document that property-related features are more predictive of prior price beliefs, while weather conditions and previous hurricane exposure do not robustly predict beliefs after adjustment for multiple testing. Our findings suggest that homeowners may not fully incorporate extreme weather events into their price expectations when shaping or updating them. Under our structural assumptions, we also find that a hypothetical hurricane-damage signal has no decision value for most homeowners because it does not change their modeled insurance choice.

We do document meaningful improvements in price expectations when we provide short messages about potential market cooling and looming hurricane and flood risks. This finding offers some hope for information campaigns as a useful tool for correcting market expectations. By December, however, the treatment effects are no longer detectable. We reject full persistence for the Economic treatment, while the Natural-treatment evidence remains too imprecise to distinguish no-persistence from full persistence. We also do not find that belief updates affect reported property-selling intentions or perceived hurricane risks.

Taken together, our findings show that information about climate risk and market movements can change housing-market expectations. However, an immediate correction does not guarantee a durable one. Information campaigns are therefore more likely to matter when they are repeated, timed to actual decisions, and paired with tools that enable households to act on them.

References

Luis Armona, Andreas Fuster, and Basit Zafar. Home price expectations and behaviour: Evidence from a randomized information experiment. *The Review of Economic Studies*, 86(4):1371–1410, 2019.

Orazio Attanasio, Andrew Leicester, and Matthew Wakefield. Do house prices drive consumption growth? the coincident cycles of house prices and consumption in the uk. *Journal of the European Economic Association*, 9(3):399–435, 2011.

Michael Bailey, Eduardo Dávila, Theresa Kuchler, and Johannes Stroebel. House price beliefs and mortgage leverage choice. *The Review of Economic Studies*, 86(6):2403–2452, 2019.

Laura A. Bakkensen and Lint Barrage. Going underwater? flood risk belief heterogeneity and coastal home price dynamics. *Review of Financial Studies*, 35(8):3666–3709, 2022.

Markus Baldauf, Lorenzo Garlappi, and Constantine Yannelis. Does climate change affect real estate prices? only if you believe in it. *Review of Financial Studies*, 33(3):1256–1295, 2020.

Asaf Bernstein, Matthew T. Gustafson, and Ryan Lewis. Disaster on the horizon: The price effect of sea level rise. *Journal of Financial Economics*, 134(2):253–272, 2019.

Carola Binder, Pei Kuang, and Li Tang. Central bank communication and house price expectations. *Journal of the European Economic Association*, 24(2):769–799, 2026.

Nicolas L. Bottan and Ricardo Perez-Truglia. Betting on the house: Subjective expectations and market choices. *American Economic Journal: Applied Economics*, 17(1): 459–500, 2025.

Leah Platt Boustan, Matthew E. Kahn, Paul W. Rhode, and Maria Lucia Yanguas. The effect of natural disasters on economic activity in us counties: A century of data. *Journal of Urban Economics*, 118:103257, 2020. doi: 10.1016/j.jue.2020.103257.

Grant Brissey. December market report: 4 predictions for 2025. Zillow Premier Agent, December 2024. URL `https://www.zillow.com/premier-agent/market-report-predictions-2025/`. Accessed July 23, 2026.

Jan K. Brueckner, Matthew E. Kahn, and Gary C. Lin. A new spatial hedonic equilibrium in the emerging work-from-home economy? *American Economic Journal: Applied Economics*, 15(2):285–319, 2023. doi: 10.1257/app.20210190.

Giovanni Burro, Alessandro Castagnetti, Alessandra Cillo, and Paolo Crespi. The impact of environmental risk information on beliefs and home insurance decisions. *Journal of Economic Psychology*, page 102903, 2026.

Colin F Camerer and Robin M Hogarth. The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of risk and uncertainty*, 19 (1):7–42, 1999.

A. Colin Cameron, Jonah B. Gelbach, and Douglas L. Miller. Bootstrap-based improvements for inference with clustered errors. *Review of Economics and Statistics*, 90(3): 414–427, 2008.

Karl E. Case, Robert J. Shiller, and Anne K. Thompson. What have they been thinking? home buyer behavior in hot and cold markets. *Brookings Papers on Economic Activity*, 43(2):265–315, 2012.

Alberto Cavallo, Guillermo Cruces, and Ricardo Perez-Truglia. Inflation expectations, learning, and supermarket prices: Evidence from survey experiments. *American Economic Journal: Macroeconomics*, 9(3):1–35, 2017.

Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. Double/debiased machine learning for treatment and structural parameters. *Econometrics Journal*, 21(1):C1–C68, 2018.

Felix Chopra, Christopher Roth, and Johannes Wohlfart. Home price expectations and spending: Evidence from a field experiment. *American Economic Review*, 115(7):2267–2305, 2025.

Citizens Property Insurance Corporation. Combined assessments and surcharge amounts, November 2023. URL `https://www.citizensfla.com/-/20231103-combined-assessments-and-surcharge-amounts`. Accessed July 23, 2026.

Olivier Coibion, Yuriy Gorodnichenko, and Michael Weber. Monetary policy communications and their effects on household inflation expectations. *Journal of Political Economy*, 130(6):1537–1584, 2022.

Benjamin Enke, Uri Gneezy, Brian Hall, David Martin, Vadim Nelidov, Theo Offerman, and Jeroen Van De Ven. Cognitive biases: Mistakes or missing stakes? *Review of Economics and Statistics*, 105(4):818–832, 2023.

Christine L Exley and Kirby Nielsen. The gender gap in confidence: Expected but not accounted for. *American Economic Review*, 114(3):851–885, 2024.

Daryl Fairweather and Chen Zhao. Redfin's 2025 housing market predictions. Redfin News, December 2024. URL `https://www.redfin.com/news/housing-market-predictions-2025/`. Accessed July 23, 2026.

Daryl Fairweather, Matthew E Kahn, Robert D Metcalfe, and Sebastian Sandoval Olascoaga. Expecting climate change: A nationwide field experiment in the housing market. Technical report, National Bureau of Economic Research, 2024.

Florida Hurricane Catastrophe Fund. Claims-paying capacity estimates. Technical report, State Board of Administration of Florida, May 2025. URL `https://fhcf.sbafla.com/media/dzcapkgh/fhcf-may-2025-bonding-capacity-final-corrected.pdf`. May 2025 bonding-capacity report; accessed July 23, 2026.

Florida Senate. Chapter 627, section 351, florida statutes: Citizens property insurance corporation, 2024. URL `https://www.flsenate.gov/Laws/statutes/2024/627.351`. Accessed July 23, 2026.

Andreas Fuster and Basit Zafar. Survey experiments on economic expectations. In *Handbook of Economic Expectations*, pages 107–130. Elsevier, 2023.

David Genesove and Christopher Mayer. Loss aversion and seller behavior: Evidence from the housing market. *The Quarterly Journal of Economics*, 116(4):1233–1260, 2001. doi: 10.1162/003355301753265561.

Matthew Gibson and Jamie T. Mullins. Climate risk and beliefs in new york floodplains. *Journal of the Association of Environmental and Resource Economists*, 7(6):1069–1111, 2020. doi: 10.1086/710240.

Edward L. Glaeser and Charles G. Nathanson. An extrapolative model of house price dynamics. *Journal of Financial Economics*, 126(1):147–170, 2017. doi: 10.1016/j.jfineco.2017.06.012.

Jesse D. Gourevitch, Carolyn Kousky, Yanjun Liao, Christoph Nolte, Adam B. Pollack, Jeremy R. Porter, and Joakim A. Weill. Unpriced climate risk and the potential consequences of overvaluation in us housing markets. *Nature Climate Change*, 13(3):250–257, 2023. doi: 10.1038/s41558-023-01594-8.

Arpit Gupta, Vrinda Mittal, Jonas Peeters, and Stijn Van Nieuwerburgh. Flattening the curve: Pandemic-induced revaluation of urban real estate. *Journal of Financial Economics*, 146(2):594–636, 2022. doi: 10.1016/j.jfineco.2021.10.008.

Ingar Haaland, Christopher Roth, and Johannes Wohlfart. Designing information provision experiments. *Journal of Economic Literature*, 61(1):3–40, 2023.

Miyuki Hino and Marshall Burke. The effect of information about climate risk on property values. *Proceedings of the National Academy of Sciences*, 118(17):e2003374118, 2021. doi: 10.1073/pnas.2003374118.

Tanjim Hossain and Ryo Okui. The binarized scoring rule. *The Review of Economic Studies*, 80(3):984–1001, 2013.

Greg Howard, Jack Liebersohn, and Adam Ozimek. The short- and long-run effects of remote work on u.s. housing markets. *Journal of Financial Economics*, 150(1):166–184, 2023. doi: 10.1016/j.jfineco.2023.103705.

Solomon Hsiang, Robert Kopp, Amir Jina, James Rising, Michael Delgado, Shashank Mohan, D. J. Rasmussen, Robert Muir-Wood, Paul Wilson, Michael Oppenheimer, Kate Larsen, and Trevor Houser. Estimating economic damage from climate change in the united states. *Science*, 356(6345):1362–1369, 2017. doi: 10.1126/science.aal4369.

Samir Huseynov. Overreaction in house price expectations. *Journal of Real Estate Finance and Economics*, 2026. doi: 10.1007/s11146-026-10060-z.

Insurance Information Institute. Triple-i: Home insurance premiums in florida increased 80% less in 2023 than initial projections, June 2024. URL `https://www.iii.org/press-release/triple-i-home-insurance-premiums-in-florida-increased-80-less-in-2023-than-initial-projections-due-in-large-part-to-legislative-legal-system-abuse-reforms-062624`. Accessed July 23, 2026.

Benjamin J. Keys and Philip Mulder. Neglected no more: Housing markets, mortgage lending, and sea level rise. NBER Working Paper 27930, National Bureau of Economic Research, 2020.

Benjamin J. Keys and Philip Mulder. Property insurance and disaster risk: New evidence from mortgage escrow data. NBER Working Paper 32579, National Bureau of Economic Research, 2024. URL `https://www.nber.org/papers/w32579`. June 2024, revised November 2025.

Theresa Kuchler and Basit Zafar. Personal experiences and expectations about aggregate outcomes. *Journal of Finance*, 74(5):2491–2542, 2019.

Theresa Kuchler, Monika Piazzesi, and Johannes Stroebel. Housing market expectations. In *Handbook of economic expectations*, pages 163–191. Elsevier, 2023.

Charles F. Manski. Measuring expectations. *Econometrica*, 72(5):1329–1376, 2004.

