Decomposing Fire Sale Discounts in Housing Markets

Anurag A. Mehrotra [*] Adam D. Nowak [†] Patrick S. Smith [‡]

SDSU WVU UNC Charlotte

September 12, 2024

Abstract

This paper decomposes raw fire sale discounts in housing markets into house quality, asset illiquidity, and residual liquidity components. Using a novel computer-assisted text analysis algorithm, we find substandard house quality represents one-quarter of raw fire sale discounts. Although owner-occupied houses can easily be redeployed as rentals, we find asset illiquidity also represents one-quarter of raw fire sale discounts. The remaining residual liquidity component represents one-half of the raw fire sale discount. We confirm the residual liquidity component captures the cost of immediacy by showing that houses liquidated via fire sales sell 12 days faster on average than non-distressed houses.

Key Words: Fire sales, House quality, Illiquidity, Redeployability, Text analysis

[*]San Diego State University, Fowler College of Business; Email: amehrotra@sdsu.edu
[†]West Virginia University, John Chambers College of Business; Email: adam.d.nowak@gmail.com
[‡]University of North Carolina at Charlotte, Belk College of Business; Email: patrick.smith@charlotte.edu

1 Introduction

The self-reinforcing nature of fire sales can have far-reaching effects during times of market-wide distress, both in housing markets and the overall economy (French et al., 2010; Guren & McQuade, 2020). Understanding the components that contribute to fire sales in housing markets is critical to crafting and evaluating policies aimed at mitigating fire sales. Yet, little is known about the individual components that comprise fire sales in housing markets.[1] This paper decomposes the *raw* or unadjusted fire sale discount for single-family houses into a house quality component, an asset illiquidity component, and a residual liquidity component.

Housing studies examining the 2007-2009 financial crisis find raw fire sale discounts ranging from 20 to 30 percent. At face value, these steep discounts imply investors could have earned significant returns over a short holding period by purchasing distressed houses and reselling them at market values. However, a few studies acknowledge the lower unobserved quality of fire sale houses likely confounds the estimated raw fire sale discount. For example, Campbell, Giglio, and Pathak (2011) note estimated foreclosure discounts capture both the "urgency of the sale and the effects of the ownership transfer on the physical quality of the house." Similarly, Anenberg and Kung (2014) note that foreclosure discounts are "difficult to interpret because while some of the effect may be due to the foreclosure (e.g., banks have relatively high holding costs and thus sell at fire sale prices), some of it may simply reflect the fact that foreclosures tend to be of lower unobserved house quality." Although quality impairments are recognized in the literature, measurement and data limitations have hindered their decomposition.

We overcome these limitations using machine learning methods to decompose the raw fire sale discount for single-family houses into a quality component and a quality-adjusted fire sale component. The quality component captures foreclosure-related deferred maintenance stemming from homeowners in or near default. These homeowners have little incentive to invest in their property as any value added to the property will flow to the lender in the event of a default (Melzer, 2017).[2] Although previous research finds homeowners at risk of

[1] Fire sale discounts in housing markets are frequently referred to as either foreclosure or real estate owned (REO) discounts. We use the terms foreclosure, REO, and fire sale interchangeably.

[2] We use the term lender to refer to the party who owns the mortgage at the time of default.

default substantially cut back on home improvements and defer maintenance (Campbell et al., 2011; Lambie-Hanson, 2015; Melzer, 2017), we are the first to quantify how this agency conflict manifests in distressed house prices.

Using just over 7 million housing transactions across ten large metropolitan statistical areas (MSAs), we estimate a raw fire sale discount of 23%. Adjusting for house quality yields a quality-adjusted fire sale discount of 16.5%, indicating that house quality constitutes over one-quarter of the raw fire sale discount. The size of the raw fire sale discount varies across the ten MSAs, ranging from 12.5% in Los Angeles to 30.9% in Saint Louis. Adjusting for quality yields quality-adjusted fire sale discounts between 9.6% in Los Angeles and 18.2% in Saint Louis. Overall, we find evidence of a sizeable house quality component, ranging from one-quarter to up to two-fifths of the raw fire sale discount, in all ten MSAs.

Assets that are redeployable (i.e., have alternative uses) tend to have higher liquidation values (Williamson, 1988). Single-family housing is, in theory, highly redeployable. Owner-occupied houses can easily be converted to rental properties and vice-versa.[3] During times of market-wide distress, owner-occupiers, who are typically the highest-value users, may be unable to purchase houses that were previously owner-occupied. In this case, these distressed houses must be sold to either local investors (second highest value user) or "deep pocket" outside investors (lowest value user). This implies that although owner-occupied housing can easily be redeployed, its liquidation price will be lower when purchased as an investment. Shleifer and Vishny (1992) define the difference between the liquidation price of redeployed assets and their value in best use as asset illiquidity.

Using a subsample of transactions where detailed occupancy information is available for the buyer, we further decompose the quality-adjusted fire sale discount into an asset illiquidity component and a residual liquidity component. We measure asset illiquidity by comparing the liquidation price paid by investors (redeployed) to the liquidation price paid by owner-occupiers (best use) for houses that were previously owner-occupied. We find asset illiquidity is a significant component of fire sale discounts. When owner-occupiers purchase previously owner-occupied fire sale properties, they sell for an average quality-adjusted fire sale discount of 11.8%. When investors purchase similar fire sale properties, they sell for

[3]Some restrictions do exist. For example, homeowners associations may limit or ban rental properties.

an additional 12.2% (24% total) quality-adjusted fire sale discount. This finding indicates that owner-occupied houses liquidated as investment properties sell for an average quality-adjusted fire sale discount that is twice the size of those that remain owner-occupied. This asset illiquidity exists when investors purchase fire sale properties for short-term (speculation) and long-term (rentals) investment horizons.

After decomposing the house quality and asset illiquidity components, we estimate that the residual liquidity component constitutes roughly one-half of raw fire sale discounts in housing markets. We interpret the residual liquidity component as the cost of immediacy or the cost of selling an asset quickly in an illiquid market (Franks, Seth, Sussman, & Vig, 2022; Shleifer & Vishny, 2011). We provide support for our interpretation by showing that houses liquidated via fire sales are sold 7 to 12 days faster on average than non-distressed houses.

1.1 Related Literature

This paper contributes to several strands of literature. First and foremost, our paper contributes to the literature on fire sale discounts in housing markets. The internet appendix provides a high-level review of over 40 studies whose sole focus is estimating fire sale discounts in housing markets. A few of these studies recognize the need to control for house quality when estimating fire sale discounts in housing markets (e.g., Campbell et al. (2011), Anenberg and Kung (2014)). However, previous attempts have been subject to measurement and data collection limitations because house quality is time-varying, difficult to quantify, and cost-prohibitive to collect at scale.[4] We overcome these limitations by quantifying the textual information in listing agents' property descriptions (i.e., public remarks) about the house. Unlike previous research, our approach allows us to reasonably approximate the condition and quality of every house in our sample at the time it is listed for sale.

Our approach adds to a growing literature in finance, economics, and related fields that use text as data. Recent studies use newspapers, patents, and documents to measure vari-

[4]One exception is Zhou, Yuan, Lako, Sklarz, and McKinney (2015), which includes time-varying property information from the Freddie Mac Appraisal Portal. However, their sample includes fewer than 13,000 transactions from 2012 and, due to the limited variability of property conditions in the data, only houses in excellent, good, or average condition.

ables that are conceptually understood but difficult to quantify in conventional data sources, including uncertainty (Baker, Bloom, & Davis, 2016), political risk (Hassan, Hollander, Van Lent, & Tahoun, 2019), innovation (Kelly, Papanikolaou, Seru, & Taddy, 2021), and climate-risk (Sautner, Van Lent, Vilkov, & Zhang, 2023). We use listing agents' written descriptions of a house provided in the public remarks field of multiple listing service (MLS) data (Liu, Nowak, & Smith, 2020; Nowak, Price, & Smith, 2021; Nowak & Smith, 2017, 2020; Shen & Ross, 2021). We contribute to the literature by introducing a new methodology that identifies and removes a vestigial set of tokens that are spurious predictors of the variable of interest. Although we focus on estimating fire sale discounts, researchers can use our computer-assisted text analysis algorithm to explore any topic.

Our paper also contributes to empirical research that estimates fire sale discounts in non-real estate settings. The fire sale liquidations in most of these studies differ from ours in that whole assets are not liquidated, assets are relatively homogeneous, unobserved asset quality is less of a concern, and the assets trade frequently (e.g., Coval and Stafford (2007), Ellul, Jotikasthira, and Lundblad (2011), Mitchell and Pulvino (2012), and Dinc, Erel, and Liao (2017)). Researchers have also examined fire sale discounts in the secondary market for used commercial aircraft (e.g., Pulvino (1998) and Franks et al. (2022)). Although aircraft liquidations are similar to housing in that they involve infrequently traded whole assets where unobserved asset quality is a concern, aircraft, by design, are mobile and can be redeployed anywhere across the world. Land, in contrast, is immobile, and it is often financially infeasible to move housing structures. These frictions make housing more susceptible to asset illiquidity during times of market-wide distress, thereby providing the ideal setting to examine the economic inefficiencies of fire sale liquidations.

In their review of the literature, Shleifer and Vishny (2011) note that the fire sale of an asset is a forced sale where the seller might wish for various reasons to sell the asset quickly. Despite the growing empirical evidence that fire sales exist, data limitations have prevented previous non-real estate studies from showing that assets sold via fire sales are sold quicker than those sold via non-distressed transactions. Unlike prior research, our data includes the date when a house is listed on the MLS and when it is taken off the market, allowing us to measure every listing's days on market. We find houses liquidated via fire sales are

sold 12 days faster on average than non-distressed houses, thereby providing support for the assertion in Shleifer and Vishny (1992) that the need to rapidly sell an asset represents a significant component of fire sale discounts.

Previous studies examining asset redeployability tend to focus on its role in capital structure outcomes such as debt maturity (Benmelech, 2009), cost of capital (Benmelech & Bergman, 2009; Ortiz-Molina & Phillips, 2014), investment (Kim & Kung, 2017), and leverage (Campello & Giambona, 2013). We contribute to the literature by examining the relationship between asset redeployability and liquidation values in residential housing markets. We show that asset illiquidity stemming from the redeployment of housing from owner-occupied to rental is a significant component of fire sale discounts in housing markets during times of market-wide distress. Our results strongly support the assertion in Shleifer and Vishny (1992) that liquidated assets are underpriced in recessions and that asset illiquidity is an important cost of leverage.

2 Data Overview

We estimate fire sale discounts using single-family detached housing transaction data from January 2000 through June 2023. We use MLS data from CoreLogic for ten large MSAs. We selected the MSAs based on the size of their population (Top 50 in the U.S.) and historical data availability (coverage starting in 2000). We also selected the MSAs to provide geographic diversification in terms of variation in location-specific attributes such as price, supply constraints, and housing stock. The ten MSAs include Boston, Massachusetts (BOS), Charlotte, North Carolina (CLT), Columbus, Ohio (COL), Detroit, Michigan (DET), Houston, Texas (HOU), Los Angeles, California (LA), Las Vegas, Nevada (LV), Minneapolis, Minnesota (MIN), Seattle, Washington (SEA), and Saint Louis, Missouri (STL).

The MLS data includes physical attributes (e.g., bedrooms, bathrooms), location attributes (e.g., street address, zip code), and transaction information (e.g., sale price, list date, sale date). We apply several filters to remove outliers and miskeyed information. The internet appendix provides a detailed description of each filter and tracks the number of observations each filter removes. After applying the filters, the sample includes just over 7

million transactions across the ten MSAs.

Table 1 presents descriptive statistics for the combined ten MSA sample. The internet appendix presents similar MSA-level descriptive statistics. Panel A of Table 1 reports summary statistics for the entire transaction sample. The average sale price across the 10 MSAs is $326 thousand. Approximately 13% of the transactions are distressed: 9% are fire sales and 4% are short sales.[5] Figure 1 plots the share of distressed transactions over time. Although distressed transactions only represent 13% of the sample, they account for more than 40% of transactions from 2008 to 2012.

Panels B, C, and D display summary statistics for the subsample of non-distressed, short sale, and fire sale transactions, respectively. As expected, the non-distressed transaction subsample (Panel B) has a higher average price than the two distressed subsamples (Panels C and D). Houses in the non-distressed subsample are, on average, older and larger than houses in the distressed subsamples. A comparison of the short sale (Panel C) and fire sale (Panel D) subsamples shows fire sales have a substantially lower average sales price than short sales.

The MLS data also includes listing agents' public remarks. The remarks are unstructured written descriptions of the house that are entered by the listing agent. The remarks summarize the listing agent's first-hand knowledge of the house's interior and exterior condition (i.e., quality) when the house is listed for sale on the MLS. The unstructured nature of the remarks allows agents to highlight important information about the house that may not be communicated through standard MLS fields.

Panel A of Table 2 provides an example of a public remark in the data. The remark conveys information about house quality (*needs TLC*). The remark also includes redundant information given the listing is identified as a fire sale (*bank owned REO*). Section 3.1 describes how we quantify the textual information in the public remarks while simultaneously dealing with the redundant information.

[5]Short sales preclude fire sale liquidations. A short sale occurs when the lender allows the delinquent borrower to sell their home for less than the outstanding mortgage balance.

3 Fire Sale Discounts

We estimate raw fire sale discounts in housing markets using the following linear hedonic model

$$p_{nt} = \beta_{FS} FS_{nt} + \beta_{SS} SS_{nt} + x'_{nt}\beta_x + \underbrace{\mu_n + \psi_{nt} + e_{nt}}_{u_{nt}} \tag{1}$$

Where p_{nt} is the log price of house n in time period t. The coefficients of interest, β_{FS} and β_{SS}, are scalar parameters indicating price differences associated with fire sale and short sale transactions relative to non-distressed housing transactions. The vector x_{nt} includes housing characteristics, financing, time, and location controls. We allow the housing characteristics to have non-linear effects on price and create indicator variables for 100-square-foot-wide bins of living space ($SQFT_{500-599,nt}, SQFT_{600-699,nt}, ...$), 10-year-wide bins of house age ($AGE_{0-9,nt}, AGE_{10-19,nt}, ...$), number of bedrooms, number of bathrooms, and number of acres in lot size. The financing controls identify cash transactions. The time and location controls include quarter-by-year-by-census tract fixed effects to control for heterogeneous market conditions across time and space. In Equation 1, μ_n captures unobserved time-invariant house quality that may be correlated with the regressors, ψ_{nt} captures unobserved time-varying house quality that may be correlated with the regressors, e_{nt} is an unobservable term uncorrelated with any of the regressors, and u_{nt} is the sum of these unobservables.

Equation 1 highlights the need to control for house quality when estimating fire sale discounts in housing markets. Controlling for μ_n is straightforward using house fixed effects and a sample of repeat-sale transactions. However, controlling for ψ_{nt} is difficult without time-varying measures of house quality. When not controlling for ψ_{nt}, it is not clear how to interpret either the absolute or relative values of $\hat{\beta}_{FS}$ and $\hat{\beta}_{SS}$. In absolute terms, $\hat{\beta}_{FS}$ captures quality, fire sale-related transaction costs, stigmas associated with fire sale transactions, and all other price effects unrelated to quality. In relative terms, $\hat{\beta}_{FS} < \hat{\beta}_{SS}$ may reflect the lower quality of fire sale properties or higher costs unrelated to quality.

3.1 House Quality Decomposition

To the extent that default causes deferred maintenance, the house quality component is time-varying. As noted earlier, we use textual information in MLS remarks to control for house quality. Identifying the set of quality-related words (dictionary) presents challenges as text is, by nature, high-dimensional. Many machine-learning methods that identify a dictionary using predictive criteria are prone to false positives. Our setting is no different. Many words in the property descriptions that predict price are redundant given the fire sale indicator variable (e.g., phrases such as *bank owned* or *foreclosure*). In contrast, manually created dictionaries rely on the researcher's knowledge about the economic environment. This approach is prone to false negatives as humans excel when asked to associate words to topics but perform poorly when asked to recall an exhaustive list of words associated with a topic (King, Lam, & Roberts, 2017). To resolve this human-versus-machine dilemma, we adapt the iterative procedure in King et al. (2017). Our approach is transparent, flexible, and can be generalized to any hedonic model where text is available and unobserved quality is a concern.

Much of the relevant information about house quality is not codified in a standard MLS data field. For example, many fire sale listings have remarks indicating the house *needs tlc*, is a *handyman's delight*, or has some type of *damage*. If we can control for the presence of these and other important words or phrases, then we can mitigate the bias caused by ψ_{nt}. Following Liu et al. (2020), we use a bag-of-words approach and seek a dictionary, $\mathcal{S}$, or set of tokens $s = 1, ..., S$ such that

$$\mu_n + \psi_{nt} = \sum_{s \in \mathcal{S}} w_{nt,s}\theta_s + r_{nt} \tag{2}$$

In Equation 2, $w_{nt,s} = 1$ if token s appears in the remarks for property n sold at time t, θ_s is the implicit price for token s, and r_{nt} is a approximation error term uncorrelated with any regressors in Equation 1. If $\mathcal{S}$ is known, we can include $w_{nt,s}$ as additional regressors in Equation 1, thereby mitigating any bias resulting from either $\mu_n + \psi_{nt}$.

