Liquidity in Residential Real Estate Markets*

Nadia Kotova† and Anthony Lee Zhang‡

December 2020

Abstract

We develop two model-based measures of housing market liquidity: time-on-market and price dispersion. We show how these arise from a search-and-bargaining model of housing markets, in which sellers face a trade-off between selling quickly and selling at high prices. The tradeoffs faced by individual sellers aggregate to a market-level relationship between average time on market and average idiosyncratic price dispersion. We measure time-on-market and price dispersion in a large sample of US counties, and show cross-sectional and time-series evidence supporting the predictions of our model. Calibrated to the data, the model can simultaneously match the macro-relationships between TOM and PD, and produce micro-estimates of the TOM-price tradeoff which are consistent with estimates in the housing literature.

Keywords: Housing, Search, Liquidity

*We appreciate comments from Mohammad Akbarpour, Sam Antill, Alina Arefeva, Dmitry Arkhangelsky, Adrien Auclert, Lanier Benkard, Tim Bresnahan, Scarlet Chen, Daniel Chen, John Cochrane, Cody Cook, Rebecca Diamond, Evgeni Drynkin, Darrell Duffie, Liran Einav, Matthew Gentzkow, Andra Ghent, Sonia Gilbukh, Adam Guren, Guido Imbens, Chad Jones, Eddie Lazear, Paul Milgrom, Jonas Mueller-Gastell, Sean Myers, Michael Ostrovsky, Amine Ouazad, Monika Piazzesi, Peter Reiss, Al Roth, Jacob Sagi, Martin Schneider, Jesse Shapiro, Erling Skancke, Paulo Somaini, Amir Sufi, Rose Tan, Zach Taylor, Chris Tonetti, Jia Wan, Joseph Vavra, Robert Wilson, Ali Yurukoglu, and Fudong Zhang, as well as seminar participants at Stanford, UChicago, the Wisconsin Business School, the Trans-Atlantic Doctoral Conference, the Urban Economics Association 2019 meeting, the Young Economists Symposium, the XX April International Academic Conference, Tsinghua PBCSF, and the 2019 West Coast Search & Matching Workshop.

†Stanford Graduate School of Business, 655 Knight Way, Stanford, CA 94305; nkotova@stanford.edu.

‡University of Chicago Booth School of Business, 5807 S Woodlawn Ave, Chicago, IL 60637; anthony.zhang@chicagobooth.edu.

1 Introduction

Residential real estate is a very illiquid asset class. Many measures designed to measure the "liquidity" or "hotness" of housing markets, in the cross-section and over time, are closely monitored and discussed by industry participants.[1] While these measures often have intuitive interpretations, there has been relatively little academic work discussing how liquidity should be measured, in a microfounded model of frictional housing markets.

In this paper we develop model-based aggregate measures of housing market liquidity. We focus on two measures: seller time-on-market and idiosyncratic house price dispersion. We show how these arise from a search-and-bargaining model of housing markets, in which sellers face a trade-off between selling quickly and selling at high prices. We use this theoretical framework to argue that neither time on market nor price dispersion alone are sufficient as metrics of market liquidity. However, we can use the empirical relationship between the two measures and the structure of our model to quantify the menu of prices and time on markets individual sellers face. Under our model estimates, a seller who chooses to spend an additional month on the market sells at a 5.33% higher price.

We start by building a tractable search-and-bargaining model of a housing market. In our model, price dispersion arises from heterogeneity in seller and buyer preferences: sellers have different utility costs of keeping their houses on the market per unit time, and buyers receive an independent match quality shock every time they match with a house. Both dimensions of heterogeneity affect distribution of prices in equilibrium. Buyers with higher match quality draws pay higher prices, and sellers with higher holding costs sell for lower prices. We study how the supply of liquidity, measured by ratio of buyers to sellers, and the demand for liquidity, measured by average holding costs of sellers, determine time on market and price dispersion in equilibrium.

In sellers' markets with many buyers, liquidity supply is high, and time on market and price dispersion are both low. Hence, shifts in liquidity supply cause time on market and price dispersion to co-move positively. On the contrary, shifts in liquidity demand cause them to co-move negatively. When sellers have higher holding costs, their demand for liquidity is higher. Sellers collectively sell faster, even though this causes prices to be lower

[1] Some examples are Realtor.com's Market Hotness Index, Zillow's report on Hottest Markets for 2019, Redfin's report on Hottest Neighborhoods 2020, and an Inman article on hotness in the Chicago housing market.

and more dispersed. This leads to lower time-on-market, but higher price dispersion. This implies that neither time on market nor price dispersion alone are sufficient for measuring housing market liquidity. In particular, time on market can be low both because liquidity supply is high or because liquidity demand is high. We next argue that patterns in the data are consistent with both supply and demand factors being important drivers of equilibrium outcomes.

To test our model predictions in the data, we first measure average time on market and average idisyncratic price dispersion at the county level. We use the CoreLogic MLS and Deed datasets, which contain US-wide microdata on house listings and sales respectively, to measure time on market and idiosyncratic price dispersion. To measure price dispersion, we regress log sale prices of houses on county-month fixed effects, house fixed effects, and a smooth function of house characteristics and time. This approach combines repeat-sales and hedonic specifications for house prices: it absorbs time-invariant house quality into the house fixed effects, and also allows observable house characteristics to affect the evolution of house prices over time. We measure price dispersion using the squared residuals from this regression, which we can aggregate flexibly across locations and over time.

This strategy for measuring price dispersion is novel to the literature. The main innovation is that we use individual house sales rather than returns between sales as the primary unit of analysis. The benefit of our approach is that our estimates of idiosyncratic price dispersion can be flexibly aggregated cross-sectionally and over time, while return-based dispersion estimates can only be aggregated cross-sectionally. This measurement strategy allows us to produce a set of novel stylized facts about idiosyncratic price dispersion that we describe next.

Before testing our model predictions, we document a number of stylized facts about the behavior of time-on-market and price dispersion in the time series. We show that time-on-market and price dispersion are countercyclical and seasonal, and track other measures of "market hotness", such as sales volume and average price. At business cycle frequencies, both price dispersion and time-on-market are countercyclical: both metrics fell during the 2000-2006 housing boom, rose during the 2006-2012 crash, and fell again during the recovery. Seasonally, price dispersion and time-on-market are lower in the summer hot season, and higher in the winter cold season. These results suggest that time-on-market and price dispersion both behave like measures of market hotness in the time series. While time on market has been extensively studied in the literature, we believe we are the first to document the seasonality and cyclicity of idiosyncratic house price dispersion.

We proceed to test the predictions of our model in the cross-section of counties. In order to test the model, we need empirical proxies for liquidity supply and demand. We use two proxies for liquidity supply: county net migration rate and vacancy rate. Net migration rate is a direct measure of buying pressure: a county which has high inmigration rate should have a large mass of buyers interested in purchasing houses. Vacancy rate is an inverse measure of liquidity supply: high vacancy rate indicates that there is a large quantity of houses waiting to be sold, relative to buyers, indicating that the supply of liquidity is low. Using these proxies, we verify the predictions of our model. Net inmigration is negatively correlated with both price dispersion and time-on-market, and vacancy rates are positively correlated with both liquidity metrics.

Our proxy for liquidity demand is the ratio of average household income to average house prices. Intuitively, household income is a direct measure of the value of house sellers' time. Holding fixed house prices, higher-income sellers should rationally be willing to sell faster, even if this lowers sale prices, since their opportunity cost of time is higher. Our model thus predicts that higher-income areas should have lower time-on-market, but actually higher price dispersion. We verify this prediction in the data.

Our empirical results survive a number of robustness checks. Our results are robust to three other methods for estimating price dispersion: a pure hedonic price model, a pure repeat-sales model, and a model in which estimated residuals are nonparametrically adjusted for the number of times a house is sold, and the average time between house sales. Our results are robust to using two alternative time-on-market measures, from Zillow and Realtor.com. Moreover, we argue that our empirical results are difficult to explain using alternative theories, such as unobserved house heterogeneity, asymmetric information, and adverse selection.

Quantitatively, we use empirical relationship between time on market and price dispersion to estimate the menu of prices and time-on-markets that sellers with different holding costs face. Under our model estimates, a seller who chooses to spend an additional month on the market sells at a price 5.33% higher. We survey a number of papers in the literature about the micro-trade off between time on market and prices, and show that estimates of the 1-month effect range from 1.9% to around 11%. Thus, our model shows that a single set of primitives can simultaneously match the macro-relationship between time on market and price dispersion, and the micro-level liquidity discounts.

1.1 Related literature

This paper is related to a number of papers that study idiosyncratic house price volatility. An early paper studying the topic is Case and Shiller (1988). Giacoletti (2017) studies the residential real estate market in California and shows that idiosyncratic house price risk does not scale with holding period, which suggests that much of idiosyncratic house price volatility is due to liquidity risk. Sagi (2015), using commercial real estate data, also documents that idiosyncratic risk does not scale one-to-one with holding periods, and constructs a search model to rationalize these results.

Peng and Thibodeau (2017) calculates price dispersion at the zipcode level using a hedonic regression specification and documents relationships between idiosyncratic price dispersion and characteristics of zipcodes such as average income. Two other papers which measure idiosyncratic price dispersion are Anenberg and Bayer (2015) and Landvoigt, Piazzesi and Schneider (2015). Another related paper is Ben-Shahar and Golan (2019), who show that improved disclosure of transaction prices reduced price dispersion in the Israeli housing market.

In terms of the modeling approach, our work fits into the literature on applying search-and-bargaining models[2] to housing markets[3] and to financial markets more generally.[4] Relative to this literature, an innovation of our model is that we allow for continuously distributed persistent seller values as well as match-specific quality shocks. In an appendix, we show that the model can also accommodate continuously distributed persistent buyer values.

A few papers which have closely related models to ours are Albrecht et al. (2007), Albrecht, Gautier and Vroman (2016), Anenberg and Bayer (2015), and Sagi (2015). Albrecht et al. (2007) study a random search model of the housing market with two seller types. Albrecht, Gautier and Vroman (2016) construct a directed housing search model with two seller types, which can predict sellers' asking prices and the sale-to-ask spread. Anenberg

[2]For earlier applications to labor markets, see, for example, Mortensen and Pissarides (1994), or the survey article Rogerson, Shimer and Wright (2005)

[3]See, for example, Wheaton (1990), Albrecht et al. (2007), Piazzesi and Schneider (2009), Genesove and Han (2012), Anenberg and Bayer (2015), Ngai and Tenreyro (2014), Head, Lloyd-Ellis and Sun (2014), Piazzesi, Schneider and Stroebel (2015), Sagi (2015), Albrecht, Gautier and Vroman (2016), Guren (2018), Gabrovski and Ortego-Marti (2019*a*), Gabrovski and Ortego-Marti (2019*b*), Arefeva (2019), Guren and McQuade (2019), and Ouazad and Rancière (2019).

[4]See, for example, Duffie, Gârleanu and Pedersen (2005), Duffie, Gârleanu and Pedersen (2007), Vayanos and Wang (2007), Vayanos and Weill (2008), Weill (2008), Afonso (2011), Duffie, Qiao and Sun (2017), Gavazza (2016), Trejos and Wright (2016), Hugonnier, Lester and Weill (2018).

and Bayer (2015) allows continuous match quality shocks and a discrete number of seller types. Sagi (2015) allows a discrete number of seller types.

The fact that volume, prices, and time-on-market are correlated in housing markets is the subject of a number of papers; see, for example, Stein (1995), Krainer (2001), Genesove and Mayer (2001), Leung, Lau and Leong (2002), Clayton, Miller and Peng (2010), Diaz and Jerez (2013), and DeFusco, Nathanson and Zwick (2017). Our contribution to this literature is to show how idiosyncratic price dispersion co-moves with these variables at the aggregate level.

Two other papers that analyze the relationship between list prices and time-on-market at the level of individual house sales are Drenik, Herreno and Ottonello (2019), who study Spanish commercial real estate, and Guren (2018). Drenik, Herreno and Ottonello (2019) attempts to rationalize the data using adverse selection, without search frictions, and Guren (2018) emphasizes strategic complementarity in sellers' price-setting decisions. Our model abstracts away from both of these forces, focusing on market tightness and its effects on price dispersion.

1.2 Outline

The rest of the paper proceeds as follows. Section 2 describes our model, theoretical results, and predictions. Section 3 describes our data and measurement strategy. Section 4 contains our empirical results, and Section 5 does robustness checks and discusses alternative explanations of our results. Section 6 describes our calibration. Section 7 discusses implications of our findings and concludes.

2 Model

2.1 Primitives

There is a unit mass of houses which are ex-ante identical. Time is continuous, and all agents discount the future at rate r. There are three kinds of agents in the model: sellers, buyers, and matched homeowners. The lifecycle of agents in the model is as follows: buyers purchase houses and become matched homeowners, matched homeowners receive separation shocks to become sellers, and sellers leave the market upon successfully selling their houses.

Sellers. We use M_S to denote the mass of sellers in the market who are waiting to sell their houses to buyers. Sellers are matched homeowners who have received separation shocks from their houses. Once a seller successfully sells her house, she permanently leaves the market, attaining a continuation value which we normalize to 0.

Sellers are heterogeneous, characterized by some time-invariant *holding cost*, c, which she incurs per unit time her house is on the market. c is drawn from $F(\cdot)$ when a homeowner unmatches from her house and becomes a seller. In the main text, we assume that $F(\cdot)$ is uniform, with mean $\bar{c}$ and support $[\bar{c} - \Delta_c, \bar{c} + \Delta_c]$. We use $V_S(c)$ to denote the expected utility of a seller with cost c in equilibrium. We use $F_{eq}(c)$ denote the distribution of holding utilities among sellers in stationary equilibrium, which will in general differ from $F(c)$.

The parameter c represents the dollar value of sellers for staying on the market a unit of time longer. Sellers with higher values of c are less willing to stay on market, and will tend to sell faster, but at lower prices. Sellers with low values of c are willing to wait longer to sell for higher prices. c thus represents a classic tradeoff documented in the housing literature, between time-on-market and prices.

There are many possible sources of differences in c. Since selling houses involves significant time and effort for owners, agents with higher incomes may have higher values of c; they may be willing to sell faster, even if this means accepting lower prices. Sellers who are moving within the same city may be willing to hold on to their houses for longer than sellers who are moving out of the city.[5] Sellers who have less home equity to extract upon sale may have a relatively high value of cash, and thus will be willing to wait longer to sell at higher prices (Genesove and Mayer (1997), Guren (2018)). We survey papers which document other drivers of sellers' effective holding costs in table 5, which we discuss in subsection 6.2. Our model remains agnostic with respect to the exact source of this heterogeneity.

Buyers. Our model of buyers is similar to Head, Lloyd-Ellis and Sun (2014). Potential buyers enter the market at some exogeneous flow rate η_B, and each buyer draws a value $\xi \sim H(\cdot)$ for entering the city. ξ can be thought of as representing the attractiveness of amenities and job opportunities in the county, which may vary idiosyncratically across

[5] This is related to Anenberg and Bayer (2015), who study a model in which sellers may buy houses before attempting to sell their houses; sellers who buy before selling and who sell before buying will thus endogenously have different flow utilities from their houses.

buyers. After observing ξ, each potential buyer can choose to either enter the city, receiving utility ξ immediately and becoming an active homebuyer, or leave forever, receiving utility normalized to 0. As in Head, Lloyd-Ellis and Sun (2014), we assume that entry decisions are irreversible.

We use M_B to denote the mass of active buyers who are present in the market. All active buyers are identical, and receive flow utility normalized to 0 while waiting to buy a house. Buyers meet sellers through a process we describe below. When a buyer meets a seller, he draws, independently across matches, some idiosyncratic *match utility* $\epsilon \sim G(\cdot)$ for the house. If the buyer buys the house, he becomes a matched homeowner, receiving ϵ from the house per unit time, until he receives a separation shock and becomes a seller. We assume that $G(\cdot)$ is non-centered exponential, with lower bound ϵ_0, and standard deviation σ_ϵ. We use V_B to denote the expected value of an unmatched buyer in equilibrium.

Matched homeowners. Matched homeowners are buyers who have purchased houses, and have not yet received separation shocks to become sellers. Each house is owned either by a matched homeowner or a seller, so the mass of matched homeowners is always $1 - M_S$. A homeowner with match utility ϵ receives flow utility ϵ from their house per unit time. Homeowners receive separation shocks at Poisson rate λ_M. When a homeowner receives a separation shock, she draws some $c \sim F(\cdot)$, and becomes a seller with cost c.

We use $G_{eq}(\epsilon)$ to denote the distribution of match utilities among matched homeowners; this will, in general, differ from $G(\epsilon)$, the distribution of match utilities that buyers draw when they meet sellers, because higher draws of ϵ are more likely to result in a successful trade.

Price determination. We assume that prices are set through bilateral Nash bargaining. Suppose that a buyer is matched with a seller with holding cost c, and the buyer draws match utility ϵ. If the buyer purchases at price P, the buyer receives $V_M(\epsilon) - P$, and the seller leaves the market and receives P. If the buyer does not purchase, the buyer receives V_B and the seller receives $V_S(c)$. The bilateral match surplus from trade, as a function of ϵ and c, is thus:

$$V_M(\epsilon) - V_B - V_S(c) \tag{1}$$

We assume trade occurs in all cases where the bilateral match surplus is nonnegative. Thus, a seller with holding cost c will trade with any buyer with match utility ϵ higher than some *trade cutoff* $\epsilon^*(c)$, which satisfies:

$$V_M(\epsilon^*(c)) = V_B + V_S(c)$$

When trade occurs, we assume that the price is set to give the seller a share θ of the bilateral match surplus. That is, the trade price $P(c,\epsilon)$ is:

$$P(c,\epsilon) = V_S(c) + \theta\left(V_M(\epsilon) - V_B - V_S(c)\right) \tag{2}$$

We will assume that ϵ_0 is sufficiently low that, in equilibrium,

$$\epsilon^*(c) \geqslant \epsilon_0 \; \forall c$$

that is, no seller type wishes to trade with all buyer types.

2.2 Stationary equilibrium

Stationary equilibrium, in our model, requires two sets of conditions to be satisfied: the decisions of entrants, buyers, sellers, and matched homeowners must be optimal; and inflows and outflows of all kinds of agents must be equal. The following proposition summarizes the equilibrium conditions in our model. Formal derivations of these conditions are in Appendix A.1.

