Aging and Housing Returns*

Natee Amornsiripanitch
Federal Reserve Bank of Philadelphia

Philip E. Strahan
Boston College and NBER

Song Zhang
University of Delaware

March 16, 2026

Abstract

Older home sellers receive lower returns than younger home sellers. Homes sold by older people have fewer major renovations but higher rates of poor upkeep. Older sellers are also more likely to sell off-MLS ("pocket listings") and to sell to investors, leading to lower prices. These patterns suggest that older sellers may be disproportionately disadvantaged by agents' incentive to maximize fees through generating high sales volume instead of maximizing sale prices. Age-related cognitive decline makes the elderly more vulnerable. For causal evidence, we show that reforms making pocket listings more transparent reduced both the prevalence of pocket listings and the magnitude of the age gap in returns.

Keywords: Aging, housing returns, incentive misalignment

JEL Classification: G5, J1, R00

*This Philadelphia Fed working paper represents preliminary research that is being circulated for discussion purposes. The views expressed in this paper are solely those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Philadelphia, the Federal Reserve Board, or the Federal Reserve System. Any errors or omissions are the responsibility of the authors. Artificial intelligence was used to assist the authors with copy editing the paper and identifying keywords for the textual analysis.

1 Introduction

For most Americans, the vast majority of their wealth accumulation comes from homeownership and home appreciation (Benetton et al., 2022; Sodini et al., 2023; Artigue et al., 2025). Public policy choices such as interest deductibility and government subsidies of housing finance from the government-sponsored enterprises (GSEs), Fannie Mae and Freddie Mac, have been designed to encourage homeownership. Buying and selling homes therefore constitute the most significant set of financial decisions for most people in determining their wealth. This paper provides the first comprehensive empirical analysis of how home-seller returns vary over the life cycle, with a particular focus on elderly homeowners.

This focus is increasingly relevant as the U.S. housing market undergoes a profound demographic shift. According to Freddie Mac, Baby Boomers—who now make up 38% of homeowners and are all over 60 years old jointly hold $17.3 trillion in housing wealth, accounting for half of the nation's home equity.[1] As this generation begins to exit homeownership, sales by elderly homeowners will comprise a growing share of residential real estate transactions—a trend often referred to as the "silver tsunami." As illustrated in Figure 1, this generational turnover is expected to reduce homeownership by 9.2 million units as aging Baby Boomers sell off their homes. Several news reports have suggested that senior home sellers are often in vulnerable positions when they participate in the real estate market.[2] Therefore, the economic implications of a large age effect in housing returns are significant for issues related to well-being in retirement and inter-generational wealth accumulation.

Studying how age affects housing returns is challenging due to limited data on seller demographics in standard housing datasets. Prior research (Davidoff, 2004; Rodda and Patrabansh, 2007) suggests that older homeowners tend to earn lower returns, potentially due to poor upkeep, but these studies often rely on self-reported home values, which are prone to measurement error. This concern is especially salient for elderly respondents, who may face cognitive decline that impairs financial assessment (Mazzonna and Peracchi, 2024). We address these limitations by constructing a new dataset that links housing transactions to voter registration records to obtain seller age, enabling a more accurate and comprehensive analysis of age-related disparities in home-sale returns.

[1] See "U.S. Economic, Housing and Mortgage Market Outlook," *Freddie Mac,* February 2024; "2024 Baby Boomer Consumer Research," *Freddie Mac,* 2024.

[2] See "Explaining the Rising Trend of Elder Abuse in Real Estate," *CBS 8 San Diego Interview*, March 2023.

We validate our dataset's representativeness by comparing the age distribution of identified sellers in our sample to estimates from the Zillow Seller Survey data and find close alignment. With this validation in place, we proceed to the empirical analysis, where we document three core findings. First, we establish a robust and large negative effect of seller age on property returns. Older sellers receive lower returns than younger sellers, even after absorbing for buy time, sell time, and market-level effects. The magnitude of the impact of age, which begins to emerge around 70 years old, is comparable to other demographic effects such as gender and race, which have been identified as being associated with housing returns by prior research (Kermani and Wong, 2021; Goldsmith-Pinkham and Shue, 2023). Figures 2, 3, and 4 establish the benchmark relationship between age and property returns, first over the life cycle (Figure 2), then state by state (Figure 3), and finally by sale year (Figure 4). Figure 2 shows a flat profile for sellers between the ages of 40 and 70, with a decline starting at age 70 and increasing in magnitude with each additional year of seller age. The pattern is striking and suggests a very large age gap for older sellers. For example, an 80-year-old seller would earn about 0.5% less per year after a home sale relative to a 45-year-old (the omitted group); this effect corresponds to a lower sales price of more than 5% for a home with the mean holding period (11 years). Figure 3 reports the average over-75 return discount, using the same basic fixed effects framework but with less granular age bins (one for each decade), on a state-by-state basis. We see a deviation from the general pattern in just three small states (Hawaii, Indiana and Maine).[3] Figure 4 reports the corresponding over-75 return discount estimated sale year by sale year. The age effects are pervasive across the country and across the sample period.

Second, after establishing the basic facts, we explore potential explanations, such as differences in holding period, property maintenance and renovation differences, and the need to sell quickly, all of which may plausibly be correlated with seller's age. Neither the holding period nor the need to sell quickly explains the age gap. Since we focus on unlevered returns, differences in borrowing or refinancing behaviors also do not explain the results. Using textual analysis of listing descriptions, however, we show that older sellers are less likely to sell updated or renovated properties and are more likely to sell properties with poor upkeep or requiring substantial maintenance. These factors explain about 25% of the age gap in returns.

Third, we explore the role of the way in which properties are marketed and sold. Because most

[3]We exclude non-disclosure states from the analysis because, for these states, CoreLogic reports imputed prices instead of actual transaction prices. The states are Alaska, Idaho, Kansas, Louisiana, Mississippi, Missouri, Montana, New Mexico, North Dakota, Texas, Utah, and Wyoming.

people buy and sell homes very infrequently, the role of real estate professionals is paramount. Agents, who facilitate real estate transactions every year, have more information and experience than the typical seller. Agents also have different incentives. For example, they may prefer quicker sales to minimize their time marketing the property, thus allowing more time to earn fees on future transactions (Levitt and Syverson, 2008). Also, agents can earn more fees if they represent both buyer and seller. The practice is called "dual agency." We argue that older sellers are less likely to consider such incentive misalignment and information asymmetry when dealing with real estate agents. Consistent with this argument, Mazzonna and Peracchi (2024) find that people are unaware of their cognitive decline as they age, and that this decline leads to more financial mistakes and larger declines in financial wealth for older people. The age gap that we find, which begins around age 70 and increases thereafter, coincides with the age at which cognitive decline starts to become significant (Hale et al., 2020).[4]

To focus on the role of real estate agents, we start by showing that older sellers are much more likely to list their property privately, meaning without using one of the Multiple Listing Services (MLS). When properties are sold privately with an agent (off-MLS), they are known in the real estate parlance as "pocket listings."[5] These transactions receive substantially lower returns, and this pocket-listing discount is much larger for older sellers than for younger ones. As we show, the over-75 age gap in housing returns rises from 0.50% for homes sold through the MLS to 0.74% for homes sold via pocket listings. This finding helps explain a substantial portion of the baseline aging effects.

Pocket listings bring some non-monetary advantages to sellers (e.g., less time on market and fewer showings), but they also make sellers more vulnerable to agents' non-profit-maximizing actions due to their lack of transparency. For example, real estate agents may want to bring a connected buyer–such as a professional investor or developer with whom they have a pre-existing relationship–to the seller in order to receive higher fees via dual agency. Consistent with this explanation, we show that older sellers are also more likely to sell to investors than younger sellers, and like the result for pocket listings, they also receive especially low returns in these transactions. Combining these two, properties sold off-MLS to investors receive returns which are about 0.86% lower when the seller is over 75 years old, compared

[4] DeLiema et al. (2020) also find more financial fraud against the elderly in the Latino community, and Carlin et al. (2023) show self-policing mechanisms in the financial industry help to reduce the impact of financial fraud against the elderly.

[5] A pocket listing refers to a property for sale that is not publicly listed on the MLS (i.e., an off-MLS sale). In practice, the terms pocket listing and private listing mostly refer to the same concept. We use the term pocket listing throughout the paper to describe off-MLS sales.

to sellers in their forties. In contrast, the age discount is just 0.46% for properties sold on the MLS to non-investors.

While the patterns observed are consistent with older sellers facing greater challenges in protecting their financial interests in real estate transactions, they can also be explained by other factors that may make pocket listings attractive, such as the desire to limit the number of people inspecting the property. Older homeowners report multiple reasons for selling, ranked from most to least important as: moving closer to friends or family, downsizing, difficulty maintaining the home due to health or financial constraints, a decline in neighborhood desirability, and changes in family circumstances.[6] These motivations tend to differ from those reported by younger cohorts, raising the concern that differences in unobservable preferences may drive older sellers' higher likelihood to list privately, as opposed to real estate agents' incentive misalignment, and lower realized financial returns. To establish causality, we exploit a transparency-enhancing policy change implemented by the Midwest Real Estate Data (MRED), the Multiple Listing Service for Illinois. This change, which we describe in detail below, led to fewer pocket listings across all sellers in Illinois. And, the return discount experienced by older sellers fell by more than half, from -0.83% before the policy to -0.35% after the policy change.

The decline from the policy is similar in magnitude to the wedge between an older seller listing on the MLS and selling to a non-investor buyer versus one selling off-MLS to an investor. Moreover, the change in pricing effects show no pre-policy trend. Together, these results support a causal effect on outcomes, whereby the policy change protected older sellers from agents' incentive misalignment by making pocket listings more transparent. Specifically, the policy reduced the frequency with which agents act against the interest of their clients by enhancing transparency of pocket listings. This change was most beneficial for older sellers, who are more vulnerable than younger ones due to cognitive decline and, as we show, had been most likely to suffer losses from pocket listings before the change.

Overall we are able to explain the age gap in part from differences in renovation/maintenance and in part from differences in the use of pocket listings, which our policy shock suggests reflects larger economic costs from incentive misalignment when sellers are elderly. But neither mechanism fully accounts for the age discount. We have also tested whether older sellers are more likely to sell urgently, perhaps due to a sudden health shock. However, we find no difference in the time on market for older sellers nor do we

[6] See "2024 Home Buyers and Sellers Generational Trends Report," *National Association of Realtors Research Group*, 2024.

find that they are more likely to be deemed a "motivated seller" in marketing materials.

Previous papers offer some evidence that houses owned by older (75 years or older) homeowners experience lower returns (Davidoff, 2004; Rodda and Patrabansh, 2007; Ong, 2009). These papers, however, rely on data from surveys that ask homeowners what they think their home is worth. Davidoff (2004) argues that the low returns reflect poor maintenance, and Begley and Lambie-Hanson (2015) provide more recent evidence of the correlation between age and poor maintenance. Through a comprehensive study of forced sales in Boston, Campbell et al. (2011) also provide some suggestive evidence of this mechanism. However, survey-based approaches can introduce significant measurement error. For instance, Mazzonna and Peracchi (2024) show that as individuals age, they often experience cognitive decline, which can impair their ability to accurately estimate financial wealth. Like these papers, we find that poor upkeep and lack of renovation help explain the age gap. Unlike these papers, we find that the way in which older sellers choose to structure their home sale (on- versus off-MLS), how they interact with real estate agents (single versus dual listing agent), and whom they choose to sell to (homeowners versus investors) are important determinants of the transaction's financial return.

Ours is the first paper to build a comprehensive dataset containing home transactions and seller age information over a long period of time (1998–2022) for a representative sample of sellers. Data scarcity has limited our understanding of the effects of age on real estate transactions because standard datasets from CoreLogic and the Home Mortgage Disclosure Act (HMDA)—despite being widely used in the literature—historically did not include information on seller age, with HMDA only beginning to report borrower age after 2017 (Amornsiripanitch, 2023). We fill this gap by merging CoreLogic's Deeds database on home transactions with the age of the seller from voter registration data. We are able to identify seller age for roughly half of all arm's length transactions in CoreLogic, with the coverage increasing over time. We focus our attention on arm's length transactions, which is important because older sellers are more likely to transfer properties to family members or for estate-planning purposes, compared to younger ones. Because the voter registration data reflect primary residential addresses, our sample is, by construction, restricted to owner-occupied properties.

Beyond age, we contribute to the literature on differential returns in housing across demographic groups. Goldsmith-Pinkham and Shue (2023) document that single women tend to underperform single men, mainly due to poor market timing (i.e., buying high and selling low). Kermani and Wong (2021)

find that Black and Hispanic homeowners receive lower returns due to higher likelihood of foreclosure. Using a total return approach, Diamond and Diamond (2024) argue that Black and Hispanic homeowners "earn higher but more volatile rates of return than White homeowners, due in part to higher rental yields. These differences are largely explained by household income and education differences." More generally, our results imply that the homeowner's age is an important determinant of the realized financial return of housing as an asset class (Bracke, 2015; Chambers et al., 2021; Amaral et al., 2021; Eichholtz et al., 2021; Demers and Eisfeldt, 2022; Halket et al., 2023; Colonnello et al., 2024).

Last, we contribute to the literature on adverse interactions between professionals and their clients (Egan et al., 2019) by showing evidence that incentive misalignment between real estate agents and their clients drives a big part of the age gap in housing returns. Gilbukh and Goldsmith-Pinkham (2024) document that real estate agent quality matters a lot for home sale outcomes. Levitt and Syverson (2008) show that, due to compensation structure, real estate agents tend to sell clients' homes too quickly and too cheaply. Johnson et al. (2015) find that dual agency is associated with a 6.35% price premium on agent-owned properties, but a 25.10% price discount on government-owned properties and a 5.14% discount on bank-owned properties. Agarwal et al. (2019) show that agents bought their own houses at prices that are 2.54% lower than comparable houses bought by other buyers, which suggests that agents have an informational advantage over non-agents. Our paper is the first to show that age is an important determinant of the ultimate financial result when people participate in the real estate market.

2 Data

2.1 Sources

To build our sample, we start with the CoreLogic deeds database, a national repository of real estate transaction information. CoreLogic sources its data by aggregating public deed records from over 3,000 county clerk and recorder offices across the United States. Coverage can exceed 50 years in some counties. The database contains property characteristics (e.g., address, land use, lot size), information on the transaction (e.g., sale date, sale price, deed type, arm's length flag, cash sale flag), and owner informa-

tion (e.g., buyer name, seller name, corporate buyer/seller flag). Ownership-level information typically includes only the names of buyers and sellers, along with the mailing address of the buyer. Additionally, it includes unique property and transaction identifiers, such as the Assessor's Parcel Number (APN) and the CoreLogic Integrated Property Number (CLIP), which facilitates the linking of records across time. Our CoreLogic deeds dataset ends in 2022.

We also use CoreLogic's Multiple Listing Service (MLS) database. The data contain information on the subset of transactions marketed through the MLS, and include details such as the asking price, original listing date, additional property characteristics (e.g., square footage, number of bedrooms and bathrooms), textual descriptions of the listing, and information on the brokerages and agents involved. Same as the deeds dataset, our MLS data are available through 2022. We utilize the MLS listing data to assess whether a house transaction is done through the MLS listing or through a pocket listing.