Miami Herald Editorial Board. A predicted hectic hurricane season + uninsured homeowners. what could go wrong? *Miami Herald*, May 2024. URL `https://www.miamiherald.com/opinion/editorials/article288826435.html`. Opinion editorial; updated May 30, 2024, 3:55 PM.

John A. Mondragon and Johannes Wieland. Housing demand and remote work. NBER Working Paper 30041, National Bureau of Economic Research, 2022.

National Hurricane Center. 2024 atlantic hurricane season, 2026a. URL `https://www.nhc.noaa.gov/data/tcr/index.php?season=2024`. Tropical Cyclone Reports and 2024 North Atlantic summary table; accessed July 23, 2026.

National Hurricane Center. Atlantic hurricane database (HURDAT2), 1851–2025. National Oceanic and Atmospheric Administration, 2026b. URL `https://www.nhc.noaa.gov/data/hurdat/hurdat2-1851-2025-02272026.txt`. Updated February 27, 2026.

Sangmin S. Oh, Ishita Sen, and Ana-Maria Tenekedjieva. Pricing of climate risk insurance: Regulation and cross-subsidies. *Journal of Finance*, 81(3):1161–1215, 2026.

Francesc Ortega and Süleyman Taspinar. Rising sea levels and sinking property values: Hurricane sandy and new york's housing market. *Journal of Urban Economics*, 106: 81–100, 2018. doi: 10.1016/j.jue.2018.06.005.

Brad J. Reinhart. Hurricane imelda (AL092025). Tropical cyclone report, National Hurricane Center, 2026. URL `https://www.nhc.noaa.gov/data/tcr/AL092025_Imelda.pdf`. January 27, 2026.

Joseph P. Romano and Michael Wolf. Stepwise multiple testing as formalized data snooping. *Econometrica*, 73(4):1237–1282, 2005.

David Roodman, Morten Ørregaard Nielsen, James G. MacKinnon, and Matthew D. Webb. Fast and wild: Bootstrap inference in stata using boottest. *Stata Journal*, 19 (1):4–60, 2019.

U.S. Census Bureau. 2020 TIGER/Line shapefiles: Zip code tabulation areas. U.S. Department of Commerce, 2021. URL `https://www.census.gov/geographies/mapping-files/2020/geo/tiger-line-file.html`.

U.S. Senate Committee on the Budget. Whitehouse launches investigation into citizens property insurance amid questions about company's long-term solvency, November 2023. URL `https://www.budget.senate.gov/chairman/newsroom/press/whitehouse-launches-investigation-into-citizens-property-insurance-amid-questions-about-companys-long-term-solvency`. Accessed July 23, 2026.

U.S. Senate Committee on the Budget. Whitehouse presses citizens property insurance for answers about company's solvency, March 2024. URL `https://www.budget.sen` `ate.gov/chairman/newsroom/press/whitehouse-presses-citizens-property-insurance-for-answers-about-companys-solvency-`. Accessed July 23, 2026.

Oliver E. J. Wing, William Lehman, Paul D. Bates, Christopher C. Sampson, Niall Quinn, Andrew M. Smith, Jeffrey C. Neal, Jeremy R. Porter, and Carolyn Kousky. Inequitable patterns of us flood risk in the anthropocene. *Nature Climate Change*, 12(2):156–162, 2022. doi: 10.1038/s41558-021-01265-6.

Alwyn Young. Channeling fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results. *Quarterly Journal of Economics*, 134(2): 557–598, 2019.

Online Appendix to
Between Scylla and Charybdis: Florida House Price Expectations amid Market Cooling and Hurricane Risk

Samir Huseynov
Auburn University

July 2026

Contents

Appendix A Registration and design transparency

Table 1 compares the public AEA RCT registration with the analysis reported in the paper. We pre-registered the broad questions about prior beliefs, immediate updating, and persistence before the data collection. The table also records additional analyses, including inferential families, or implementation details that were not specified in the public registration. We report these distinctions so readers can distinguish the registered research questions from the analyses developed after data collection.

Table 1: Public registration and reported analysis crosswalk

Domain	Public version 1.0	Reported in this paper	Status
Sample and assignment	1,100 planned observations; computer randomization of individual property owners to three conditions.	648 recruited after the eligible platform pool was exhausted; realized arm sizes 226/212/210; assignment remains individual.	Registered design; realized sample deviation.
Prior beliefs and hurricane exposure	Analyze whether ZIP-level priors vary with previous hurricane, rain, and flood exposure.	The analysis estimates weather/NFIP, ZIP-market, property, and personal blocks; local hazard variables show little robust adjusted evidence.	Registered broad question; expanded covariate architecture.
Immediate updating	Analyze treatment-induced belief updating and whether prior hurricane exposure moderates it.	Increase-probability, entropy, eleven-bin reallocation, accuracy, and robustness families.	Broad question registered; exact outcomes and families specified later.
Persistence	Analyze whether treatment effects persist in a follow-up timed with weather developments.	December follow-up; prior-adjusted contrasts, attrition analyses, persistence ratios, and reporting-noise benchmark.	Broad question registered; timing and estimands specified later.
Forecast horizon	Registration text refers to predicting *6-month-ahead annual house price changes*.	Elicitation is interpreted as twelve-month growth; scoring uses the fixed January 2025–January 2026 Zillow window.	Material clarification/deviation.
Social connectedness	Use Facebook Social Connectedness Index to study information.	Not reported. The source is a high-dimensional network of place-to-place ties, and the registration did not specify how to collapse it into a respondent-level measure using within-Florida or nationwide connections.	Registered extension left outside the paper; no aggregation rule was prespecified, and we did not construct a post hoc index.
Inference and multiplicity	No public specification of co-primary outcomes, clustering, randomization inference, wild bootstrap, or multiplicity.	ZIP-clustered CR1S, within-wave RI, null-imposed WCB, and sample-specific Romano–Wolf families recorded in versioned ledgers.	Specified in analysis pipeline, not preregistered.
Registry status	Version 1.0; public record remained *In development* when independently checked in August 2026.	The present paper uses the fixed sample of 648 respondents and the completed January 2026 scoring window.	The record remains open until the broader project is closed.

Notes: The crosswalk distinguishes the public registration from the paper's final analysis. The persistence question and three-arm design were registered; the exact current estimands and inferential families were not. It uses only public version 1.0 and does not infer hidden fields. "Specified later" means documented in the versioned replication pipeline, not preregistered. Registry citation: Huseynov (2025).

Appendix B Sample construction and treatment balance

The main study includes 648 respondents. The December follow-up contains 341 completed submissions, but only 339 with identifiable Prolific IDs. One subject submitted three entries. Two additional entries with Prolific IDs do not match the main study data, leaving 337 linked follow-up respondents. These linkage rules are applied before examining treatment effects.

Table 2 reports selected pre-treatment characteristics and the study's balance tests. The within-wave permutation test does not reject joint balance across the registered covariates. Age is the clearest individual imbalance: respondents in the Natural treatment are younger than respondents in the Baseline arm. The covariate-adjusted specifications reported below show that this difference does not explain the treatment effects.

Table 2: Design and balance across treatment arms

	Baseline mean	Economic mean	Natural mean	Econ.−Base (SE)	Nat.−Base (SE)
Panel A. Pre-treatment beliefs					
Prior increase probability (pp)	63.32	60.77	59.28	−2.50 (3.44)	−4.02 (3.30)
Prior entropy (normalized)	0.62	0.66	0.66	0.04 (0.02)	0.03 (0.02)
Prior mean growth (pp)	0.99	0.70	0.71	−0.29 (0.20)	−0.28 (0.19)
Panel B. Property and finances					
Property value midpoint ($000)	392.3	409.0	396.4	16.19 (22.40)	3.82 (21.15)
Has mortgage	0.60	0.65	0.68	0.05 (0.05)	0.08 (0.05)
Considers selling	0.40	0.42	0.40	0.03 (0.05)	0.00 (0.05)
Does not rent property to others	0.99	0.98	0.96	−0.01 (0.01)	−0.02 (0.02)
Panel C. Demographics					
Age (years)	49.22	46.81	44.63	−2.46 (1.26)	−4.61 (1.29)
Woman	0.62	0.61	0.61	−0.01 (0.05)	−0.01 (0.05)
College degree or above	0.65	0.63	0.64	−0.02 (0.05)	−0.01 (0.05)
Household income $\geq$ $50,000	0.73	0.77	0.75	0.03 (0.04)	0.01 (0.04)
Respondents	226	212	210	Total 648; 411 ZIP clusters	
Panel D. Design tests					
Within-wave permutation omnibus over 26-covariate registry				stat. 0.552, $p = 0.162$	
Treatment arm × field wave (χ^2, df 4)				$p = 0.989$	
December follow-up response: Economic vs. Baseline				0.00 (0.05), $p = 0.931$	
December follow-up response: Natural vs. Baseline				0.07 (0.05), $p = 0.136$	

Notes: Selected rows from the versioned 26-covariate balance registry; the complete version is reported in the Supplementary Results Archive. Differences are coefficients from regressions of each covariate on arm indicators with field-wave fixed effects; standard errors are ZIP-clustered CR1S. Covariate-specific N varies with item availability: 648 for the selected belief and property rows, 643 for age, and 646 for the other selected demographic rows. No covariate is imputed. The omnibus test permutes arm labels within field wave (9,999 draws). Age shows the largest single imbalance: the Natural arm is 4.61 years younger than Baseline (joint $p = 0.002$). The post-treatment risk and patience measures are excluded from the omnibus and appear only as arm-invariance diagnostics. The registered covariate-adjusted specifications are insensitive to this difference.

Appendix C Prior beliefs and market benchmarks

Table 3 reports the variables most closely related to respondents' prior probability of a ZIP price increase. We estimate each group of variables separately and then include all four groups in the same regression. Property characteristics provide the strongest joint explanatory contribution, although the models have limited out-of-sample predictive power. The complete unscreened coefficient tables, including the entropy outcome and common-sample versions of each specification, appear in the Supplementary Results Archive.