Proposition 1. *Given primitives:*

$$r, \alpha, \phi, \theta, \lambda_M, \eta_B, \bar{c}, \Delta_c, \epsilon_0, \sigma_\epsilon, H(\cdot)$$

a stationary equilibrium of the model is described by buyer and seller masses M_B, M_S*, stationary distributions* $F_{eq}(c)$ *and* $G_{eq}(\epsilon)$*, matching rates* λ_S, λ_B*, value functions* $V_S(c), V_M(\epsilon), V_B$*, and a trade cutoff function* $\epsilon^*(c)$*, which satisfy the following conditions:*

Buyer, seller, and matched owner Bellman equations:

$$rV_B = \lambda_B \int\int_{\epsilon > \epsilon^*(c)} \left[(1-\theta)\left(V_M(\epsilon) - V_B - V_S(c)\right)\right] dG(\epsilon)\, dF_{eq}(c) \tag{3}$$

$$rV_S(c) = -c + \lambda_S \int_{\epsilon > \epsilon^*(c)} \theta\left(V_M(\epsilon) - V_B - V_S(c)\right) dG(\epsilon) \tag{4}$$

$$rV_M(\epsilon) = \epsilon + \lambda_M \left(\int V_S(c)\, dF(c) - V_M(\epsilon) \right) \tag{5}$$

Trade cutoffs:

$$V_M(\epsilon^*(c)) = V_S(c) + V_B \tag{6}$$

Matching rates:

$$M_S \lambda_S = M_B \lambda_B = \alpha M_B^{\phi} M_S^{1-\phi} \tag{7}$$

Flow equality:

$$(1 - M_S)\lambda_M f(c) = \lambda_S M_S f_{eq}(c)(1 - G(\epsilon^*(c))) \tag{8}$$

$$G_{eq}(\epsilon) = \frac{\int_c \lambda_S M_S \left[\int_{\tilde{\epsilon}=\epsilon_0}^{\epsilon} 1(\tilde{\epsilon} > \epsilon^*(c))\, dG(\tilde{\epsilon}) \right] dF_{eq}(c)}{\int_c \lambda_S M_S (1 - G(\epsilon^*(c)))\, dF_{eq}(c)} \tag{9}$$

$$(1 - M_S)\lambda_M = \eta_B (1 - H(-V_B)) \tag{10}$$

2.3 Time-on-market and price dispersion

Claim 1. In stationary equilibrium, the variance of $P(c, \epsilon)$ among trading sellers and buyers is:

$$Var(P(c, \epsilon)) = \underbrace{Var_{c \sim F(\cdot)}(V_S(c))}_{\text{Seller holding costs}} + \underbrace{\left(\frac{\theta \sigma_\epsilon}{r + \lambda_M} \right)^2}_{\text{Buyer match utility}} \tag{11}$$

Expected time-on-market for a seller of type cis:

$$TOM(c) = E\left[\frac{1}{\lambda_S (1 - G(\epsilon^*(c)))} \right] \tag{12}$$

Expression (11) shows that equilibrium price dispersion arises from two sources: differences in sellers' value functions $V_S(c)$, caused by differences in sellers' holding utilities; and differences in buyers' match utilities ϵ. Equilibrium time-on-market of seller of type c depends on matching rate, λ_S, and the probability of trade conditional on a match, $1 - G(\epsilon^*(c))$.

The seller value term $V_S(c)$ is an equilibrium object which we cannot analytically characterize, but there is a simple analytical expression for its derivative $V_S'(c)$.

Claim 2. The seller value function $V_S(c)$ satisfies:

$$V_S'(c) = \frac{-TOM(c)}{rTOM(c) + \theta} \tag{13}$$

Qualitatively, Claim 2 shows that $V_S'(c)$ is higher, and hence, the dispersion in seller values is higher, when equilibrium time-on-market $TOM(c)$ is higher. Intuitively, suppose M_B/M_S is high, so markets are tight. All sellers spend relatively little time on the market, so the relative benefit of being a low-cost seller is smaller.

2.4 Comparative statics and predictions

In Figure 1, we simulate our model. The top two panels show the micro-level tradeoff between time-on-market and price menu faced by individual sellers. Each line is one equilibrium. The dots represent the TOM and price pairs chosen by 0th, 30th, 60th, and 100th percentile sellers.

The top left panel shows what happens when we shift the buyer inflow rate, η_B. When η_B increases, there are more buyers. We can think of this as liquidity supply being higher. Hence, time-on-market decreases for all sellers: the entire curve shifts towards the left. Moreover, the menu compresses vertically: the prices attained by different types of sellers are closer to each other. Hence, price dispersion also declines. Thus, when η_B – liquidity supply – increases, both time-on-market and price dispersion decrease.

The top right panel shows what happens when we change the seller average cost, $\bar{c}$. This parameter can be thought of as liquidity demand parameter: the higher the average cost $\bar{c}$, the more costly it is for sellers to stay on the market, and the higher their demand is for low time on market. When $\bar{c}$ increases, sellers have higher urgency to sell. This decreases time-on-market, as sellers try to sell faster, but it actually increases price dispersion: the gaps between the black dots increase. Increases in $\bar{c}$ cause average prices to decrease, increasing relative price dispersion. Thus, when liquidity demand increases, time-on-market decreases, but price dispersion actually increases.

These effects are summarized in the bottom panel of Figure 1. This plot shows the macro-quantities: average time-on-market on the x-axis, and market-level price dispersion as a percentage of average price on the y-axis. Each curve is a fixed value of η_B, and shows outcomes as we vary $\bar{c}$. We see that, holding fixed liquidity supply η_B, varying liquidity demand, $\bar{c}$, causes time-on-market and price dispersion to co-move negatively.

A simple, though informal, intuition is that, as sellers are more impatient and have higher liquidity demand, holding fixed liquidity supply, sellers move along a kind of "liquidity supply curve", trading off aggregate lower time-on-market for higher aggregate price dispersion. These "liquidity supply curves" are downward sloping lines. Conversely, if η_B increases holding fixed $\bar{c}$, the entire frontier shifts: the market is able to attain both lower time-on-market and lower price dispersion.

Together, these comparative statics allow us to make two core predictions, which we can bring to the data.

Prediction 1. *Suppose a variable Z_i is correlated with liquidity supply: it is positively correlated with η_B, the buyer entry rate. Z_i should be negatively correlated with price dispersion and time-on-market.*

Prediction 1 states that any variables which are correlated with liquidity supply should be correlated with price dispersion and time-on-market in the same direction.

Prediction 2. *Suppose a variable Z_i is correlated with liquidity demand: it is positively correlated with $\bar{c}$, the average seller cost. Z_i should be negatively correlated with time-on-market, but positively correlated with price dispersion.*

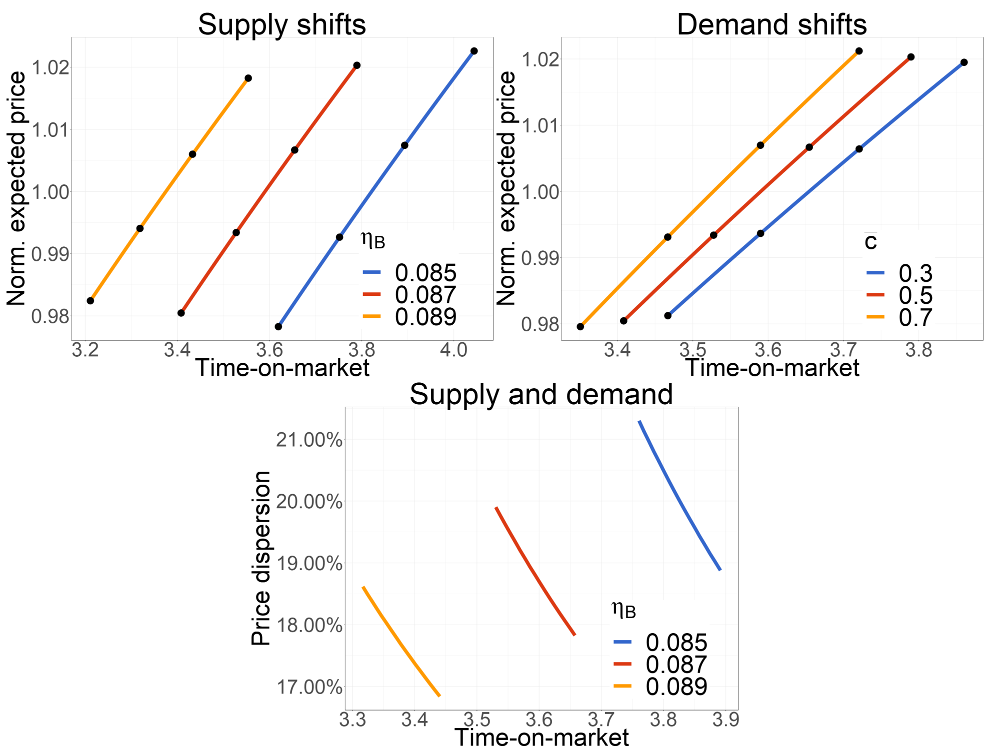
The intuition for Prediction 2 is from the comparative statics in Figure 1: increases in liquidity demand will tend to lead to lower time-on-market, but actually higher price dispersion.

2.5 Extensions

Our baseline model assumes buyers are undifferentiated for simplicity, as it implies that V_B is a scalar, so the model is easier to solve computationally. A natural extension of the model is to allow buyers to have persistent heterogeneous types so that buyers' flow utilities for remaining unmatched can differ. However, the model is robust to including buyer heterogeneity. Appendix A.6 derives equilibrium conditions under persistent buyer heterogeneity, and shows that the main theoretical results, Claims 1 and 2, are unchanged.

In the model, we assume the only drivers of price dispersion are sellers' and buyers' preferences. In reality, there are many other possible drivers of price dispersion, besides these two. For example, realtor quality affects time-on-market and prices Gilbukh and Goldsmith-Pinkham (2019). We abstract away from these for simplicity; they can be thought of as adding additional error terms to the price equation (2) in the model.

Figure 1: Liquidity supply and demand



Notes. Model comparative statics.

Table 1: Descriptive statistics, county sample

Name	Mean	SD	P10	P90
Average monthly sales	414	595	70	918
Mean price (x1000 USD)	488.4	1454.5	142.7	579.2
Mean TOM (Months)	3.25	0.75	2.44	4.20
Total counties	477			
Total sales	11,835,915			

Notes. Monthly sales and mean price data are from the CoreLogic Deed dataset, and mean time-on-market data is from the CoreLogic MLS dataset.

3 Data and measurement

3.1 Data sources

The main data source we use for house prices is microdata on single-family house sales and house characteristics from the CoreLogic Tax and Deed database, spanning the time period 2000-2017. For time-on-market, we use the Corelogic MLS dataset, and we use demographic data on county-years and counties from ACS 1-year and 5-year samples. Further details of data sources we use and data cleaning steps are described in Appendix B.

Since we estimate price dispersion using a repeat-sales specification, we filter to counties with a large enough number of sales; details of how we select counties are described in Appendix B.1. Descriptive statistics for counties in our primary estimation sample are shown in Table 1. Our primary dataset comprises 11 million house sales within 477 counties, over the period 2000-2017. As Appendix Table A1 shows, our estimation sample contains 14.8% of all counties, but covers 61.8% of the US population. Our sample is concentrated in relatively large, dense, and high-income counties, but is representative in terms of other demographic characteristics.

3.2 Measuring idiosyncratic price dispersion

We measure house price dispersion by regressing observed house sale prices on a set of predictors, and taking the residual. Our preferred specification for log house prices is:

$$p_{it} = \gamma_i + \eta_{ct} + f_c\left(x_i, t\right) + \epsilon_{it} \tag{14}$$

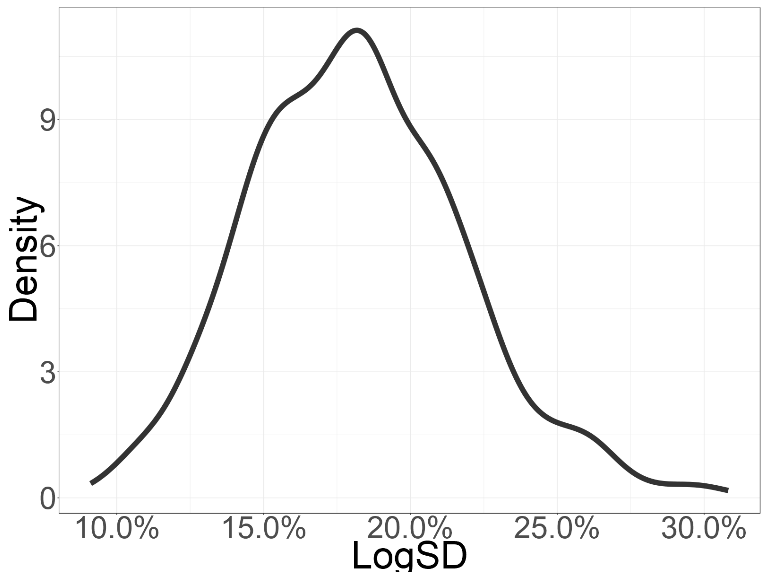
Where i indexes properties, c indexes counties, and t indexes months. In words, (14) says that log prices p_{it} are determined by a time-invariant house fixed effect, γ_i, a county-month fixed effect, η_{ct}, a smooth function $f_c\left(x_i, t\right)$ of observable house characteristics x_i and time t, and a mean-0 error term ϵ_{it}.

Specification (14) combines elements of repeat-sales and hedonic models of house prices. The house fixed effect term, γ_i, absorb all features of a house, observed and unobserved, which have time-invariant effects on the price of house i. The η_{ct} term absorbs parallel shifts in log house prices in a county over time. The $f_c\left(x_i, t\right)$ term allows houses with different observable characteristics x_i to appreciate at different rates: for example, the $f_c\left(x_i, t\right)$ term allows larger houses to appreciate faster than smaller houses, or houses in the east of a certain county to appreciate faster than houses in the west. Additional details on how we implement specification (14) are described in Appendix B.6.

We estimate price dispersion at the level of individual house sales using the estimated residuals, $\hat{\epsilon}_{it}$, from (14). While each individual estimate is very noisy, these estimates can be flexibly aggregated over time and across geographical regions. For example, we will use $\hat{\sigma}_c$ to denote the empirical estimate of standard deviation of all $\hat{\epsilon}_{it}$ terms in county c. $\hat{\sigma}_c$ can be thought of as the log standard deviation of house prices, after controlling for features in (14), so we will sometimes refer to these estimates as logSD. As we describe in Appendix B.6, we apply a degrees-of-freedom adjustment to $\hat{\epsilon}_{it}$, so the squared error estimates are unbiased at the county level.

In Figure 2, we plot the distribution of $\hat{\sigma}_c$. The mean of $\hat{\sigma}_c$ is 18.2% of house prices, and the standard deviation is 3.64%. The 10th percentile is 13.8% and the 90th percentile is 22.6%.

Figure 2: Distribution of $\hat{\sigma}_c$



Notes. Distribution of $\hat{\sigma}_c$ across counties, 2000-2017. Each data point is a county.

4 Empirical results

4.1 The time-series of housing market liquidity

In this section, we show that, over time, price dispersion and time-on-market robustly co-move positively with each other and negatively with prices and volume. While the joint dynamics of time on market, prices and volume is well-studied in the literature, we believe that our results for price dispersion are a novel contribution.

Figure 3 shows the seasonal behavior of prices, total sales, time-on-market, and logSD, aggregated to the level of calendar months over the period 2010-2016. All four variables are seasonal: during summer, sales and prices are higher, and time-on-market and price dispersion are lower. [6]

On average, comparing June values to January values, summer prices are on average 4.03% higher, sales are 66.7% higher, time-on-market is 0.664 months (19.6%) higher, and price dispersion is 1.15% of house prices lower (6.47% in relative percentage points).

Figure 4 analyzes this further by dividing counties into 3 quantile buckets, based on

[6] Ngai and Tenreyro (2014) analyze seasonal trends in sales, prices, and time-on-market.

how seasonal prices are. This plot shows that more seasonal counties are more seasonal in all variables: that is, when seasonal price variation in larger, seasonal variation in sales, time-on-market, and price dispersion also tends to be larger.

Figure 5 shows the behavior of all four variables at the yearly level, across counties. Once again, all four variables co-move robustly. In the 2000-2005 boom, prices and sales increased, and time-on-market and price dispersion decreased. During the crash, prices and sales decreased, and time-on-market and price dispersion increased. During the recovery, we observe the reverse. As of 2016, on average across counties, sales, time-on-market and price dispersion are now roughly back to their level in 2000, though prices have increased somewhat.

Quantitatively, average time-on-market falls from 2.66 months in 2000 to 2.46 months in 2004, rises to 3.5 months in 2011, and falls to 2.62 months in 2016. Price dispersion falls from 16.6% in 2000 to 15.7% in 2004, rises to 18.2% in 2011, and falls to 16.8% in 2016.

Figure 6 divides counties into three quantile buckets, based on the size of the housing cycle, measured as the ratio between average prices in 2000 and 2005. Similar to the right panel of Figure 3, we see that counties which had bigger price booms also had larger decreases in time-on-market and price dispersion during the boom, and larger increases during the bust.

Together, we have shown that, seasonally and over the business cycle, all four liquidity measures co-move positively. Through the lens of our model, this suggests that seasonal and business cycle movements in housing market liquidity are largely driven by shifts in liquidity supply, rather than demand.

4.2 Cross-sectional relationships

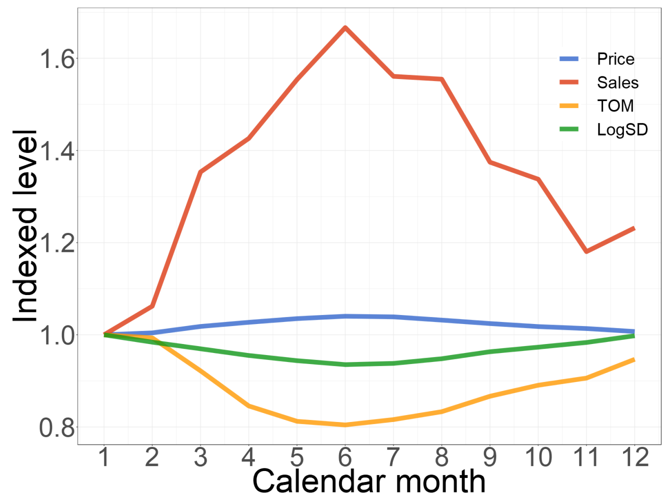
Liquidity supply. In Table 2, we test Prediction 1 in the cross-section, by estimating the following two specifications:

$$\log SD_c = Z_c^S \alpha_1 + X_c \beta_1 + \epsilon_c \tag{15}$$

$$TOM_c = Z_c^S \alpha_2 + X_c \beta_2 + \epsilon_c \tag{16}$$

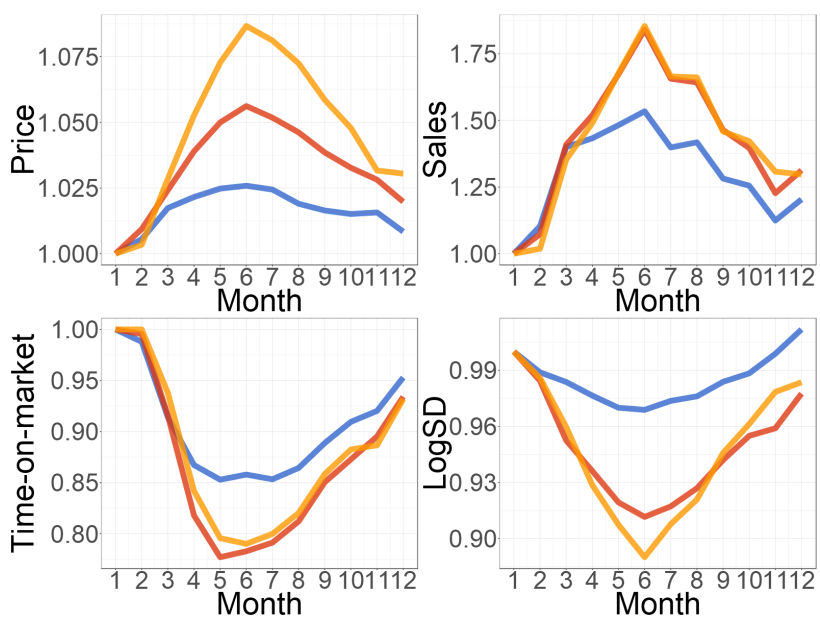
Where, Z_c^S is a county-level liquidity supply shifter, and X_c is a vector of controls. We control for third-order polynomials in a number of control variables: the average price, average age, average square footage, average bedroom and bathroom counts of sold

Figure 3: Seasonal variation in sales, prices, logSD, and time-on-market



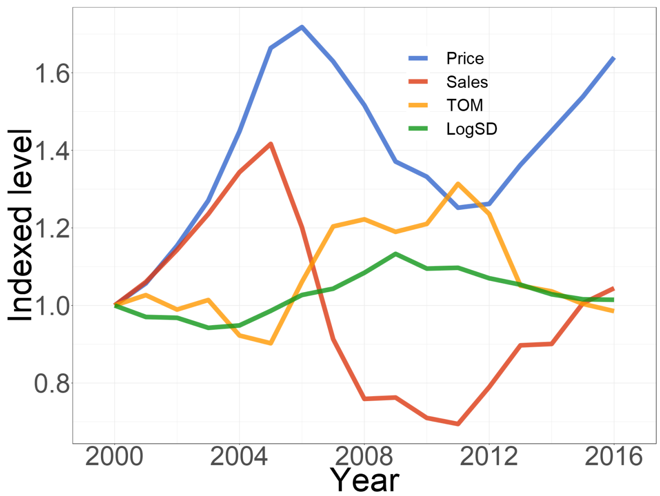
Notes. Total sales, average prices, logSD, and time-on-market (TOM) by calendar month, over the time period 2010-2016. All variables are indexed by dividing by their January level.

