

# Liquidity in Residential Real Estate Markets\*

Nadia Kotova<sup>†</sup> and Anthony Lee Zhang<sup>‡</sup>

April 2021

## Abstract

We build a rich panel dataset tracking two measures of housing market liquidity: time-on-market and price dispersion. The two measures co-vary closely at seasonal and business-cycle frequencies, but there is substantial independent variation in the cross-section of counties. This suggests that the two measures reflect different dimensions of market liquidity. Using a housing search model, we show that time-on-market and price dispersion can be thought of as equilibrium outcomes from a supply and demand system for liquidity. Consistent with the model's predictions, proxies for liquidity supply are negatively correlated with both measures, whereas a proxy for liquidity demand is negatively correlated with time-on-market, but positively correlated with price dispersion.

**Keywords:** Housing, Search, Liquidity

---

\*This paper was previously circulated under the title “Search Frictions and Idiosyncratic Price Dispersion in the US Housing Market”. 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, Cathy Ruochen Wang, Joseph Vavra, Robert Wilson, Jiro Yoshida, Chenyanzi (Shannen) Yu, Ali Yurukoglu, and Fudong Zhang, Eric Zwick, 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, the 2019 WCSM Workshop, UChicago, the AREUEA Virtual Seminar, the Columbia HULM Conference 2020, and ASSA 2021.

<sup>†</sup>Stanford Graduate School of Business, 655 Knight Way, Stanford, CA 94305; [nkotova@stanford.edu](mailto:nkotova@stanford.edu).

<sup>‡</sup>University of Chicago Booth School of Business, 5807 S Woodlawn Ave, Chicago, IL 60637; [anthony.zhang@chicagobooth.edu](mailto:anthony.zhang@chicagobooth.edu).

# 1 Introduction

Assets are liquid if they can be sold quickly and at stable prices. According to this definition, the residential real estate market in the US is highly illiquid. Houses take many months to sell, and realtor commissions and other fees amount to 6-10% of house prices. Illiquidity is also reflected in idiosyncratic price dispersion: similar houses can sell for very different prices at similar points in time. This idiosyncratic price component is quantitatively large, as it accounts for more than half of total capital gains volatility for individual homes. However, there is little understanding in the literature of how market liquidity varies across regions and over time, and what forces determine liquidity in residential real estate markets.

In this paper, we build a rich panel dataset tracking two housing market liquidity measures: time-on-market and idiosyncratic price dispersion. The two measures co-vary closely at seasonal and business-cycle frequencies, but there is substantial independent variation in the cross-section of counties. This observation suggests that time-on-market and price dispersion measure different aspects of liquidity. We build a search-and-bargaining model of the housing market, and show that time-on-market and price dispersion arise as equilibrium outcomes from a supply and demand system for liquidity. Increases in liquidity supply cause both measures to decrease. Increases in liquidity demand cause time-on-market to decrease, but price dispersion to increase. Consistent with the model's predictions, proxies for liquidity supply are negatively correlated with both liquidity measures in the data, whereas a proxy for liquidity demand is negatively correlated with time-on-market but positively correlated with price dispersion.

We start our analysis by measuring time-on-market and price dispersion for a large set of US counties over time. Time-on-market is defined as the average time between when a house is listed and when it is sold. Idiosyncratic price dispersion is the extent to which similar houses sold at similar points in time sell for different prices. To measure this, we first regress log sale prices of houses on county-month fixed effects, house fixed effects, and a smooth function of house characteristics and time. We then take the residuals from this regression, which can be flexibly aggregated cross-sectionally and over time, as our measure of idiosyncratic price dispersion.

We show that time-on-market and price dispersion are countercyclical and seasonal, and co-move with other measures of housing market "hotness", such as sales volume and average price. Counties which are more cyclical or seasonal in one variable also tend to

be more cyclical or seasonal in other variables. In the cross-section of counties, there is large variation in both time-on-market and price dispersion. The 10th and 90th percentiles of time-on-market are, respectively, 2.40 and 4.20, and the 10th and 90th percentiles of price dispersion are 14.0% and 22.7%. While time-on-market and price dispersion are positively correlated, there is substantial idiosyncratic variation. Time-on-market and price dispersion are not very well explained by each other, or by the level of house prices. For example, house prices and time-on-markets are virtually identical in Rochester and Baltimore, but the two cities exhibit large differences in idiosyncratic price dispersion: 14.1% and 27.5%, respectively. This difference is similar to the difference in annual return volatilities of US high yield bonds and emerging markets stocks.