This dictionary approach mitigates the omitted-variable bias in a way that is not possible with other methods previously used in the literature. For instance, property-specific fixed

effects control for μ_n in Equation 2. However, ψ_{nt} will still generate a bias if house quality deteriorates during the foreclosure process, as suggested in Lambie-Hanson (2015) and Melzer (2017). Likewise, including census tract-by-time period fixed effects will control for ψ_{nt} assuming every house transacting in a given census tract and time period is of similar quality. However, in all ten MSAs, we find evidence of renovations and fix-and-flips that transact alongside fire sales in the same census tract and quarter.

Identifying $\mathcal{S}$ requires identifying a subset of important tokens from among the thousands of unique tokens appearing across all listings in our sample. We cast this dictionary-creation problem as a high-dimensional variable-selection problem and use the post double-selection estimator for statistical inference in high-dimensional models (Belloni, Chernozhukov, & Hansen, 2014). Given the large number of candidate tokens, the variable-selection procedure will almost surely make mistakes. Despite this reality, the double-selection approach provides for valid asymptotic inference by selecting tokens that are either strong predictors of price or strong predictors of distressed transactions. By construction and appealing to the formula for omitted-variable bias, any token not in the estimated dictionary has minimal effect on the coefficient estimates $\hat{\beta}_{FS}$ and $\hat{\beta}_{SS}$.

We begin by creating indicator variables for the $M = 2,000$ most frequent tokens in the remarks where $w_{nt,m} = 1$ if token m is in the remarks when house n is listed at time t, and $w_{nt,m} = 0$ otherwise. Next, we perform variable selection by applying the heteroscedastic least absolute shrinkage and selection operator (LASSO) to the hedonic model in Equation 1 and solve

$$\min_{b_{FS},b_{SS},b_x,\theta} \sum (p_{nt} - FS_{nt}b_{FS} - SS_{nt}b_{SS} - x_{nt}b_x - \sum_m w_{nt,m}\theta_m)^2 + \lambda \sum_m \upsilon_m |\theta_m| \qquad (3)$$

where υ_m is a token-specific penalty parameter that controls for heteroscedasticity in the hedonic equation. Define $\{\hat{b}_{FS}, \hat{b}_{SS}, \hat{b}_x, \hat{\theta}\}$ as the solution to Equation 3. The objective in Equation 3 is a penalized regression where the ℓ_1 penalty on θ yields a solution where $\hat{\theta}_m = 0$ for some m. Define $\hat{\mathcal{S}}_p \subseteq \{1, ..., M\}$ as the index of the coefficients such that $\hat{\theta}_m \neq 0$. We can interpret the set of tokens in $\hat{\mathcal{S}}_p$ as the dictionary of tokens that are strong predictors of price. Unfortunately, $\hat{\mathcal{S}}_p$ may omit tokens that are not strong predictors of price but are,

nonetheless, important for resolving the omitted-variable bias in $\hat{\beta}_{FS}$ and $\hat{\beta}_{SS}$.

To immunize against variable-selection mistakes, we perform a double-selection step. We use LASSO two more times, replacing p_{nt} on the left-hand side with FS_{nt} and SS_{nt}. Doing so yields dictionaries $\hat{\mathcal{S}}_{FS}$ and $\hat{\mathcal{S}}_{SS}$, respectively. Tokens in these dictionaries are not necessarily strong predictors of price but are strong predictors of FS_{nt} or SS_{nt}. The union $\hat{\mathcal{S}}_2 = \hat{\mathcal{S}}_p \cup \hat{\mathcal{S}}_{FS} \cup \hat{\mathcal{S}}_{SS}$ is a set of $|\hat{\mathcal{S}}_2| = \hat{S}$ tokens that are either strong predictors of price or distressed housing transactions.

The post double-selection step estimates $\hat{\beta}_{FS}$ and $\hat{\beta}_{SS}$ using least squares with the tokens in $\hat{\mathcal{S}}_2$ included as controls. Asymptotic statistical inference is valid so long as $\mu_n + \psi_{nt}$ can be approximated using a dictionary $\mathcal{S}$ of $S \ll M$ tokens.[6] Given the size limits of the remarks section and agents' incentive to convey the most salient information about the property, it is reasonable to believe that such a dictionary exists. Ultimately, this assumption is not verifiable. However, Liu et al. (2020) provide simulation evidence indicating the post double-selection estimator performs well in a similar environment. Simulation evidence in our setting is available upon request.

Text is unstructured and its quantification requires researchers to make decisions ex ante that must be evaluated ex post (Gentzkow, Kelly, & Taddy, 2019). We provide two means of validating our dictionary for the purposes of statistical inference on the fire sale coefficient: one theoretical and one based on inspecting coefficient estimates. The theoretical validation is based on the double-selection procedure in Belloni et al. (2014). Under suitable assumptions on the tokens and their implicit prices, the dictionary includes a set of tokens sufficient to mitigate the effect of unobserved quality on the fire sale discount up to an order of magnitude. The inspection validation is similar to Gelbach (2016) and Nowak and Smith (2020) in that we validate the dictionary using not only the implicit prices of tokens but also the set of controls used in the regression. We find tokens with negative implicit prices (houses that *need work* or *require repairs*) and tokens with positive implicit prices (houses with *upgrades* and *custom features*) both play a role. Specifically, deferred maintenance (lower quality) of fire sale properties and improvements and renovations (higher

[6] We also need the dictionary to sufficiently approximate FS_{nt} and SS_{nt} and technical assumptions on the relationship between the regressors.

quality) to nearby non-distressed properties affect the quality-adjusted fire sale discount. Other validation methods, including human audits (Baker et al., 2016), are possible but cost-prohibitive given the size and scope of our data.

3.2 Vestigial Predictor Problem

When creating $\hat{\mathcal{S}}_2$, it is important to only include tokens that convey new information about house quality and not include any redundant tokens that indicate a fire sale or short sale. One possibility is to examine $\hat{S}_{FS}$ and $\hat{S}_{SS}$, remove problematic tokens, and repeat the double-selection procedure. However, when a group of highly correlated predictors is considered, LASSO selects only one from among the group, thereby masking the effects of the other highly correlated predictors (Zou & Hastie, 2005).

Take, for example, the remark in Panel A of Table 2. The remark includes the phrases *this is a bank owned REO* and *please include REO addendum*, where *bank owned* and *REO* clearly indicate a fire sale in the remark. Not surprisingly, we find that the *REO* token is a strong predictor of fire sale transactions, but the *addendum* token is a much weaker predictor of fire sale transactions. However, if we clean the remarks by removing *REO* and then repeat the procedure, *addendum* becomes a strong predictor.

Although it is tempting to repeat this iterative cleaning process, we find purging redundant words and phrases from the remarks leaves behind a vestigial set of tokens that are spurious predictors of the fire sale indicator variable. For example, after purging *REO* and *addendum* from the remarks, we find *include* is, surprisingly, a strong predictor of fire sales as most remarks for fire sales indicate the buyer must *include [the] REO addendum*. This is problematic because the token for *include* will have unknown effects when included as a regressor in the hedonic regression. We refer to this as the vestigial predictor problem.

The example above highlights complexities in cleaning the remarks token by token. Instead of using an iterative process, we address the complexities using a text-cleaning algorithm that removes entire sentences that contain at least one redundant token. Algorithm 1 presents our computer-assisted text-cleaning procedure. Panel B of Table 2 displays the remark from Panel A after we clean it using Algorithm 1. The internet appendix provides three additional applications of the algorithm.

The cleaning procedure in Algorithm 1 all but eliminates the vestigial predictor problem. Moreover, we find the algorithm requires no more than a few rounds to identify redundant tokens in each of the ten MSAs. Also, despite this more aggressive cleaning approach, we find no significant impact on the number of tokens selected or their impact on price. That said, removing entire sentences may inadvertently remove non-redundant house quality tokens. For example, Panel B of Table 2 shows the algorithm removed the sentence "this is a bank owned reo property in good condition but needs tlc." Although the sentence contains redundant tokens (e.g., *bank owned reo*), it also includes information about the quality of the house (e.g., *needs tlc*). We address this issue by using the full set of cleaned remarks to create the token dictionary and then using the full set of raw remarks for the post-double-selection regression analysis. This ensures redundant tokens, here *bank owned reo*, are not included in the dictionary but allows the non-redundant tokens in $\hat{\mathcal{S}}_2$ that appear in the same sentences as *bank owned reo*, here *needs TLC*, to affect the post-selection regression results. Panel C of Table 2 displays the remark from Panels A and B after we reintroduce the full remark for the post-double-selection regression.

3.3 Quality-adjusted Fire Sale Discounts

We estimate Equation 1 for the combined ten MSA transaction sample as well as each individual MSA. Table 3 displays fire sale coefficient estimates for the combined ten MSA transaction sample. Panel A uses the full transaction sample and Panel B uses a subsample of houses that transacted at least twice (i.e., repeat sales) during the study period. The dependent variable in every column is the log of sales price. The fire sale coefficient estimates are relative to non-distressed transactions. Every column includes house controls (housing characteristics in Panel A or house fixed effects in Panel B), financing controls (i.e., an indicator for cash purchases), and quarter-by-year-by-census tract fixed effects.

We estimate Equation 1 with and without controls for the tokens in $\hat{\mathcal{S}}_2$. Column 1 of Table 3 does not include tokens that control for house quality. We interpret the fire sale coefficient estimates in column 1 as the raw fire sale discount. Columns 2 through 5 include tokens that control for house quality. We interpret the fire sale coefficient estimates that include the tokens as the quality-adjusted fire sale discount. We interpret the difference between the

two coefficient estimates as the quality component of fire sales in housing markets.

Columns 2 and 4 of Table 3 include tokens in $\hat{\mathcal{S}}_p$ from a single-selection variable selection process. Whereas columns 3 and 5 include tokens in $\hat{\mathcal{S}}_2$ from a double-selection variable selection process. We report results using both the cleaned remarks (e.g., Panel B of Table 2) in columns 2 and 3 and the cleaned full remarks (e.g., Panel C of Table 2) in columns 4 and 5 for informational purposes. For the sake of brevity, we focus our discussion on the post-double-selection results for the cleaned full remarks in column 5.

In Panel A of Table 3, the raw fire sale coefficient estimate in column 1 indicates houses liquidated via fire sales sell for an average discount of 23% relative to non-distressed houses.[7] Adjusting for house quality in column 5 yields an average quality-adjusted fire sale discount of 16.5%. The 6.5 percentage point decrease indicates that house quality constitutes over one-quarter of the raw fire sale discount. Figure 2 plots the corresponding raw and quality-adjusted fire sale coefficient estimates for each of the ten MSAs. Raw fire sale discounts range from 12.5% in Los Angeles to 30.9% in Saint Louis. Adjusting for house quality reduces the average fire sale discount to 9.6% in Los Angeles and 18.2% in Saint Louis.

The 12.7 percentage point price difference in Saint Louis represents 41% of the raw fire sale price discount. In other words, over two-fifths of the raw fire sale price discount can be attributed to lower house quality. The remaining 59% represents the quality-adjusted fire sale price discount. In Los Angeles, the 2.9 percentage point price difference represents 23% of the raw fire sale price discount. This implies approximately one-quarter of the raw fire sale price discount can be attributed to lower house quality. Figure 2 shows the magnitude of the raw and quality-adjusted fire sale discounts vary considerably across the 10 MSAs. That said, we find house quality is a sizeable component in all 10 MSAs, representing one-quarter to two-fifths of the raw fire sale discount.

In Panel B of Table 3, the raw repeat-sales fire sale coefficient estimate in column 1 indicates fire sale houses sell for an average discount of 23.5% relative to non-distressed houses. This repeat-sales estimate is similar to the full sample estimate in column 1 of Panel A. Note, however, that the raw fire sale coefficient estimate in column 1 of Panel B is

[7]The dependent variable in Equation 1 is the log of sales price. Accordingly, we calculate the fire sale price discounts as the exponent of the coefficient estimate minus one.

markedly different than the quality-adjusted fire sale coefficient estimate in column 5 of Panel A. These findings support our earlier assertions that (i) controlling for μ_n in Equation 1 is straightforward using house fixed effects and (ii) controlling for ψ_{nt} is difficult without time-varying measures of house quality but important for unbiased fire sale discount estimates. After adjusting for house quality in column 5 of Panel B, we find an average quality-adjusted repeat-sales fire sale discount of 15.9%. The 7.6 percentage point decrease indicates that house quality constitutes almost one-third of the raw repeat-sales fire sale discount.

3.4 Quality-adjusted Short Sale Discounts

Although not the primary focus of this paper, we compare fire sale liquidations to short sale liquidations to further highlight the magnitude of the house quality component. In column 1 of Table 3, we estimate an average raw short sale discount of 14.4% relative to non-distressed transactions. Adjusting for house quality in column 5 yields an average quality-adjusted short sale discount of 11%. The 3.4 percentage point decrease indicates that house quality constitutes almost one-quarter of the raw short sale discount.

The quality-adjusted fire sale and short sale coefficient estimates in Table 3 provide two important insights into the difference between the liquidation channels. First, houses liquidated via a fire sale sell for steeper discounts than houses liquidated via short sales, even after adjusting for quality. Short sales sell for more than fire sales because the short sale liquidation process aligns the interests of the delinquent borrower with the loan holder. Delinquent borrowers have a vested interest in maintaining their house during the short sale listing because they are likely living in the house and consuming its services. Maintaining the quality of the house also increases the probability of a successful short sale, thereby averting foreclosure and its negative effect on the delinquent borrower's credit.

Second, including the tokens significantly reduces the estimated price difference between fire sale and short sale liquidations from 10 percentage points in column 1 to 6.2 percentage points in column 5. The reduced spread indicates almost two-fifths of fire sales' steeper discount (relative to short sales) results from further declines in quality that occur during the fire sale liquidation process. This further decline in quality is not surprising. In all probability, the distressed homeowner abandons the house after foreclosure, and if they do

not leave, they are soon evicted. It is common for distressed homeowners to cause damage and leave the house in disrepair under such circumstances.[8]

Figure 3 plots the raw and quality-adjusted short sale coefficient estimates for each of the ten MSAs. The average raw short sale price discount ranges from 9.3% in Charlotte to 16.7% in Seattle. Adjusting for house quality reduces the average short sale price discount to 6.5% in Charlotte and 12.4% in Seattle. The 2.8 and 4.3 percentage point price differences in Charlotte and Seattle represent 28-42% of their respective raw short sale price discounts. In other words, over one-quarter of the raw short sale price discount can be attributed to lower house quality. The remaining three-quarters represents the quality-adjusted short sale price discount. Overall, the findings in this section suggest lenders (or servicers acting on their behalf) should consider alternative liquidation methods, such as short sales, to incentivize the continued maintenance of distressed properties.

3.5 Token Adjustments

This section provides a method for validating the dictionary by demonstrating how including tokens as additional regressors in Equation 1 changes the fire sale and short sale coefficient estimates. An alternative approach similar to that used in the text analysis literature is to validate the dictionary based on the sign or magnitude of the token coefficient estimates. Although informative, in a hedonic model, this approach is inherently based on the predictive power of the tokens whereas our focus is on the debiasing properties of the tokens. We demonstrate that changes to the fire sale discount when including tokens as controls come not only from the token coefficient estimates but also from the relative frequency of the tokens across the controls. We also provide a graphical means of visualizing the contribution of each token to the adjusted fire sale discount.

For simplicity, we focus on a single indicator for distressed sales, but the results generalize to the fire sale and short sale indicators. We also exclude time, location, and house controls. This is also without loss of generality as we can appeal to the Frisch-Waugh-Lovell Theorem

[8] Although short sales benefit lenders and MBS investors, research finds free rent during foreclosure causes more borrowers to select foreclosures (Zhang, 2019) and securitized loan servicers have a financial incentive to pursue fire sale liquidations (Mehrotra, Nowak, & Smith, 2023).

and interpret the distressed sale vector and price vector as residuals after partialling out the effects of these controls. For transaction $i = 1, ..., I$, p_i is the log sale price and d_i is an indicator for distressed sales. Define $\mathbf{P} = (p_1, ..., p_I)'$ as the $I \times 1$ vector of prices, $\mathbf{D} = (d_1, ..., d_I)'$ as the $I \times 1$ vector of distressed indicators, and $\mathbf{X} = [\mathbf{1}, \mathbf{D}]$ as the $I \times 2$ matrix with a vector of 1s in the first column and D in the second column.

Assuming a linear hedonic model, the baseline distressed sale coefficient estimate, $\hat{\tau}^0$, is given by

$$\hat{\tau}^{baseline} = [\mathbf{X}'\mathbf{X}]^{-1}\mathbf{X}'\mathbf{p} = \bar{p}_{d=1} - \bar{p}_{d=0}$$

Here $\bar{p}_{d=1}$ is the average log price of distressed transactions and $\bar{p}_{d=0}$ is the average log price of non-distressed transactions.

Time-varying unobservables can bias $\hat{\tau}^{baseline}$ if distressed houses are lower quality than non-distressed houses. The tokens in $\hat{\mathcal{S}}_2$ control for both deferred maintenance prior to distress and any possible improvements made by the subsequent purchaser of the distressed property. Define $\mathbf{W}$ as the $I \times |\hat{\mathcal{S}}_2|$ matrix of indicator variables for the tokens $s \in \{1, ..., |\hat{\mathcal{S}}_2|\}$ in $\hat{\mathcal{S}}_2$. Define $\hat{\tau}^{token}$ as the distressed sale coefficient estimate when we include W in the hedonic regression and $\hat{\tau}^{adj} = \hat{\tau}^{token} - \hat{\tau}^{baseline}$ as the difference between the estimate that includes the tokens and the baseline estimate. If $\hat{\mathcal{S}}_2$ only includes a single token, evaluating the impact of this token on $\hat{\tau}^{adj}$ is straightforward. However, even after using the LASSO procedure to reduce the dimension of the tokens, $\hat{\mathcal{S}}_2$ contains a few hundred tokens that each contribute to $\hat{\tau}^{adj}$.