Table 3: Determinants of the prior probability of a ZIP price increase

	Within block	All blocks
Block A. Weather and NFIP		
2024 average monthly precipitation	-3.750^{**} (1.495)	-2.515 (1.635)
Log(1 + NFIP policies in force, 2025-07-01)	-5.461^{***} (1.871)	-4.553 (3.075)
Log(1 + premium per insured unit)	-2.453 (2.259)	-1.716 (2.316)
10-yr claim units per 1,000 policies (annualized)	1.997 (2.429)	3.506 (2.340)
Current month-to-date precipitation	2.083 (1.926)	2.325 (2.187)
Block B. ZIP market history		
Trailing 12-month ZHVI growth	7.980^{***} (1.586)	5.611 (2.209)
Trailing 12-month growth volatility	0.991 (1.444)	-0.916 (1.579)
Log ZIP ZHVI level	0.163 (1.407)	-0.560 (2.215)
Log ZIP population	-1.866 (1.457)	-1.344 (2.076)
Block C. Property characteristics		
Purchased within last 5 years	10.762^{***} (3.226)	10.467^{*} (3.423)
Built within last 10 years	-8.020^{**} (3.917)	-10.544 (4.220)
Self-reported market value	2.611^{*} (1.578)	5.265 (1.990)
Positive probability of selling	-5.911^{**} (2.858)	-6.449 (2.879)
House or townhouse	-6.076 (4.415)	-6.083 (4.729)
Has a mortgage	5.397 (5.112)	2.731 (5.585)
Block D. Personal characteristics		
Age	-1.906 (1.467)	-0.371 (1.514)
Woman	3.353 (2.900)	2.087 (2.823)
Bachelor's degree or higher	2.490 (3.431)	0.452 (3.450)
Household income $\geq$ \$50,000	-4.645 (3.655)	-5.012 (3.793)
Employed full-time	0.135 (3.138)	0.867 (3.189)
N (within-block cols. A–D)	648 / 648 / 643 / 603	
ZIP clusters (A–D)	411 / 411 / 408 / 390	
N; clusters (all blocks)	598; 387	
R^2 (all blocks)	0.133 (adj. 0.063)	

Notes: Outcome is the prior probability (0–100) that own-ZIP prices rise over twelve months. Continuous determinants are standardized; coefficients are per-SD effects in probability points. "Within block" estimates each block separately on its maximum available sample; "All blocks" is the joint model on the common sample. All specifications include survey-month fixed effects; CR1S standard errors clustered by ZIP in parentheses. The table shows 20 of the 36 registered determinants (those with $|t| > 1$ in either column or of direct economic interest); the complete table, the entropy counterpart, and common-sample variants of every column are in the Supplementary Results Archive. Within-block markers use raw CR1S p-values; all-block markers use Romano–Wolf adjusted p-values from the registered determinant family. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$; markers on binary determinants with a cell below 30 observations are suppressed by rule.

Table 4 compares the four groups of variables using joint tests, partial R^2, Shapley shares, and out-of-sample prediction. These measures answer different questions. The joint tests ask whether a block contributes after the other blocks are included. The Shapley shares divide the model's in-sample explanatory power across blocks. The cross-validation column asks whether the same block improves prediction for respondents in ZIP codes that were not used to fit the model.

Table 4: Relative importance of determinant blocks (common sample, $N = 598$)

Block	Joint Wald p (Holm)	Partial R^2 (full model)	Shapley share (%)	CV R^2 gain over month FE
Panel A. Prior increase probability (full-model adj. $R^2 = 0.063$)				
Weather and NFIP	0.764	0.025	30.4	−0.019
ZIP market history	0.318	0.015	22.9	+0.025
Property characteristics	0.002	0.050	36.1	+0.010
Personal characteristics	0.764	0.014	10.6	−0.024
Panel B. Prior entropy (full-model adj. $R^2 = 0.035$)				
Weather and NFIP	0.189	0.034	30.6	−0.061
ZIP market history	0.189	0.014	13.5	−0.008
Property characteristics	0.064	0.030	28.8	−0.014
Personal characteristics	0.092	0.030	27.1	−0.013

Notes: Shapley shares decompose the full model's R^2 across blocks (Shorrocks, 2013) and sum to 100 within outcome. Joint Wald tests are CR1S-based with Holm adjustment across the four blocks within outcome (Holm, 1979). The final column reports the gain in ZIP-grouped cross-validated R^2 from adding the block to survey-month fixed effects, averaged over 20 repetitions of five-fold cross-validation. The full models themselves do not predict out of sample (CV R^2 -0.050 for the increase prior, -0.100 for entropy).

Figure 1 presents the same comparison visually. Property characteristics account for the largest share of the explained variation in the prior probability of a price increase. However, the negative cross-validated R^2 for the full model shows that the measured characteristics do not provide a reliable prediction rule outside the estimation sample.

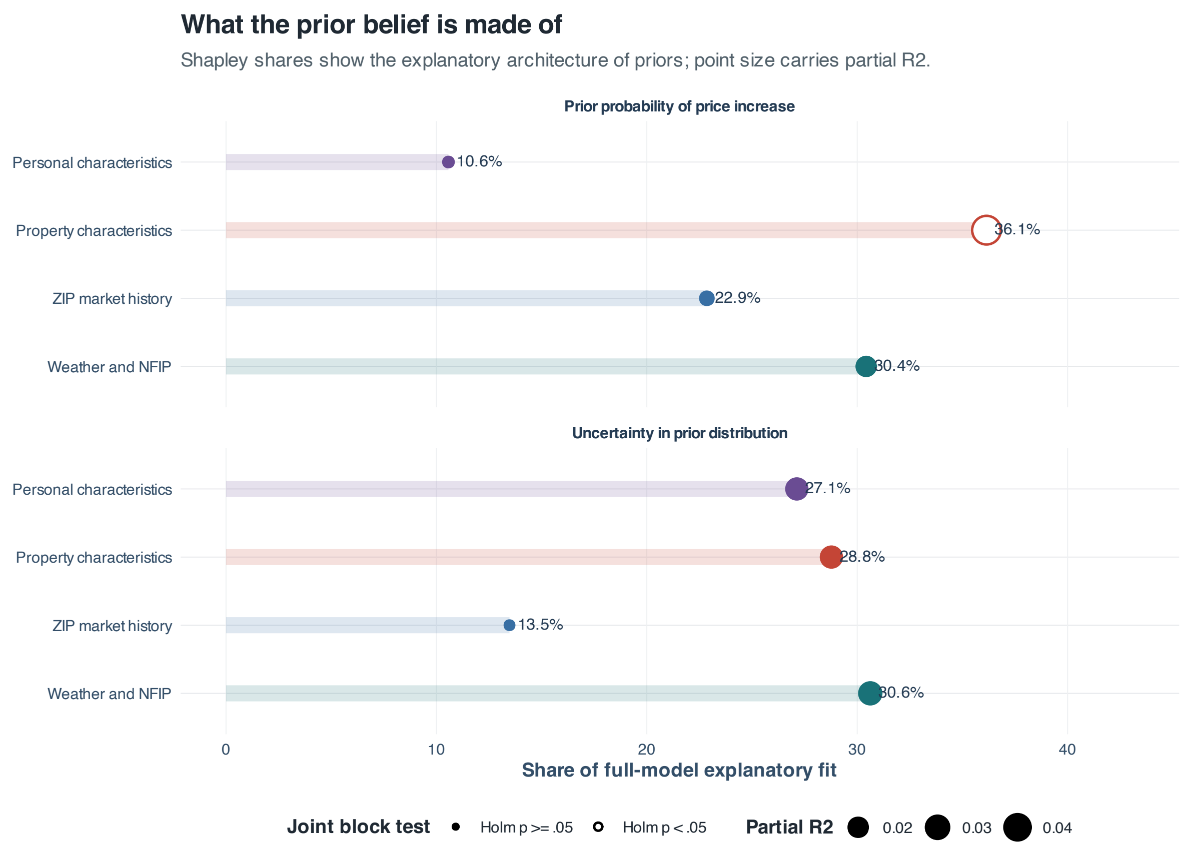

Figure 1: Measured characteristics and prior beliefs. *Notes:* The figure compares the four groups of variables using their contribution to the model's in-sample fit, their partial R^2, and the evidence from the adjusted joint tests. The exercise is descriptive. It shows which observed characteristics are associated with prior beliefs but does not assign those associations a causal interpretation.

Table 5 compares owner beliefs with four simple forecasts based on Zillow market history. The raw forecasts are not restricted to the survey's elicitation range. We therefore also censor them to the same support and map them to the eleven survey bins. This adjustment makes the owner and statistical forecasts more comparable. The statistical benchmarks remain more accurate after the adjustment.

Table 5: Forecast accuracy of owner beliefs and publication-lag-truncated statistical benchmarks

Forecaster	RMSE	Censored RMSE	MAE	Bias	Point-bin RPS	N
Owner prior mean	6.178	6.178	5.451	+5.380	5.067	642
Owner posterior mean	6.096	6.096	5.352	+5.265	4.894	642
Banked random walk	3.393	3.526	2.739	+2.596	1.816	642
Trailing 12-month momentum	1.691	2.613	1.288	+0.481	0.836	642
ZIP AR(1)	2.189	2.610	1.671	−0.881	0.737	642
Equal-weight mixture	1.709	2.563	1.283	+0.732	0.740	642

Notes: Accuracy against realized own-ZIP growth, $100 \cdot (\text{ZHVI}_{2026\text{-}01}/\text{ZHVI}_{2025\text{-}01} - 1)$, for the 642 respondents with Zillow coverage. Owner means are interval midpoints of the elicited 11-bin distribution. Censored RMSE clips every forecast to $[-4.5, +4.5]$. Point-bin RPS maps a point forecast to a deterministic elicitation bin; lower is better. Benchmarks use only Zillow history through the last complete month released at least 45 days before each respondent's survey date; the AR(1) is fitted per ZIP on pre-cutoff log changes; the mixture is the equal-weight mean of the three statistical benchmarks. Benchmarks are computed from the May 2026 ZHVI record after applying the lag rule. Because Zillow can revise historical observations, this procedure does not recreate a strictly real-time data vintage. Four AR windows contain a missing intervening calendar month. Panel B of Table 10 reports the results after excluding them. The exploratory reversion-carry forecast remains in the Supplementary Results Archive but is omitted from this table. Positive bias means that a forecast exceeds the realized outcome.

Appendix D Immediate treatment-effect diagnostics

The average treatment effects in the main text are not tied to one narrow sample choice. Table 6 repeats the prior-adjusted specification after applying a series of transparent sample restrictions. Both treatments continue to reduce the probability assigned to a price increase. The smallest Economic-treatment estimate is -9.11 percentage points, and the smallest Natural-treatment estimate is -6.29 points.

