Geographic Price Extrapolation, Learning, and Housing Search: Evidence from Danish Movers

Matteo Saccarola*

May, 2026

[CLICK HERE FOR LATEST VERSION]

Abstract

Using population-wide Danish administrative registers on housing transactions, I document an asymmetric, hockey-stick relationship between origin market prices and overpayment for comparable homes. Quantitatively, the elasticity of overpayment with respect to the origin-destination price difference is 3.9 percent ($p < 0.01$) when movers relocate from more expensive to cheaper housing markets. In contrast, buyers moving to more expensive locations exhibit little systematic overpayment, and their purchase prices are unrelated to prices at origin. I interpret these patterns through a housing search model in which buyers enter with price beliefs anchored in their origin market and update those beliefs gradually during search. Despite homogeneous learning, endogenous stopping generates the observed asymmetry at purchase: buyers predisposed to overpay transact quickly before fully learning the local price level, while those predisposed to underpay search longer and converge toward local prices. The model yields additional predictions that I test using administrative and survey data. The evidence supports origin-based price extrapolation with subsequent learning rather than preference-based explanations such as reference dependence.

*UC Berkeley (msaccarola@berkeley.edu).

I am especially indebted to my advisors, mentors, and coauthors Luigi Butera, Stefano DellaVigna, Amir Kermani, Chen Lian, Enrico Moretti, Francesco Ruggieri, and Dmitry Taubinsky for their guidance and encouragement at every stage of this project. I also benefited from valuable conversations with Ned Augenblick, Matteo Benetton, Elin Colmsjö, Francesco Fabbri, Nicola Gennaioli, Dominic Kassirra, Henrik Kleven, Grady Killeen, Jianmeng Lyu, George Loewenstein, Ulrike Malmendier, Claudia Marangon, Peter Maxted, Tim McQuade, Ismir Mulalic, Bailey Palmer, Alexander Sebald, Afras Sial, Tim de Silva, Jon Steinsson, Dustin Swonder, Davide Taglialatela, Nadav Tadelis, Steve Tadelis, Pietro Tebaldi, and Sam Wang. I thank audiences at Copenhagen Business School, Behavioral Economics Annual Meeting (Student Workshop), the Haas Seminar in Real Estate, and UC Berkeley Psychology and Economics Seminar for feedback on this work. All remaining errors are my own.
I gratefully acknowledge financial support from the NBER Fellowship in Behavioral Macroeconomics, the BB90 Grant for Monetary Economics, and the Fisher Center for Real Estate and Urban Economics. I am also very thankful to Copenhagen Business School for hosting me during the period in which this project was developed.

Introduction

Individuals' past environments and experiences influence their economic choices in a wide range of settings. Early-life scarcity leaves a durable imprint on spending behavior (Mullainathan and Shafir, 2013); lifetime exposure to inflation alters saving and portfolio decisions (Malmendier and Nagel, 2016); and trading behavior of professional investors is too sensitive to recent changes in market conditions (Greenwood and Shleifer, 2014). One explanation for such history dependence is rooted in preferences: past experiences can shape current preferences through mechanisms such as habit formation (Charness and Gneezy, 2009) or the establishment of reference points (Kahneman and Tversky, 1979). An alternative explanation emphasizes the role of beliefs: experiences may affect expectations about the environment through channels such as experience-based updating (Kuchler and Zafar, 2019) or systematic distortions in memory and recall (Bordalo et al., 2022).

In this paper, I investigate how preferences and beliefs shape backward-looking behavior during the search for a home, the largest financial transaction in most households' lifetime (Gomes et al., 2021). I examine how previously experienced house price levels influence the prices homebuyers ultimately pay. To do so, I link Danish administrative registers covering the universe of housing transactions with detailed population-wide household data. I measure overpayment by comparing the prices paid by non-local movers with those paid by local buyers for comparable properties sold in the same markets and periods. I then relate these overpayments to house price levels in buyers' origin markets. The data reveal a striking asymmetry: buyers arriving from more expensive housing markets pay systematically higher prices than locals for similar homes, and the extent of overpayment increases with the price gap between origin and destination markets. By contrast, movers from cheaper origins pay only slightly more than locals, and their purchase prices show no systematic relationship with prices at origin.

The linkage between house transactions, household-level demographic information, and detailed tax records containing income, assets, and liabilities, is essential to identify the relationship between previous house prices and current overpayment. The richness of the linked data allows me to compare homebuyers arriving from different origins while holding constant their financial circumstances at the time of purchase, making it possible to separate the influence of house price levels at origin from other determinants of housing choices such as current wealth and home equity gains (Aiello et al., 2024).[1]

[1]Studies leveraging U.S. data often use multiple-listing service (MLS) datasets. While these data provide rich information on housing transactions, they are less well suited for addressing my research question. First, they include limited information on buyer demographics and financial circumstances, making it difficult to disentangle the effect of origin house price levels from other factors such as household wealth. Second, MLS data record an origin location only when the buyer was previously a homeowner, sold that property, and both transactions occurred within the same MLS system. As I discuss in Section 2, analyzing renters' choices is important for credible identification.

I interpret my findings through the lens of a housing search model in which buyers enter the destination market with beliefs that are not fully adjusted to local conditions. Specifically, the initial beliefs of homebuyers are modeled as a linear combination of prices in their origin and destination markets, a phenomenon I refer to as geographic extrapolation. Extrapolating buyers gradually update toward the true local price during the search process. This parsimonious search-and-learning structure is sufficient to produce the asymmetric overpayment patterns observed in the data. The asymmetry does not arise from heterogeneous learning rates or nonstandard preferences, but from the endogenous acceptance of buyers' offers by sellers. Buyers from more expensive origins initially bid above local prices and are quickly matched with sellers before they have time to learn. By contrast, buyers from cheaper origins tend to bid below local prices, face rejection from sellers unwilling to transact, and remain in search longer. These repeated rejections give them time to update their beliefs, so that they eventually converge toward correct prices before transacting or exiting the search. As discussed below, alternative explanations based on reference-dependent preferences are less parsimonious and less consistent with the empirical evidence.

I begin the empirical analysis with a movers design that compares the overpayment of households who purchase a house in the same ZIP code at the same time but arrived from different origins and thus observed different house prices. Across all specifications, moving from an expensive location to a cheaper one tends to induce overpayment proportional to the origin–destination price gap. In my preferred specification, the elasticity of overpayment with respect to the price difference when moving *down* in the house price distribution is 3.9% ($p < 0.01$).[2] In nominal terms, the effect is sizable: moving to a less expensive ZIP code by one standard deviation in the house price distribution leads to an overpayment of roughly USD 11,300 on the median Danish house. In contrast, buyers moving to more expensive locations display little systematic overpayment.[3]

There are, however, two reasons to be cautious about this design. First, overpayment is measured relative to locals who purchase an observably similar house. Therefore, estimates may be biased upward if movers from expensive origins systematically buy homes of higher unobserved quality as I would misattribute the discrepancy in value to overpayment. To address this concern, I turn to a repeated-sales design that compares transactions of the same house by different buyers

[2]To highlight the importance of conditioning on household-level characteristics, I begin by estimating a specification that controls only for a vector of origin ZIP code characteristics, replicating the type of controls typically available in U.S. data. Omitting buyer-level demographics and finances, the estimated effects are 30.1% larger than in my preferred specification. Once I add precise measures of income and net wealth from tax records, the coefficients stabilize, and the inclusion of additional wealth controls does not meaningfully alter the estimates.

[3]I directly assess the asymmetry in slopes by testing the null hypothesis that the slope of overpayment with respect to origin-destination price discrepancies is the same for movers from cheap to expensive location and from expensive to cheap location. The null hypothesis of equal slopes is strongly rejected in the data ($p < 0.01$).

at different points in time. This approach controls for any time-invariant property characteristics by design, while ZIP code-by-time fixed effects control for any changes in local housing values and amenities. The only remaining concern is therefore time-varying unit-specific unobserved housing quality. To gauge the importance of this channel, I add controls for all observable time-varying unit characteristics; the stability of the coefficients between movers and repeated sales design as well as between specifications suggests that unobserved housing quality is not driving the results.

A second set of concerns stems from the absence of random assignment of origin markets to homebuyers. Exogenous location shifters are notoriously scarce (see e.g. Finkelstein et al., 2021), and residual confounders could influence both household migration decisions and how much homebuyers are willing to pay for a given home. Still, the set of confounders that affect overpayment is much more limited than those that affect the overall house choice.[4] Moreover, although selection may induce a correlation between origin prices and overpayment, it is difficult to rationalize how selection alone could generate the hockey stick overpayment patterns observed in the data. Nevertheless, to strengthen the credibility of my estimates, I adapt an instrumental variable approach commonly used in the real estate literature (Guren, 2018; Aiello et al., 2024). As originally noted by Stein (1995), down-payment constraints in the mortgage market imply that sellers with low home equity face a high marginal utility of cash at hand for financing their next purchase. Constrained sellers therefore set higher listing prices and wait longer in order to realize larger sale proceeds that can be reinvested as the down payment on their subsequent home. In practice, I exploit plausibly exogenous variation in seller composition by using the median ratio of home equity to house value among sellers in the origin market as an instrument for origin house price levels. Once again, my findings are robust to the change in the design.

To further examine the mechanisms and evaluate geographic extrapolation in a different decision context, I use an inheritance design comparing only children who inherit similar homes in the same location but currently reside in different ZIP codes. Children living in more expensive areas tend to list comparable inherited homes at higher prices, are more likely to revise their initial listings downward, and spend longer on the market, yet they do not achieve higher sale premia. These patterns are consistent with initial overpricing driven by extrapolative beliefs that is subsequently corrected during the selling process.

Motivated by this evidence, I develop a model of housing search to obtain clear and testable predictions and to clarify how belief-driven extrapolative behavior affects buyer decisions. The model begins with a rational baseline in which a non-local buyer searches for a home in a des-

[4]Many confounders–for example, heterogeneous tastes for housing and homeownership–may influence both the type of unit purchased and the overall level of housing expenditure. They need not, however, affect the willingness to pay above the market price for a given unit.

tination market characterized by frictional matching. Buyers face a trade-off between offering a higher price and bearing the cost of continued search, while match-specific taste shocks capture heterogeneous valuations of housing units. Sellers differ in their urgency to sell, represented by heterogeneous waiting costs, and accept offers that exceed their outside option, defined by the prevailing local market price net of waiting costs.

I then consider two extensions of this baseline model. In one, I allow buyers to extrapolate from house price levels in their origin markets and gradually update toward local price levels according to a tractable learning process in the spirit of Calvo (1983). In an alternative specification, I enrich the model by introducing reference dependence and loss aversion into buyers' preferences, in the spirit of Barberis and Xiong (2012). A buyer who is reference dependent derives utility not only from the consumption value of the house net of price, but also from comparing the purchase price to a reference point. When the transaction price lies below the reference point, the buyer experiences a utility gain, while paying above the reference point generates a utility loss. A loss-averse buyer places greater weight on losses than on equivalent gains, so that the disutility from overpaying relative to the reference point exceeds the utility of gain from underpaying.

Returning to the movers design, I show that the evidence supports this belief-based interpretation. First, a reference-point mechanism would predict an S-shaped rather than a hockey-stick relationship between overpayment and origin prices. Second, overpayment is more affected by previous prices among renters than among homeowners, indicating that mental accounting or reference points tied to the proceeds of a prior home sale cannot account for the findings. Third, the heterogeneity analysis further reinforces this interpretation, as the hockey-stick pattern is attenuated among more educated, wealthier, and older buyers—groups that are likely to be more experienced and financially sophisticated. In a further heterogeneity exercise, I show that the strength of the relationship between overpayment and the change in prices depends on the degree of price dispersion: the relationship weakens when origin markets exhibit greater variation in house prices, consistent with faster belief updating, and strengthens when destination markets display higher price dispersion, consistent with slower learning.

To complement these observational designs, I conducted an online survey of prospective U.S. homebuyers. The survey elicited respondents' beliefs about median house prices in both their current location (origin) and up to three ZIP codes they were actively considering for relocation (destinations). By comparing these belief elicitations to true prices measured using Redfin transaction data, I show that forecast errors about destination house prices are systematically predicted by the discrepancy between origin and destination prices—a pattern consistent with geographic extrapolative beliefs. To address concerns about idiosyncratic and respondent-specific measurement error in reported beliefs, I leverage both the structure of the survey design and an instrumental-variables strategy that exploits only the exogenous variation in respondents' beliefs induced by variation in true prices at origin.

This paper relates to pioneering studies showing that the rental choices of U.S. movers are predicted by rental prices in their origin cities (Simonsohn and Loewenstein, 2006; Bordalo et al., 2019). These contributions provide valuable evidence of history dependence in housing decisions, but they differ in important ways. They analyze rental rather than purchase decisions, study total rental expenditure rather than overpayment relative to comparable properties, and relate outcomes to origin rent levels rather than to the gap between origin and destination prices. Both papers rely on self-reported survey data from the Panel Survey of Income Dynamics, which lacks measures of wealth and is limited in its ability to capture heterogeneity in movers. Further, due to data limitations, the authors cannot fully partial out the effects of current market conditions on house choice, which confounds identification.[5] By contrast, my analysis leverages administrative data with full population coverage, which allows me to control flexibly for neighborhood-by-time effects and to uncover richer empirical features such as the asymmetric overpayment patterns. Finally, whereas the earlier studies interpret correlations between origin prices and outcomes without modeling the market, I embed movers in a search-and-bargaining framework, which disciplines the mechanism behind the observed patterns and links them to their equilibrium implications.

This work also contributes to recent developments in the behavioral real estate literature (Genesove and Mayer, 2001; Andersen et al., 2020b; Bracke and Tenreyro, 2021; Andersen et al., 2022). The key differences are twofold. First, while most of this literature focuses on the behavior of sellers, I study the decisions of homebuyers. Second, rather than emphasizing mechanisms rooted in reference dependence and loss aversion, I highlight a belief-based explanation grounded in buyers' expectations about local prices. The fact that the evidence in my context favors extrapolative beliefs is entirely consistent with reference dependence playing a key role in other settings. Finally, the paper also connects to Badarinza et al. (2025), who show how behavioral forces can shape the dynamics and equilibrium outcomes of housing markets.

This paper relates to the growing literature on the role of expectations in housing markets (for a review, see Kuchler et al., 2023). My work is most closely related to recent research highlighting the geographic component of extrapolative beliefs, especially Bailey et al. (2018). They show that homebuyers update expectations based on housing returns experienced by geographically distant friends. I complement this evidence by studying the opposite phenomenon: movers extrapolate

[5]Because of limited sample sizes, these studies cannot saturate the data with fine-grained destination ZIP code-by-time fixed effects that would absorb local market conditions. Instead, they rely on yearly fixed effects common across all locations and control for local conditions only through the inclusion of current rental prices in the destination market. Yet, as emphasized in the real estate literature, prices alone are not a sufficient statistic for contemporaneous market conditions (Genesove, 2024) and housing markets are highly seasonal.

the house price levels they observe in their origin markets onto their destinations.

The rest of the paper is structured as follows. Section 1 describes the data. Section 2 details the main empirical designs and presents the main empirical findings. Section 3 provides a framework to discuss the alternative mechanisms and their predictions. Section 4 brings the model predictions to the data. Section 5 concludes.

1 Data and Measurement

This section introduces the data sources, the restrictions I use to construct the main sample, and my approach to measuring overpayment and exposure to real estate prices. Appendix Section B contains additional details on data sources, construction, and cleaning.

1.1 Registry Data

My analysis draws on comprehensive Danish administrative registers that link house transactions to the financial, geographic, and demographic characteristics of both buyers and sellers. In the following paragraphs, I describe each data source in detail.

House Transactions and Characteristics of the Housing Stock I use records from the Danish Tax and Customs Administration (SKAT) to track the universe of house transactions between 1996 and 2017. Information on property ownership, physical and hedonic characteristics, and biennially assessed property values is obtained from the official housing register (Bygnings- og Boligregisteret) and the property ownership registry. Appendix Section B.2 documents the full set of property characteristics included in the analysis.

Beginning in 2008, I complement these registers with housing listings data provided by RealView. These data allow me to observe listing dates, asking prices, subsequent revisions, and retractions for all properties offered for sale.

Household Financial and Demographic Information Annual income and balance sheet information on all Danish residents come from administrative tax records maintained by SKAT. These data are regarded as exceptionally reliable, as income and household financial accounts are subject to third-party reporting, and tax evasion is minimal in Denmark (Kleven et al., 2011). Additional information on mortgage borrowing is incorporated from data compiled by Statistics Denmark based on reports submitted by mortgage banks to Danmarks Nationalbank.

Demographic information is obtained from the Civil Registration System (CPR), which records household composition and demographic characteristics on a quarterly basis. Educational attainment is obtained from the Ministry of Education (Undervisningsministeriet), which records each individual's highest completed level of education and corresponding professional qualifications.

All data are recorded at the individual level using unique, anonymized CPR identifiers.[6] Following Statistics Denmark's definition of households, I aggregate individuals into households, which constitute my unit of analysis. Appendix Section B.1 provides further details on the household definition and a full list of household-level characteristics used in the analysis.

Geography and Moves Drawing on the quarterly residential address history from the CPR, I identify individuals who relocate across ZIP codes and purchase a housing property in the destination ZIP code within one year of their official move date. I define these individuals as movers, while those who purchase a property within their original ZIP code are defined as locals.

Death Data and Inheritances To identify property owners who have passed away, I use the Danish Cause-of-Death Register maintained by the National Board of Health (Sundhedsstyrelsen). This register is based on official death certificates issued by physicians immediately after a death. I merge these records with the property ownership registry to establish whether the deceased owned real estate. Using family identifiers from the CPR registries, I then link deceased owners to their descendants, which enables me to study subsequent selling decisions of the beneficiaries.

1.2 Measurement

My primary independent variable captures the discrepancy between house prices previously observed by homebuyers in their origin housing market and the prevailing house prices in their destination ZIP code. To measure house prices for a given ZIP code and calendar month, I use the natural logarithm of the median price per square foot from realized sales in that ZIP code over the preceding 12 months. This approach smooths transitory fluctuations while reflecting the information environment most relevant to buyers at the time of their move. To compute the percent difference between an origin and destination market, I simply take the difference in the house price measures between the two locations at the time of the house transaction.[7]

My main dependent variable is a buyer's *quality-adjusted* overpayment for a given house. Following standard hedonic approaches in this literature (e.g. Genesove and Mayer, 2001; Bajari et al., 2013; Guren, 2018; Andersen et al., 2022), I estimate

$$p_{h,t} = f(H_{h,t}) + \xi_{t,\text{ZIP}(h)} + e_{h,t}, \tag{1}$$

[6] CPR identifiers are the Danish equivalent of U.S. Social Security numbers.

[7] For the remainder of the paper $\bar{p}_{z,t}$ denotes such measure of house price level measured in ZIP code z at time t.

where $p_{h,t}$ denotes the natural logarithm of the sale price of home h at time t. The function $f(H_{h,t})$ flexibly controls for observable housing characteristics (e.g., square footage, lot size, age, number of bedrooms, and tax assessed value) using a polynomial in attributes; the complete list of hedonic attributes and the exact functional form of $f(\cdot)$ are provided in Appendix B.2. The fixed effects $\xi_{t,\text{ZIP}(h)}$ absorb ZIP-by-quarter variation in house prices. Let $\hat{p}_{h,t}$ denote the predicted log sale price for house h at time t. The regression residual $\hat{e}_{h,t} \equiv p_{h,t} - \hat{p}_{h,t}$ captures the buyer's quality-adjusted overpayment, measured on a log-points scale.

The hedonic model performs well, as reflected in its high explanatory power ($R^2 = 0.89$). By construction, hedonic predictions capture only the *observed* quality of a housing unit, which may differ from its true quality. For example, a property may have undergone recent high-end renovations that are not recorded in the available data. In such cases, unobserved quality, defined as the difference between true and observed quality, is mechanically embedded in my overpayment measure.

Formally, in the presence of unobserved housing quality, the measured overpayment can be decomposed as

$$\hat{e}_{h,t} = e_{h,t} + u_{h,t},$$

where $e_{h,t}$ represents the true overpayment relative to quality-adjusted value, and $u_{h,t}$ denotes the contribution of unobserved quality characteristics not captured by $f(H_{h,t})$. If $u_{h,t}$ is unrelated to my independent variable of interest, then this mismeasurement simply adds noise to the estimates without introducing bias. However, if unobserved quality is systematically related to the origin location of the buyer, this may bias my estimates. I return to this issue when I discuss robustness of my results in Section 2.

1.3 Sample Construction

I begin my sample construction from the universe of house transactions in Denmark between 1996 and 2017. To ensure that the dataset reflects valid market transactions of owner-occupied housing, I apply a series of restrictions. First, I exclude properties with missing or implausible characteristics (e.g., total square footage below 100 square feet). Second, I drop transactions flagged as anomalous by Statistics Denmark, as well as those with extreme prices (below 15,000 DKK or above 20 million DKK). Third, I omit intra-family transfers and transactions involving government agencies, firms, or nonprofit organizations. Fourth, I exclude summer houses, cooperative housing, and other properties designated as vacation units. Finally, I restrict the sample to transactions in which the buying household can be identified in the registry, contains no missing information, and in which the oldest household member is at least 25 years old.

After applying these filters, I obtain 571,603 cleaned transactions. Of these, 257,222 involve buyers I classify as movers, which form the analysis sample for the movers design. In addition, 190,123 transactions correspond to properties that are transacted at least twice in the data, irrespective of the moving status of the buyer; these constitute the repeated-sales sample. Appendix Table A.1 reports summary statistics for both the movers sample and the repeated-sales sample.

2 Overpayment and House Prices at Origin

In this section, I present the main empirical finding: a "hockey-stick" relationship between buyers' quality-adjusted overpayment and the gap between house prices in their origin market and those in the destination. The analysis begins with a movers design that compares the overpayment of similar households who buy in the same ZIP code at the same time but arrive from different origin markets, and thus faced different pre-move price environments. I then discuss the main limitations of this design, present a series of robustness checks, and introduce two alternative research strategies that directly address the two central concerns: unobserved housing quality and selection correlated with origin-market house prices.

2.1 Movers Design

Movers Design: Specification To investigate the causal effect of origin house prices on buyers' overpayment, the ideal experiment would compare two otherwise identical households from similar origin locations who, prior to being randomly assigned to the same destination, were exposed to different house prices in their respective ZIP codes. While such an experiment is infeasible, I approximate it with a movers design. This design compares the overpayment of households who are similar across a rich set of demographic and financial characteristics. These households move to the same destination ZIP code at the same time but come from origins that differ in prevailing house prices at the time of the move.

In practice, I implement this intuition by fitting the following regression equation:

$$\hat{e}_{i,t} = \beta\left[\bar{p}_{o(i),t} - \bar{p}_{d(i),t}\right] + \delta_{d(i),t} + \gamma_O O_{i,t} + \gamma_X X_{i,t} + \varepsilon_{i,t}, \tag{2}$$

where $\hat{e}_{i,t}$ is the measured overpayment for the house purchased by household i in calendar quarter t, constructed as described in Section 1.2. The main regressor of interest, $\bar{p}_{o(i),t} - \bar{p}_{d(i),t}$ captures the discrepancy in house prices between the buyer's origin ZIP code, $o(i)$, and their destination ZIP code, $d(i)$. A positive (negative) value indicates that the buyer moved from a more expensive (cheaper) origin to a cheaper (more expensive) destination.

The regression also includes fixed effects for destination ZIP code interacted with calendar quarter of the purchase, $\delta_{d(i),t}$. $O_{i,t}$ is a rich set of controls for characteristics of the origin ZIP code, including median income, median wealth, demographic composition of the ZIP code population, and characteristics of the local housing stock. The vector $X_{i,t}$ contains detailed household-level covariates, such as household composition, income, and multiple measures of financial and real estate wealth.

Standard errors are clustered two ways, by destination ZIP code and by calendar quarter, to account for correlated shocks within local housing markets and across locations in a given period (Cameron et al., 2011; Abadie et al., 2022; MacKinnon et al., 2023).