We use a partitioned-regression result to evaluate how each token in $\hat{\mathcal{S}}_2$ contributes to $\hat{\tau}^{adj}$. The first-order conditions for the regression that includes $\mathbf{W}$ satisfy

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{W} \\ \mathbf{W}'\mathbf{X} & \mathbf{W}'\mathbf{W} \end{bmatrix} \begin{bmatrix} \hat{\tau}^{token} \\ \hat{\theta}^{token} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{p} \\ \mathbf{W}'\mathbf{p} \end{bmatrix} \tag{4}$$

Here, $\hat{\theta}^{token}$ is the $|\hat{\mathcal{S}}_2| \times 1$ vector of implicit prices for the tokens in $\hat{\mathcal{S}}_2$. Solving this systems

of equations for $\hat{\tau}^{token}$

$$\hat{\tau}^{token} = [\mathbf{X}'\mathbf{X}]^{-1}\mathbf{X}'\mathbf{p} - [\mathbf{X}'\mathbf{X}]^{-1}\mathbf{X}'\mathbf{W}\hat{\theta}^{token} \tag{5}$$

$$= \hat{\tau}^{baseline} + \hat{\tau}^{adj} \tag{6}$$

The second term in Equation 5 can be written as

$$\sum_s [\mathbf{X}'\mathbf{X}]^{-1}\mathbf{X}' w_s \hat{\theta}_s^{token} = \sum_s (\bar{w}_{s,d=1} - \bar{w}_{s,d=0})\hat{\theta}_s^{token} = \sum_s \hat{\pi}_s \hat{\theta}_s^{token} \tag{7}$$

where w_s is the s^{th} column of $\mathbf{W}$.

Equation 7 indicates token s can have an impact on $\hat{\tau}^{adj}$ if and only if $\hat{\pi}_s \neq 0$ *and* $\hat{\theta}_s^{token} \neq 0$. First, $\hat{\theta}_s$ is the implicit price for token s. Not surprisingly, when the implicit price is 0, the token does not affect the price. Second, $\hat{\pi}_s = \bar{w}_{s,d=1} - \bar{w}_{s,d=0}$ is the difference in the frequency of token s between distressed and non-distressed sales. When there is no difference in the frequency of token s between these two groups, $\hat{\pi}_s = 0$, and token s cannot affect $\hat{\tau}^{adj}$ regardless of the implicit price of the token. Equation 7 clearly indicates the impact of token s on $|\hat{\tau}^{adj}|$ is increasing in $|\hat{\pi}_s|$. This result is intuitive, as it is generally accepted that distressed houses are more likely to be of lower quality than non-distressed houses.

Importantly, this result also implies that omitting tokens that appear more often in non-distressed sales, $\hat{\pi}_s > 0$, can also impact coefficient estimates. For example, even though distressed houses are less likely to be recently renovated, omitting tokens that control for upgrades to the house can bias fire sale estimates. Suppose many fix-and-flip transactions occur alongside distressed transactions. In that case, the distressed coefficient estimate that controls only for the time-varying quality of distressed sales may still have a significant bias. The double-selection estimation and the availability of remarks for all transactions in the data, of both high and low house quality, address these concerns by controlling for all relevant tokens in both distressed and non-distressed transactions.

Figure 4 provides insight into how tokens with non-negligible effects affect $\hat{\beta}_{FS}$ and $\hat{\beta}_{SS}$ for four of the MSAs in this study. The internet appendix provides similar token adjustment

plots for the remaining six MSAs. The figure indicates tokens with negative and positive implicit prices impact short sale and fire sale price coefficient estimates. Many of the tokens with negative implicit prices identify distressed houses that *need work*, *could use some tlc*, or *require repairs*. In contrast, tokens with positive implicit prices identify *granite countertops*, *upgrades*, and *custom features*. Accordingly, including the tokens in the post double-selection estimation helps control for deferred maintenance and disrepair, which are common in distressed houses, and improvements and upgrades made to the entire housing stock. The effect of the tokens on the fire sale and short sale coefficient estimates highlights the need to control for house quality in both distressed and non-distressed transactions.

4 Asset Illiquidity

Although owner-occupied houses can, in most cases, easily be converted to rentals, their fundamental value may differ when redeployed. We examine how asset redeployability affects liquidation values in housing markets by decomposing the quality-adjusted fire sale discounts in Section 3.3 into an asset illiquidity component and a residual liquidity component. Following Shleifer and Vishny (1992), we define asset illiquidity as the difference between the liquidation price of redeployed assets and their value in best use. In our setting, we can approximate the asset illiquidity component by comparing the price of owner-occupied houses that are liquidated to investors (i.e., redeployed) relative to those that are liquidated to owner-occupiers.

We estimate the following linear hedonic model and then include interactions between the $Redeploy_{k,nt}$ and distressed indicator variables to capture the difference between the liquidation price of redeployed assets and their value in best use.

$$p_{nt} = \sum_{k} \gamma_{Redeploy,k} Redeploy_{k,nt} + \beta_{FS} FS_{nt} + \beta_{SS} SS_{nt} + x'_{nt}\beta_x + e_{nt} \tag{8}$$

In Equation 8, $Redeploy_{k,nt}$ is a series of indicators comparing redeployability type $k \in \{Own\text{-}to\text{-}Invest,\ Invest\text{-}to\text{-}Own,\ Invest\text{-}to\text{-}Invest\}$. We construct the $Redeploy_{k,nt}$ indicator variables by comparing the previous owner's tenure to that of the buyer. For example,

Own-to-Invest equals one when a house that was previously owner-occupied is purchased by an investor and zero otherwise. To avoid perfect multicollinearity, we omit the indicator variable for owner-occupied houses that remain owner-occupied (*Own-to-Own*) after they are purchased.

We identify the tenure of the previous owner and buyer using several data fields.[9] First, we use the "Buyer Occupancy Code" in CoreLogic. If the tenure is unknown, we update it using the "Buyer Corporate Indicator" field in CoreLogic. If the tenure is still unknown, we use the "Occupancy" field in HMDA for financed transactions that we match. Finally, if the tenure is still unknown, we compare the owner's mailing address to the house's street address using an approximate string matching technique that allows the addresses to differ slightly (e.g., Street vs. St.).

We drop the transactions where the tenure status of the current owner is unavailable, leaving over 4.6 million transactions. We identify the tenure of the previous owner in just under 1.5 million of the remaining transactions. When running the empirical analysis we assume the remaining 3.1 million transactions were previously owner-occupied. The internet appendix shows that this assumption does not materially impact our results. Specifically, we show our results are similar using the subsample of 1.5 million transactions where we can reliably identify the previous owner's tenure.

4.1 Asset Illiquidity Results

Table 4 reports coefficient estimates from Equation 1 for a subsample of 4.6 million transactions where we can reliably identify the tenure of the buyer. The dependent variable in every column is the log of sales price. The fire sale coefficient estimates are relative to non-distressed transactions. Every column includes housing characteristics, financing controls, and quarter-by-year-by-census tract fixed effects. The fire sale coefficient estimates in columns 1 and 2 of Table 4 are very comparable to the estimates for the full sample of 7 million transactions in columns 1 and 5 of Table 3. This suggests the remaining subsample analysis we perform in columns 3 to 6 of Table 4 should generalize to the full sample.

[9]For fire sale properties, the previous owner is not the seller. The previous owner is the party that was foreclosed on.

Columns 3 and 4 of Table 4 report coefficient estimates from Equation 8 where we introduce a series of indicator variables identifying whether a house has been redeployed. The redeploy coefficient estimates are relative to previously owner-occupied houses that were not redeployed after being purchased (*Own-to-Own*). Column 4 includes the tokens in $\hat{\mathcal{S}}_2$, whereas column 3 does not. The *Own-to-Invest* coefficient estimate in column 3 indicates houses purchased by investors that were previously owner-occupied sell, on average, for 4.6% less than similar houses that are purchased by owner-occupiers.

Not all redeployed houses sell for a discount. Houses that were previously investment properties that were redeployed as owner-occupied (*Invest-to-Own*) sell, on average, for 1.2% more than houses purchased by owner-occupiers that were previously owner-occupied. We also find that houses purchased by investors that were previously owned by other investors sell, on average, for 1.6% less than previously owner-occupied houses purchased by owner-occupiers. After adjusting for quality, the *Own-to-Invest* fire sale discount in column 4 decreases from 4.6% to 3.1%, indicating that investors purchase previously owner-occupied houses of lower quality than those purchased by owner-occupiers.

Next, we examine the effect of asset redeployability on liquidation values by modifying Equation 8 to include interactions between the redeploy and distressed sale indicator variables. Column 6 of Table 4 includes the tokens in $\hat{\mathcal{S}}_2$, whereas column 5 does not. We interpret the difference between the quality-adjusted fire sale coefficient estimates in columns 6 and 2 as the asset illiquidity component of fire sale discounts.

The fire sale coefficient estimate in column 6 of Table 4 is significantly different than both the raw fire sale discount in column 1 and the quality-adjusted fire sale discount in column 2. Taken together, the fire sale coefficient estimates in columns 1, 2, and 6 indicate that asset illiquidity represents roughly one-quarter of the raw fire sale discount and one-third of the quality-adjusted fire sale discount. A comparison of the quality-adjusted fire sale discounts for owner-occupied houses that were redeployed further highlights the economic significance of the asset illiquidity component. In column 6, the quality-adjusted discount for *Own-to-Invest* fire sale transactions is over twice the size as the quality-adjusted fire sale discount for *Own-to-Own* transactions (24.0% vs. 11.8%).

One potential concern is that investors and owner-occupiers may purchase houses via fire

sales that differ in quality. In this case, the asset illiquidity component will capture both the effect of the redeployment to a lower-value user as well as the lower (or higher) quality of the houses that are redeployed. We can rule out this concern by comparing the coefficient estimates for the raw and quality-adjusted fire sale interaction terms in columns 5 and 6 of Table 4. We find the coefficient estimates are fairly similar, indicating houses that are redeployed (*Own-to-Invest*) during fire sales are similar quality as houses that remain owner-occupied (*Own-to-Own*) during fire sales. This finding indicates that our asset illiquidity component estimate is mostly unrelated to house quality, and confirms that we capture the difference between the liquidation price of redeployed houses and their value in best use.

Figure 5 plots the corresponding quality-adjusted *Own-to-Own* and *Own-to-Invest* fire sale coefficient estimates for each of the ten MSAs. The quality-adjusted *Own-to-Own* fire sale discounts - which represent the residual liquidity component of fire sales - range from 7.5% in Los Angeles to 18.3% in Columbus. The corresponding quality-adjusted *Own-to-Invest* fire sale discounts are 15% in Los Angeles and 27.6% in Columbus. The difference between the two coefficient estimates ranges from 5.4 percentage points in Detroit to 16.7 percentage points in Charlotte. Overall, we find asset illiquidity is a significant component in all ten MSAs, with investors receiving a discount that is 1.3 to 2.8 times the size of the discount received by owner-occupiers when they purchase previously owner-occupied housing in fire sale transactions.

4.2 Investment Horizon

Table 5 examines whether the asset illiquidity component varies by investor type. We further delineate the $Redeploy_{nt}$ indicator variables in Equation 8 to identify whether the investor has a short-term or long-term investment horizon. The indicator variable *Own-to-ST Invest* equals one if an investor purchases a previously owner-occupied property and holds it for less than three years and zero otherwise. Whereas the indicator variable *Own-to-LT Invest* equals one if an investor purchases a previously owner-occupied property and holds it for three or more years and zero otherwise. We interpret investors with a short-term investment horizon as speculative investors and investors with a long-term investment horizon as landlords who rent the property.

The coefficient estimates in column 4 of Table 5 indicate that short-term (i.e., speculative) investors pay significantly less than long-term (i.e., rental) investors for previously owner-occupied houses, even after adjusting for house quality. That said, both long-term (6.9%) and short-term (2.0%) investors pay significantly less than owner-occupiers for previously owner-occupied houses. Turning our attention to the quality-adjusted coefficient estimates in column 6, we find that both short-term and long-term investors receive a fire sale discount that is roughly 2 to 2.4 times larger than owner-occupiers when they purchase previously owner-occupied housing.

5 Residual Liquidity Component

After decomposing the house quality and asset illiquidity components, we can approximate the residual liquidity component. We interpret the fire sale coefficient estimate in column 6 of Table 4 - which represents the quality-adjusted fire sale discount for owner-occupied houses that are not redeployed - as the residual liquidity component of fire sale discounts. We attribute the 11.8% discount, which is roughly one-half the size of the raw fire sale discount, to the cost of quickly liquidating houses via fire sales.

We provide support for our interpretation by comparing the days on market for fire sale transactions to non-distressed transactions. Days on market is the length of time, as measured by the number of days, it takes to sell a house after it is listed for sale on the MLS. Table 6 reports coefficient estimates from modified versions of Equations 1 and 8 that substitute days on market for log of sales price as the dependent variable. The fire sale coefficient estimates in columns 1 to 6 of Table 6 indicate that houses liquidated via fire sales are sold 7 to 12 days faster on average than non-distressed houses. This finding indicates that houses sold via fire sale are liquidated quickly, supporting our claim that the residual liquidity component represents a cost of immediacy.

Although not our primary focus, Table 6 also examines the days on the market for short sale liquidations relative to non-distressed transactions. Despite finding a sizeable residual liquidity price component for short sale liquidations in Table 4, the coefficient estimates in Table 6 indicate that houses liquidated via short sales sell over 50 days slower on average

than non-distressed houses. We attribute this finding mostly to procedural delays associated with short sale liquidations.[10] That said, there is also less immediacy for the lender because they have not yet taken ownership of the house. Accordingly, their holding costs are much lower for short sales than fire sales.

In Section 4.1, we find that houses that are redeployed from owner-occupiers to investors (*Own-to-Invest*) sell for a quality-adjusted fire sale discount that is roughly twice the size as the quality-adjusted fire sale discount for owner-occupied houses that are not redeployed (*Own-to-Own*). We attribute the steeper discount to asset illiquidity stemming from the fact that the house was redeployed to a lower value user. Alternatively, it is possible that a portion of the steeper discount can be explained by quicker liquidations afforded by investor purchases. In other words, if lenders value immediacy and investors can purchase houses faster than owner-occupiers, then a portion of the steeper discount for *Own-to-Invest* fire sale transactions may be related to quicker liquidations, not asset illiquidity.

The results in column 6 of Table 6 allow us to rule out the alternative explanation above. In column 6, the quality-adjusted coefficient estimate for *Own-to-Invest* fire sale transactions indicates that houses that are redeployed from owner-occupiers to investors take roughly 5 days longer on average to sell than similar owner-occupied houses that are not redeployed (*Own-to-Own*) during fire sale transactions. This finding implies the steeper fire sale discount for *Own-to-Invest* transactions in Table 4 can be attributed to asset illiquidity.

6 Conclusion

Understanding the components that contribute to fire sales in housing markets and their magnitudes is critical to formulating future housing crisis mitigation strategies. This section briefly summarizes our key findings and discusses their policy implications.

Using a computer-assisted text analysis algorithm, we find substandard house quality, stemming from deferred maintenance and reduced investment, represents a significant com-

[10]There are more parties involved in a short sale transaction than a non-distressed transaction making the process complicated. In a traditional non-distressed transaction, price negotiations happen between the buyer and seller, not the lender. Although short sales have longer days on market (off market date minus list date) they are faster than REOs in terms of overall liquidation times (off market date minus default date) and they do not delay the foreclosure process (Mehrotra et al., 2023).

ponent (roughly one-quarter) of raw fire sale discounts in housing markets. This is a concern because research shows disamenities and reduced prices associated with lower house quality spillover to neighboring properties (e.g., Campbell et al. (2011); Gerardi, Rosenblatt, Willen, and Yao (2015); Harding, Rosenblatt, and Yao (2009)). It is also a concern because home improvements are an important economic activity, representing between 1 to 2 percent of the United State's gross domestic product (Choi, Hong, & Scheinkman, 2014). Our findings suggest policies targeted at incentivizing home maintenance and improvements among homeowners at risk of default (or creditors who take ownership) can help reduce fire sale discounts while simultaneously stimulating the economy during times of market-wide distress.

We also find asset illiquidity is a significant component (roughly one-quarter) of raw fire sale discounts in housing markets. We estimate that previously owner-occupied houses that are redeployed as rentals sell for an average quality-adjusted fire sale discount twice as large as similar houses that remain owner-occupied. This finding suggests policies targeted at owner-occupiers, such as a first-time homebuyer policy (Berger, Turner, & Zwick, 2020), can help reduce the asset illiquidity component of fire sale discounts by increasing demand among high-value users. Alternatively, policies targeted at stimulating competition among investors, such as a tax break on rental income generated by fire sale properties, can also help reduce the asset illiquidity component of fire sale discounts.[11]

After decomposing house quality and asset illiquidity, we find residual liquidity remains a sizeable component (roughly one-half) of fire sale discounts in housing markets. This finding implies that policies aimed at addressing the house quality and asset illiquidity components of fire sale discounts could cut fire sale discounts in half. The equilibrium effect of the policies would likely be larger because the policies will also presumably attenuate foreclosure spillover effects and price-default spirals, and the residual liquidity component will be partially competed away as the market recovers.