Table 6: Immediate treatment effects across alternative samples

Sample	Treatment	Effect (SE)	p-value	N
Full randomized sample	Economic	−11.45 (1.90)	< 0.001	648
	Natural	−7.29 (1.77)	< 0.001	648
Exclude five ZIP codes without a Census ZCTA match	Economic	−11.49 (1.91)	< 0.001	643
	Natural	−7.38 (1.79)	< 0.001	643
Exclude priors at zero or 100	Economic	−12.69 (2.29)	< 0.001	429
	Natural	−8.45 (2.15)	< 0.001	429
Exclude the longest one percent of survey durations	Economic	−11.52 (1.90)	< 0.001	641
	Natural	−7.14 (1.73)	< 0.001	641
Exclude the fastest five percent of survey durations	Economic	−11.76 (1.98)	< 0.001	615
	Natural	−7.51 (1.82)	< 0.001	615
Exclude respondents who placed mass in every prior bin	Economic	−9.11 (2.16)	< 0.001	506
	Natural	−6.29 (2.01)	0.002	506
Require a weather-data match	Economic	−11.51 (1.92)	< 0.001	639
	Natural	−7.35 (1.80)	< 0.001	639
Require a Zillow-data match	Economic	−11.49 (1.91)	< 0.001	642
	Natural	−7.25 (1.79)	< 0.001	642

Notes: The outcome is the posterior probability of a ZIP price increase. Each row compares a treatment with the Baseline arm while controlling for the prior and field wave. Effects are measured in percentage points, and standard errors are CR1S and clustered by ZIP. Each restriction removes only the observations described in the first column. The estimates remain negative across every sample, including the smallest estimate of −9.11 points for the Economic treatment after excluding respondents who placed positive prior probability in all eleven bins.

Table 7: Flexible adjustment and the specification range

Method	Treatment	Effect or range	Standard error	N
ZIP-grouped double machine learning	Economic	-12.01	1.94	438
	Natural	-7.88	1.90	434
Post-double-selection lasso controls	Economic	-11.10	1.99	648
	Natural	-6.73	1.75	648
ZIP-population weighting	Economic	-9.50	2.21	643
	Natural	-7.00	1.87	643
Leave one county out	Economic	$[-12.31, -10.61]$	—	57 omissions
	Natural	$[-7.87, -6.38]$	—	57 omissions
Range across the prespecified specification paths				
Change in increase probability	Economic	$[-13.11, -11.10]$	—	18 paths
	Natural	$[-8.66, -6.70]$	—	18 paths
Prior-adjusted posterior	Economic	$[-14.26, -11.24]$	—	18 paths
	Natural	$[-10.75, -7.01]$	—	18 paths

Notes: The first three methods allow the relationship between the outcome and observed characteristics to be more flexible than in the primary regression. In the double-machine-learning exercise, respondents from the same ZIP code remain together when the sample is divided for model fitting and out-of-sample prediction. The leave-one-county-out rows report the smallest and largest estimates obtained after omitting each represented county in turn. The final four rows summarize every combination of the registered sample, outcome, control, and timing choices. All estimates remain negative.

Table 8 asks whether the treatment response can be interpreted as the compression predicted by a simple common-signal model. Under that model, respondents with stronger priors should respond less to the same signal, producing a negative treatment interaction with the prior. The estimated interactions do not display that pattern consistently. We therefore use the experiment to estimate treatment effects on beliefs rather than to recover a structural signal weight.

Table 8: The prior–posterior slope does not flatten under treatment

Estimator	Arm	Slope interaction (SE)	95% CI	N
OLS, full sample	Economic	0.86 (0.62)	$[-0.36, 2.08]$	648
	Natural	1.30** (0.56)	$[0.20, 2.39]$	648
OLS, drop prior boundaries	Economic	0.95 (1.01)	$[-1.05, 2.94]$	429
	Natural	1.66* (0.89)	$[-0.08, 3.40]$	429
Fractional logit	Economic	0.05 (0.05)	$[-0.05, 0.15]$	648
	Natural	0.11** (0.05)	$[0.01, 0.21]$	648
Quantile, $\tau = 0.25$	Economic	−1.30 (0.84)	$[-2.95, 0.34]$	648
	Natural	−1.11** (0.44)	$[-1.97, -0.26]$	648
Quantile, $\tau = 0.50$	Economic	0.00 (0.32)	$[-0.63, 0.63]$	648
	Natural	0.00 (0.15)	$[-0.29, 0.29]$	648
Quantile, $\tau = 0.75$	Economic	2.00*** (0.59)	$[0.85, 3.15]$	648
	Natural	2.00*** (0.59)	$[0.84, 3.16]$	648

Notes: Each row reports the treatment × standardized-prior interaction from a posterior-on-prior regression for the increase probability, by estimator. A negative interaction would indicate compression of strong priors toward the message; none appears. Boundary-robust rows drop priors at 0 or 100 (219 respondents). Fractional logit coefficients are on the log-odds scale. Quantile rows use a ZIP-pairs cluster bootstrap (1,999 draws). The identical $\tau = 0.75$ point estimates reflect the discrete token grid used in the belief elicitation rather than duplicated calculations. OLS and fractional rows use CR1S ZIP-clustered standard errors; 411 clusters (312 in the boundary-robust rows). $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$ (unadjusted; this taxonomy is a specification exhibit, not a co-primary family).

We also examine whether treatment effects differ with measured hurricane exposure, NFIP conditions, recent ZIP-level price growth, home value, mortgage status, tenure, and prior uncertainty. None of the 26 interactions survives the family or global adjustment. These estimates do not show robust heterogeneity, but their minimum detectable effects are often large. The complete interaction tables and the timing and spatial-correlation checks are reported in the Supplementary Results Archive.

Appendix E Forecast accuracy and benchmark sensitivity

Table 9 reports two descriptive ways to interpret the improvement in forecast accuracy. Panel A translates the average log-odds shift into the precision respondents acted as if the treatments had. Panel B separates the change in ranked probability scores into calibration and resolution. Most of the improvement comes from calibration: respondents move the overall distribution closer to observed outcomes, but do not become substantially better at distinguishing which ZIP codes decline more than others.

Table 9: Implied treatment precision and the calibration–resolution decomposition

Panel A. Precision owners acted as if treatments had				
	Log-odds shift b (SE)	p	Implied $\hat{q}$ (SE)	N
Economic treatment	−1.132 (0.221)	< 0.001	0.756 (0.041)	648
Natural treatment	−0.302 (0.171)	0.079	0.575 (0.042)	648
Ex-post objective content	—	—	0.969	642
Panel B. Murphy decomposition of the within-arm RPS improvement				
	RPS improv.	Calibration	Resolution	N
Baseline	−0.147	−0.143	−0.003	225
Economic	0.383	0.390	−0.005	211
Natural	0.123	0.138	−0.012	206
Pooled	0.114	0.119	−0.004	642

Notes: Panel A is a reduced-form calibration, not a causal mechanism. b is the treatment coefficient from the prior-adjusted log-odds update regression; $\hat{q} = 1/(1 + e^b)$ with delta-method SEs and CR1S clustering by ZIP. The objective-content row is the share of matched respondent-ZIP observations (622 of 642) that declined. It is an ex-post state frequency, not a verified truth probability for either message sentence, and carries no SE. Panel B: within-arm prior-to-posterior RPS improvement decomposed into calibration, resolution, and a grouping residual (omitted; ≤ 0.003 everywhere) following Murphy (1973); positive = improvement.

Figure 2 shows the same decomposition for each study arm. The Economic treatment has the largest improvement, and this change comes mainly from the lower calibration penalty. The figure describes prior-to-posterior changes within each arm and should not be read as a separate causal decomposition.

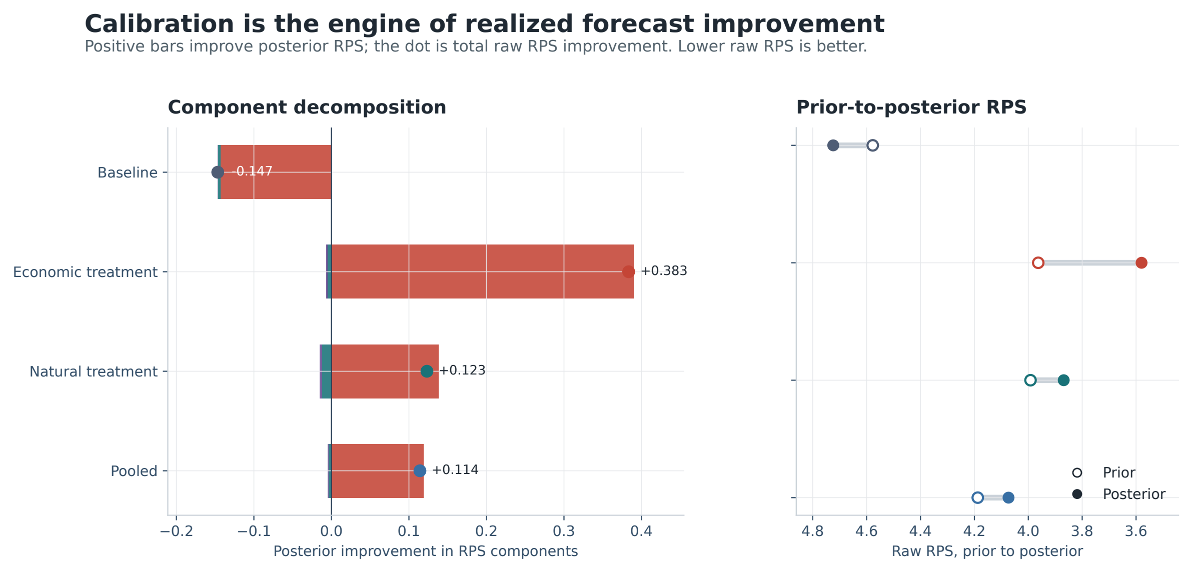

Figure 2: Calibration and resolution of the elicited price forecasts. *Notes:* Bars decompose the within-arm change in the ranked probability score into calibration, resolution, and the small remaining grouping component. Positive values indicate an improvement. Because each bar compares priors and posteriors within the same arm, the figure is descriptive rather than a treatment-effect decomposition.

Figure 3 places all immediate accuracy outcomes on a common display. The Economic treatment improves the probability assigned to the realized sign and the full-distribution ranked probability score. The Natural treatment moves the same measures in the same direction, although the evidence is weaker. Neither treatment places more probability on the exact realized bin.