Motivated by theory and by the asymmetry in the relationship between $\bar{p}_{o(i),t} - \bar{p}_{d(i),t}$ and overpayment, I allow the effect of origin-destination house price discrepancies to differ by the direction of the move. I distinguish between households moving from relatively cheaper to more expensive markets (where $\bar{p}_{o(i),t} - \bar{p}_{d(i),t} \leq 0$) and those moving from more expensive to cheaper markets (where $\bar{p}_{o(i),t} - \bar{p}_{d(i),t} > 0$). Accordingly, Equation 2 becomes

$$\begin{aligned} \hat{e}_{i,t} = & \beta_{\leq 0} \left(\bar{p}_{o(i),t} - \bar{p}_{d(i),t}\right) \cdot \mathbf{1}\{\bar{p}_{o(i),t} - \bar{p}_{d(i),t} \leq 0\} \\ & \beta_{>0} \left(\bar{p}_{o(i),t} - \bar{p}_{d(i),t}\right) \cdot \mathbf{1}\{\bar{p}_{o(i),t} - \bar{p}_{d(i),t} > 0\} \\ & + \delta_{\mathbf{1}\{\bar{p}_{o(i),t} - \bar{p}_{d(i),t} \leq 0\}} + \delta_{d(i),t} + \gamma_O O_{i,t} + \gamma_X X_{i,t} + \varepsilon_{i,t}. \end{aligned} \tag{3}$$

Movers Design: Results Table 1 and Figure 1 report results from the movers design.

Figure 1 provides a graphical illustration of the hockey-stick relationship between overpayment and house prices at origin. The figure shows a binned scatterplot of measured overpayment against the house price discrepancy between origin and destination ZIP codes for the movers sample.[8] The plot overlays estimates from multiple specifications that sequentially add controls, for visual assessment of the robustness of the relationship.

To the left of zero on the horizontal axis, Figure 1 shows overpayment for movers arriving from cheaper origins into more expensive destinations. Their overpayment is small in levels and changes little with the prices observed in their origin locations. By contrast, the right side of the plot forms the handle of the hockey stick: for movers arriving from more expensive origins and trading down into cheaper destinations, overpayment rises sharply with the house prices they faced prior to moving.

Table 1 reports ordinary least squares estimates of the regression specified in Equation 3. The first row reports coefficients for movers relocating from cheaper to more expensive markets (i.e., "up" in the house price distribution), while the second row presents coefficients for movers relocating from more expensive to cheaper markets (i.e., "down" in the house price distribution).

[8]The dependent and independent variables are residualized by partialling out fixed effects for calendar quarter of transaction interacted with the destination ZIP code, ensuring that market conditions at the time of the transaction are appropriately demeaned for ease of interpretation. Furthermore, for all binned scatterplot figures, I residualize the variables of interest with respect to any controls following the procedures introduced in Cattaneo et al. (2024).

The contrast between these two rows mirrors the asymmetry in slopes documented in Figure 1. Consistent with the graphical evidence, the estimates for down-movers are positive and sizable, whereas those for up-movers are small and, with the exception of the first column, not statistically significant.

Across columns, the first three specifications gradually incorporate additional controls. Column (1) reports a baseline specification that already includes ZIP-code-of-origin characteristics but excludes household-level controls, similar to what would typically be used with U.S. listings data. Adding household-level controls in Column (2) substantially attenuates the coefficients, indicating that part of the raw relationship reflects compositional differences across households rather than pure origin price effects. By contrast, moving from Column (2) to Column (3), which introduces origin ZIP code fixed effects, leaves point estimates unaffected, suggesting that unobserved heterogeneity at the origin ZIP level contributes little beyond the observable controls.

Columns (4) and (5) report estimates from the preferred specification, restricting the sample to buyers who were renters and prior homeowners, respectively. The persistence of the pattern among renters is particularly informative. First, it rules out explanations based on reference-point or mental accounting mechanisms (Hastings and Shapiro, 2013), whereby homebuyers simply reinvest the proceeds from their previous home in the new location. Second, the renter subsample offers a cleaner measure of household wealth, since it is not affected by variation in housing equity—the main source of wealth that closely covaries with home prices at origin.

Because both the dependent and independent variables are measured in log units, the estimated coefficients can be interpreted as elasticities of overpayment with respect to the price discrepancy between origin and destination. For example, the coefficient of 0.037 in the second row of Column (3) implies that a mover arriving from an origin where housing is twice as expensive as the destination pays, on average, 3.7% more than a comparable local household for the same housing unit.

To contextualize the magnitude of the estimates, I translate the elasticity into an implied nominal overpayment.[9] A buyer moving down by one standard deviation in the Danish house price distribution would overpay for the median home by roughly 11,300 USD relative to a comparable local buyer. For reference, consider a homebuyer relocating from San Francisco to Austin in June 2020, where average house prices per square foot were about three times higher in the origin than in the destination. Applying my estimates implies an overpayment of approximately 15,300 USD for the median Austin home relative to otherwise similar local buyers.[10]

[9]For comparability, all amounts are expressed in U.S. dollars using the prevailing exchange rate in the year of each transaction.

[10]The San Francisco–to–Austin migration is a salient example, as many tech workers relocated during the pandemic in response to widespread work-from-home policies. U.S. home price data are obtained from `redfin.com`.

Movers Design: Robustness The richness of the Danish administrative data enables me to rule out several potential confounding explanations.

First, households from different origins might systematically select into different destinations or time their moves differently, for example due to seasonal migration patterns. However, because the specification includes destination-by-time fixed effects, all comparisons are made *within* the same housing market and quarter, eliminating such confounding by construction.

Second, wealthier households may simply be more willing to pay more for a given housing unit to shorten search time, as suggested by Aiello et al. (2024). This concern is mitigated by detailed controls for financial and real estate wealth, as well as for mortgage and non-mortgage liabilities. The vector of wealth controls includes prior homeownership status and, where applicable, for the proceeds from concurrent sales of previous properties.[11] The stability of the estimates after including these controls indicates that residual unobserved wealth is unlikely to explain the results. Consistent with this, the "hockey-stick" overpayment pattern is even more pronounced for the renter subsample.

Appendix C also shows that, conditional on the rich vector of household-level controls in my preferred specification, movers facing different origin–destination house price differentials follow similar income trajectories in the years around the move. Appendix Figure A.1 Panel (a) presents the income trajectories without partialling out the controls. Panel (b) demonstrates that these differences disappear once controls are included.

I also note that it is difficult to generate the pronounced asymmetry observed in the data through selection alone. If unobserved buyer characteristics correlated with origin prices were driving the results, we would expect similar effects for movers arriving from slightly more expensive and slightly cheaper origins. The asymmetry in overpayment therefore provides an implicit robustness check against such concerns.

Nevertheless, as discussed in Section 1, measuring overpayment using a hedonic residual may introduce bias if, even after conditioning on all observables, households from more expensive origins systematically purchase homes of unobservably higher quality. Since observed home quality is positively correlated with origin prices in the raw data, addressing this concern is essential for validating the main result. I therefore turn next to a repeated-sales design.

2.2 Repeated Sales Design

Repeated Sales: Specification and Estimation The repeated-sales design compares multiple transactions of the same housing unit conducted by different buyers at different points in time. By holding the unit fixed, all unit characteristics that remain constant over time—including the

[11] I also note that components of household net wealth are subject to third-party reporting by financial institutions for tax purposes, and tax evasion is exceedingly rare in my setting (Kleven et al., 2011).

time-invariant component of unobserved quality—are differenced out by design and therefore cannot confound the results.

To estimate housing-unit fixed effects while retaining sufficient statistical power, I implement this specification using the repeated-sales sample, which includes transactions conducted by both movers and non-movers. Because the sample is not limited to movers, I include an indicator for mover status to allow movers and non-movers to have different transaction outcomes, for example due to systematic differences in search costs (Lambson et al., 2004).[12]

To bring this intuition to the data, I adapt the specification in Equation 2 as follows:

$$p_{h(i),t} = \beta\left[\bar{p}_{o(i),t} - \bar{p}_{d(i),t}\right] + \alpha_{h(i)} + \gamma H_{h(i),t} + \text{Controls} + \varepsilon_{i,t}. \tag{4}$$

Here, $p_{h(i),t}$ denotes the log sale price of housing unit $h(i)$ purchased by household i at time t. The term $\alpha_{h(i)}$ is a housing-unit fixed effect that absorbs all time-invariant characteristics of the property. The vector $H_{h(i),t}$ captures time-varying housing attributes.[13] As additional controls, I include the same origin-ZIP code characteristics and demographic and financial household variables as in the movers design, augmented with an indicator for mover status. Finally, to allow for potential asymmetries between moves up and down the price distribution, I also estimate a version of this specification that mirrors the asymmetric formulation in Equation 3.

Compared to the movers design, this specification is estimated in a single step rather than relying on the hedonic residual as a separate measure of overpayment. Importantly, the main coefficient of interest retains the same interpretation: it captures how the buyer's quality-adjusted overpayment—measured as the price paid after partialling out all price-relevant characteristics of the unit—responds to the price discrepancy between the origin and destination markets.

Repeated Sales: Results and Discussion Table 2 reports the results from the repeated-sales design. Column (1) includes only housing-unit fixed effects, while Column (2) additionally incorporates time-varying property characteristics. As discussed above, the main coefficient of interest retains the same interpretation as in the movers design and can therefore be directly compared to the estimates in Table 1. Column (3) further adds fixed effects for the ZIP code of origin among movers. As shown by comparing the first two sets of coefficients in Figure 2, the results are very similar to those from the baseline design: once again, I find the same asymmetric relationship between overpayment and the price discrepancy experienced by buyers between their origin and

[12] Interacting the price discrepancy with an indicator for movers is neither necessary nor feasible. By construction, origin and destination coincide for non-movers, so $\bar{p}_{o(i),t} - \bar{p}_{d(i),t} = 0$ for all such households.

[13] The time-varying housing characteristics I include are: (i) the log of the most recent tax-assessed market value of the property, interacted with calendar-year-of-assessment fixed effects and updated biennially by the Danish tax authority; (ii) fixed effects for building age, grouped in five-year intervals; and (iii) an indicator for whether the property was unoccupied in the year prior to sale.

destination locations. For ease of comparison across designs, Figure 2 also plots the preferred coefficients side by side.

Notably, when comparing the second rows of Tables 1 and 2, the estimated elasticities of overpayment with respect to the origin–destination price discrepancy are slightly larger in the repeated-sales specification. This suggests that, if anything, buyers from more expensive origins tend to purchase homes of lower unobserved quality after conditioning on buyer characteristics. Such a pattern is consistent with interpreting these buyers as less informed and making suboptimal purchases, whereas bias arising from a positive correlation between unobserved quality and origin prices would instead attenuate the estimates when comparing the baseline movers design with the repeated-sales specification.

The repeated-sales approach controls for all time-invariant property characteristics by design, while ZIP code-by-time fixed effects absorb changes in local housing values and amenities that affect all properties within a ZIP code. This leaves only one potential source of concern: unobserved quality that is both unit-specific and time-varying. To assess the importance of this confound, I compare the estimates in Column (1) and Column (2). The addition of time-varying housing characteristics substantially increases explanatory power—as reflected in the near-doubling of the within-R^2—while the main coefficients remain stable in both magnitude and significance. Following the logic of Oster (2019) and Finkelstein et al. (2021), I conclude that any remaining bias is likely small.

As an additional robustness check, I exploit a tax deduction for home renovations that was in place between 2011 and 2022.[14] I estimate regressions where the dependent variable is an indicator for whether the housing unit was renovated either shortly before or shortly after the transaction. Specifically, I construct two measures of renovation activity: renovations undertaken by the seller in the years immediately preceding the sale and renovations undertaken by the buyer shortly thereafter. The results, reported in Appendix Table A.3, show no systematic tendency for movers from more expensive origins either to purchase recently renovated units, as shown in Panel a, or to renovate their homes shortly after purchase, as shown in Panel b.

2.3 Instrumental Variable Approach

To address any remaining concerns about selection, I adapt the seller's home equity instrumental variables strategy originally developed in Guren (2018). In practice, the instrument exploits changes in the composition of sellers in the origin ZIP codes of households migrating to the same destination at the same time. Put differently, I compare households that are observably similar and originate from observably similar ZIP codes, but who observe different origin house prices

[14]Details on the deduction scheme and the construction of renovation variables are provided in Appendix Section C.2.

due to variation in the financial positions of sellers active in their *origin* market at the time of their move.

As emphasized in prior work (Genesove and Han, 2012; Andersen et al., 2022), search frictions in housing markets generate a fundamental trade-off: sellers can set higher list prices and remain on the market longer in pursuit of higher sale prices, or set lower prices to secure faster transactions. The optimal sale price thus depends on the marginal value of liquidity at the time of sale. Down payment constraints are central to this trade-off.[15] Sellers with low home equity are more likely to face binding constraints and therefore have stronger incentives to hold out for higher prices, which can be used to finance the down payment on their next purchase. By contrast, households with ample equity are less constrained and can sell more quickly, since financing the next purchase does not hinge on extracting maximum value from the current sale.[16]

Instrumental Variable: Specification and Estimation I begin by constructing a ZIP code–quarter measure of predicted home equity at sale. I include all sellers who completed a transaction within the previous twelve months. For each seller, I compute home equity at sale as the ratio of the property's tax-assessed market value to the outstanding mortgage balance at the time of sale. Because this measure is defined at the seller level, I then aggregate it to the ZIP code level by taking the median value across all sellers in that ZIP code during the relevant time horizon. For brevity, I refer to this quantity as the home equity among sellers in a given ZIP code throughout the remainder of the paper.

Letting $\bar{Z}_{o(i),t}$ denote my measure of home equity among sellers in household i's origin ZIP code $o(i)$ at time t, the first-stage equation is given by

$$\left[\bar{p}_{o(i),t} - \bar{p}_{d(i),t}\right] = \tilde{\delta}_{d(i),t} + h(\bar{Z}_{o(i),t}) + \tilde{\gamma}_X X_{i,t} + \tilde{\gamma}_O O_{i,t} + u_{i,t}, \tag{5}$$

where $h(\cdot)$ is some function of the instrument. The second-stage equation is entirely analogous to Equation 2.[17] The controls included in the household characteristics, $X_{i,t}$, and ZIP code of origin characteristics, $O_{i,t}$, are defined analogously to those in the movers design.[18]

[15]For a model of the seller's house sale with down payment constraints, see Stein (1995). Among Danish households these constraints are salient, as the maximum loan-to-value ratio has remained fixed at 80 percent throughout the period of analysis. For a comprehensive description of the Danish mortgage market, and a comparison with the U.S., see Andersen et al. (2020b).

[16]Two edge cases are worth noting. First, if equity is so low that it triggers default, households may sell in distressed "fire sales" (Andersen and Nielsen, 2017; Guren and McQuade, 2020). Second, in particularly hot markets sellers may deliberately set low list prices to induce bidding wars (Han and Strange, 2014). These mechanisms could weaken the first stage of my instrument but need not introduce bias in the estimates.

[17]When applying the instrumental variable strategy to the asymmetric specification (Equation 3), two separate first-stage equations are defined, corresponding to the instrumented variables: prices interacted with the positive price-change dummy and prices interacted with the negative discrepancy dummy. I omit them here for brevity.

[18]Note that the inclusion of destination-time fixed effects, $\tilde{\delta}_{d(i),t}$, in the first stage fully absorbs variation in $\bar{p}_{d(i),t}$.

As with all instrumental variable approaches, identification requires that the instrument satisfies both relevance and exclusion. Figure 3 illustrates the relevance of my instrument. Consistent with theoretical predictions, the relationship between the instrument—home equity among sellers in the origin ZIP code—and house prices in the ZIP code is negative. Similarly, the first-stage F-statistic of 45.07 is above conventional thresholds in the literature Stock and Yogo (2005).

The exclusion restriction merits further discussion, as two additional considerations arise. The first relates to potential spillover effects that may influence the wealth of movers. If movers sell their own home in the origin ZIP code prior to purchasing a new one, then the listing and sale prices of other homes in that ZIP code could mechanically affect the sale price of the mover's own property. This, in turn, may influence the household's available wealth at the time of purchase and thereby affect overpayment through channels unrelated to the mechanism of interest.[19] To account for this, the baseline specification includes controls for whether the mover sells a home and, if so, the corresponding transaction price. Furthermore, the analysis is replicated on a subsample of buyers who were renters prior to the move, for whom such spillover effects are mechanically ruled out.

A second consideration is that shifts in the composition of house sellers within a ZIP code could coincide with local shocks that also influence movers' purchasing decisions. For instance, a negative labor market shock may lead homeowners to sell despite low home equity and relocate while simultaneously affecting other movers' income trajectories. These factors are largely absorbed in the baseline specification through the inclusion of time-varying ZIP code characteristics. In addition, I refine the specification by making comparisons within municipalities, so that common local shocks are fully absorbed and the instrument captures plausibly idiosyncratic variation in seller composition across ZIP codes within the same municipality. In practice, this is implemented by including municipality-of-origin fixed effects in the regression specification.

Instrumental Variable: Results and Discussion Table 3 reports results from two-stage least squares (2SLS) estimations of the instrumental variables design, where I split the sample by the sign of the price discrepancy. Odd-numbered columns report estimates for movers relocating to more expensive destinations, while even-numbered columns report estimates for movers relocating to cheaper markets. Columns (3) and (4) restrict the sample to buyers who were not homeowners prior to the move. As expected, the coefficients remain stable in magnitude and significance, confirming that the results are not driven by potential spillover effects from concurrent housing sales in the origin ZIP code. Columns (5) and (6) add municipality-of-origin fixed

This implies that the specification is essentially equivalent to instrumenting origin prices directly, rather than the price differential.

[19]In related work, Aiello et al. (2024) show that households with higher home equity are more willing to overpay in order to shorten search duration.

effects to account for common local shocks that could simultaneously influence seller composition and movers' purchasing decisions. The estimates are again robust, indicating that my vector of time-varying ZIP-code characteristics of origin is largely sufficient to absorb such local shocks.

As expected, the use of 2SLS reduces statistical power, but the findings broadly replicate those from the baseline analysis reported in Table 1. Figure 2 compares estimates across the three main designs. The estimated elasticities remain asymmetric and point estimates are similar across specifications, reinforcing the interpretation that buyers from more expensive origins tend to overpay relative to comparable locals.

2.4 Inheritance Design

To shed additional light on the underlying mechanism explaining the observed hockey-stick pattern of overpayment, I exploit a unique feature of the Danish data: the ability to construct family links. This feature allows me to examine how households anchor their price norms to local housing markets by analyzing the behavior of children who inherit a house after their parent passes away.[20]

My empirical design compares the listing decisions of observably similar children living in the same location and inheriting comparable properties within the same parental ZIP codes. Identification is obtained from time-series variation in the discrepancy between house prices in the parental ZIP code and those in the beneficiary's own ZIP code.

This setting offers several key advantages that make it highly complementary to the main movers analysis. First, the parental house is fixed at the time of inheritance: beneficiaries only decide whether to list the property and at what price, while all other margins of the housing choice—such as location, size, and quality—are predetermined. Second, the timing of parental death is plausibly exogenous, conditional on parental age and income. Third, I observe a richer set of seller-side outcomes—such as time on the market and revisions to the initial listing price—that provide additional insights into behavioral mechanisms.

Details of data construction are provided in Appendix Section B.3, and summary statistics for this sample are reported in Appendix Table A.2.

Inheritance Design: Specification In practice, I estimate a regression specification similar to the one introduced in Equation 2, with two key differences. The first difference lies in the measurement of geographic house-price discrepancy. Specifically, I define the price discrepancy as the difference between the log median house price per square foot in the ZIP code of the beneficiary's current residence and the corresponding measure in the ZIP code of the inherited

[20]Following Andersen et al. (2021), I restrict the analysis to single children to avoid complications arising from intra-sibling bargaining over listing decisions.

property. The interpretation is analogous to the definition used in the movers design: a positive value indicates that beneficiaries are exposed to higher prevailing prices in their own residential market relative to the inherited property's market, and a negative value indicates the opposite.

The second difference concerns the set of outcomes used in the analysis. Linking the inheritance sample with listing data allows me to examine a richer set of outcomes. The first is the *log listing premium*, defined as the difference between the initial listing price and the predicted house value based on the hedonic model described in Section B. The second outcome is the probability that the initial listing price is revised downward. Finally, for the subset of inherited properties that eventually sell, I examine *time on the market* and the realized *sale premium*, measured as the sale price minus the predicted hedonic value of the property.

Formally, I regress a listing outcome $y_{i,t}$ for beneficiary i at time t on the price discrepancy between the beneficiary's ZIP code and the inherited property's ZIP code, controlling for the same vector of covariates as in the baseline specification (Equation 2).

Inheritance Design: Results Table 4 reports the main results. Although the sample is smaller than in the main movers analysis, the estimates display a clear and internally consistent pattern: beneficiaries living in more expensive areas initially list inherited properties at systematically higher premia relative to predicted quality (Column 1). These sellers are also more likely to revise listing prices downward (Column 2). When sales occur, properties tend to remain on the market longer before sale (Column 3), yet the realized sale prices do not yield a significant premium relative to predicted value (Column 4).[21]

Taken together, the results point to a mechanism of initial extrapolation from local price norms, followed by gradual correction through learning while on the market. This interpretation contrasts with a purely preference-driven explanation, which would imply that sellers remain firm on their initial listing prices rather than systematically revising them downward.

3 A Conceptual Framework for Housing Search

To make progress in identifying the source of overpayment, I develop a partial-equilibrium model of housing search in which a homebuyer arriving from some origin location o searches over properties in some destination location d. I assume that the decision to move to a particular location and the type of house being searched for have already been determined in a prior optimization stage.[22] Finally, I assume that homebuyers are homogeneous in all respects except for their origin

[21]Due to the more limited number of inheritance cases, I do not explore heterogeneity by the sign of the price difference for this subsection.

[22]This assumption is consistent with recent empirical evidence showing that most households search within relatively narrow segments of the housing market (Piazzesi et al., 2020).

location and the induced effects this heterogeneity has on subsequent choices.[23]

3.1 Rational Benchmark

I begin by describing a rational benchmark to serve as a reference for the behavioral extensions introduced in the following subsections. Specifically, I consider a discrete-time infinite-horizon search environment in which a buyer seeks to purchase an indivisible unit of housing in some location d. The buyer discounts future payoffs at the exponential rate $\beta \in (0,1)$. Each period begins with the buyer incurring a search cost $c_b > 0$. Then, with probability $m \in (0,1)$, the buyer is randomly matched with a seller in d; if the buyer is unmatched, the search continues into the next period. The parameter m can be interpreted as the equilibrium match probability arising from the equilibrium market tightness of a bilateral search model with random matching between buyers and sellers (Genesove, 2024).

When a match occurs, the buyer encounters a seller whose outside option reflects the payoff from waiting for another buyer. Sellers are risk-neutral and face a heterogeneous cost of waiting, c_s, drawn from a distribution F known to buyers, but with realizations unobserved by them. This heterogeneity captures, in reduced form, differences in sellers' urgency to sell.[24] Because sellers can always wait for another local buyer willing to pay the prevailing market price in d, denoted $\bar{p}_d$, the seller's outside option is $O_s = \bar{p}_d - c_s$. Intuitively, a seller facing a higher waiting cost has a stronger incentive to sell quickly and is therefore more willing to accept a discount relative to the market price. The prevailing price in d, the distribution of active sellers, and the buyer–seller matching probability are all endogenous equilibrium objects. In the remainder of this section, however, I treat $\bar{p}_d$, F, and m as given, thereby focusing on the buyer's problem in partial equilibrium.

Upon meeting a seller, the buyer observes the house's permanent quality v, which is common across all buyers and homes in the market segment. The buyer also draws a match-specific valuation for the current house, ε, distributed according to cumulative distribution function G. After observing the realization of the shock, the buyer makes a take-it-or-leave-it offer p. If the seller accepts, the buyer purchases the house at price p, obtaining utility and thus receives continuation utility $v - p + \varepsilon$. If the seller rejects the offer, the buyer continues to search in the next period, while the seller receives the outside option O_s.

It is immediate to see that, in partial equilibrium, the seller's problem reduces to a simple acceptance rule: accept any offer that yields a price above the outside option. From the buyer's

[23]This assumption mirrors the empirical analysis, which examines overpayment after partialling out a rich set of controls.