These stylized facts suggest that housing market liquidity appears to be a multi-dimensional object, which time-on-market and price dispersion measure different aspects of. We build a search-and-bargaining model to explain the variation in time-on-market and price dispersion. 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 the distribution of prices in equilibrium. Buyers with higher match quality draws pay higher prices, and sellers with higher holding costs sell for lower prices.

In the model, the main drivers of market outcomes are the *supply* and *demand* for liquidity. *Liquidity supply* is buying pressure: the inflow rate of buyers, which determines equilibrium market tightness. *Liquidity demand* is the average urgency of sellers to sell their houses. Liquidity supply changes create positive co-movement in time-on-market and price dispersion: when there are more buyers, sellers can sell faster and at more stable prices, causing price dispersion and time-on-market to both decrease. Liquidity demand changes create negative co-movement: when sellers' urgency increases, sellers collectively sell faster, causing prices to be lower and more dispersed, so time-on-market decreases, but price dispersion actually increases.

Through the lens of the model, markets with high liquidity supply will have low time-on-market and price dispersion, whereas markets with high liquidity demand will have low time-on-market but high price dispersion. Thus, neither metric alone is sufficient as a measure of market liquidity. Time-on-market can be low, either because there are many buyers and markets are tight, or because sellers have high urgency, and sell quickly despite incurring large liquidity discounts.

Our model makes testable predictions about cross-sectional variation in liquidity measures: proxy variables for liquidity supply should be negatively associated with both price dispersion and time-on-market, whereas measures of liquidity demand should correlate negatively with time-on-market, but positively with price dispersion. We test the predictions of our model in the data. The first proxy that we use is county net migration rate. Intuitively, a county that has a high immigration rate should have a large mass of buyers interested in purchasing houses. Consistent with our model’s predictions, net immigration is negatively correlated with both price dispersion and time-on-market.

Next, we use an empirical approach based on Schubert (2021) to construct a plausibly exogenous shock to liquidity supply: *migration spillovers* from high-productivity areas. If a certain county experiences large productivity shocks, house prices will increase. This will tend to cause outmigration from the county, which will create net immigration flows to areas with strong historical migration links to the county. These migration flows are plausibly exogenous shocks to housing market tightness. As our model predicts, areas which experience large migration shocks have lower time-on-market and lower price dispersion.

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 be willing to sell faster, even if this lowers sale prices, since their opportunity cost of keeping their houses on the market is higher. Our model thus predicts that higher-income areas should have lower time-on-market, but higher price dispersion. We verify this prediction in the data.

Our empirical results survive a number of robustness checks. The results largely hold in panel regressions, though some coefficients lose significance. Our results about price dispersion hold using three other estimation methods: a pure hedonic price model, a pure repeat-sales model, and a model in which estimated residuals are non-parametrically adjusted for the number of times a house is sold, and the average time between house sales. Our results about time-on-market hold using two other 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.

Finally, we calibrate the model to the data. We use the 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, and the “liquidity discounts” that sellers

incur if they sell their houses quickly. Under our model estimates, a seller who chooses to spend an additional month on the market sells at a price 5% higher. This is consistent with estimates of liquidity discounts in the housing market microstructure literature, which range from 1.9% to around 11%. Hence, our model can quantitatively match the aggregate relationship between time-on-market and price dispersion, with parameters that imply reasonable values of sellers' liquidity discounts relative to previous literature.

Our results have implications for academics and policymakers. There is a large academic literature on housing market liquidity, and many industry participants use various measures to analyze the "hotness" or liquidity of housing markets cross-sectionally and over time.<sup>1</sup> Time-on-market is commonly used as a measure of liquidity in both the academic literature and industry practice. However, while it is less commonly studied by practitioners, the academic literature has shown that idiosyncratic price dispersion is very high for residential real estate, and is important in accounting for risk and returns of housing as an asset class.