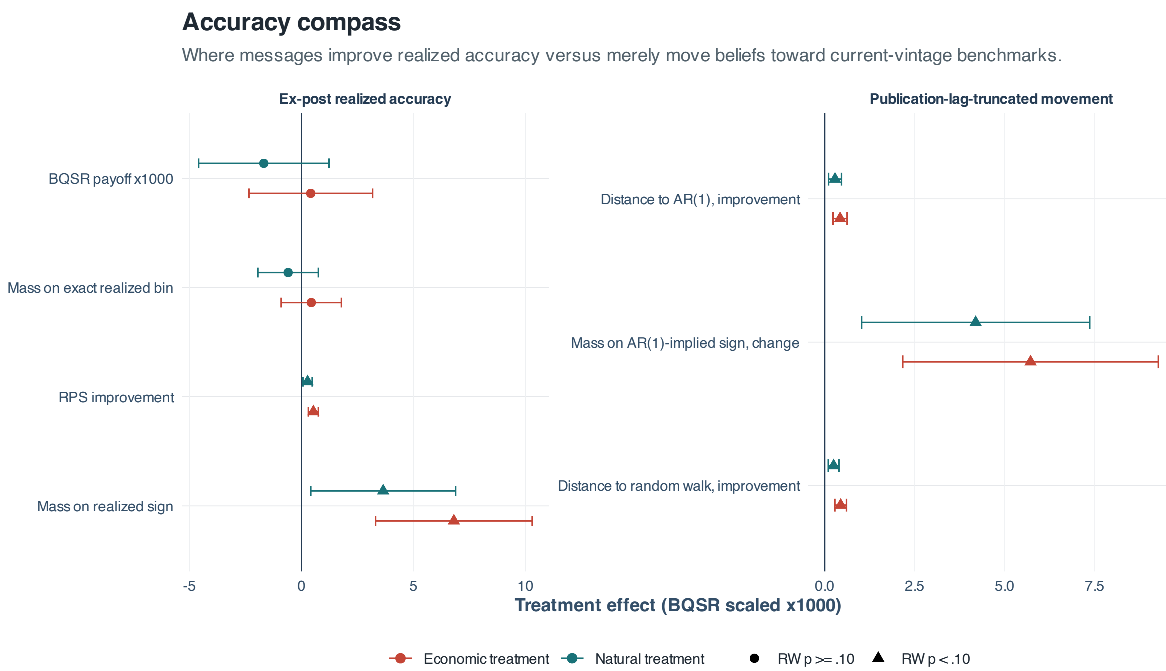

Figure 3: Treatment effects on forecast accuracy. *Notes:* The figure reports treatment effects on ex-post accuracy and movement toward forecasts constructed from Zillow observations dated early enough to satisfy each respondent's publication-lag cutoff. Markers also show whether each estimate remains statistically distinguishable from zero after the paper's family adjustment.

Tables 10 and 11 examine whether the benchmark comparison depends on the assumed publication lag or on treating the benchmark as a point forecast. The statistical forecasts outperform owner priors under 30-, 45-, and 60-day lags. They also outperform owner priors when converted to probability distributions over the same eleven bins used in the survey.

Table 10: Forecast benchmarks under alternative publication lags

Model	Publication lag	Raw RMSE	Censored RMSE	Point-bin RPS	N
Panel A. Alternative publication lags					
Banked random walk	30 days	2.908	3.129	1.456	642
Trailing twelve-month momentum	30 days	1.387	2.448	0.621	642
ZIP AR(1)	30 days	2.078	2.519	0.670	642
Equal-weight mixture	30 days	1.354	2.415	0.553	642
Banked random walk	45 days	3.393	3.526	1.816	642
Trailing twelve-month momentum	45 days	1.691	2.613	0.836	642
ZIP AR(1)	45 days	2.189	2.610	0.737	642
Equal-weight mixture	45 days	1.709	2.563	0.740	642
Banked random walk	60 days	3.483	3.608	1.922	642
Trailing twelve-month momentum	60 days	1.749	2.652	0.891	642
ZIP AR(1)	60 days	2.204	2.618	0.748	642
Equal-weight mixture	60 days	1.771	2.593	0.802	642
Panel B. Primary 45-day lag after excluding four AR windows with a calendar gap					
Banked random walk	45 days	3.384	3.518	1.812	638
Trailing twelve-month momentum	45 days	1.696	2.612	0.842	638
ZIP AR(1)	45 days	2.051	2.498	0.697	638
Equal-weight mixture	45 days	1.668	2.540	0.730	638

Notes: Each benchmark uses Zillow observations dated early enough to satisfy the publication lag shown in the second column. The primary analysis uses 45 days. Censored RMSE limits every point forecast to the survey support of $[-4.5, 4.5]$, while point-bin RPS places each benchmark forecast into one of the eleven elicitation bins. Lower values indicate greater accuracy. Panel A uses the same 642 respondents with realized Zillow outcomes. Panel B excludes four respondents whose AR estimation windows skip an intervening calendar month. The comparison is nearly unchanged.

Table 11: Probabilistic benchmark scores and paired comparisons with owner priors

Forecast	Mean forecast	Probability of decline	Full-distribution RPS	Owner-prior RPS difference
Banked random walk	−1.981	0.619	1.528	2.66
Trailing twelve-month momentum	−4.096	0.709	1.041	3.15
ZIP AR(1)	−5.458	0.868	0.582	3.61
Equal-weight mixture	−3.845	0.732	0.918	3.27
Owner prior distribution	0.804	0.303	4.188	—
Owner posterior distribution	0.689	0.330	4.073	—

Notes: The statistical models are converted to predictive distributions over the same eleven bins used in the survey. Lower ranked probability scores indicate greater accuracy. The final column reports the owner's prior score minus the benchmark score. Positive values therefore indicate that owner priors incur greater loss. All four paired differences have Holm-adjusted $p < 0.001$. The comparison uses 642 respondents.

Appendix F The December follow-up

About half of the original sample returns in December. Table 12 first reports response rates by randomized arm and then examines whether the December estimates change when we reweight or impute outcomes using observed main-study information. The resulting point estimates remain close to zero. These exercises reduce concern about selection related to observed characteristics, but they cannot rule out selection based on unobserved differences between respondents who return and those who do not.

Table 12: Follow-up response and missing-outcome sensitivity

Panel	Method or arm	Outcome	Estimate (SE)	95% interval	N
Panel A. Follow-up response					
	Baseline	Response rate	112 of 226 (49.6%)	—	226
	Economic	Response rate	106 of 212 (50.0%)	—	212
	Natural	Response rate	119 of 210 (56.7%)	—	210
Panel B. Inverse-probability weighting					
	Economic	Increase probability	−0.44 (4.35)	$[-8.97, 7.81]$	337
	Natural	Increase probability	−1.71 (4.16)	$[-9.86, 6.31]$	337
	Economic	Entropy	2.56 (2.49)	$[-2.36, 7.51]$	337
	Natural	Entropy	−0.16 (2.63)	$[-5.45, 4.92]$	337
Panel C. Multiple imputation					
	Economic	Increase probability	0.30 (5.25)	$[-10.01, 10.62]$	648
	Natural	Increase probability	−1.45 (5.12)	$[-11.52, 8.62]$	648
	Economic	Entropy	2.07 (3.10)	$[-4.02, 8.16]$	648
	Natural	Entropy	−0.03 (3.04)	$[-6.01, 5.94]$	648

Notes: Panel A reports the number of linked follow-up respondents in each randomized arm. The regression-adjusted response-rate differences relative to Baseline are 0.4 percentage points for the Economic treatment ($p = 0.931$) and 7.1 points for the Natural treatment ($p = 0.136$). Panel B gives greater weight to respondents whose observed characteristics predict a lower probability of returning. Panel C fills in missing December outcomes using repeated draws from models based on observed main-study information. These approaches address selection related to observed characteristics. They do not remove uncertainty about selection on unobserved characteristics.

Table 13: Additional diagnostics for the December follow-up

Diagnostic	Treatment or group	Estimate	Uncertainty or p-value	N
Panel A. Choice to use the December report for payment				
Replacement rate	All follow-up respondents	283 of 337 (84.0%)	—	337
Replacement rate	Baseline	84.8%	—	112
Replacement rate	Economic	84.0%	—	106
Replacement rate	Natural	83.2%	—	119
Joint treatment-arm test	Economic and Natural	—	$p = 0.947$	337
December RPS of replacers	Replacer difference	−0.05 (0.29)	$p = 0.858$	335
Panel B. Forecast accuracy over time				
Prior to immediate RPS improvement	Economic	0.394 (0.160)	Holm $p = 0.085$	335
	Natural	0.123 (0.155)	Holm $p = 1.000$	335
Immediate to December RPS improvement	Economic	−0.614 (0.333)	Holm $p = 0.332$	335
	Natural	−0.362 (0.344)	Holm $p = 1.000$	335
Prior to December RPS improvement	Economic	−0.220 (0.333)	Holm $p = 1.000$	335
	Natural	−0.239 (0.338)	Holm $p = 1.000$	335
Panel C. Sensitivity of the mechanical-reversion calculation				
Observed rebound	Economic	10.34 (4.90)	—	337
Predicted mechanical rebound	Economic	4.27	95% CI $[-0.91, 8.25]$	337
Observed minus predicted rebound	Economic	6.07	95% CI $[-3.56, 17.59]$	337
Observed rebound	Natural	3.92 (4.57)	—	337
Predicted mechanical rebound	Natural	2.26	95% CI $[-0.57, 4.79]$	337
Observed minus predicted rebound	Natural	1.65	95% CI $[-6.91, 11.06]$	337

Notes: In Panel A, respondents could choose whether their December report replaced both main-study reports for payment. Panel B reports improvements in the ranked probability score; positive values indicate greater accuracy and negative values indicate that an earlier gain was given back. Panel C repeats the main mechanical-reversion calculation. It uses the Baseline relationship between the immediate and December updates to predict how much of each treatment effect would ordinarily reverse. Because both updates are elicited with error, reversing the regression and placing the resulting coefficient on the same scale produces a much wider sensitivity range. Combining that uncertainty with follow-up selection widens the range further. We therefore use the displayed calculation as a benchmark rather than as an identified decomposition of reporting error and belief change.

The additional diagnostics in Table 13 address three questions raised by the main text. First, the choice to use the December report for payment is nearly identical across treatment arms. Second, the initial improvement in ranked probability scores is not detectable by December. Third, the mechanical-reversion calculation is highly sensitive to assumptions about measurement error. We therefore use that calculation only to show why an observed rebound cannot automatically be interpreted as forgetting.

Appendix G Downstream outcomes and insurance-related responses

Figure 4 compares downstream estimates after scaling each one by the minimum detectable effect for its own outcome. This scaling allows outcomes measured in different units to appear on the same figure. None of the treatment effects survives the paper's family adjustment.