Figure 4: Seasonal variation in sales, prices, logSD, and time-on-market, heterogeneity



Notes. Indexed prices, sales, time-on-market, and logSD for three quantile buckets of counties, divided based on the ratio between summer and winter prices. Data is from 2010-2016.

Figure 5: Business cycle variation in sales, prices, logSD, and time-on-market



Notes. Total sales, average prices, and logSD, 2000-2016, and time-on-market (TOM), 2010-2016. All variables are indexed so that they are equal to 1 in 2000.

houses, county's population density, and the fractions of the county's population which are aged 18-35, 35-64, black, high school and college graduates, married, unemployed, and homeowners.

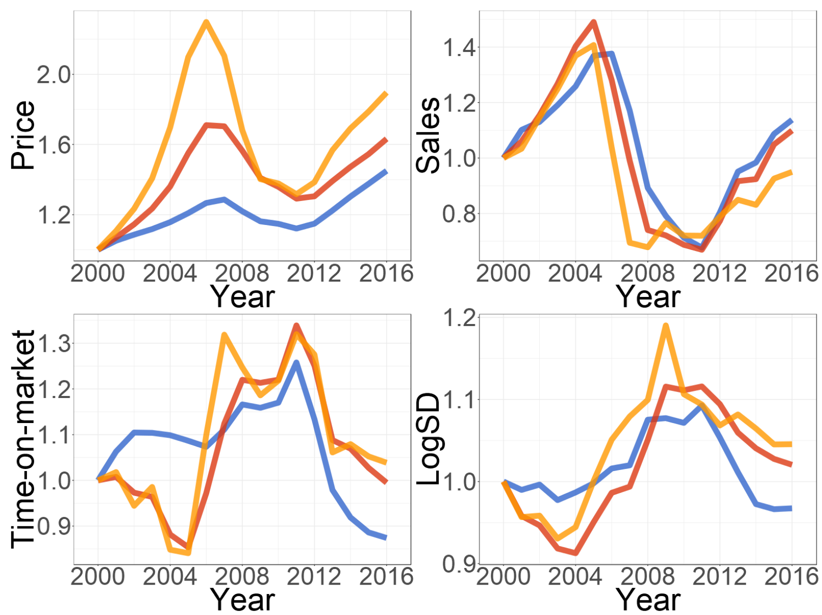
We use two liquidity supply shifters for Z_c^S. First, we use population growth rates. High population growth should indicate high η_B, so both α_1 and α_2 should be negative. Second, we use vacancy rates. High vacancy rates should indicate low η_B, so both α_1 and α_2 should be positive.

The results from this are shown in Table 2. Columns 1 and 6 show results from regressing $\log SD$ and TOM, respectively, on vacancy rates. Both coefficients are positive, supporting Prediction 1. Columns 2 and 7 show results from regressing both liquidity measures on population growth rates. Both coefficients are negative, again supporting Prediction 1. Columns 4 and 8 include both vacancy rates and population growth rates together, and columns 5 and 9 add state fixed effects: the signs and significance of all coefficients are unchanged. Hence, both liquidity supply shifters appear to be robustly correlated with price dispersion and time-on-market, supporting Prediction 1.

In addition, column 3 estimates the following specification:

$$\log SD_c = TOM_c \alpha + X_c \beta + \epsilon_c$$

Figure 6: Business cycle variation in sales, prices, logSD, and time-on-market, heterogeneity



Notes. Indexed prices, sales, time-on-market, and logSD for three quantile buckets of counties, divided into three buckets based on the ratio between average prices in 2000 and 2005. All variables are indexes so they are equal to 1 in 2000.

The estimated α is positive. Thus, $\log SD$ and TOM are positively correlated in the cross-section of counties, controlling for other features of counties. Informally, this suggests that the cross-sectional variation in liquidity measures is driven somewhat more by liquidity supply than liquidity demand.

Liquidity demand. Prediction 2 of the model states that variables which are correlated with liquidity demand – sellers' costs $\bar{c}$ – should drive time-on-market and price dispersion in opposite directions. To test this, we run the following regressions.

$$\log SD_c = Z_c^D \alpha_1 + X_c \beta_1 + \epsilon_c \tag{17}$$

$$TOM_c = Z_c^D \alpha_2 + X_c \beta_2 + \epsilon_c \tag{18}$$

We control for the same variables as in (15) and (16). Our main liquidity demand shifter is income. If homeowners' income is higher, their value of time should be higher. Hence, they should be willing to sell faster, even if that means they have to sell for lower prices. Thus, Prediction 2 states that α_1 should be positive, and α_2 should be negative: higher income should be associated with lower time-on-market, but higher price dispersion.

Columns 1 and 2 of Table 2 show that this holds in the data. Controlling for prices, higher average income is associated with lower time-on-market, but actually higher price dispersion. The magnitudes of these correlations are nontrivial: a 10% increase in average income is associated with approximately a 0.5% increase in price dispersion, and a 0.15 month (roughly 4.5 day) decrease in time-on-market.

Note also that, holding fixed income, higher prices are associated with higher time-on-market, and lower price dispersion. Prediction 2 does not produce this result directly, but this is intuitive. What matters is how high incomes are relative to prices. When prices are high relative to incomes, waiting costs are relatively lower, so sellers wait longer, but prices are more stable. When house prices are low relative to incomes, sellers' holding costs are higher relative to house prices, so sellers sell faster and at less stable prices.

5 Robustness checks and alternative explanations

In this section, we conduct a number of robustness checks. In Subsection 5.1, we discuss concerns about how we measure price dispersion and time-on-market, and we show that our results are robust to a number of other ways to measure both variables. In

Table 2: County cross-sectional regressions, liquidity supply shifters

	Dependent variable:								
	LogSD	LogSD	LogSD	LogSD	LogSD	TOM	TOM	TOM	TOM
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Vac rate	16.479***			18.187***	9.903***	3.648***		3.945***	2.610***
	(2.903)			(2.878)	(2.516)	(0.580)		(0.578)	(0.478)
Pop growth		−21.784***		−28.573***	−25.966***		−3.485**	−4.958***	−3.468***
		(7.481)		(7.191)	(6.754)		(1.516)	(1.445)	(1.284)
TOM			1.071***						
			(0.254)						
Controls	X	X	X	X	X	X	X	X	X
Fixed effects					State				State
Observations	405	405	405	405	405	405	405	405	405
Adjusted R^2	0.630	0.606	0.616	0.644	0.793	0.573	0.533	0.585	0.784

Notes. Each data point is a county. Regressions are weighted by the number of sales within the county over the time period 2012-2016.

Table 3: County cross-sectional regressions, preferences

	Dependent variable:			
	LogSD	LogSD	TOM	TOM
	(1)	(2)	(3)	(4)
Mean log price	−4.051***	−2.075***	0.500***	0.771***
	(0.655)	(0.772)	(0.136)	(0.145)
Log income	5.625***	8.125***	−1.575***	−0.869**
	(1.987)	(2.081)	(0.414)	(0.391)
Controls	X	X	X	X
Fixed effects		State		State
Observations	405	405	405	405
Adjusted R^2	0.638	0.786	0.546	0.782

Notes. Each data point is a county. Regressions are weighted by the number of sales within the county over the time period 2012-2016.

Subsection 5.2, we discuss alternative explanations for our findings, and argue that they have difficulty explaining all the stylized facts we document in Section 4.

5.1 Measurement concerns

Alternative measurement strategies. In Appendix D.1, we consider three different ways to estimate price dispersion: a pure repeat-sales specification for prices, a pure hedonic specification, and a nonparametric adjustment for time-between-sales and the number of times a house is sold. Appendix D.2 discusses the relationship of our methodology with other papers in the literature on idiosyncratic house price dispersion, and shows that our measures of price dispersion are largely in line with the literature. In Appendix D.3, we consider time-on-market measures from Zillow and Realtor.com. In all cases, results are qualitatively unchanged.

Other factors which affect price dispersion. Specification (14) is designed to capture price dispersion generated by search frictions, taking out as much as possible of price variation which is generated by house characteristics. Since specification (14) includes both house fixed effects and time-varying effects of observable characteristics, (14) can absorb both observed and unobserved characteristics of houses which have time-invariant effects on prices, and observable characteristics with time-varying effects on prices.

There are two kinds of house characteristic effects on prices which (14) cannot capture. First, our data only allow us to observe characteristics at a single point in time, so specification (14) cannot capture price changes caused by time-varying house characteristics. For example, we cannot account for the effects of house renovations or improvements on prices. Second, while specification (14) absorbs time-invariant effects of unobservables into the house fixed effects, γ_i, (14) cannot account for time-varying effects of unobservable characteristics. For example, if some houses have better construction quality than others, and the effect of construction quality on prices changes over time, this would be attributed to the error term in (14).

We believe that both effects are likely to be quantitatively small. First, Giacoletti (2017) observes data on remodeling expenditures for houses in California. Accounting for these expenditures has a relatively small effect on estimated price dispersion: comparing Figure E.vi to Figure 2 in Giacoletti (2017), accounting for remodelling decreases the estimated standard deviation of returns by only around 2% of house prices.

Second, in Appendix D.1, we show that the $f_c(x_i, t)$ term only slightly decreases our estimated residuals, implying that time variation in the market value of observable house characteristics plays a relatively small role in our data. The features we include in x_i are the main variables used in most hedonic regressions, so time variation in the market value of unobservables is likely to play a similarly small role. Thus, we believe that both issues are unlikely to have quantitatively large effects on our estimates of standard errors.

A broader issue is that, even if one believes that our error estimation using (14) in fact captures price dispersion generated by search frictions, our model is not the only model possible of search frictions. Other kinds of frictions also exist: for example, heterogeneity in realtor quality, or asymmetric information, and other forces could also influence price dispersion. We discuss various other theories of market frictions in Subsection 5.2, and argue that these alternative theories have difficulty explaining the stylized facts that we document.

5.2 Alternative explanations

5.2.1 Liquidity demand

Asymmetric information and adverse selection. The Nash bargaining model that we use implicitly assumes that agents have perfect information. In practice, the seller could be more informed about the quality of her house than the buyer. A number of papers argue this (Kurlat and Stroebel (2015), Stroebel (2016)). This could affect time-on-market and price dispersion.

Qualitatively, however, asymmetric information and adverse selection should create both delays in trade, and increased price dispersion. Thus, if liquidity differences across high- and low-income counties is largely driven by differences in the degree of adverse selection, time-on-market and price dispersion should co-move positively, which is not consistent with our stylized facts.

House heterogeneity. Another explanation for the positive association between income and price dispersion is that high-income counties could have more a more heterogeneous housing stock. Local housing markets could be thinner, since each house is slightly different, causing price dispersion to be higher. Within our model, there are two ways to rationalize this relationship: first, the effective buyer mass, M_B, could be lower when houses are more heterogeneous, since any individual house has less interested buyers. Second, the variance of buyer values, σ_ϵ, could be larger, if buyers have more specific preferences over different house characteristics.

However, our comparative statics show that, within our model, both of these parameter changes would increase time-on-market as well as price dispersion. Intuitively, decreasing M_B lowers the number of buyers, so sellers have to wait longer. Increasing the variance of buyer values increases sellers' returns from waiting longer, causing sellers to optimally wait longer. Thus, neither force can explain the fact that higher income, controlling for prices, correlates with lower time-on-market.

Higher-income households could live in more heterogeneous houses. This could explain the positive correlation between income and price dispersion. However, if houses are more heterogeneous, we would also expect time-on-market to be higher: in contrast, we find that high-income areas, controlling for prices, have lower time-on-market. Thus, house heterogeneity alone cannot explain the cross-Sectional relationship that we observe between income, time-on-market, and price dispersion.

Intuitively, the reason our stylized facts are hard to explain with other factors is that most forces tend to move time-on-market and price dispersion in the same direction. There is a robust negative relationship, once we analyze variation in prices and income. In our model, we explain this as follows: variation in preferences creates a negative relationship between price dispersion and time-on-market, as sellers rationally trade off "dollar liquidity" for more "time liquidity" as their value of time increases. We are not aware of any other theory which can produce this observed negative relationship.

5.2.2 Liquidity supply

Sophistication and cycles. A number of papers argue that, during housing booms, the fraction of unsophisticated market participants – agents who are irrational, uninformed, or inexperienced – tends to increase during housing booms, and that these agents tend to achieve worse outcomes than more sophisticated participants. For example, a number of papers argue that housing booms involve increasing dispersion in market participants' beliefs (Glaeser and Nathanson (2017), Nathanson and Zwick (2018)), and an increase in the market share of short-term speculators (Bayer et al. (2011), DeFusco, Nathanson and Zwick (2017)). Another strand of literature shows that more informed market participants, such as realtors and local buyers, are better able to time their trades, and thus achieve higher returns on average (Kurlat and Stroebel (2015), Chinco and Mayer (2015)). Finally, Gilbukh and Goldsmith-Pinkham (2019) shows that more experienced realtors sell houses faster, and that the share of inexperienced realtors tends to increase during housing booms.

Based on this literature, one might expect one or both measures of housing market liquidity to worsen during booms, as markets are increasingly dominated by unsophisticated agents. Our results show that exactly the opposite occurs: both time-on-market and price dispersion decreased during the boom, and the decrease is larger in counties with larger volume and price movements. Our explanation for this fact is that, in boom periods, the increase in market tightness means that the supply of liquidity is greater, allowing agents to do better on both dimensions. The increased in unsophisticated agents' activity may still play a role, but it seems to be overwhelmed by the increase in liquidity supply in the data.

We note that all of our statements about liquidity deal with price *dispersion*, not about the *level* of prices. Average prices can depart from fundamentals while maintaining high levels of liquidity – in particular, for bubble goods such as money, liquidity can be very high while prices are very disconnected from fundamentals. Our contribution is to show that, while booms may well involve irrationality and departure from fundamentals, booms also improve liquidity and stabilize relative prices.

Other liquidity models. Ours is not the only model which predicts that thicker markets should have lower price dispersion; this prediction follows from much of the literature on search frictions. In particular, the idea that price dispersion and time-on-market are lower in booms, when volumes and prices are higher, seems not to rely strongly on assumptions about perfect rationality. While we are not aware of any papers which have made this point, boundedly rational things, such as using "comparable sales" or benchmarking to past returns, could also produce faster sales and lower price dispersion in tighter markets. For example, if someone just sells based on returns of her neighboring house, there would be more comparables in thicker markets. Thus, our results should not be interpreted as proving that market participants are rational on average: only that price dispersion and time-on-market are lower in hot markets.

6 Calibration

In this section, we calibrate our model to data to estimate the menu of prices and time-on-markets that sellers with different holding costs face. Time is measured in years. We calibrate the model to average US-level data in 2016.

Externally calibrated parameters. We set the yearly discount rate $r = 0.052$, so that the annual discount factor is 0.95. We assume symmetric bargaining power, so $\theta = 0.5$; this is also used by, for example, Anenberg and Bayer (2015) and Arefeva (2019). We choose the matching function elasticity ϕ to 0.84, based on Genesove and Han (2012), who estimate this elasticity using the National Association of Realtors survey data, which captures both buyer and seller time-on-market. This estimate is also used by Anenberg and Bayer (2015). Since we do not observe buyer time-on-market, we cannot separately identify the matching efficiency parameter α and the inflow of buyers η_B, so we normalize the match efficiency parameter to $\alpha = 1$.

Moment matching. The remaining parameters in our model are the rate at which matched homeowners become sellers λ_M, the buyer entry rate η_B, the parameters of seller holding cost distribution, $\bar{c}, \Delta_c$, and the parameters of the buyer match value distribution $\epsilon_0, \sigma_\epsilon$. First, we choose $\lambda_M, \eta_B, \bar{c}, \epsilon_0, \sigma_\epsilon$ to match target moments from our data, and from the literature. The three moments that we calculate from our data are the sales-weighted sample averages of average house price (Zillow's ZHVI) and average time-on-market (Zillow), and the turnover rate – total house sales as a fraction of the total housing stock (Corelogic deed and tax). Moments we use from the literature are the average number of houses that buyers visit before buying, from Genesove and Han (2012), and the dispersion in buyers' values for houses, from Anundsen, Larsen and Sommervoll (2019). Table 4 shows the values of empirical moments that we target. Next, we choose Δ_c to match the empirical relationship between time on market and price dispersion. We perturb η_B around the equilibrium to generate a grid of model-implied price dispersions and time on markets. Then, we calibrate Δ_c such that the model-implied relationship between price dispersion and time on market matches the empirical coefficient. We describe our moment matching procedure in detail in Appendix E.1.

6.1 Parameter estimates

Table 4 shows the parameter values that we estimate. We estimate that match quality ϵ has a lower bound of \$557 monthly, or \$6,689 annually, and a standard deviation of \$299 monthly, or \$3,593 annually. While these seem somewhat low, buyers see many houses before buying, so the average value of ϵ among successful buyers is higher.

We find that sellers' values are uniformly distributed on $[\bar{c} + \Delta_c, \bar{c} - \Delta_c]$, with the mean

$\bar{c}$ equal to \$6,137 per month, and the range parameter Δ_c equal to \$3,262.