Time-on-market would be a reasonable measure of housing market liquidity in a world in which all liquidity metrics co-move closely, since we could predict any other welfare-relevant liquidity measure using time-on-market. We find that this is not the case: there are many markets with low time-on-market but high price dispersion, and vice versa, and our model explains why this can happen. Practically, our findings imply that academics and policymakers who are interested in studying and monitoring liquidity conditions in housing markets should calculate and track idiosyncratic price dispersion, in addition to existing metrics such as time-on-market.

## 1.1 Related literature

This paper is related to several literatures. Most closely related is work studying idiosyncratic house price dispersion. Case and Shiller (1988) is one of the early papers to show that prices of individual houses are much more volatile than city-wide average prices. Giacoletti (2017), using California residential real estate data, and Sagi (2015), using commercial real estate data, show that idiosyncratic house price risk does not scale linearly with holding periods, suggesting that much of idiosyncratic house price risk is caused by market illiquidity at the time of sale. Several papers provide evidence

---

<sup>1</sup>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.

in support of this hypothesis. Giacoletti (2017) shows that contractions in local credit availability increase idiosyncratic risk. Ben-Shahar and Golan (2019) show that improved disclosure of transaction prices reduces price dispersion in the Israeli housing market. Landvoigt, Piazzesi and Schneider (2015) show that idiosyncratic variance increased in San Diego following the 2008 housing bust.

Our contribution to this literature is two-fold. First, we propose a novel strategy of measuring idiosyncratic price dispersion at the level of individual sales rather than returns. The advantage of this approach is that idiosyncratic price dispersion can then be flexibly aggregated cross-sectionally and over time. This measurement strategy allows us to document a set of new stylized facts about idiosyncratic price dispersion described in Section 3. Second, we construct a search-and-bargaining model of housing markets that provides microfoundation for the relationship between time-on-market and idiosyncratic price dispersion.

Our results also relate to the literature studying liquidity measurement in housing markets. Lippman and McCall (1986) is an early paper proposing time-on-market as a liquidity measure. Kluger and Miller (1990) proposes a Cox proportional hazard model to measure liquidity in housing markets. Lin and Vandell (2007) constructs a model to justify using aggregate time-on-market as a measure of housing market liquidity. Carrillo (2013), constructs a measure of housing market hotness using time-on-market data, as well as the sale-to-list price ratio. Our findings suggest 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. Hence, both metrics are important for assessing the performance of the US housing market.

Another strand of literature studies seasonal and cyclical co-movements in housing markets. Several papers document volume, prices, and time-on-market are correlated in housing markets; see, for example, Stein (1995), Krainer (2001*a*), Genesove and Mayer (2001), Leung, Lau and Leong (2002), Clayton, Miller and Peng (2010), Diaz and Jerez (2013), Ngai and Tenreyro (2014), and DeFusco, Nathanson and Zwick (2017). We contribute to this literature by showing that idiosyncratic price dispersion is also seasonal, cyclical and is correlated with time-on-market and other measures of market tightness. 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.

In terms of the modeling approach, our work fits into the literature on applying search-and-bargaining models to housing markets<sup>2</sup> and to financial markets more generally.<sup>3</sup> Our modeling approach is closely related to 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 and Bayer (2015) allows continuous match quality shocks and a discrete number of seller types. Sagi (2015) allows a discrete number of seller types. 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. The model, nevertheless, remains analytically tractable.

Finally, several studies provide empirical estimates of price versus time-on-market trade-off faced by individual sellers. See Genesove and Mayer (1997), Genesove and Mayer (2001), Levitt and Syverson (2008), Hendel, Nevo and Ortalo-Magné (2009), Guren (2018), Buchak et al. (2020) for estimates of non-foreclosure liquidity discounts, and Pennington-Cross (2006), Clauretje and Daneshvary (2009), Campbell, Giglio and Pathak (2011), Harding, Rosenblatt and Yao (2012), Zhou et al. (2015) for estimates of foreclosure discounts. We provide a comprehensive survey of this literature in Appendix 3. When calibrated to the average US data, our model is able to match estimates of liquidity discounts from this literature.