[24]For example, some sellers might need to relocate quickly because they experienced a liquidity shock (Guren and McQuade, 2020), or a recent change in household composition (Tscharaktschiew and Hirte, 2010). Throughout, it is assumed that cumulative distribution function F exhibits non-decreasing hazard rate.

perspective, the probability that an offer p is accepted is therefore

$$\psi(p) := \Pr(\text{Seller accepts offer } p) = 1 - F(\bar{p}_d - p).$$

Given this acceptance probability, the matched buyer's problem can be expressed through the following recursive Bellman equation, conditional on the realization of the taste shock ε:

$$V_b^m(\varepsilon) = \max_p \left\{ -c_b + \psi(p)\big[v + \varepsilon - p\big] + \big[1 - \psi(p)\big]\beta\mathbb{E}[V_b^\omega] \right\}. \tag{6}$$

Here, $\mathbb{E}[V_b^\omega]$ denotes the expected continuation value from search, which integrates over the probability of being matched again in the next period and the distribution of match-specific valuations for other available houses.[25]

The buyer's policy function maps realizations of the match-specific shock into optimal offers for the current house. On the one hand, the buyer prefers to offer a higher price to increase the probability of acceptance, especially when the realized shock is high. On the other hand, conditional on acceptance, lowering the price raises the surplus from the transaction. This trade-off is captured by the first-order condition for interior solutions:

$$p^*(\varepsilon): \quad \psi'(p^*)\big(v + \varepsilon - p^* - \beta\mathbb{E}[V_b^\omega]\big) = \psi(p^*), \tag{7}$$

where the left-hand side is the marginal increase in the acceptance probability, $\psi'(p)$, weighted by the surplus from closing the current transaction rather than continuing the search, $\big(v + \varepsilon - p - \beta\mathbb{E}[V_b^\omega]\big)$. The right-hand side is the marginal utility of money (normalized to one) multiplied by the probability of acceptance, $\psi(p)$.[26]

Finally, it is straightforward to show that the expected search duration at the optimal offer policy is just:

$$D^* = \frac{1}{m\mathbb{E}[\psi(p^*(\varepsilon))]}. \tag{8}$$

The comparative statics of the rational buyer's policy function are straightforward. The optimal offer is increasing in the realization of the match-specific shock ε: a higher valuation of the current match raises the buyer's surplus from purchasing this particular house relative to

[25] Formally, we have

$$\mathbb{E}[V_b^\omega] = m\,\mathbb{E}_\varepsilon[V_b^m(\varepsilon)] + (1-m)V_b^u,$$

where $\mathbb{E}_\varepsilon[V_b^m(\varepsilon)]$ denotes the expectation of the matched buyer's ($\omega = b$) value function with respect to the realization of the match-specific shock, and

$$V_b^u = -c_b + \beta\mathbb{E}[V_b^\omega]$$

is the continuation value for an unmatched buyer ($\omega = u$).

[26] As I formally show in Appendix E, the solution to the search problem is unique and the policy function is differentiable at interior solutions under mild regularity conditions.

continuing the search, leading the buyer to submit a higher offer to increase the probability of acceptance. The optimal offer is also increasing in the buyer's search cost c_b, since higher search costs make waiting for future matches more expensive and tilt the balance toward closing the current transaction more quickly. Finally, the optimal offer is increasing in the prevailing market price $\bar{p}_d$, because a higher $\bar{p}_d$ raises the seller's outside option and thus requires the buyer to bid more aggressively to secure acceptance.[27]

Crucially, in the rational search benchmark the buyer's place of origin plays no role in determining search behavior. Accordingly, as illustrated by the blue lines in Figure 4, both overpayment (shown in Panel (a)) and expected search duration (shown in Panel (b)) remain flat across the price discrepancy between origin and destination.

3.2 Housing Search with Extrapolation and Learning

To capture how prior price experiences shape buyer behavior, I extend the rational benchmark by introducing extrapolative beliefs about destination prices. This modification not only relaxes the assumption of full information but also provides a behavioral mechanism that links past market exposure to current decision-making. Consider a buyer arriving from location o, where the prevailing price level is denoted by $\bar{p}_o$. This buyer's belief about the price level in the destination market d is given by a convex combination of the origin and destination prices:

$$\tilde{p}_d \;:=\; \theta\bar{p}_o \;+\; (1-\theta)\bar{p}_d, \tag{9}$$

where $\theta \in (0,1)$ captures the degree of extrapolation from the origin. When $\theta = 1$, the buyer fully projects local price conditions onto the destination and perceives prices in d as identical to those in o. Conversely, as θ approaches zero, the buyer's beliefs converge to the true destination price $\bar{p}_d$, in which case the model reverts to the full-information rational benchmark.

I remain agnostic about the precise psychological microfoundations underlying extrapolation. Some explanations require only minimal departures from full rationality. A leading example is the difficulty of learning the joint distributions of multiple stochastic processes (Farmer et al., 2024). In this interpretation, extrapolation arises as a near-rational response to limited information: buyers rely on their prior experience to form beliefs before acquiring new information. As a result, initial beliefs continue to assign weight to priors centered at $\bar{p}_o$. Other explanations for extrapolative beliefs may stem from more serious violation of Bayesian belief formation. For example, irrelevant prices encountered previously may intrude into memory-based recall due to associative reasoning (Bordalo et al., 2022).[28]

[27] Formal statements and derivations are provided in Appendix E.

[28] While it is realistic to assume that buyers extrapolate multiple features of the origin market, I adopt the simpli-

Misperceiving destination prices distorts two key model quantities. First, buyers from more expensive (cheaper) origins misperceive the seller's outside option, leading them to raise (lower) their current offer. Second, they misperceive their own continuation value $\mathbb{E}[V_b^\omega]$, since they underestimate (overestimate) their expected search duration. This pessimism (optimism) induces them to further reduce (increase) their current offer.

As shown in Panel (a) of Figure 4, introducing extrapolation generates a positive relationship between overpayment and the gap between $\bar{p}_o$ and $\bar{p}_d$.[29] However, Panel (b) highlights a counterfactual implication: under realistic calibration of the variance in sellers' outside options, search times diverge rapidly for movers from substantially cheaper origins. These buyers consistently make offers well below market prices, face repeated rejections from sellers, and—despite receiving this feedback through prolonged and unsuccessful search—fail to adjust their beliefs.

To address this limitation, I introduce a simple learning process. At the beginning of each period, a misperceiving buyer receives a binary signal. With probability α, the signal fully corrects their bias and reveals the true price distribution in d; with probability $1 - \alpha$, the signal is uninformative and buyer's beliefs remain unchanged.[30]

Although the probability of learning is the same for all buyers, its effects are markedly asymmetric, depending on the direction of the initial misperceptions. As illustrated in Panels (a) and (b) of Figure 4, outcomes for buyers arriving from more expensive origins show only minor adjustments when learning is introduced: their acceptance probabilities are already high, leaving little scope for updating. By contrast, buyers from cheaper origins—depicted to the left of the origin in the figure—face longer search durations due to their higher likelihood of early rejection. This extended search process provides greater opportunity to learn and ultimately converge toward the true price distribution. Consequently, even a simple learning rule is sufficient to generate the characteristic "hockey stick" pattern observed in the empirical analysis.

While the process I introduce is tractable and parsimonious, it is not meant as a literal description of buyer behavior. However, the key insights are robust to changes in the specification of the learning technology: virtually any learning process in which longer search yields additional information and gradual convergence of beliefs delivers qualitatively similar predictions. This includes a Bayesian learning process where buyers draw a new signal of destination prices in each period, or a search model with constant-gain learning in the spirit of Evans et al. (2021).

fying assumption that extrapolation occurs only through the price level. In practice, extrapolation matters primarily insofar as it distorts perceptions of buyers' own outside option and that of the seller. Allowing a richer set of extrapolations would generate similar qualitative predictions.

[29] A formal derivation is provided in Appendix E.

[30] This structure parallels the staggered price adjustment technology first introduced in Calvo (1983).

3.3 Housing Search with Reference Points

To derive formal predictions of a loss-averse model, I extend the rational framework introduced in Section 3.1 by incorporating reference-dependent preferences. Following prior work in the behavioral real estate literature (e.g. Genesove and Mayer, 2001; Andersen et al., 2022), I adopt the realization utility framework originally developed by Barberis and Xiong (2012). In this setting, buyers form a reference price as a weighted average of origin and destination prices,

$$\hat{p} = \gamma \bar{p}_o + (1 - \gamma)\bar{p}_d, \tag{10}$$

where γ determines the relative weight placed on origin prices.

Consistent with the realization utility framework, the reference-dependent component of preferences is triggered upon the execution of a transaction, when the buyer experiences a burst of gain–loss utility. Specifically, after purchasing at price p, the buyer's utility flow is

$$u^{LA}(p) = v - p - \begin{cases} \eta(p - \hat{p}), & \text{if } p < \hat{p}, \\ \eta\lambda(p - \hat{p}), & \text{if } p \geq \hat{p}. \end{cases} \tag{11}$$

While the term $v - p$ corresponds to the same quasi-linear net utility component as in the rational benchmark, the second term highlights the role of gain-loss considerations. Specifically, $\eta > 0$ captures the degree of reference dependence, scaling the impact of deviations from the reference point, while $\lambda > 1$ governs the extent of loss aversion by amplifying the disutility from losses relative to equivalent gains. This formulation introduces a kink in utility at the reference price, consistent with standard models of reference-dependent preferences.

Reference points can generate a positive relationship between origin prices and overpayment, but the shape of this relationship differs from the "hockey-stick" pattern predicted by extrapolative beliefs. Instead, it follows an S-shaped curve. Overpayment increases with origin prices only for intermediate values, flattening at both extremes where buyers are either always in the loss domain or always in the gain domain.

At the lower extreme, consider buyers originating from locations where housing is so inexpensive that it is virtually impossible to find a property at the destination that is cheaper than their reference point. Because the marginal utility of money is high in the loss region, these buyers are willing to wait longer to secure a favorable price. However, for buyers from sufficiently inexpensive origins—such that the probability of being in the gain domain approaches zero—the marginal rate of substitution between search time and money remains constant. As a result, overpayment is negative but flat with respect to origin prices at this lower end.

At the upper extreme, the argument is symmetric. Buyers from very expensive origins are almost always in the gain domain, where the marginal utility of money is relatively low. These buyers search less intensively than locals, but once origin prices are sufficiently high, their marginal rate of substitution also stabilizes, producing another flat (but positive) segment of the overpayment curve at the upper tail.

To summarize, loss aversion shapes behavior most strongly for households originating from intermediate price levels, while its influence vanishes at the extremes. This generates an S-shaped relationship between overpayment and origin prices, and a reverse S-shaped relationship between origin prices and search duration.

Finally, note that enriching the loss-averse model with adaptive reference points in the spirit of DellaVigna et al. (2017) and Thakral and Tô (2021) could help flatten the overpayment curve to the left of the origin, much as learning does in the extrapolative-beliefs framework. However, such modifications would do nothing to eliminate the flattening of overpayment for movers arriving from sufficiently expensive origins and thus the overall shape of the relationship would remain S-shaped.

4 The Cause of Overpayment: Preferences or Beliefs?

The empirical evidence gathered so far is more consistent with an explanation rooted in beliefs and learning. First, the effects are, if anything, larger among renters—who are predominantly first-time homebuyers—than among previous homeowners. This pattern is difficult to reconcile with explanations based on reference-dependent preferences, as the purchase price of a previous home should be more salient for homeowners, and instead suggests that experience in the housing market reduces the likelihood of overpayment. Second, the inheritance design points to a process of initial mispricing followed by subsequent adjustment, again consistent with agents updating their beliefs about local market conditions over time. Finally, the overall shape of the relationship between overpayment and origin prices—characterized by a pronounced hockey-stick pattern rather than the S-shape predicted by reference dependence—further supports an interpretation based on extrapolative beliefs rather than stable reference points. In the remainder of this section, I return to the movers design to provide additional evidence on the underlying mechanism.

4.1 Movers Design: Additional Empirical Evidence

Panel (a) of Table 5 provides additional support for the hypothesis that the hockey-stick pattern is attenuated among more experienced buyers. The first two columns split the sample by educational attainment and show that the effect is weaker for households in which the household head holds a postgraduate degree. Similarly, comparing columns (3) and (4), the effect is stronger among younger buyers, who are likely to have less experience in the housing market. Finally, the last two columns group households into terciles of the national wealth distribution as a proxy for financial sophistication and show that the hockey-stick pattern is more pronounced among those in the lower net wealth terciles.[31]

Panel (b) instead splits the sample by the degree of house price heterogeneity in the origin ZIP code (Columns 1–3) and in the destination ZIP code (Columns 4–6). To measure heterogeneity within a ZIP code, I compute the coefficient of variation of house prices and divide ZIP codes into terciles of this measure. If beliefs drive the results, greater heterogeneity in origin prices should weaken movers' priors about destination prices, leading to faster updating upon arrival. Consistent with this prediction, the first three columns show that the hockey-stick pattern is attenuated as price dispersion at origin increases. A test of the null hypothesis that the three second-row coefficients in these columns are equal is rejected in the data ($p = 0.006$).[32] Conversely, higher dispersion in destination prices should be associated with a stronger degree of extrapolation, as learning the true price distribution becomes more difficult. While the differences across destination terciles are not statistically significant in this second exercise, the point estimates still move in the direction predicted by the theory.

4.2 Survey Evidence

Survey Design Because the study targeted active homebuyers, respondents first completed a multiple-choice screening question verifying that they were currently looking for a house at the time of the survey; those who were not were screened out, as they were not part of the target population. Appendix G reproduces the full questionnaire.

After the screening, respondents reported the duration of their housing search and then listed up to three ZIP codes they were considering as potential locations for purchasing a home. They received the instruction: *"Please list up to three U.S. ZIP codes you are currently considering for buying a house. If you list more than one, put them in order from most likely to least likely based on your current preference."* Responses were entered as five-digit ZIP codes. In a follow-up screen, respondents also reported the ZIP code of their current residence.

I then elicited beliefs about recent housing prices in each ZIP code of interest. The set of ZIP codes of interest consisted of all destination ZIP codes listed by the respondent as well as the

[31] Lusardi and Mitchell (2014) documents how financial sophistication increases with age and education, while wealthier households generally exhibit higher levels of financial literacy. Moreover, Fagereng et al. (2020) show that wealthier households also earn higher returns on their financial investments, supporting the use of household net wealth as a proxy for financial sophistication.

[32] Formally, using the notation introduced in Equation 3, let $\beta_{>0}^{\text{Dispersion Tercile}}$ denote the estimated elasticity of overpayment for movers relocating from more expensive to cheaper markets (i.e., moving down in prices). I test the joint null hypothesis $\beta_{>0}^{\text{Bottom}} = \beta_{>0}^{\text{Middle}} = \beta_{>0}^{\text{Top}}$, corresponding to the null hypothesis that the magnitude of the elasticity does not change with price dispersion.

respondent's origin ZIP code. For each of these ZIP codes, I elicited beliefs about both listing and sale prices using analogous prompts. Respondents were asked: *"What is your best guess for the median sale price for a single-family home in ZIP code [current elicitation ZIP code] over the past three months?"* The order in which ZIP codes appeared was randomized across participants, and each elicitation screen included short reminders explaining the definitions and differences between listing and sale prices.[33]

For a randomly assigned half of the sample, these price elicitations were also incentivized. I implemented a quadratic scoring rule similar to the one used in Enke et al. (2024). Following the recommendations in Danz et al. (2022), incentivized respondents received a simplified on-screen explanation accompanied by an optional "More information" button. They were informed that one of their elicitations would be randomly selected for payment and that they could earn up to one extra dollar. The full formula used to compute the bonus, along with additional details on the calculation, was accessible through the additional information button.[34]

In the second part of the survey, I introduced a hypothetical choice module that experimentally varied hypothetical house prices at destination to test for reference dependence. Appendix Section D.3 describes the design of this module and presents the main findings. The remainder of the survey collected demographic characteristics, self-assessed real-estate expertise, and several proxies for current and expected household income as well as current household wealth.

Survey Logistics In November 2025, I administered the survey online to 1,000 U.S. prospective homebuyers recruited through Prolific. Respondents received a fixed participation payment of $2, plus a bonus payment to incentivize forecasts as described above. The median completion time was 10 minutes and 58 seconds. For the main expectations analysis, I require that respondents report price expectations for at least one destination ZIP code different from their origin ZIP code. I also trim belief-elicitation responses at the 2.5th and 97.5th percentiles.[35] After applying these criteria, the final sample for the main analysis consists of 640 respondents, yielding a total of 1,334 price elicitations. Summary statistics for the survey sample are reported in Table A.4. Importantly, respondents' beliefs are well calibrated: elicited prices closely track true median prices for both origin and destination ZIP codes, as shown in Appendix Figure A.3.

[33] If a ZIP code is mentioned twice, for example because origin and destination coincide, I only elicit beliefs once.

[34] Let t denote the true value of house prices for the elicitation selected to be relevant for the bonus payment, measured using Redfin data available to researchers, and let $\hat{t}_i$ denote the respondent's answer for that elicitation. The bonus payment is computed as

$$\max\left\{1 - \tfrac{1}{2500}\,(t-\hat{t}_i)^2,\ 0\right\}.$$

Respondents receive the full $1 for an exact guess, with the bonus decreasing quadratically in the prediction error and reaching zero once the error exceeds $50,000.

[35] I additionally dropped two respondents due to a Qualtrics data-restriction malfunction that allowed them to enter values in formats inconsistent with the survey's input filters.

Extrapolation Regression and Elicitation Noise Define respondent i's forecast error for house prices in ZIP code z as

$$FE(i,z) := F_i(p_z) - p_z,$$

where $F_i(p_z)$ denotes the respondent's reported belief about the median price in z, and p_z is the corresponding true median price over the previous three months (computed using Redfin transactions). Intuitively, if forecast errors systematically increase with the gap between a respondent's origin price and the price in a prospective destination, then respondents tend to overpredict prices in destinations that are more expensive than their own origin and underpredict prices in destinations that are cheaper. Such a pattern is consistent with extrapolative belief formation: respondents anchor their beliefs at their origin ZIP code and insufficiently adjust toward the true price in the destination.

This intuition motivates the empirical specification in which I regress destination-specific forecast errors on the discrepancy between respondents' beliefs about origin prices and true destination prices. Proposition 3 shows that, when origin beliefs are measured without noise, this regression directly identifies the structural degree of extrapolation θ.

In practice, however, survey responses may contain elicitation noise. If the econometrician only observes $\tilde{F}_i(p_z) = F_i(p_z) + \eta_{i,z}$, where $\eta_{i,z}$ captures elicitation noise, then both the regressor and the dependent variable are measured with error. When the elicitation noise is mean-zero and independent across questions, the resulting attenuation bias pushes the estimated degree of extrapolation toward zero. Intuitively, if all responses were pure white noise, there would be no extrapolation because beliefs would exhibit no correlation across questions. In contrast, if part of the survey noise is respondent-specific—that is, correlated across all responses from the same respondent—the bias may instead be positive. Respondent-specific noise mechanically inflates the covariance between beliefs about origin and destination locations. Proposition 4 formally characterizes the sign of this bias and shows that it is, a priori, ambiguous, depending on the relative magnitudes of idiosyncratic versus respondent-level noise.

Finally, Proposition 5 shows that the structural parameter θ can be recovered even in the presence of both types of survey noise. If the econometrician observes the true origin–destination price difference $p_{o(i)} - p_d$, then this variable is a valid instrument for the noisy regressor $\tilde{F}_i(p_{o(i)}) - p_d$. Because the true price difference is uncorrelated with all survey-noise components, the IV estimator purges both attenuation and the spurious positive correlation induced by common noise.

Main Survey Results Table 6 reports the estimates from the regression of forecast errors on the price discrepancy between origin and destination ZIP codes introduced in the previous paragraph. Figure 6 complements these results with binned-scatter plots based on the raw data. Panel (a) displays the relationship for median sale-price elicitations, while Panel (b) presents the corresponding results for median listing-price elicitations. As expected, forecast errors are systematically related to the price difference between origin and destination ZIP codes. This relationship remains highly robust to the inclusion of additional controls, as shown by comparing columns (1)–(3).

Moving from the OLS estimates in columns (1)–(3) to the IV specifications in columns (4)–(6), we observe a slight attenuation in the coefficients. This pattern is consistent with the interpretation that correlated survey noise induces an upward bias in the OLS estimates, and that the IV procedure partially corrects for this bias. Appendix Table A.5 reports the corresponding first-stage regressions, which confirm that the instrument is strong and precisely estimated across all specifications.

Appendix Table A.6 provides additional robustness checks. Restricting the sample to homebuyers searching for single-family homes—those most likely to be knowledgeable about the prices elicited in the survey—yields nearly identical results. Likewise, restricting the sample to respondents who passed an unannounced attention check requiring recall of earlier survey conditions further strengthens the results. Appendix Table A.7 reports results from a log-log specification.

5 Conclusion

This paper shows that geographic price extrapolation meaningfully distorts housing search and bargaining. Using administrative data that link the universe of Danish house transactions to buyers' prior addresses, I document a pronounced "hockey-stick" relationship between quality-adjusted overpayment and the price gap between a mover's origin and destination markets. Movers arriving from more expensive origins systematically overpay relative to comparable locals for the same houses, whereas movers from cheaper origins relocating to more expensive destinations do not. These patterns are robust to the inclusion of a rich set of controls, hold under alternative measures of both local prices and overpayment, and apply to renters as well as homeowners.

To distinguish preferences from beliefs, I combine these reduced-form facts with a search-and-bargaining framework that allows for learning about local prices. The model rationalizes the asymmetric overpayment as the consequence of extrapolative beliefs that converge only gradually with market experience. Consistent with this mechanism, evidence from an inheritance setting—where beneficiaries price and sell inherited parental homes outside their own neighborhoods—shows that inherited properties are listed high, remain longer on the market, and do not realize sale premia. Together, the movers and inheritance designs point to belief-driven misperception rather than stable reference-dependent preferences.

The implications of belief extrapolation extend to market equilibrium. When some buyers overpay, they raise sellers' outside options, allowing even a small share of transactions to shift bargaining outcomes. In markets where supply is not perfectly elastic, these stronger outside options translate into higher equilibrium prices that affect both movers and locals. By contrast, the arrival of movers whose beliefs imply underpayment cannot drive prices down, since sellers can simply ignore low offers until these buyers either learn or exit the market. As a result, extrapolation can only push local prices upward, amplifying gentrification dynamics across housing markets.

Some limitations remain. First, the Danish context may not translate directly to the U.S., where institutions, financing, and mobility patterns differ. Second, the distinction between belief-driven learning and loss aversion becomes less clear once adaptive reference points are introduced. The survey evidence provides complementary support for extrapolative beliefs among U.S. homebuyers, but it cannot fully resolve how these forces translate into realized transaction outcomes.

Finally, I note that future work could deploy randomized information interventions for movers and non-movers to explore implications for geographic mobility and gentrification dynamics. Additionally, studying how intermediaries—and especially online listing platforms—shape belief formation would be of first-order importance.

Tables and Figures

Table 1: Movers Design, Overpayment and House Prices at Origin

	Overpayment				
	(1)	(2)	(3)	(4)	(5)
$1(O \leq D) \times$ (Log Price Origin $-$ Log Price Dest.)	0.010***	0.004*	0.002	0.004	0.000
	(0.002)	(0.002)	(0.006)	(0.005)	(0.003)
$1(O > D) \times$ (Log Price Origin $-$ Log Price Dest.)	0.051***	0.039***	0.037***	0.050***	0.037***
	(0.003)	(0.003)	(0.005)	(0.004)	(0.004)
ZIP House $\times$ Quarter FE	Yes	Yes	Yes	Yes	Yes
Zip Origin Controls	Yes	Yes	Yes	Yes	Yes
Household Controls	No	Yes	Yes	Yes	Yes
Zip Origin FE	No	No	Yes	No	No
Observations	257,222	257,222	257,222	93,762	154,535
Sample	Movers	Movers	Movers	Movers	Movers
Restriction				Renters	Owners

Notes: This table reports regressions of buyers' quality-adjusted *Overpayment* on the difference in house price indices between the buyers' origin and destination ZIP codes. The dependent variable *Overpayment* is the residual from the hedonic regression in Equation 1 and is measured in log-points. The main regressor, *Log Price Origin − Log Price Destination*, captures the level difference in house prices between the buyers' origin and destination ZIP code and is interacted with indicators for the sign of the origin–destination price gap to allow for asymmetric effects. The first row reports estimates for moves from cheaper to more expensive markets ($O \leq D$); the second row reports estimates for moves from more expensive to cheaper markets ($O > D$). Figure 1 provides a graphical illustration of the result.

All specifications include destination ZIP code–by–calendar quarter fixed effects. *ZIP Origin Controls* denotes a vector of characteristics of the origin ZIP code at the time of the move. *Household Controls* denotes demographic and financial characteristics of the buyer. *ZIP Origin FE* denotes fixed effects for the origin ZIP code. The regression specification is given in Equation 3; a complete list of all control variables and their definitions appear in Appendix B.1. The *Movers* sample comprises all housing transactions from 1996–2017 in which the buyer moved across ZIP codes prior to the transaction. Columns (4) and (5) restrict the sample to buyers who were renters and homeowners, respectively, before the move.

Standard errors are two-way clustered by destination ZIP code and calendar quarter.