[11] In the internet appendix, we perform a series of back-of-the-envelope calculations to show that a 5 to 10-year tax break on rental income generated by fire sale properties can theoretically increase investors' (low-value users) valuation to similar levels as owner-occupiers (high-value users), thereby offsetting the asset illiquidity component of fire sale discounts.

References

Anenberg, E., & Kung, E. (2014). Estimates of the size and source of price declines due to nearby foreclosures. *American Economic Review*, *104*(8), 2527–2551.

Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics*, *131*(4), 1593–1636.

Belloni, A., Chernozhukov, V., & Hansen, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies*, *81*(2), 608–650.

Benmelech, E. (2009). Asset salability and debt maturity: Evidence from nineteenth-century American railroads. *The Review of Financial Studies*, *22*(4), 1545–1584.

Benmelech, E., & Bergman, N. K. (2009). Collateral pricing. *Journal of Financial Economics*, *91*(3), 339–360.

Berger, D., Turner, N., & Zwick, E. (2020). Stimulating housing markets. *The Journal of Finance*, *75*(1), 277–321.

Campbell, J. Y., Giglio, S., & Pathak, P. (2011). Forced sales and house prices. *The American Economic Review*, *101*(5), 2108–2131.

Campello, M., & Giambona, E. (2013). Real assets and capital structure. *Journal of Financial and Quantitative Analysis*, *48*(5), 1333–1370.

Choi, H.-S., Hong, H., & Scheinkman, J. (2014). Speculating on home improvements. *Journal of Financial Economics*, *111*(3), 609–624.

Coval, J., & Stafford, E. (2007). Asset fire sales (and purchases) in equity markets. *Journal of Financial Economics*, *86*(2), 479–512.

Dinc, S., Erel, I., & Liao, R. (2017). Fire sale discount: Evidence from the sale of minority equity stakes. *Journal of Financial Economics*, *125*(3), 475–490.

Ellul, A., Jotikasthira, C., & Lundblad, C. T. (2011). Regulatory pressure and fire sales in the corporate bond market. *Journal of Financial Economics*, *101*(3), 596–620.

Franks, J. R., Seth, G., Sussman, O., & Vig, V. (2022). Decomposing fire sale discounts. *ECGI Working Paper Series in Finance #722/2022*.

French, K., Baily, M., Campbell, J., Cochrane, J., Diamond, D., Duffie, D., ... others (2010). The Squam Lake report: Fixing the financial system. *Journal of Applied Corporate Finance*, *22*(3), 8–21.

Gelbach, J. B. (2016). When do covariates matter? And which ones, and how much? *Journal of Labor Economics*, *34*(2), 509–543.

Gentzkow, M., Kelly, B., & Taddy, M. (2019). Text as data. *Journal of Economic Literature*, *57*(3), 535–74.

Gerardi, K., Rosenblatt, E., Willen, P. S., & Yao, V. (2015). Foreclosure externalities: New evidence. *Journal of Urban Economics*, *87*, 42–56.

Guren, A. M., & McQuade, T. J. (2020). How do foreclosures exacerbate housing downturns? *The Review of Economic Studies*, *87*(3), 1331–1364.

Harding, J. P., Rosenblatt, E., & Yao, V. W. (2009). The contagion effect of foreclosed properties. *Journal of Urban Economics*, *66*(3), 164–178.

Hassan, T. A., Hollander, S., Van Lent, L., & Tahoun, A. (2019). Firm-level political risk: Measurement and effects. *The Quarterly Journal of Economics*, *134*(4), 2135–2202.

Kelly, B., Papanikolaou, D., Seru, A., & Taddy, M. (2021). Measuring technological innovation over the long run. *American Economic Review: Insights*, *3*(3), 303–320.

Kim, H., & Kung, H. (2017). The asset redeployability channel: How uncertainty affects corporate investment. *The Review of Financial Studies*, *30*(1), 245–280.

King, G., Lam, P., & Roberts, M. E. (2017). Computer-assisted keyword and document set discovery from unstructured text. *American Journal of Political Science*, *61*(4), 971–988.

Lambie-Hanson, L. (2015). When does delinquency result in neglect? Mortgage distress and property maintenance. *Journal of Urban Economics*, *90*, 1–16.

Liu, C. H., Nowak, A. D., & Smith, P. S. (2020). Asymmetric or incomplete information about asset values? *The Review of Financial Studies*, *33*(7), 2898–2936.

Mehrotra, A. A., Nowak, A. D., & Smith, P. S. (2023). The effect of mortgage securitization on asset liquidation decisions. *SSRN Working Paper #3957658*.

Melzer, B. T. (2017). Mortgage debt overhang: Reduced investment by homeowners at risk of default. *The Journal of Finance*, *72*(2), 575–612.

Mitchell, M., & Pulvino, T. (2012). Arbitrage crashes and the speed of capital. *Journal of Financial Economics*, *104*(3), 469–490.

Nowak, A. D., Price, B. S., & Smith, P. S. (2021). Real estate dictionaries across space and time. *The Journal of Real Estate Finance and Economics*, *62*, 139–163.

Nowak, A. D., & Smith, P. S. (2017). Textual analysis in real estate. *Journal of Applied Econometrics*, *32*(4), 896–918.

Nowak, A. D., & Smith, P. S. (2020). Quality-adjusted house price indexes. *American Economic Review: Insights*, *2*(3), 339–56.

Ortiz-Molina, H., & Phillips, G. M. (2014). Real asset illiquidity and the cost of capital. *Journal of Financial and Quantitative Analysis*, *49*(1), 1–32.

Pulvino, T. C. (1998). Do asset fire sales exist? An empirical investigation of commercial aircraft transactions. *The Journal of Finance*, *53*(3), 939–978.

Sautner, Z., Van Lent, L., Vilkov, G., & Zhang, R. (2023). Firm-level climate change exposure. *The Journal of Finance*, *78*(3), 1449–1498.

Shen, L., & Ross, S. (2021). Information value of property description: A machine learning approach. *Journal of Urban Economics*, *121*, 103–299.

Shleifer, A., & Vishny, R. (1992). Liquidation values and debt capacity: A market equilibrium approach. *The Journal of Finance*, *47*(4), 1343–1366.

Shleifer, A., & Vishny, R. (2011). Fire sales in finance and macroeconomics. *Journal of Economic Perspectives*, *25*(1), 29–48.

Williamson, O. E. (1988). Corporate finance and corporate governance. *The Journal of Finance*, *43*(3), 567–591.

Zhang, C. (2019). A shortage of short sales: Explaining the underutilization of a foreclosure alternative. *Federal Reserve Board of Philadelphia Working Paper 19-13*.

Zhou, H., Yuan, Y., Lako, C., Sklarz, M., & McKinney, C. (2015). Foreclosure discount: definition and dynamic patterns. *Real Estate Economics*, *43*(3), 683–718.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (statistical methodology)*, *67*(2), 301–320.

Tables and Figures

Table 1: Summary Statistics

	Panel A: Full sample		Panel B: Non-distressed		Panel C: Short sale		Panel D: Fire sale	
	mean	stdev	mean	stdev	mean	stdev	mean	stdev
Price (000s)	326.03	259.29	346.91	264.80	227.50	167.82	157.88	134.58
Days on market	63.84	84.21	59.83	79.83	129.52	126.44	75.75	91.06
Sqft (000s)	1.90	0.72	1.92	0.73	1.81	0.68	1.70	0.65
Age	38.71	27.39	38.96	27.45	34.93	26.43	37.77	27.07
Acres	0.31	0.52	0.32	0.53	0.24	0.44	0.26	0.50
Bedrooms	3.32	0.78	3.33	0.78	3.30	0.77	3.22	0.76
Bathrooms	2.37	0.88	2.39	0.88	2.37	0.84	2.18	0.85
Cash purchase	0.11	-	0.09	-	0.20	-	0.25	-
Fire sale	0.09	-	0.00	-	0.00	-	1.00	-
Short sale	0.04	-	0.00	-	1.00	-	0.00	-
Observations		7,091,684		6,209,518		267,637		614,529

Note: Table 1 reports summary statistics for the combined ten MSA transaction sample. Panel A of Table 1 reports summary statistics for the entire transaction sample. Panel B reports the same summary statistics for a subsample of non-distressed transactions, Panel C for a subsample of short sale transactions, and Panel D for a subsample of fire sale (i.e., REO) transactions.

Table 2: Remarks Cleaning Procedure

Panel A: Raw Remark
this is a bank owned reo property in good condition but needs tlc. must see. outstanding opportunity to own a fabulous house in a historic neighborhood at this price. roof has been replaced in recent years, red oak hardwood floors are in excellent condition, siding is hardyplank and needs to be pressure cleaned. vaulted ceilings in upstairs bedrooms. new marvin windows throughout. large eat in kitchen with granite countertops. sold as is. please include reo addendum and proof of funds with all offers.
Panel B: Cleaned Remark
must see. outstanding opportunity to own a fabulous house in a historic neighborhood at this price. roof has been replaced in recent years, red oak hardwood floors are in excellent condition, siding is hardyplank and needs to be pressure cleaned. vaulted ceilings in upstairs bedrooms. new marvin windows throughout. large eat in kitchen with granite countertops.
Panel C: Cleaned Full Remark
this is a ~~bank~~ ~~owned~~ ~~reo~~ in good condition but needs tlc. must see. outstanding opportunity to own a fabulous house in a historic neighborhood at this price. roof has been replaced in recent years, red oak hardwood floors are in excellent condition, siding is hardyplank and needs to be pressure cleaned. vaulted ceilings in upstairs bedrooms. new marvin windows throughout. large eat in kitchen with granite countertops. ~~sold~~ ~~asis~~. please include ~~reo~~ ~~addendum~~ and proof of ~~funds~~ with all ~~offers~~.

Note: Table 2 provides an example of a remark for a fire sale liquidation. Panel A displays the raw remark before applying our cleaning procedure. Panel B displays the cleaned remark after running the computer-assisted text cleaning algorithm. Panel C reintroduces the full remark after running the variable selection procedure on the cleaned remark in Panel B. Boxed text indicates the token is included in the post-selection estimation procedure. ~~Strikethrough~~ text indicates the token is redundant. The example in Table 2 includes stopwords and does not stem the tokens to increase readability. The internet appendix provides a detailed overview of the text preprocessing we perform and a comprehensive list of the redundant tokens for each of the ten MSAs.

Table 3: Quality-adjusted Fire Sale Discounts

		Cleaned Remarks		Cleaned Full Remarks	
	Raw	Single	Double	Single	Double
	(1)	(2)	(3)	(4)	(5)
Panel A: Full sample					
Fire sale	−0.261***	−0.200***	−0.197***	−0.183***	−0.180***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Short sale	−0.156***	−0.122***	−0.121***	−0.117***	−0.116***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Observations	7,091,684	7,091,684	7,091,684	7,091,684	7,091,684
Adjusted R^2	0.920	0.933	0.934	0.935	0.935
Panel B: Repeat-sales sample					
Fire sale	−0.268***	−0.190***	−0.187***	−0.177***	−0.173***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Short sale	−0.174***	−0.126***	−0.125***	−0.120***	−0.120***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Observations	4,275,177	4,275,177	4,275,177	4,275,177	4,275,177
Adjusted R^2	0.954	0.967	0.968	0.968	0.969
House Controls	✓	✓	✓	✓	✓
Financing Controls	✓	✓	✓	✓	✓
Tract×Qtr×Yr FE	✓	✓	✓	✓	✓

*p<0.1; **p<0.05; ***p<0.01

Note: Table 3 displays fire sale and short sale coefficient estimates for the combined ten MSA transaction sample. Panel A uses the full transaction sample and Panel B uses a repeat-sales subsample. The dependent variable in every column is the log of sales price. The fire sale and short sale coefficient estimates are relative to non-distressed transactions. Every column includes house controls, financing controls, and quarter-by-year-by-census tract fixed effects. The house controls in Panel A include housing characteristics (e.g., beds, baths). Whereas the house controls in Panel B include house fixed effects. Columns 2 through 5 include tokens that control for house quality. Column 2 (3) includes tokens from a single-selection (double-selection) variable selection process using the cleaned remarks. Column 4 (5) includes the same single-selection (double-selection) variables as column 2 (3). However, the tokens are populated using the agents' cleaned full remarks. Standard errors are clustered at the census tract level.

Table 4: Asset Illiquidity

	Raw (1)	Double (2)	Raw (3)	Double (4)	Raw (5)	Double (6)
Fire sale	−0.252***	−0.178***	−0.246***	−0.175***	−0.195***	−0.125***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
Short sale	−0.146***	−0.111***	−0.141***	−0.109***	−0.129***	−0.097***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Invest-to-Invest			−0.016***	−0.013***	−0.001	0.000
			(0.001)	(0.001)	(0.001)	(0.001)
Own-to-Invest			−0.047***	−0.031***	−0.032***	−0.017***
			(0.001)	(0.001)	(0.001)	(0.001)
Invest-to-Own			0.012***	0.003***	0.014***	0.004***
			(0.001)	(0.000)	(0.001)	(0.000)
Invest-to-Invest×Short sale					−0.020***	−0.017***
					(0.004)	(0.004)
Invest-to-Invest×Fire sale					−0.147***	−0.134***
					(0.004)	(0.003)
Own-to-Invest×Short sale					−0.039***	−0.040***
					(0.002)	(0.001)
Own-to-Invest×Fire sale					−0.113***	−0.113***
					(0.002)	(0.002)
Invest-to-Own×Short sale					−0.011***	−0.004
					(0.003)	(0.003)
Invest-to-Own×Fire sale					−0.002	−0.006**
					(0.003)	(0.003)
Observations	4,644,689	4,644,689	4,644,689	4,644,689	4,644,689	4,644,689
Adjusted R^2	0.931	0.945	0.932	0.945	0.932	0.945
House Characteristics	✓	✓	✓	✓	✓	✓
Financing Controls	✓	✓	✓	✓	✓	✓
Tract×Qtr×Yr FE	✓	✓	✓	✓	✓	✓
Cleaned Full Remarks		✓		✓		✓

*p<0.1; **p<0.05; ***p<0.01

Note: Table 4 displays fire sale, short sale, and redeploy coefficient estimates for a subsample of transactions where we can reliably identify the buyer tenure type (i.e., owner-occupied vs. investor). The dependent variable in every column is the log of sales price. The fire sale and short sale coefficient estimates are relative to non-distressed transactions. The redeploy coefficient estimates are relative to *Own-to-Own* transactions. Every column includes housing characteristics and financing controls as well as quarter-by-year-by-census tract fixed effects. Even numbered columns include tokens from a double-selection variable selection process using the cleaned full remarks. Standard errors are clustered at the census tract level.

Table 5: Asset Illiquidity by Investment Horizon

	Raw (1)	Double (2)	Raw (3)	Double (4)	Raw (5)	Double (6)
Fire sale	−0.252***	−0.178***	−0.242***	−0.173***	−0.192***	−0.124***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
Short sale	−0.146***	−0.111***	−0.139***	−0.108***	−0.127***	−0.096***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Invest-to-ST Invest			−0.058***	−0.043***	−0.037***	−0.023***
			(0.002)	(0.001)	(0.002)	(0.001)
Invest-to-LT Invest			−0.005***	−0.005***	0.008***	0.006***
			(0.001)	(0.001)	(0.001)	(0.001)
Own-to-ST Invest			−0.107***	−0.072***	−0.089***	−0.052***
			(0.001)	(0.001)	(0.001)	(0.001)
Own-to-LT Invest			−0.030***	−0.020***	−0.017***	−0.007***
			(0.001)	(0.001)	(0.001)	(0.001)
Invest-to-Own			0.011***	0.002***	0.013***	0.004***
			(0.001)	(0.000)	(0.001)	(0.000)
Invest-to-ST Invest×Short sale					−0.093***	−0.078***
					(0.010)	(0.009)
Invest-to-ST Invest×Fire sale					−0.180***	−0.172***
					(0.007)	(0.006)
Invest-to-LT Invest×Short sale					−0.007*	−0.007*
					(0.004)	(0.004)
Invest-to-LT Invest×Fire sale					−0.137***	−0.123***
					(0.004)	(0.004)
Own-to-ST Invest×Short sale					−0.065***	−0.066***
					(0.003)	(0.002)
Own-to-ST Invest×Fire sale					−0.111***	−0.122***
					(0.002)	(0.002)
Own-to-LT Invest×Short sale					−0.030***	−0.031***
					(0.002)	(0.001)
Own-to-LT Invest×Fire sale					−0.110***	−0.108***
					(0.002)	(0.002)
Invest-to-Own×Short sale					−0.009***	−0.004
					(0.003)	(0.003)
Invest-to-Own×Fire sale					−0.001	−0.006**
					(0.003)	(0.003)
Observations	4,644,689	4,644,689	4,644,689	4,644,689	4,644,689	4,644,689
Adjusted R^2	0.931	0.945	0.932	0.945	0.933	0.946
House Characteristics	✓	✓	✓	✓	✓	✓
Financing Controls	✓	✓	✓	✓	✓	✓
Tract×Qtr×Yr FE	✓	✓	✓	✓	✓	✓
Cleaned Full Remarks		✓		✓		✓

*p<0.1; **p<0.05; ***p<0.01

Note: Table 5 displays fire sale, short sale, and redeploy coefficient estimates for a subsample of transactions where we can reliably identify the buyer tenure type (i.e., owner-occupied vs. investor). The dependent variable in every column is the log of sales price. The fire sale and short sale coefficient estimates are relative to non-distressed transactions. The redeploy coefficient estimates are relative to *Own-to-Own* transactions. Every column includes housing characteristics and financing controls as well as quarter-by-year-by-census tract fixed effects. Even numbered columns include tokens from a double-selection variable selection process using the cleaned full remarks. Standard errors are clustered at the census tract level.