---

<sup>2</sup>See, for example, Wheaton (1990), Krainer (2001*b*), Albrecht et al. (2007), Novy-Marx (2009), 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).

<sup>3</sup>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).

## 1.2 Outline

The rest of the paper proceeds as follows. Section 2 describes our data and measurement strategy. Section 3 describes stylized facts about housing market liquidity, in the cross-section and time-series. Section 4 describes our model, theoretical results, and predictions. Section 5 contains tests of the model’s predictions. Section 6 contains robustness checks and alternative explanations of our findings. Section 7 contains our calibration. Section 8 discusses implications of our findings and concludes.

## 2 Data and measurement

### 2.1 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 to 2016. For time-on-market, we use the Corelogic MLS dataset. For demographic data on county-years and counties, as well as data on county-to-county migration, we use ACS 1-year and 5-year samples. For data on industry-specific wages and employment by county, we use Quarterly Census of Employment and Wages (QCEW). Further details of data sources and data cleaning steps are described in Appendix A.

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 A.1. Descriptive statistics for counties in our primary estimation sample are shown in Table 1. Our primary dataset comprises 11 million house sales within 472 counties. As Appendix Table A1 shows, our estimation sample contains 14.7% of all counties, but covers 61.7% 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.

We analyze two liquidity metrics: average time-on-market and idiosyncratic price dispersion. We define time-on-market as the difference between the closing date and the original listing date of sold houses. We describe our strategy of measuring idiosyncratic price dispersion below.



## 2.2 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:

$$p_{it} = \gamma_i + \eta_{ct} + f_c(x_i, t) + \epsilon_{it} \quad (1)$$

Where  $i$  indexes properties,  $c$  indexes counties, and  $t$  indexes months. In words, (1) says that log prices  $p_{it}$  are determined by a time-invariant house fixed effect,  $\gamma_i$ , a county-month fixed effect,  $\eta_{ct}$ , a smooth function  $f_c(x_i, t)$  of observable house characteristics  $x_i$  and time  $t$ , and a mean-0 error term  $\epsilon_{it}$ .

Specification (1) combines elements of repeat-sales and hedonic models of house prices. The house fixed effect term,  $\gamma_i$ , absorbs all features of a house, observed and unobserved, which have time-invariant effects on the price of house  $i$ . The  $\eta_{ct}$  term absorbs parallel shifts in log house prices in a county over time. The  $f_c(x_i, t)$  term allows houses with different observable characteristics  $x_i$  to appreciate at different rates: for example, the  $f_c(x_i, t)$  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 (1) are described in Appendix A.7.

We estimate price dispersion at the level of individual house sales using the estimated residuals,  $\hat{\epsilon}_{it}$ , from (1). 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 (1), so we will sometimes refer to these estimates as logSD. As we describe in Appendix A.7, we apply a degrees-of-freedom adjustment to  $\hat{\epsilon}_{it}$ , so the squared error estimates are unbiased at the county level.

## 3 Stylized facts

In this section, we demonstrate a number of stylized facts about price dispersion and time-on-market. The first set of facts concerns time-series patterns in our liquidity measures, sales volume, and prices.

**Fact 1.** *In the time series,*

- *Price dispersion and time-on-market are seasonal: both measures are lower in the summer hot season, and higher in the winter cold season of the housing market.*
- *Price dispersion and time-on-market are countercyclical: both measures decreased in the 2000-2005 housing boom, increased in the bust, and decreased in the recovery.*

Figure 1 shows the seasonal behavior of prices, total sales, time-on-market, and logSD, aggregated to the level of calendar months over the period 2000-2016. All four variables are seasonal: during summer, sales and prices are higher, and time-on-market and price dispersion are lower. On average, comparing June values to January values, summer prices are 2.30% higher, sales are 64.8% higher, time-on-market is 0.517 months (16.2%) lower, and price dispersion is 1.08% of house prices lower (6.43% in relative percentage points). Figure 2 analyzes this further by dividing counties into 3 quantile buckets, based on how seasonal prices are. This plot shows that more seasonal counties are more seasonal in all variables: that is, when seasonal price variation is larger, seasonal variation in sales, time-on-market, and price dispersion also tends to be larger.