To assign seller age to transactions, we merge individual-level data from the L2 voter registration database into the CoreLogic deeds data. L2 builds a national file containing the records of registered voters across the United States. The company compiles these data by acquiring official voter registration lists from state and county election officials and standardizing them into a uniform format. L2 provides extensive nationwide coverage and includes a detailed record for each individual, which contains their full name, address, birth date, party registration, and voting participation record in specific federal, state, and local elections.

The L2 data include two types of voter registration records: (1) formatted voter registration data spanning 2014 to 2024 for all fifty states and the District of Columbia, and (2) raw, unformatted voter files available for a more limited set of states. The exact coverage of the raw data varies by state and generally correlates with state population size. For larger states—such as California, Florida, Illinois, and North Carolina—coverage extends back to the early 2000s. In contrast, smaller states typically have coverage beginning around 2010. From the unformatted dataset, we extract and standardize individual name, birth year, and address information from the raw data and integrate it with the formatted L2 records to maximize sample coverage. In the Appendix Table B.1, we report the years in which data are available in each state.

2.2 Matching CoreLogic to L2 Voter Registration Data

To obtain information on seller age, we link individuals in the CoreLogic transaction data to the L2 voter registration records using property owner's name, residential address, and ZIP code. Specifically, we rely on exact matches after standardizing variables across the two datasets. Because L2 data only cover registered U.S. citizens and primary residences, our matched sample is composed of owner-occupied housing transactions by sellers who are U.S. citizens.

Overall, we successfully match approximately 51 million unique transactions, which represent approximately 40% of all CoreLogic transactions in the state-by-year cells that our L2 dataset covers. Several factors help explain why a substantial portion of transactions remain unmatched. First, our matching algorithm is intentionally conservative. We use exact matching in all stages of the name-matching procedure. Second, the L2 data exclude non-citizens, so housing transactions involving non-citizens are omitted. Third, L2 does not capture individuals who are not registered to vote. According to U.S. Census estimates, only 65 to 75% of adult citizens are registered to vote in a given year.[7] Fourth, the voter file is restricted to individuals' primary residential addresses and so transactions involving secondary or investment properties are excluded. For a detailed description of the matching procedure, see Appendix B.

Table 1 reports the number of transactions by year, separated by whether or not we are able to match the transaction with seller age information from L2. The matching frequency increases over time, starting from less than 5% in the late 1990s and rising to nearly 50% by early 2021. In the early years, match rates are low due to limited overlap between the voter registration data and the CoreLogic transaction records. Both datasets' coverage increases over time such that, by 2010, they are capturing all states.

Table 1 also reports mean and median home prices for our matched sample compared with transactions that we were not able to match. Since the CoreLogic deeds data do not provide seller demographics, we can only compare prices for the matched versus non-matched data. Other than the first two years, when the match rate is very low, the matched sample exhibits higher prices on average. This may reflect several factors. First, our sample includes only owner-occupied homes, while the unmatched sample thus

[7] See "Historical Reported Voting Rates: Table A-1. Reported Voting and Registration by Race, Hispanic Origin, Sex and Age Groups: November 1964 to 2024," *United States Census Bureau,* accessed July 2025.

includes second- and investment-home properties, which are likely to have lower value than primary residences. Second, the unmatched data include homes purchased by non-citizens, which may have lower value than homes owned by citizens. Third, the unmatched sample also has many homeowners who are not registered voters, who tend to have lower income than registered voters and so are likely to own lower priced homes on average.[8]

We further refine the sample by focusing on the age of the first seller listed on the deed and whether the transaction was conducted at arm's length. Since we can assign only one age per transaction, we use the age of the first-listed seller who, in cases with multiple sellers, is disproportionately male. Transactions where the first seller's age cannot be identified are excluded, reducing the matched sample to 49 million unique transactions. We then exclude non-arm's-length sales, which reduces the sample to approximately 24 million transactions. This step also removes distressed sales (e.g., foreclosures and short sales), which have been shown to be a key driver of racial disparities in housing returns (Kermani and Wong, 2021). Appendix Table C.1 reports the breakdown of arm's-length and non-arm's-length transactions by age group. We observe a monotonic decline in the share of arm's-length transactions across age groups, consistent with the idea that older home sellers are more likely to transfer properties to relatives. Nevertheless, a substantial share—nearly 40%—of transactions involving sellers aged 76 and above are still conducted at arm's length.

We then identify repeat sales by aggregating transactions based on the CLIP, which uniquely identifies each property in the CoreLogic data set. To ensure continuity of ownership, we require that the buyer in the initial transaction be the same individual listed as the seller in the subsequent transaction, following Kermani and Wong (2021) and Goldsmith-Pinkham and Shue (2023). We implement this match by comparing the buyer's full name from the first sale and the seller's full name from the second sale using the Jaro-Winkler string distance metric, which ranges from 0 to 1, with 1 indicating an exact match. We consider a pair to be a match if the Jaro-Winkler score exceeds 0.8. This procedure yields approximately 14 million repeat sales.

Lastly, we apply several additional filters. First, we exclude fiduciary sales—transactions conducted using a fiduciary deed, in which the grantor acts in a legal or trust-based capacity on behalf of another party. Fiduciaries may include executors, administrators, trustees, guardians, receivers, or

[8] See "Voting and Registration in the Election of November 2022: Table 7 Reported Voting and Registration of Family Members, by Age and Family Income: November 2022," *United States Census Bureau,* accessed July 2025.

commissioners, and such transactions typically arise in the context of estates, trusts, or court proceedings. Since the fiduciary is not the asset owner and therefore may have different economic incentives from the owner, these sales may not reflect typical homeowner behavior. Second, we restrict the sample to transactions in which ages of both the purchaser and seller are greater than 18. Third, we exclude sales with holding periods of less than three years, a commonly used cutoff to filter out home flipping activity (DeFusco et al., 2022). Fourth, we require both the purchase and sale prices to exceed $10,000 to ensure that the analysis is not affected by data entry errors (Amornsiripanitch, 2020). Finally, we exclude transactions from non-disclosure states—Alaska, Idaho, Kansas, Louisiana, Mississippi, Missouri, Montana, New Mexico, North Dakota, Texas, Utah, and Wyoming because the sale prices in these states are often not part of the public record and may be based on CoreLogic's imputation rather than actual transaction data. After applying these filters, our final sample for the baseline regression analysis consists of approximately 10 million repeat sales.

2.3 Age Distribution Comparison with the Zillow Consumer Housing Trends Report

To validate the representativeness of our matched sample, we benchmark its age distribution against survey data collected by Zillow. In each wave, Zillow's Consumer Housing Trends survey is nationally representative and includes responses from approximately 6,000 recent home sellers in the United States. Figure 5 overlays our sample's age breakdown with Zillow's published age distributions from 2022 and 2024.[9]

Visually, the two sets of distributions exhibit strong concordance across most age bins, suggesting that our matching procedure captures the core demographic structure of the country's seller population. Nevertheless, two systematic differences emerge. First, the 18–to-29 cohort accounts for a noticeably smaller share in our sample than reported by Zillow—in our data, this group constitutes roughly 2.5 percent of sellers versus 11 to 13 percent in the Zillow surveys. Second, the late middle-aged brackets (50–59 and 60–69) are over-represented in our matched data relative to Zillow's findings.

We attribute these deviations to two sample construction choices. The first stems from our in-

[9] See "Sellers: Results from the Zillow Consumer Housing Trends Report 2022," *Zillow*, August 2022; "Sellers: Results from the Zillow Consumer Housing Trends Report 2024," *Zillow*, October 2024.

clusion criteria. By requiring each seller to have been at least 18 at the time of purchase and to hold the property for a minimum of three years, we effectively shift the lower bound of our youngest cohort to ages 22 to 29, thereby truncating representation among the youngest sellers. The second source of divergence arises from our reliance on voter-registration records to infer age. Younger adults register to vote at significantly lower rates than older adults (see footnote 7).

3 OLS Fixed Effects Results

3.1 Baseline Model

In our baseline model, we start by regressing the seller's return on seller-age indicator variables, using 10-year age bins to partition the age distribution, because we not do want to impose a linear relationship between seller's age and returns. We use the 36-to-45 age group as the reference group. We then introduce fixed effects capturing zip-code location, buy-year-quarter, sell-year-quarter, and their interactions. To assess how the results vary when we introduce control variables for seller demographics, we add indicator variables for race, ethnicity, gender, and marital status. Last, we add the age of the property, the holding period, and cash sale indicator. Standard errors are clustered by state. Formally, we use ordinary least squares (OLS) to estimate the following regression equation:

$$Y_i = \alpha + \sum_{j}^{J} \beta_j \times \mathbb{1}(Age\,Group_{\,j})_i + \gamma' \mathbf{x_i} + FE + \epsilon_i. \tag{1}$$

i is the index for transactions. j is the index for age groups that sellers can be assigned to. $\gamma' \mathbf{x_i}$ is a vector of control variables, described above. The β_j coefficients capture the possibly non-linear relationship between the outcome variable Y and the seller's age. Our dependent variable of interest is the seller's unlevered, annualized holding period return (HPR), defined as the ratio of the sales price to the buy price, annualized by the holding period in years:

$$HPR = \left(\frac{Sale\,Price}{Buy\,Price} \right)^{\frac{1}{Holding\,Period}}. \tag{2}$$

Returns, as opposed to the actual sale price, help remove property-level unobserved heterogeneity by differencing out time-invariant characteristics such as square footage, lot size, number of bedrooms, number of bathrooms, neighborhood characteristics, etc. We do not use levered returns because the variable construction would require us to impose strong assumptions about down payments, monthly payments, the path of interest rates, and refinancing behavior, all of which are likely to vary strongly with seller age. Moreover, unlevered returns offer a more direct measure of whether older home sellers receive a fair deal in the housing market.

Panel A of Table 2 reports summary statistics for the annualized returns, seller age, as well as demographic characteristics for our matched baseline regression sample. The average return is approximately 4.8% per year, ranging from 1.4% at the 25^{th} percentile to 6.7% at the 75^{th} percentile. Sellers between the ages of 36 and 75 comprise about 77% of the sample, with sellers above age 75 comprising about 8%. Recall that the lower proportion reflects, in part, that older sellers are more likely to transfer properties to family members in non-arm's-length transactions, which we leave out of the analysis. Appendix Table D.1 presents summary statistics for the gender and race control variables used in the analysis. Note also that the share of minority sellers, especially Black and Hispanic sellers, fall well below their population shares, in part because these groups are less likely to own homes than White and Asian Americans, and in part because they are less likely to appear in the voter registration rolls.

Table 3 reports our baseline regression results. Column 1 presents the result from a specification with no control variables. We can see a clear monotonic decline in returns across the age distribution, with the youngest sellers earning the highest returns, and returns falling consistently with seller age thereafter. This model has little explanatory power, however, and may largely reflect the fact that older sellers, on average, own older properties which usually fetch low sales prices. In column 2, we add *zip-code* $\times$ *buy-year-quarter* and *zip-code* $\times$ *sell-year-quarter* fixed effects to capture most of this heterogeneity, raising the adjusted R^2 from 1% to 43%. Here, the age effects are now prominent only for the youngest and oldest sellers. Introducing demographic control variables lowers the over-75 coefficient slightly (columns 3 and 4).[10] The change, or the lack thereof, in the coefficient of interest suggests that the age effect is largely orthogonal to the race and gender effects that have been previously studied. The last four

[10] We include control variables for gender and race. For gender, the categories are single male, single female, couple, and unknown, with single male serving as the omitted category in the regression. For race, the categories are Asian, White, Black, Latino, and unknown, with White serving as the reference group. See Appendix A for a detailed description of these variables.

columns, where we incorporate an indicator for cash buyers, holding period in years, and even more granular fixed effects, reduce the over-75 coefficient somewhat further, to about 0.57% to 0.59% lower returns per year, compared with sellers between age 36 and 45, the omitted category. The last column contains the most stringent set of fixed effects, interacting *zip-code* × *buy-year-quarter* and *zip-code* × *sell-year-quarter*. Introducing the most stringent set of fixed effects has no meaningful impact on the age gap in returns.[11] Therefore, unlike the gender effect documented by Goldsmith-Pinkham and Shue (2023), the age effect is not driven by market timing.

As noted in the introduction, Figure 2 shows that the age gap holds even when we use year-by-year age indicator variables instead of age group indicator variables; that is, in old age, the age effect grows with each passing year of seller age. Figure 3 shows that the age gap in return is present in almost every state. In Figures 6 and 7, we deconstruct the returns by modeling log sale price and log buy price separately; this approach separates the numerator and the denominator used to build the holding period return outcome variable. These two figures plot the age effect for both prices, estimated in separate models. In Figure 7, age represents the buyer's age, calculated as the seller's age at sale minus the holding period.[12] The figures show that the 75+ age gap in returns is principally driven by the variation in the sale price and not the buy price. Furthermore, the figures also give some assurance that the old-age returns disadvantage that we document is unlikely to be driven by person-specific omitted variable bias that just so happens to be correlated with selling age in our sample; that is, the older sellers in our sample are not particularly "bad buyers" such that, when they were younger, they all bought houses at relatively higher prices.[13]

3.2 MLS Properties

To explore other potential mechanisms, we next focus on the subset of transactions that we can identify on one of the Multiple Listing Services (MLS). Each MLS is a separate local or regional chartered corporation.

[11] The sample size drops sharply because of singletons.

[12] To estimate the age effects in Figures 6 and 7, we control for property fixed effects for both regressions, *zip-code* × *buy-year-quarter* effects for buy-price, and *zip-code* × *sell-year-quarter* for the sale price. Property fixed effects are used to make the results from these regressions comparable to the baseline returns regressions because, by construction, time-invariant property characteristics are differenced out when annualized HPR is used as the outcome variable.

[13] We see a small age discount on the buy price for the youngest buyers, perhaps because young buyers, on average, are more financially constrained and thus are more likely to purchase lower-price properties. In fact, while we do not focus on the youngest cohorts, these two figures show that their higher returns reflect both lower prices at purchase and higher prices at sale. This is consistent with the idea that young people are more likely to purchase low-priced "starter homes," which Damen et al. (2025) show to have higher returns both over time and across different countries.

Most are non-profit organizations, while a minority (e.g., Northwest MLS) are broker-owned cooperatives. Member brokers elect a board of directors that sets rules and budgets. Paid staff manage the listing database, data feeds, training, and compliance. Listing on the standard MLS ensures that sellers receive broad visibility by making their listing accessible to all MLS subscribers, and by distributing them to public portals such as Zillow and Realtor.com. These public listings provide detailed data on location, asking price, days on market, physical characteristics of the property, as well as a textual description of the property. Such data are not available for properties sold off-MLS, which we examine and compare with MLS transactions below.[14]

Our base-case models include all transactions, regardless of whether or not they were marketed publicly on the MLS. As we have shown, controlling for seller demographics does little to affect the magnitude of age on returns. In this subsection, we focus on the MLS sub-sample in order to incorporate addition co-variates to the returns models. In particular, we test whether differences in renovations and/or property maintenance help explain the age gap. In addition, we consider whether dual agency, the same agent represents both the buyer and the seller, matters for returns.