6.2 Results

Quantifying liquidity discounts. Using our estimated model, we can estimate "liquidity discounts": how much faster impatient sellers sell their houses relative to patient sellers, and how much lower impatient sellers' prices are as a result. To show this, figure 7 shows the "menu" of time-on-market and average prices attained by sellers with different values.

Sellers with 75th percentile value of holding cost c spend on average 3.05 months on the market, and attain expected sale prices of \$270,302. Sellers with 25th percentile holding cost spend 1.95 months, and attain expected sale prices of \$254,850. That is, 75th percentile sellers take 1.10 more months to sell, and achieve \$15,452 higher prices – in percentage terms, 5.88% higher prices. Dividing these, the implied effect of spending an extra month on the market is 5.33% higher prices.

Comparison to literature. Are these results consistent with previous studies? To check this, we survey the literature on the price vs time-on-market tradeoff. We survey two sets of papers: a set of papers which discuss various factors that affect the tradeoff between price and time-on-market, which we call "Non-foreclosure", and then a literature on foreclosure discounts. We present the results in Table 5. The foreclosure discounts can be interpreted simply as the percentage price difference between a foreclosed house, relative to a comparable non-foreclosed house.

To compare the results across papers, we attempt to convert all the papers' estimates of liquidity discounts into a percentage gain in prices of spending an extra month on the market. Essentially, we divide price effects by time-on-market effects. Details on how we arrive at these numbers are described in Appendix E.3.

We observe that the range of estimates of 1-month price effects is somewhat large, but most estimates fall within the range 1.9% to around 11%. Hence, our one-month effect, of 5.33%, is well within the range of these estimates.

We note that we do not use the liquidity discount as a target moment, in our estimation. Liquidity discounts essentially depend on the distribution of sellers' holding costs, which we estimate using aggregate-level estimates of buyer value heterogeneity, prices, volume, time-on-market, as well as the estimated correlation between price dispersion and time-on-market.

Table 4: Moment and parameter values

Moment	Value	Parameter	Value
Price	266.72	λ_M	0.052
LogSD	16.51%	ϵ_0	0.5574
TOM (Months)	2.554	σ_ϵ	0.2994
Turnover rate	0.051	η_B	0.05124
Num. visits	9.960	$\bar{c}$	6.137
PD-TOM Corr	0.596	Δ_c	3.262

Notes. Target moments and estimated parameter values. $\epsilon_0, \sigma_\epsilon, \bar{c}, \Delta_c$ are reported in thousands of US dollars per month. Turnover rate and λ_M are yearly turnover and separation rates respectively, and η_B is a fraction of the unit mass of houses per year.

I-buyer premia. How much would an i-buyer be able to gain, in our model? We can measure how much an agent with low holding costs would attain.

Suppose an i-buyer has annual holding costs of $20,000, or monthly holding costs of $1,666. For reference, the lowest-cost seller in our calibration has cost of $34,000 per year. We find that this i-buyer would hold the house for an average of 5.94 months, and sell for an average price of $293,226, or 16.4% higher than average prices.

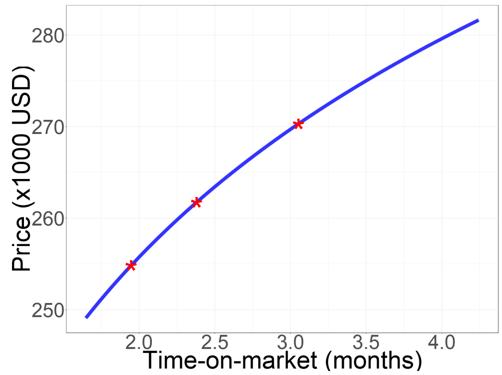
Similarly, if an I-buyer had annual holding costs of $10,000, the I-buyer would take approximately 5.94 months to sell, and would sell for an average of $293,226, or 16.4% higher than average prices. These numbers are close to, for example, the findings in Buchak et al. (2020).

This illustrates a strength of our model. It is fairly simple and tractable, and it is able to both match macro quantities for time-on-market, price dispersion, and the estimated relationship between the two; as well as produce estimates of micro quantities – liquidity discounts – which are in line with estimates from the literature.

7 Conclusion

In this paper, we have constructed a search-and-bargaining model, designed to illustrate how the tradeoffs that individual sellers face between time-on-market and prices determine aggregate liquidity measures of time-on-market and price dispersion. We showed that

Figure 7: Liquidity discounts



Notes. The left panel shows the average price and time-on-market achieved by sellers with different holding utilities c. The red stars represent, from left to right, the 25th, 50th, and 75th percentiles of seller holding utilities.

Table 5: Liquidity discounts in the literature

Paper	Type	1-month effect
Genesove and Mayer (1997)	Non-foreclosure	11.02%
Genesove and Mayer (2001)	Non-foreclosure	1.92%-24%
Levitt and Syverson (2008)	Non-foreclosure	11.68%
Hendel, Nevo and Ortalo-Magné (2009)	Non-foreclosure	4.50%
Guren (2018)	Non-foreclosure	2.46%-6.15%
Buchak et al. (2020)	Non-foreclosure	1.92%-4.46%
Pennington-Cross (2006)	Foreclosure	22%
Clauretie and Daneshvary (2009)	Foreclosure	10%
Campbell, Giglio and Pathak (2011)	Foreclosure	27%
Harding, Rosenblatt and Yao (2012)	Foreclosure	5%
Zhou et al. (2015)	Foreclosure	11%-26%

Notes. Estimates of 1-month price effects and foreclosure discounts from the literature. For the non-foreclosure lines, the estimates correspond to how much prices would increase if time-on-market increased by a month. The foreclosure discount estimates compare foreclosusure prices or returns, to prices on comparable houses which were not foreclosed on. Further details of how we calculated these are described in Appendix E.3.

these measures can be thought of as equilibrium outcomes within a supply-demand system for liquidity: supply and demand shifters drive the outcomes to co-move differentially. These predictions are supported in the data. Moreover, calibrated to the data, the model can simultaneously match the macro-relationship between aggregate time-on-market and price dispersion, and estimates of liquidity discounts faced by individual sellers at the micro-level in the housing microstructure literature.

There are a number of directions for further research. One is to analyze the effects of different market mechanisms on idiosyncratic price dispersion. We use Nash bargaining as a simple reduced-form model of price-setting. In practice, price-setting mechanisms differ somewhat in different housing markets: most houses are sold via bilateral bargaining based on a posted list price,[7] but in some markets explicit auctions are used.[8] A natural extension of our results is to analyze whether different trading mechanisms are associated with higher or lower levels of time-on-market and idiosyncratic price dispersion.

Another question is how the composition of housing market participants influences aggregate liquidity measures. A number of papers show that different classes of participants in housing markets achieve different prices and average returns. Housing markets may be more efficient, and thus idiosyncratic price dispersion may be lower if participants are more sophisticated. For example, Chinco and Mayer (2015) shows that out-of-town second-home buyers achieve lower capital gains than local buyers, Myers (2004), Ihlanfeldt and Mayock (2009), and Bayer, Ferreira and Ross (2016) study racial price gaps in the housing market, and Goldsmith-Pinkham and Shue (2019) that men attain higher returns in housing markets than women. Bayer et al. (2011) and Giacoletti and Westrupp (2017) study the effects of house flippers, and Gilbukh and Goldsmith-Pinkham (2019) study the performance of experienced versus inexperienced realtors.

[7]See Han and Strange (2016) for a discussion of the role of house list prices. Guren (2018) analyzes the relationship between house list prices and sale prices, arguing that sellers face strategic complementarities in adjusting list prices.

[8]Han and Strange (2014) analyzes time-series and cross-sectional trends in the prevalence of auctions. Arefeva (2019) constructs a dynamic search model which suggests that auctions amplify house price volatility.

References

Afonso, Gara. 2011. "Liquidity and congestion." *Journal of Financial Intermediation*, 20(3): 324–360.

Albrecht, James, Axel Anderson, Eric Smith, and Susan Vroman. 2007. "Opportunistic Matching in the Housing Market*." *International Economic Review*, 48(2): 641–664.

Albrecht, James, Pieter A. Gautier, and Susan Vroman. 2016. "Directed search in the housing market." *Review of Economic Dynamics*, 19: 218 – 231. Special Issue in Honor of Dale Mortensen.

Anenberg, Elliot, and Patrick Bayer. 2015. "Endogenous Sources of Volatility in Housing Markets: The Joint Buyer-Seller Problem." National Bureau of Economic Research Working Paper 18980.

Anundsen, Andre K., Reling Roed Larsen, and Dag Einar Sommervoll. 2019. "Price-setting and incentives in the housing market." Working Paper.

Arefeva, Alina. 2019. "How Auctions Amplify House-Price Fluctuations." Social Science Research Network SSRN Scholarly Paper ID 2980095, Rochester, NY.

Bayer, Patrick, Christopher Geissler, Kyle Mangum, and James Roberts. 2011. "Speculators and Middlemen: The Strategy and Performance of Investors in the Housing Market." National Bureau of Economic Research w16784, Cambridge, MA.

Bayer, Patrick, Fernando Ferreira, and Stephen L Ross. 2016. "What Drives Racial and Ethnic Differences in High Cost Mortgages? The Role of High Risk Lenders." National Bureau of Economic Research Working Paper 22004.

Ben-Shahar, Danny, and Roni Golan. 2019. "Improved information shock and price dispersion: A natural experiment in the housing market." *Journal of Urban Economics*, 112: 70–84.

Buchak, Greg, Gregor Matvos, Tomasz Piskorski, and Amit Seru. 2020. "iBuyers: Liquidity in Real Estate Markets?" *Available at SSRN 3616555*.

Campbell, John Y, Stefano Giglio, and Parag Pathak. 2011. "Forced sales and house prices." *American Economic Review*, 101(5): 2108–31.

Carrillo, Paul E, William M Doerner, and William D Larson. 2019. "House Price Markups and Mortgage Defaults." 49.

Case, Karl, and Robert Shiller. 1988. "The Efficiency of the Market for Single-Family Homes." National Bureau of Economic Research w2506, Cambridge, MA.

Chinco, Alex, and Christopher Mayer. 2015. "Misinformed speculators and mispricing in the housing market." *The Review of Financial Studies*, 29(2): 486–522.

Clauretie, Terrence M, and Nasser Daneshvary. 2009. "Estimating the house foreclosure discount corrected for spatial price interdependence and endogeneity of marketing time." *Real Estate Economics*, 37(1): 43–67.

Clayton, Jim, Norman Miller, and Liang Peng. 2010. "Price-volume Correlation in the Housing Market: Causality and Co-movements." *The Journal of Real Estate Finance and Economics*, 40(1): 14–40.

DeFusco, Anthony A, Charles G Nathanson, and Eric Zwick. 2017. "Speculative Dynamics of Prices and Volume." National Bureau of Economic Research Working Paper 23449.

Diaz, Antonia, and Belen Jerez. 2013. "House Prices, Sales, and Time on the Market: A Search-Theoretic Framework." *International Economic Review*, 54(3): 837–872.

Drenik, Andres, Juan Herreno, and Pablo Ottonello. 2019. "Measuring Information Frictions: Evidence from Capital Markets." 54.

Duffie, Darrell, Lei Qiao, and Yeneng Sun. 2017. "Continuous time random matching."

Duffie, Darrell, Nicolae Gârleanu, and Lasse Heje Pedersen. 2005. "Over-the-counter markets." *Econometrica*, 73(6): 1815–1847.

Duffie, Darrell, Nicolae Gârleanu, and Lasse Heje Pedersen. 2007. "Valuation in over-the-counter markets." *The Review of Financial Studies*, 20(6): 1865–1900.

Gabrovski, Miroslav, and Victor Ortego-Marti. 2019*a*. "The cyclical behavior of the Beveridge Curve in the housing market." *Journal of Economic Theory*, 181: 361–381.

Gabrovski, Miroslav, and Victor Ortego-Marti. 2019*b*. "Search and Credit Frictions in the Housing Market." 27.

Gavazza, Alessandro. 2016. “An empirical equilibrium model of a decentralized asset market.” *Econometrica*, 84(5): 1755–1798.

Genesove, David, and Christopher J Mayer. 1997. “Equity and Time to Sale in the Real Estate Market.” *The American Economic Review*, 255–269.

Genesove, David, and Christopher Mayer. 2001. “Loss Aversion and Seller Behavior: Evidence from the Housing Market.” *The Quarterly Journal of Economics*, 116(4): 1233–1260.

Genesove, David, and Lu Han. 2012. “Search and matching in the housing market.” *Journal of Urban Economics*, 72(1): 31–45.

Giacoletti, Marco. 2017. “Idiosyncratic Risk in Housing Markets.” Social Science Research Network SSRN Scholarly Paper ID 2995323, Rochester, NY.

Giacoletti, Marco, and Victor Westrupp. 2017. “The Risk-Adjusted Performance of Asset Flippers.” Social Science Research Network SSRN Scholarly Paper ID 2892439, Rochester, NY.

Gilbukh, Sonia, and Paul S Goldsmith-Pinkham. 2019. “Heterogeneous real estate agents and the housing cycle.” *Available at SSRN 3436797*.

Glaeser, Edward L, and Charles G Nathanson. 2017. “An extrapolative model of house price dynamics.” *Journal of Financial Economics*, 126(1): 147–170.

Goldsmith-Pinkham, Paul, and Kelly Shue. 2019. “The Gender Gap in Housing Returns.”

Guren, Adam M. 2018. “House price momentum and strategic complementarity.” *Journal of Political Economy*, 126(3): 1172–1218.

Guren, Adam M, and Timothy J McQuade. 2019. “How Do Foreclosures Exacerbate Housing Downturns?” National Bureau of Economic Research Working Paper 26216.

Han, Lu, and William C Strange. 2014. “Bidding wars for houses.” *Real Estate Economics*, 42(1): 1–32.

Han, Lu, and William C Strange. 2016. “What is the role of the asking price for a house?” *Journal of Urban Economics*, 93: 115–130.

Harding, John P, Eric Rosenblatt, and Vincent W Yao. 2012. "The foreclosure discount: Myth or reality?" *Journal of Urban Economics*, 71(2): 204–218.

Head, Allen, Huw Lloyd-Ellis, and Hongfei Sun. 2014. "Search, Liquidity, and the Dynamics of House Prices and Construction." *American Economic Review*, 104(4): 1172–1210.

Hendel, Igal, Aviv Nevo, and François Ortalo-Magné. 2009. "The relative performance of real estate marketing platforms: MLS versus FSBOMadison. com." *American Economic Review*, 99(5): 1878–98.

Hugonnier, Julien, Benjamin Lester, and Pierre-Olivier Weill. 2018. "Frictional intermediation in over-the-counter markets." National Bureau of Economic Research.

Ihlanfeldt, Keith, and Tom Mayock. 2009. "Price discrimination in the housing market." *Journal of Urban Economics*, 66(2): 125 – 140.

Krainer, John. 2001. "A Theory of Liquidity in Residential Real Estate Markets." *Journal of Urban Economics*, 49(1): 32 – 53.

Kurlat, Pablo, and Johannes Stroebel. 2015. "Testing for information asymmetries in real estate markets." *The Review of Financial Studies*, 28(8): 2429–2461.

Landvoigt, Tim, Monika Piazzesi, and Martin Schneider. 2015. "The Housing Market(s) of San Diego." *American Economic Review*, 105(4): 1371–1407.

Leung, Charles, Garion Lau, and Youngman Leong. 2002. "Testing Alternative Theories of the Property Price-Trading Volume Correlation." *Journal of Real Estate Research*, 23(3): 253–264.

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

Mortensen, Dale T, and Christopher A Pissarides. 1994. "Job creation and job destruction in the theory of unemployment." *The review of economic studies*, 61(3): 397–415.

Myers, Caitlin Knowles. 2004. "Discrimination and neighborhood effects: understanding racial differentials in US housing prices." *Journal of Urban Economics*, 56(2): 279 – 302.

Nathanson, Charles G, and Eric Zwick. 2018. "Arrested Development: Theory and Evidence of Supply-Side Speculation in the Housing Market." *The Journal of Finance*, 73(6): 2587–2633.

Ngai, L. Rachel, and Silvana Tenreyro. 2014. "Hot and Cold Seasons in the Housing Market." *American Economic Review*, 104(12): 3991–4026.

Ouazad, Amine, and Romain Rancière. 2019. "Market Frictions, Arbitrage, and the Capitalization of Amenities." National Bureau of Economic Research Working Paper 25701.

Peng, Liang, and Thomas G. Thibodeau. 2017. "Idiosyncratic Risk of House Prices: Evidence from 26 Million Home Sales: Idiosyncratic Risk of House Prices." *Real Estate Economics*, 45(2): 340–375.

Pennington-Cross, Anthony. 2006. "The value of foreclosed property." *Journal of Real Estate Research*, 28(2): 193–214.

Piazzesi, Monika, and Martin Schneider. 2009. "Momentum traders in the housing market: survey evidence and a search model." *American Economic Review*, 99(2): 406–11.

Piazzesi, Monika, Martin Schneider, and Johannes Stroebel. 2015. "Segmented Housing Search." National Bureau of Economic Research Working Paper 20823.

Rogerson, Richard, Robert Shimer, and Randall Wright. 2005. "Search-Theoretic Models of the Labor Market: A Survey." *Journal of Economic Literature*, 43(4): 959–988.

Sagi, Jacob S. 2015. "Asset-Level Risk and Return in Real Estate Investments." *SSRN Electronic Journal*.

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

Stroebel, Johannes. 2016. "Asymmetric information about collateral values." *The Journal of Finance*, 71(3): 1071–1112.

Trejos, Alberto, and Randall Wright. 2016. "Search-based models of money and finance: An integrated approach." *Journal of Economic Theory*, 164: 10–31.

Vayanos, Dimitri, and Pierre-Olivier Weill. 2008. "A search-based theory of the on-the-run phenomenon." *The Journal of Finance*, 63(3): 1361–1398.

Vayanos, Dimitri, and Tan Wang. 2007. "Search and endogenous concentration of liquidity in asset markets." *Journal of Economic Theory*, 136(1): 66–104.

Weill, Pierre-Olivier. 2008. "Liquidity premia in dynamic bargaining markets." *Journal of Economic Theory*, 140(1): 66–96.

Wheaton, William C. 1990. "Vacancy, Search, and Prices in a Housing Market Matching Model." *Journal of Political Economy*, 98(6): 1270–1292.

Zhou, Hanqing, Yuan Yuan, Christopher Lako, Michael Sklarz, and Charles McKinney. 2015. "Foreclosure discount: definition and dynamic patterns." *Real Estate Economics*, 43(3): 683–718.

Appendix

A Supplementary material for section 2

A.1 Stationary equilibrium conditions

A.1.1 Bellman equations

Given the buyer match rate λ_B, trade cutoffs $\epsilon^*(c)$, the equilibrium distribution of seller values $F_{eq}(c)$, and the seller value function $V_S(c)$, the equilibrium value of buyers, V_B, must satisfy:

$$rV_B = \lambda_B \int \int_{\epsilon > \epsilon^*(c)} \left[(1-\theta)\left(V_M(\epsilon) - V_B - V_S(c)\right)\right] dG(\epsilon)\, dF_{eq}(c) \tag{19}$$

In words, expression (19) can be interpreted as follows. At rate λ_B, the buyer is matched to a seller with type randomly drawn from $F_{eq}(\cdot)$, and the buyer draws match quality ϵ from $G(\cdot)$. If the buyer's match quality draw, ϵ, is higher than the seller's match quality cutoff, $\epsilon^*(c)$, trade occurs, and the buyer receives a share $(1-\theta)$ of the bilateral match surplus.