Figure 3 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 16.5% in 2000 to 15.7% in 2004, rises to 18.2% in 2011, and falls to 17.1% in 2016.

Figure 4 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 1, 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.

In summary, Fact 1 says that, at both seasonal and business-cycle frequencies, time-on-market and price dispersion co-move with each other, and with volumes and prices, in intuitive ways: when markets are hotter, prices and volumes are higher, and time-on-market and price dispersion are lower. Moreover, these co-movements appear to be

driven by a single underlying factor, since counties which experience larger changes in one variable also experience larger changes in others. Thus, in the time series, price dispersion and time-on-market appear to measure market hotness, in a manner similar to price or volume increases. Next, we analyze the behavior of these liquidity measures in the cross-section of counties.

**Fact 2.** *In the cross-section of US counties,*

- *There is substantial variation in idiosyncratic price dispersion and time-on-market.*
- *Price dispersion and time-on-market are positively correlated, but there is substantial independent variation.*
- *There is substantial variation in price dispersion and time-on-market for counties with similar house prices.*

The observations in Fact 2 are based on Figure 5, in which we plot the cross-section of time-on-market and idiosyncratic price dispersion,  $\hat{\sigma}_c$ , across counties. There is substantial dispersion in both liquidity measures. The mean of  $\hat{\sigma}_c$  is 18.3% of house prices, and the standard deviation is 3.65%. The 10th percentile is 14.0% and the 90th percentile is 22.7%. For time-on-market, the mean is 3.24 months, the standard deviation is 0.755, the 10th percentile is 2.40, and the 90th percentile is 4.20.

Moreover, in the cross-section of counties, time-on-market and price dispersion are not well explained by each other, or by the level of house prices. Time-on-market and price dispersion are positively correlated, but the  $R^2$  from a univariate regression is only 0.199. Both measure are correlated with prices: high-price counties tend to be in the lower left quadrant of Figure 5, so high-price counties tend to have lower price dispersion and time-on-market. However, there is substantial variation in time-on-market and price dispersion which is not explained by average prices: the  $R^2$  values from regressing time-on-market and price dispersion on mean prices, respectively, are 0.0511 and 0.233.

These observations suggest that, in the cross-section of counties, “liquidity”, as measured by time-on-market and price dispersion, appears to vary significantly across counties with similar average prices. Moreover, the fact that time-on-market and price dispersion have substantial independent variation suggests that housing market liquidity is, in some sense, a multi-dimensional object, which time-on-market and price dispersion measure different aspects of. To better understand these facts, we proceed to construct a search-and-bargaining model of the housing market.

## 4 Model

To rationalize the stylized facts described in Section 3, we build an analytically 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. Sellers, therefore, face a trade-off between selling quicker or selling at a higher price. We use this model to study how the supply of liquidity, measured by the buyers' inflow rate, which results in a higher buyers-to-sellers ratio, and the demand for liquidity, measured by average holding costs of sellers, determine time-on-market and price dispersion in equilibrium.

### 4.1 Setup

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. Once a seller successfully sells her house, she permanently leaves the market, attaining a continuation value which we normalize to 0. Each seller has 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 unmatched from her house and becomes a seller. 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. Let  $F_{eq}(c)$  denote the distribution of holding utilities among sellers in stationary equilibrium, which will in general differ from  $F(c)$ .

Sellers with higher holding costs  $c$  will tend to sell faster, but at lower prices. There are many possible sources of differences in  $c$ . Since selling houses involves significant time and effort for owners, sellers with higher incomes may have higher values of  $c$ . 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 (Anenberg and Bayer (2015)). 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)). Our model is agnostic with respect to the exact source of this heterogeneity.