Panel B of Table 2 reports summary statistics for the MLS sample. Here, we introduce four indicator variables to capture renovation and general quality of the home, based on keywords taken from the textual description of each listing on the MLS. To define the keywords, we start with a random sample of 1,000 MLS property descriptions that we scraped from the internet and asked ChatGPT to produce a list of keywords based on the text from the MLS listing, as well as ChatGPT's general knowledge base, as follows:

- *High positive*: Terms indicating major upgrades to core structural components (e.g., new roof, foundation repair) or essential home systems (e.g., updated HVAC, remodeled kitchen),
- *Low positive*: Terms indicating minor cosmetic updates or upgrades to individual components and surface finishes (e.g., fresh paint, upgraded finishes),
- *Neutral*: Marketing terms with no condition implication (e.g., charming, cozy, pride of ownership),
- *High negative*: Signs of deferred maintenance or poor condition (e.g., fixer-upper, as-is).

[14] Days on market is not correlated with seller age, so we leave it out of the analysis below.

We then use the keywords to define the following variables: *High positive*, *Low positive*, *Neutral*, and *High negative*.[15] As shown in Panel B of Table 2, *High positive* represents 20% of the sample, *Low positive* represents 43%, *Neutral* represents 67%, and *High negative* represents just 3%. Note that some of these categories overlap (e.g., *Low positive* and *Neutral*), while others almost never do (e.g., *High positive* and *High negative*). Beyond the textual analysis, the MLS sample also allows us to identify transactions associated with dual agency whereby the same agent receives fees on both the sell-side and the buy-side, which is associated with about 8% of the MLS transactions.

Table 4 reports linear probability models where the four quality metrics and the dual agency indicator are used as outcome variables. The likelihood of *High positive*, *Low positive* and *Neutral* all decline throughout the life cycle, with homes sold by the youngest people exhibiting more renovations and the highest overall level of upkeep. Most strikingly, the *High negative* indicator increases sharply among the oldest group of sellers, although it shows little variation across age groups otherwise. The coefficient for the 76-and-older cohort is more than 4%, which, when compared with the sample mean of 3%, means that the oldest sellers are 1.3 times more likely than the average seller to sell a property requiring substantial repair. As such, this factor sharply distinguishes very old sellers from everyone else. Interestingly, the likelihood of dual agency increases monotonically with age.[16]

We next explore how property quality and dual agency affect the relationship between seller age and returns. Table 5 reports the returns regressions for the MLS sample.[17] Column 1 presents the estimates for the baseline model on the MLS sample, which turns out to be comparable to the baseline estimates for the larger sample presented in columns 5 through 8 from Table 3. The return discount for the oldest group is nearly identical in the MLS sample (-0.59% versus -0.58%), as both samples cover the same sample period. In the subsequent columns, we introduce the dual agency indicator and the four indicators for quality of upkeep and renovations, first individually and then interacted with the age group indicators. Results presented in column 2 show that both *High positive* and *High negative* help explain returns, with the other three indicators having relatively little impact. Adding the quality indicators to

[15] See Appendix A.4 for the complete list of keywords used to define each variable.

[16] In untabulated analysis, we examine whether sellers aged 76 and older are more likely to be classified as motivated sellers based on listing descriptions containing keywords such as "motivated seller," "urgent sale," and other similar phrases indicating seller urgency. We find no evidence that individuals in this age group are more likely to be motivated sellers than younger sellers.

[17] Since using holding period return as the outcome effectively removes fixed property characteristics, we leave out property characteristic control variables, such as number of bedrooms, number of bathrooms, square footage, and lot size. However, adding these controls has very little effect on the age effects. See Appendix Table D.2.

the model attenuates the age coefficient by about 10% (from -0.59% to -0.53%) because older sellers on average sell properties with fewer major improvements and poorer maintenance. These effects, however, are amplified in the models with interactions. The last column includes interactions of *Dual agency*, *High positive* and *High negative* with the age indicators. In this model, the oldest age group coefficient declines to 0.43%, meaning that an elderly person selling a property with "normal" quality, meaning both *High positive* and *High negative* are set to zero, and without dual agency experience lower annual returns of about one-half percent per year. In contrast, an older person selling a poorly maintained home (*High positive* = 0 and *High negative* = 1) receives returns that are about 1.09% lower per year than the reference group (=0.43+0.66). Moreover, the age discount increases from 0.43% to 0.63% for properties sold under dual agency.

3.3 The Role of Pocket Listings

In this subsection, we compare private (off-MLS) to public (MLS) sales. Properties sold off-MLS, known as "pocket listings" in real estate parlance, are marketed privately, typically within the listing agent's personal network.[18] As a result, exposure to potential buyers is severely limited. In addition, pocket listings are fundamentally less transparent than sales made on the MLS because the listing is not subjected to the MLS's listing standards. As such, the information about these pocket listing transactions that are observable to us is much more limited. This means that we are not able to directly compare, in the same model, the impact of the quality metrics from the MLS data with the impact of pocket listings, since detailed information on the property is not available for off-MLS transactions.

Sellers choose to list off-MLS with the advice and influence of their listing agents. Sellers may prefer pocket listings for legitimate reasons. For example, high-profile sellers, such as celebrities or public figures, may prefer pocket listings to protect their privacy or personal safety. Other sellers may use them to show the property to a limited number of possible buyers while preparing the home for a fully public presentation, including completing renovations or staging. Some agents employ pocket listings as part of a phased marketing approach, starting with peer-to-peer promotion to build interest before launching a full public campaign on the MLS. Agents may also use off-MLS marketing to test market pricing for

[18] Public advertising for pocket listing was explicitly prohibited after the National Association of Realtors adopted its Clear Cooperation Policy in 2020. For more details, see "MLS Clear Cooperation Policy," *National Association of Realtors,* 2020.

unique or luxury properties before committing to a public list price.

Despite these benefits, pocket listings can also have negative effects on the final sale price due to lack of transparency (e.g., no information for the public to see on the MLS platform) and market depth (i.e., not as many people will know that the property is for sale). Less transparency and higher degrees of information asymmetry can lead to worse financial outcomes for sellers (Levitt and Syverson, 2008; Agarwal et al., 2019). Therefore, if older sellers are more likely to use pocket listing, this practice could help explain the large age gap in return.

Panel C of Table 2 reports summary statistics, comparing MLS properties with those listed off-MLS. We identify the on- and off-MLS sales by merging the CoreLogic deeds data to the MLS listing data. A transaction is sold on the MLS if the transaction can be matched with the MLS listing data. If the transaction cannot be matched, we classify it as off-MLS or pocket listing. Because a non-match can arise not only from true pocket listings but also from for-sale-by-owner deals, wholesale transactions, incomplete MLS coverage in CoreLogic's feed, or record-linkage errors (e.g., address variants or missing identifiers), our off-MLS indicator should be viewed as an upper-bound estimate of the prevalence of genuine pocket listings. According to survey data from the National Association of Realtors (NAR), for-sale-by-owner (FSBO) transactions account for approximately 6 to 10% of all housing sales, with the primary reason sellers choose the FSBO route being that they are selling to a relative, friend, or neighbor.[19] This analysis contains fewer observations than our baseline tests from Table 3 because the CoreLogic MLS listing data are less complete than the deeds data, e.g., some counties might not be covered by the MLS data collected by CoreLogic. Hence, we restrict this sample to counties in which there is at least one MLS transaction in a given month. Overall, 38% of the transactions are off-MLS. In the raw data, the off-MLS transactions have lower average returns than on-MLS ones: 4.55% versus 4.91%. Moreover, off-MLS transactions are more likely to be sold to an investor (23% versus 19%), and they also are more likely to be sold in all-cash deals (21% versus 17%).[20]

Table 6 presents regression results that test whether seller age correlates with the method of sale.

[19] See "FSBOs Reach All-Time Low, More Sellers Rely on Agents," *NAR*, November 2024.

[20] To identify investors in housing transactions, we use two complementary approaches. The first categorizes transactions based on short holding periods, labeling homes resold within three years as "short-term" sales, which are indicative of speculative activity (Bayer et al., 2020). The second approach classifies buyers based on occupancy status, identifying them as non-occupants—and thus more likely to be investors—if their mailing address at purchase differs from the property address (Chinco and Mayer, 2016). Both proxies aim to capture investment-driven behavior through either quick resale or lack of intent to occupy the home.

The regressions estimate the likelihood that a property is sold off-MLS and the likelihood that it is sold to an investor, conditioning on similar control variables and fixed effects as in prior regressions. In the last two columns, we model the likelihood that a property is sold to investors conditional on both seller demographics like age as well as whether or not the property was a pocket listing (i.e., off-MLS). Results in columns 1 and 2 show that the likelihood of pocket listings and sale to investors increases with age, with large increases for sellers aged 76 and older. The magnitude suggests that the oldest cohort is 2.54% more likely to sell to an investor than sellers in the reference group, and 1.84% more likely to sell off-MLS. These are large magnitudes compared to the unconditional means presented in the summary statistics table. Results in column 3 show that sales to investors are more likely when transactions are off-MLS; that is, agents bring investors to the table more in pocket listings than properties sold on the MLS. And, from column 4, this effect is much larger for older sellers. For example, an over-75 seller using a pocket listing is 3.27% more likely to sell to an investor than a middle aged (36 to 45) seller.

Table 7 shows how the method of sale is correlated with returns, and whether these patterns vary across the age distribution. Column 1 presents regression results that replicate the baseline age effect result from Table 3 for this subsample. Results presented in column 2 show that both pocket listings (off-MLS) and properties sold to investors receive lower returns: off-MLS sales earn 0.29% lower returns and homes sold to investors earn 0.09% lower returns. These discounts, as shown in columns 3 through 5, are *much* larger for older sellers. For example, the discount from a pocket listing more than doubles for the oldest sellers, from -0.23% for middle-aged sellers to -0.47% (=0.23+0.24) for those over the age of 75 (Column 3). If we combine the coefficients, the full model from column 5 implies that an elderly seller receives returns that are 0.86% lower than a middle-aged seller when they use a pocket listing and sell to an investor (=0.46+0.23+0.17), compared to a 0.46% lower return when they list their property on the MLS and sell to a non-investor. Combining the evidence presented above, our analysis offers a potential explanation for why real estate middlemen are often able to purchase properties at substantially below market value (Bayer et al., 2020).

4 Aging and Agent Incentive Structure: Establishing Causality

As we have shown, properties sold as pocket listings receive lower returns on average, and this effect is much larger for older sellers, as well as when buyers are professional investors. To establish causality, we focus on a policy change designed to increase transparency and thus reduce improper use of private/pocket listings.

4.1 Pocket Listings Reform

We focus on a policy change that was implemented by the Midwest Real Estate Data (MRED), which operates one of the largest MLS platforms in the United States and the largest in Illinois. In recognition of the growing use of and concerns associated with pocket listings, MRED launched the Private Listing Network (PLN) in April 2016. This initiative was intended to allow brokers to pre-market properties, while preserving enough transparency to protect sellers' interest. The PLN was MRED's response to the increasing presence of pocket listing groups, private Facebook exchanges, brokerage-specific platforms, and unregulated "coming soon" marketing that bypassed the MLS entirely.[21]

The PLN operates as a separate listing channel within MRED's connectMLS system. It allows agents to share limited listing details with fellow MRED subscribers prior to full public exposure, hence, preserving many of the benefits of a pocket listing. Unlike the Standard Listing Network (SLN)—the standard MLS platform—which requires a full suite of listing information, the PLN permits the submission of minimal property details, such as address, basic location fields, expiration date, showing instructions, listing brokerage, and compensation terms. Listings in the PLN may include no price, a fixed price, or a price range, whereas SLN listings must include a definitive asking price. Importantly, days on market (DOM) do not accrue while a property is on the PLN. DOM begins only when the listing transitions to the SLN. Listings can remain in the PLN indefinitely, provided that they are still covered by a valid listing agreement.[22]

[21] See "MRED's Private Listing Network (PLN) FAQ'S," *Midwest Real Estate Data (MRED)*, April 2016, accessed July 2025; "Private, Not A Secret An inside look at off-MLS listing solutions," *Midwest Real Estate Data (MRED)*, 2019, accessed July 2025.

[22] Real estate agents are concerned with managing a property's days on market because properties that have been listed on the MLS for many months may lead buyers to draw negative inferences about their quality. For evidence that buyer stigma can harm real estate values from failed real estate auctions, see Cortés et al. (2022).

As we have argued, the visibility of a property depends on where it is listed. A pocket listing, which exists entirely outside of MRED's system, is typically visible only to the listing agent and their immediate network. A PLN listing, by contrast, is visible to all MRED subscribers within the secure connectMLS database but is not distributed to client-facing tools or public platforms such as Zillow or Redfin. A listing on the SLN receives the widest exposure; it is accessible to all MRED subscribers and is automatically syndicated to third-party real estate portals, websites, and client auto-search systems.

MRED imposes clear rules regarding the timing and documentation of listing entries. A property must be entered into either the PLN or SLN within 48 hours of the effective listing date, which is the date the listing agreement is signed, or within 24 hours of any public marketing, such as a yard sign or online advertisement, whichever occurs first. This timing requirement includes weekends and holidays, and excuses such as computer failure are not acceptable. Failure to list a property on the PLN or SLN within the required 48- or 24-hour window, depending on the conditions described above, would result in an automatic $1,000 fine if the property is later listed on the MRED system. If a property that is listed on the PLN goes under contract and closes without ever transitioning to the SLN, the listing agent is required to move the listing onto the SLN and complete all required data fields so that it can be used for statistical and appraisal purposes. Failure to report a closed PLN transaction to the SLN within 48 hours of closing would result in a $250 fine.

In sum, MRED's creation of the Private Listing Network (PLN) balances legitimate reasons for pocket listings against potential abuse by real estate agents by providing a "middle ground" between a traditional pocket listing and a fully public MLS listing. Importantly, the policy broadens exposure from the limited reach of a single agent's network to the pool of agents subscribed to MRED's MLS platform.

While other MLS platforms have implemented similar approaches, such as "coming soon" policies, MRED's PLN policy offers two notable advantages for us.[23] First, MRED operates one of the largest MLS systems in the United States, covering the majority of Illinois—including the Chicago metropolitan area—and serving over 45,000 real estate professionals, so we define all Illinois transactions after April 2016 as being treated in our difference-in-differences (DiD) regression models. Second, the policy is strictly enforced, with fines for non-compliance. Although we cannot directly identify which transactions

[23] Many MLS services have implemented policies or best practice guidelines designed to limit abuse related to pocket listings. This helps explain why the overall age gap has diminished somewhat in recent years, as we show later. See the "How are Other MLSs Around the Country Addressing the Off-MLS Problem?" section in "Private, Not A Secret An inside look at off-MLS listing solutions," *Midwest Real Estate Data (MRED)*, 2019, accessed July 2025.

originated on the PLN versus the SLN, the rule enables us to observe shifts in the relative volume of off- and on-MLS transactions around the policy's implementation. Since the PLN offers a middle ground between the public listings on the MLS and pocket listings, we expect the change to lead to an increase in overall MLS usage. In addition, if the PLN mitigates the risk of agent abuse from non-transparent pocket listings, older sellers should benefit most from the policy change.