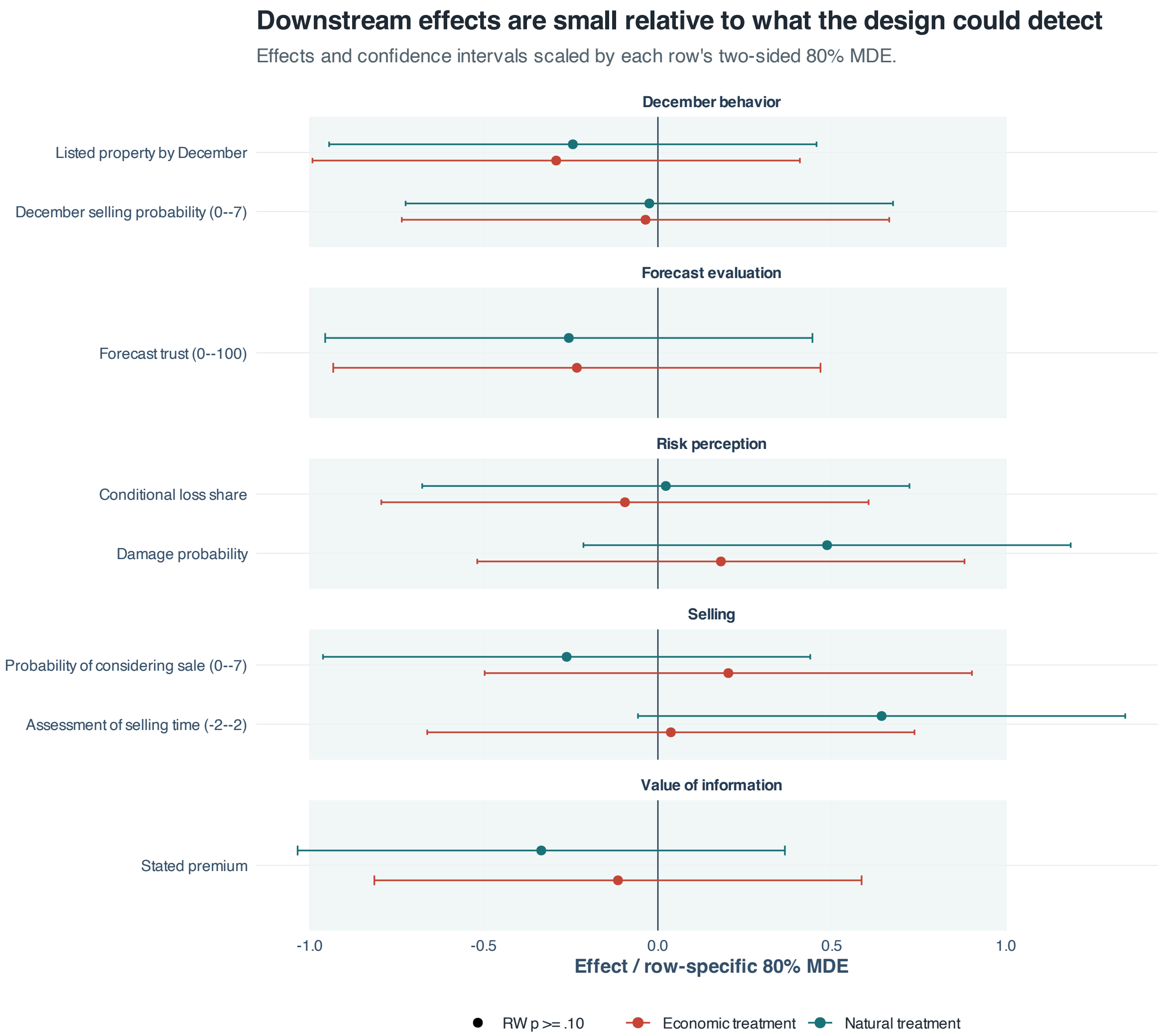

Figure 4: Downstream treatment effects relative to detectable effect sizes. *Notes:* Each estimate is divided by its outcome-specific two-sided 80-percent minimum detectable effect. Values near zero therefore indicate estimates that are small relative to what the study could reliably detect. The underlying outcomes include selling assessments, hurricane-loss beliefs, stated premium tolerance, forecast trust, and December selling outcomes.

Table 14 provides two additional pieces of context. The immediate selling-consideration effect lies inside the prespecified equivalence region after adjustment. Respondents also state annual premium tolerances that are high relative to their own subjective expected losses. The second comparison is descriptive because the premium question does not reproduce an actual insurance contract.

Table 14: Selling-intention equivalence and stated premium tolerance

Panel	Comparison	Estimate	90% interval	Adjusted p	Interpretation	N
Panel A. Equivalence test for immediate selling consideration						
	Economic versus Baseline	0.315	$[-0.599, 1.228]$	0.003	Equivalent	648
	Natural versus Baseline	-0.460	$[-1.489, 0.569]$	0.007	Equivalent	648
	Either treatment versus Baseline	-0.071	$[-0.900, 0.759]$	< 0.001	Equivalent	648
Panel B. Stated annual premium tolerance, as a share of home value						
	All respondents	13.84%	Median 5.00%	—	P90 42.60%	648
	Baseline	14.61%	Median 5.00%	—	P90 41.50%	226
	Economic	14.02%	Median 5.00%	—	P90 50.90%	212
	Natural	12.85%	Median 5.00%	—	P90 37.20%	210
Panel C. Stated tolerance and subjective expected loss						
	Mean stated tolerance	13.84%	—	—	—	648
	Mean subjective expected loss	6.95%	—	—	—	648
	Mean difference	6.90%	Median 3.00%	—	76.9% above expected loss	648

Notes: Panel A linearly rescales the eight-category selling-consideration item to 0–100 and tests whether each effect lies inside a fixed $[-2, 2]$ region. The adjusted p-values come from two one-sided tests with Holm correction across the two treatment-specific and pooled comparisons. The result supports a narrow equivalence statement for the immediate selling item only. Panel B describes the survey question asking what percentage of home value respondents would be willing to pay each year for hurricane-related insurance. Panel C defines subjective expected loss as reported damage probability multiplied by reported loss conditional on damage. These comparisons are descriptive and do not measure actual insurance demand.

An earlier analysis used a risk-neutral rule in which an owner purchases insurance whenever perceived damage probability exceeds a premium-to-loss threshold. We retain that exercise in the Supplementary Results Archive because it was part of the analysis history. It is not the model used in the paper. The square-root-utility exercise in Appendix H is the paper's canonical insurance calibration.

Appendix H Insurance and information value

This appendix presents the insurance calibration used in the main text. We first place respondents' stated premium tolerance beside an external market benchmark. We then show how the modeled insurance decision changes with the assumptions about wealth, premiums, and signal accuracy. The earlier risk-neutral threshold exercise remains in the Supplementary Results Archive, while the square-root-utility model below is the paper's primary decision exercise. The replication package records every external-premium, utility, signal, and home-value rule used in these calculations.

H.1 External premium benchmark

The November 2025 revision of Keys and Mulder (2024) reports national 2024 mortgage-escrow premiums of \$862.79 at p10, \$2,712.98 on average, and \$5,131.79 at p90. Relative to purchase price, the corresponding shares are 0.25, 0.96, and 1.98 percent. These values are external scale markers, not Florida quotes. The \$425 increment is the paper's estimate for top-decile catastrophe-exposure ZIP codes and is not applied as a universal shock.

Table 15: Stated premium tolerance relative to external market benchmarks

Benchmark	External value	Above	Below	Median headroom
Panel A. Premium shares				
2024 p10 premium share	0.25%	98.3%	1.7%	4.75 pp
2024 mean premium share	0.96%	98.3%	1.7%	4.04 pp
2024 p90 premium share	1.98%	82.9%	17.1%	3.02 pp
Panel B. Mean annual premium in dollars				
Lower home-value rule	$2,712.98	86.6%	13.4%	$12,287
Bracket midpoint	$2,712.98	93.4%	6.6%	$16,037
Adjacent-width upper rule	$2,712.98	97.8%	2.2%	$19,787

Notes: $N = 648$. External values are from the November 2025 revision of Keys and Mulder (2024), a national mortgage-escrow sample; they are not respondent-specific Florida quotes. Panel A compares the stated share directly. Panel B converts stated shares using survey home-value brackets. The open top category uses the adjacent-width rule here; the $1.2 million alternative is archived. Midpoint estimates are shown only beside lower and upper sensitivity.

The dollar panel converts whole-percentage stated tolerance using the lower endpoint, midpoint, and adjacent-width upper endpoint of each survey home-value bracket. The open top category is also evaluated at \$1.2 million in the archived outputs. Relative to the external mean premium, 86.6 percent are above even under the lower endpoint, 2.2 percent are below even under the upper endpoint, and 11.3 percent cannot be classified without a within-bracket value. The midpoint share above is 93.4 percent; 92.9 percent remain above after adding the top-decile \$425 increment. Only 0.46 percent lie between those two midpoint thresholds. These are bracket sensitivities, not measured reservation payments.

H.2 Implemented insurance model

All monetary quantities are normalized by home value. For damage probability d, conditional loss share l, premium share c, and wealth ratio λ, the insured and uninsured expected utilities are

$$U_I(\lambda, c) = \sqrt{\lambda - c}, \tag{1}$$

$$U_U(\lambda, d, l) = (1 - d)\sqrt{\lambda} + d\sqrt{\lambda - l}. \tag{2}$$

The no-information value is $V_0 = \max\{U_I, U_U\}$, and the reservation-premium share is

$$c^* = \lambda - \left[(1 - d)\sqrt{\lambda} + d\sqrt{\lambda - l}\right]^2. \tag{3}$$

The model requires $\lambda > l$ and $\lambda > c$. At $\lambda = 1$, 29 observations have conditional loss at modeled wealth and are classified as invalid rather than truncated or repaired. All observations are valid at larger declared wealth ratios.

For sensitivity s and specificity t, the warning and reassurance probabilities are

$$q_W = sd + (1 - t)(1 - d), \qquad q_R = (1 - s)d + t(1 - d), \tag{4}$$

and the posteriors are $d_W = sd/q_W$ and $d_R = (1 - s)d/q_R$. The implementation verifies $q_W d_W + q_R d_R = d$ numerically. Conditional on each signal, the insurance choice is re-optimized. Thus

$$V_I = q_W \max\{U_I, U_U(\lambda, d_W, l)\} + q_R \max\{U_I, U_U(\lambda, d_R, l)\}. \tag{5}$$

Actionability equals one only when the optimizing choice differs across the two signals. The monetary-equivalent share κ solves

$$V_0(\lambda + \kappa, d, l, c) = V_I, \tag{6}$$

where the wealth increment enters every insured and uninsured state and the no-information choice is re-optimized. This definition prevents the comparison from holding a suboptimal baseline action fixed.