$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 2: Repeated Sales, Overpayment and House Prices at Origin

	Log Sale Price		
	(1)	(2)	(3)
$1(\mathrm{O} \leq \mathrm{D}) \times$	-0.001	-0.002	0.003
Log Price Origin - Log Price Dest	(0.003)	(0.003)	(0.005)
$1(\mathrm{O} > \mathrm{D}) \times$	0.052***	0.048***	0.050***
Log Price Origin - Log Price Dest	(0.004)	(0.004)	(0.005)
Property FE	Yes	Yes	Yes
ZIP House × Quarter FE	Yes	Yes	Yes
Origin and Household Controls	Yes	Yes	Yes
Time Varying Hedonics	No	Yes	Yes
Zip Origin FE	No	No	Yes
R-sq.	0.953	0.956	0.956
Within R-sq.	0.064	0.116	-
Observations	190,264	190,264	190,264
Sample	Repeated Sales	Repeated Sales	Repeated Sales

Notes: This table reports the main result for the Repeated Sales design introduced in Section 2.2. The table reports regressions of total *Log Sale Price* paid by a buyer on the difference in house price indices between the buyers' origin and destination ZIP codes. The dependent variable, *Log Sale Price*, is the logarithm of the final sale price paid by the buyer. The main regressor, *Log Price Origin − Log Price Destination*, captures the level difference in house prices between buyer origin and destination ZIP codes and is interacted with indicators for the direction of the price difference to allow for asymmetric effects (Equation 3). The first row reports estimates for moves from cheaper to more expensive markets ($O \leq D$), and the second row reports estimates for moves from more expensive to cheaper markets ($O > D$).
All specifications include *Property FE*, ensuring within-property comparisons the main feature of the repeated sales design. I also always include destination ZIP code–by–calendar quarter fixed effects. The same control vectors as in the Movers specification are included in all columns. *ZIP Origin Controls* denotes a vector of characteristics of the origin ZIP code at the time of the move. *Household Controls* denotes demographic and financial characteristics of the buyer. Variable definitions are provided in Appendix B.1. *Time-Varying Property Controls* denotes the inclusion of time-varying property characteristics measured at the time of the move. I include fixed effects for the building age, the property's occupancy status, and the most recent tax-assessed market value interacted with the assessment year. The *Repeated-Sale* sample comprises all housing transactions from 1996–2017 for which at least two transactions are observed for the same housing unit. Standard errors are two-way clustered by destination ZIP code and calendar quarter.
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 3: Home Equity Instrument, Overpayment and House Prices at Origin

	Overpayment					
	(1)	(2)	(3)	(4)	(5)	(6)
1(O $\leq$ D) $\times$ (Log Price Origin - Log Price Dest)	-0.005 (0.012)		0.020 (0.022)		-0.047 (0.053)	
1(O $>$ D) $\times$ (Log Price Origin - Log Price Dest)		0.039*** (0.007)		0.052*** (0.010)		0.055** (0.023)
ZIP House $\times$ Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
Origin and Household Controls	Yes	Yes	Yes	Yes	Yes	Yes
Municipality of Origin FE	No	No	No	No	Yes	Yes
Observations	100,933	156,333	27,670	71,550	100,933	156,333
Sample	Movers	Movers	Movers	Movers	Movers	Movers
Restriction	-	-	Renters	Renters	-	

Notes: This table presents results from two-stage least squares regressions of overpayment on the difference in house price indexes between a buyer's origin and destination ZIP codes. The dependent variable, *Overpayment*, is measured as the residual from the hedonic regression in Equation 1 and is measured in log-points. In odd-numbered columns I report estimates for the subsample of households moving from cheaper to more expensive housing markets ($O \leq D$), while in even-numbered columns I report estimates for households moving from more expensive to cheaper housing markets ($O > D$).

Log Price Origin $-$ *Log Price Destination* is instrumented with a fifth-order polynomial in the median home equity among sellers in the origin location at the time of the sale. Details on the construction of the instrument and the first-stage specification are provided in Section 2.3. Figure 3 provides a graphical illustration of the first-stage relationship. All specifications include destination ZIP code–by–calendar quarter fixed effects. *ZIP Origin Controls* denotes a vector of time-varying characteristics of the origin ZIP code. *Household Controls* denote demographic and financial characteristics of the buyer. *Municipality Origin FE* denotes fixed effects for the municipality of origin. All detailed variable definitions are provided in Appendix B.1. In parentheses, I report standard errors clustered two-way by ZIP code of destination and quarter of move.

The *Movers* sample which consists of all house transactions between 1996 and 2017 where the buyer moved across ZIP codes before the transaction.

$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 4: Inheritance Design, Main Results

	Log Listing Premium	Revise Listing Down	Log Time on Market	Log Realized Sale Premium
	(1)	(2)	(3)	(4)
Log Price Child - Log Price Home	0.026***	0.032*	0.037*	-0.003
	(0.008)	(0.018)	(0.021)	(0.005)
ZIP House × Quarter FE	Yes	Yes	Yes	Yes
Household Controls	Yes	Yes	Yes	Yes
Observations	10231	10231	8656	8656
Sample	Inheritance	Inheritance	Inheritance	Inheritance
Restriction	Listed	Listed	Sold	Sold

Notes: This table presents main results for the inheritance design presented in Section 2.4. The dependent variables are defined as follows. *Log Listing Premium* is the log difference between the property's initial list price and its predicted hedonic quality as resulting from the hedonic specification in Equation 1. *Revise Listing Down* is an indicator equal to one if the list price was revised downward during the listing period. *Log Time on Market* is the log of the number of days between the initial listing and the sale. *Log Realized Sale Premium* is the log difference between the final sale price and the predicted hedonic quality. The main regressor, *Log Price Child − Log Price Home*, denotes the current difference in house price indexes between the child's ZIP code of residence and the ZIP code of the inherited property.
All specifications include house ZIP code–by–calendar quarter fixed effects. *Household Controls* denote demographic and financial characteristics of the seller (i.e. the child inheriting the property). Variable definitions are provided in Appendix B.1.
The *Inheritance* sample consists of all house transactions between 2008 and 2017 where the seller is a single child who inherited the parental property. The *Listed* subsample denotes all cases where the house appears for listing. The *Sold* subsample denotes transactions where a subsequent sale is observed. Robust standard errors are clustered by destination ZIP code and by calendar quarter.
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 5: Mechanisms: Heterogeneity of Overpayment Elasticity

(a) Heterogeneity by Buyer Sophistication

	Overpayment					
	(1)	(2)	(3)	(4)	(5)	(6)
$1(O \leq D) \times$ (Log Price Origin − Log Price Dest.)	0.004	-0.001	0.008*	0.001	0.009*	-0.001
	(0.005)	(0.003)	(0.004)	(0.003)	(0.005)	(0.003)
$1(O > D) \times$ (Log Price Origin − Log Price Dest.)	0.036***	0.022***	0.049***	0.028***	0.042***	0.032***
	(0.004)	(0.004)	(0.003)	(0.005)	(0.004)	(0.006)
Observations	198,797	51,657	131,359	111,672	79,004	79,205
Heterogeneity	Master's Degree		Age of HH Head		Wealth Tercile	
	No	Yes	Age $\leq$ 40	Age $>$ 40	Bottom	Top

(b) House Price Dispersion and Overpayment

	Overpayment					
	(1)	(2)	(3)	(4)	(5)	(6)
$1(O \leq D) \times$ (Log Price Origin - Log Price Dest)	-0.001	-0.000	-0.001	0.004	-0.009*	0.008**
	(0.006)	(0.005)	(0.006)	(0.006)	(0.005)	(0.003)
$1(O > D) \times$ (Log Price Origin - Log Price Dest)	0.039***	0.035***	0.030***	0.039***	0.040***	0.043***
	(0.007)	(0.006)	(0.006)	(0.005)	(0.004)	(0.006)
Observations	80,063	85,342	74,225	86,201	85,317	85,748
Heterogeneity	House Price Dispersion at Origin			House Price Dispersion at Destination		
Tercile	Bottom	Middle	Top	Bottom	Middle	Top

Notes: This table examines heterogeneity in buyers' quality-adjusted *Overpayment* responses to the house price gap between origin and destination ZIP codes. Panel (a) explores heterogeneity by buyer sophistication, while Panel (b) investigates heterogeneity by local house price dispersion in the buyers' origin and destination ZIP codes.
In Panel (a), the sample is split by three proxies for buyer financial sophistication at the time of purchase: education (Master's degree), age of the household head (age ≤ 40 vs. age > 40), and net-wealth terciles (bottom vs. top). In Panel (b), Columns (1)–(3) report separate estimates by terciles of house price dispersion at the origin ZIP code, while Columns (4)–(6) re-estimate the specification by dispersion at the destination ZIP code. Price dispersion is measured as the coefficient of variation of log sale prices within a ZIP code.
All estimates correspond to the preferred specification from the baseline Movers Design (Equation 3; Column (2) of Table 1). The analysis sample is the baseline *Movers* sample covering years 1996–2017. Standard errors, reported in parentheses, are two-way clustered by destination ZIP code and calendar quarter.

$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 6: Survey Forecast Errors and Origin Destination House Price Discrepancy

(a) Sale Price

	Forecast Error in Destination Median Sale Price					
	(1)	(2)	(3)	(4)	(5)	(6)
Forecasted Origin Sale Price - True Destination Sale Price	0.492*** (0.029)	0.489*** (0.029)	0.486*** (0.029)	0.379*** (0.035)	0.423*** (0.048)	0.330*** (0.053)
Demographic Controls	No	Yes	Yes	Yes	Yes	Yes
Housing Taste Controls	No	No	Yes	Yes	Yes	Yes
Elicitations	1334	1334	1334	1334	655	679
Unique Respondents	640	640	640	640	319	321
Estimation	OLS	OLS	OLS	2SLS	2SLS	2SLS
First-stage F-stat	-	-	-	998.750	562.587	445.587
R-sq.	0.440	0.458	0.466	0.412	0.432	0.386
Sample	All Responses	All Responses	All Responses	All Responses	Incentivized Price Elicit.	Non-Incentivized Price Elicit.

(b) List Price

	Forecast Error in Destination Median List Price					
	(1)	(2)	(3)	(4)	(5)	(6)
Forecasted Origin List Price - True Destination List Price	0.591*** (0.027)	0.592*** (0.027)	0.590*** (0.027)	0.449*** (0.038)	0.488*** (0.050)	0.376*** (0.058)
Demographic Controls	No	Yes	Yes	Yes	Yes	Yes
Housing Taste Controls	No	No	Yes	Yes	Yes	Yes
Elicitations	1347	1347	1347	1347	654	693
Unique Respondents	647	647	647	647	322	325
Estimation	OLS	OLS	OLS	2SLS	2SLS	2SLS
First-stage F-stat	-	-	-	998.239	491.404	468.499
R-sq.	0.498	0.514	0.520	0.461	0.507	0.393
Sample	All Responses	All Responses	All Responses	All Responses	Incentivized Price Elicit.	Non-Incentivized Price Elicit.

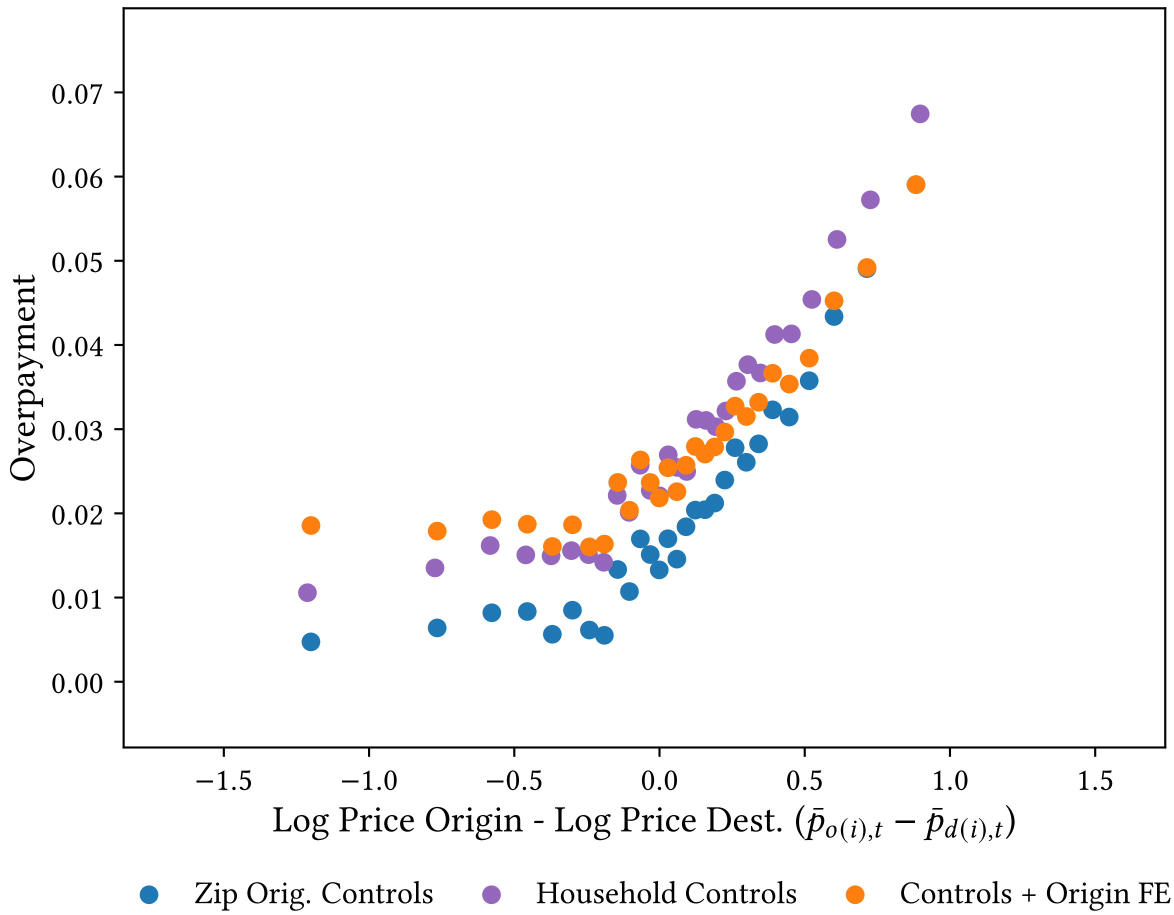
Notes: This table reports regressions of forecast errors for respondents' median house price elicitations in destination ZIP codes onto the respondent's elicited belief about median prices in their origin ZIP code minus the true median price in the destination ZIP code. Forecast errors are defined as the respondent's survey response for the destination minus the true price in the destination, where true prices are computed using Redfin data. All prices are measured in U.S. dollars.

Panel (a) uses median *sale*-price elicitations, and Panel (b) uses median *listing*-price elicitations. The survey question was worded as follows: *"What is your best guess for the median [sale/listing] price for a single-family home in ZIP code [current elicitation ZIP code] over the past three months?"*. Columns (1)–(3) report OLS estimates with varying sets of controls. Columns (4)–(6) report two-stage least squares (2SLS) estimates, instrumenting the elicited belief about origin prices minus the true destination price with the true origin price minus the true destination price. Column (5) includes only respondents whose price elicitations were incentivized. Column (6) includes only respondents whose price elicitations were not incentivized.

Demographic controls include fixed effects for five-year age bins, gender, marital status, and self-reported income decile. *House-taste controls* include fixed effects for the respondent's preferred type of home and desired number of bedrooms. Each observation corresponds to an individual destination elicitation, and the number of unique respondents contributing to the estimation is reported at the bottom of the table. Standard errors, clustered at the respondent level, are shown in parentheses. First-stage regression results for Columns (4)–(6) are reported in Appendix Table A.5.

$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Figure 1: Movers Design, Overpayment and House Prices at Origin



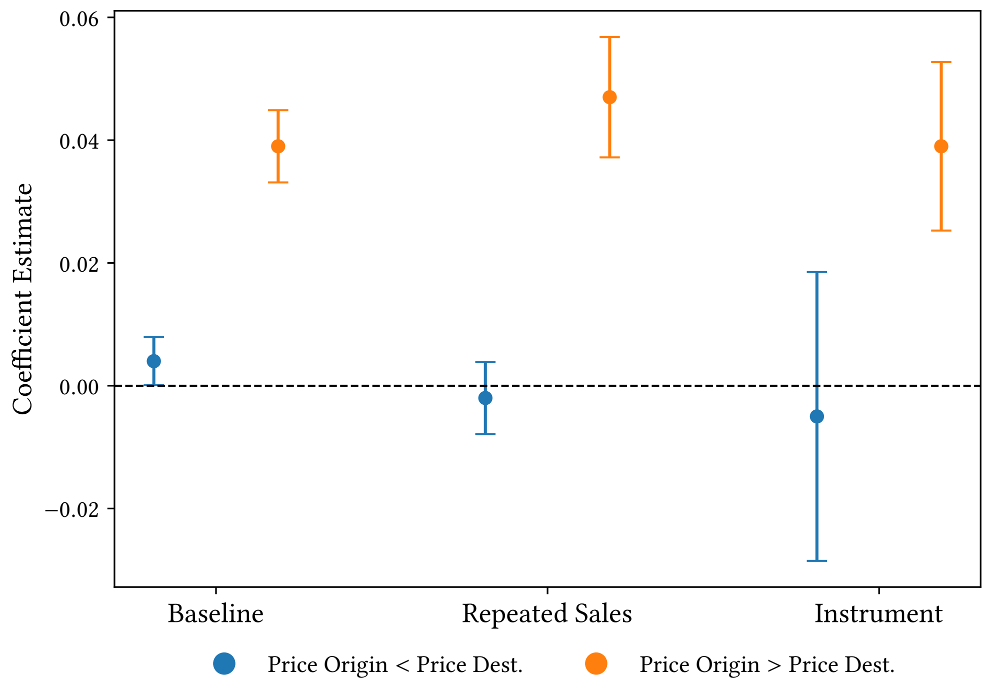
Notes: This figure presents a binned scatterplot of buyers' quality-adjusted *Overpayment* on the difference in house price indices between the buyers' origin and destination ZIP codes. The dependent variable *Overpayment* is the residual from the hedonic regression in Equation 1 and is measured in log-points. The main regressor, *Log Price Origin − Log Price Destination*, captures the level difference in house prices between the buyers' origin and destination ZIP codes.

All specifications partial out destination ZIP code by calendar quarter fixed effects. The *ZIP Origin Controls* specification (in blue) also partials out a vector of characteristics of the origin ZIP code, corresponding to Column (1) of Table 1. The *Household Controls* specification (in green) further partials out demographic and financial characteristics of the buyer, corresponding to Column (2) of Table 1. The *All Controls* specification (in yellow) also partials out ZIP code of origin fixed effects and thus corresponds to Column (3) of Table 1. Variable definitions are provided in Appendix B.1.

The figure is constructed using the *Movers* sample which consists of all house transactions between 1996 and 2017 where the buyer moved across ZIP codes before the transaction.

To partial out controls, I use the commands and procedures described in Cattaneo et al. (2024).

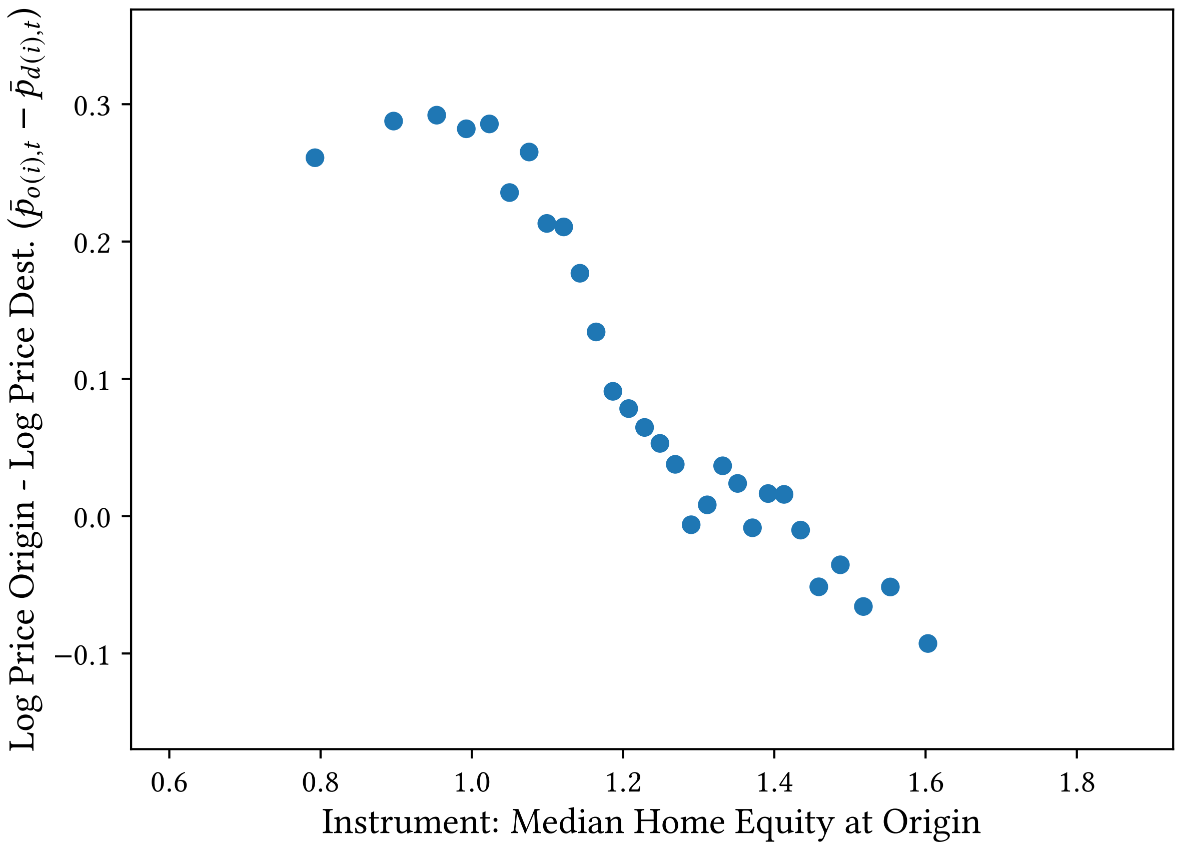
Figure 2: Elasticity of Overpayment: Comparison of Designs



Notes: This figure reports estimates of the elasticity of quality-adjusted overpayment with respect to the origin–destination house price discrepancy from the preferred specification of each design.
Blue bars correspond to moves up the house price distribution (origin ZIP cheaper than destination ZIP); orange bars correspond to moves down (origin ZIP more expensive than destination ZIP). *Baseline* estimates correspond to the Movers Design discussed in Section 2.1. *Repeated Sales* estimates refer to the design discussed in Section 2.2. *Instrument* reports estimates from the design outlined in Section 2.3.

Dots denote point estimates while the bars denote 95 percent confidence intervals constructed with standard errors two-way clustered by destination ZIP and calendar quarter.

Figure 3: First Stage, House Prices and Sellers' Home Equity



Notes: This figure presents the graphical relationship behind the first stage for the results presented in Table 3. I show a binned scatterplot of the instrumented variable, the difference in house price indexes between a buyer's origin and destination ZIP codes, against the instrument.

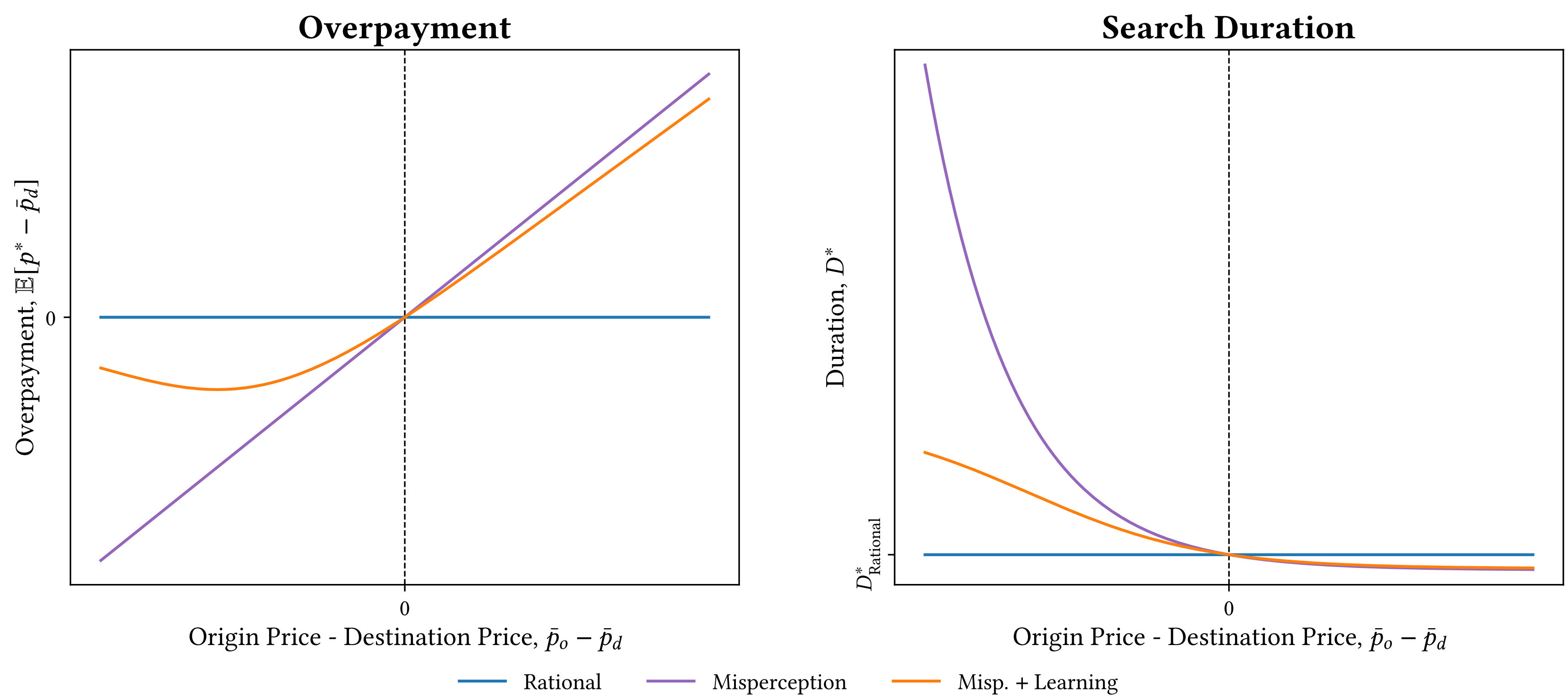
The dependent variable, *Log Price Origin − Log Price Destination*, is defined as the change in house price indexes between origin and destination for a mover. The instrument, on the x-axis, is a measure of home equity among sellers in the movers' origin ZIP code. The instrument is constructed as described in Section 2.3.