4.2 The Impact of MRED's Private-Listing-Network Reform

Figure 8 plots how the fraction of properties listed on the MLS varies over time for transactions before and after the MRED policy change, comparing homes sold by people aged over 75 years in Illinois (treated) to those sold by people from the same age group in other states (control).[24] The effects are estimated relative to the base-period of 2014 using a dynamic difference-in-differences regression setup with the three-way fixed effects being zip-codes, which absorb time-invariant differences between Illinois and non-Illinois sellers, buy year-quarters, and sell year-quarters, which absorb the time trend. We plot the MLS fraction based on each transaction's closing date, rather than its listing date, because the former is available for all of the transactions, while the latter is observable only for observations that appear in the MLS data. As such, the effect of the policy, which went into effect in April of 2016, should show up in the data with a lag of one to two quarters, which is approximately equivalent to the average amount of time that a home gets sold.[25] As is clear from Figure 8, the policy led to a 10 percentage point (pp) increase in the probability that a for-sale property in Illinois gets listed on the MLS, compared with the control group. There is no pre-policy trend and the increase in on-MLS sales is persistent throughout the event window.

Table 8 reports the difference-in-differences (DiD) regression results for the MRED policy change, including the full set of fixed effects and control variables. Columns 1 through 3 present regression results for Illinois transactions, and compare the policy's effect across the age distribution. In these models, *Post* is set to one for properties sold after the policy goes into effect. In column 4, we present regression results from a triple difference regression that includes observations from all states. Here, *Post* continues to represent transactions after the policy and *Treated(IL)* refers to transactions from Illinois.

[24] For the MRED analysis in this section, we restrict the sample to 2014–2019 to avoid any effects related to COVID-19 and the Clear Cooperation Policy—implemented by NAR in 2020 and described below.

[25] While the time from listing to closing has shortened in the post-pandemic period, it was significantly longer before the pandemic. For instance, in 2010, the average number of days on market was 140. See "What Is the Average Time to Sell a House?" *Zillow*, 2019.

Results presented in column 1 suggest that the MRED policy, which substantially raised the overall frequency of MLS listings for older sellers in Illinois, had similar effects on the MLS listings across the age distribution within Illinois. While the policy effect is smaller for sellers between the ages of 46 and 65, the magnitude is small compared to the policy's overall effect of more than 10 pp (Figure 8). Results presented in the last three columns of Table 8 show that the age gap in returns falls sharply after the policy. Before the policy, older sellers in Illinois earned 0.83% lower annualized returns than middle-aged sellers (Column 3); but after the policy, the age gap fell to 0.35% (=0.83 - 0.48). Column 4, which incorporates all states, shows similar effects. Relative to other states, the Illinois age gap fell by 0.31%. This smaller relative improvement occurs because, as we have mentioned, there is a national trend towards reforms that address potential abuse of pocket listings, although with less specificity and clarity compared to the MRED reform (see footnote 23).

Figure 9 presents the dynamic difference-in-differences return results, showing the return discount for sellers over age 75 in Illinois around the policy change, relative to the benchmark group of Illinois sellers aged 36–45. Consistent with Figure 8, the return gap shrinks, again with a lag of about 2 quarters; and, also consistent with Figure 8, there is no pre-policy trend in returns between older and younger sellers.

The treatment effect from the DiD estimator should be interpreted as a lower bound, as there has been a general effort to increase the transparency of pocket listings. This effort may also explain why the age effect has declined in recent years, as shown in Figure 4. For example, in 2020, the National Association of Realtors initiated its Clear Cooperation Policy, which attempts to mimic some of the changes initiated earlier by MRED in Illinois (see footnote 18). This change works against us finding a statistically significant effect because the control also gets treated during the window of analysis. However, the goal of this section is to use the MRED policy to show that part of the age gap in housing returns that we document is causally driven by the adverse effects of pocket listings. Although we are only able to recover the lower bound of the treatment effect of the MRED policy, the results presented in this section help support our causal claim of the mechanism at play.

5 Conclusion

This paper provides the first comprehensive evidence that older people receive lower returns when they sell their homes. The age gap starts to emerge soon after age 70 and then increases with each additional year of seller age. We find that two mechanisms help explain the results. First, homes sold by older people tend to have lower rates of major renovations and higher rates of poor upkeep. Second, older sellers are more likely to sell their homes off-MLS and sell to investors. Both of these choices are associated with lower returns, and we show that their return effects are much larger for older sellers. These patterns are consistent with the theory that the real estate market is set up such that agents earn more income by maximizing total sales volume as opposed to maximizing the sale price of each transaction (Levitt and Syverson, 2008). Cognitive decline with age may make the elderly more vulnerable to such actions (Mazzonna and Peracchi, 2024), thus widening the age gap in returns. To provide causal evidence, we show that reforms designed to make pocket listings more transparent reduced both the prevalence of pocket listings and the magnitude of the age gap by a large and statistically significant amount.

References

Agarwal, S., J. He, T. F. Sing, and C. Song (2019). Do real estate agents have information advantages in housing markets? *Journal of Financial Economics 134*(3), 715–735.

Amaral, F., M. Dohmen, S. Kohl, and M. Schularick (2021). Superstar returns.

Amornsiripanitch, N. (2020). Why are residential property tax rates regressive? *Available at SSRN 3729072*.

Amornsiripanitch, N. (2023). The age gap in mortgage access. *Federal Reserve Bank of Philadelphia Working Paper*.

Artigue, H., P. Bayer, F. V. Ferreira, and S. Ross (2025). Does Homeownership Matter? The Long-Term Consequences of Losing a House during the Great Recession. Technical report, National Bureau of Economic Research.

Bayer, P., C. Geissler, K. Mangum, and J. W. Roberts (2020). Speculators and middlemen: The strategy and performance of investors in the housing market. *Review of Financial Studies 33*(11), 5212–5247.

Begley, J. and L. Lambie-Hanson (2015). The home maintenance and improvement behaviors of older adults in Boston. *Housing Policy Debate 25*(4), 754–781.

Benetton, M., M. Kudlyak, and J. Mondragon (2022). Dynastic home equity. *Available at SSRN 4158773*.

Bracke, P. (2015). House prices and rents: Microevidence from a matched data set in Central London. *Real Estate Economics 43*(2), 403–431.

Campbell, J. Y., S. Giglio, and P. Pathak (2011). Forced sales and house prices. *American Economic Review 101*(5), 2108–31.

Carlin, B., T. Umar, and H. Yi (2023). Deputizing financial institutions to fight elder abuse. *Journal of Financial Economics 149*(3), 557–577.

Chambers, D., C. Spaenjers, and E. Steiner (2021). The rate of return on real estate: Long-run micro-level evidence. *Review of Financial Studies 34*(8), 3572–3607.

Chinco, A. and C. Mayer (2016). Misinformed speculators and mispricing in the housing market. *Review of Financial Studies 29*(2), 486–522.

Colonnello, S., R. Marfè, Q. Xiong, et al. (2024). Housing yields. *University of Venice Working Paper*.

Cortés, K., M. Singh, D. H. Solomon, and P. Strahan (2022). The stench of failure: How perception affects house prices. Technical report, National Bureau of Economic Research.

Damen, S., M. Korevaar, and S. Van Nieuwerburgh (2025). An alpha in affordable housing? *Available at SSRN*.

Davidoff, T. (2004). Maintenance and the home equity of the elderly. *Haas School of Business Working Paper*.

DeFusco, A. A., C. G. Nathanson, and E. Zwick (2022). Speculative dynamics of prices and volume. *Journal of Financial Economics 146*(1), 205–229.

DeLiema, M., M. Deevy, A. Lusardi, and O. S. Mitchell (2020). Financial fraud among older Americans: Evidence and implications. *Journals of Gerontology: Series B 75*(4), 861–868.

Demers, A. and A. L. Eisfeldt (2022). Total returns to single-family rentals. *Real Estate Economics 50*(1), 7–32.

Diamond, R. and W. F. Diamond (2024). Racial differences in the total rate of return on owner-occupied housing. Technical report, National Bureau of Economic Research.

Egan, M., G. Matvos, and A. Seru (2019). The market for financial adviser misconduct. *Journal of Political Economy 127*(1), 233–295.

Eichholtz, P., M. Korevaar, T. Lindenthal, and R. Tallec (2021). The total return and risk to residential real estate. *Review of Financial Studies 34*(8), 3608–3646.

Gilbukh, S. and P. Goldsmith-Pinkham (2024). Heterogeneous real estate agents and the housing cycle. *Review of Financial Studies 37*(11), 3431–3489.

Goldsmith-Pinkham, P. and K. Shue (2023). The gender gap in housing returns. *Journal of Finance 78*(2), 1097–1145.

Hale, J. M., D. C. Schneider, N. K. Mehta, and M. Myrskylä (2020). Cognitive impairment in the US: Lifetime risk, age at onset, and years impaired. *SSM-Population Health 11*, 100577.

Halket, J., L. Loewenstein, and P. S. Willen (2023). The cross-section of housing returns. *Unpublished manuscript*.

Johnson, K. H., Z. Lin, and J. Xie (2015). Dual agent distortions in real estate transactions. *Real Estate Economics 43*(2), 507–536.

Kermani, A. and F. Wong (2021). Racial disparities in housing returns. Technical report, National Bureau of Economic Research.

Levitt, S. D. and C. Syverson (2008). Market distortions when agents are better informed: The value of information in real estate transactions. *Review of Economics and Statistics 90*(4), 599–611.

Mazzonna, F. and F. Peracchi (2024). Are older people aware of their cognitive decline? Misperception and financial decision-making. *Journal of Political Economy 132*(6), 1793–1830.

Ong, R. (2009). House price appreciation among elderly home owners in Australia. *CRAE Research Paper* (200904).

Rodda, D. T. and S. Patrabansh (2007). Homeowner age and house price appreciation. *Cityscape*, 123–151.

Shu, T., F. Xie, S. Zhang, and W. Zhang (2025). Political ideology in patent examination: Examiner partisanship and green innovation. *Available at SSRN 5305584*.

Sodini, P., S. Van Nieuwerburgh, R. Vestman, and U. von Lilienfeld-Toal (2023). Identifying the benefits from homeownership: A Swedish experiment. *American Economic Review 113*(12), 3173–3212.

Spenkuch, J. L., E. Teso, and G. Xu (2023). Ideology and performance in public organizations. *Econometrica 91*(4), 1171–1203.

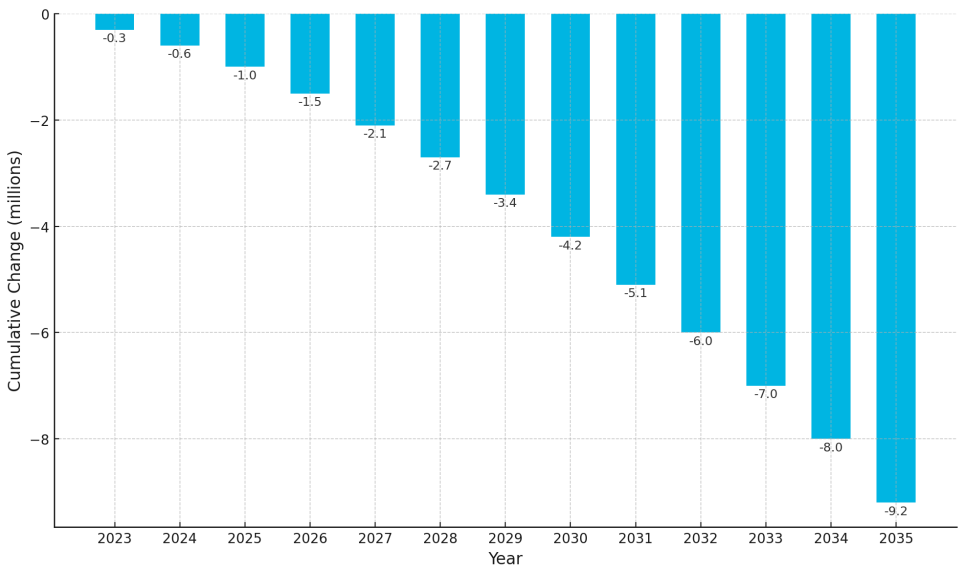


Figure 1: Cumulative Change in Baby Boomer Homeowner Households

This figure shows the annual cumulative change (in millions) in the number of Baby Boomer homeowner households from 2023 to 2035. Source: Exhibit 6 from Freddie Mac's February 2024 Economic, Housing and Mortgage Market Outlook.

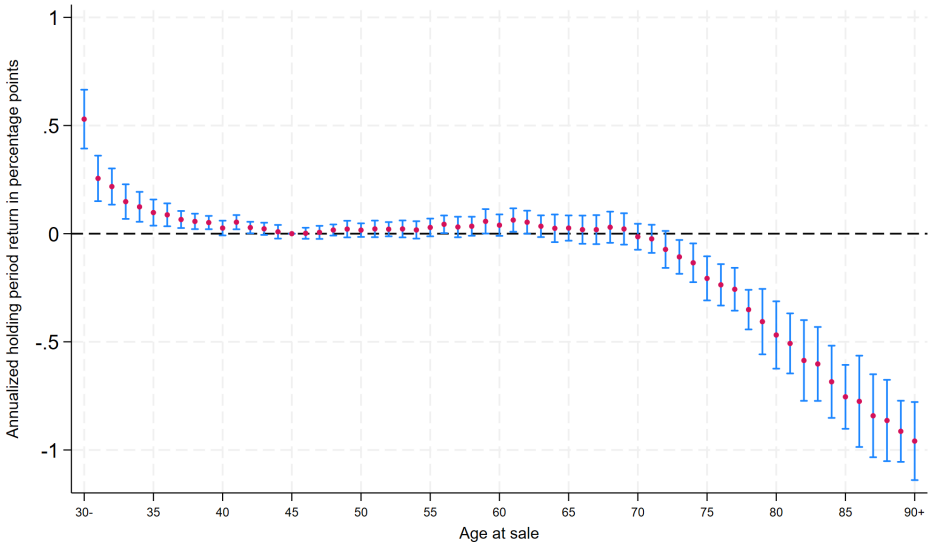


Figure 2: Age and Annualized Return

This figure plots the estimated annualized returns by seller age group. The dependent variable is the unlevered holding period return, calculated using repeat sales and expressed in percentage points. "30−" refers to sellers aged 30 or younger, and "90+" refers to sellers aged 90 or older. The estimates are produced by a regression that includes the same set of control variables and fixed effects as the specification presented in column 7 of Table 3. Heteroskedasticity-robust standard errors are clustered at the state level. Data source: CoreLogic and L2.