Similarly, given the seller match rate λ_S, trade cutoffs $\epsilon^*(c)$, and the buyer value V_B, the seller value function $V_S(c)$ satisfies:

$$rV_S(c) = \nu + \lambda_S \int_{\epsilon > \epsilon^*(c)} \theta\left(V_M(\epsilon) - V_B - V_S(c)\right) dG(\epsilon) \tag{20}$$

In words, expression (20) states that a seller of type c receives flow value $-c$ from their house while they are waiting for buyers. At rate λ_S, the seller meets a buyer with match value ϵ randomly drawn from $G(\cdot)$. If $\epsilon > \epsilon^*(c)$, trade occurs, and the seller receives a share θ of the bilateral match surplus.

The expected value V_M of matched owners is determined by the Bellman equation:

$$rV_M(\epsilon) = \epsilon + \lambda_M \left(\int V_S(c)\, dF(c) - V_M(\epsilon) \right) \tag{21}$$

In words, expression (21) states that matched owners get flow value ϵ while matched to their house and receive separation shocks at rate λ_M, at which point they become sellers and attain the expectation of the seller value function $V_S(c)$ over the seller holding cost distribution $F(c)$.

Finally, the marginal entering buyer must have 0 expected value. Hence, buyers enter if the sum of their utility from entering the city, ξ, and their expected value from being an active homebuyer, V_B, is positive:

$$\xi^* + V_B \geqslant 0 \tag{22}$$

A.1.2 Flow equality conditions

First, consider flow equality for sellers. In equilibrium, the rate at which matched homeowners receive separation shocks and become sellers of type c is:

$$(1 - M_S)\, \lambda_M f(c) \tag{23}$$

In words, this is the product of the total mass of matched homeowners, $1 - M_S$; the rate at which shocks homeowners receive separation shocks, λ_M; and the density $f(c)$ of entering sellers with value c.[9] The equilibrium rate at which sellers of type c sell their houses and leave the market is:

$$M_S f_{eq}(c)\, \lambda_S\, (1 - G(\epsilon^*(c))) \tag{24}$$

In words, this is the product of the mass of sellers, M_S; the density of values among sellers in equilibrium, $f_{eq}(c)$; the rate at which sellers are matched to buyers in equilibrium, λ_S; and the probability that the match utility draw ϵ exceeds the trade cutoff $\epsilon^*(c)$ for a seller of type c, which is $1 - G(\epsilon^*(c))$. In stationary equilibrium, expressions (23) and (24) must be equal.

Flow equality for individual seller types implies that the total rate at which matched homeowners become sellers is equal to the total rate at which sellers sell and exit; that is, integrating (23) and (24) over c, we have:

$$(1 - M_S)\, \lambda_M = \int_{\nu} \lambda_S M_S\, (1 - G(\epsilon^*(c)))\, f_{eq}(c)\, d\nu \tag{25}$$

Moreover, since each successful sale turns a buyer into a matched homeowner, the RHS of (25) is also equal to the rate at which buyers turn into matched homeowners.

[9] Since the distribution $F(c)$ of holding costs does not depend on matched homeowners' match utility ϵ, we do not need to explicitly integrate over the distribution $G_{eq}(\epsilon)$ in expression (23).

Second, inflows and outflows for matched homeowners with match utility ϵ must be equal. Matched homeowners' separation rate λ_M does not depend on their match utility ϵ, so the distribution of match utilities among matched homeowners is equal to the distribution of match utilities among successful home buyers, which is:

$$G_{eq}(\epsilon) = \frac{\int_c \lambda_S M_S \left[\int_{\tilde{\epsilon}=\epsilon_0}^{\epsilon} 1\left(\tilde{\epsilon} > \epsilon^*(c)\right) dG(\tilde{\epsilon})\right] dF_{eq}(c)}{\int_c \lambda_S M_S \left(1 - G\left(\epsilon^*(c)\right)\right) dF_{eq}(c)}$$

In words, the numerator is the flow rate at which a seller of value c successfully trades with a buyer with match utility below ϵ, integrated over the equilibrium distribution $f_{eq}(c)$ of holding costs c among sellers. The denominator is the RHS of (25), the total flow rate at which buyers become matched homeowners.

Finally, the rate at which buyers enter the market must be equal to all other flow rates. If all buyers with value above some cutoff ξ^* enter, the inflow rate of buyers is equal to:

$$\eta_B\left(1 - F_\xi\left(\xi^*\right)\right)$$

Hence, we must have:

$$(1 - M_S)\lambda_M = \eta_B\left(1 - F_\xi\left(\xi^*\right)\right)$$

Substituting for the cutoff ξ^* using (22), this simplifies to:

$$(1 - M_S)\lambda_M = \eta_B\left(1 - F_\xi\left(-V_B\right)\right) \tag{26}$$

A.2 Proof of Claim 1

A.2.1 Price dispersion

From (2), prices are:

$$P(\epsilon, c) = \theta\left(V_M(\epsilon) - V_B - V_S(c)\right) + V_S(c) \tag{27}$$

We wish to take the variance of expression (27) with respect to the joint distribution of holding costs c and match utilities ϵ within the set of pairs of buyers and sellers that match and trade in any given moment; call this joint distribution $F_{tr}(c, \epsilon)$.

First, let $F_{tr}(c)$ be the marginal distribution of seller holding costs c, among the stationary mass of seller types that trade in any given time period. By flow equality in expression (8) of proposition 1, the marginal distribution of c among sellers who trade and exit the market at any moment must be the same as the distribution of c among sellers that enter the platform; thus, we simply have:

$$F_{tr}(c) = F(c) \tag{28}$$

Thus, to characterize $F_{tr}(c, \epsilon)$, we need only characterize

$$F_{tr}(\epsilon \mid c)$$

for all c; that is, the distributions of buyer match utilities, conditional on trade occurring and conditional on a given seller holding cost c. Each time a seller of holding cost c meets a buyer, a random match quality $\epsilon \sim G(\cdot)$ is drawn; trade occurs if $\epsilon > \epsilon^*(c)$. Thus,

$$F_{tr}(\epsilon \mid c) = G(\epsilon \mid \epsilon > \epsilon^*(c)) \tag{29}$$

that is, the conditional distribution of match qualities ϵ, conditional on a seller having holding cost c and trade occurring, is simply the distribution of ϵ conditional on it being above the trade cutoff $\epsilon^*(c)$.

Having characterized $F_{tr}(c, \epsilon)$, we can now take the variance of expression (27) for prices. Applying the law of iterated expectations, price variance can be written as:

$$Var(P(\epsilon, c)) = E_{c \sim F_{tr}(c)}\left[Var_{\epsilon \sim F_{tr}(\epsilon|c)}(P(\epsilon, c) \mid c)\right] + Var_{c \sim F_{tr}(c)}\left(E_{\epsilon \sim F_{tr}(\epsilon|c)}[P(\epsilon, c) \mid c]\right) \tag{30}$$

Substituting (28) and (29), we can write this as:

$$Var(P(\epsilon, c)) = E_{c \sim F(c)}\left[Var_{\epsilon \sim G(\epsilon|\epsilon > \epsilon^*(c))}(P(\epsilon, c) \mid c)\right] + Var_{c \sim F(c)}\left(E_{\epsilon \sim G(\epsilon|\epsilon > \epsilon^*(c))}[P(\epsilon, c) \mid c]\right) \tag{31}$$

First, we characterize the left term on the RHS of (30). Conditional on c, the only random term in $P(\epsilon, c)$ conditional on c is the buyer's match utility ϵ; thus, substituting expression

(44) for $P(\epsilon, c)$ and ignoring constant terms, we have:

$$Var_{\epsilon \sim G(\epsilon|\epsilon>\epsilon^*(c))}\left(P(\epsilon, c) \mid c\right) = \left(\frac{\theta}{r+\lambda_M}\right)^2 Var_{\epsilon \sim G(\epsilon|\epsilon>\epsilon^*(c))}(\epsilon)$$

In words,

$$Var_{\epsilon \sim G(\epsilon|\epsilon>\epsilon^*(c))}(\epsilon)$$

is the variance of an exponential random variable ϵ, conditional on ϵ being above some cutoff $\epsilon^*(c)$, which is greater than its lower bound ϵ_0. This conditional distribution has variance equal to the unconditional variance of ϵ, σ_ϵ^2, for any cutoff $\epsilon^*(c)$; thus, we have:

$$Var_{\epsilon \sim G(\epsilon|\epsilon>\epsilon^*(c))}\left(P(\epsilon, c) \mid c\right) = \left(\frac{\theta}{r+\lambda_M}\right)^2 \sigma_\epsilon^2 \tag{32}$$

Since expression (32) is independent of c, we also have:

$$E_{c \sim F(c)}\left[Var_{\epsilon \sim G(\epsilon|\epsilon>\epsilon^*(c))}\left(P(\epsilon, c) \mid c\right)\right] = \left(\frac{\theta}{r+\lambda_M}\right)^2 \sigma_\epsilon^2 \tag{33}$$

Now we move to the right term in expression (30). Substituting expression (44) for prices, we have:

$$\begin{aligned} Var_{c \sim F(c)}\left(E_{\epsilon \sim G(\epsilon|\epsilon>\epsilon^*(c))}\left[P(\epsilon, c) \mid c\right]\right) = \\ Var_{c \sim F(c)}\left(E_{\epsilon \sim G(\epsilon|\epsilon>\epsilon^*(c))}\left[V_S(c) + \theta\left(\frac{\epsilon - \epsilon^*(c)}{r+\lambda_M}\right) \Big| \, c\right]\right) \end{aligned} \tag{34}$$

Rearranging, and moving $V_S(c)$ out of the conditional expectation, this is equal to:

$$Var_{c \sim F(c)}\left(V_S(c) + \frac{\theta}{r+\lambda_M} E_{\epsilon \sim G(\epsilon|\epsilon>\epsilon^*(c))}\left[\epsilon - \epsilon^*(c) \mid c\right]\right) \tag{35}$$

Since we have assumed $G(\cdot)$ is exponential, and $\epsilon^*(c) \geqslant \epsilon_0$, the term:

$$E_{\epsilon \sim F_{tr}(\epsilon|c)}\left[\epsilon - \epsilon^*(c) \mid c\right]$$

is equal to σ_ϵ, the standard deviation of ϵ. It is thus constant with respect to $\epsilon^*(c)$ and thus c, and can be ignored when calculating the variance in (35). Hence,

$$Var_{c\sim F(c)}\left(E_{\epsilon\sim G(\epsilon|\epsilon>\epsilon^*(c))}\left[P\left(\epsilon,c\right)\mid c\right]\right)=Var_{c\sim F(c)}\left(V_S\left(c\right)\right) \tag{36}$$

Substituting (33) and (36) into (31), we have

$$Var\left(P\left(\epsilon,c\right)\right)=Var_{c\sim F(\cdot)}\left(V_S\left(c\right)\right)+\left(\frac{\theta}{r+\lambda_M}\right)^2\sigma_\epsilon^2 \tag{37}$$

Now, taking the expectation of prices from (44) below, we have:

$$E\left[P\left(\epsilon,c\right)\right]=E\left[V_S\left(c\right)\right]+\frac{\theta\sigma_\epsilon}{r+\lambda_M} \tag{38}$$

Using (37) and (38), we get (11).

A.2.2 Time-on-market

Time-on-market for type c is the inverse of $\lambda_S\left(1-G\left(\epsilon^*\left(c\right)\right)\right)$, which is the product of the equilibrium rate at which sellers meet buyers, λ_S, and the fraction of meetings for a seller of type c that result in trade, $\left(1-G\left(\epsilon^*\left(c\right)\right)\right)$. Average time on market, (12), is the expectation of this.

A.3 Expressions for $V_M\left(\epsilon\right), \epsilon^*\left(c\right), P\left(\epsilon,c\right)$

To begin with, we analytically characterize $V_M\left(\epsilon\right)$. From expression (21), we have:

$$rV_M\left(\epsilon\right)=\epsilon+\lambda_M\left(\int V_S\left(c\right)dF\left(c\right)-V_M\left(\epsilon\right)\right)$$

Solving for $V_M\left(\epsilon\right)$, we have:

$$V_M\left(\epsilon\right)=\frac{\epsilon}{r+\lambda_M}+\frac{\lambda_M}{r+\lambda_M}\int V_S\left(c\right)dF\left(c\right) \tag{39}$$

Using expression (39), we can also characterize the trade cutoff function $\epsilon^*\left(c\right)$. Trade occurs if:

$$V_M\left(\epsilon\right)\geqslant V_B+V_S\left(c\right)$$

$$\implies \frac{\lambda_M}{r+\lambda_M}\int V_S(c)\,dF(c) + \frac{\epsilon}{r+\lambda_M} \geqslant V_B + V_S(c) \tag{40}$$

Since we have assumed that $\epsilon^*(c)$ is greater than ϵ_0, the lower bound of $G(\cdot)$, we can treat expression (40) as an equality. Solving for $\epsilon^*(c)$, we have:

$$\epsilon^*(c) = (r+\lambda_M)\left[V_B + V_S(c)\right] - \lambda_M \int V_S(c)\,dF(c) \tag{41}$$

Using (39) and (41) we can also characterize equilibrium prices. From (2), we have:

$$P(\epsilon, c) = V_S(c) + \theta\left(V_M(\epsilon) - V_B - V_S(c)\right)$$

Substituting for $V_M(\epsilon)$ using (39), we have:

$$P(\epsilon, c) = V_S(c) + \theta\left(\frac{\epsilon}{r+\lambda_M} + \frac{\lambda_M}{r+\lambda_M}\int V_S(c)\,dF(c) - V_B - V_S(c)\right) \tag{42}$$

Now, we can write (41) as:

$$\frac{\lambda_M}{r+\lambda_M}\int V_S(c)\,dF(c) - V_B - V_S(c) = -\frac{\epsilon^*(c)}{r+\lambda_M} \tag{43}$$

Hence, substituting (43) into (42), we get:

$$P(\epsilon, c) = V_S(c) + \theta\left(\frac{\epsilon - \epsilon^*(c)}{r+\lambda_M}\right) \tag{44}$$

A.4 Proof of Claim 2

From expression (4) in proposition 1, the seller value function $V_S(c)$ is:

$$rV_S(c) = -c + \lambda_S \int_{\epsilon > \epsilon^*(c)} \theta\left(V_M(\epsilon) - V_B - V_S(c)\right) dG(\epsilon)$$

Differentiating with respect to c, using the Leibniz rule, we have:

$$\begin{aligned} rV_S'(c) = -1 - \lambda_S\theta\Big(V_M(\epsilon^*(c)) - V_B - V_S(c)\Big) g(\epsilon^*(c))\,\frac{d\epsilon^*(c)}{dc} + \\ \lambda_S \int \theta\left(-V_S'(c)\right) 1\left(\epsilon > \epsilon^*(c)\right) dG(\epsilon) \end{aligned} \tag{45}$$

By definition of $\epsilon^*(c)$ in (6):

$$V_M(\epsilon^*(c)) - V_B - V_S(c) = 0$$

so the middle term is 0. Hence, (45) becomes:

$$rV_S'(c) = -1 + \lambda_S\theta\left(-V_S'(c)\right)(1 - G(\epsilon^*(c)))$$

Solving for $V_S'(c)$, we have:

$$V_S'(c) = \frac{-1}{r + \lambda_S\theta(1 - G(\epsilon^*(c)))} \tag{46}$$

Substituting expression (12) for $TOM(c)$ in the denominator of (46), we have:

$$V_S'(c) = \frac{-1}{r + \frac{\theta}{TOM(c)}}$$

Rearranging, we have (13).

A.5 Derivation of model quantities

In this Appendix, we derive expressions for average prices, time-on-market, and price dispersion, which are plotted in Figure 1. The average transaction price conditional on trade is:

$$\int\int P(\epsilon, c)\ dG(\epsilon \mid \epsilon > \epsilon^*(c))\ dF(c)$$

This is the expectation of the price function $P(\epsilon, c)$ over the joint distribution of ϵ, c among successfully trading buyers and sellers.

Average time-on-market, over the distribution of realized sales, is:

$$\int TOM_S(c)\ dF(c)$$

This is simply the average of time on market for a seller of holding cost c over the distribution of holding costs $F(c)$; note that we showed in (28) above that the distribution of holding costs among trading sellers is simply $F_{tr}(c) = F(c)$.