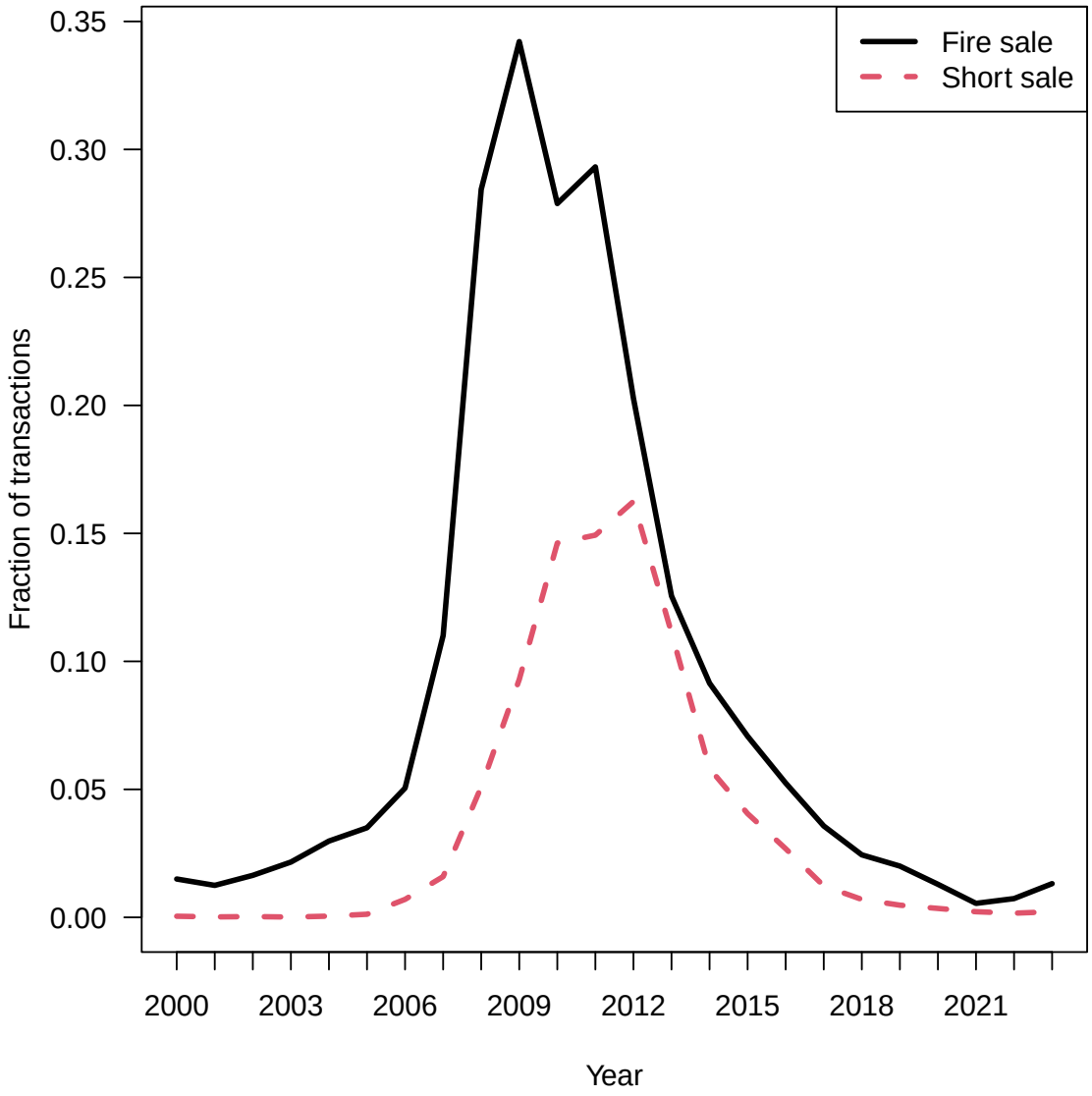
Table 6: Days On Market

	Raw (1)	Double (2)	Raw (3)	Double (4)	Raw (5)	Double (6)
Fire sale	−7.684***	−9.477***	−7.839***	−9.552***	−10.212***	−12.099***
	(0.249)	(0.231)	(0.249)	(0.231)	(0.283)	(0.266)
Short sale	55.299***	54.035***	55.182***	53.971***	55.058***	53.870***
	(0.362)	(0.367)	(0.362)	(0.367)	(0.437)	(0.437)
Invest-to-Invest			−0.980***	−1.244***	−1.564***	−1.763***
			(0.177)	(0.169)	(0.176)	(0.169)
Own-to-Invest			1.008***	0.491***	0.453***	−0.137
			(0.095)	(0.091)	(0.097)	(0.094)
Invest-to-Own			−0.828***	−0.974***	−1.022***	−1.191***
			(0.123)	(0.121)	(0.123)	(0.122)
Invest-to-Invest×Short sale					−3.621**	−4.783***
					(1.519)	(1.521)
Invest-to-Invest×Fire sale					7.707***	7.500***
					(0.587)	(0.576)
Own-to-Invest×Short sale					0.961	1.028
					(0.653)	(0.653)
Own-to-Invest×Fire sale					4.795***	5.362***
					(0.292)	(0.284)
Invest-to-Own×Short sale					−1.024	−1.029
					(1.469)	(1.463)
Invest-to-Own×Fire sale					3.992***	4.373***
					(0.680)	(0.672)
Observations	4,644,689	4,644,689	4,644,689	4,644,689	4,644,689	4,644,689
Adjusted R^2	0.197	0.218	0.197	0.218	0.197	0.218
House Characteristics	✓	✓	✓	✓	✓	✓
Financing Controls	✓	✓	✓	✓	✓	✓
Tract×Qtr×Yr FE	✓	✓	✓	✓	✓	✓
Cleaned Full Remarks		✓		✓		✓

*p<0.1; **p<0.05; ***p<0.01

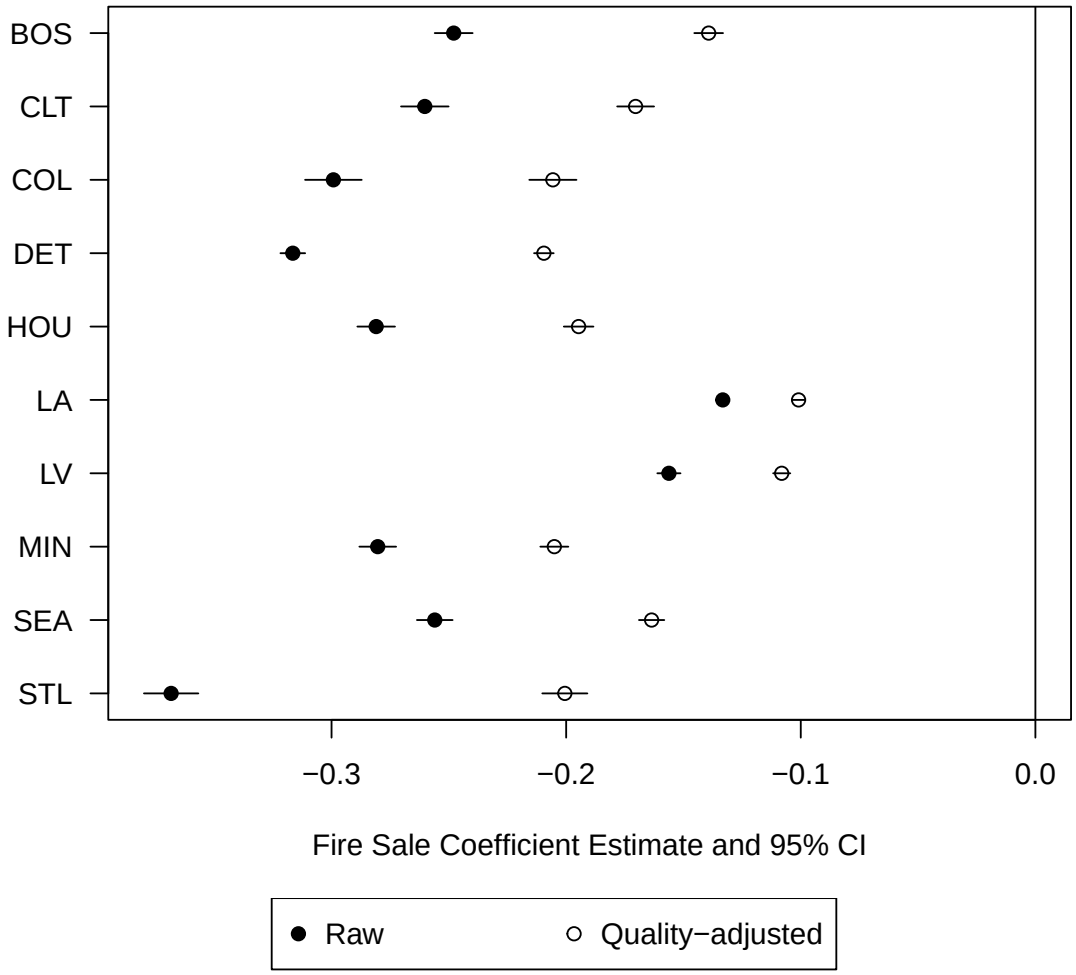
Note: Table 6 displays fire sale, short sale, and redeploy coefficient estimates for a subsample of transactions where we can reliably identify the buyer tenure type (i.e., owner-occupied vs. investor). The dependent variable in every column is the number of days the house is listed on the market before it sells. The fire sale and short sale coefficient estimates are relative to non-distressed transactions. The redeploy coefficient estimates are relative to *Own-to-Own* transactions. Every column includes housing characteristics and financing controls as well as quarter-by-year-by-census tract fixed effects. Even numbered columns include tokens from a double-selection variable selection process using the cleaned full remarks. Standard errors are clustered at the census tract level.

Figure 1: Fraction of Distressed Transactions



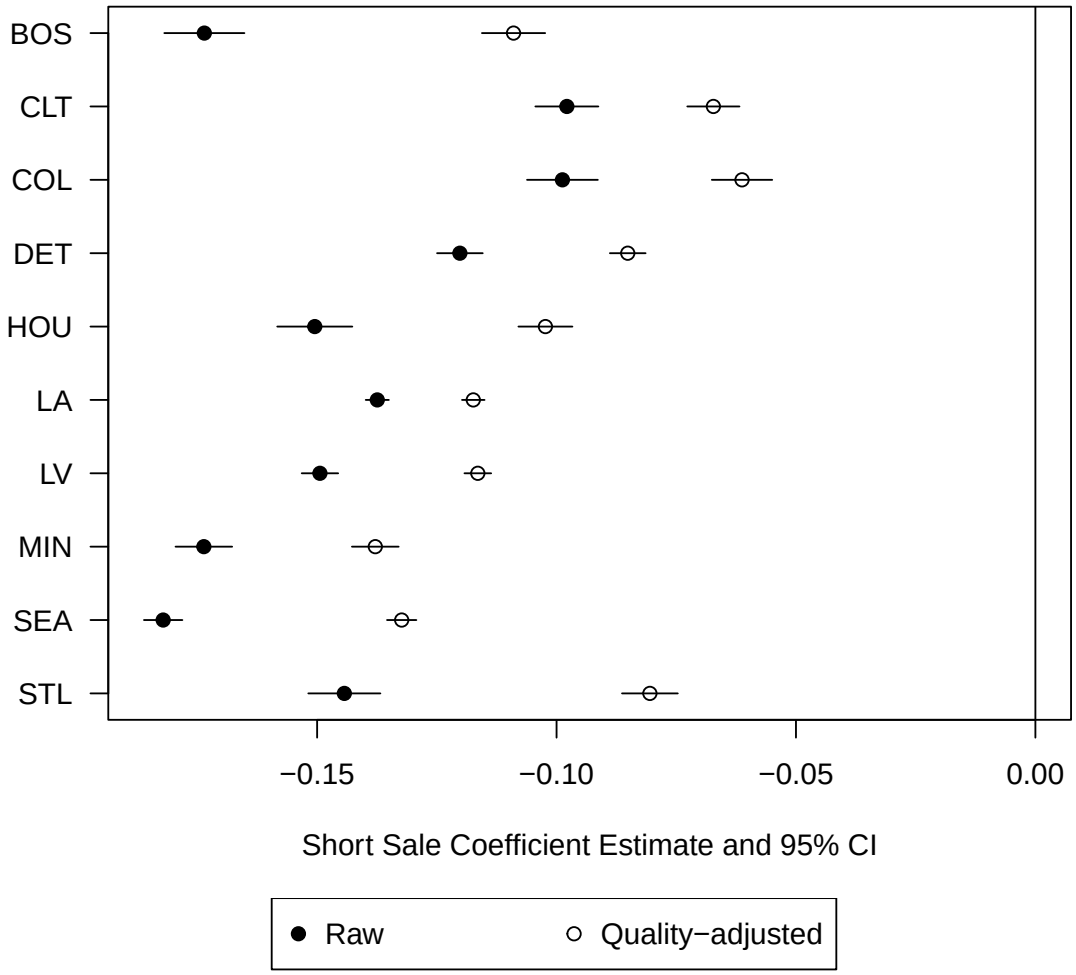
Notes: Figure 1 plots the percent of transactions that were liquidated via a fire sale (black, solid line) or short sale (red, dotted line) each year in the combined ten MSA transaction sample.

Figure 2: Fire Sale Coefficient Estimates By MSA



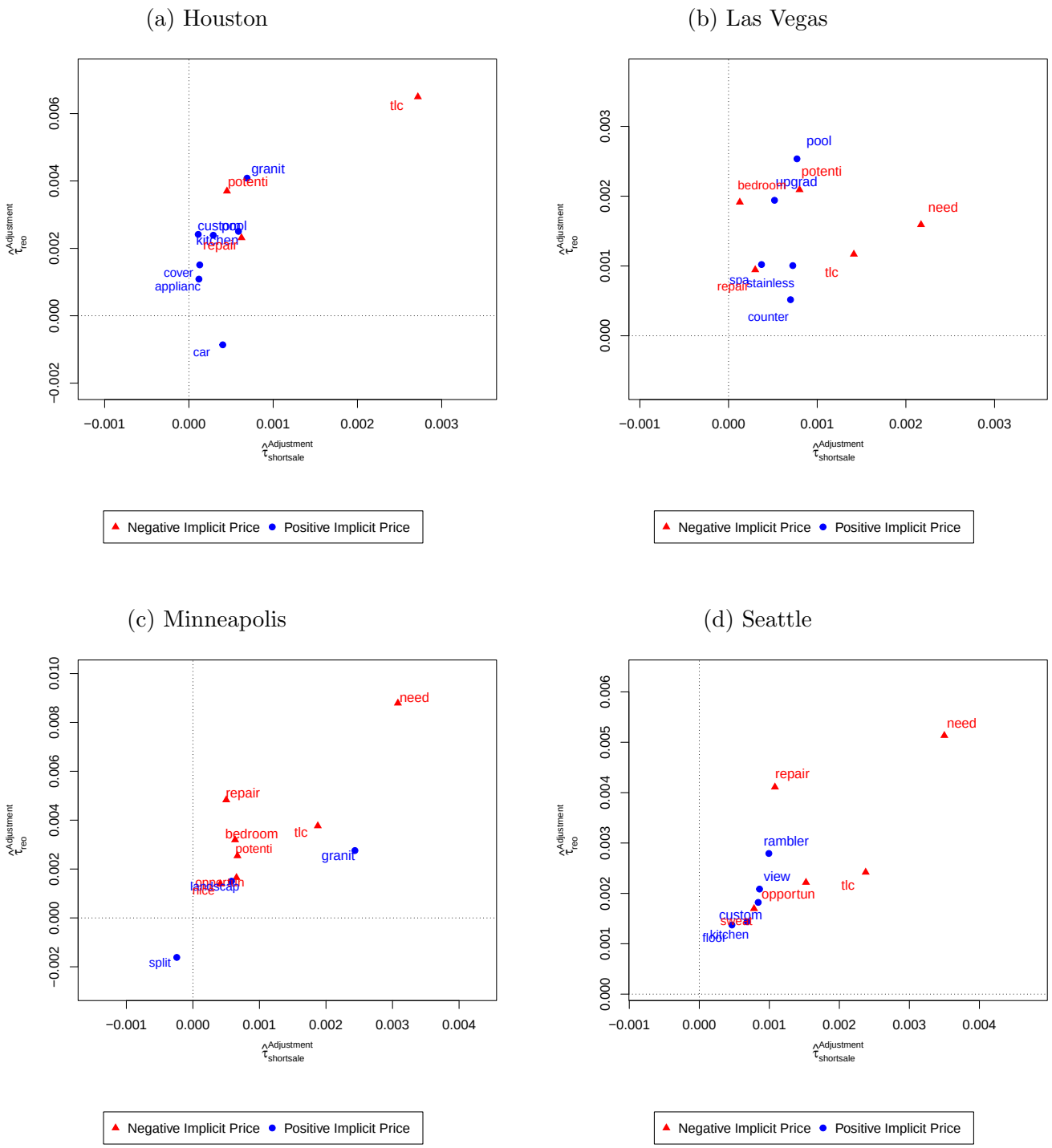
Notes: Figure 2 plots the raw and quality-adjusted fire sale coefficient estimates and 95% confidence intervals individually for all ten MSAs.