**Buyers.** Potential buyers enter the market at some exogeneous flow rate  $\eta_B$ , and each buyer draws a value  $\xi \sim H(\cdot)$  for entering the city.  $\xi$  can be thought of as representing the attractiveness of amenities and job opportunities in the county, which may vary idiosyncratically across buyers. After observing  $\xi$ , each potential buyer can choose to either enter the city, receiving utility  $\xi$  immediately and becoming an active homebuyer, or leave forever, receiving utility normalized to 0. Entry decisions are irreversible. We use  $V_B$  to denote the expected value of an unmatched buyer in stationary equilibrium. Buyers will only enter if their expected utility from entry is positive, that is:

$$\xi + V_B > 0 \quad (2)$$

Hence, the inflow rate of buyers in stationary equilibrium is:

$$\eta_B (1 - H(-V_B))$$

These assumptions imply that the entry rate of buyers responds to market conditions: the inflow rate of buyers increases when buyers' expected value  $V_B$  is high. Similar assumptions are made in a number of other papers (Novy-Marx (2009), Head, Lloyd-Ellis and Sun (2014)).

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 matching 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  $\epsilon$  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  $\epsilon_0$ , and standard deviation  $\sigma_\epsilon$ .

Our baseline model assumes buyers, after they have entered the market, are undifferentiated; in Appendix B.7, we extend the model to include persistent buyer heterogeneity, and show that the main theoretical results are unchanged.

**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 type  $\epsilon$  receives flow utility  $\epsilon$  from their house per unit time. At Poisson rate  $\lambda_M$ , homeowners receive separation shocks: they draw 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, which will in general differ from  $G(\epsilon)$ .

**Price determination.** 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  $\epsilon$ . 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  $\epsilon$  and  $c$ , is thus:

$$V_M(\epsilon) - V_B - V_S(c) \quad (3)$$

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  $\epsilon$  higher than some *trade cutoff*  $\epsilon^*(c)$ , which satisfies:

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

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

$$P(c, \epsilon) = V_S(c) + \theta(V_M(\epsilon) - V_B - V_S(c)) \quad (4)$$

We assume that  $\epsilon_0$  is sufficiently low that, in equilibrium,

$$\epsilon^*(c) \geq \epsilon_0 \quad \forall c$$

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

We note that, in the model, the only drivers of price dispersion are sellers' and buyers' preferences. There are other possible drivers of price dispersion in practice. For example, realtor quality affects time-on-market and prices (Gilbukh and Goldsmith-Pinkham (2019)). We abstract away from these other factors for simplicity; they can be thought of as adding additional error terms to the price equation (4) in the model.

**Match formation.** Matches between buyers and sellers are generated at a flow rate  $m(M_B, M_S)$ , which depends on the masses of buyers  $M_B$  and sellers  $M_S$  present in the market. We assume that  $m(M_B, M_S)$  is Cobb-Douglas with constant returns to scale:

$$m(M_B, M_S) = \alpha M_B^\phi M_S^{1-\phi}$$

From the perspective of any given buyer or seller, matching occurs at Poisson rates  $\lambda_B$  and  $\lambda_S$ , given by:

$$\lambda_B = \frac{m(M_B, M_S)}{M_B}, \lambda_S = \frac{m(M_B, M_S)}{M_S}$$

## 4.2 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 states the equilibrium conditions. Formal derivations of these conditions are in Appendix B.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  $\lambda_S, \lambda_B$ , 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:*

$$rV_B = \lambda_B \int \int_{\epsilon > \epsilon^*(c)} [(1 - \theta)(V_M(\epsilon) - V_B - V_S(c))] dG(\epsilon) dF_{eq}(c) \quad (5)$$

$$rV_S(c) = -c + \lambda_S \int_{\epsilon > \epsilon^*(c)} \theta(V_M(\epsilon) - V_B - V_S(c)) dG(\epsilon) \quad (6)$$

$$rV_M(\epsilon) = \epsilon + \lambda_M \left( \int V_S(c) dF(c) - V_M(\epsilon) \right) \quad (7)$$