From Claim 1, equilibrium price variance as:

$$Var_{c\sim F(\cdot)}\left(V_S\left(c\right)\right)+\left(\frac{\theta}{r+\lambda_M}\right)^2\sigma_\epsilon^2$$

A.6 Heterogeneous buyer urgency

We can extend the main model to accomodate persistent buyer heterogeneity. Suppose that buyers have some persistent type $u\sim H\left(u\right)$, drawn at the point that buyers enter the market. Unmatched buyers receive flow utility u per unit time they are waiting to purchase their houses. Transaction prices become a function of sellers' holding cost c, buyers' urgency u, and buyers' match utility ϵ:

$$P\left(c,u,\epsilon\right)=V_S\left(c\right)+\theta\left(V_M\left(\epsilon\right)-V_B\left(u\right)-V_S\left(c\right)\right) \tag{47}$$

Thus, the match quality cutoff condition becomes:

$$V_M\left(\epsilon^*\left(c,u\right)\right)=V_B\left(u\right)+V_S\left(c\right)$$

Analogous to the main text, for our theoretical results, we will need to assume that

$$\epsilon^*\left(c,u\right)\geqslant\epsilon_0\ \forall c,u$$

Buyers' and sellers' value functions become, respectively:

$$rV_B\left(u\right)=u+\lambda_B\int_c\int_{\epsilon\geqslant\epsilon^*(c,u)}\left[\left(1-\theta\right)\left(V_M\left(\epsilon\right)-V_B\left(u\right)-V_S\left(c\right)\right)\right]dG\left(\epsilon\right)dF_{eq}\left(c\right)$$

$$rV_S\left(c\right)=-c+\lambda_S\int_u\int_{\epsilon>\epsilon^*(c,u)}\left[\theta\left(V_M\left(\epsilon\right)-V_B\left(u\right)-V_S\left(c\right)\right)\right]dG\left(\epsilon\right)dH_{eq}\left(u\right) \tag{48}$$

The flow equality conditions for sellers and matched owners must now integrate over the equilibrium distribution $H_{eq}\left(u\right)$ of buyer urgencies:

$$\left(1-M_S\right)\lambda_M f\left(c\right)=\lambda_S M_S f_{eq}\left(c\right)\int_u\left[1-G\left(\epsilon^*\left(c,u\right)\right)\right]dH_{eq}\left(u\right)$$

$$G_{eq}\left(\epsilon\right)=\frac{\int_u\int_c\lambda_S M_S\left[\int_{\tilde{\epsilon}=\epsilon_0}^{\epsilon}1\left(\tilde{\epsilon}>\epsilon^*\left(c,u\right)\right)dG\left(\tilde{\epsilon}\right)\right]dF_{eq}\left(c\right)dH_{eq}\left(u\right)}{\int_u\int_c\lambda_S M_S\left(1-G\left(\epsilon^*\left(c,u\right)\right)\right)dF_{eq}\left(c\right)dH_{eq}\left(u\right)}$$

Moreover, there is an additional flow equality constraint requiring inflows and outflows of all buyer types to be equal:

$$\eta_B h(u) = \lambda_B M_B h_{eq}(u) \int_c \left[1 - G\left(\epsilon^*(c,u)\right)\right] dF_{eq}(c)$$

Somewhat surprisingly, despite these changes to stationary equilibrium conditions, claims 1 and 2 continue to hold. To prove claim 1, note that, when buyers have heterogeneous values, the matched owner value function is unchanged:

$$V_M(\epsilon) = \frac{\epsilon}{r+\lambda_M} + \frac{\lambda_M}{r+\lambda_M} \int_c V_S(c)\, dF(c) \tag{49}$$

The derivations in Appendix A.3 thus imply that:

$$P(c,u,\epsilon) = V_S(c) + \theta\left(\frac{\epsilon - \epsilon^*(c,u)}{r+\lambda_M}\right) \tag{50}$$

Now, similar to (30), we take the variance of prices, applying the law of iterated expectations with respect to c and u, to get:

$$\begin{aligned} Var\left(P(c,u,\epsilon)\right) = E_{c,u\sim F_{tr}(c,u)}\left[Var_{\epsilon\sim G_{tr}(\epsilon|c,u)}\left(P(c,u,\epsilon) \mid c,u\right)\right] + \\ Var_{c,u\sim F_{tr}(c,u)}\left(E_{\epsilon\sim G_{tr}(\epsilon|c,u)}\left[P(c,u,\epsilon) \mid c,u\right]\right) \end{aligned} \tag{51}$$

Where, analogously to expression (30), $F_{tr}(c,u)$ is the joint distribution of c and u among trading buyers and sellers, and $F_{tr}(\epsilon \mid c,u)$ is the conditional distribution of ϵ given c,u among trading buyers and sellers. Analogously to the argument to Appendix A.2, we have:

$$F_{tr}(\epsilon \mid c,u) = G\left(\epsilon \mid \epsilon > \epsilon^*(c,u)\right)$$

The joint distribution $F_{tr}(c,u)$ is more complicated to characterize; however, by flow equality, the marginal distributions of $F_{tr}(c,u)$ must be equal to the distributions of entering buyer and seller types, $F(c)$ and $H(u)$. This implies that the following steps in Appendix A.2 go through essentially unchanged. Going through the steps, for the top term of (51), we have:

$$\mathrm{Var}_{\epsilon\sim G(\epsilon|\epsilon>\epsilon^*(c,u))}\left(P\left(c,u,\epsilon\right)\mid c,u\right)=$$
$$\left(\frac{\theta}{r+\lambda_M}\right)^2\mathrm{Var}_{\epsilon\sim G(\epsilon|\epsilon>\epsilon^*(c,u))}\left(\epsilon\right)=\left(\frac{\theta}{r+\lambda_M}\right)^2\sigma_\epsilon^2$$

For the bottom term, substituting expression (50) for prices, we have:

$$\mathrm{Var}_{c,u\sim F_{tr}(c,u)}\left(E_{\epsilon\sim G(\epsilon|\epsilon>\epsilon^*(c,u))}\left[P\left(c,u,\epsilon\right)\mid c,u\right]\right)$$
$$=\mathrm{Var}_{c,u\sim F_{tr}(c,u)}\left(E_{\epsilon\sim G(\epsilon|\epsilon>\epsilon^*(c,u))}\left[V_S\left(c\right)+\theta\left(\frac{\epsilon-\epsilon^*\left(c,u\right)}{r+\lambda_M}\right)\mid c,u\right]\right)$$
$$=\mathrm{Var}_{c,u\sim F_{tr}(c,u)}\left(V_S\left(c\right)+\frac{\theta}{r+\lambda_M}E_{\epsilon\sim G(\epsilon|\epsilon>\epsilon^*(c,u))}\left[\epsilon-\epsilon^*\left(c,u\right)\mid c,u\right]\right)\tag{52}$$

Again, the left term of (52),

$$E_{\epsilon\sim G(\epsilon|\epsilon>\epsilon^*(c,u))}\left[\epsilon-\epsilon^*\left(c,u\right)\mid c,u\right]$$

is equal to σ_ϵ, which is independent of c,u, so we can ignore it in the variance calculation; (52) thus simplifies to

$$\mathrm{Var}_{c,u\sim F_{tr}(c,u)}\left(V_S\left(c\right)\right)$$

which is independent of u, so this simplifies further to the variance with respect to the marginal distribution of c, that is,

$$\mathrm{Var}_{c\sim F(c)}\left(V_S\left(c\right)\right)$$

This proves (11). The proof of (12) is identical to the baseline model. Finally, differentiating (48), we have:

$$rV_S'\left(c\right)=1+\lambda_S\theta\int_u\int_{\epsilon>\epsilon^*(c,u)}\left(-V_S'\left(c\right)\right)1\left(\epsilon>\epsilon^*\left(c,u\right)\right)dG\left(\epsilon\right)dH_{eq}\left(u\right)\tag{53}$$

The total match rate facing a seller of type c is the inverse of time-on-market, so we have:

$$\mathrm{TOM}\left(c\right)=\frac{1}{\lambda_S\int_u\int_{\epsilon>\epsilon^*(c,u)}1\left(\epsilon>\epsilon^*\left(c,u\right)\right)dG\left(\epsilon\right)dH_{eq}\left(u\right)}\tag{54}$$

Combining (53) and (54), we have:

$$V_S'(c) = \frac{1}{r + \frac{\theta}{\mathrm{TOM}(c)}} = \frac{\mathrm{TOM}(c)}{r\mathrm{TOM}(c) + \theta}$$

proving claim 2.

B Supplementary material for section 3

B.1 CoreLogic tax, deed, and MLS data

Our data on house sales comes from the CoreLogic deed dataset, which is derived from county government records of house transactions. Corelogic records the price and date of each sale, and housing units are uniquely identified, within a FIPS county code, by an Assessor Parcel Number (APN), which is assigned to each plot of land by tax assessors. Our data on house characteristics comes from the CoreLogic tax assessment data for the fiscal year 2016-2017, which contains, for each APN, its latitude, longitude, year built, square footage, and numbers of bedrooms and bathrooms, as of 2016-2017. We merge the tax data to the CoreLogic deed data by APN and FIPS county code.

We clean the datasets using a number of steps. First, we use only arms-length new construction or resales of single-family residences, which are not foreclosures, which have non-missing sale price, date, APN, and county FIPS code in the CoreLogic deed data, and which have non-missing year built and square footage in the CoreLogic tax data. As mentioned in the main text, we use only data from 2000 onwards, as we find that CoreLogic's data quality is quite low prior to around 2000. Even after throwing out pre-2000 data, we find that some counties have very low total sales for early years, suggesting that some data is missing. To address this, we filter out all county-years for which the total number of sales we observe is below 20% of average annual sales over all county-years in our sample.

We use the dataset that results from these cleaning steps to measure monthly sales by county. This subsample is, however, unsuitable for estimating price dispersion, and we apply a few additional cleaning steps for the subsample we use to estimate price dispersion regressions in Subsection 3.2.

First, our measurement of price dispersion uses a repeat-sales specification, so we can only use houses that were sold multiple times. Moreover, we wish to filter for "house flips", as well as instances where reported sale price seems anomalous. If a house is ever sold twice within a year, we drop all observations of the house. Most of these kinds of transactions appear to be either flips, which are known to be a peculiar segment of the real estate market (Bayer et al. (2011), Giacoletti and Westrupp (2017)), or duplication bugs in the data, where a single transaction is recorded twice or more. To filter for potentially anomalous prices, similarly to Landvoigt, Piazzesi and Schneider (2015) and Giacoletti (2017), if we ever observe a property whose annualized appreciation or depreciation is above 50% for any given pair of sales, we drop all observations of the property. We also drop houses with prices that seem like significant outliers: if a house is ever sold at a price which is more than 50% times higher or lower than the median house price in the same county-year, we drop all observations of the house from our dataset.

Second, specification (14) involves a fairly large number of parameters: house and county-month fixed effects, as well as many parameters in the $f_z(x_i, t)$ polynomial term. We thus require a fairly large number of house sales in order to precisely estimate (14); thus we filter to counties with at least 1000 house sales remaining, and with at least 10 sales per month on average, after applying the filtering steps described above.

Appendix Table A1 shows characteristics of the counties in our estimation sample, compared to the universe of counties in the 2012-2016 ACS 5-year sample. Our dataset constitutes approximately 14.8% of all counties. Counties in our sample are larger and denser than average, so our sample constitutes around 61.8% of the total US population. In terms of demographics, the average income is somewhat higher than average for counties in our sample, but our sample are quite representative in terms of age, race, and the fraction of the population that is married.

We measure time on market using CoreLogic MLS dataset that contains data on individual house listings. Housing units are uniquely identified, within a FIPS county code, by an Assessor Parcel Number (APN). We only use listings of single-family residences that eventually were sold with non-missing original listing and closing dates and non-missing FIPS county code. We define time on market as the difference between closing date and original listing date. We drop listings with time on market longer than 900 days, and winsorize listings with time on market longer than 550 days. We then use listing-level time on market to compute county-year-month and county-year average time on market. We require county-year-month triplets to have at least 10 listings, and county-year pairs to have at least 50 listings. Otherwise we record county-year-month or county-year average

Table A1: Characteristics of counties in our dataset

	Sample mean	All counties mean
Population	417,315	100,027
Pop / Sq mile	984	290
Housing units	170,127	42,120
Avg income	$77,974	$61,995
% Age 18-35	22.4%	20.7%
% Married	49.9%	51.3%
% Black	11.0%	9.00%
Total counties	477	3,220
Total pop (1000's)	199,059	322,088

Notes. Characteristics of counties in primary estimation sample, compared to all counties in ACS 2012-2016 5-year sample.

time on market as missing.

B.2 ACS

We use demographic information about counties from the ACS. For our cross-Sectional regressions, we use the ACS 5-year sample spanning the years 2012-2016. For our panel regressions, we use ACS 1-year samples spanning the years 2006-2017. We accessed the data using Social Explorer, a commercial provider of pre-aggregated ACS data. The demographic and housing stock characteristics we use are total population, population growth rate, total number of housing units, log average income, unemployment rate (calculated as one minus the fraction of population which is employed, divided by the fraction of the population in the labor force), the vacancy rate (calculated as the fraction of all surveyed houses which are vacant), the fraction of population aged 18-35 and 35-64, and the fractions of the population which are black, married, high school graduates, and college graduates.

B.3 Other time-on-market data sources

We use two other sources for time-on-market. First, we use Zillow Research data on time-on-market, which is available at the county-month level from 2010-2018. Second, we use Realtor.com time-on-market, which is available at the county-month level from 2012-2018.

B.4 County-year and county-year-month data construction

To construct the county-year dataset, we first aggregate the estimated errors $\hat{\epsilon}_{it}$ from specification (14) to the county-year level. We then get average prices, sales, and time-on-market by aggregating Corelogic data. The price line in the year plots, Figures 5 and 6, is the Zillow ZHVI. Results are qualitatively very similar if we instead use the Corelogic price index, or a price index which we construct using the Corelogic data.

To construct the monthly dataset, used in Figures 3 and 4, we aggregate total house sales from Corelogic and estimated errors $\hat{\epsilon}_{it}$ from specification (14) to the county month level, for all months within 2010-2018. We also take average time-on-market from the Corelogic MLS data, for the same sample period. For monthly prices, we do not use Zillow or Corelogic's house price indices, as both are seasonally adjusted; instead, we estimate a price index at the county-month level by regressing log house prices on county-month and house fixed effects, and taking the exponent of the county fixed effects as our price index. We normalize the index so it is equal to 1 in January 2010 for all counties.

B.5 Yearly and seasonal data construction

To construct the dataset used in Figures 5 and 6, we first filter to counties which we observe every year from 2000-2016; this leaves us with 361 counties, comprising approximately 38.8 million home sales. To construct the LogSD line in Figures 5 and 6, we average $\hat{\epsilon}_{it}^2$ over all observations within a given county-year, then take the square root of the resultant average. The time-on-market line represents the sales-weighted average of time-on-market across county-months in a given year, and the price line represents the sales-weighted average of the Zillow Home Value Index for single-family residences across county-months in a given year.

To construct the seasonal dataset, we filter to the years 2010-2016, as we only observe time-on-market for these years. This leaves us with 262 counties, comprising approximately 13.8 million home sales over this time period. We first collapse the data to year-month level. To collapse, we take the sum over sales in all counties, the mean over all $\hat{\epsilon}_{it}^2$ terms that we estimate, and the sales-weighted average of time-on-market. For the price line, we take the monthly FHFA house price index for the entire US, also for the time horizon 2010-2016. The Zillow ZHVI is seasonally adjusted, hence is not appropriate for studying seasonal variation in prices.

Since all four variables – prices, price dispersion, sales, and time-on-market – have low-frequency trends over time, we filter out these trends by regressing the collapsed year-month dataset outcomes on a third-order polynomial in year-month, subtracting away the predicted values, and adding back the mean. We then average the filtered series over years to the calendar month level, index each series to its January level, and plot the resultant series in Figures 3 and 4.

B.6 Implementation of specification (14)

When we estimate specification (14), it is computationally infeasible to estimate a fully interacted polynomial in all house characteristics for $f_c(x_i, t)$, so we use an additive functional form for:

$$f_c(x_i, t) = g_c^{latlong}(t, lat_i, long_i) + g_c^{sqft}(t, sqft_i) + g_c^{yrbuilt}(t, yrbuilt_i) + \\ g_c^{bedrooms}(t, bedrooms_i) + g_c^{bathrooms}(t, bathrooms_i) \quad (55)$$

The functions $g_c^{latlong}$, g_c^{sqft}, and $g_c^{yrbuilt}$ are fully-interacted third-order polynomials in their constituent components, and the functions $g_c^{bedrooms}$ and $g_c^{bathrooms}$ interact dummies for a given house having 1, 2, 3 or more bedrooms and 1, 2, 3 or more bathrooms respectively with third-order polynomials in time.

Additivity in specification (55) rules out many interaction effects between characteristics. Older or larger houses can appreciate faster or slower than newer or smaller houses. However, houses which are both large and old are constrained to appreciate at a rate which is the sum of the "old house" and "large house" effects on prices. The only interaction term we include is the $g_c^{latlong}(t, lat_i, long_i)$ function, which interacts latitude and longitude. This is important because it is implausible that latitude and longitude have additive effects on prices; effectively, this specification allows house prices to vary smoothly as a function of a house's geographic location over time.

Given this functional form for $f_c(x_i, t)$, specification (14) is a standard fixed effects regression, and we estimate specification (14) using OLS separately for each county in our sample. Once we have estimated specification (14), we estimate squared residuals $\hat{\epsilon}_{it}^2$ for each house sale as:

$$\hat{\varepsilon}_{it}^2 = \frac{N_c}{N_c - K_c} \left(p_{it} - \hat{p}_{it}\right)^2 \tag{56}$$

where N_c is the number of house sales in county c, and K_c is the number of parameters estimated from specification (14). The term $\frac{N_c}{N_c - K_c}$ is a degrees-of-freedom correction, which causes variance estimates to be unbiased at the county level; this is important to include because most houses are sold relatively few times, so the number of parameters K_c is nontrivially large relative to the number of house sales N_c in our dataset.

More formally, assuming homoskedasticity within counties, $\sigma_{it}^2 = \sigma_c^2$, the degrees-of-freedom correction in expression (56) causes the expectation of $\hat{\varepsilon}_{it}^2$ to be equal to the true variance, σ_c^2. We thus apply the homoskedastic variance adjustment term here, as we are not aware of any computationally tractable way to implement a degrees-of-freedom correction in the general heteroskedastic case. However, in Appendix D.1, we further adjust the estimated residuals $\hat{\varepsilon}_{it}$ to account for the number of times a house is sold and the average time-between-sales, and show that our results are robust to this adjustment.

C Supplementary material for section 4

C.1 Panel regressions

Table A2 shows county-year panel regressions of $logSD$ and TOM on various liquidity supply measures. This largely confirms findings from table 2 in the main text. Columns 1 and 5 show that vacancy rates are positively correlated with $logSD$ and TOM. Columns 2 and 6 show that population growth is populated with both liquidity measures. Columns 4 and 7 show that the coefficients are unchanged if both variables are included together. Finally, column 3 shows that time-on-market is also positively correlated with $logSD$ in the panel specification.

Table A3 shows results from regressing $logSD$ and TOM on income and prices. The results are mostly analogous to 3: price increases predict decreases in $logSD$, and income increases predict increases in $logSD$. Price increases predict increases in TOM, but the coefficient of TOM on income is not significant.

Table A2: County-year panel regressions

	Dependent variable:						
	LogSD	LogSD	LogSD	LogSD	TOM	TOM	TOM
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Vac rate	17.820^{***}			16.718^{***}	2.172^{***}		1.982^{***}
	(2.777)			(2.607)	(0.621)		(0.613)
Pop growth		-12.434^{***}		-9.969^{***}		-2.003^{***}	-1.711^{***}
		(3.502)		(3.355)		(0.603)	(0.582)
TOM			0.583^{***}				
			(0.147)				
County fixed effects	X	X	X	X	X	X	X
Year fixed effects	X	X	X	X	X	X	X
Sample period	2007-2016	2007-2016	2007-2016	2007-2016	2007-2016	2007-2016	2007-2016
Observations	3,360	3,360	3,360	3,360	3,360	3,360	3,360
Adjusted R^2	0.903	0.900	0.900	0.904	0.871	0.871	0.872

Notes. Each data point is a county-year. Regressions are weighted by the average number of sales in a given county over all years we observe. Standard errors are clustered at the county level.

Table A3: County-year panel regressions

	Dependent variable:	
	LogSD	TOM
	(1)	(2)
Log price	-3.424^{***}	0.266^{***}
	(0.657)	(0.097)
Log income	4.045^{***}	0.034
	(1.165)	(0.244)
County fixed effects	X	X
Year fixed effects	X	X
Sample period	2007-2016	2007-2016
Observations	3,360	3,360
Adjusted R^2	0.909	0.872

Notes. Each data point is a county-year. Regressions are weighted by the average number of sales in a given county over all years we observe. Standard errors are clustered at the county level.

D Supplementary material for section 5

D.1 Price dispersion robustness checks

In this appendix, we show that our main results are robust to using three alternative methods for estimating price dispersion.

Pure repeat sales specification: First, we omit the polynomial $f_c(x_i, t)$ term from (14), estimating residuals using the specification:

$$p_{it} = \gamma_i + \eta_{ct} + \epsilon_{it} \tag{57}$$

This corresponds to a pure repeat-sales specification for log prices.