Figure 3: Short Sale Coefficient Estimates By MSA



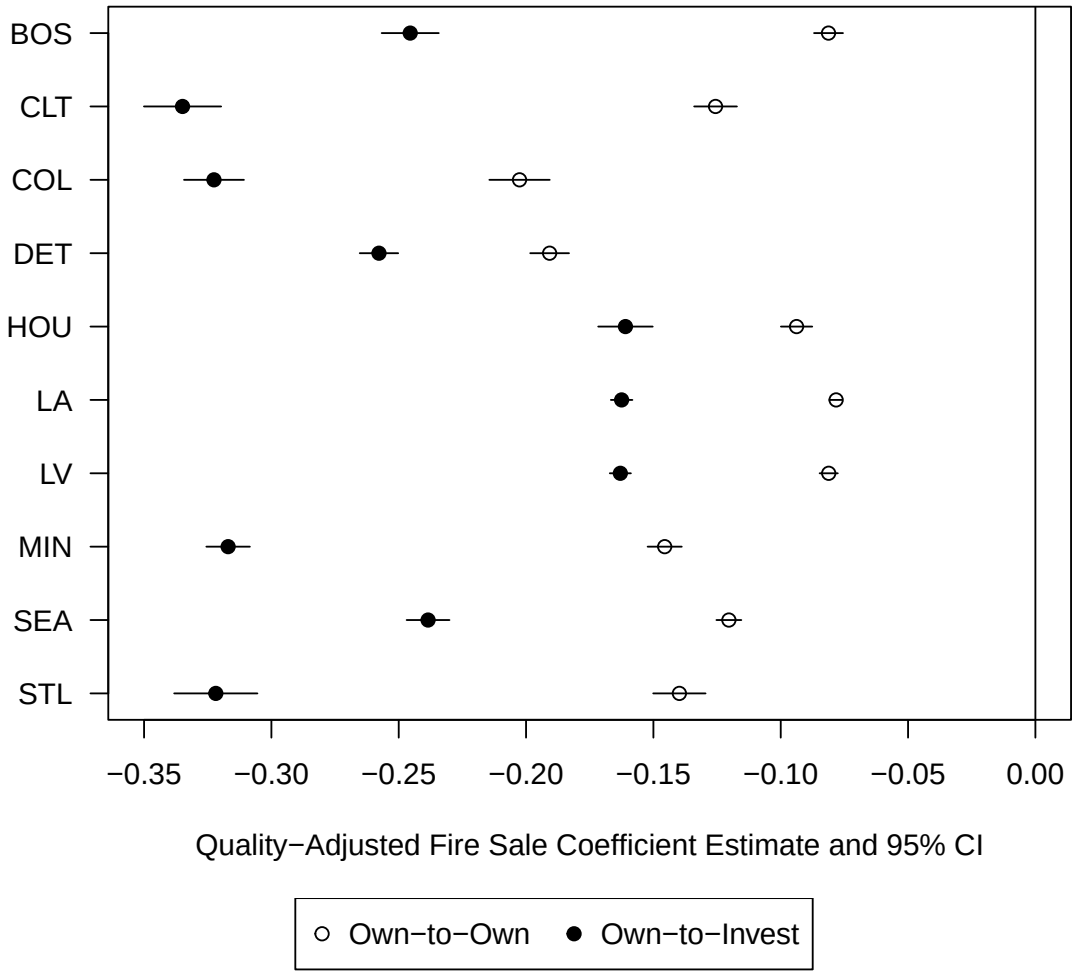
Notes: Figure 3 plots the raw and quality-adjusted short sale coefficient estimates and 95% confidence intervals individually for all ten MSAs.

Figure 4: Fire Sale and Short Sale Coefficient Adjustments



Notes: Figure 4 plots fire sale and short sale token adjustments individually for four MSAs in Panels A to D. Tokens that increase price, $\hat{\theta}_s > 0$, are plotted alongside blue circles. Tokens that decrease price, $\hat{\theta}_s < 0$, are plotted alongside red triangles.

Figure 5: Fire Sale Coefficient Estimates By Redeployability Type



Notes: Figure 5 plots the quality-adjusted *Own-to-Own* and *Own-to-Invest* fire sale coefficient estimates and 95% confidence intervals individually for all ten MSAs. The *Own-to-Own* coefficient estimates represent the average quality-adjusted fire sale discount when owner-occupiers purchase houses that were previously owner-occupied (i.e., the house is not redeployed). The *Own-to-Invest* coefficient estimates represent the average quality-adjusted fire sale discount when investors purchase houses that were previously owner-occupied (i.e., the house is redeployed).

Algorithm 1: Computer-Assisted Text Cleaning Procedure

The following algorithm describes our computer-assisted text-cleaning procedure. For observations $n = 1, ..., N$, we partition the agent remarks about the house for observation n, $\mathcal{R}_n$, into a set of $i = 1, ..., I_n$ sentences. Define $\mathcal{R} = \{\mathcal{R}_n\}_{n=1}^{n=N}$ and $E(\mathcal{R}_n, \mathcal{E})$ as the set of sentences that remain after eliminating all sentences containing any of the tokens in $\mathcal{E}$.

Algorithm 1: Computer-Assisted Text Cleaning for Double-Selection

Result: A set of redundant tokens for short sales ($\mathcal{E}_{SS}$) and fire sales ($\mathcal{E}_{FS}$)

Set $j = 0$, $\mathcal{E}^0 = \emptyset$, and $D = 1$. Define $\mathcal{R}^0$ as the initial remarks before any computer-assisted text cleaning. ;

while $D > 0$ **do**

Set $j = j + 1$;

Run LASSO with the dependent variable set to the short sale or fire sale indicator and the $2,000$ most-frequent tokens in $\mathcal{R}^j$ as the predictors ;

Define the set of tokens with non-zero coefficients from the previous step as $\mathcal{S}^j$;

Define $\mathcal{E}^j_{NEW} \subseteq \mathcal{S}^j$ as the set of tokens that are redundant for the dependent variable where redundancy is based on the researcher's knowledge of the economic environment;

If there are no redundant tokens, set $\mathcal{E}^j_{NEW} = \emptyset$;

Set $D = |\mathcal{E}^j_{NEW}|$;

if $\mathcal{E}^j_{NEW} \neq \emptyset$ **then**

Update the set of redundant tokens $\mathcal{E}^j = \mathcal{E}^j_{NEW} \cup \mathcal{E}^{j-1}$;

For $n = 1, ..., N$, update the remarks $\mathcal{R}^j_n = E(\mathcal{R}^{j-1}_n, \mathcal{E}^j)$;

end

end

Appendices

Contents

A Data Overview (Internet Appendix) **41**

A.1 Data Filters . . . 41

A.2 Text Preprocessing . . . 43

A.3 Redundant Tokens . . . 44

A.4 Computer-Assisted Text Cleaning Examples . . . 46

B Additional Results (Internet Appendix) **49**

B.1 MSA-level Summary Statistics . . . 49

B.2 Repeat Sales . . . 54

B.3 Distressed Transactions by MSA . . . 56

B.4 Fire Sale and Short Sale Coefficient Adjustments . . . 59

C Miscellaneous **62**

C.1 Stimulating Housing Markets . . . 62

A Data Overview (Internet Appendix)

A.1 Data Filters

The MLS data used to examine fire sale discounts in housing markets was obtained from CoreLogic. Prior to estimating the fire sale discounts, we apply several filters to the single-family detached residential transaction data. We list the filters below and provide a detailed overview of the number of transactions that are dropped by MSA in Table A1. Records are dropped that do not meet the following criteria:

1. filter 1: listing successfully sold
2. filter 2: $2000 \leq$ list year ≤ 2023
3. filter 3: \$20,000 $\leq$ sale price $\leq$ \$2,000,000
4. filter 4: $0 \leq$ days-on-market ≤ 730
5. filter 5: $1 \leq$ bathrooms ≤ 6
6. filter 6: $1 \leq$ bedrooms ≤ 6
7. filter 7: square feet of living area missing
8. filter 8: $500 \leq$ square feet of living area $\leq$ 4,000
9. filter 9: lot size missing
10. filter 10: lot size ≤ 5 acres
11. filter 11: $1 \leq$ age
12. filter 12: $1900 \leq$ year built
13. filter 13: latitude or longitude missing
14. filter 14: census tract not in county
15. filter 15: not duplicate listing
16. filter 16: num sales in zip code each year ≥ 50

Table A1: Filtered transaction data by MSA

	MSA									
Filter	BOS	CTL	COL	DET	HOU	LA	LV	MIN	SEA	STL
no filter	1,319,784	996,643	777,878	3,012,190	2,421,564	3,599,217	716,578	2,054,851	1,585,331	1,105,014
filter 1	901,840	652,953	528,976	1,470,278	1,606,585	2,650,962	552,901	1,348,734	1,102,954	750,404
filter 2	735,294	624,139	451,933	1,261,168	1,481,763	2,558,527	552,891	1,058,598	1,099,606	701,139
filter 3	722,890	619,939	451,698	1,160,057	1,470,300	2,251,912	550,767	1,056,990	1,072,271	698,170
filter 4	722,318	617,219	448,982	1,153,790	1,465,057	2,249,069	550,295	1,054,626	1,070,561	695,306
filter 5	713,508	614,131	446,734	1,149,354	1,459,222	2,215,825	549,104	1,052,568	1,061,999	677,212
filter 6	710,647	597,054	446,250	1,148,396	1,458,333	2,209,504	548,485	1,051,381	1,060,024	675,836
filter 7	706,759	597,041	434,377	1,148,212	1,448,359	2,110,967	547,844	1,051,370	1,059,964	576,938
filter 8	669,454	562,659	419,756	1,133,787	1,362,096	2,069,831	532,450	996,525	1,010,126	567,252
filter 9	669,237	552,273	320,460	1,076,074	1,206,826	2,031,795	532,229	995,935	1,009,739	439,597
filter 10	661,299	546,194	313,726	1,055,673	1,203,377	2,024,583	532,161	979,633	996,893	432,495
filter 11	639,907	485,541	304,995	1,020,676	1,132,141	1,993,704	519,446	931,148	903,255	416,451
filter 12	593,551	485,027	301,755	1,016,115	1,131,397	1,991,521	519,446	917,577	901,588	408,991
filter 13	589,819	484,067	301,671	984,189	1,129,397	1,985,338	519,441	902,111	901,113	408,008
filter 14	588,659	484,058	301,671	984,016	1,129,185	1,977,462	519,163	902,111	900,760	408,008
filter 15	577,739	479,118	300,569	901,610	1,128,380	999,181	519,048	899,354	893,270	402,669
filter 16	572,191	478,621	300,161	901,112	1,128,086	998,716	518,659	898,986	892,964	402,188

Notes: Table A1 tabulates the number of records that are dropped for each filter across the ten MSAs examined in this study. The final row for each column identifies the number of transactions that are included in the MSA-level analysis.

A.2 Text Preprocessing

Before cleaning and tokenizing the remarks, we perform a minimal amount of preprocessing. The primary goal of the preprocessing procedure is to clean and standardize the remarks. We preprocess the remarks as follows:

1. Convert to lower case.
2. Combine key bigrams into unigrams: *short sale* $\rightarrow$ *shortsale*; *as is* $\rightarrow$ *asis*.
3. Replace commas (,), periods (.), ampersands (&), and the word *and* with a space.
4. Replace all special characters with a space.
5. Remove apostrophes.
6. Remove all remaining single letters.
7. Replace all numbers with a space. Numbers can be in either numeric or character form.
8. Remove duplicate empty spaces.
9. Stem the remarks.
10. Trim empty spaces at the beginning and end of the remark.
11. Run computer-assisted text cleaning algorithm.

A.3 Redundant Tokens

When creating the token dictionaries, it is vital to only include tokens that convey new information about the house. Specifically, we want to ensure the dictionary does not include tokens conveying information about the transaction type or explicitly indicating the house is an fire sale or short sale. We identify redundant tokens using the computer-assisted text cleaning algorithm in the paper. We run the text cleaning algorithm separately for each MSA. After identifying the redundant tokens, we remove entire sentences that contain at least one redundant token to address the vestigial predictor problem discussed in the paper. This section displays the redundant tokens we identify and remove for each MSA. We will post the computer-assisted text cleaning algorithm code on our personal websites as soon as this paper is accepted for publication.

Boston: addenda, addendum, appli, approval, asi, asis, auction, bank, cash, contact, contract, corpor, escrow, fannie, fha, fhmc, firm, fnma, foreclosure, freddie, fund, highest, homepath, homestep, hud, impli, information, instruct, lender, loan, occup, paid, preforeclosure, property, purchase, qualifi, reo, represent, representation, response, restrict, shortsale, sold, submit, verifi, website

Charlotte: accur, addenda, addendum, addm, approval, asi, asis, auction, bank, bid, case, contribut, corpor, earnest, elig, emd, equal, exempt, fannie, fha, fhmc, fnma, foreclosure, freddie, fund, harm, homepath, homestep, hud, information, instruct, lender, minimum, occup, period, preforeclosure, prequal, property, qual, regard, reo, represent, representation, response, shortsale, sold, title, verbal, verif, website

Columbus: addenda, addendum, appli, approval, appt, asi, asis, auction, bank, bcm, bid, case, contact, corpor, deadlin, deal, disclosure, doc, earnest, elig, equal, escrow, exist, fannie, fha, fhmc, final, fnma, foreclosure, form, freddie, fund, harm, highest, homepath, homestep, hud, import, incent, information, instruct, insure, lender, letter, lockbox, multipl, naid, occup, period, preforeclosure, process, property, purchase, remark, reo, requir, response, select, shortsale, sold, strict, subject, submit, website

Detroit: addenda, addendum, anytime, appli, applic, approval, asi, asis, assist, auction, bank, bid, charg, click, com, comm, commiss, compli, contact, coordin, corp, corpor, deadline, disclosure, doc, document, earnest, elig, email, emd, escrow, fannie, fha, fhmc, financ, fnma, foreclosure, freddie, fund, homepath, homestep, hud, idrbng, incent, info, inform, information, instruct, lbp, lender, mcb, municip, net, obtain, onlin, preforeclosure, prop, property, purchase, reo, represent, request, response, restrict, shortsale, sold, status, subject, submit, website

Houston: addenda, addendum, approval, asi, asis, auction, bank, bid, case, corpor, daili,

disclosur, fannie, fha, fhmc, fnma, foreclosure, freddie, fund, homepath, homestep, hud, information, late, lender, preforeclosure, property, reo, shortsale, sold, website

Los Angeles: addendum, approval, asi, auction, bank, com, consider, corporate, cost, countrywide, disclosure, escrow, express, expresspath, fannie, fargo, fha, fhmc, final, finance, fnma, foreclosure, freddie, fund, ginnie, gnma, guarantee, herein, highest, homepath, hud, information, instruction, investigate, lender, nationstar, nsp, offer, paid, preforeclosure, property, read, receive, record, rely, remark, reo, request, respond, response, review, sfr, shortsale, sold, subject, submit, termite, thank, themselves, verify, warranty, website

Las Vegas: addenda, addendum, anytime, approval, asi, asis, assist, auction, bank, bid, corpor, document, fannie, fha, fhmc, finance, flood, fnma, foreclosure, freddie, fund, highest, homepath, homestep, hud, info, information, insure, lead, lender, measure, nrs, preforeclosure, property, purchase, reo, report, restrict, review, shortsale, sold, stuff , submit, toward, verify, waiver, warranti, website

Minneapolis: addenda, addendum, approval, asi, asis, auction, bank, corpor, estimate, fannie, fha, fhmc, fnma, foreclosure, freddie, homepath, homestep, hud, information, lender, mandatori, preforeclosure, property, reo, response, restrict, shortsale, sold, verifi, website

Seattle: accord, addenda, addendum, agent, approval, asis, assist, auction, bank, bid, broker, case, corpor, date, earnest, elig, em, emd, express, fannie, fee, fha, fhmc, fnma, foreclosure, freddie, homepath, homestep, hud, info, information, insur, lender, occup, occupi, preforeclosure, property, purchase, record, reo, request, restrict, review, shortsale, sold, warranti, website

Saint Louis: accept, addenda, addendum, appli, approval, asis, assist, auction, bank, bid, case, com, commission, consider, corpor, count, disclosure, earnest, elig, escrow, fannie, fha, fhmc, flashlight, fnma, foreclosure, freddie, fund, highest, homepath, homestep, hour, hud, incent, inform, information, initial, lead, lender, obtain, payable, preforclosure, property, purchase, refer, regardless, reo, represent, representation, response, shortsale, sold, submit, title, turn, verifi, website, written

A.4 Computer-Assisted Text Cleaning Examples

Text Cleaning Example 1

The remark below includes at least one sentence with redundant tokens for fire sale transactions.

$$R_n^0 = \underbrace{\text{This home needs TLC.}}_{i=1}\underbrace{\text{Please call before showing.}}_{i=2}\underbrace{\text{This is a bank property.}}_{i=3}$$

We begin without knowledge of the redundant tokens, $\mathcal{E}^0 = \emptyset$. The first iteration of the cleaning process in Algorithm 1 runs the LASSO procedure and finds *bank* is in the dictionary, $\mathcal{S}^1$. This term is a redundant token for *bank property*, *bank owned property*, and other variants. Based on this knowledge, we include *bank* in the set of redundant tokens, *bank* $\in \mathcal{E}^1_{NEW}$. We then update the dictionary of redundant tokens $\mathcal{E}^1 = \mathcal{E}^1_{NEW} \cup \mathcal{E}^0$ and update the remarks

$$R_n^1 = E(R_n^0, \mathcal{E}^1) = \underbrace{\text{This home needs TLC.}}_{i=1}\underbrace{\text{Please call before showing.}}_{i=2}$$

In the second round, the algorithm continues searching for strong predictors of fire sale transactions using R^1. Assuming there are no redundant tokens in $\mathcal{S}^2$, the algorithm terminates. We then use R^1 to create the token dictionary that we include in the regression analysis.