The figure partials out destination ZIP code by calendar quarter fixed effects, effectively plotting the relationship between house price at origin and the instrument.

I construct the figure using the *Movers* sample which consists of all house transactions between 1996 and 2017 where the buyer moved across ZIP codes before the transaction.

To partial out controls, I use the commands and procedures described in Cattaneo et al. (2024).

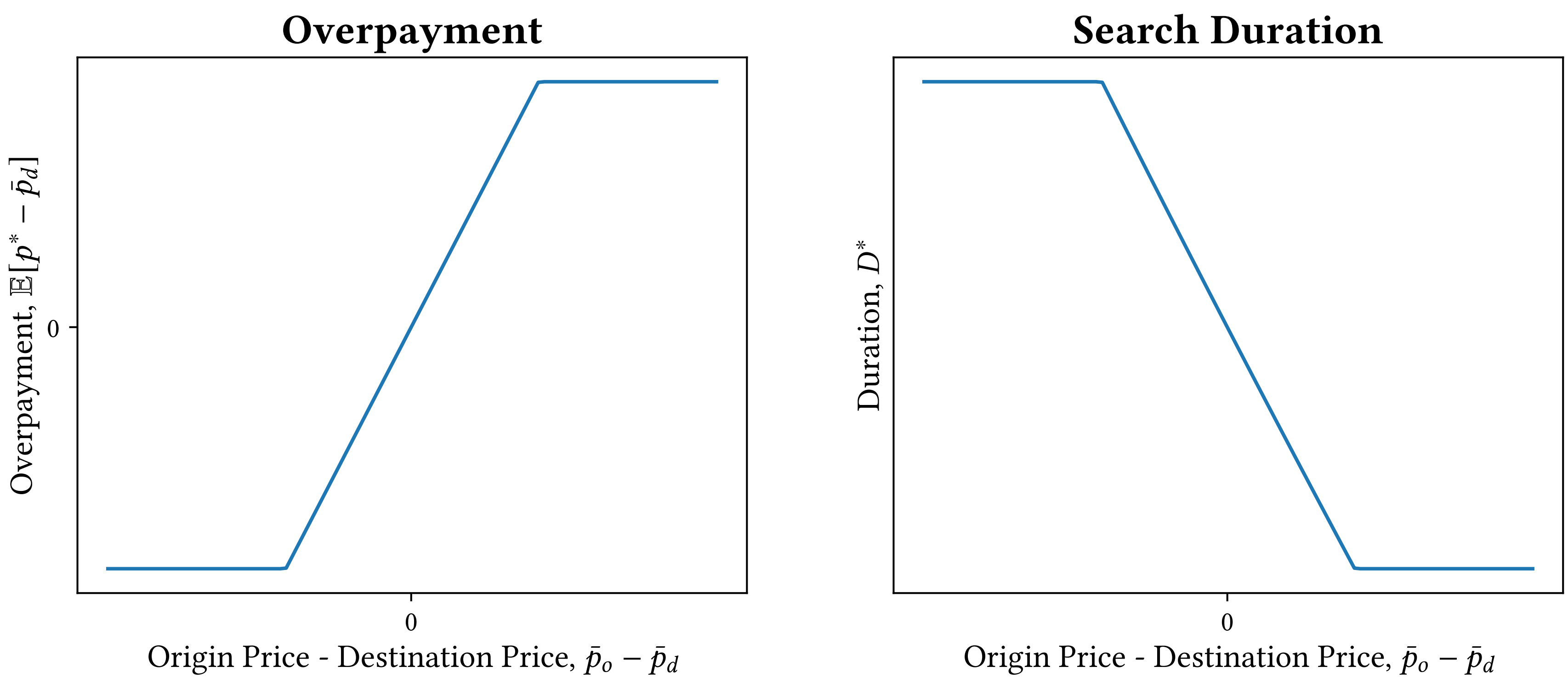
Figure 4: Model Predictions, Extrapolative Beliefs and Learning



Notes: Predictions from the buyer housing search model with extrapolation and learning (Section 3).
Panel (a) plots expected overpayment for movers arriving in market d against the origin-destination house price difference, for comparison with the empirical counterpart in Figure 1. Panel (b) plots search duration against the same variable, unobserved in the empirical data.
The blue line shows the full-information rational benchmark ($\theta = 0$). The green line allows for geographic extrapolation without learning ($\theta > 0, \alpha = 0$). The yellow line introduces both extrapolation and learning ($\theta > 0, \alpha > 0$).

Details on the parameters and procedures used to produce the figure are discussed in Appendix E.3.

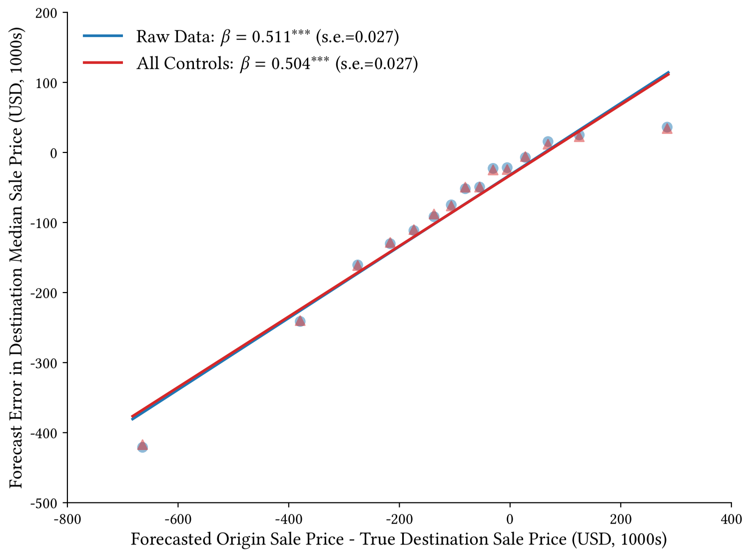
Figure 5: Model Predictions, Loss Aversion



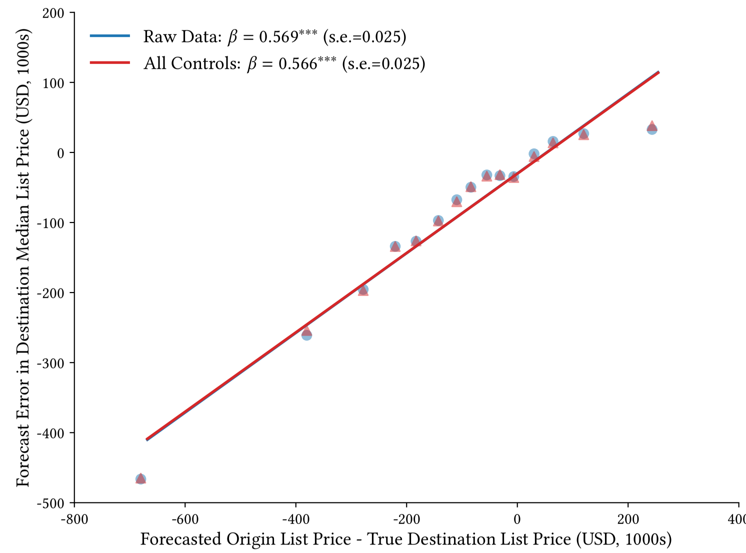
Notes: Predictions from the buyer housing search model with loss aversion (Section 3). Panel (a) plots expected overpayment for movers arriving in market d against the origin-destination house price difference, for comparison with the empirical counterpart in Figure 1. Panel (b) plots search duration against the same variable, unobserved in the empirical data. The blue line shows the model prediction for a loss-averse buyer. Details on the parameters and procedures used to produce the figure are discussed in Appendix E.3.

Figure 6: Price Discrepancy and Forecast Errors in Median House Price at Destination

(a) Sale Prices



(b) List Prices



Notes: This figure presents binned scatterplots of respondents' forecast errors for median house prices in destination ZIP codes against the respondent's elicited belief about median prices in their origin ZIP code minus the true median price in the destination ZIP code. Forecast errors are defined as the respondent's survey response for the destination minus the true price in the destination, with true prices measured using Redfin data. The horizontal axis, *Forecasted Origin Price - True Destination Price*, measures the difference between the respondent's elicited belief about median house prices in their origin ZIP code and the true median price in the elicited destination ZIP code. All prices are measured in thousands of U.S. dollars. Panel (a) displays the raw relationship using no additional controls. Panel (b) presents the specification that partials out the full set of controls. The *Demographic Controls* include fixed effects for five-year age bins, gender, marital status, and self-reported income decile. The *House-Taste Controls* include fixed effects for the respondent's preferred type of home and desired number of bedrooms. The *All Controls* specification partials out both vectors of controls simultaneously. Each observation corresponds to an individual destination elicitation.

Partialling-out procedures follow the recommendations and implementation in Cattaneo et al. (2024).

References

Abadie, Alberto, Susan Athey, Guido W Imbens, and Jeffrey M Wooldridge, "When Should You Adjust Standard Errors for Clustering?," *The Quarterly Journal of Economics*, December 2022, *138* (1), 1–35.

Aiello, Darren, Jason Kotter, and Gregor Schubert, "When Money Moves In: The Consequences of Housing Wealth," 2024.

Andersen, Asger Lau, Niels Johannesen, and Adam Sheridan, "Bailing out the Kids: New Evidence on Informal Insurance from one Billion Bank Transfers," *SSRN Electronic Journal*, 2020.

Andersen, Steffen and Kasper Meisner Nielsen, "Fire Sales and House Prices: Evidence from Estate Sales Due to Sudden Death," *Management Science*, January 2017, *63* (1), 201–212.

__ , Cristian Badarinza, Lu Liu, Julie Marx, and Tarun Ramadorai, "Reference Dependence in the Housing Market," *American Economic Review*, October 2022, *112* (10), 3398–3440.

__ , John Y Campbell, Kasper Meisner Nielsen, and Tarun Ramadorai, "Sources of Inaction in Household Finance: Evidence from the Danish Mortgage Market," *American Economic Review*, 2020, *110* (10), 3184–3230. Publisher: American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

__ , Julie Marx, Kasper Meisner Nielsen, and Lise Vesterlund, "Gender Differences in Negotiation: Evidence from Real Estate Transactions," *The Economic Journal*, August 2021, *131* (638), 2304–2332.

Badarinza, Cristian, Tarun Ramadorai, Juhana Siljander, and Jagdish Tripathy, "Behavioral Lock-In: Housing Market Taxation with Reference Dependent Agents," 2025.

Bailey, Michael, Ruiqing Cao, Theresa Kuchler, and Johannes Stroebel, "The Economic Effects of Social Networks: Evidence from the Housing Market," *Journal of Political Economy*, December 2018, *126* (6), 2224–2276.

Bajari, Patrick, Phoebe Chan, Dirk Krueger, and Daniel Miller, "A Dynamic Model of Housing Demand: Estimation and Policy Implications," *International Economic Review*, 2013, *54* (2), 409–442. Publisher: [Economics Department of the University of Pennsylvania, Wiley, Institute of Social and Economic Research, Osaka University].

Barberis, Nicholas and Wei Xiong, "Realization utility," *Journal of Financial Economics*, 2012, *104* (2), 251–271.

Benveniste, L. M. and J. A. Scheinkman, "On the Differentiability of the Value Function in Dynamic Models of Economics," *Econometrica*, May 1979, *47* (3), 727.

Bordalo, Pedro, John J Conlon, Nicola Gennaioli, Spencer Y Kwon, and Andrei Shleifer, "Memory and Probability," *The Quarterly Journal of Economics*, December 2022, *138* (1), 265–311.

__ , Nicola Gennaioli, and Andrei Shleifer, "Memory and Reference Prices: An Application to Rental Choice," *AEA Papers and Proceedings*, May 2019, *109*, 572–576.

Bracke, Philippe and Silvana Tenreyro, "History Dependence in the Housing Market," *American Economic Journal: Macroeconomics*, April 2021, *13* (2), 420–443.

Calvo, Guillermo A., "Staggered prices in a utility-maximizing framework," *Journal of Monetary Economics*, 1983, *12* (3), 383–398.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller, "Robust Inference With Multiway Clustering," *Journal of Business & Economic Statistics*, April 2011, *29* (2), 238–249.

Cattaneo, Matias D, Richard K Crump, Max H Farrell, and Yingjie Feng, "On Binscatter," *American Economic Review*, 2024.

Charness, Gary and Uri Gneezy, "Incentives to exercise," *Econometrica : journal of the Econometric Society*, 2009, *77* (3), 909–931.

Danz, David, Lise Vesterlund, and Alistair J. Wilson, "Belief Elicitation and Behavioral Incentive Compatibility," *American Economic Review*, September 2022, *112* (9), 2851–2883.

DellaVigna, Stefano, Attila Lindner, Balázs Reizer, and Johannes F. Schmieder, "Reference-dependent job search: Evidence from Hungary," *The Quarterly Journal of Economics*, May 2017, *132* (4), 1969–2018. tex.eprint: https://academic.oup.com/qje/article-pdf/132/4/1969/30637685/qjx015.pdf.

Enke, Benjamin, Frederik Schwerter, and Florian Zimmermann, "Associative Memory, Beliefs, and Market Interactions," *Journal of Financial Economics*, 2024.

Evans, David, George W. Evans, and Bruce McGough, "Learning when to say no," *Journal of Economic Theory*, June 2021, *194*, 105240.

Fagereng, Andreas, Luigi Guiso, Davide Malacrino, and Luigi Pistaferri, "Heterogeneity and Persistence in Returns to Wealth," *Econometrica*, 2020, *88* (1), 115–170.

Farmer, Leland E., Emi Nakamura, and Jón Steinsson, "Learning about the Long Run," *Journal of Political Economy*, October 2024, *132* (10), 3334–3377.

Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams, "Place-Based Drivers of Mortality: Evidence from Migration," *American Economic Review*, August 2021, *111* (8), 2697–2735.

Genesove, David, "Search and Matching, and Price Formation in Real Estate Markets," in "Oxford Research Encyclopedia of Economics and Finance," Oxford University Press, November 2024.

— **and Christopher Mayer**, "Loss Aversion and Seller Behavior: Evidence from the Housing Market," *The Quarterly Journal of Economics*, 2001, *116* (4), 1233–1260. Publisher: Oxford University Press.

— **and Lu Han**, "Search and matching in the housing market," *Journal of Urban Economics*, July 2012, *72* (1), 31–45.

Gomes, Francisco, Michael Haliassos, and Tarun Ramadorai, "Household Finance," *Journal of Economic Literature*, September 2021, *59* (3), 919–1000.

Greenwood, Robin and Andrei Shleifer, "Expectations of returns and expected returns," *Review of Financial Studies*, 2014, *27* (3), 714–746.

Guren, Adam M., "House Price Momentum and Strategic Complementarity," *Journal of Political Economy*, June 2018, *126* (3), 1172–1218.

Guren, Adam M and Timothy J McQuade, "How Do Foreclosures Exacerbate Housing Downturns?," *The Review of Economic Studies*, May 2020, *87* (3), 1331–1364.

Han, Lu and William C. Strange, "Bidding Wars for Houses," *Real Estate Economics*, March 2014, *42* (1), 1–32.

Hastings, Justine S. and Jesse M. Shapiro, "Fungibility and Consumer Choice: Evidence from Commodity Price Shocks*," *The Quarterly Journal of Economics*, November 2013, *128* (4), 1449–1498.

Jayachandran, Seema, Lea Nassal, Matthew J Notowidigdo, Marie Paul, Heather Sarsons, and Elin Sundberg, "Moving to Opportunity, Together," *National Bureau of Economic Research Working Paper Series*, 2024, *No. 32970.*

Kahneman, Daniel and Amos Tversky, "Prospect theory: An analysis of decision under risk," *Econometrica : journal of the Econometric Society*, 1979, *47* (2), 263–292.

Kleven, Henrik Jacobsen, Martin B. Knudsen, Claus Thustrup Kreiner, Søren Pedersen, and Emmanuel Saez, "Unwilling or Unable to Cheat? Evidence From a Tax Audit Experiment in Denmark," *Econometrica*, 2011, *79* (3), 651–692.

Kuchler, Theresa and Basit Zafar, "Personal Experiences and Expectations about Aggregate Outcomes," *The Journal of Finance*, 2019, *74* (5), 2491–2542. Publisher: Wiley Online Library.

__ , Monika Piazzesi, and Johannes Stroebel, "Expectations and housing markets," in R. Bachmann, G. Topa, and W. van der Klaauw, eds., *Handbook of economic expectations*, Elsevier, 2023, pp. 431–456.

Lambson, Val E., Grant R. McQueen, and Barrett A. Slade, "Do Out-of-State Buyers Pay More for Real Estate? An Examination of Anchoring-Induced Bias and Search Costs," *Real Estate Economics*, March 2004, *32* (1), 85–126.

Lusardi, Annamaria and Olivia S. Mitchell, "The Economic Importance of Financial Literacy: Theory and Evidence," *Journal of Economic Literature*, March 2014, *52* (1), 5–44.

MacKinnon, James G., Morten Ørregaard Nielsen, and Matthew D. Webb, "Cluster-robust inference: A guide to empirical practice," *Journal of Econometrics*, February 2023, *232* (2), 272–299.

Malmendier, Ulrike and Stefan Nagel, "Learning from Inflation Experiences," *The Quarterly Journal of Economics*, 2016, *131* (1), 53–87. Publisher: Oxford University Press.

Mullainathan, Sendhil and Eldar Shafir, *Scarcity: Why having too little means so much*, Times Books, 2013.

Oster, Emily, "Unobservable Selection and Coefficient Stability: Theory and Evidence," *Journal of Business & Economic Statistics*, April 2019, *37* (2), 187–204.

Piazzesi, Monika, Martin Schneider, and Johannes Stroebel, "Segmented Housing Search," *American Economic Review*, March 2020, *110* (3), 720–759.

Simonsohn, Uri and George Loewenstein, "Mistake: The Effect of Previously Encountered Prices on Current Housing Demand," *The Economic Journal*, January 2006, *116* (508), 175–199.

Stein, Jeremy C., "Prices and Trading Volume in the Housing Market: A Model with Down-Payment Effects," *The Quarterly Journal of Economics*, May 1995, *110* (2), 379–406.

Stock, James and Motohiro Yogo, “Testing for weak instruments in linear IV regression,” in “Identification and Inference for Econometric Models,” New York: Cambridge University Press, 2005, pp. 80–108.

Thakral, Neil and Linh T. Tô, “Daily Labor Supply and Adaptive Reference Points,” *American Economic Review*, August 2021, *111* (8), 2417–2443.

Tscharaktschiew, Stefan and Georg Hirte, “How does the household structure shape the urban economy?,” *Regional Science and Urban Economics*, November 2010, *40* (6), 498–516.

Online Appendix

Table of Contents

A Summary Statistics

Table A.1: Summary Statistics, Movers Sample and Repeated Sales Sample

	Movers Sample		Repeated Sales Sample	
	Mean	Std. Dev.	Mean	Std. Dev.
Buyer Demographics				
Age of Household Head	41.1	12.7	41.4	12.8
Female Household Head (%)	45.0	49.7	45.5	49.8
Has Partner (%)	77.3	41.9	74.5	43.6
Bachelor Degree (%)	43.6	49.6	38.4	48.6
Number of Children	0.8	0.9	0.8	1.0
Home Ownership (%)	61.8	48.6	65.1	47.6
Movers Zip (%)			52.6	50.0
Buyer Financial Situation				
Household Income (2015-USD)	105,122	59,470	102,261	58,959
Household Financial Assets (2015-USD)	55,789	111,416	51,581	100,956
Household Liabilities (2015-USD)	274,724	376,412	287,912	375,577
Tax Assessed House Value (2015-USD)	191,432	258,224	195,360	251,103
Property Characteristics				
Living Area (Sq. feet)	1,227	476	1,144	447
Real Sale Price (2015-USD)	258,917	147,300	232,652	131,327
Number of Rooms	4.0	1.5	3.7	1.4
Observations	257,222		190,123	

Notes: This table presents summary statistics for the main analysis samples. An observation is a house purchase matched with the household characteristics of the buyer. The *Movers Sample* refers to housing transactions where the buyer moved across ZIP codes before purchasing the house. The *Repeated Sales* sample refers to all transactions where we observe at least two different transactions for the same house, irrespective of the buyer type. For details on the definition of household and the construction of each variable refer to Appendix B.1. All nominal quantities in local currency (Danish Kroner) are deflated using the Consumer Price Index (base year 2015) reported by Statistics Denmark and converted to US dollars using the average exchange rate in 2015. The data covers the years 1996 to 2017.

Table A.2: Summary Statistics, Inheritance Design Sample

	Inheritance Sample	
	Mean	Std. Dev.
Seller Demographics		
Age of Household Head	51.3	9.5
Female Household Head (%)	46.8	49.9
Bachelor Degree (%)	43.5	49.6
Has Partner (%)	66.1	47.3
Nr. Children	0.7	1.2
Seller Financial Situation		
Household Income (2015-USD)	102,993	62,854
Household Liquid Assets (2015-USD)	92,394	140,891
Homeowner (%)	75.1	43.3
Tax Assessed House Value (2015-USD)	360,032	302,261
Property Characteristics		
Sale Price (2015-USD)	242,785	191,558
Nr. Rooms	4.0	1.3
Living Area (Sq. Ft)	1,223	398
Observations	10,231	

Notes: This table presents summary statistics for the Inheritance Design sample. An observation is a house purchase matched with the household characteristics of the seller. The *Inheritance Design* refers to the analysis where I study house sales by single children who inherit parental properties and subsequently list them for sale. For details on the definition of household and the construction of each variable refer to Appendix B.1. All nominal quantities in local currency (Danish Kroner) are deflated using the Consumer Price Index (base year 2015) reported by Statistics Denmark and converted to US dollars using the average exchange rate in 2015. The inheritance design data covers the years 2008 to 2017.

B Data Appendix

B.1 Household Data Details

Definition of Household I follow Statistics Denmark and define a household as either a single person or a couple (married, in a registered partnership, or cohabiting under CPR-based criteria) together with their cohabiting children. Children are included in the parental family if they live at the same address, are under 25, unmarried, without children, and not part of a cohabiting couple. Cohabiting couples are identified either (i) if they have a child together registered in the CPR, or (ii) if they do not have children together, they are of opposite sex, not closely related, have an age difference of less than 15 years, and constitute the only two adults in the household.

Aggregation of Individual Administrative Records at the Household Level Financial and income data are reported at the individual level in personal tax records and must therefore be aggregated to the household level for analysis. When aggregating financial data, I simply sum each source of income and each financial account balance across all household members.

Consistent with previous studies, I define the household head as the oldest member of the household (Andersen and Nielsen, 2017; Andersen et al., 2020a).