Pure hedonic specification: Second, we omit house fixed effects from (14), estimating residuals using the following specification:

$$p_{it} = \eta_{ct} + f_c(x_i, t) + \epsilon_{it} \tag{58}$$

Adjusting for time-between-sales and times sold: Specification (14) implies that idiosyncratic price variance does not depend on the holding period. Also, when estimating (14), $\hat{\epsilon}_{it}^2$ will tend to be larger for houses which are sold more times, because the house fixed effect γ_i is estimated more precisely.

Let tbs_i be the average time-between-sales for house i, and let $sales_i$ be the total number of times we see house i being sold. Figure A1 plots a kernel regression fit of our estimated residuals, $|\hat{\epsilon}_{it}|$, against tbs_i, separately for $sales_i$ equal to 2, 3 and 4, for houses with tbs_i between the 1st and 99th percentiles for each value of $sales_i$. We see that the estimated logSD, $|\hat{\epsilon}_{it}|$, is on average higher when $sales_i$ and tbs_i are larger.

To ensure that these measurement issues are not driving our results, we attempt to purge $\hat{\epsilon}_{it}^2$ of any variation which can be explained by tbs_i and $sales_i$. First, we filter to houses sold at most four times over the whole sample period, with estimated values of $\hat{\epsilon}_{it}^2$ below 0.25. We then run the following regression, separately for each county:

$$\hat{\epsilon}_{it}^2 = g_c(sales_i, tbs_i) + \zeta_{it} \tag{59}$$

Where, $g_c(sales_i, tbs_i)$ interacts a vector of $sales_i$ dummies with a fifth-order polynomial in tbs_i. The residual $\hat{\zeta}_{it}$ from this regression can be interpreted as the component of the house's price variance which is not explainable by $sales_i$ and tbs_i. We then add back the mean of $\hat{\epsilon}^2_{it}$ within county c:

$$\hat{\epsilon}^2_{TBSadj,it} = \hat{\zeta}_{it} + E_c\left[\hat{\epsilon}^2_{it}\right] \tag{60}$$

$\hat{\epsilon}^2_{TBSadj,it}$ can be interpreted as the baseline estimates, $\hat{\epsilon}^2_{it}$, nonparametrically purged of all variation which is explainable by a smooth function of $sales_i$ and tbs_i. We use these estimates in the regressions of table A7.

Qualitative results: The top row of figure A2 compares residuals from the pure repeat-sales and pure hedonic specifications, (57) and (58) respectively, to our baseline residuals. The top left panel shows that the difference between repeat-sales residual estimates and the estimates from our baseline specification are quantitatively quite small, implying that the polynomial term $f_c(x_i, t)$ plays a relatively small role in fitting prices.

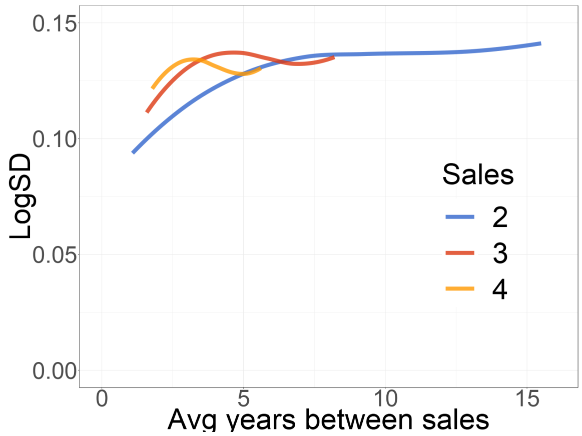
The top right panel of figure A2 shows that residual estimates from the pure hedonic specification are substantially higher than from our baseline specification, implying that house fixed effects are very important for accurately fitting prices. Note that we have included a degrees-of-freedom correction in all specifications, so this bias is not mechanically caused by estimating a larger number of parameters. Practically, the top row of figure A2 implies that idiosyncratic price dispersion can be estimated fairly well simply by taking the average residuals from a repeat-sales regression. We do not show time-between-sales adjusted residuals, because they are on average equal to residuals from the baseline specification, due to (60).

The bottom row of figure A2 shows how the different estimates of price dispersion behave seasonally and over the business cycle. All four measures are seasonal, with changes of similar magnitudes. Over the business cycle, the baseline, repeat-sales, and TBS-adjusted estimates behave very similarly. The hedonic estimate behaves somewhat differently, but is also noticeably countercyclical.

Our regression results are robust to all three different ways of measuring price dispersion. Table A4 collects specifications using logSD as the dependent variable from Tables 2 and 3 in the main text. In addition, in column 6, we include all dependent variables in a combined specification, and in column 7 we add state fixed effects. The signs of all coefficients in both specifications are the same as those in the main text, and most variables are significant.

In Tables A5, A6, and A7, we estimate all specifications in Table A4, using each of

Figure A1: Effect of $\mathtt{sales}_i$ and $\mathtt{tbs}_i$ on $\hat{\varepsilon}^2_{it}$



Notes. Predictions from kernel regressions of $\hat{\varepsilon}^2_{it}$ on $\mathtt{tbs}_i$, separately for each value of $\mathtt{sales}_i$.

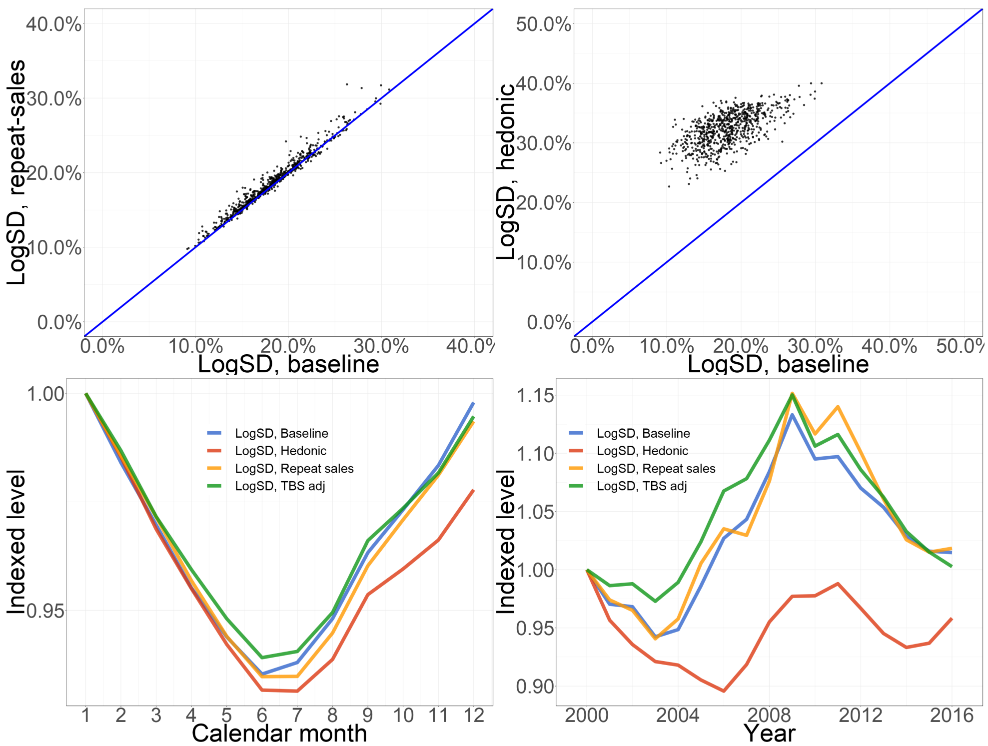
the three alternative estimates of price dispersion. Most results are qualitatively and quantitatively similar to those from the baseline specification. Across all specifications, prices, time-on-market, vacancy rates, and population growth rates are correlated with price dispersion significantly and in the expected directions.

D.2 Comparison of price dispersion estimates to literature

A number of other papers have attempted to measure idiosyncratic house price dispersion. Giacoletti (2017) uses the same Corelogic data that we use to measure idiosyncratic price dispersion in the metropolitan areas of San Francisco, San Deigo, and Los Angeles. Unlike our specification (14), Sagi (2015) and Giacoletti use returns, rather than individual house sales, as the primary unit of analysis.[10] Using returns, rather than sales, as the unit of analysis is more appropriate to the extent that the difference between an individual house's price and the county index follows a random walk. However, Sagi (2015) and Giacoletti show that the random walk assumption is rejected in the data; idiosyncratic variance does

[10]There are a number of other differences between Giacoletti's methodology and ours. First, Giacoletti measures returns with respect to Zillow's home value index, rather than adding county-month fixed effects as we do in this paper. Second, Giacoletti does not allow returns to flexibly vary over time as a function of house characteristics – characteristics are allowed to affect returns, but not in a time-dependent manner. Third, Giacoletti incorporates data on remodeling expenses in measuring price dispersion, which we do not do in this paper.

Figure A2: Effect of $sales_i$ and tbs_i on $\hat{\epsilon}_{it}^2$



Notes. The top left plot shows repeat-sales residuals, from (57), against the baseline residual estimates, and the top right plot shows residuals from the hedonic specification (58). We aggregate residuals to the county level from 2012-2016.

Table A4: County cross-sectional regressions, price dispersion

	Dependent variable:						
	LogSD						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log price	-4.051^{***}					-4.729^{***}	-3.339^{***}
	(0.655)					(0.655)	(0.752)
Log income	5.625^{***}					8.027^{***}	10.752^{***}
	(1.987)					(1.886)	(1.917)
Vac rate		16.479^{***}			18.187^{***}	15.788^{***}	7.827^{***}
		(2.903)			(2.878)	(2.850)	(2.441)
Pop growth			-21.784^{***}		-28.573^{***}	-7.036	-20.542^{***}
			(7.481)		(7.191)	(7.329)	(6.646)
TOM				1.071^{***}		0.926^{***}	1.626^{***}
				(0.254)		(0.250)	(0.282)
Controls	X	X	X	X	X	X	X
Fixed effects							State
Observations	405	405	405	405	405	405	405
Adjusted R^2	0.638	0.630	0.606	0.616	0.644	0.691	0.824

Notes. Each data point is a county. Regressions are weighted by the number of sales within the county over the time period 2012-2016.

Table A5: County cross-sectional regressions, price dispersion, repeat-sales

	Dependent variable:						
	LogSD						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log price	-4.588^{***}					-5.292^{***}	-3.755^{***}
	(0.688)					(0.699)	(0.828)
Log income	8.279^{***}					10.453^{***}	13.601^{***}
	(2.089)					(2.013)	(2.110)
Vac rate		16.898^{***}			18.295^{***}	17.322^{***}	9.571^{***}
		(3.070)			(3.070)	(3.041)	(2.686)
Pop growth			-16.536^{**}		-23.366^{***}	-2.506	-17.437^{**}
			(7.938)		(7.673)	(7.820)	(7.314)
TOM				0.754^{***}		0.645^{**}	1.534^{***}
				(0.271)		(0.267)	(0.310)
Controls	X	X	X	X	X	X	X
Fixed effects							State
Observations	405	405	405	405	405	405	405
Adjusted R^2	0.642	0.629	0.603	0.607	0.637	0.685	0.809

Notes. Each data point is a county. Regressions are weighted by the number of sales within the county over the time period 2012-2016.

Table A6: County cross-sectional regressions, price dispersion, hedonic

	Dependent variable:						
	LogSD						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log price	-3.253^{***}					-3.643^{***}	-2.688^{***}
	(0.518)					(0.515)	(0.644)
Log income	2.043					3.729^{**}	5.347^{***}
	(1.573)					(1.484)	(1.641)
Vac rate		14.263^{***}			15.925^{***}	14.336^{***}	10.730^{***}
		(2.322)			(2.278)	(2.242)	(2.089)
Pop growth			-21.847^{***}		-27.792^{***}	-9.697^{*}	-25.675^{***}
			(5.987)		(5.692)	(5.766)	(5.688)
TOM				0.801^{***}		0.575^{***}	0.675^{***}
				(0.205)		(0.197)	(0.241)
Controls	X	X	X	X	X	X	X
Fixed effects							State
Observations	405	405	405	405	405	405	405
Adjusted R^2	0.630	0.614	0.589	0.591	0.636	0.688	0.790

Notes. Each data point is a county. Regressions are weighted by the number of sales within the county over the time period 2012-2016.

Table A7: County cross-sectional regressions, price dispersion, time-between-sales adjusted

	Dependent variable:						
	LogSD						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log price	-3.174^{***}					-3.528^{***}	-2.240^{***}
	(0.485)					(0.495)	(0.548)
Log income	5.301^{***}					6.596^{***}	7.668^{***}
	(1.471)					(1.426)	(1.396)
Vac rate		11.486^{***}			12.617^{***}	12.231^{***}	5.950^{***}
		(2.166)			(2.157)	(2.155)	(1.778)
Pop growth			-14.206^{**}		-18.916^{***}	-5.064	-11.198^{**}
			(5.568)		(5.391)	(5.541)	(4.841)
TOM				0.463^{**}		0.353^{*}	1.155^{***}
				(0.191)		(0.189)	(0.205)
Controls	X	X	X	X	X	X	X
Fixed effects							State
Observations	405	405	405	405	405	405	405
Adjusted R^2	0.649	0.635	0.614	0.613	0.646	0.687	0.835

Notes. Each data point is a county. Regressions are weighted by the number of sales within the county over the time period 2012-2016.

scale with holding periods, but much more slowly than under a random walk model. A related paper is Carrillo, Doerner and Larson (2019), which finds that excess returns of individual houses over market averages are mean-reverting in subsequent transactions.

The results of these papers thus support the use of our specification (14) to measure price dispersion. Specification (14) goes further, assuming that idiosyncratic variance has no relationship with holding period; this is violated in the data, but we relax this in appendix D.1. The benefit of our measurement strategy is that, since we can measure errors at the level of individual house sales, rather than pairs of purchases and sales, our estimates of idiosyncratic price dispersion can be flexibly aggregated cross-sectionally and over time. This is necessary for us to produce our stylized facts, which we believe are new to the literature: that idiosyncratic price dispersion is countercyclical, seasonal, and correlated with time-on-market and other measures of market tightness. These results build on and complement Giacoletti (2017), who shows that contractions to mortgage credit availability at the zipcode level are associated with increased idiosyncratic variance, and Landvoigt, Piazzesi and Schneider (2015), who show that idiosyncratic variance increased in San Diego following the 2008 housing bust.

Peng and Thibodeau (2017) uses a purely hedonic specification to measure price dispersion, analyzing the relationship between idiosyncratic price dispersion and various other variables in the cross-section of zipcodes. To address the possibility that the hedonic model determining prices changes over time, Peng and Thibodeau (2017) runs separate hedonic regressions for different time periods. We address this issue through the hedonic $f_c(x_i, t)$ term in specification (14), which effectively allows the hedonic coefficients on different characteristics to change continuously over time. In appendix D.1, we measure price dispersion using a purely hedonic specification for log prices, similar to Peng and Thibodeau (2017); this does not substantially change our results.

Two other papers which measure idiosyncratic price dispersion are Anenberg and Bayer (2015) and Landvoigt, Piazzesi and Schneider (2015). Anenberg and Bayer (2015), as an input moment for estimating their structural model, estimate the idiosyncratic volatility of house prices using a repeat-sales specification with zipcode-month and house fixed effects, without allowing characteristics to affect prices over time. Landvoigt, Piazzesi and Schneider (2015) estimates idiosyncratic price dispersion assuming that the only characteristic that affects mean returns is a house's previous sale price. Our specification (14) does not nest that of Landvoigt, Piazzesi and Schneider (2015), since we do not include previous sale prices in specification (14); however, to the extent that the factors which affect prices are summarized by our house characteristics x_t, our specification will also be able to capture these trends.

Quantitatively, Giacoletti (2017), using data from 1989 to 2013, finds that the standard deviation of idiosyncratic component of returns is approximately 9.6%-11.8% in San Diego, 13.9%-16.5% in Los Angeles, and 13.7-17.6% in San Francisco. Landvoigt, Piazzesi and Schneider (2015) finds a similar SD of 8.8%-13.8% for San Diego over the time horizon 1999-2007. In our sample, over the time period 2000-2017, we estimate return standard deviations of 15.7% for San Diego, 16.8% for Los Angeles, and 19.1% for San Francisco.[11] Our estimates are thus roughly in line with the estimates from Giacoletti (2017) and Landvoigt, Piazzesi and Schneider (2015), preserving the ordering of idiosyncratic price dispersion between the three regions, although our estimates are somewhat higher than theirs. Moreover, similar to our findings, Landvoigt, Piazzesi and Schneider (2015) finds that price dispersion increased during the 2008 housing bust, though their sample does not include the subsequent recovery. Thus, our paper, Giacoletti (2017), and Landvoigt, Piazzesi and Schneider (2015) arrive at similar estimates using different methodologies, datasets, time horizons, and geographic definitions, suggesting that the stylized facts we document are fairly robust to different measurement strategies.

D.3 Time-on-market robustness checks

For robustness, we repeat our main analyses using two different data sources for time-on-market: Realtor.com time-on-market, which is available at the county-month level from 2012 to 2017, and Zillow Research time-on-market, which is available from 2010 to 2018.

In the top two panels of Figure A3, we aggregate both time-on-market sources to the county level, using data within the interval 2012-2016, and show how they correlate with Corelogic time-on-market across counties. Realtor.com time-on-market is somewhat lower than Corelogic, and Zillow time-on-market is somewhat higher, but all three measures are very positively correlated.

The bottom row of Figure A3 shows how Zillow and Realtor.com time-on-market behave seasonally and over the business cycle. All three data sources display seasonality,

[11] To calculate these quantities, we take sales-weighted averages of $\hat{\sigma}_c^2$ for all counties within the San Diego-Carlsbad-San Marcos, San Francisco-Oakland-Fremont, and Los Angeles-Long Beach-Anaheim CBSAs. We then multiply $\hat{\sigma}_c$ by a factor of $\sqrt{2}$, to convert standard deviations of prices at each sale to standard deviations of returns, which can then be compared directly to the estimates in Giacoletti (2017) and Landvoigt, Piazzesi and Schneider (2015).

although the patterns and magnitudes are somewhat different for the Realtor.com data. While the Zillow and Realtor.com data do not go very far back in time, all data sources display a decrease in time-on-market from around 2010 onwards.

Table A8 collects specifications using TOM as the dependent variable from Tables 2 and 3 in the main text. In column 5, we include all dependent variables in a combined specification, and in column 6 we add state fixed effects. The signs of all coefficients in both specifications are the same as those in the main text, and most variables are significant.

In Tables A9 and A10, we repeat the regressions of Table A8 using Zillow and Realtor.com time-on-market respectively as dependent variables. Most results are qualitatively unchanged.