Text Cleaning Example 2

The remark below highlights the vestigial predictor problem and challenges associated with eliminating redundant tokens without eliminating the associated sentences.

$$R_n^0 = \underbrace{\text{This home needs TLC.}}_{i=1}\underbrace{\text{Buyer must sign included addendum.}}_{i=2}\underbrace{\text{This is a bank property.}}_{i=3}$$

The first round of the cleaning process in Algorithm 1 runs the LASSO procedure and finds *bank, addendum* $\in \mathcal{E}^1$. Instead of using the eliminator function E to remove entire sentences that contain one of the redundant tokens, let's assume we only remove the redundant tokens

themselves.

$$R_n^{1*} = E^*(R_n^0, \mathcal{E}^1) = \underbrace{\text{This home needs TLC.}}_{i=1} \underbrace{\text{Buyer must sign included.}}_{i=2} \underbrace{\text{This is a property.}}_{i=3}$$

In the second round, the algorithm identifies *included* as a strong predictor of fire sale transactions because it has a non-zero LASSO coefficient estimate. Although not as clear as *bank* or *addendum*, the researcher could manually inspect the remarks and reasonably infer that *included* is redundant for fire sale transactions. The researcher would then include *included* in $\mathcal{E}^2$. The algorithm then cleans the remarks yielding

$$R_n^{2*} = E^*(R_n^{1*}, \mathcal{E}^2) = \underbrace{\text{This home needs TLC.}}_{i=1} \underbrace{\text{Buyer must sign.}}_{i=2} \underbrace{\text{This is a property.}}_{i=3}$$

Continuing the algorithm would identify *sign* as a strong predictor of fire sale transactions. However, it is less clear to the researcher if *sign* is a redundant token for fire sale transactions because remarks for both fire sale and non-fire sale listings include phrases such as *buyer must sign a disclosure* or *there is no sign out front*. Thus, it is unclear to the researcher if *sign* is a redundant token in some, none, or all instances. We refer to this ambiguity as the vestigial predictor problem where the predictive power of *sign* is entirely derived from its inclusion in the same sentence as *Buyer must sign included addendum*. For this reason, Algorithm 1 eliminates entire sentences that include at least one redundant token.

The cleaning procedure in Algorithm 1 all but eliminates the vestigial predictor problem. Moreover, we find the algorithm requires no more than a few rounds to identify redundant tokens in each of the ten MSAs. Also, despite this more aggressive cleaning approach, we find no significant impact on the number of tokens selected or their impact on price. That said, removing entire sentences may inadvertently remove non-redundant house quality tokens. We highlight this potential issue and how we address it in the following example.

Text Cleaning Example 3

The remark below includes a sentence that contains both redundant and non-redundant tokens for fire sale transactions.

$$R_n^0 = \underbrace{\text{This is a bank owned home that needs TLC.}}_{i=1} \underbrace{\text{Please call before showing.}}_{i=2}$$

Similar to the previous examples, the first round of the cleaning process in Algorithm 1 finds *bank* $\in \mathcal{E}^1$. In which case, the cleaning procedure removes the first sentence leaving

$$R_n^1 = E(R_n^0, \mathcal{E}^1) = \underbrace{\text{Please call before showing.}}_{i=1}$$

Unlike the previous examples, the phrase *needs TLC* is no longer present in the cleaned remark because it is part of a sentence that contains redundant information. We address this issue by using the full set of cleaned remarks, here R_n^1, to create the token dictionary and then use the full set of raw remarks, here R_n^0, for the post-double-selection regression analysis. This ensures redundant tokens, here *bank owned*, are not included in the dictionary but allows the non-redundant tokens in $\hat{\mathcal{S}}_2$ that appear in the same sentences as *bank owned*, here *needs TLC*, to affect the post-selection regression results. Panel C of Table 2 displays the remark from Panels A and B after we reintroduce the full remark for the post-selection regression.

B Additional Results (Internet Appendix)

B.1 MSA-level Summary Statistics

Table B1 provides descriptive statistics for select housing characteristics for each of the ten MSAs examined in this study. In addition to the descriptive statistics, we also list the time period of the data and the represented counties that comprise the MSA in the body of the paper.

Table B1: Descriptive statistics by MSA

	Min	Pctl(25)	Mean	Median	Pctl(75)	Max
Boston (N = 572,191)						
Price (000s)	20.00	287.00	449.76	390.00	550.00	2,000.00
Days on market	1	17	57.53	34	74	730
Sqft (000s)	0.50	1.36	1.90	1.76	2.31	4.00
Age	1	31	53.21	52	72	123
Acres	0.00	0.17	0.58	0.32	0.72	5.00
Bedrooms	1	3	3.30	3	4	6
Bathrooms	1	2	2.18	2	3	6
Cash purchase	0	0	0.09	0	0	1
Fire sale	0	0	0.03	0	0	1
Short sale	0	0	0.02	0	0	1
Years:	2000-2023					
Counties:	Essex MA, Middlesex MA, Norfolk MA, Plymouth MA, Rockingham MA, Strafford MA, Suffolk MA					
Charlotte (N = 478,621)						
Price (000s)	20.00	134.00	243.57	205.00	314.60	2,000.00
Days on market	1	10	73.16	36	95	730
Sqft (000s)	0.50	1.44	2.06	1.93	2.59	4.00
Age	1	8	25.50	17	38	123
Acres	0.00	0.20	0.45	0.29	0.47	5.00
Bedrooms	1	3	3.41	3	4	6
Bathrooms	1	2	2.60	3	3	6
Cash purchase	0	0	0.13	0	0	1
Fire sale	0	0	0.04	0	0	1
Short sale	0	0	0.02	0	0	1
Years:	2000-2023					
Counties:	Cabarrus NC, Gaston NC, Iredell NC, Lincoln NC, Mecklenburg NC, Rowan NC, Union NC, York NC					

Table A2: Descriptive statistics by MSA (cont.)

	Min	Pctl(25)	Mean	Median	Pctl(75)	Max
Columbus (N = 300,161)						
Price (000s)	20.00	122.00	211.99	185.00	276.00	2,000.00
Days on market	1	7	68.56	24	84	730
Sqft (000s)	0.50	1.32	1.85	1.70	2.24	4.00
Age	1	18	42.48	38	60	123
Acres	0.00	0.15	0.37	0.20	0.30	5.00
Bedrooms	1	3	3.30	3	4	6
Bathrooms	1	2	2.38	2	3	6
Cash purchase	0	0	0.14	0	0	1
Fire sale	0	0	0.09	0	0	1
Short sale	0	0	0.03	0	0	1
Years:	2000-2023					
Counties:	Delaware OH, Fairfield OH, Franklin OH, Licking OH					
Detroit (N = 901,112)						
Price (000s)	20.00	94.90	187.56	158.93	245.00	2,000.00
Days on market	1	12	65.22	36	84	730
Sqft (000s)	0.50	1.10	1.62	1.43	2.00	4.00
Age	1	30	48.04	51	64	123
Acres	0.00	0.00	0.28	0.13	0.26	5.00
Bedrooms	1	3	3.21	3	4	6
Bathrooms	1	1	2.15	2	3	6
Cash purchase	0	0	0.18	0	0	1
Fire sale	0	0	0.13	0	0	1
Short sale	0	0	0.04	0	0	1
Years:	2000-2023					
Counties:	Lapeer MI, Livingston MI, Macomb MI, Oakland MI, St. Clair MI, Wayne MI					
Houston (N = 1,128,086)						
Price (000s)	20.00	125.00	227.14	187.00	283.00	2,000.00
Days on market	1	12	72.80	36	96	730
Sqft (000s)	0.50	1.63	2.19	2.10	2.66	4.00
Age	1	9	25.17	21	38	123
Acres	0.00	0.14	0.25	0.17	0.22	5.00
Bedrooms	1	3	3.45	3	4	6
Bathrooms	1	2	2.60	2	3	6
Cash purchase	0	0	0.0005	0	0	1
Fire sale	0	0	0.09	0	0	1
Short sale	0	0	0.01	0	0	1
Years:	2000-2023					
Counties:	Brazoria TX, Fort Bend TX, Galveston TX, Harris TX, Montgomery TX					

Table A2: Descriptive statistics by MSA (cont.)

	Min	Pctl(25)	Mean	Median	Pctl(75)	Max
Los Angeles (N = 998,716)						
Price (000s)	20.00	350.00	600.48	518.78	750.00	2,000.00
Days on market	1	12	60.63	38	80	730
Sqft (000s)	0.50	1.23	1.72	1.57	2.09	4.00
Age	1	36	50.77	52	64	123
Acres	0.00	0.12	0.19	0.15	0.18	5.00
Bedrooms	1	3	3.29	3	4	6
Bathrooms	1	2	2.20	2	3	6
Cash purchase	0	0	0.11	0	0	1
Fire sale	0	0	0.07	0	0	1
Short sale	0	0	0.06	0	0	1
Years:	2000-2023					
Counties:	Los Angeles CA, Orange CA					
Las Vegas (N = 518,659)						
Price (000s)	20.00	155.00	269.21	242.00	341.00	2,000.00
Days on market	1	10	52.67	27	69	730
Sqft (000s)	0.50	1.47	1.94	1.80	2.30	4.00
Age	1	8	19.02	15	25	116
Acres	0.00	0.10	0.15	0.14	0.17	5.00
Bedrooms	1	3	3.38	3	4	6
Bathrooms	1	2	2.66	3	3	6
Cash purchase	0	0	0.25	0	0	1
Fire sale	0	0	0.20	0	0	1
Short sale	0	0	0.12	0	0	1
Years:	2000-2023					
Counties:	Clark NV					
Minneapolis (N = 898,986)						
Price (000s)	20.01	170.00	257.45	229.90	315.00	2,000.00
Days on market	1	13	67.01	33	81	730
Sqft (000s)	0.50	1.46	1.96	1.86	2.36	4.00
Age	1	15	39.78	34	59	123
Acres	0.00	0.00	0.32	0.19	0.31	5.00
Bedrooms	1	3	3.23	3	4	6
Bathrooms	1	2	2.25	2	3	6
Cash purchase	0	0	0.11	0	0	1
Fire sale	0	0	0.08	0	0	1
Short sale	0	0	0.02	0	0	1
Years:	2000-2023					
Counties:	Anoka MN, Carver MN, Chisago MN, Dakota MN, Hennepin MN, Ramsey MN, Scott MN, Sherburne MN, St. Croix MN, Washington MN, Wright MN					

Table A2: Descriptive statistics by MSA (cont.)

	Min	Pctl(25)	Mean	Median	Pctl(75)	Max
Seattle (N = 892,964)						
Price (000s)	20.50	245.00	440.61	361.00	550.00	2,000.00
Days on market	1	8	55.11	25	69	730
Sqft (000s)	0.50	1.42	1.96	1.86	2.41	4.00
Age	1	13	36.92	31	55	123
Acres	0.00	0.12	0.33	0.18	0.27	5.00
Bedrooms	1	3	3.34	3	4	6
Bathrooms	1	2	2.44	3	3	6
Cash purchase	0	0	0.09	0	0	1
Fire sale	0	0	0.05	0	0	1
Short sale	0	0	0.03	0	0	1
Years:	2000-2023					
Counties:	King WA, Pierce WA, Snohomish WA					
Saint Louis (N = 402,188)						
Price (000s)	20.00	117.00	211.43	179.90	270.00	1,950.00
Days on market	1	10	64.64	37	83	729
Sqft (000s)	0.50	1.12	1.66	1.49	2.06	4.00
Age	1	22	45.02	44	62	123
Acres	0.00	0.17	0.39	0.22	0.32	5.00
Bedrooms	1	3	3.21	3	4	6
Bathrooms	1	2	2.38	2	3	6
Cash purchase	0	0	0.12	0	0	1
Fire sale	0	0	0.08	0	0	1
Short sale	0	0	0.03	0	0	1
Years:	2000-2023					
Counties:	Franklin MO, Jefferson MO, Madison IL, St. Charles MO, St. Clair IL, St. Louis County MO, St. Louis MO					

B.2 Repeat Sales

Table B2: Asset Illiquidity (Repeat Sales)

	Raw	Double	Raw	Double	Raw	Double
	(1)	(2)	(3)	(4)	(5)	(6)
Fire sale	−0.242***	−0.173***	−0.231***	−0.168***	−0.181***	−0.122***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
Short sale	−0.158***	−0.119***	−0.148***	−0.114***	−0.130***	−0.098***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Invest-to-Invest			0.023***	0.012***	0.035***	0.023***
			(0.002)	(0.001)	(0.002)	(0.001)
Own-to-Invest			−0.028***	−0.021***	−0.014***	−0.007***
			(0.001)	(0.001)	(0.001)	(0.001)
Invest-to-Own			0.044***	0.022***	0.045***	0.023***
			(0.001)	(0.001)	(0.001)	(0.001)
Invest-to-Invest×Short sale					−0.052***	−0.040***
					(0.005)	(0.004)
Invest-to-Invest×Fire sale					−0.149***	−0.135***
					(0.004)	(0.004)
Own-to-Invest×Short sale					−0.046***	−0.044***
					(0.002)	(0.002)
Own-to-Invest×Fire sale					−0.101***	−0.096***
					(0.002)	(0.002)
Invest-to-Own×Short sale					−0.037***	−0.022***
					(0.003)	(0.003)
Invest-to-Own×Fire sale					−0.015***	−0.014***
					(0.003)	(0.003)
Observations	2,446,113	2,446,113	2,446,113	2,446,113	2,446,113	2,446,113
Adjusted R^2	0.967	0.974	0.968	0.974	0.968	0.975
House FE	✓	✓	✓	✓	✓	✓
Financing Controls	✓	✓	✓	✓	✓	✓
Tract×Qtr×Yr FE	✓	✓	✓	✓	✓	✓
Cleaned Full Remarks		✓		✓		✓

*p<0.1; **p<0.05; ***p<0.01

Note: Table B2 displays fire sale, short sale, and redeploy coefficient estimates for a subsample of repeat-sales transactions where we can reliably identify the buyer tenure type (i.e., owner-occupied vs. investor). The dependent variable in every column is the log of sales price. The fire sale and short sale coefficient estimates are relative to non-distressed transactions. The redeploy coefficient estimates are relative to *Own-to-Own* transactions. Every column includes house fixed effects and financing controls as well as quarter-by-year-by-census tract fixed effects. Even numbered columns include tokens from a double-selection variable selection process using the cleaned full remarks. Standard errors are clustered at the census tract level.

Table B3: Asset Illiquidity by Investment Horizon (Repeat Sales)

	Raw (1)	Double (2)	Raw (3)	Double (4)	Raw (5)	Double (6)
Fire sale	−0.242***	−0.173***	−0.219***	−0.163***	−0.173***	−0.118***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
Short sale	−0.158***	−0.119***	−0.139***	−0.110***	−0.124***	−0.096***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Invest-to-ST Invest			−0.041***	−0.031***	−0.017***	−0.010***
			(0.002)	(0.002)	(0.002)	(0.002)
Invest-to-LT Invest			0.024***	0.014***	0.032***	0.022***
			(0.001)	(0.001)	(0.002)	(0.001)
Own-to-ST Invest			−0.114***	−0.076***	−0.093***	−0.056***
			(0.002)	(0.001)	(0.002)	(0.001)
Own-to-LT Invest			0.004***	−0.001	0.013***	0.008***
			(0.001)	(0.001)	(0.001)	(0.001)
Invest-to-Own			0.033***	0.016***	0.033***	0.017***
			(0.001)	(0.001)	(0.001)	(0.001)
Invest-to-ST Invest×Short sale					−0.103***	−0.082***
					(0.010)	(0.008)
Invest-to-ST Invest×Fire sale					−0.188***	−0.176***
					(0.006)	(0.006)
Invest-to-LT Invest×Short sale					−0.021***	−0.019***
					(0.005)	(0.004)
Invest-to-LT Invest×Fire sale					−0.117***	−0.110***
					(0.004)	(0.004)
Own-to-ST Invest×Short sale					−0.050***	−0.053***
					(0.003)	(0.003)
Own-to-ST Invest×Fire sale					−0.114***	−0.112***
					(0.003)	(0.002)
Own-to-LT Invest×Short sale					−0.032***	−0.031***
					(0.002)	(0.002)
Own-to-LT Invest×Fire sale					−0.081***	−0.079***
					(0.002)	(0.002)
Invest-to-Own×Short sale					−0.022***	−0.014***
					(0.003)	(0.003)
Invest-to-Own×Fire sale					−0.010***	−0.012***
					(0.003)	(0.003)
Observations	2,446,113	2,446,113	2,446,113	2,446,113	2,446,113	2,446,113
Adjusted R^2	0.967	0.974	0.969	0.975	0.970	0.975
House FE	✓	✓	✓	✓	✓	✓
Financing Controls	✓	✓	✓	✓	✓	✓
Tract×Qtr×Yr FE	✓	✓	✓	✓	✓	✓
Cleaned Full Remarks		✓		✓		✓

*p<0.1; **p<0.05; ***p<0.01

Note: Table B3 displays fire sale, short sale, and redeploy coefficient estimates for a subsample of repeat-sales transactions where we can reliably identify the buyer tenure type (i.e., owner-occupied vs. investor). The dependent variable in every column is the log of sales price. The fire sale and short sale coefficient estimates are relative to non-distressed transactions. The redeploy coefficient estimates are relative to *Own-to-Own* transactions. Every column includes house fixed effects and financing controls as well as quarter-by-year-by-census tract fixed effects. Even numbered columns include tokens from a double-selection variable selection process using the cleaned full remarks. Standard errors are clustered at the census tract level.