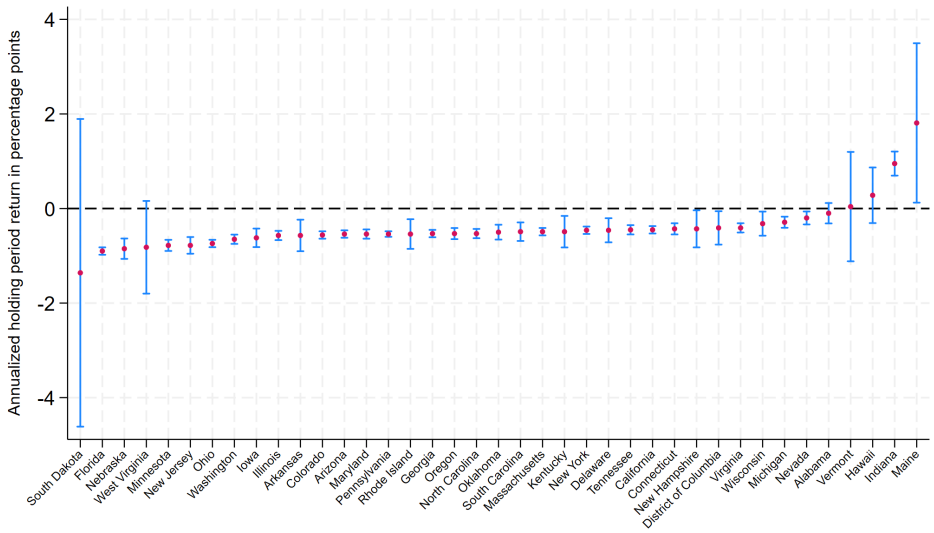


Figure 3: Annualized Returns for Sellers Aged Over 75 Across States

This figure plots the estimated annualized holding period returns for home sellers aged 76 and older across states. The estimates are produced from state-by-state regressions that include the same set of control variables and fixed effects as the specification presented in column 7 of Table 3. Heteroskedasticity-robust standard errors are clustered at the zip-code level. Data source: CoreLogic and L2.

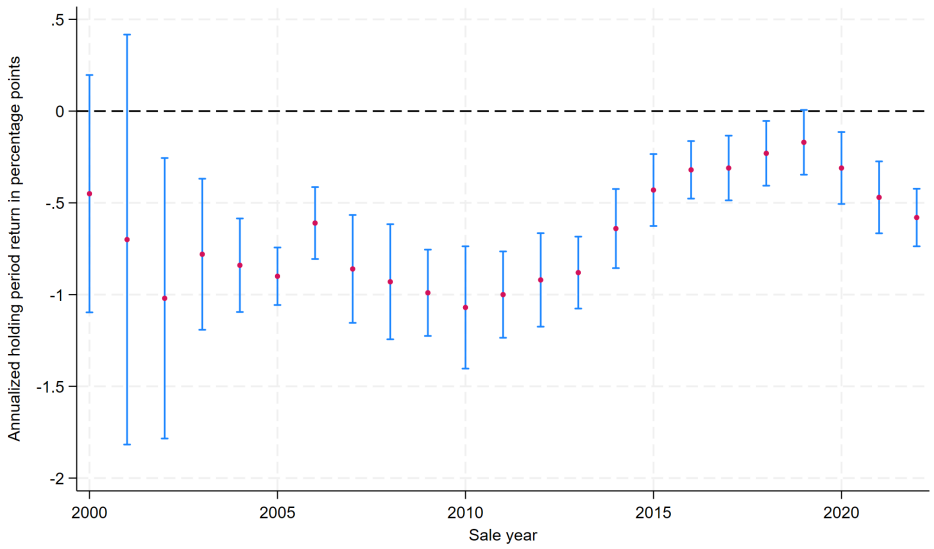


Figure 4: Annualized Returns for Sellers Aged Over 75 Across Sale Year

This figure shows the estimated annualized holding period returns for home sellers aged 76 and older by sale year. The estimates come from separate regressions run for each sale year, using the same control variables as in column 7 of Table 3. Because each observation corresponds to a sale year, the specification includes only zip-code-by-buy-year-quarter fixed effects. Our full sample spans 1998–2022. In this figure, we plot coefficients only from 2000 onward because the early years contain relatively few observations (see Table 1). Standard errors are heteroskedasticity-robust and clustered at the state level. Data sources: CoreLogic and L2.

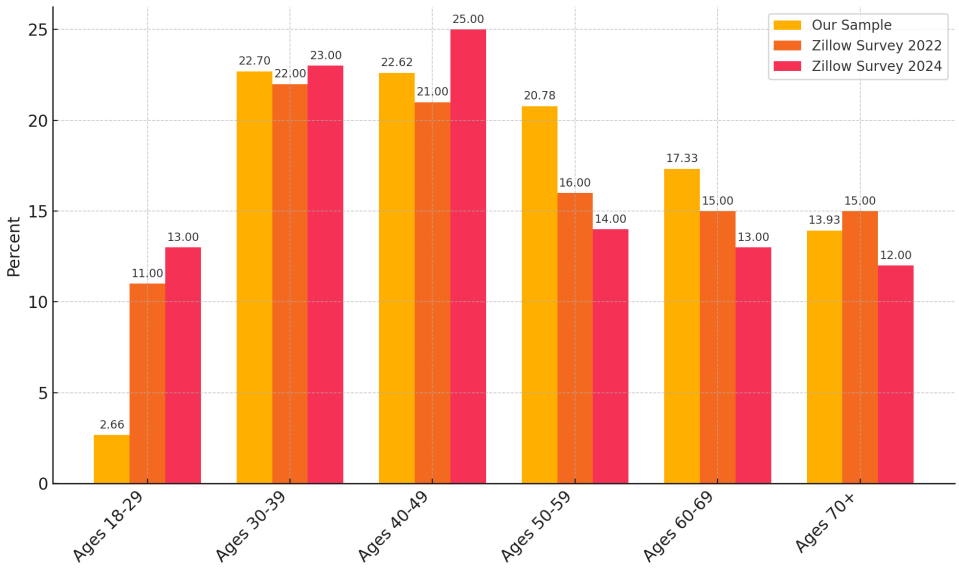


Figure 5: Seller Age Distribution

This figure shows home seller age distribution from different samples, with age bins on the horizontal axis and the percentage of sellers in each bin on the vertical axis. Each bar indicates the share of total sellers within each age group. The plot overlays our sample's distribution with those from Zillow's Consumer Housing Trends Reports for 2022 and 2024. Data source: CoreLogic, L2, and Zillow.

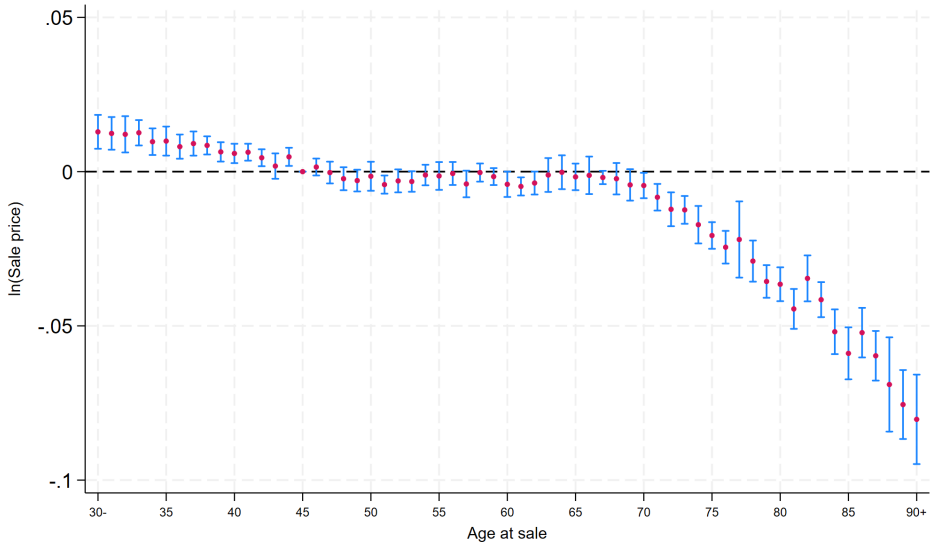


Figure 6: Sell Price and Seller's Age

This figure presents OLS estimates of the natural log of sell price across seller age bins. "30–" refers to sellers aged 30 or younger and "90+" refers to sellers aged 90 or older. The regression includes fixed effects for *zip code* × *sell-year-quarter* and *property*. The regression includes the same set of control variables as the specification presented in column 7 of Table 3. Heteroskedasticity-robust standard errors are clustered at the state level. Data source: CoreLogic and L2.

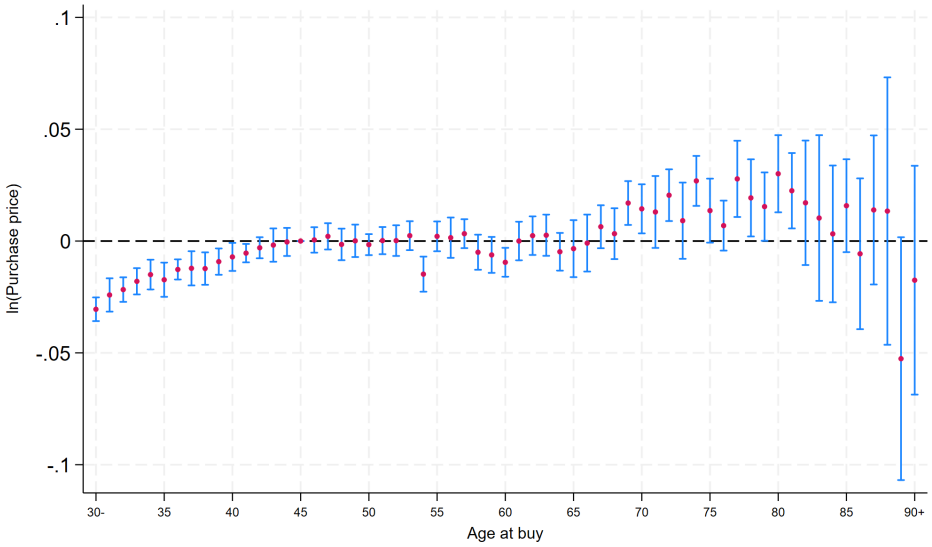


Figure 7: Purchase Price and Buyer's Age

This figure presents OLS estimates of the natural log of purchase price across buyer age bins at the time of purchase. "30–" refers to sellers aged 30 or younger and "90+" refers to sellers aged 90 or older. The regression includes fixed effects for *zip code* × *buy-year-quarter* and *property*. The regression includes the same set of control variables as the specification presented in column 7 of Table 3, except the *Holding period* control variable. Heteroskedasticity-robust standard errors are clustered at the state level. Data source: CoreLogic and L2.

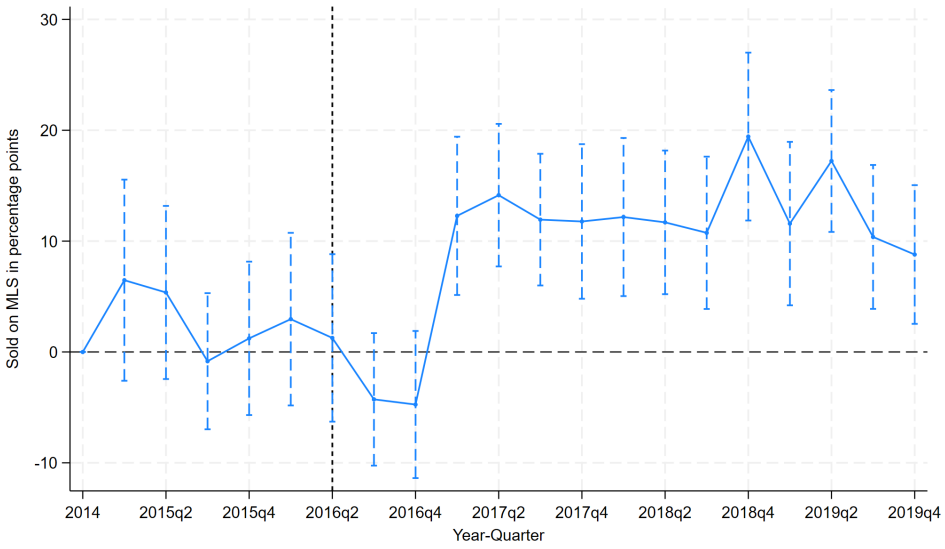


Figure 8: Event Study Plot for MRED's PLN Policy – MLS Sale

This figure presents the event study plot for the likelihood that sellers aged 76 or older listed their properties on the MLS around the time that MRED's PLN policy was implemented. The sample is restricted to sellers aged 76 or older. The dependent variable, *Sold on MLS* $\times$ *100*, is a binary variable that equals 100 if the transaction occurred on the MLS and zero otherwise. The treated group consists of elderly sellers in Illinois. The control group consists of elderly sellers from other states. The coefficients are from the interaction between the treatment indicator and the time indicator variables. The benchmark period is the full year of 2014. The regression includes the same set of control variables as the specification presented in column 7 of Table 3. The vertical line indicates the policy implementation quarter. The regression includes fixed effects for *zip code*, *buy-year-quarter*, and *sell-year-quarter*. Heteroskedasticity-robust standard errors are clustered at the zip-code level. Data source: CoreLogic and L2.

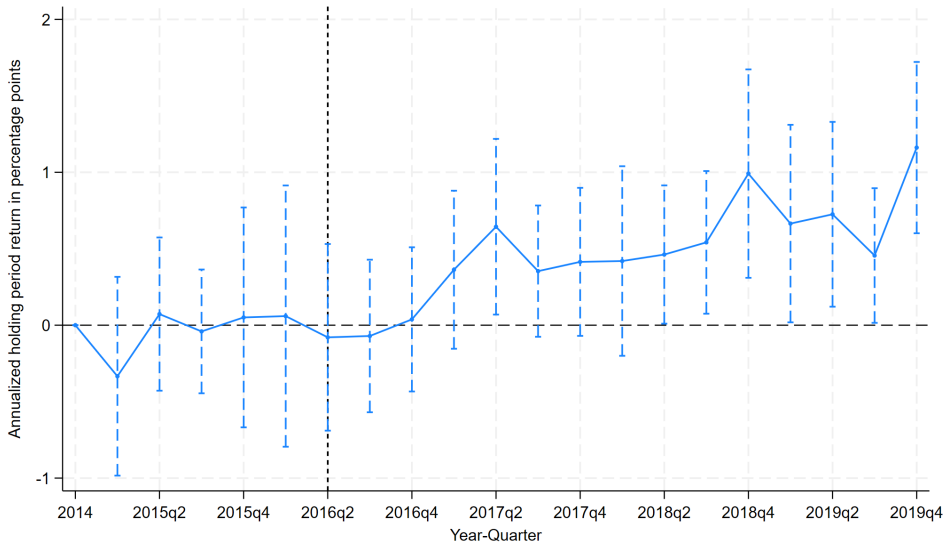


Figure 9: Event Study Plot for MRED's PLN Policy – Returns

This figure presents the event study plot for the annualized holding period return, expressed in percentage points, dynamic around the time that MRED's PLN policy was implemented. The sample is restricted to Illinois sellers. The dependent variable is the annualized holding period return. The treated group consists of sellers aged 76 or older, while the benchmark group includes sellers aged 36 to 45; we also control for other age groups and their corresponding interaction terms in the regression. The coefficients are from the interaction between the treatment indicator and the time indicator variables. The benchmark period is the full year of 2014. The regression includes the same set of control variables as the specification presented in column 7 of Table 3. The vertical line marks the policy implementation quarter. The regression includes *zip-code* × *buy-year-quarter* and *zip-code* × *sell-year-quarter* fixed effects. Heteroskedasticity-robust standard errors are clustered at the zip-code level. Data source: CoreLogic and L2.

Table 1: Match Rate

This table reports the year-by-year match statistics of our sample. Columns 1 through 3 present summary statistics for transactions that are not matched between transactions in CoreLogic and the L2 voter registration data. Columns 4 through 6 present the same statistics for the matched sample. The sample includes both arm's-length and non-arm's-length transactions. All years are based on the year of sale. Data source: CoreLogic and L2.