Table 16: Square-root insurance calibration across wealth ratios

λ	Valid N	Actionable	Zero VOI	Mean κ	Median gap	Stated $\geq c_i^*$
1.0	619	43.8%	56.2%	0.145%	3.49 pp	79.8%
1.5	648	42.4%	57.6%	0.143%	3.00 pp	77.5%
2.0	648	42.7%	57.3%	0.144%	3.00 pp	77.5%
3.0	648	43.1%	56.9%	0.144%	3.00 pp	77.8%

Notes: Square-root utility, $W_i = \lambda H_i$, binary full insurance, the \$2,712.98 external mean premium, midpoint home values, and symmetric signal accuracy 0.85. κ is the home-value share added to wealth in every state before the no-information choice is re-optimized. At $\lambda = 1$, 29 observations with conditional loss at modeled wealth are invalid and are reported rather than truncated.

H.3 Signal interpretation and model sensitivity

The symmetric calibration sets $s = t = a$, where a is either fixed or a respondent's reported forecast accuracy. That mapping is maintained, not identified. If reported accuracy instead constrains only unconditional correct classification, then

$$a = sd + t(1 - d), \tag{7}$$

which admits many sensitivity-specificity pairs. The implementation uses a dense enumeration of the feasible set for each respondent. It also reports literal values below one half, an equivalent relabeling of anti-informative signals, and a conservative rule that sets those reports to one half.

Table 17: Signal-structure sensitivity at $\lambda = 1.5$

Signal rule	Actionability	Mean κ	Zero VOI	Valid N
Fixed symmetric $a = .60$	10.0%	0.008%	90.0%	648
Fixed symmetric $a = .70$	22.4%	0.038%	77.6%	648
Fixed symmetric $a = .80$	34.9%	0.099%	65.1%	648
Fixed symmetric $a = .85$	42.4%	0.143%	57.6%	648
Fixed symmetric $a = .90$	52.5%	0.203%	47.5%	648
Individual accuracy: literal	29.2%	0.093%	70.8%	648
Individual accuracy: relabel below .50	29.2%	0.093%	70.8%	648
Individual accuracy: set below .50 to uninformative	21.9%	0.078%	78.1%	648
Asymmetric feasible envelope	4.9–69.4%	0.019–0.276%	30.6%	648

Notes: Mean external premium, midpoint home value, square-root utility, and full insurance. The last row spans 401 sensitivity/specificity pairs per respondent satisfying $a_i = s_i d_i + t_i(1 - d_i)$. Its zero entry is the necessarily-zero share, not a point-calibration zero mass. Literal and relabeled anti-informative symmetric signals have identical information value because only their names change.

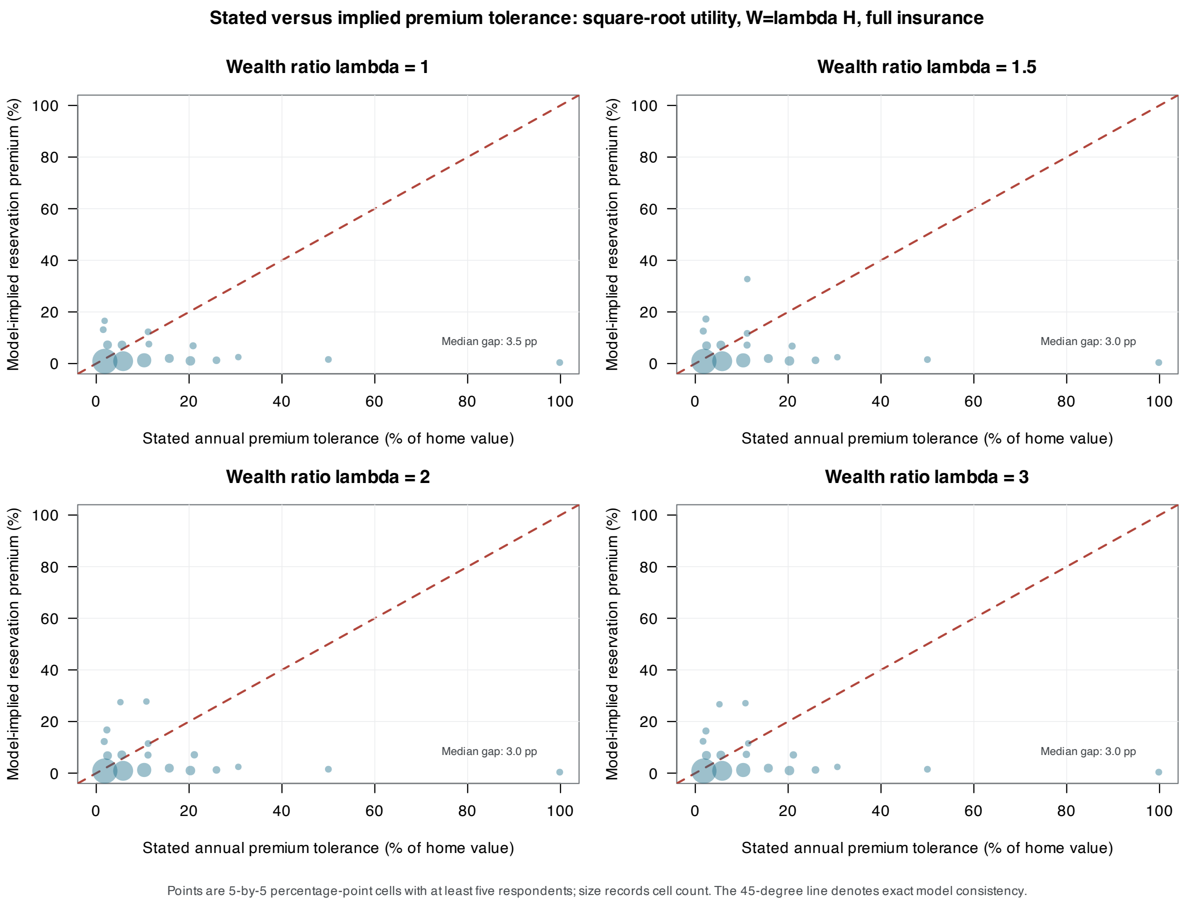

Figure 5: Stated and model-implied premium tolerance. *Notes:* Square-root utility, $W = \lambda H$, and full insurance. Stated tolerance is not revealed demand, and the distance from the model cutoff is not a measure of irrationality. Points are aggregate 5-by-5 percentage-point cells containing at least five respondents; point size records the cell count. The public plot-data file reproduces the figure without respondent rows.

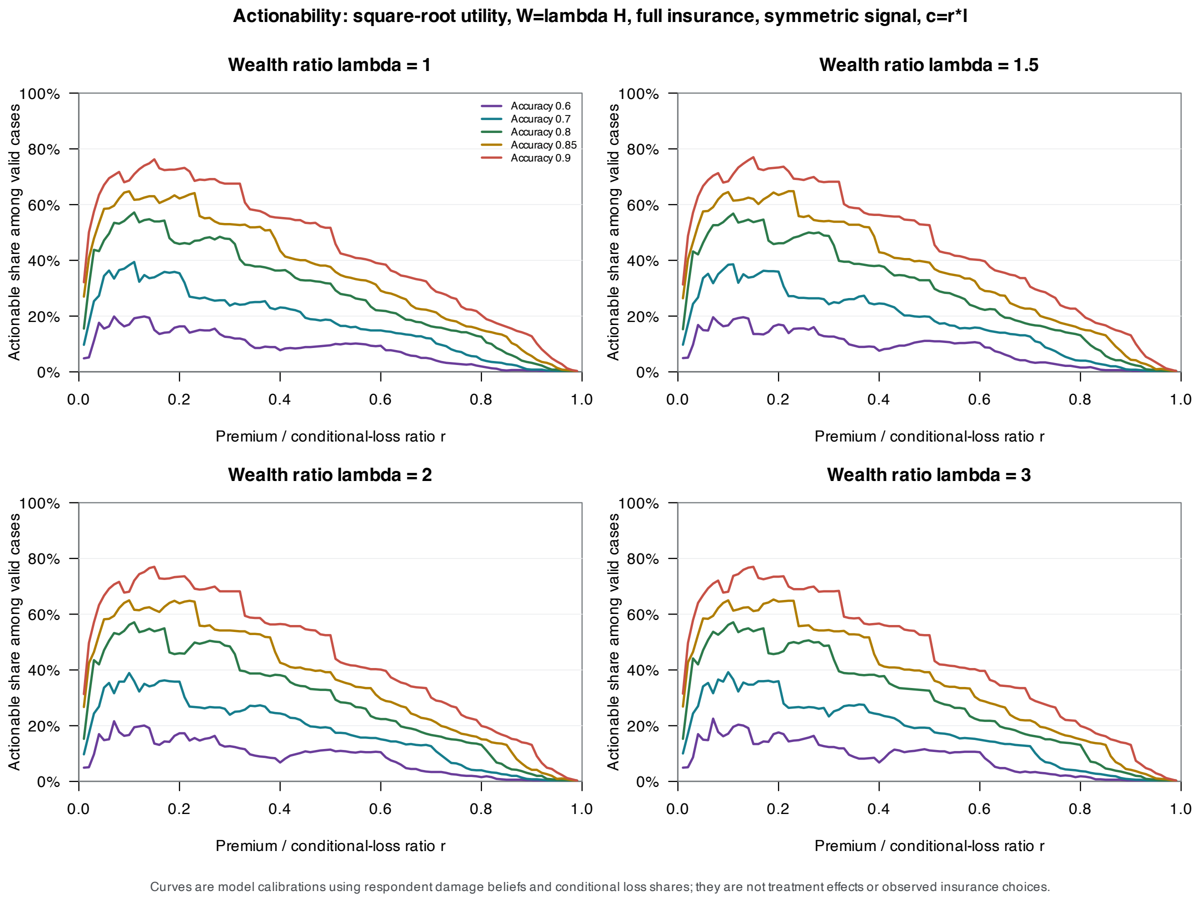

Figure 6: Actionability over premium-loss ratios and symmetric signal accuracy. *Notes:* Each surface point is model implied under the declared square-root/full-insurance calibration. Reported accuracy does not identify symmetric signal precision.