B.3 Distressed Transactions by MSA

Figure B1 plots the percent of transactions that were liquidated via a fire sale (black, solid line) or short sale (red, dotted line) from 2000 through 2023 for each of the 10 MSAs. The paper provides a similar plot for the combined ten MSA transaction sample.

Figure B1: Fraction of Distressed Transactions by MSA

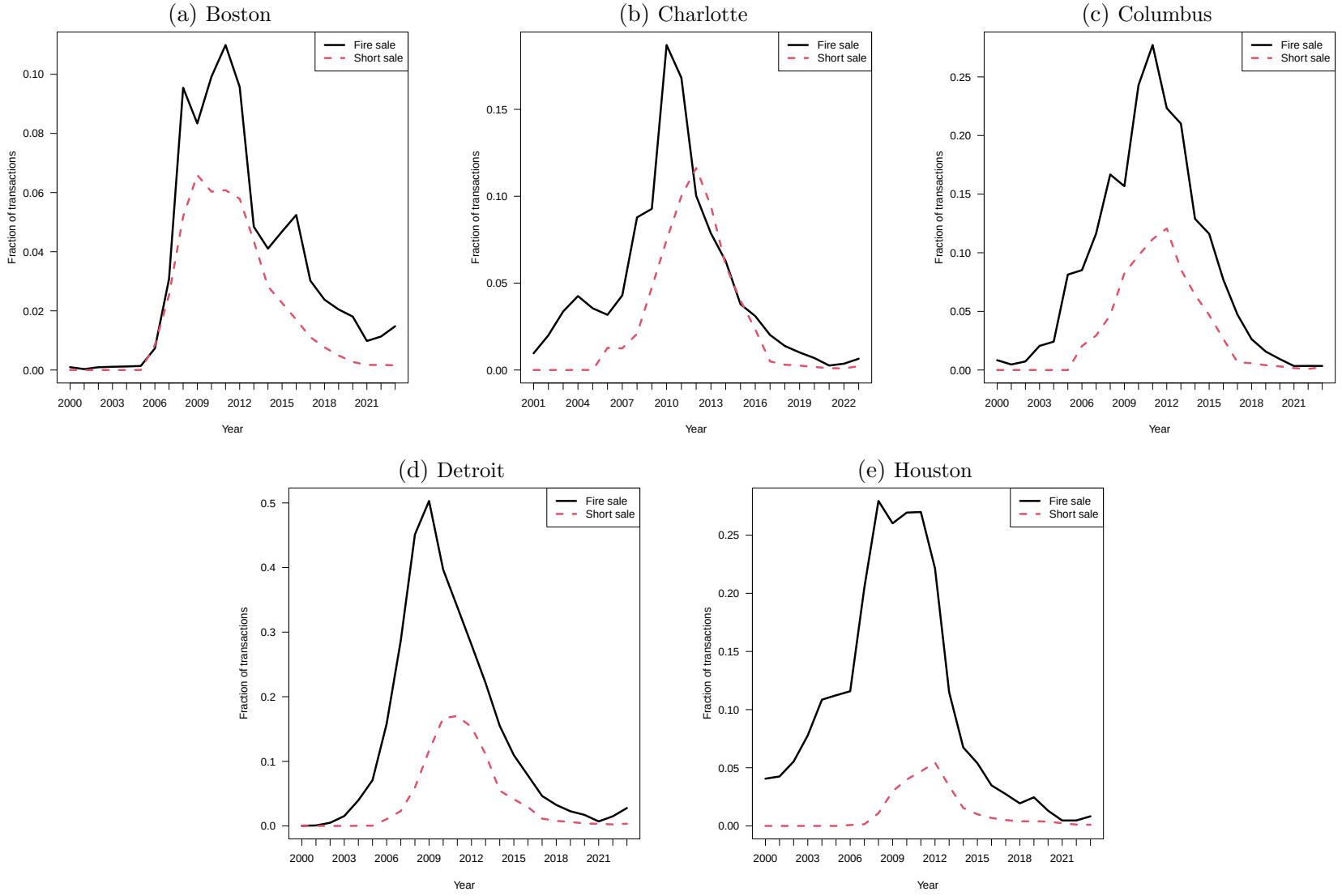
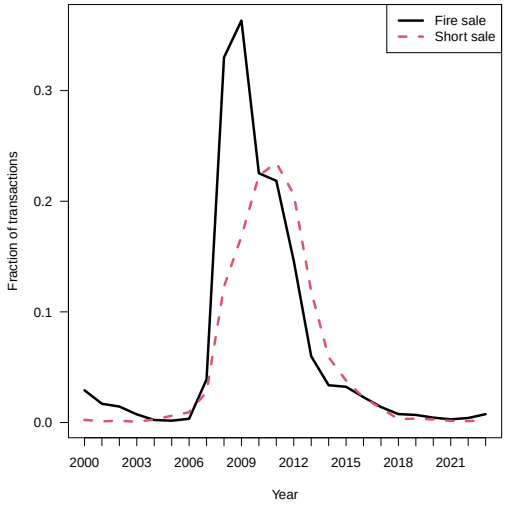
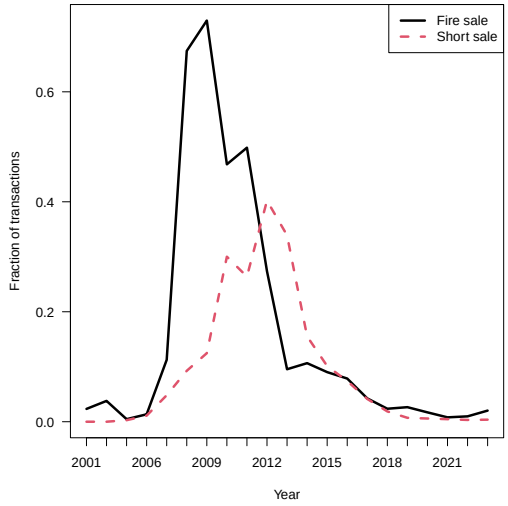


Figure B1: Fraction of Distressed Transactions by MSA (cont.)

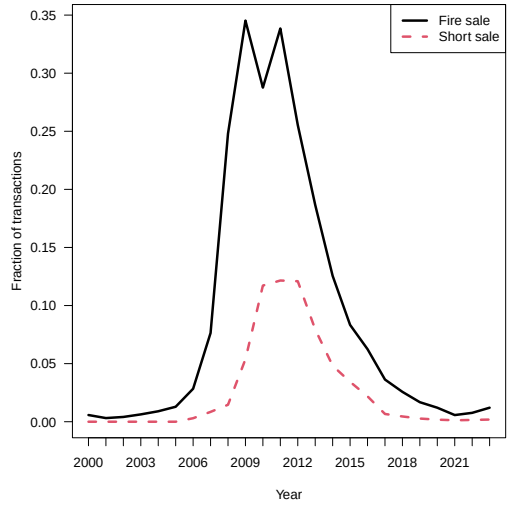
(f) Los Angeles



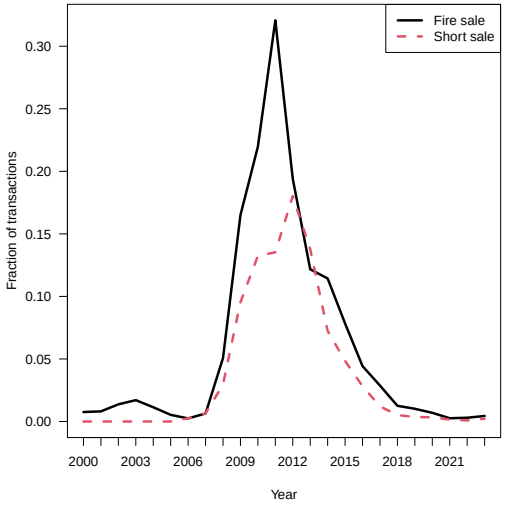
(g) Las Vegas



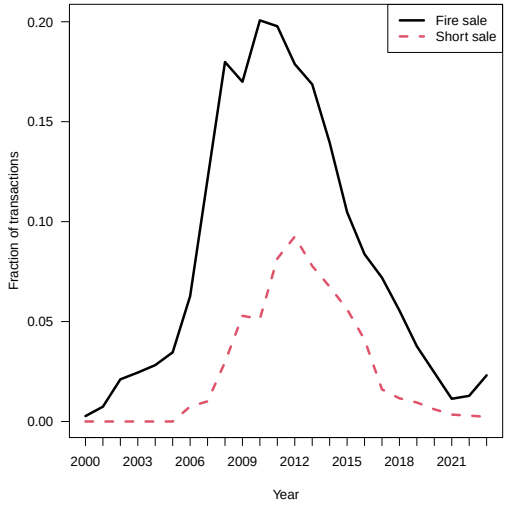
(h) Minneapolis



(i) Seattle



(j) Saint Louis



B.4 Fire Sale and Short Sale Coefficient Adjustments

Figure B3 plots short sale and fire sale (i.e., REO) token adjustments individually for the six MSAs not displayed in the body of the paper. Tokens that increase price, $\hat{\theta}_s > 0$, are plotted alongside blue circles. Tokens that decrease price, $\hat{\theta}_s < 0$, are plotted alongside red triangles. The paper provides a similar plot for the other four MSAs.

Figure B3: Fire sale and Short Sale Coefficient Adjustments



Table B4: Asset Illiquidity Subsample Results

	Raw (1)	Double (2)	Raw (3)	Double (4)	Raw (5)	Double (6)
Fire sale	−0.258***	−0.190***	−0.253***	−0.187***	−0.198***	−0.135***
	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Short sale	−0.160***	−0.126***	−0.158***	−0.125***	−0.141***	−0.110***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Invest-to-Invest			−0.038***	−0.027***	−0.026***	−0.016***
			(0.001)	(0.001)	(0.001)	(0.001)
Own-to-Invest			−0.036***	−0.026***	−0.020***	−0.011***
			(0.001)	(0.001)	(0.001)	(0.001)
Invest-to-Own			−0.003***	−0.004***	−0.002***	−0.003***
			(0.001)	(0.000)	(0.001)	(0.000)
Invest-to-Invest×Short sale					−0.020***	−0.021***
					(0.005)	(0.004)
Invest-to-Invest×Fire sale					−0.151***	−0.144***
					(0.005)	(0.004)
Own-to-Invest×Short sale					−0.070***	−0.063***
					(0.003)	(0.002)
Own-to-Invest×Fire sale					−0.126***	−0.118***
					(0.003)	(0.003)
Invest-to-Own×Short sale					−0.001	0.001
					(0.004)	(0.003)
Invest-to-Own×Fire sale					0.007**	−0.002
					(0.003)	(0.003)
Observations	1,492,098	1,492,098	1,492,098	1,492,098	1,492,098	1,492,098
Adjusted R^2	0.940	0.952	0.940	0.952	0.941	0.953
House Characteristics	✓	✓	✓	✓	✓	✓
Financing Controls	✓	✓	✓	✓	✓	✓
Tract×Qtr×Yr FE	✓	✓	✓	✓	✓	✓
Cleaned Full Remarks		✓		✓		✓

*p<0.1; **p<0.05; ***p<0.01

Note: Table B4 displays fire sale, short sale, and redeploy coefficient estimates for a subsample of transactions where we can reliably identify both the buyer and seller tenure type (i.e., owner-occupied vs. investor). The dependent variable in every column is the log of sales price. The fire sale and short sale coefficient estimates are relative to non-distressed transactions. The redeploy coefficient estimates are relative to *Own-to-Own* transactions. Every column includes housing characteristics and financing controls as well as quarter-by-year-by-census tract fixed effects. Even numbered columns include tokens from a double-selection variable selection process using the cleaned full remarks. Standard errors are clustered at the census tract level.

C Miscellaneous

C.1 Stimulating Housing Markets

We find house quality and asset illiquidity (i.e., redeployability) are both sizeable components of fire sale discounts in housing markets. In Section 6, we note that the sizeable contribution of redeployability to fire sale discounts suggests policies aimed at stimulating demand from owner-occupiers and investors may have a sizeable effect on market prices. Berger et al. (2020) find a first-time homebuyer policy had a sizeable effect on increasing demand from first-time homebuyers, often resulting in the reallocation of housing from low-value sellers to high-value buyers. To the best of our knowledge, no study nor policy has considered the effects of stimulating demand from investors.

The policies we consider are variations on reductions of the marginal tax rate from rental income at the federal level. This analysis uses the valuation equation for the price of the property at time 0 as the present value of after-tax rental income.

$$P(\boldsymbol{\tau}) = \sum_{0 \leq t \leq 30} \delta_t (1 - \tau_t) Rent_t \tag{9}$$

In Equation 9, δ_t is the discount factor for time period t, $\boldsymbol{\tau} = (\tau_{t=0}, \ldots, \tau_{t=30})'$ is the vector of marginal tax rate for the investor at times $t = 0 \ldots, 30$ where τ_t is assumed constant for the relevant $Rent_t$ at time t, and $Rent_t$ is rent at time t net of any operating expenses.

We consider the effect of alternative tax policies, holding all other variables constant, on the implied price and compare to the baseline price implied by the current tax rates. For example, if the baseline tax policy satisfies $\tau_t = \tau_B, \forall t$, the baseline price, P_B, is

$$P_B = (1 - \tau_B) \sum_{0 \leq t \leq 30} \delta_t Rent_t = (1 - \tau_B) K \tag{10}$$

If we limit counterfactual tax policies to a constant $\tau_t = \tau, \forall t$, the relative price effect of the tax policy for generic τ is the ratio $\frac{1-\tau}{1-\tau_B}$. The tax policy that eliminates all current and

future taxes on rental income implies a relative price increase of $\frac{1}{1-\tau_B}$.

For the baseline tax rates, we assume constant marginal tax rates of $\tau_{Bt} = \tau_B, \forall t$. We use statutory marginal tax rates for federal personal income tax in 2009, $\tau_B \in \{0.25, 0.28, 0.33, 0.35\}$. We also assume $\delta_t = \delta^t, \forall t$ and $Rent_t = Rent, \forall t$. For simplicity, we limit the counterfactual tax polities to the form

$$\tau_t = \begin{cases} 0 & \text{if } t \leq t^* \\ \tau_B & \text{if } t > t^* \end{cases} \tag{11}$$

Using Equation 9, the price under the counterfactual tax policies is

$$P(\boldsymbol{\tau}) = \sum_{0 \leq t \leq t^*} \delta^t Rent + \sum_{t^* < t \leq 30} \delta^t (1 - \tau_B) Rent \tag{12}$$

Table C1 displays the relative price effects of counterfactual price policies. In the body of the paper, we find previously owner-occupied fire sale properties purchased as investments sell for a (quality-adjusted) fire sale discount that is roughly twice the size of those that remain owner-occupied (22.5% vs. 11.8%). Depending on the investor's tax bracket, the results in Table C1 suggest that a 5 to 10-year tax break could increase the investors' valuation of fire sale properties enough to offset the asset illiquidity component of fire sale discounts that we document in the data.

Table C1: Tax Policy and House Prices

δ	t^*	τ_B	$\%\Delta P(\boldsymbol{\tau})$
0.96	1.00	0.25	1.89
0.98	1.00	0.25	1.47
0.96	5.00	0.25	8.72
0.98	5.00	0.25	7.05
0.96	10.00	0.25	15.82
0.98	10.00	0.25	13.42
0.96	30.00	0.25	33.33
0.98	30.00	0.25	33.33
0.96	1.00	0.28	2.20
0.98	1.00	0.28	1.71
0.96	5.00	0.28	10.17
0.98	5.00	0.28	8.22
0.96	10.00	0.28	18.46
0.98	10.00	0.28	15.65
0.96	30.00	0.28	38.89
0.98	30.00	0.28	38.89
0.96	1.00	0.33	2.79
0.98	1.00	0.33	2.17
0.96	5.00	0.33	12.88
0.98	5.00	0.33	10.41
0.96	10.00	0.33	23.38
0.98	10.00	0.33	19.82
0.96	30.00	0.33	49.25
0.98	30.00	0.33	49.25
0.96	1.00	0.35	3.05
0.98	1.00	0.35	2.37
0.96	5.00	0.35	14.08
0.98	5.00	0.35	11.38
0.96	10.00	0.35	25.56
0.98	10.00	0.35	21.67
0.96	30.00	0.35	53.85
0.98	30.00	0.35	53.85

Note: Table C1 reports model primitives, counterfactual tax policies, and relative price effects of counterfactual price policies. Baseline prices are given by

$$P_B = \sum_{0 \leq t \leq 30} \delta^t (1 - \tau_B) Rent \tag{13}$$

Counterfactual prices are given by

$$P(\boldsymbol{\tau}) = \sum_{0 \leq t \leq 30} \delta^t (1 - \tau_t) Rent \tag{14}$$

with $\tau_t = 0, t \leq t^*$ and $\tau_t = \tau_B, t > t^*$. Column 4 is calcualted as $\%\Delta P(\boldsymbol{\tau}) = 100\% \times \left(\frac{P(\boldsymbol{\tau})}{P_B} - 1 \right)$