	Unmatched			Matched			Fraction matched
Sale year	N	Mean	Median	N	Mean	Median	
1998	1,148,845	316,868	150,000	39,789	230,801	102,000	0.03
1999	1,465,399	216,752	139,000	123,431	183,948	136,200	0.08
2000	1,518,558	278,971	145,000	170,671	209,018	155,000	0.10
2001	2,379,996	191,537	130,000	398,286	212,369	158,000	0.14
2002	2,853,293	224,433	147,500	634,950	231,626	169,000	0.18
2003	3,261,945	235,874	156,000	1,181,591	239,228	176,500	0.27
2004	3,211,856	282,305	179,500	1,353,898	267,816	199,900	0.30
2005	3,176,410	309,170	210,000	1,608,450	305,145	227,000	0.34
2006	2,706,200	305,436	195,000	1,568,324	321,141	239,000	0.37
2007	3,471,736	252,920	153,000	1,647,342	314,222	225,000	0.32
2008	2,697,628	233,694	138,300	1,400,694	293,555	200,000	0.34
2009	2,746,256	200,611	135,000	1,624,366	257,228	183,465	0.37
2010	2,829,114	236,769	130,000	1,790,093	258,777	180,381	0.39
2011	3,073,774	260,002	128,850	1,971,774	246,433	172,000	0.39
2012	3,433,871	213,223	133,000	2,469,750	254,644	178,400	0.42
2013	3,641,604	227,437	145,000	2,749,895	267,567	194,500	0.43
2014	3,470,136	232,155	142,900	2,845,521	283,717	208,000	0.45
2015	3,618,982	242,717	150,000	3,212,405	291,921	219,500	0.47
2016	3,779,018	234,752	156,100	3,425,145	299,015	228,000	0.48
2017	3,892,158	243,178	160,000	3,466,699	315,338	240,500	0.47
2018	4,158,678	241,443	160,000	3,467,547	328,036	249,000	0.45
2019	4,224,314	238,082	160,000	3,675,172	318,984	253,000	0.47
2020	4,467,221	263,064	179,900	4,191,316	361,579	285,000	0.48
2021	5,184,891	310,489	202,400	4,707,138	421,978	329,000	0.48
2022	1,938,195	340,626	225,000	1,602,277	503,329	364,000	0.45
Total	78,350,078			51,326,524			0.40

Table 2: Summary Statistics

This table presents summary statistics for the samples used in our analysis. Panel A covers the baseline regression sample from 1998 to 2022. Panel B reports statistics for on-MLS transactions only, for which we can identify textual descriptions of property conditions from 1998 to 2022. Panel C reports statistics for both on- and off-MLS transactions over the same period. Panel D summarizes the sample used for the MRED analysis, restricted to the period from 2014 to 2019. See Appendix A for additional details on variable definitions. Time periods are defined based on the year of sale. Data source: CoreLogic and L2.

	N	Mean	Median	S.D.
Panel A: Baseline sample				
Annualized return	10,200,671	4.77	3.56	5.90
Seller age 18 to 35	10,200,671	0.15	0.00	0.36
Seller age 36 to 45	10,200,671	0.25	0.00	0.43
Seller age 46 to 55	10,200,671	0.21	0.00	0.41
Seller age 56 to 65	10,200,671	0.19	0.00	0.40
Seller age 66 to 75	10,200,671	0.12	0.00	0.33
Seller age 76+	10,200,671	0.08	0.00	0.26
Cash buyer	10,200,671	0.18	0.00	0.38
Built year	9,008,261	1978	1986	31
Holding period (years)	10,200,671	11.17	9.02	7.42
Panel B: On-MLS and textual analysis sample				
Annualized return	3,001,317	4.70	3.49	5.57
High positive	3,001,317	0.20	0.00	0.40
Low positive	3,001,317	0.43	0.00	0.49
Neutral	3,001,317	0.67	1.00	0.47
High negative	3,001,317	0.03	0.00	0.16
Dual agent	3,001,317	0.08	0.00	0.27
Panel C: Off- and on-MLS sample				
Annualized return (full sample)	6,559,869	4.77	3.58	5.72
Sold to investor (full sample)	6,559,869	0.21	0.00	0.41
Off-MLS sale (full sample)	6,559,869	0.38	0.00	0.49
Cash buyer (full sample)	6,559,869	0.18	0.00	0.39
Annualized return (off-MLS subsample)	2,500,554	4.55	3.33	5.88
Sold to investor (off-MLS subsample)	2,500,554	0.23	0.00	0.42
Cash buyer (off-MLS subsample)	2,500,554	0.21	0.00	0.41
Annualized return (on-MLS subsample)	4,059,315	4.91	3.73	5.61
Sold to investor (on-MLS subsample)	4,059,315	0.19	0.00	0.39
Cash buyer (on-MLS subsample)	4,059,315	0.17	0.00	0.37
Panel D: MRED analysis sample				
Annualized return	131,390	1.99	1.31	4.22
Sold on MLS	125,087	0.54	1.00	0.50
Treated (IL)	3,841,791	0.03	0.00	0.18

Table 3: Seller's Age and Return

This table presents OLS regression results fpr annualized unlevered holding period returns on seller age groups at the time of sale. Returns are expressed in percentage points. The benchmark group is the 36-to-45 age group. *Cash buyer* is a binary variable that equals one if the transaction was completed in cash. *Year built* refers to the year that the property was constructed. *Holding period* denotes the number of years the seller owned the property before selling. Heteroskedasticity-robust standard errors are clustered at the state level. See Appendix A for detailed variable definitions. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Data source: CoreLogic and L2.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
18 – 35	1.05***	0.28***	0.28***	0.28***	0.28***	0.28***	0.20***	0.24***
	[0.11]	[0.05]	[0.05]	[0.05]	[0.05]	[0.05]	[0.03]	[0.04]
46 – 55	-0.36***	-0.08***	-0.06**	-0.06**	-0.05**	-0.05**	-0.02	-0.04
	[0.06]	[0.03]	[0.03]	[0.03]	[0.02]	[0.02]	[0.02]	[0.03]
56 – 65	-0.35***	-0.05	-0.02	-0.03	-0.01	-0.01	0.00	0.05
	[0.10]	[0.04]	[0.04]	[0.04]	[0.03]	[0.03]	[0.03]	[0.05]
66 – 75	-0.40***	-0.12***	-0.09*	-0.09*	-0.06	-0.06	-0.07*	-0.01
	[0.13]	[0.04]	[0.04]	[0.04]	[0.04]	[0.04]	[0.04]	[0.05]
76+	-1.13***	-0.74***	-0.67***	-0.67***	-0.58***	-0.59***	-0.59***	-0.57***
	[0.15]	[0.08]	[0.08]	[0.08]	[0.07]	[0.07]	[0.07]	[0.08]
Cash buyer					-0.44***	-0.44***	-0.43***	-0.34***
					[0.05]	[0.05]	[0.04]	[0.05]
Year built					-0.00***	-0.00***	-0.00***	-0.01***
					[0.00]	[0.00]	[0.00]	[0.00]
Holding period						0.23***		
						[0.04]		
Gender controls			Yes	Yes	Yes	Yes	Yes	Yes
Race and ethnicity controls				Yes	Yes	Yes	Yes	Yes
Zip-BuyYQ FE		Yes	Yes	Yes	Yes	Yes	Yes	
Zip-SellYQ FE		Yes	Yes	Yes	Yes	Yes	Yes	
BuyYQ-SellYQ FE							Yes	
Zip-BuyYQ-SellYQ FE								Yes
Observations	10,200,671	9,612,442	9,612,442	9,612,442	8,467,277	8,467,277	8,467,073	3,277,147
Adjusted R-squared	0.01	0.43	0.43	0.44	0.44	0.44	0.46	0.46

Table 4: Seller's Age and Property Condition

This table presents linear probability model regression results where property condition indicator variables and dual agency indicator variable are regressed onto seller age groups. Outcome variables are multiplied by 100. The reference group consists of sellers aged 36 to 45. Heteroskedasticity-robust standard errors are clustered at the state level. See the main text and Appendix A for detailed variable definitions. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Data source: CoreLogic and L2.

	(1)	(2)	(3)	(4)	(5)
	High positive	Low positive	Neutral	High negative	Dual agent
18 – 35	1.04***	1.42***	-0.65***	0.04	-0.26***
	(0.28)	(0.24)	(0.16)	(0.05)	(0.05)
46 – 55	-1.80***	-2.50***	-0.44***	0.20***	0.57***
	(0.27)	(0.16)	(0.14)	(0.04)	(0.07)
56 – 65	-2.22***	-2.87***	-0.52***	0.05	1.03***
	(0.43)	(0.33)	(0.18)	(0.07)	(0.09)
66 – 75	-2.90***	-4.00***	-0.79***	-0.12	1.49***
	(0.56)	(0.47)	(0.17)	(0.16)	(0.16)
76+	-6.88***	-7.88***	-2.24***	4.04***	2.40***
	(0.75)	(0.59)	(0.23)	(0.37)	(0.34)
Gender controls	Yes	Yes	Yes	Yes	Yes
Race and ethnicity controls	Yes	Yes	Yes	Yes	Yes
Other controls	Yes	Yes	Yes	Yes	Yes
Zip-BuyYQ FE	Yes	Yes	Yes	Yes	Yes
Zip-SellYQ FE	Yes	Yes	Yes	Yes	Yes
Observations	3,001,317	3,001,317	3,001,317	3,001,317	3,001,317
Adjusted R-squared	0.07	0.04	0.07	0.07	0.08

Table 5: Seller's Age, Property Condition, and Return

This table presents OLS regression results for annualized unlevered returns on property conditions interacted with seller age. The reference group consists of sellers aged 36 to 45. Heteroskedasticity-robust standard errors are clustered at the state level. See the main text and Appendix A for detailed variable definitions. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Data source: CoreLogic and L2.

	(1)	(2)	(3)	(4)	(5)	(6)
18 – 35	0.28***	0.28***	0.19***	0.28***	0.29***	0.19***
	(0.06)	(0.06)	(0.07)	(0.06)	(0.06)	(0.07)
46 – 55	-0.02	-0.01	0.03	-0.01	-0.02	0.03
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
56 – 65	0.02	0.02	0.07**	0.02	0.01	0.06**
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
66 – 75	-0.05	-0.04	0.01	-0.05	-0.05	0.01
	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)
76+	-0.59***	-0.53***	-0.53***	-0.51***	-0.56***	-0.43***
	(0.10)	(0.10)	(0.11)	(0.10)	(0.10)	(0.10)
Dual agent		-0.00			-0.00	-0.00
		(0.00)			(0.00)	(0.00)
High positive		0.36***	0.44***			0.43***
		(0.03)	(0.03)			(0.03)
Low positive		0.00				
		(0.00)				
Neutral		-0.00***				
		(0.00)				
High negative		-0.77***		-0.65***		-0.63***
		(0.07)		(0.08)		(0.08)
18 – 35 × Dual agent					-0.09**	-0.09**
					(0.04)	(0.04)
46 – 55 × Dual agent					0.09**	0.09**
					(0.04)	(0.04)
56 – 65 × Dual agent					0.07	0.06
					(0.04)	(0.04)
66 – 75 × Dual agent					-0.01	-0.02
					(0.07)	(0.07)
76+ × Dual agent					-0.21***	-0.20***
					(0.06)	(0.06)
18 – 35 × High positive			0.38***			0.38***
			(0.06)			(0.06)
46 – 55 × High positive			-0.20***			-0.20***
			(0.02)			(0.02)
56 – 65 × High positive			-0.24***			-0.24***
			(0.03)			(0.03)
66 – 75 × High positive			-0.29***			-0.29***
			(0.05)			(0.05)
76+ × High positive			-0.18***			-0.23***
			(0.03)			(0.03)
18 – 35 × High negative				0.29***		0.29***
				(0.10)		(0.10)
46 – 55 × High negative				-0.17***		-0.18***
				(0.06)		(0.06)
56 – 65 × High negative				-0.02		-0.02
				(0.08)		(0.07)
66 – 75 × High negative				-0.01		-0.02
				(0.08)		(0.08)
76+ × High negative				-0.66***		-0.66***
				(0.07)		(0.07)
Gender controls	Yes	Yes	Yes	Yes	Yes	Yes
Race and ethnicity controls	Yes	Yes	Yes	Yes	Yes	Yes
Other controls	Yes	Yes	Yes	Yes	Yes	Yes
Zip-SellYQ FE	Yes	Yes	Yes	Yes	Yes	Yes
Zip-BuyYQ FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,001,317	3,001,317	3,001,317	3,001,317	3,001,317	3,001,317
Adjusted R-squared	0.46	0.46	0.46	0.46	0.46	0.46

Table 6: Seller's Age and Transaction Type

This table presents linear probability model regression results where transaction type outcomes are regressed onto seller's age. Outcome variables are multiplied by 100. The reference group consists of sellers aged 36 to 45. Heteroskedasticity-robust standard errors are clustered at the state level. See Appendix A for detailed variable definitions. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Data source: CoreLogic and L2.

	(1)	(2)	(3)	(4)
	Sold to investor	Off-MLS sale	Sold to investor	Sold to investor
18 – 35	0.46***	-0.02	0.46***	0.48***
	(0.07)	(0.10)	(0.07)	(0.08)
46 – 55	0.04	0.33***	0.03	0.04
	(0.07)	(0.09)	(0.07)	(0.07)
56 – 65	0.58**	0.59***	0.57**	0.63***
	(0.21)	(0.15)	(0.21)	(0.23)
66 – 75	1.31***	0.76***	1.30***	1.34***
	(0.41)	(0.23)	(0.42)	(0.42)
76+	2.54***	1.84***	2.52***	2.04***
	(0.69)	(0.42)	(0.70)	(0.65)
Off-MLS sale			1.06*	1.01**
			(0.55)	(0.48)
18 – 35 × Off-MLS sale				-0.05
				(0.15)
46 – 55 × Off-MLS sale				-0.02
				(0.09)
56 – 65 × Off-MLS sale				-0.16
				(0.15)
66 – 75 × Off-MLS sale				-0.08
				(0.28)
76+ × Off-MLS sale				1.23***
				(0.43)
Gender controls	Yes	Yes	Yes	Yes
Race and ethnicity controls	Yes	Yes	Yes	Yes
Other controls	Yes	Yes	Yes	Yes
Zip-BuyYQ FE	Yes	Yes	Yes	Yes
Zip-SellYQ FE	Yes	Yes	Yes	Yes
Observations	6,559,869	6,559,869	6,559,869	6,559,869
Adjusted R-squared	0.34	0.41	0.34	0.34

Table 7: Seller's Age, Transaction Type, and Return

This table presents OLS regression results for annualized unlevered returns regressed onto seller's age interacted with the transaction type. The dependent variable is annualized holding period return, expressed in percentage points. *Sold to investor* is a binary variable that equals one if the property was sold to an investor and zero otherwise. *Off-MLS sale* is a binary variable that equals one if the transaction occurred off the MLS and zero otherwise. The reference group consists of sellers aged 36 to 45. Heteroskedasticity-robust standard errors are clustered at the state level. See Appendix A for detailed variable definitions. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Data source: CoreLogic and L2.