Household Characteristics I measure all household characteristics of buyers at the end of the calendar year preceding the housing transaction. Below I describe each variable and its construction:

- **Age of household head:** age of the oldest member of the household.
- **Marital status:** indicator equal to one if the household head lives with a partner. This includes all partnerships recognized by Statistics Denmark: cohabiting couples, registered partnerships, and formal marriages.
- **Female household head:** indicator equal to one if the household head is female.
- **Educational attainment:** highest educational attainment among all household members, measured using single-digit ISCED codes (2011 revision).
- **Number of children:** number of children below age 25 living in the same housing unit as the household head.
- **Average past household income:** total household income is measured before taxes and labor market contributions, and includes labor income, public transfers, property income, and other taxable income attributable to household members. Labor income encompasses

taxable wage income, benefits, bonuses, severance pay, and stock option values. I compute average household income as the mean of total household income over the three calendar years preceding the housing transaction.

- **Homeownership status:** indicator equal to one if any household member is flagged as a homeowner by tax authorities.
- **Liquid financial assets:** constructed following Andersen et al. (2022). I include the total value of bank deposits, stocks, and bonds as reported by Danish financial institutions to SKAT. Physical cash and foreign assets are not included, as they are unobserved by Danish tax authorities.
- **Value of household real estate:** market value as assessed every other year by tax authorities of all Danish properties owned by household members, including houses, commercial real estate, farmland, and undeveloped land.
- **Outstanding mortgage balances:** obtained from Danmarks Nationalbank, which collects information from mortgage banks via Finance Denmark. For households with multiple mortgaged properties or properties with multiple mortgages, I sum all mortgage balances.
- **Outstanding non-mortgage household debts:** following Andersen et al. (2022), I include all non-mortgage liabilities such as student loans, unsecured consumer credit, tax arrears, and other debts.
- **Concurrent real estate transactions:** For households identified as homeowners, I record whether the household sells any real estate property within a one-year window around the move. I include an indicator equal to one if such a sale is observed, as well as the logarithm of the proceeds from the transaction interacted with this indicator.

ZIP-Code Characteristics In some analyses I also control for a vector of demographic and financial characteristics of the household's origin ZIP code. I also add controls for the characteristics of the housing stock at origin. All variables are measured with yearly frequency. I describe each variable below:

- **Median number of children per household**: median across households in the ZIP code of the number of children below age 25 living with their parents.
- **Median age of residents**: median age of all residents in the ZIP code.
- **Fraction married**: share of residents in the ZIP code who are married.

- **Log of median household income**: log of the median household income among residents, measured consistently with the household-level definition.
- **Asinh of median household net wealth**: median of the inverse hyperbolic sine transformation of household net wealth, where net wealth is defined as assets minus liabilities, as defined in the household-level definitions.
- **Homeownership rate**: share of households in the ZIP code that own their residence.
- **Fraction apartments**: share of dwellings in the ZIP code that are apartments.
- **Fraction unoccupied dwellings**: share of dwellings in the ZIP code that are unoccupied at the end of the calendar year.
- **Log of median living area size**: log of the median dwelling size in square meters.
- **Urban/rural indicator**: indicator equal to one if the ZIP code is classified as rural by Statistics Denmark.

B.2 Hedonic Specification and Property Characteristics

In the hedonic specification, I estimate unit quality by fitting

$$f(H_{h,t}) = \sum_{\text{Type}\in\{\text{House},\text{Apt}\}} \left[\Psi^{\text{Type}}_{\text{tax}}(\tau_{h,t}) + \Psi^{\text{Type}}_{\text{Living area}}(a_h) + \Psi^{\text{Type}}_{\text{Lot size}}(l_h) + \beta^{\text{Type}}\tilde{H}_{h,t}\right],$$

where:

- All variables are interacted with the unit type (apartment or detached house).
- $\Psi^{\text{Type}}_{\text{tax}}(\tau_{h,t})$ is a third-order polynomial in the log of the most recent tax-assessed value, interacted with the calendar year of assessment.
- $\Psi^{\text{Type}}_{\text{Living area}}(a_h)$ is a third-order polynomial in the log of the unit's living area.
- $\Psi^{\text{Type}}_{\text{Lot size}}(l_h)$ is a third-order polynomial in the log of the lot size.
- $\tilde{H}_{h,t}$ is a vector of additional unit characteristics, including:
  1. fixed effects for building age in five-year bins,
  2. indicators for whether the unit is unoccupied at the time of sale, designated as historic, or located in a rural area,

3. fixed effects for number of rooms and number of bathrooms,
4. fixed effects for the floor of the unit (if apartment) and number of floors in the building (if detached house)

B.3 Inheritance Design: Sample Construction

To identify decedents and their beneficiaries I largely follow the procedures outlined in Andersen et al. (2021). I begin by identifying recently deceased individuals using administrative data from the Danish cause-of-death registry (Dodsaag). These data are collected by the National Board of Health from death certificates that record the time, place, and cause of death. From detailed tax records, I then identify decedents who owned a house at the time of passing. I restrict the sample to decedents who are either widowed or single and have one living single child.

Through a data agreement with RealView Denmark, I have access to the universe of housing listings from 2007 through 2020. Statistics Denmark anonymized the identifiers in these data, which makes it possible to merge them with the registry data.

For beneficiaries, I impose the following restrictions: they must be above 25 years of age, own fewer than three properties (excluding the inherited one), have complete demographic and financial information, and sell the inherited house within two calendar years of the decedent's death.

B.4 A Note on Empirical Cumulative Distribution of Continuous Variables

Our data agreement with Statistics Denmark allows us to export only statistics computed in samples of at least five individuals. This restriction is only binding when we plot empirical cumulative distribution functions (CDF) of continuous functions where each pixel represents information of a single individual. To comply with the data provider, we adopt the following procedure whenever we plot a CDF. First we order the data in increasing order with respect to the variable we are studying. Then, we collapse the data in ordered bins of ten observations and substitute individual values with bin averages. We then plot the empirical CDF of this collapsed data.

B.5 Data Citations

- Statistics Denmark, Befolkningen (BEF; Population Register)
- Statistics Denmark, Indkomst (IND; Income from tax returns)
- Statistics Denmark, Uddannelser (UDDA; Education)

- Statistics Denmark, Boligtællingen (BOL; Housing census)
- Statistics Denmark, Ejendomme salgsoplysninger (EJSA; Property sales information)
- Statistics Denmark, Ejendomsskatter (EJSK; Property taxes)
- Statistics Denmark, Ejere af ejendomme (EJER; Property owners)
- Statistics Denmark, Detaljeret lønmodtagerdata fra e-Indkomst (BFL; Detailed employee data from e-Income)
- Statistics Denmark, Landspatientregistret (LPR; National Patient Register)
- Statistics Denmark, Døde i Danmark (DOD; Deaths in Denmark)
- Statistics Denmark, Boligfinansiering (REAL; Housing finance)

C Robustness and Additional Empirical Results

C.1 Are Movers Comparable?

In this section, I assess whether movers from different who experience different house prices differences follow comparable income trajectories. Leveraging a staggered event-study design around the time of the move, I show that, once my preferred set of controls is included, income evolves similarly across movers, regardless of the house price discrepancy they experienced between origin and destination.

Income Trajectory Around the Move: Estimation Let t denote calendar time measured in years and τ_i indicate the move year for household i. Further define $r_{i,t} := t - \tau_i$ as the relative event time. Our main outcome of interested is the evolution of total log household income $y_{i,t}$ around the move.

To assess whether movers are comparable, I estimate the event study separately for each quintile $q \in \{1, 2, 3, 4, 5\}$ of the origin-destination price discrepancy, $\bar{p}_{o(i),t} - \bar{p}_{d(i),t}$ and compare the resulting estimates of dynamic effects at each relative time across the five quintiles. If the movers are on similar income trajectories we should observe that at any relative time r the estimate of the dynamic effect is similar for all quintiles q.

I note that estimating the true causal effect of moving on income is not necessary for the validity of this exercise, as the objective is simply to assess heterogeneity in income dynamics across movers. For a more thorough discussion of the identification of the treatment effect heterogeneity using a staggered event study, I refer to (Jayachandran et al., 2024), who employ a similar specification to study heterogeneity in the effects of moving on income by gender.

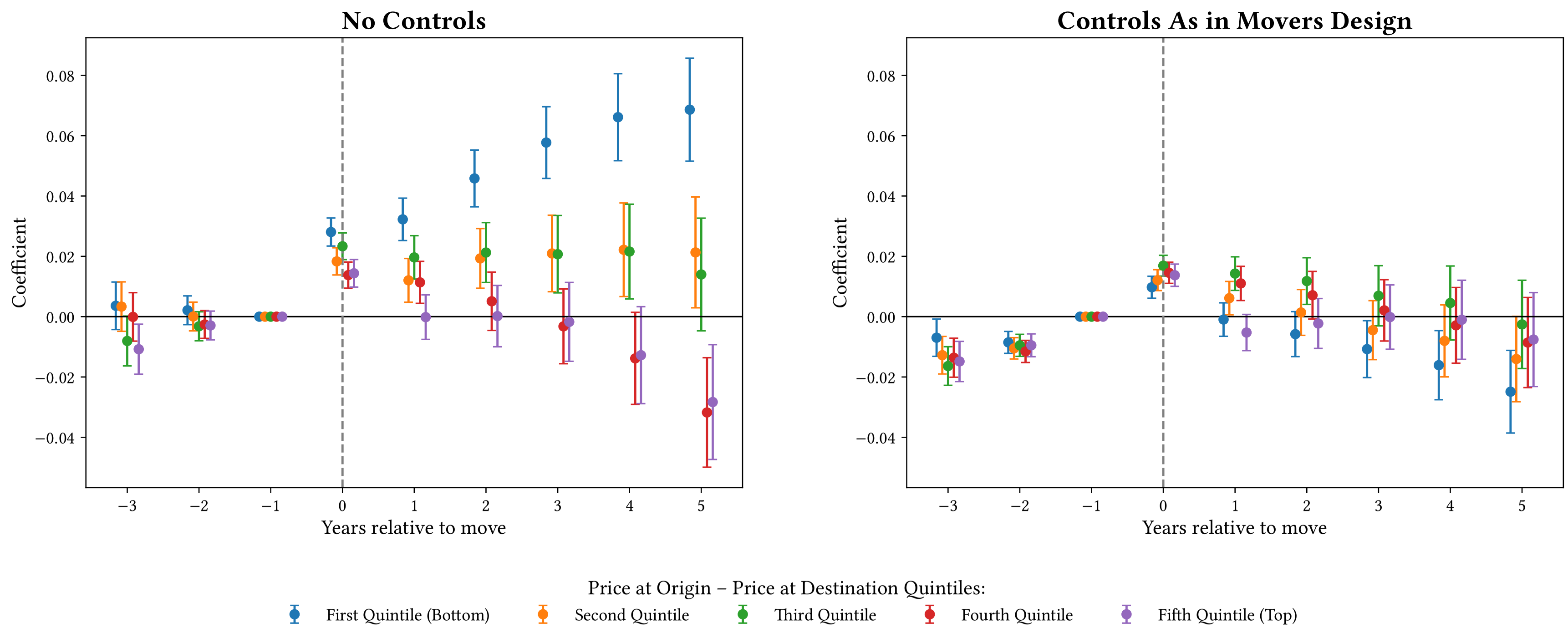
To estimate dynamic effects I estimate the following regression specification:

$$y_{it}^{q} = \sum_{r \in \mathcal{R},\, r \neq -1} \alpha_r^q \, \mathbf{1}\{r_{it} = r\} \;+\; \delta_t^q \;+\; \gamma_{g(i)}^q \;+\; \gamma_{\text{age}(i,t)}^q \;+\; X'_{i,r-1}\beta^q \;+\; \varepsilon_{it}^q, \tag{A.1}$$

where the coefficients $\{\alpha_r^q\}_{r \in \mathcal{R}}$ trace the dynamic path of household income around the move. Showing comparability amounts to showing that for a given r the estimate does not vary with q; δ_t^q are calendar-year fixed effects; and $\gamma_{g(i)}^q$ are cohort of treatment fixed effects defined as destination ZIP code interacted with the calendar quarter of the move for comparability with the main specification. $\gamma_{\text{age}(i,t)}^q$ are calendar age fixed effects to absorb life-cycle income changes. Finally, the vector $X_{i,r-1}$ contains the same pre-move controls as in the cross-sectional specification, measured at the end of year $r-1$ for comparability with Section 2. Event-time dummies are binned for leads preceding $r = -3$ and I omit $r = -1$ as the reference period. Standard errors are clustered at the household level.

Income Trajectory Around the Move: Results Figure A.1 presents the estimated income trajectories from the staggered event-study design. Panel (a) reports estimates from a specification of Equation A.1 that excludes the main controls introduced in the movers design. In this specification, income trajectories differ systematically across movers who experienced different discrepancies in origin-destination house price levels. Consistent with prior findings in the labor literature, movers relocating to more expensive destinations (the first and second quintiles) tend to experience increases in income around the move, reflecting migration toward areas with greater economic opportunity. By contrast, households moving to cheaper destinations exhibit modest declines in income. When the full set of controls is introduced in Panel (b), however, the estimated trajectories align closely across all quintiles, indicating that these controls effectively capture compositional differences in movers' characteristics.

Figure A.1: Movers Income Trajectories by Quintile of Price Origin - Price Destination: Event Study



Notes: This figure plots dynamic effects estimates of household income around the time of the move, separately by quintile of the origin-destination house price discrepancy. The event study specification is introduced in Equation A.1. Panel (a) reports estimates from a specification excluding additional household-level control variables, while Panel (b) includes the preferred vector of household and origin ZIP code characteristics corresponding to Column (2) of Table 1, measured at the end of the year prior to the move.

The outcome variable is total log household income, as reported in annual tax filings. Relative event time equal to minus one is omitted as the reference period. Dots denote point estimates, and vertical bars indicate 95 percent confidence intervals with standard errors clustered at the household level.

The sample is a panel extension of the *Movers Sample* described in Section 1.

C.2 Renovations around Transaction

Home Renovations Reform The Danish *Boligjobordningen* ("Home Job Scheme") was introduced in 2011 as a temporary stimulus measure to boost consumption following the 2008–2009 recession. The reform allowed households to deduct labour expenses for approved home services—such as maintenance, renovations, and cleaning—performed by external service providers on primary residences or summer houses, excluding material costs. Between 2011 and 2015, the maximum deductible amount was DKK 15,000 per adult household member. From 2016 to 2018, the deduction was divided into two categories: up to DKK 12,000 for maintenance and renovation work, and up to DKK 6,000 for other household services. Administrative tax data record all claims under this scheme, providing a high-quality proxy for household-level renovation and maintenance spending. [36]

Construction of Renovation Data To construct a consistent measure of household renovation activity, I use individual-level tax records capturing annual claims under the *Boligjobordningen* scheme. For each household, I aggregate the claimed deductions across all adult members within a calendar year to obtain a household-level measure of renovation expenditure. Using these data, I define two groups of variables. For *sellers*, I measure renovation activity in the year preceding the property transaction and two years before the sale. For *buyers*, I analogously compute indicators for whether the household undertook renovations in the year following the purchase and in the subsequent year. Each variable is constructed in two versions: (i) an *indicator variable* equal to one if total deductions exceed USD 1,000 in the relevant year(s), and (ii) a *continuous variable* reflecting the total nominal value of the claimed household renovations.[37]

Analysis and Results To examine whether households from different origins are more likely to purchase recently renovated properties—or, conversely, whether they are more likely to undertake renovations after moving—I regress the renovation variables on a specification analogous to my preferred model (see Equation 3 and Column (2) of Table 1).

Table A.3 reports the results of this analysis, with Panel (a) referring to sellers' renovations and Panel (b) to buyers' renovations. Because the analysis covers a different set of years than the baseline sample, Column (1) in each panel replicates the main result to ensure comparability. Columns (2) and (3) show that there is no evidence that sellers are more likely to renovate immediately before the sale when the buyer originates from a more expensive area. These columns use the indicator variables for renovation activity in $t-1$ only and in $t-1$ and $t-2$, respectively.

[36] For further official detail, see the Danish Tax Agency guidance. Danish Tax Agency – Home Improvements Deduction.

[37] All nominal amounts are converted into U.S. dollars using the prevailing average exchange rate for the corresponding calendar year.

Columns (4) and (5) present analogous results using the total dollar amount of seller renovations in years $t-1$ and $t-2$. The estimated coefficients are not only statistically insignificant but also economically small: the upper bounds of their confidence intervals are on the order of a few hundred U.S. dollars. Similar patterns are observed in Panel (b) for buyers, indicating no systematic relationship between origin prices and post-purchase renovation behavior.

Table A.3: Renovations Around Purchase Date

(a) Seller's Renovations Before Purchase

	(1) Overpayment	(2) 1(Seller renovations) in $t-1$	(3) 1(Seller renovations) in $t-1$ and $t-2$	(4) Seller renovations USD value in $t-1$	(5) Seller renovations USD value in $t-1$ and $t-2$
1(O≤D) ×	0.002	-0.000	-0.003	-22.815	-26.951
Log Price Origin - Log Price Dest	(0.004)	(0.010)	(0.012)	(32.693)	(46.373)
1(O>D) ×	0.041***	0.010	0.009	38.548	76.482
Log Price Origin - Log Price Dest	(0.005)	(0.012)	(0.013)	(40.232)	(50.546)
ZIP House × Quarter FE	Yes	Yes	Yes	Yes	Yes
Origin Controls	Yes	Yes	Yes	Yes	Yes
Household Controls	Yes	Yes	Yes	Yes	Yes
Observations	56,601	53,921	45,858	53,921	45,858
Sample	Movers	Movers	Movers	Movers	Movers
Years	2013-2017	2013-2017	2014-2017	2013-2017	2014-2017

(b) Buyer's Renovations After Purchase

	(1) Overpayment	(2) 1(Buyer renovations) in $t+1$	(3) 1(Buyer renovations) in $t+1$ and $t+2$	(4) Buyer renovations USD value in $t+1$	(5) Buyer renovations USD value in $t+1$ and $t+2$
1(O≤D) ×	0.000	0.013	0.009	22.386	-24.822
Log Price Origin - Log Price Dest	(0.004)	(0.009)	(0.010)	(34.431)	(50.824)
1(O>D) ×	0.039***	0.004	0.015	-8.260	-16.599
Log Price Origin - Log Price Dest	(0.005)	(0.009)	(0.010)	(31.510)	(46.750)
ZIP House × Quarter FE	Yes	Yes	Yes	Yes	Yes
Origin Controls	Yes	Yes	Yes	Yes	Yes
Household Controls	Yes	Yes	Yes	Yes	Yes
Observations	73,292	73,257	73,128	73,257	73,128
Sample	Movers	Movers	Movers	Movers	Movers
Years	2011-2017	2011-2017	2011-2017	2011-2017	2011-2017

Notes: This table reports regressions of renovation activity around the property transactions on the difference in house price indices between the buyers' origin and destination ZIP codes. The analysis follows the specification in Equation 3, using the same control vectors as in the movers regressions (Table 1).
Panel (a) reports results for sellers' renovations prior to the transaction, while Panel (b) reports results for buyers' renovations following the purchase. Variables are defined both as (i) an indicator equal to one if total deductions exceed USD 1,000 in the relevant year(s), and (ii) the total nominal value of claimed renovations. Column (1) in each panel re-estimates the main specification within the relevant subsample to verify that subsequent null results are not driven by changes in sample composition. Columns (2) and (3) in Panel (a) use the indicator version of the renovation variable, equal to one if the selling household's renovation activity exceeds USD 1,000 in the year before the sale or in the two years before the sale, respectively. Columns (4) and (5) report corresponding regressions using the total nominal value of renovations over the same periods. Results for buyers in Panel (b) are constructed analogously for renovations undertaken in the year and two years after purchase.
Each specification includes destination ZIP code–by–calendar quarter fixed effects and controls for origin ZIP code characteristics (*ZIP Origin Controls*) and buyer demographics and finances (*Household Controls*). The estimation sample covers 2011–2018, corresponding to the years when the home renovation deduction scheme was in place. Standard errors are two-way clustered by destination ZIP code and calendar quarter.
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

D Survey: Additional Results

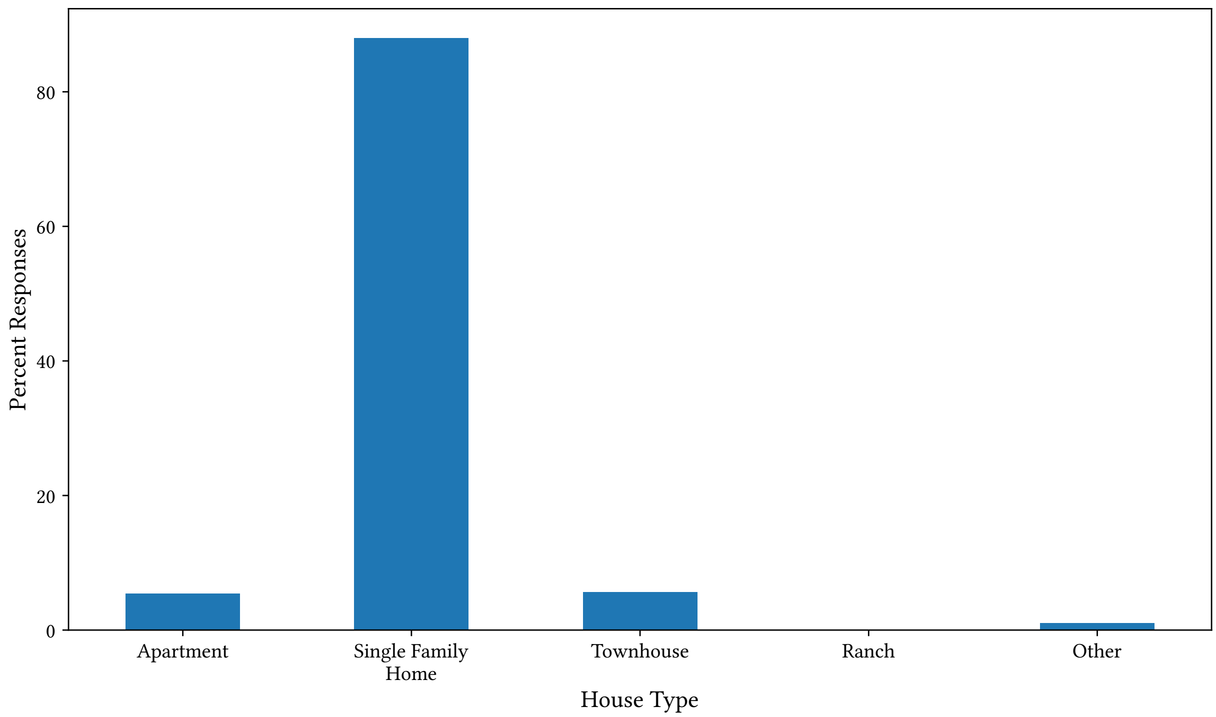
D.1 Summary Statistics

Table A.4: Survey Responses: Summary Statistics

	Mean	Std. Dev
Age (in years)	41.0	11.7
Bachelor or Higher (%)	60.2	49.0
Married (%)	78.1	41.4
Household Income (USD)	112,101	194,374
Homeowner (%)	39.6	48.9
Respondents	998	

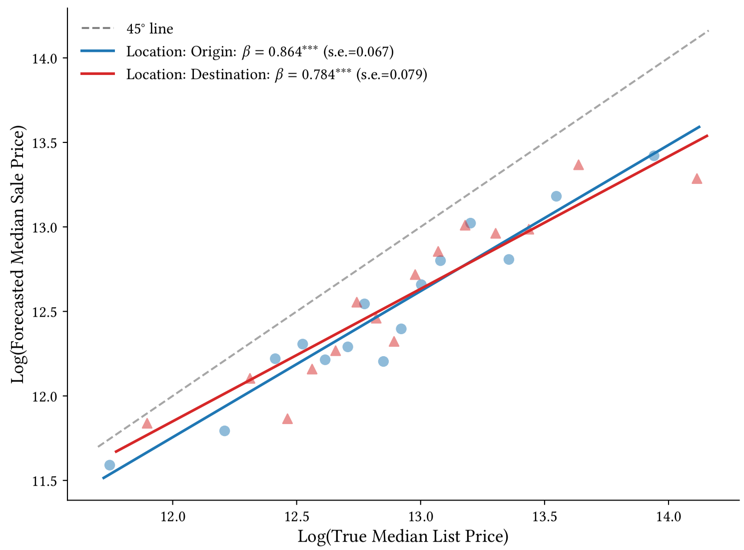
Notes: This table displays summary statistics for the US survey sample. "Age", "Married", and "Bachelor or higher", "Household Income" and "Homeowner" refer to self-reported demographic information among survey respondents.