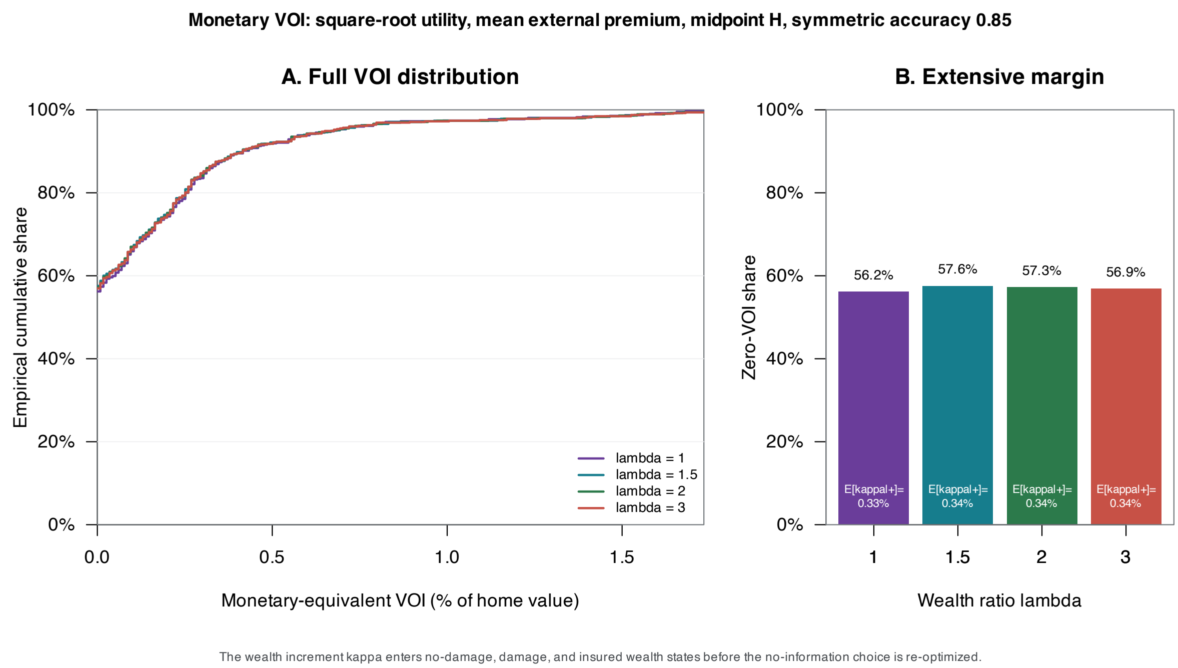

Figure 7: Distribution and zero mass of monetary-equivalent information value. *Notes:* κ is a share of home value. The displayed concentration is conditional on the external mean premium, midpoint home values, square-root utility, full insurance, and symmetric accuracy 0.85. Curves are drawn from the public aggregate ECDF grid rather than a retained respondent-level panel.

H.4 Descriptive mortgage split and randomized-message family

Mortgage status is not randomized. The split below is included as a descriptive diagnostic, not a heterogeneity estimand. The model also abstracts from actual mortgage covenants, coverage mandates, flood and wind exclusions, deductibles, insurer availability, and contract limits.

Table 18: Structural calibration by mortgage status

λ	Group	Valid N	Actionable	Zero VOI	Mean κ
1.0	Mortgaged	399	43.6%	56.4%	0.119%
1.0	Mortgage-free	220	44.1%	55.9%	0.193%
1.5	Mortgaged	415	42.4%	57.6%	0.115%
1.5	Mortgage-free	233	42.5%	57.5%	0.194%
2.0	Mortgaged	415	42.7%	57.3%	0.115%
2.0	Mortgage-free	233	42.9%	57.1%	0.195%
3.0	Mortgaged	415	42.7%	57.3%	0.115%
3.0	Mortgage-free	233	43.8%	56.2%	0.196%

Notes: Descriptive subgroup results under the main structural calibration. Mortgage status is not randomized, and this table is not a causal heterogeneity test.

Table 19: Randomized-message effects on model-implied insurance outcomes

λ	Outcome and arm	Effect	CR1 SE	Uniform 95% interval	RI p	Max-t p	N
1.0	Actionability: Economic	-4.384	4.853	[-17.165, 8.396]	0.366	0.918	619
1.0	Actionability: Natural	-5.095	4.735	[-17.839, 7.648]	0.286	0.850	619
1.0	Monetary-equivalent VOI share: Economic	-0.026	0.027	[-0.101, 0.049]	0.354	0.907	619
1.0	Monetary-equivalent VOI share: Natural	0.006	0.031	[-0.068, 0.081]	0.825	1.000	619
1.0	Stated minus square-root implied premium share: Economic	-0.403	1.873	[-5.458, 4.651]	0.835	1.000	619
1.0	Stated minus square-root implied premium share: Natural	-1.403	1.911	[-6.465, 3.658]	0.463	0.969	619
1.5	Actionability: Economic	-4.832	4.829	[-17.614, 7.950]	0.315	0.877	619
1.5	Actionability: Natural	-5.055	4.782	[-17.825, 7.714]	0.292	0.856	619
1.5	Monetary-equivalent VOI share: Economic	-0.026	0.027	[-0.101, 0.049]	0.359	0.913	619
1.5	Monetary-equivalent VOI share: Natural	0.006	0.031	[-0.069, 0.081]	0.834	1.000	619
1.5	Stated minus square-root implied premium share: Economic	-0.496	1.855	[-5.522, 4.530]	0.794	1.000	619
1.5	Stated minus square-root implied premium share: Natural	-1.508	1.895	[-6.539, 3.524]	0.429	0.954	619
2.0	Actionability: Economic	-5.300	4.829	[-18.079, 7.478]	0.269	0.828	619
2.0	Actionability: Natural	-5.522	4.750	[-18.281, 7.237]	0.251	0.802	619
2.0	Monetary-equivalent VOI share: Economic	-0.026	0.027	[-0.101, 0.049]	0.360	0.915	619
2.0	Monetary-equivalent VOI share: Natural	0.006	0.031	[-0.069, 0.080]	0.846	1.000	619
2.0	Stated minus square-root implied premium share: Economic	-0.511	1.850	[-5.528, 4.506]	0.786	1.000	619
2.0	Stated minus square-root implied premium share: Natural	-1.525	1.890	[-6.548, 3.498]	0.422	0.952	619
3.0	Actionability: Economic	-5.756	4.832	[-18.545, 7.033]	0.232	0.771	619
3.0	Actionability: Natural	-5.978	4.775	[-18.755, 6.799]	0.214	0.741	619
3.0	Monetary-equivalent VOI share: Economic	-0.026	0.027	[-0.101, 0.049]	0.360	0.915	619
3.0	Monetary-equivalent VOI share: Natural	0.005	0.031	[-0.070, 0.080]	0.859	1.000	619
3.0	Stated minus square-root implied premium share: Economic	-0.521	1.846	[-5.531, 4.489]	0.782	1.000	619
3.0	Stated minus square-root implied premium share: Natural	-1.537	1.886	[-6.552, 3.479]	0.418	0.950	619

Notes: The table uses the sample for which the structural calculation is defined. It assumes the mean external premium, midpoint home values, square-root utility, full insurance, and symmetric signal accuracy of 0.85. Each estimate compares a treatment with the Baseline arm and includes field-wave fixed effects. The 95-percent intervals and adjusted p-values come from one 9,999-draw within-wave randomization family covering all 24 contrasts; the joint test gives $p = 0.741$. The sample excludes 29 respondents (9 in the Natural treatment, 10 in the Economic treatment, 10 in the Baseline arm) whose reported conditional loss equals all modeled wealth when $\lambda = 1$. These estimates therefore describe the 619 respondents for whom every displayed calibration is defined. Coherence-gap and κ effects are in percentage points of home value; actionability effects are percentage-point changes in the share of respondents whose choice differs across signals.

The treatment table estimates treatment-minus-Baseline contrasts on coherence gaps, actionability, and κ over the common structural sample. Its 24 hypotheses form one 9,999-draw within-wave randomization family with simultaneous intervals and max-t p-values. The global p-value is 0.741. This joint test prevents selection of a favorable wealth ratio or model-implied outcome after inspecting pointwise estimates. It does not convert calibrated outcomes into direct welfare effects of the messages. The common-valid sample excludes 29 respondents whose post-treatment conditional-loss report equals 100 percent (10 Baseline, 10 Economic, and 9 Natural). Point estimates are therefore conditional on this selected sample rather than population average treatment effects; the sharp-null within-wave randomization test remains exact because the validity set is assignment-invariant under the null.

H.5 Interpretive boundaries

The extension distinguishes five evidentiary classes. Randomized arm contrasts are causal for the fielded communication packages. Sample distributions and mortgage splits are descriptive. Survey quantities placed beside Keys–Mulder statistics are externally benchmarked but not contract-matched. Insurance cutoffs, actionability, and κ are model implied. Ranges across home-value, wealth, premium, and signal rules are sensitivity analyses. No result identifies actual insurance demand, affordability, insurer supply, coverage, capitalization, or willingness to pay for the randomized messages.

Appendix I Guide to the Supplementary Archive

The Online Appendix reports the analyses most helpful for evaluating the paper's economic argument. The Supplementary Results Archive preserves the complete generated output, including tables that are useful for auditing but too detailed to reproduce here. Its five parts cover the determinants of prior beliefs, immediate updating, the December follow-up, forecast accuracy, and downstream decision outcomes. The archive also contains the paper-wide multiplicity maps and the earlier risk-neutral insurance exercise. No result removed from the Online Appendix has been deleted from the release.

The replication package contains the data and scripts used to rebuild both documents. Reader-facing labels in this appendix follow the terminology of the paper. The archive retains original computational identifiers so that each result can be traced directly to its producing script and machine-readable output.

References

Sture Holm. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2):65–70, 1979.

Samir Huseynov. Housing market expectations and environment. AEA RCT Registry, 2025. AEARCTR-0016346, first published July 11, 2025.

Benjamin J. Keys and Philip Mulder. Property insurance and disaster risk: New evidence from mortgage escrow data. NBER Working Paper 32579, National Bureau of Economic Research, 2024. URL `https://www.nber.org/papers/w32579`. June 2024, revised November 2025.

Allan H. Murphy. A new vector partition of the probability score. *Journal of Applied Meteorology*, 12(4):595–600, 1973.

Anthony F. Shorrocks. Decomposition procedures for distributional analysis: A unified framework based on the shapley value. *Journal of Economic Inequality*, 11(1):99–126, 2013.