	(1)	(2)	(3)	(4)	(5)
18 – 35	0.31***	0.31***	0.31***	0.31***	0.31***
	(0.05)	(0.05)	(0.06)	(0.06)	(0.06)
46 – 55	-0.05*	-0.04*	-0.02	-0.05**	-0.03
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
56 – 65	-0.01	-0.01	0.02	-0.02	0.02
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
66 – 75	-0.07*	-0.07*	-0.02	-0.07*	-0.02
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
76+	-0.59***	-0.59***	-0.50***	-0.55***	-0.46***
	(0.09)	(0.09)	(0.08)	(0.07)	(0.07)
Off-MLS sale		-0.29***	-0.23***		-0.22***
		(0.05)	(0.05)		(0.05)
Sold to investor		-0.09***		-0.08***	-0.08***
		(0.03)		(0.03)	(0.03)
18 – 35 × Off-MLS sale			-0.00		-0.00
			(0.04)		(0.04)
46 – 55 × Off-MLS sale			-0.06		-0.06
			(0.04)		(0.04)
56 – 65 × Off-MLS sale			-0.09**		-0.09**
			(0.03)		(0.03)
66 – 75 × Off-MLS sale			-0.14***		-0.14***
			(0.04)		(0.04)
76+ × Off-MLS sale			-0.24***		-0.23***
			(0.04)		(0.04)
18 – 35 × Sold to investor				0.01	0.01
				(0.04)	(0.04)
46 – 55 × Sold to investor				0.03	0.03
				(0.05)	(0.05)
56 – 65 × Sold to investor				0.04	0.04
				(0.04)	(0.04)
66 – 75 × Sold to investor				-0.01	-0.01
				(0.05)	(0.05)
76+ × Sold to investor				-0.19**	-0.17**
				(0.08)	(0.08)
Gender controls	Yes	Yes	Yes	Yes	Yes
Race and ethnicity controls	Yes	Yes	Yes	Yes	Yes
Other controls	Yes	Yes	Yes	Yes	Yes
Zip-BuyYQ FE	Yes	Yes	Yes	Yes	Yes
Zip-SellYQ FE	Yes	Yes	Yes	Yes	Yes
Observations	6,559,869	6,559,869	6,559,869	6,559,869	6,559,869
Adjusted R-squared	0.46	0.46	0.46	0.46	0.46

Table 8: MRED's PLN Policy Shock Analysis

This table presents difference-in-differences regression results based on MRED's adoption of the Private Listing Network (PLN) policy. The sample period spans 2014 to 2019. *Sold on MLS* is a binary variable that equals 100 if the transaction is completed on the MLS and zero otherwise. *Post* is a binary variable that equals one if the transaction's closing date is in or after the second quarter of 2016 and zero otherwise. *Treated (IL)* is a binary variable that equals one if the transaction occurred in Illinois and zero otherwise. Columns 1 to 3 restrict the sample to transactions in Illinois. Column 4 includes transactions from all states. The reference group consists of sellers aged 36 to 45. Heteroskedasticity-robust standard errors are clustered at the ZIP code level in columns 1–3, and at the state level in column 4. See Appendix A for detailed variable definitions. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Data source: CoreLogic and L2.

	(1)	(2)	(3)	(4)
	Sold on MLS	Annualized return	Annualized return	Annualized return
18 – 35	-0.88	0.46***	0.42***	0.23***
	(0.97)	(0.06)	(0.10)	(0.07)
46 – 55	1.88**	-0.02	-0.05	0.08**
	(0.80)	(0.04)	(0.07)	(0.03)
56 – 65	1.19	-0.00	-0.05	0.09**
	(0.82)	(0.04)	(0.07)	(0.04)
66 – 75	1.27	-0.09*	-0.18**	0.03
	(1.01)	(0.05)	(0.08)	(0.05)
76+	1.92	-0.48***	-0.83***	-0.55***
	(1.64)	(0.07)	(0.11)	(0.09)
18 – 35 × Post	1.18		0.05	0.10*
	(1.18)		(0.11)	(0.05)
46 – 55 × Post	-2.52***		0.05	-0.11***
	(0.96)		(0.07)	(0.02)
56 – 65 × Post	-1.78*		0.07	-0.04*
	(0.95)		(0.08)	(0.02)
66 – 75 × Post	-0.78		0.13	-0.02
	(1.13)		(0.09)	(0.03)
76+ × Post	-2.26		0.48***	0.17***
	(1.82)		(0.12)	(0.06)
Treated (IL) × 18 – 35				0.19***
				(0.07)
Treated (IL) × 46 – 55				-0.14***
				(0.03)
Treated (IL) × 56 – 65				-0.14***
				(0.04)
Treated (IL) × 66 – 75				-0.22***
				(0.05)
Treated (IL) × 76+				-0.28***
				(0.09)
Treated (IL) × Post × 18 – 35				-0.05
				(0.05)
Treated (IL) × Post × 46 – 55				0.15***
				(0.02)
Treated (IL) × Post × 56 – 65				0.11***
				(0.02)
Treated (IL) × Post × 66 – 75				0.15***
				(0.03)
Treated (IL) × Post × 76+				0.31***
				(0.06)
Gender controls	Yes	Yes	Yes	Yes
Race and ethnicity controls	Yes	Yes	Yes	Yes
Other controls	Yes	Yes	Yes	Yes
Zip-BuyYQ FE	Yes	Yes	Yes	Yes
Zip-SellYQ FE	Yes	Yes	Yes	Yes
Observations	125,087	131,390	131,390	3,841,791
Adjusted R-squared	0.14	0.33	0.33	0.38

Appendix

A Variable Definitions

This appendix section provides detailed definitions for the variables used in the regression analyses. The data are primarily sourced from CoreLogic's deeds and Multiple Listing Service (MLS) databases, supplemented with demographic information from L2 voter registration records.

A.1 Main Dependent Variables

Annualized returns: The unlevered, annualized return on a property, calculated as the ratio of the sale price to the purchase price, annualized by the holding period in years using repeat sales.

ln(Purchase price): The natural logarithm of the property's purchase price from the pair of transactions used to compute annualized return.

ln(Sell price): The natural logarithm of the property's sale price from the pair of transactions used to compute annualized return.

A.2 Seller's Demographics

Seller's demographics are identified by matching property transaction records from CoreLogic to individual records in the L2 voter registration database.

Seller Age Groups: A set of mutually exclusive indicator variables categorizing the seller's age at the time of sale. In cases with multiple sellers, we use the age of the first seller listed on the deed. The age groups are 18 to 35, 36 to 45, 46 to 55, 56 to 65, 66 to 75, and 76 and older. We use the 36-to-45 age group as the reference group.

Seller Gender and Marital Status: A set of mutually exclusive indicator variables based on a textual analysis of seller names to infer gender and marital status. These variables are constructed using information from all sellers listed on the deed. The groups are heterosexual couple, single male, single female, and other, which includes individuals whose gender we cannot identify and same-sex couples. Single male is the omitted category in regression models. For a detailed description of the variable construction, see Appendix C.2.

Seller Race and Ethnicity: A set of mutually exclusive indicator variables based on a textual analysis of seller names to infer race and ethnicity. In cases with multiple sellers, we use the race of the first seller listed on the deed. The categories are Black, Asian, Hispanic, White, and unknown. White is the

omitted reference category in regression models. For a detailed description of the variable construction, see Appendix C.2.

A.3 Transaction and Property Characteristics

These variables capture details about the property and the transaction, sourced from the CoreLogic deeds and MLS databases.

Cash buyer: An indicator variable that equals one if the transaction was an all-cash purchase and zero otherwise. We infer that the transaction is an all-cash transaction if no mortgage was recorded at the time of sale.

Year Built: The year the property was originally constructed.

Holding period: A count variable that measures the number of years between the current seller's purchase year and sale year.

Dual agent: An indicator variable that equals one if the same real estate agent represented both the buyer and the seller in the transaction and zero otherwise.

Off-MLS sale: An indicator variable that equals one if a transaction in the deeds dataset cannot be matched with a transaction in the MLS listing dataset and zero otherwise. The variable is defined only when there is at least one active MLS board in the transaction's county-year cell.

Sold to investor: An indicator variable that equals one if the buyer is classified as an investor and zero otherwise. A buyer is considered to be an investor if (1) their mailing address differs from the property's physical address at the time of purchase, or (2) the property is resold within a three-year period.

Sold on MLS: An indicator variable that equals one if a transaction in the deeds dataset can be matched with a transaction in the MLS listing dataset and zero otherwise. The variable is defined only when there is at least one active MLS board in the transaction's county-year cell.

Treated (IL): An indicator variable that equals one for properties located in Illinois and zero otherwise.

A.4 Property Condition from MLS Listing Descriptions

These variables are derived from a textual analysis of the public remarks section of MLS listings and they are used to classify the property's condition. To compile the keyword lists below, we start with a random sample of 1,000 MLS property descriptions that we scraped from the internet and asked ChatGPT to produce a list of keywords based on both the text itself and its general knowledge base. The four, not mutually exclusive, categories are as follows.

High positive: An indicator variable that equals one if the listing description includes keywords signifying major renovations or structural upgrades and zero otherwise. The keywords are: "new roof" "roof replacement" "replaced roof" "new gutters" "upgraded gutters" "new siding" "siding replacement" "replaced siding" "foundation repair" "repaired foundation" "stabilized foundation" "new deck" "replaced deck" "deck replacement" "new HVAC" "HVAC replacement" "replaced HVAC" "new furnace" "furnace replacement" "replaced furnace" "new central air" "new heat pump" "new water heater" "water heater replaced" "rewired electrical" "new wiring" "updated electrical" "electrical rewiring" "new electrical" "replumbed pipes" "new plumbing" "updated plumbing" "new sewer line" "new water main" "remodeled kitchen" "kitchen remodel" "new kitchen" "gut renovated kitchen" "new cabinets" "remodeled bathroom" "new bathroom" "bath remodel" "updated bath" "remodeled basement" "fully renovated" "completely remodeled" "recently renovated" "gut renovation" "down to the studs remodel" "new well pump" "new well holding tank" "new windows" "window replacement" "replaced windows" "new septic system" "new drain field."

Low positive: An indicator variable that equals one if the listing text indicates minor or cosmetic updates and zero otherwise. The keywords are: "fresh paint" "repainted" "new paint" "painted" "upgraded finishes" "new fixtures" "modern fixtures" "stylish updates" "new hardware" "updated hardware" "refinished cabinets" "new cabinets" "granite countertops" "quartz countertops" "marble countertops" "solid surface countertops" "new backsplash" "backsplash installed" "new flooring" "hardwood floors" "engineered hardwood" "laminate flooring" "luxury vinyl plank" "lvp installed" "new tile" "new marble" "carpet installed" "brand new carpet" "polished hardwood" "refinished floors" "new countertops" "tile backsplash" "sparkling clean" "clean" "move-in ready" "turnkey" "ready to move" "ready to go" "freshly updated" "recently updated" "like new" "mint condition" "pristine."

Neutral: An indicator variable that equals one if the description uses general, subjective, or marketing-focused language without specific information about the property's condition and zero otherwise. The keywords are: "charming" "cozy" "inviting" "bright" "sun-filled" "sun filled" "airy" "must-see" "must see" "spacious" "open concept" "open layout" "floor plan" "elegant" "breathtaking" "picture-perfect" "picture perfect" "showstopper" "one-of-a-kind" "one of a kind" "unique" "pride of ownership" "ideal for entertaining" "perfect for entertaining" "garden views" "cul-de-sac" "cul de sac" "desirable location" "luxury feel" "resort style" "amenity rich" "amenity-rich" "walkable" "quiet neighborhood" "prime location" "historic charm" "architectural details" "high ceilings" "vaulted ceilings" "exposed beams" "hardwood accents" "molding" "crown molding."

High negative: An indicator variable that equals one if the listing description contains terms that suggest that the property is in poor condition or requires significant repairs and zero otherwise. The keywords are: "fixer-upper" "fixer upper" "needs tlc" "as is" "as-is" "deferred maintenance" "handyman special" "structural issue" "structural damage" "structural problem" "renovation needed" "outdated" "distressed property" "great potential" "investor" "diamond in the rough" "undervalued" "contractor's special" "requires work" "poor condition" "tender loving care" "needs work" "requires renovations" "bring your contractor."

B Matching Transaction Records to Voter Data

In this section, we describe the matching process used to link CoreLogic's transaction data with L2's voter demographic records in order to identify seller age information. We begin by standardizing the L2 data across years, and then perform the matching based on individuals' names, residential addresses, and ZIP codes.

B.1 L2 Data Coverage

To identify sellers' age information, we utilize individual-level data from the L2 voter registration database, a comprehensive national file of registered voters across the United States. L2 compiles these data by acquiring official voter registration lists directly from state and county election offices and then standardizing them into a uniform format. Each dataset represents a point-in-time snapshot of the active voter population at the time of acquisition. The database offers extensive nationwide coverage and includes detailed records for each individual. The information includes full name, residential address, date of birth, and party affiliation. The L2 data have been widely used in political campaigns and academic research (Spenkuch et al., 2023; Shu et al., 2025).

Table B.1: L2 Data Coverage by State

This table shows the years of data availability for each state in the L2 voter registration database, including coverage in both the raw and formatted datasets. For the early years, in some states, voter registration data are available only for selected counties rather than the entire state.

State	Coverage	State	Coverage
AL	2009, 2012 – 2023	MT	2012 – 2023
AK	2014 – 2023	NE	2009, 2011, 2013 – 2023
AZ	2014 – 2024	NV	2012 – 2023
AR	2013 – 2023	NH	2014 – 2023
CA	2001, 2002, 2006 – 2024	NJ	2001, 2005 – 2024
CO	2012 – 2024	NM	2012 – 2023
CT	2008, 2010 – 2023	NY	2002, 2004 – 2024
DE	2013 – 2023	NC	2003 – 2024
DC	2014 – 2023	ND	2014 – 2023
FL	2004 – 2023	OH	2006 – 2024
GA	2010 – 2024	OK	2010 – 2023
HI	2012, 2014 – 2023	OR	2004, 2005, 2007, 2008, 2010 – 2023
ID	2012 – 2023	PA	2004 – 2024
IL	2005 – 2024	RI	2010 – 2023
IN	2006 – 2024	SC	2010 – 2024
IA	2010, 2012 – 2023	SD	2012 – 2023
KS	2010 – 2023	TN	2010 – 2023
KY	2012 – 2023	TX	2010 – 2024
LA	2010 – 2023	UT	2012, 2014 – 2023
ME	2011 – 2023	VT	2012 – 2023
MD	2010 – 2023	VA	2014 – 2023
MA	2013 – 2023	WA	2006 – 2024
MI	2010 – 2023	WV	2012 – 2023
MN	2012 – 2023	WI	2014 – 2024
MS	2013 – 2023	WY	2014 – 2023
MO	2009 – 2024		

We obtain a complete version of the L2 database, which consists of two types of voter registration records: (1) formatted voter registration files covering all 50 states and Washington, D.C., from 2014 to 2024, and (2) raw, unformatted voter files available for a subset of states. Coverage of the raw files varies by state and generally correlates with population size. For example, larger states such as California, Florida, Illinois, and North Carolina have records dating back to the early 2000s, while coverage in smaller states typically begins around 2010. To maximize coverage, we extract and standardize individual-level

variables – name, date of birth, residential address, and ZIP code – from the raw data and integrate them with the formatted L2 files. Appendix Table B.1 details the availability of formatted and raw voter registration data across states.