E Supplementary material for section 6

E.1 Moment matching

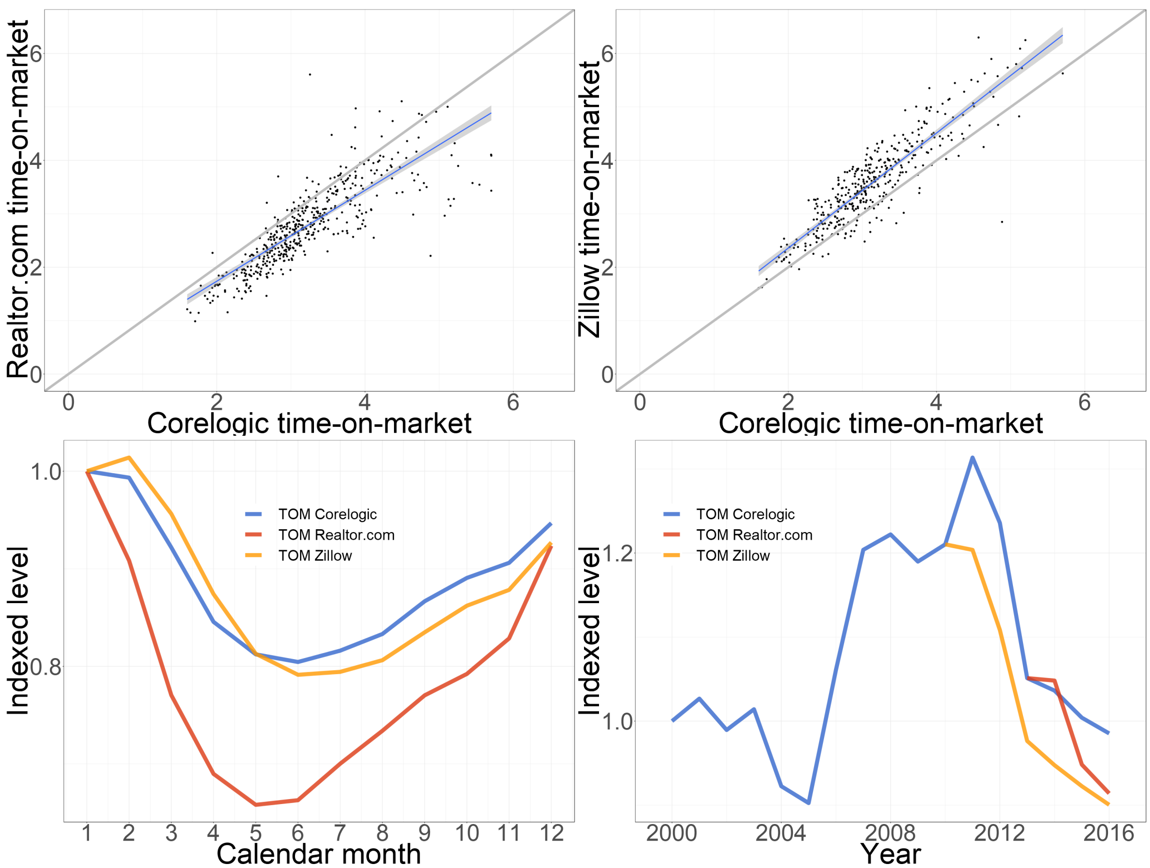
The free parameters in our model are $\lambda_M, \eta_B, \bar{c}, \Delta_c, \epsilon_0, \sigma_\epsilon^2$. We match these parameters to data moments using an inner-outer loop procedure. In the outer loop, we match Δ_c, and in the inner loop we solve for all other parameters conditional on Δ_c. In the inner loop, for any guess for Δ_c, we choose other parameters to exactly match five data moments. In the outer loop, we generate a model-implied relationship between time-on-market and price dispersion using a procedure described in Subsection E.1.2, and we choose Δ_c to match the model-implied TOM-PD relationship to the data. Computational details of how we solve the model for a given choice of parameters are described in Appendix E.1.3.

E.1.1 Inner loop

Given any value of Δ_c, we choose $\lambda_M, \eta_B, \bar{c}, \epsilon_0, \sigma_\epsilon^2$ to match the average values of five moments. Three of the target moments come from a sample of counties in 2016.

We target two moments from other papers in the literature. The first is the average number of houses that buyers visit before buying, which Genesove and Han (2012) find to be 9.96, in the US. The second is the dispersion in buyer values for houses, from Anundsen, Larsen and Sommervoll (2019). Using data on Norwegian residential real

Figure A3: TOM robustness



Notes. The top left panel plots Realtor.com time-on-market against Corelogic time-on-market, and the top right panel plots Zillow time-on-market against corelogic. Each data point is a county, and time-on-market is in months. The bottom left panel shows seasonal variation in each time-on-market measure, where each measure is indexed to 1 in January. The bottom right panel shows time-on-market over the business cycle. Realtor.com and Zillow time-on-market are indexed so that they are equal to Corelogic time-on-market, in the first year that we observe them.

Table A8: County cross-sectional regressions

	Dependent variable:					
	Time-on-market					
	(1)	(2)	(3)	(4)	(5)	(6)
Log price	0.500***				0.588***	0.753***
	(0.136)				(0.134)	(0.142)
Log income	−1.575***				−1.469***	−0.528
	(0.414)				(0.389)	(0.376)
Vac rate		3.648***		3.945***	3.734***	2.285***
		(0.580)		(0.578)	(0.567)	(0.463)
Pop growth			−3.485**	−4.958***	−6.649***	−5.236***
			(1.516)	(1.445)	(1.502)	(1.275)
Controls	X	X	X	X	X	X
Fixed effects						State
Observations	405	405	405	405	405	405
Adjusted R^2	0.546	0.573	0.533	0.585	0.605	0.803

Notes. Each data point is a county. Regressions are weighted by the number of sales within the county over the time period 2012-2016. The dependent variable is Zillow time-on-market.

Table A9: County cross-sectional regressions, Zillow time-on-market

	Dependent variable:					
	Time-on-market					
	(1)	(2)	(3)	(4)	(5)	(6)
Log price	0.267				0.388^{**}	0.912^{***}
	(0.178)				(0.174)	(0.174)
Log income	-1.318^{**}				-1.210^{**}	-0.620
	(0.541)				(0.501)	(0.457)
Vac rate		4.294^{***}		4.829^{***}	4.637^{***}	2.670^{***}
		(0.768)		(0.752)	(0.750)	(0.578)
Pop growth			-6.377^{***}	-8.224^{***}	-9.126^{***}	-7.131^{***}
			(1.924)	(1.823)	(1.937)	(1.550)
Controls	X	X	X	X	X	X
Fixed effects						State
Observations	324	324	324	324	324	324
Adjusted R^2	0.602	0.635	0.610	0.658	0.664	0.855

Notes. Each data point is a county. Regressions are weighted by the number of sales within the county over the time period 2012-2016. The dependent variable is Zillow time-on-market.

Table A10: County cross-sectional regressions, Realtor.com time-on-market

	Dependent variable:					
	Time-on-market					
	(1)	(2)	(3)	(4)	(5)	(6)
Log price	0.386***				0.488***	0.725***
	(0.134)				(0.124)	(0.137)
Log income	−2.004***				−1.850***	−0.606*
	(0.406)				(0.359)	(0.361)
Vac rate		4.563***		5.095***	4.822***	3.897***
		(0.569)		(0.551)	(0.535)	(0.459)
Pop growth			−5.785***	−7.858***	−8.682***	−6.812***
			(1.486)	(1.351)	(1.388)	(1.227)
Controls	X	X	X	X	X	X
Fixed effects						State
Observations	385	385	385	385	385	385
Adjusted R^2	0.614	0.652	0.605	0.682	0.704	0.841

Notes. Each data point is a county. Regressions are weighted by the number of sales within the county over the time period 2012-2016. The dependent variable is Realtor.com time-on-market.

estate auctions,[12] Table 1 of Anundsen, Larsen and Sommervoll (2019) reports a variety of summary statistics about bid-ask, bid-appraisal, and bid-sell spreads. We target the lowest spread in Table 1, the difference between the opening bid price and the ask price, which is 6.45% of house prices on average. This will tend to produce conservative estimates of buyer value-induced price variance.

We match buyer-induced price variance in the model by requiring the square root of buyer match utility-induced price variance to equal 6.45% of house prices; that is, we will choose σ_ϵ such that:

$$\frac{1}{E\left(P\left(c,\epsilon\right)\right)}\sqrt{\left(\frac{\theta}{r+\lambda_M}\right)^2\sigma_\epsilon^2} = 0.0645$$

While the mapping between moments and parameters is complex, roughly speaking, the input parameters determine the output moments as follows. The lower bound, ϵ_0, and dispersion, σ_ϵ^2, of buyer utilities jointly determine the level of buyer-induced price dispersion and the overall level of house prices. The mean of seller holding utilities, $\bar{c}$, and the lower bound, ϵ_0, affect average gains-from-trade and thus move prices and the average number of house visits by buyers before purchasing. The entry rate η_B of buyers determines market tightness, which determines average time-on-market. The separation rate λ_M is tightly linked to the turnover rate.

Since buyer- and seller-induced price variance do not add up to total price variance, there is a residual term. This could be driven by a number of factors: such as unobserved house-level heterogeneity or renovations, or realtor bargaining frictions. Using our estimates, we can do a simple accounting of roughly how much of total price dispersion, under our estimates, is attributable to preferences. That is, we can decompose total logSD in the data into components attributable to buyer values, seller values, and residuals:

$$(0.165)^2 \approx \underbrace{\frac{Var_{c\sim F(\cdot)}\left(V_S\left(c\right)\right)}{\left(E\left(P\left(c,\epsilon\right)\right)\right)^2}}_{\text{Seller holding value}} + \underbrace{\frac{1}{\left(E\left(P\left(c,\epsilon\right)\right)\right)^2}\left(\frac{\theta}{r+\lambda_M}\right)^2\sigma_\epsilon^2}_{\text{Buyer match value}} + \sigma_{residual}^2 \tag{61}$$

In our baseline model estimates, the seller holding value component has a standard deviation of 3.52% of prices, the buyer component is 6.45% of house prices, and the unobserved heterogeneity component is 14.77% of house prices. In terms of variance fractions, 4.5% of total logSD is attributable to seller values, 15.3% to buyer values, and

[12]We use these data because we are unaware of publically available bid data on housing auctions in the US.

80.2% to unobserved heterogeneity. Thus, in our baseline calibration, preferences can account for a nontrivial component of total idiosyncratic dispersion, but a large fraction is attributed to residual factors.

E.1.2 Outer loop

In the data, we have several ways to measure the relationship between price dispersion and time-on-market. The coefficient is relatively similar in the panel, seasonal and time-series regression specifications. For our calibration, we use the panel regression coefficient.

To most closely match the data, holding fixed $\lambda_M, \bar{c}, \Delta_c, \epsilon_0, \sigma_\epsilon^2$, we perturb η_B to simulate model implied grids of dollar price dispersion and time-on-market. We divide dollar price dispersion by the average price at the initial set of parameters $\lambda_M, \eta_B, \bar{c}, \Delta_c, \epsilon_0, \sigma_\epsilon^2$ to obtain percentage price dispersion. We then regress simulated percentage price dispersion from our model on simulated time-on-market, to generate a model-predicted regression coefficient. We choose Δ_c to match this model-predicted TOM-logSD relationship to the county panel coefficient.

E.1.3 Computation

Computationally, we solve the model by varying M_B rather than η_B; this is computationally simpler, and for any equilibrium in terms of M_B, we can back out an η_B which implements this equilibrium, through (10). Given a vector of parameters $r, \alpha, \phi, \theta, \lambda_M, M_B, \bar{c}, \Delta_c, \epsilon_0, \sigma_\epsilon^2$ we solve the model by iteratively solving the Bellman equations and flow equality conditions until convergence. Given guesses for $f_{eq}(c), M_S$, we calculate λ_S and λ_B, and then numerically solve (3), (4), (5) for $V_S(c), V_M(\epsilon), V_B, \epsilon^*(c)$. Given guesses for the trade cutoff $\epsilon^*(c)$, we can then use (8) to solve for $f_{eq}(c)$. We iterate these equations, updating in a penalized matter; if the result of one iteration on M_S implies some new value $\tilde{M}_S$, for the next iteration, we update M_S to:

$$(1-t) M_S + t\tilde{M}_S$$

For small enough t, the iteration converges. While we were not able to prove that the model admits a unique solution, in our simulations, the model reached the same equilibrium point from many different starting values.

E.2 Log price variance approximation

In our data, idiosyncratic price variance corresponds to the residual from a regression in which the dependent variable is the log sale price; hence, the residual can be interpreted as the variance of log prices. In the model, the variance of prices is computationally easy to calculate, using the analytical result of claim 1, but the variance of log prices is more complex. For computational simplicity, in generating the variance of log prices in the model, we use the following linear approximation, based on the Taylor expansion of $\log(P)$ around its mean $\bar{P}$:

$$Var(\log(P)) \approx Var\left(\log\left(\bar{P}\right) + \frac{P-\bar{P}}{\bar{P}}\right) = Var\left(\frac{P-\bar{P}}{\bar{P}}\right) = \frac{1}{\bar{P}^2} Var(P)$$

Hence, we generate the variance of prices as the variance of model-generated prices, divided by the squared mean of model-generated prices, where we calculate the variance of model-generated prices using expression (11) of Claim 1.

E.3 Literature estimates of liquidity discounts

A number of papers in the literature have documented various factors which affect sale prices and time-on-market, through a channel which is plausibly related to seller patience. For each of these papers, we calculate the implied 1-month effects, essentially by dividing the estimated price effects by estimated time-on-market effects. We describe our calculation methodology for each row of Table 5 below. We divide the papers into two groups: papers which estimate foreclosure discounts, and those which estimate "liquidity discounts" driven by factors other than foreclosure. The reason for this is that foreclosure discount estimates are systematically higher than liquidity discounts.

Liquidity discounts

- Genesove and Mayer (1997), using data from Boston, MA from 1990-1992, analyzes the relationship between owners' equity position, time-on-market, and prices. Intuitively, owners who have higher home equity set higher list prices, take longer to sell, and achieve higher sale prices. They find that a homeowner with loan-to-value 1 sells for 4.3% higher than a homeowner with loan-to-value 0.8, and remains on market 15% longer. Assuming average time-on-market is 2.6 months, the estimated

1-month effect is:

$$\frac{4.3\%}{(0.15)(2.6)} = 11.02\%$$

- Genesove and Mayer (2001), using data from condos in Boston, MA from 1990-1997, analyze the behavior of sellers subject to different amounts of nominal losses, due to the time they purchased their houses. Sellers subject to nominal losses set higher list prices, sell more slowly, and sell for higher prices. They find that sellers pass through around 3-18% of nominal losses; hence, with a 10% higher nominal loss, sellers set asking prices between 0.3% and 1.8% higher. Time-on-market is around 3-6% higher. In our data, average time-on-market is around 2.6 months. We can calculate an upper bound on the 1-month effect by taking the upper estimate of the price effect, and the lower estimate of the time-on-market effect, of a 10% nominal loss:

$$\frac{1.8\%}{(0.03)(2.6)} = 24\%$$

  As a lower estimate, we can plug in the lower estimate of the price effect and the upper estimate of the time-on-market effect:

$$\frac{0.3\%}{(0.06)(2.6)} = 1.92\%$$

- Levitt and Syverson (2008), using data from Cook County, IL from 1992-2002, analyze sales of realtor-owned houses. They find that realtors tend to sell more slowly, but for higher prices: realtors spend around 9.5 extra days on market, and sell for 3.7% higher prices. The estimated 1-month effect is:

$$\frac{3.7}{\left(\frac{9.5}{30}\right)} = 11.68\%$$

  Note that this estimate assumes that all realtors do is set higher list prices. Realtors are likely to have a better selling technology – for example, they may be more effective at finding buyers. In this case, this is likely to be an overestimate of the 1-month effect.

- Hendel, Nevo and Ortalo-Magné (2009) analyze FSBO transactions, using data spanning 1998-2005 in Wisconsin, WI. They find that FSBO sales take around 20 days longer to sell, and sell at the same price, but without sellers' realtor commissions. If we assume realtor commissions are 3%, this gives an estimated 1-month effect of:

$$\frac{3}{\left(\frac{20}{30}\right)} = 4.5\%$$

There are clearly other differences between FSBO sales and MLS sales, but this corresponds to the 1-month discount if we simply think of FSBO sales as a slower, but higher-price way to sell a house.

- Guren (2018) uses data from Los Angeles, San Diego, and San Francisco, from 1988-2013. Similar to Genesove and Mayer (1997), Guren uses price appreciation since purchase as an instrument for sellers' marginal utility for cash, and thus sellers' urgency. While Guren emphasizes the curvature of demand, we can use the slope of demand to estimate the 1-month price effect, using the IV estimates in his Figure 2. In this figure, a relative markup change of 4% – that is, from -0.02 to 0.02 – changes the 13-week sale probability of a house from 0.5 to 0.4. Assuming that house sales follow a Poisson process, and assuming 4 weeks in a month, the sale probabilities map to expected time-on-markets of 6.5 months and 8.125 months, respectively. Assuming list prices pass through perfectly to sale prices, this gives a lower estimate of the 1-month effect of:
$$\frac{4}{8.125 - 6.5} = 2.46\%$$
We note that the estimated time-on-markets, using this method, are much higher than our estimate of 2.6 months. If we instead assume time-on-market increases by 25%, using a base of 2.6 months, we can calculate an upper estimate of the 1-month effect as:
$$\frac{4}{\left(\frac{8.125}{6.5} - 1\right)(2.6)} = 6.15\%$$
Another caveat to note is that our assumption that list prices pass through perfectly to sale prices is likely an overestimate. In support of this assumption, however, in appendix D.1, Guren writes that "the modal house sells at its list price" in his sample.

- Buchak et al. (2020) analyze I-buyers, using data from Phoenix, Las Vegas, Dallas, Orlando, and Gwinnet Atlanta, from 2013-2018. They find that I-buyers purchase houses at at around 3.6% lower prices. In addition, we found that Zillow charges the standard 6% realtor commission, as well as around 1.4-8% extra fees, depending on the region in question. Thus, combining the price discount and the explicit fee, a buyer is effectively paying around 5-11.6% more than they would pay if they used a realtor, in order to sell instantly. Assuming time-on-market is 2.6 months, we can calculate upper and lower estimates on the implied 1-month effect as:

$$\frac{5}{2.6} = 1.92\%, \; \frac{11.6}{2.6} = 4.46\%$$

Foreclosure discounts

- Pennington-Cross (2006), using confidential data, finds foreclosure discounts or around 22% of house prices.
- Clauretie and Daneshvary (2009), using data from Las Vegas from 2004-2007, find foreclosure discounts averaging around 10%.
- Campbell, Giglio and Pathak (2011), using data from Massachusetts from 1987-2009, find foreclosure discounts of around 27%. Note that Campbell, Giglio and Pathak also analyze discounts from "forced sales", driven by deaths or bankruptcies of sellers, but which are not foreclosures. These discounts are much smaller, at 3-7%. This is in the range of the 1-month effects from our model and other papers, but the paper does not report the average time-on-market difference between forced sales and normal sales, so we cannot calculate a 1-month effect.
- Harding, Rosenblatt and Yao (2012), using data from 13 MSAs from 1990-2008, calculate foreclosure discounts using a "holding period returns" methodology. Figure 2 of Harding, Rosenblatt and Yao (2012) shows that that, while returns are very high for foreclosures held for 1 year, foreclosed houses held for 2-4 years only make around 5% excess returns on average. The estimates vary somewhat across specifications, but tend to be lower than other papers in the literature.
- Zhou et al. (2015), using data from 16 CBSAs from 2000-2012, finds foreclosure discounts ranging from around 11% (Los Angeles) to 26% (Chicago). The average across CBSAs is around 15%. There is also substantial time-series variation in the estimates.