Figure A.2: Desired House Type

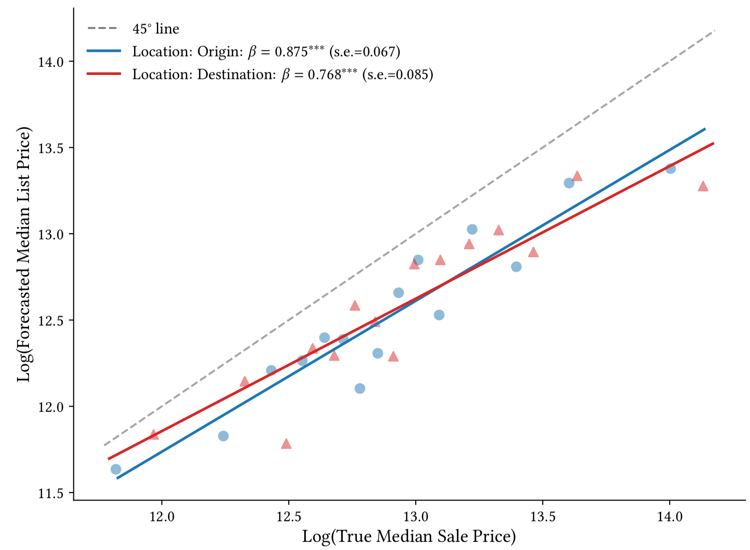


Notes: This figure displays the desired house type among house buyers for the US survey sample. I tabulate answers to the multiple choice survey elicitation *"What type of property are you looking for?/What type of property were you looking for while searching for the house you recently purchased?"*.

Figure A.3: Quality of Price Elicitations: Forecasted Median Housing Price and True Median Housing Price



(a) Log Sale Price



(b) Log List Price

Notes: This figure displays binned scatterplots of respondents' elicited median house prices against true median prices from Redfin, separately for origin and destination ZIP codes. Panel (a) shows log sale prices and Panel (b) shows log list prices. The sample is restricted to respondents included in the main regression sample. "Location: Origin" refers to elicitations for the respondent's current ZIP code, while "Location: Destination" refers to elicitations for alternative ZIP codes. The dashed gray line represents the 45-degree line indicating perfect calibration. Regression coefficients, standard errors, and significance levels are displayed in the legend for each subsample. Standard errors are computed clustering response at the respondent level.

D.2 Expectations Module

Table A.5: First Stage Regressions: Belief about Origin Price Discrepancy

(a) Sale Price

	Forecasted Origin Sale Price - True Destination Sale Price		
	(1)	(2)	(3)
Sale Price Origin - Sale Price Dest.	0.679*** (0.036)	0.688*** (0.056)	0.657*** (0.051)
Demographic Controls	Yes	Yes	Yes
Housing Taste Controls	Yes	Yes	Yes
Elicitations	1334	676	705
Unique Respondents	640	329	331
R-sq.	0.464	0.447	0.469
Sample	All Responses	Incentivized Price Elicit.	Non-Incentivized Price Elicit.

(b) List Price

	Forecasted Origin List Price - True Destination List Price		
	(1)	(2)	(3)
List Price Origin - List Price Dest.	0.640*** (0.036)	0.646*** (0.052)	0.635*** (0.046)
Demographic Controls	Yes	Yes	Yes
Housing Taste Controls	Yes	Yes	Yes
Elicitations	1347	679	714
Unique Respondents	647	330	336
R-sq.	0.475	0.473	0.488
Sample	All Responses	Incentivized Price Elicit.	Non-Incentivized Price Elicit.

Notes: This table reports first-stage regressions for the IV specifications in Table 6. The dependent variable is the respondent's elicited belief about median prices in their origin ZIP code minus the true median price in the destination ZIP code. The instrument is the true median price in the origin ZIP code minus the true median price in the destination ZIP code. Panel (a) uses median *sale*-price elicitations, and Panel (b) uses median *listing*-price elicitations. *Demographic controls* include fixed effects for five-year age bins, gender, marital status, and self-reported income decile. *House-taste controls* include fixed effects for the respondent's preferred type of home and desired number of bedrooms. Each observation corresponds to an individual destination elicitation, and the number of unique respondents contributing to the estimation is reported at the bottom of the table. Standard errors, clustered at the respondent level, are shown in parentheses.

$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.6: Forecast Errors and Price Discrepancy: Robustness

(a) Sale Price

	Forecast Error in Destination Median Sale Price		
	(1)	(2)	(3)
Forecasted Origin Sale Price - True Destination Sale Price	0.385*** (0.038)	0.393*** (0.042)	0.327*** (0.067)
Demographic Controls	Yes	Yes	Yes
Housing Taste Controls	Yes	Yes	Yes
Elicitations	1158	869	465
Unique Respondents	546	402	230
Estimation	2SLS	2SLS	2SLS
First-stage F-stat	915.743	753.067	257.282
R-sq.	0.419	0.417	0.361
Sample	Single Fam. Home Buyer	Attention Check: Pass	Attention Check: Fail

(b) List Price

	Forecast Error in Destination Median List Price		
	(1)	(2)	(3)
Forecasted Origin List Price - True Destination List Price	0.460*** (0.041)	0.481*** (0.043)	0.347*** (0.076)
Demographic Controls	Yes	Yes	Yes
Housing Taste Controls	Yes	Yes	Yes
Elicitations	1158	878	469
Unique Respondents	540	402	227
Estimation	2SLS	2SLS	2SLS
First-stage F-stat	1026.090	798.124	245.439
R-sq.	0.456	0.474	0.385
Sample	Single Fam. Home Buyer	Attention Check: Pass	Attention Check: Fail

Notes: This table reports robustness checks for regressions of forecast errors for respondents' median house price elicitations in destination ZIP codes onto the respondent's elicited belief about median prices in their origin ZIP code minus the true median price in the destination ZIP code. Forecast errors are defined as the respondent's survey response for the destination minus the true price in the destination, where true prices are computed using Redfin data. All prices are measured in U.S. dollars. Panel (a) uses median *sale*-price elicitations, and Panel (b) uses median *listing*-price elicitations. Column (1) restricts the sample to respondents searching for single-family homes. Column (2) restricts the sample to respondents who passed the attention check. Column (3) includes only respondents who failed the attention check.

Demographic controls include fixed effects for five-year age bins, gender, marital status, and self-reported income decile. *House-taste controls* include fixed effects for the respondent's preferred type of home and desired number of bedrooms. Each observation corresponds to an individual destination elicitation, and the number of unique respondents contributing to the estimation is reported at the bottom of the table. All specifications use two-stage least squares (2SLS) estimation, instrumenting the elicited belief about origin prices minus the true destination price with the true origin price minus the true destination price. Standard errors, clustered at the respondent level, are shown in parentheses.

$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.7: Forecast Errors and Price Discrepancy: Log-Log Specification

(a) Sale Price Extrapolation

	Forecast Error in Log Destination Median Sale Price					
	(1)	(2)	(3)	(4)	(5)	(6)
Log(Forecasted Origin Sale Price) - Log(True Destination Sale Price)	0.441^{***}	0.427^{***}	0.420^{***}	0.350^{***}	0.319^{***}	0.360^{***}
	(0.048)	(0.048)	(0.047)	(0.040)	(0.059)	(0.054)
Demographic Controls	No	Yes	Yes	Yes	Yes	Yes
Housing Taste Controls	No	No	Yes	Yes	Yes	Yes
Elicitations	1302	1302	1302	1302	651	651
Unique Respondents	634	634	634	634	320	314
Estimation	OLS	OLS	OLS	2SLS	2SLS	2SLS
First-stage F-stat	-	-	-	537.129	248.634	288.864
R-sq.	0.284	0.310	0.319	0.251	0.226	0.268
Sample	All Responses	All Responses	All Responses	All Responses	Incentivized Price Elicit.	Non-Incentivized Price Elicit.

(b) List Price Extrapolation

	Forecast Error in Log Destination Median List Price					
	(1)	(2)	(3)	(4)	(5)	(6)
Log(Forecasted Origin List Price) - Log(True Destination List Price)	0.517^{***}	0.511^{***}	0.508^{***}	0.339^{***}	0.299^{***}	0.333^{***}
	(0.049)	(0.050)	(0.050)	(0.046)	(0.064)	(0.066)
Demographic Controls	No	Yes	Yes	Yes	Yes	Yes
Housing Taste Controls	No	No	Yes	Yes	Yes	Yes
Elicitations	1304	1304	1304	1304	638	666
Unique Respondents	632	632	632	632	320	312
Estimation	OLS	OLS	OLS	2SLS	2SLS	2SLS
First-stage F-stat	-	-	-	512.050	213.522	303.171
R-sq.	0.352	0.370	0.376	0.292	0.253	0.300
Sample	All Responses	All Responses	All Responses	All Responses	Incentivized Price Elicit.	Non-Incentivized Price Elicit.

Notes: This table reports regressions of log forecast errors for respondents' median house price elicitations in destination ZIP codes onto the log of the respondent's elicited belief about median prices in their origin ZIP code minus the log of the true median price in the destination ZIP code. Forecast errors are defined as the respondent's survey response for the destination minus the true price in the destination, where true prices are computed using Redfin data. All prices are measured in U.S. dollars and transformed to natural logarithms. Panel (a) uses median *sale*-price elicitations, and Panel (b) uses median *listing*-price elicitations. The survey question was worded as follows: *"What is your best guess for the median [sale/listing] price for a single-family home in ZIP code [current elicitation ZIP code] over the past three months?"*.
Columns (1)–(3) report OLS estimates with varying sets of controls. Columns (4)–(6) report two-stage least squares (2SLS) estimates, instrumenting the log of the elicited belief about origin prices minus the log of the true destination price with the log of the true origin price minus the log of the true destination price. *Demographic controls* include fixed effects for five-year age bins, gender, marital status, and self-reported income decile. *House-taste controls* include fixed effects for the respondent's preferred type of home and desired number of bedrooms. Each observation corresponds to an individual destination elicitation, and the number of unique respondents contributing to the estimation is reported at the bottom of the table. Standard errors, clustered at the respondent level, are shown in parentheses.
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

D.3 Hypothetical Choice Module

Survey Design The hypothetical choice module presents respondents with two scenarios featuring different price levels in their main ZIP code of interest. After being shown an introductory screen explaining the exercise, respondents view a realistic house listing accompanied by information about local market conditions. The house listing shown is tailored to match each respondent's stated preferences regarding property type (apartment, townhouse, or single-family home) and desired number of bedrooms.[38] In one scenario, respondents are told that the median single-family home price in the destination ZIP code matches their earlier prediction exactly. In the other scenario, the median price is experimentally manipulated to deviate from their expectation. Specifically, the alternative price is set to be T_i percent higher or lower than the respondent's reported expectation for house prices in the ZIP code, where T_i is randomly drawn from a list ranging from -10 to 10 in 2.5 percent increments. The two scenarios are presented in random order to avoid order effects.

For each scenario, respondents answer up to four questions: whether they would consider making an offer on the displayed property, what initial offer they would make (conditional on answering yes to the first question), their maximum willingness to pay for the property, and their prediction of the property's likely sale price. These responses generate a within-respondent panel: for each individual, we observe choices under both the baseline price scenario (matching expectations) and the alternative price scenario (deviating from expectations).

Main Analysis and Results The module exploits within-respondent variation in stated median prices to test for asymmetries in housing demand responses. If reference dependence shapes housing choices, willingness to pay and offer amounts should respond asymmetrically to price increases versus price decreases around the reference point formed by initial price beliefs. To test this hypothesis, I construct a respondent-scenario panel by stacking each respondent's observations from the two scenarios. For each outcome variable—an indicator equal to unity if the respondent would make an offer amount and the maximum willingness to pay for the housing unit under current market conditions—I compute the first difference across scenarios: $\Delta y_i = y_i^{\text{alt}} - y_i^{\text{base}}$, where y_i^{base} denotes respondent i's response when prices match expectations and y_i^{alt} denotes the response in the alternative price scenario.

I then estimate the following specification:

$$\Delta y_i = \beta^+ \cdot \Delta \log(p_i) \cdot \mathbf{1}[p_i^{\text{alt}} > p_i^{\text{base}}] + \beta^- \cdot \Delta \log(p_i) \cdot \mathbf{1}[p_i^{\text{alt}} \leq p_i^{\text{base}}] + \varepsilon_i,$$

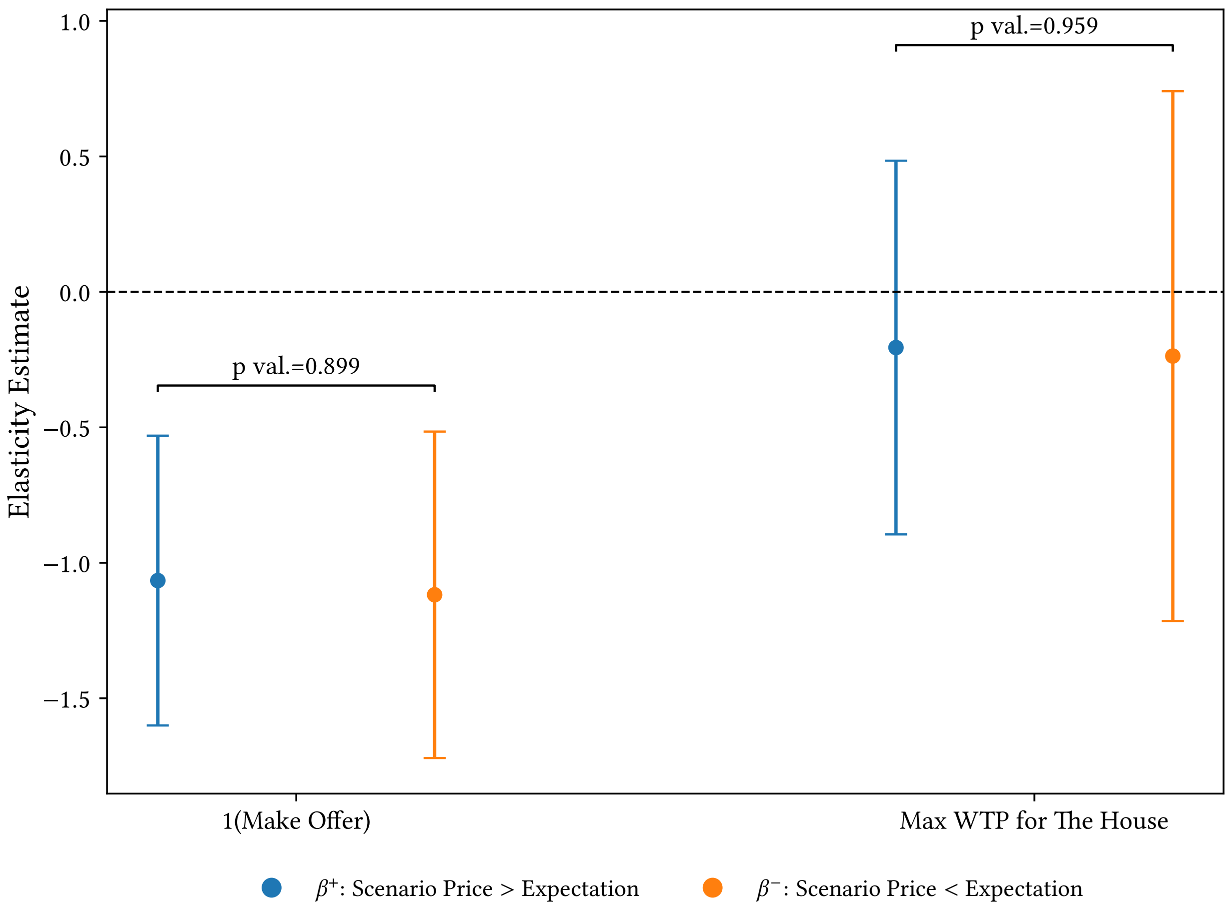
where $\Delta \log(p_i) = \log(p_i^{\text{alt}}) - \log(p_i^{\text{base}})$ is the log difference in median prices across scenar-

[38] Figure A.5 presents an example listing.

ios, with p_i^{alt} denoting the median price in the alternative scenario and p_i^{base} the baseline price matching the respondent's expectation. The specification separately identifies the elasticity of demand with respect to price increases above expectations (β^+) and price decreases below expectations (β^-). Because both the dependent variable and the regressor are in first differences, the specification implicitly controls for all time-invariant respondent characteristics, ensuring that identification comes exclusively from within-respondent variation across the two price scenarios.

Figure A.4 presents coefficient estimates for β^+ and β^- across the four outcomes, along with p-values from tests of the null hypothesis $\beta^+ = \beta^-$. Evidence of asymmetric elasticities—with demand responding more strongly to price increases than to price decreases of equal magnitude—would be consistent with reference dependence in housing demand. The figures shows estimate elasticities in line with economic intuition. The semi-elasticity of the probability of making an offer is roughly minus one, the elasticity of maximal offer is however close to zero. Overall, I conclude that there is no evidence of asymmetric behavior around the candidate reference point.

Figure A.4: Coefficient Plot: Asymmetry of Elasticities



Notes: This figure presents coefficient estimates for the elasticity of survey responses with respect to median price changes in the hypothetical choice module. In this module, respondents were presented with two scenarios showing the same house listing but featuring different median single-family home prices in their destination ZIP code of interest. In one scenario, the stated median price matched the respondent's earlier elicited expectation exactly; in the other scenario, the median price was experimentally manipulated to deviate from the respondent's expectation by a randomly assigned percentage ranging from -10% to $+10\%$ in 2.5% increments. Each set of two coefficients presents estimates for one of the two outcomes: an indicator for whether the respondent would make an offer on the displayed property and the maximum willingness to pay for the house. For each outcome, the figure displays two coefficients: β^+ (labeled "Above Expectations"), which captures the elasticity when the median price in the alternative scenario exceeds the respondent's stated expectation, and β^- (labeled "Below Expectations"), which captures the elasticity when the median price falls below expectations. The specification regresses the first difference in outcomes across scenarios on the first difference in log median prices, separately for price increases and decreases: $\Delta y_i = \beta^+ \cdot \Delta \log(p_i) \cdot \mathbf{1}[p_i^{\text{alt}} > p_i^{\text{base}}] + \beta^- \cdot \Delta \log(p_i) \cdot \mathbf{1}[p_i^{\text{alt}} \leq p_i^{\text{base}}] + \varepsilon_i$. Each point estimate is accompanied by 95% confidence intervals constructed using robust standard errors. The p-value displayed above each set of coefficients tests the null hypothesis $\beta^+ = \beta^-$ against the two-sided alternative, where rejection would indicate asymmetric price responses consistent with reference-dependent preferences.

E Model Derivations

E.1 Regularity Conditions, Existence and Uniqueness of Solutions

We impose the following regularity conditions throughout:

R1 The distribution of seller's waiting costs F has a non-decreasing hazard rate.

R2 Time discounting satisfies $\beta < 1$.

R3 F is continuously differentiable.

R4 G satisfies the conditions of Fubini–Tonelli. That is, when differentiating $\mathbb{E}[V_b^m(\varepsilon)]$ with respect to the model parameters, expectations and differentiation operators can be interchanged.

While assumptions R2–R4 are relatively straightforward, R1 merits further discussion. The requirement that the seller's cost distribution has a non-decreasing hazard rate guarantees that the distribution is thin-tailed enough to behave well at the extremes. In particular, it ensures that the probability of acceptance converges to 1 as the buyer's offer price diverges to $+\infty$, and conversely converges to 0 as the offer price tends to $-\infty$. Intuitively, this rules out situations where the buyer might find it optimal to make extreme offers—either arbitrarily high or arbitrarily low—in the hope of matching with a seller of some "extreme" type. In other words, this assumption guarantees that the buyer's value is single peaked in the current offer. In practice, this assumption is mild: most standard continuous distributions employed in economics satisfy it.[39]

If the distribution of seller waiting costs exhibits a non-decreasing hazard rate, then the buyer's search problem is log-concave. Consequently, the Bellman operator implied by Equation 6 is a contraction, and by Banach's Contraction Theorem we obtain existence and uniqueness of the buyer's value function.

Moreover, under assumptions R1–R4, we can apply the results of Benveniste and Scheinkman (1979) to the seller's search problem. This ensures that the seller's value function is at least once differentiable in all parameters at interior solutions, and that the dynamic analogue of the Envelope Theorem holds. In particular, when differentiating the value function with respect to the parameter vector, one may ignore indirect effects through the optimal policy function and simply take the partial derivative with respect to the parameters.

[39]Specifically, the Uniform (and more generally the Beta distribution family for appropriate scale parameters), Normal, Logistic and Gumbel distribution satisfy this property.

E.2 Comparative Statics in the Rational Problem

Proposition 1. *The policy function p^* for the rational search problem is increasing in ε and the derivative is bounded from above by one.*

Proof. Recall that the optimal price $p^*(\varepsilon)$ satisfies the f.o.c. in Equation 7.

Rearrange and define

$$G(p, \varepsilon) := p - v - \varepsilon + \beta EV^* + \frac{\psi(p)}{\psi'(p)},$$

so that $p^*(\varepsilon)$ solves $G(p^*(\varepsilon), \varepsilon) = 0$.

Compute the partial derivatives to apply the implicit function theorem:

$$G_\varepsilon(p, \varepsilon) = -1, \qquad G_p(p, \varepsilon) = 1 + \frac{(\psi'(p))^2 - \psi(p)\psi''(p)}{(\psi'(p))^2}.$$

where it should be noted that continuation value EV^* is invariant of current ε since the draws are iid over the search periods. Further, by the implicit function theorem,

$$\frac{dp^*(\varepsilon)}{d\varepsilon} = -\frac{G_\varepsilon}{G_p} = \frac{1}{G_p(p^*(\varepsilon), \varepsilon)}.$$

Finally, by log concavity of ψ, $G_p(p, \varepsilon) \geq 1 > 0$ so that

$$0 < \frac{dp^*(\varepsilon)}{d\varepsilon} \leq 1.$$

□

Proposition 2. *The policy function p^* is increasing in the local price level $\bar{p}_d$.*

Proof. Recall that by definition $\psi(p) = 1 - F(p_d - p)$.

Applying Benveniste and Scheinkman (1979), Since $\partial\psi/\partial p_d = -f(p_d - p)$, the envelope gives

$$\frac{dV_m(\varepsilon)}{dp_d} = -\,\psi\big(p^*(\varepsilon)\big) + \beta\,(1 - \psi\big(p^*(\varepsilon)\big))\,H_d, \quad H_d \equiv \frac{\partial EV}{\partial p_d}.$$

Differentiating EV and noting $\partial V_n/\partial p_d = \beta H_d$,

$$H_d = \alpha\Big(-\bar{\psi} + \beta(1 - \bar{\psi})H_d\Big) + (1 - \alpha)\beta H_d = -\alpha\bar{\psi} + \beta(1 - \alpha\bar{\psi})H_d,$$

whence

$$H_d = \frac{-\alpha\,\bar{\psi}}{D}.$$

Thus the pointwise derivative is

$$\frac{dV_m(\varepsilon)}{dp_d} = -\,\psi\big(p^*(\varepsilon)\big) - \beta\left(1 - \psi\big(p^*(\varepsilon)\big)\right)\frac{\alpha\,\bar{\psi}}{D}.$$

Rearrange the FOC to make dependence on p_d explicit

$$G(p, p_d) := p - v - \varepsilon + \beta EV^* + \frac{\psi(p|p_d)}{\psi'(p|p_d)},$$

□

E.3 Figure details

E.3.1 Solving for the Value Function

Throughout, I solve for the value function numerically with a value function iteration algorithm.

Step 0 Set the initial conditions as follows:

- Assume F and G are logistic with mean zero and scale parameters s_F and s_G respectively.
- Set initial guess VF_0 as the solution to the rational search problem with F as defined above, $\beta = 1$ and no iid match specific shocks.
- Set vector of primitives $(m, \beta, c_b, v, \bar{p_d}, s_F, s_G)$
- Draw a large vector of match specific shocks $(\varepsilon_i)_{i=1}^I$ from G

Step $k+1$ Suppose you know EV^{k+1}. Then:

1. Compute by solving optimal prices at current step by numerically solving the FOC give the realization of ε_i and guess EV^k

$$p_i^{k+1} = v + \varepsilon_i - \beta EV^k - \frac{\psi\left(p_i^{k+1}\right)}{\psi'\left(p_i^{k+1}\right)}$$

2. Compute $V_{m,i}^{k+1} = -c_b + \psi(p_i^{k+1})[v + \varepsilon_i - p_i^{k+1}] + (1 - \psi(p_i^{k+1}))\beta VF^k$
3. Compute $EV_m^{k+1} = \frac{1}{I}\sum_{i=1}^I V_{m,i}^{k+1}$
4. Compute $V_n^{k+1} = -c_b + EV^k$
5. $VF^{k+1} = \alpha VF_n^{k+1} + (1-\alpha)EV_m^{k+1}$

Stopping Repeat until $|VF^{k+1} - VF^{k}| \leq \delta$ or the maximum number of iteration is reached.