B.2 Matching Seller Age

We match CoreLogic deed transfer records, which include buyer and seller names and property addresses, with the combined L2 voter registration data listed above, which contain voter names, residential addresses, and birth dates. Matches are based on exact matches of individual names, residential addresses, and Zip Code across the two datasets. Importantly, we match only sellers (not buyers) to the L2 data, so our sample is conditional on completed sales. Because some transactions list multiple sellers on the deed, we conduct the matching at the transaction-by-seller level. In total, we successfully match approximately 68 million observations at this level. The matching process involves three main steps. Detailed descriptions of each step are provided below.

Step 1: Standardizing Addresses and Names

In the first step, we standardize residential addresses and individual names to harmonize the reporting formats between the CoreLogic and the L2 databases. Our objective is to create two versions of each residential address—one that includes the apartment number and one that does not—and to extract standardized name components: first name, middle initial, last name, and name suffix.

Residential addresses in CoreLogic and L2 are often recorded inconsistently. For example, one source may use "Avenue," "Northwest," and "Unit," while another uses "Ave," "NW," and "Apartment." To resolve these discrepancies, we harmonize the common street and directional terms across both datasets based on postal abbreviations. Specifically, we make the following changes:

- Street suffixes: Avenue → Ave, Boulevard → Blvd, Circle → Cir, Court → Ct, Cove → Cv, Drive → Dr, Freeway → Fwy, Highway → Hwy, Lane → Ln, Place → Pl, Point → Pt, Ridge → Rdg, Road → Rd, Street → St, Terrace → Ter, Trace → Trce, and Trail → Trl.
- Directionals: East → E, West → W, North → N, South → S, Northeast → NE, Northwest → NW, Southeast → SE, and Southwest → SW.
- Apartment: Apartment → #, Apt → #, Suite → #, Room → #, and Unit → #,

After the standardization step, each dataset retains two versions of the addresses. For example, consider the address "1810 Commonwealth Avenue Apartment 10, Brighton, MA 02135." The standardized output includes one version with unit information (e.g., "1810 Commonwealth Ave #10, Brighton, MA 02135") and another version without it (e.g., "1810 Commonwealth Ave, Brighton, MA 02135").

For seller names, the two datasets differ significantly in formatting. The L2 voter registration data generally provide higher-quality name fields, with clear separation of first name, middle name, last name, and name suffix. By contrast, CoreLogic names are less consistently structured. The first and middle names are often combined into a single field, with middle names frequently reduced to initials. Name suffixes are typically embedded at the end of the name string rather than stored in a separate field. When multiple sellers are listed on a deed, their names are often combined using an ampersand ("&"), without a standardized format. Additionally, the two sources may differ in how they report first names. For instance, one dataset may list "Benjamin" while the other records the abbreviated form "Ben." To address these inconsistencies, we standardize CoreLogic seller names to align them with the L2 names using the following steps:

1. Separate co-sellers into individual name entries if joined by symbols such as "&."
2. Extract last names by taking the first word from the full name string. The remaining text is assumed to include the first name, middle name, and suffix. We rely on the full name field because many observations lack values in the structured name components, and because the separate last name field often omits suffixes.
3. Identify and extract suffixes (e.g., Jr., Sr., II, III, IV, VI, VII) from the end of the remaining string.
4. Parse first and middle names by treating the first word as the first name and the rest as the middle name(s).
5. Create a truncated version of the first name (first three letters), which is later used in relaxed matching rounds. We generate this variable for both CoreLogic and L2 records.

Although we extract last names from the full name string, this method has limitations. For instance, it may incorrectly parse multi-word surnames such as "De Sierra" by capturing only "De" as the last name. Similarly, although CoreLogic typically lists first names before middle names, the order is occasionally reversed. We address these issues in Step 3 by constructing no-space full names, by concatenating all name components without whitespace, to improve consistency and robustness in matching.

Step 2: Selecting the Appropriate L2 Data Year for Matching

After standardizing both individual names and residential addresses, the next step involves matching each housing transaction with the appropriate L2 voter registration dataset. Ideally, we use the voter registration dataset corresponding to the year of the transaction. However, this process is complicated by the fact that voter records are not always promptly updated, and individuals may remain registered at outdated addresses or be inactive in certain years. To account for these discrepancies, we developed a temporal matching strategy. The following steps are performed sequentially, and the search for a given transaction stops as soon as the first match is found:

1. We first attempt to match a transaction with the L2 dataset from the same year. For example, a 2015 transaction is matched with the 2015 L2 dataset.
2. If no match is found, we search the L2 datasets from the preceding three years (e.g., 2014, 2013, 2012).
3. If a match is still not found, we expand the search to the subsequent three years of L2 data (e.g., 2016, 2017, 2018).
4. To account for transactions occurring before our L2 data begin, we match them against the first available year of L2 data, applying the three-year forward search. For example, if the first L2 dataset is from 2014, transactions from as early as 2011 can be matched to it.
5. If a transaction remains unmatched after all steps are completed, it is removed from the analysis because the seller's age cannot be identified.

This approach maximizes the chances of identifying the corresponding voter while acknowledging the limitations of voter registration data. The matching composition by L2 year and transaction year difference is shown below in Table B.2.

Table B.2: Matching Result by L2 Dataset

This table shows the distribution of matched housing transactions based on the time difference between the transaction year and the year of the L2 voter registration data used for the match. A "0" indicates the transaction was matched with L2 data from the same year, while negative and positive values indicate matches with L2 data from preceding or subsequent years, respectively.

L2 Year Minus Transaction Year	Frequency	Percent
-3	1,365,983	2
-2	2,133,825	3.12
-1	6,808,277	9.95
0	49,733,590	72.65
1	3,725,743	5.44
2	2,543,577	3.72
3	2,145,714	3.13
Total	68,456,709	100

Step 3: Matching Criteria

We conduct the matching between each CoreLogic transaction and the selected L2 voter registration dataset in 16 sequential rounds, progressing from the strictest to most relaxed criteria. The steps are

performed in order, and for each transaction, the search stops as soon as a match is found. Note that all address-based matching variables include the property's Zip code.

Rounds 1 Through 6: Standard Name and Address Matching

1. Match on first name, middle name initials, last name, name suffix, and address, including apartment number. We match based on middle name initials rather than full middle names, as the CoreLogic database typically records middle names as initials.
2. Match on first name, middle name initials, last name, name suffix, and address, excluding apartment number.
3. Match on first name, last name, name suffix, and address, including apartment number.
4. Match on first name, last name, name suffix, and address, excluding apartment number.
5. Match on first name, last name, and address, including apartment number.
6. Match on first name, last name, and address, excluding apartment number.

Rounds 7 Through 10: No-Space Full Name Matching

We further address name formatting issues, including cases where multi-word last names are incorrectly parsed. For example, "Del Rio" being truncated to "Del." To mitigate this, we implement a procedure that compares a version of the full name with all spaces removed, allowing us to more reliably detect and correct such parsing errors. Specifically, the no-space full name in the CoreLogic data is constructed as last name followed by first name, middle name, and suffix. In the L2 data, the no-space full name is similarly constructed as last name followed by first name, middle name (or middle initial), and suffix.

7. Match on no-space full name (using full middle name in L2) and address, including apartment number.
8. Match on no-space full name (using middle initial in L2) and address, including apartment number.
9. Match on no-space full name (using full middle name in L2) and address, excluding apartment number.
10. Match on no-space full name (using middle initial in L2) and address, excluding apartment number.

Rounds 11 Through 14: Adjusted No-Space Full Name Matching (Excluding Single-Letter Name Components)

To further address cases where the first and middle names are reversed across datasets, for example, "H Grant Erling" in L2 versus "Grant H Erling" in CoreLogic, we exclude any name component consisting

of a single letter before constructing the no-space full name. Specifically, we remove all single-letter components prior to concatenating the name fields. In the example above, the adjusted no-space full name becomes "GrantErling."

11. Match on adjusted no-space full name (using full middle name in L2) and address including apartment number.
12. Match on adjusted no-space full name (using middle initial in L2) and address including apartment number.
13. Match on adjusted no-space full name (using full middle name in L2) and address excluding apartment number.
14. Match on adjusted no-space full name (using middle initial in L2) and address excluding apartment number.

Rounds 15 Through 16: Relaxed First Name Matching

To account for inconsistencies in first name abbreviations (e.g., "Benjamin" vs. "Ben"), we apply a more flexible rule that uses only the first three letters of the first name.

15. Match on the first three letters of the first name, full last name, and address including apartment number.
16. Match on the first three letters of the first name, full last name, and address excluding apartment number.

Duplicate matches are rare, well below 1 percent. In cases where a seller matches to multiple L2 voter records, we break ties by first computing the Levenshtein distance between names and selecting the record with the closest string match. If multiple records are equally close, we choose the one with the older age, based on the assumption that older individuals are more likely to be the property owners. The composition of matches by round is reported in Table B.3.

Table B.3: Matching Results Break-down by Rounds

This table presents the frequency and percentage of successful matches between CoreLogic transaction records and L2 voter data, broken down by the 16 sequential matching rounds. The rounds are ordered from strictest (Round 1) to most relaxed criteria (Round 16).

Round	Frequency	Percent
1	45,975,660	67.16
2	1,466,590	2.14
3	16,104,262	23.52
4	600,467	0.88
5	1,560,869	2.28
6	51,725	0.08
7	333,623	0.49
8	147,138	0.21
9	13,227	0.02
10	3,893	0.01
11	136,922	0.2
12	69,613	0.1
13	5,918	0.01
14	2,097	0
15	1,916,080	2.8
16	68,625	0.1
Total	68,456,709	100

Finally, we remove all personally identifiable information (PII) from the dataset before any analysis is performed.

C Further Processing

C.1 Arm's-Length Transactions

The 68 million matched observations are at the transaction-by-seller level, as some transactions involve multiple sellers. In total, the matched sample contains 51,326,524 unique transactions. Since we can assign only one age per transaction, we use the age of the first seller listed on the deed who, in cases with multiple sellers, is disproportionately male. If the age of the first seller cannot be identified, we drop the transaction. This reduces the matched sample to 48,787,113 unique transactions.

The matched sample includes both arm's-length and non-arm's-length sales. Although these non-arm's-length transactions are excluded from our regression analyses (e.g., intra-family, foreclosure, short sales), they offer valuable descriptive insight into how older homeowners manage their housing assets. For sellers aged 76 and above, 38.75% of transactions are sold on the market. The breakdown of the arm's-length transactions by age buckets is as follows in Table C.1.

Restricting the sample to arm's-length transactions yields 24,327,194 observations. Additional sample filters—such as identifying repeat sales and imposing age restrictions—are described in the data section in the main text.

Table C.1: Arm's-Length and Non-Arm's-Length Transaction by Age Group

This table presents the breakdown of arm's-length and non-arm's-length transactions categorized by seller age group.

Age Group	Non Arm's Length (Freq.)	Arm's Length (Freq.)	Arm's Length (%)	Total Transactions
18 – 35	2,483,051	3,763,508	60.25	6,246,559
36 – 45	3,858,043	5,237,942	57.59	9,095,985
46 – 55	4,493,299	4,694,638	51.1	9,187,937
56 – 65	4,868,174	4,596,213	48.56	9,464,387
66 – 75	4,339,871	3,239,669	42.74	7,579,540
76+	4,417,481	2,795,224	38.75	7,212,705
Total	24,459,919	24,327,194		48,787,113

C.2 Identify the Gender and Race of Home Sellers

We infer the gender and race of home sellers using name-based algorithms. For gender classification, we follow the approach of Goldsmith-Pinkham and Shue (2023) to identify family structures. Similarly, race is inferred from names using established name–race prediction methods.

Gender Identification

We infer sellers' gender using the data provided by Goldsmith-Pinkham and Shue (2023). Sellers are grouped into four categories: *Single Male*, *Single Female*, *Couple*, and *Unknown Gender*. These variables are constructed using information from all sellers listed on the deeds.

- A transaction is classified as *Single Male* if there is only one seller listed on the deed and the seller's first name is identified as male with at least 95% confidence.
- A transaction is classified as *Single Female* if there is only one seller and the name is identified as female with at least 95% confidence.

- A transaction is classified as a *Couple* if the deed lists at least one confidently identified male name and at least one confidently identified female name.

We also retain transactions for which gender cannot be reliably inferred and label them as *Unknown Gender*. Same-sex couples are also included in this group. Gender classification is primarily used as a control variable in our analysis.

Race and Ethnicity Identification

Race is inferred using the `ethnicolr` algorithm, which predicts racial and ethnic categories based on last names. The model classifies individuals into one of four groups: *Asian Pacific Islander*, *Black*, *White*, and *Hispanic*. The `ethnicolr` algorithm is trained on U.S. Census data and other administrative sources.[26] In cases with multiple sellers, we use the race of the first seller listed on the deed. As with gender, we retain observations with unclassified race and categorize them as *Unknown Race*. Race is also used as a control variable.

Our method differs from the approach in Kermani and Wong (2021), which identifies race by merging deed records with the HMDA data. Since race is not a central variable in our study, we adopt a name-based approach to ensure broader coverage and ease of implementation. Moreover, Kermani and Wong (2021) find that racial disparities in housing returns are largely driven by distressed sales. Since we exclude all distressed transactions from our analysis, our setting is less likely to reflect the same mechanisms.

[26] For detail, see "ethnicolr: Predict Race and Ethnicity From Name."

D Additional Results

Table D.1: Summary Statistics on Gender and Race Control Variables

This table presents summary statistics for the gender and race control variables used in the baseline regression sample from 1998 to 2022. The variables include classifications for gender and marital status (Couple, Single male, Single female, Unknown gender) and race/ethnicity (Black, Asian, Hispanic, White, Unknown race). Data source: CoreLogic and L2.

	N	Mean	Median	S.D.
Couple	10,200,671	0.39	0.00	0.49
Single male	10,200,671	0.23	0.00	0.42
Single female	10,200,671	0.20	0.00	0.40
Unknown gender	10,200,671	0.19	0.00	0.39
Black	10,200,671	0.01	0.00	0.11
Asian	10,200,671	0.03	0.00	0.17
Hispanic	10,200,671	0.05	0.00	0.23
White	10,200,671	0.90	1.00	0.29
Unknown race	10,200,671	0.00	0.00	0.01

Table D.2: Robustness Test – Control for Property Characteristics

This table presents OLS regression results from a robustness test examining annualized unlevered returns across seller age groups. The benchmark group comprises sellers aged 36 to 45. The regression includes additional controls for property characteristics—such as square footage, number of bedrooms and bathrooms, and indicators for fireplace, poor condition, garage, AC unit, basement, and waterfront view. Heteroskedasticity-robust standard errors are clustered at the state level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Data source: CoreLogic and L2.

	Annualized Returns
18 - 35	0.36***
	[0.04]
46 - 55	-0.09***
	[0.03]
56 - 65	-0.04
	[0.03]
66 - 75	-0.07
	[0.04]
76 +	-0.56***
	[0.08]
Gender controls	Yes
Race and ethnicity controls	Yes
Other controls	Yes
Property characteristic controls	Yes
Zip-BuyYQ FE	Yes
Zip-SellYQ FE	Yes
Observations	6,438,116
Adjusted R-squared	0.45