F Survey Sections Derivations

Proposition 3. *Suppose respondent i's beliefs about house prices at their origin $o(i)$ are given by $F_i(p_{o(i)})$ and observed without survey noise. Let beliefs about destination prices p_d for $d \neq o(i)$ follow*

$$\tilde{F}_i(p_d) = (1-\theta)p_d + \theta F_i(p_{o(i)}) + \tilde{\eta}_{d,i},$$

where $\tilde{\eta}_{d,i}$ is survey noise. Consider the regression

$$\tilde{F}_i(p_d) - p_d = \alpha + \beta\big(F_i(p_{o(i)}) - p_d\big) + u_{i,d}.$$

Then the OLS estimator satisfies $\operatorname{plim}(\hat{\beta}) = \theta$.

Proof. The probability limit of the OLS estimator is

$$\operatorname{plim}\ \hat{\beta} = \frac{\operatorname{Cov}\Big(\tilde{F}_i(p_d) - p_d,\ F_i(p_{o(i)}) - p_d\Big)}{\operatorname{Var}\big(F_i(p_{o(i)}) - p_d\big)}.$$

Because

$$\tilde{F}_i(p_d) - p_d = \theta\big(F_i(p_{o(i)}) - p_d\big) + \tilde{\eta}_{d,i},$$

and the survey noise is mean-independent of beliefs, the covariance term reduces to

$$\theta\ \operatorname{Var}\big(F_i(p_{o(i)}) - p_d\big)\,,$$

yielding the result. □

Proposition 4. *Now suppose that origin beliefs are elicited with noise:*

$$\tilde{F}_i(p_{o(i)}) = F_i(p_{o(i)}) + \eta_{o(i),i},$$

with the decomposition

$$\eta_{o(i),i} = \tilde{\eta}_i + \tilde{\eta}_{o(i),i}, \qquad \eta_{d,i} = \tilde{\eta}_i + \tilde{\eta}_{d,i},$$

where:

- *$\tilde{\eta}_i$ is a respondent-specific noise term common across all elicited beliefs,*
- *$\tilde{\eta}_{o(i),i}$ is idiosyncratic noise in the origin elicitation,*

- $\tilde{\eta}_{d,i}$ *is idiosyncratic noise in the destination elicitation.*

Destination beliefs satisfy

$$\tilde{F}_i(p_d) = (1-\theta)p_d + \theta F_i(p_{o(i)}) + \eta_{d,i}.$$

Consider the regression

$$\tilde{F}_i(p_d) - p_d = \alpha + \beta\big(\tilde{F}_i(p_{o(i)}) - p_d\big) + u_{i,d}.$$

Then the OLS estimator does not *recover the true extrapolation parameter.*

Proof. The OLS probability limit is

$$\text{plim } \hat{\beta} = \frac{\text{Cov}\Big(\tilde{F}_i(p_d) - p_d,\ \tilde{F}_i(p_{o(i)}) - p_d\Big)}{\text{Var}\Big(\tilde{F}_i(p_{o(i)}) - p_d\Big)}.$$

Substituting the measurement equations:

$$\tilde{F}_i(p_d) - p_d = \theta\big(F_i(p_{o(i)}) - p_d\big) + \tilde{\eta}_i + \tilde{\eta}_{d,i},$$

$$\tilde{F}_i(p_{o(i)}) - p_d = \big(F_i(p_{o(i)}) - p_d\big) + \tilde{\eta}_i + \tilde{\eta}_{o,i}.$$

Expanding covariances and using independence of idiosyncratic noise gives

$$\text{plim } \hat{\beta} = \frac{\theta\,\text{Var}\big(F_i(p_{o(i)}) - p_d\big) + \text{Var}(\tilde{\eta}_i)}{\text{Var}\big(F_i(p_{o(i)}) - p_d\big) + \text{Var}(\tilde{\eta}_i) + \text{Var}(\tilde{\eta}_{o,i})}.$$

If $\text{Var}(\tilde{\eta}_i) = 0$, this reduces to classical attenuation bias. With common noise, the estimated coefficient exceeds θ whenever

$$\text{Var}(\tilde{\eta}_i) > \frac{\theta}{1-\theta}\ \text{Var}(\tilde{\eta}_{o,i}).$$

Intuitively, common survey noise raises both the regressor and dependent variable in the same direction, creating a spurious positive correlation that is mistaken for extrapolation. □

Proposition 5. *Suppose the setting is as in Proposition 4, where origin beliefs are measured with noise,*

$$\tilde{F}_i(p_{o(i)}) = F_i(p_{o(i)}) + \tilde{\eta}_i + \tilde{\eta}_{o,i},$$

destination beliefs satisfy

$$\tilde{F}_i(p_d) = (1-\theta)p_d + \theta F_i(p_{o(i)}) + \tilde{\eta}_i + \tilde{\eta}_{d,i},$$

and price discrepancy beliefs are contaminated with common and idiosyncratic noise.

Assume, however, that the researcher observes the true *price difference*

$$Z_i = p_{o(i)} - p_d,$$

which is measured without survey noise. Consider estimating θ using IV:

$$\tilde{F}_i(p_d) - p_d = \alpha + \beta\big(\tilde{F}_i(p_{o(i)}) - p_d\big) + u_{i,d},$$

instrumenting $\tilde{F}_i(p_{o(i)}) - p_d$ with $Z_i = p_{o(i)} - p_d$.

Then the IV estimator recovers the true extrapolation parameter:

$$\operatorname{plim}(\hat{\beta}^{IV}) = \theta.$$

Proof. The IV probability limit is

$$\operatorname{plim}\ \hat{\beta}^{IV} = \frac{\operatorname{Cov}\Big(Z_i,\ \tilde{F}_i(p_d) - p_d\Big)}{\operatorname{Cov}\Big(Z_i,\ \tilde{F}_i(p_{o(i)}) - p_d\Big)}.$$

Substituting the measurement equation for the dependent variable yields

$$\tilde{F}_i(p_d) - p_d = \theta\big(F_i(p_{o(i)}) - p_d\big) + \tilde{\eta}_i + \tilde{\eta}_{d,i}.$$

Since $Z_i = p_{o(i)} - p_d$ is measured without noise and is independent of all survey noise components,

$$\operatorname{Cov}\Big(Z_i, \tilde{F}_i(p_d) - p_d\Big) = \theta\ \operatorname{Cov}\big(Z_i,\ F_i(p_{o(i)}) - p_d\big)\,.$$

For the regressor,

$$\tilde{F}_i(p_{o(i)}) - p_d = \big(F_i(p_{o(i)}) - p_d\big) + \tilde{\eta}_i + \tilde{\eta}_{o,i},$$

and again, because Z_i is noise-free and independent of the survey noise,

$$\operatorname{Cov}\Big(Z_i, \tilde{F}_i(p_{o(i)}) - p_d\Big) = \operatorname{Cov}\big(Z_i,\ F_i(p_{o(i)}) - p_d\big)\,.$$

Taking the ratio of these two expressions yields $\operatorname{plim}\ \hat{\beta}^{IV} = \theta$ which establishes the result. □

G Survey Instrument

This section describes each survey module in detail and reproduces the exact wording of all elicitations as they were presented to respondents.

Verbatim survey text is enclosed in framed boxes. Editorial notes and clarifications to the reader—such as the conditions under which specific questions were displayed—are shown in square brackets and typeset in `monospaced` font.

G.1 Survey Module: Introduction and Consent

Upon clicking the Prolific link, respondents encounter a captcha verification followed by the consent form.

Key Information

You are being invited to participate in a research study about why people decide to move and the houses they buy after moving. Participation in this research is completely voluntary and you may withdraw your consent at any time. The study will take about 10 minutes to complete and pays 2 USD upon completion. Some respondents might receive an additional bonus payment up to 1USD. You will be asked to answer survey questions about your past and future housing choices, personal experiences, and the places where you have lived. Possible risks or discomforts include feeling slightly uncomfortable with some questions if the topic of buying a house is stressful to you; you may stop participating at any time. There is no direct benefit to you from taking part in this study. The results of this research may improve understanding of housing decisions and help inform housing policy.

Consent Form

Introduction and Purpose We are researchers at the University of California, Berkeley in the Department of Economics. The main researcher, Matteo Saccarola, is conducting this research under the supervision of Professor Dmitry Taubinsky. We would like to invite you to take part in our research study, which concerns why people decide to move and the houses they buy after moving.

Procedures If you agree to participate in our research, we will ask you to complete the attached online survey/questionnaire. The survey will involve questions about your past and future housing choices, personal experiences, and the places where you have lived. The survey should take about 10 minutes to complete.

Compensation You will be paid 2 USD if you complete the full survey questionnaire. Payment will be received within 7 days of submission through the Prolific payment system. If you are selected to receive a bonus payment, it will be paid at the same time as the completion payment. Maximum bonus payment is 1USD.

Benefits There is no direct benefit to you from taking part in this study. It is hoped that the research will contribute to a better understanding of housing decisions, which may inform housing policy and planning in the future.

Risks/Discomforts We do not anticipate any significant risks from participating in this study. You may feel mild discomfort if the topic of buying a house is stressful to you. You are free to withdraw your participation at any time.

Confidentiality Your study data will be handled as confidentially as possible. Your data will be anonymous and will not be linked to your identity. We will use your Prolific ID to pay you and we will then strip the data of this information. When the research is completed, we may save the anonymous data for possible use in future research conducted by us or others. The same measures described above will be taken to protect confidentiality of this study data. Your personal information may be released if required by law. Authorized representatives from the University of California may review your research data for purposes such as monitoring or managing the conduct of this study.

Rights Participation in research is completely voluntary. You are free to decline to take part in the project. You can decline to answer any question. You are free to stop taking part in the project at any time. Whether or not you choose to participate, to answer any particular question, or continue participating in the project, there will be no penalty to you or loss of benefits to which you are otherwise entitled.

Questions If you have any questions about this research, please feel free to contact the research team at berkeleyhousingsurvey@gmail.com. If you have any questions about your rights or treatment as a research participant in this study, please contact the University of California at Berkeley's Committee for Protection of Human Subjects at 510-642-7461, or e-mail subjects@berkeley.edu.

Agreement to Participate Please answer below to indicate that you have read this consent form and that you voluntarily agree to participate in the study. You may print a copy of this page for your record.

- I read the consent form and agree to take part in this research.
- Before we begin, please enter your Prolific ID below. Note that this response should

auto-fill with the correct ID.

Screened Out: Consent Refusal Participants who decline consent are screened out and shown the following message:

Thank you for your interest in this research. Your informed consent is necessary for participation. Please return this submission to Prolific.

Prolific ID Entry Participants who consent proceed to enter their Prolific ID.

Before we begin, please enter your Prolific ID below. Note that this response should auto-fill with the correct ID.

`[Text Box Entry]`

G.2 Housing Search Stage

Current Search Stage Respondents first indicate their current stage in the house search process using the following multiple-choice question.

Are you currently in the process of buying a new house?

- I am **not** interested in buying a house at this time.
- I am considering buying a house but have not started the process.
- I am actively looking for a house to buy.
- I have made an offer on a house.
- I am in the process of closing on a house.
- I have recently purchased a house.

Screened Out: Not Searching for a House Participants who indicate they are not searching for a house are screened out and receive the following message.

> You are ineligible for this study as you have provided information which is inconsistent with your Prolific prescreening responses.
>
> **Please return your submission on Prolific by selecting the 'stop without completing' button**

Search Length Elicitation Respondents at the "I am considering buying" stage are asked whether they have started looking at listings.

> Have you started looking at home listings?
>
> - Yes
> - No

Respondents who are actively looking at listings, or further along in the process but have not yet purchased, are asked how long ago they began their search:

> Approximately how long ago did you start looking at home listings? *Please answer in months*
>
> `[Text/Number Box Entry]`

Respondents who have already purchased a home receive the following alternative elicitation:

> How long did it take you to buy your home?
>
> *Please count the number of months from when you first started browsing listings until you officially closed on the purchase*
>
> `[Text/Number Box Entry]`

G.3 Destinations

This module elicits up to three candidate destination ZIP codes from each respondent.

Please list up to three U.S. ZIP codes you are currently considering for buying a house. If you list more than one, put them in order from most likely to least likely based on your current preference.

Please only use standard 5-digit US ZIP codes. If you need help in locating the ZIP codes you can look up ZIP codes by searching for addresses near the areas you are interested in on Google Maps or by looking at ZIP-code maps on websites such as www.unitedstateszipcodes.org.

ZIP code 1 [`Text/Number Box Entry`]
ZIP code 2 [`Text/Number Box Entry`]
ZIP code 3 [`Text/Number Box Entry`]

For respondents who have already purchased a house, the first paragraph is revised as follows:

Please list up to three U.S. ZIP codes you were considering for buying a house while you were looking for your current one. If you listed more than one, put them in order from most likely to least likely based on your preference at that time.

G.4 Current Location

This module elicits the respondent's current location or previous location if the respondent has already purchased a home and moved.

Where do you currently live?
City: [`Text/Number Box Entry`]
State: [`Text/Number Box Entry`]
ZIP code (5 digits): [`Text/Number Box Entry`]

In the following screen, respondents report their tenure at the current location.

How many years have you spent living in that ZIP code?
Please enter a number in years.
[`Text/Number Box Entry`]

For respondents who have already purchased a house, the wording and tenses are adjusted for consistency, but the elicitations remain substantively the same.

G.5 Housing Price Beliefs Elicitations

What is your best guess for the **median listing price** for a single-family home in ZIP code `[zipcode]` over the past three months?

The listing price is the price you see in online ads or real estate listings, as advertised by the seller. Please enter an answer in whole US dollars.

`[Text/Number Box Entry]`

What is your best guess for the **median sale price** for a single-family home in ZIP code `[zipcode]` over the past three months?

The sale price is the final price paid by the buyer to the seller. Please enter an answer in whole US dollars.

`[Text/Number Box Entry]`

`[This information is only shown if the respondent is randomized to the "Incentivized" condition]`

Bonus Payment Information: You may earn an additional bonus of up to $1 USD. The bonus is higher the closer your guess is to the truth. To maximize your payment, you should simply state your best guess. Tick the box below to see the equations used to determine your bonus.

[ ] Show the additional information

Remember: If you need help in locating the ZIP codes you can look up ZIP codes by searching for addresses near the areas you are interested in on Google Maps or by looking at ZIP-code maps on websites such as Zillow or www.unitedstateszipcodes.org.

Respondents who select the box for additional information receive the following explanation on the same screen.

At the end of the survey, one of the questions about home prices in a ZIP code you answered will be randomly selected for payment. For that question, we will compare your answer to the true median home price in that ZIP code (measured in thousands of U.S. dollars). Your bonus depends on how close your answer is to the true value. It is calculated as follows:

$$1 - (1/2500) * (\text{True price in 1000 USD} - \text{Your answer in 1000 USD})^2$$

If your answer is exactly correct, you earn the full $1.00 bonus. If you are off by $10,000, your bonus is about $0.96. If you are off by $50,000 or more, you receive no bonus. The bonus can never be negative.

Important: Because each ZIP code question can be chosen for payment, your best strategy is simply to give your most accurate guess for every question.

G.6 House Type Module

Respondents first indicate the property type they are seeking.

What type of property are you looking for?

- apartment or condo
- townhouse
- single-family home
- estate, ranch, or other very large property
- other

In a follow-up screen, respondents provide additional details about the property's specific characteristics. Only characteristics relevant to the property type selected in the previous question are displayed.

Please provide more details about the property you are looking for by answering the questions below

Size of interior living space — square feet

`[Text/Number Box Entry]`

Number of bedrooms

- 0 (Studio)
- 1
- 2
- 3
- 4
- 5 or more

Number of bathrooms (use .5 to indicate half baths, if any)

- 1
- 1.5
- 2
- 2.5
- 3
- 3.5
- 4
- 4.5
- 5 or more

`[This question is only shown if property type is "Apartment" or "Townhouse"]`

Floor number of the unit

- Ground floor
- 1st
- 2nd
- 3rd

- 4th
- 5th floor or higher

[This question is only shown if property type is "Apartment" or "Townhouse"]
Elevator to access floor

- Yes
- Not sure
- No

[This question is only shown if property type is "Single Family Home" or "Very Large Property"]
Approximate lot (yard) size — square feet

[Text/Number Box Entry]

Garage and/or parking spaces

- No Parking Space
- 1 car
- 2 cars
- 3 or more cars

[This question is only shown if property type is "Other"]
Please, briefly describe key features of the ideal property to you

[Text/Number Box Entry]

Is there any other property characteristic that you feel is important to your house choice but was not covered in the previous questions? Please describe below

[Text/Number Box Entry]

G.7 Purchase Plans Module

Why are you interested in purchasing a home?

- Primary residence for self
- Primary residence for family member (e.g. adult child, parent)
- Vacation home or second residence
- An investment property intended for rental income or resale
- Other

How familiar are you with the home-buying/selling process?

- New to the process
- Somewhat familiar
- Moderately familiar
- Very familiar
- Expert/industry professional

Including yourself, how many people will live full-time in the home you are about to purchase?

- 1 (just you)
- 2
- 3
- 4
- 5 or more

G.8 Hypothetical Choice Module

In this subsection, I describe the elicitations from the hypothetical choice module. Respondents were first presented with an introductory screen, followed by two scenarios in which I varied housing prices in the main ZIP code of interest. In one scenario, respondents were told that prices in the destination ZIP code were exactly as they had predicted. In the other scenario, I randomly assigned an alternative price level for the destination ZIP code. The alternative price was set to $(1 + T_i/100) \times \widehat{p}_i$, where $\widehat{p}_i$ denotes respondent i's expected price for that ZIP code and T_i is randomly drawn from $\{-10, -7.5, -5, -2.5, 2.5, 5, 7.5, 10\}$. The two scenarios were shown in random order to avoid potential order effects.

The module concluded with an unannounced attention check asking respondents to recall which of the two scenarios featured higher prices. This attention check was not tied to payment and is used solely to classify high-quality responses for robustness analyses.

Introductory Screen The module opens with an introductory screen describing the content of the module.

> In the next few screens, we will ask you about your choices in a few different housing market scenarios. You'll be presented with hypothetical listings and asked how you might respond under various market conditions. There are no right or wrong answers, we're simply interested in your opinions and decision-making.
>
> In each case, you will first be presented with a housing market scenario. Then, you will answer several questions about how you would respond under those conditions.

Scenario 1: Prices In Line with Expectations The screen presents a scenario followed by up to four elicitations.

> Imagine you are currently searching for a home in ZIP code `[Main ZIP of Interest for respondent]`. You've been working with a real estate agent to explore options in the area. Your agent has just sent you a new listing that they think could be a good fit.
>
> **Market Conditions:** The local housing market in ZIP code `[Main ZIP of Interest for respondent]` is such that the median single-family home sells for about `[Previous Response to Median Price in ZIP]` USD.

Note that your agent is showing you a house that they think would be a good for you, not necessarily a median house in the ZIP code.

Please review the listing carefully and answer the following questions as if you were making real decisions in this situation.

`[Here, the respondent sees the listing of a house with the correct house type (apartment, single-family home, townhouse) and with the number of bedrooms they are looking for. Refer to Figure A.5 for an example]`

Would you consider making an offer on this house under these market conditions?

- Yes
- No

`[ This elicitation is only shown if the respondents answered "Yes" to wanting to make an offer]`

What offer would you be willing to make for this house under these market conditions?

Please enter your answer in whole US dollars.

`[Text/Number Box Entry]`

Suppose the seller made you an offer to purchase this house. What is the highest price you would be willing to pay for this house under these market conditions?

Please enter your answer in whole US dollars.

`[Text/Number Box Entry]`

If this house were listed on the market under these same conditions, how much do you think it would sell for?

Please enter your answer in whole US dollars.

`[Text/Number Box Entry]`

Scenario 2: Alternative Price Scenario The alternative price scenario is identical to the previous one except that the price is adjusted according to the experimental manipulation introduced above. The screen otherwise contains the same listing and elicitations.

> Imagine you are currently searching for a home in ZIP code `[Main ZIP of Interest for respondent]`. You've been working with a real estate agent to explore options in the area. Your agent has just sent you a new listing that they think could be a good fit.
>
> Market Conditions: The local housing market in ZIP code `[Main ZIP of Interest for respondent]` is such that the median single-family home sells for about `[Treatment Price, see beginning of this subsection]` USD.

Attention Check After completing both scenarios, respondents proceed to a separate screen where they are asked to recall which scenario featured higher house prices.

> What was the relationship between the median single-family home price in the first and second scenarios?
>
> - The price in Scenario 1 was higher than in Scenario 2
> - The price in Scenario 1 was the same as in Scenario 2
> - The price in Scenario 1 was lower than in Scenario 2

G.9 Income and Wealth

In this subsection, I reproduce the elicitations on household income and wealth. For brevity, I report only the wording shown to respondents who had not yet purchased a home. The corresponding version for respondents who had already bought a house is analogous, except that all elicitations referring to the previous home and its sale are phrased in the past tense and refer to the period immediately preceding the move.

Income Elicitations The module begins by eliciting current and future expected household income.

On this screen we will ask you a couple of question about your household income. Please enter answers in whole US dollars.

What is your best guess of your total pre-tax household income for 2024?

`[Text/Number Box Entry]`

Looking ahead five years, what is your best guess of your total pre-tax household income for 2030?

`[Text/Number Box Entry]`

Wealth Elicitation: Introductory Screen Before eliciting wealth, respondents answer preliminary questions to tailor the subsequent screen.

Do you own your current home, rent, or live under some other arrangement?

- Own
- Rent
- Other arrangement

`[This question is only shown if response "Own" is chosen in the previous question]`

Do you currently have any outstanding mortgages or loans on this property?

- Yes
- No

`[This question is only shown if "Yes" to outstanding mortgage question]`

What is the total amount of outstanding mortgage or other loans on your home? Please enter an answer in whole US dollars.

`[Text/Number Box Entry]`

[This question is only shown if response "Own" is chosen to initial question on this screen]

Are you planning to sell this property to purchase your next one?

- Yes
- No

[This question is only shown if response "Yes" is chosen for planning to sell question]

What sale price do you expect your current home to fetch? Please enter an answer in whole US dollars without considering any mortgages, taxes, and fees.

[Text/Number Box Entry]

[This question is only shown if response "Yes" is chosen for planning to sell question]

Approximately, how much do you expect to net from the sale after paying off mortgages, taxes, and fees? Please enter an answer in whole US dollars.

[Text/Number Box Entry]

Wealth Elicitation: Main Screen The survey then elicits wealth components for a subset of asset and liability categories.

On this screen we will ask you to think about your assets and debt. For every category, please provide your best estimate of its value at the present moment. Please enter answers in whole US dollars.

What is your best guess about the value of your household retirement savings? Examples of retirement savings: 401(k) account balance, other retirement accounts such as IRA, pensions, annuities

[Text/Number Box Entry]

What is your best guess about the value of your household's other financial assets, excluding retirement savings? Examples of financial assets: bank account balances, non-retirement financial investments, investments in financial apps, cryptocurrency wallets, savings for a rainy day

[Text/Number Box Entry]

What is your best guess about the total value of your non-mortgage household debt? Please exclude mortgages since we already considered them in previous questions. Examples of non mortgage debt: credit card balances, personal loans, payday loans

[Text/Number Box Entry]

G.10 Demographics Module

The demographics module is presented on a single page and contains the following elicitations.

How old are you in years?

[Text/Number Box Entry]

What is your gender?

- Male
- Female
- Non-binary / third gender
- Other
- Prefer not to answer

Which of the following best describes your marital status?

- Single
- Married
- Domestic partnership / Civil union
- Separated
- Divorced
- Widowed

Please select the highest level of education that you have completed.

- Less than high school
- High school diploma or GED
- Some college, no degree
- Associate degree (AA/AS)
- Bachelor's degree (BA/BS)
- Master's degree (MA/MS)
- Professional degree (e.g., JD, MD, DVM)
- Doctorate (PhD, EdD, etc.)

Does any of the following describes your primary occupation? If more than one applies, please select the most important one.

- Financial analyst, banker or investment professional
- Real estate agent
- Mortgage loan officer, mortgage broker, or similar profession
- Real estate appraiser
- Real estate attorney, title officer, or escrow officer
- Construction Contractor
- Architect or engineer
- Property manager
- Real estate investor, developer, or private-equity real estate professional
- CPA, accountant, or tax professional
- City planner, zoning official, or property assessor
- Housing data analyst or economist
- None of the above

Figure A.5: Example of House Listing: One Bedroom Apartment

Notes: This figure presents an example of a listing as it appeared to respondents. All images were generated using the GPT–5 image generation engine. Each respondent was shown a listing for a unit that matched the house type and number of bedrooms they were searching for.