Trade cutoffs:

$$V_M(\epsilon^*(c)) = V_S(c) + V_B \quad (8)$$

Matching rates:

$$M_S \lambda_S = M_B \lambda_B = \alpha M_B^\phi M_S^{1-\phi} \quad (9)$$

Flow equality:

$$(1 - M_S) \lambda_M f(c) = \lambda_S M_S f_{eq}(c) (1 - G(\epsilon^*(c))) \quad (10)$$

$$G_{eq}(\epsilon) = \frac{\int_c \lambda_S M_S \left[ \int_{\tilde{\epsilon}=\epsilon_0}^{\epsilon} 1(\tilde{\epsilon} > \epsilon^*(c)) dG(\tilde{\epsilon}) \right] dF_{eq}(c)}{\int_c \lambda_S M_S (1 - G(\epsilon^*(c))) dF_{eq}(c)} \quad (11)$$

$$(1 - M_S) \lambda_M = \eta_B (1 - H(-V_B)) \quad (12)$$

Our model admits simple expressions for two liquidity measures: time-on-market and price dispersion.

*Claim 1.* In stationary equilibrium, expected time-on-market for a seller of type  $c$  is:

$$TOM(c) = \frac{1}{\lambda_S (1 - G(\epsilon^*(c)))} \quad (13)$$

Expected time-on-market across all seller types is thus:

$$TOM = E \left[ \frac{1}{\lambda_S (1 - G(\epsilon^*(c)))} \right] \quad (14)$$

the variance of  $P(c, \epsilon)$  among trading sellers and buyers is:

$$\text{Var}(P(c, \epsilon)) = \underbrace{\text{Var}_{c \sim F(\cdot)}(V_S(c))}_{\text{Seller holding costs}} + \underbrace{\left( \frac{\theta \sigma_\epsilon}{r + \lambda_M} \right)^2}_{\text{Buyer match utility}} \quad (15)$$

Expression (15) shows that  $TOM(c)$ , the equilibrium time-on-market of seller of type  $c$  depends on the rate at which the seller meets buyers,  $\lambda_S$ , and the probability of trade conditional on a match,  $1 - G(\epsilon^*(c))$ .  $TOM(c)$  is also decreasing: sellers with higher urgency  $c$  always tend to sell faster. Thus, two forces can change time-on-market. Time-on-market will be lower when markets are thicker,  $\lambda_S$  is higher, and all sellers meet buyers faster. Time-on-market will also be lower if sellers' average holding costs  $c$  are high, since the buyer cutoffs  $\epsilon^*(c)$  will decrease, and sellers will sell faster by being less selective.

Expression (15) shows that equilibrium price dispersion arises from two sources:



differences in sellers' value functions  $V_S(c)$ , caused by differences in sellers' holding costs; and differences in buyers' match utilities  $\epsilon$ . When market thickness changes, in our model, the buyer match utility term stays constant, but the seller cost term changes. The following claim helps illustrate the effects of market thickness on the seller cost term. First, define:

$$P(c) \equiv E[P(\epsilon, c) | c]$$

as the expected sale price attained by a seller of type  $c$ .

*Claim 2.* We have:

$$P'(c) = V'_S(c) = \frac{-\text{TOM}(c)}{r\text{TOM}(c) + \theta} \quad (16)$$

where  $\text{TOM}(c)$  is expected time-on-market for a seller of type  $c$ .

Sellers face a tradeoff between prices and time-on-market. Urgent sellers, with high values of  $c$ , sell faster, but get lower average prices as a result. The derivative  $P'(c)$  measures how much a seller's average sale price would decrease if her holding cost  $c$  increased by a small amount. Claim 2 shows that  $P'(c)$  depends on average time-on-market. Intuitively, if markets are thick and all sellers sell quickly, sellers will also sell at similar prices, so  $P'(c)$  is low. When markets are thin and time-on-market is high for all sellers,  $P'(c)$  is high, and differences in  $c$  will translate into larger price dispersion. In fact, we show in the following claim that if time-on-market  $\text{TOM}(c)$  increases, point-wise for every  $c$ , then equilibrium price dispersion also increases.

*Claim 3.* Fix  $r, \theta, F(c), \sigma_\epsilon^2$ . Consider two sets of model parameters,

$$\Theta_1 = (\alpha^1, \phi^1, \lambda_M^1, \eta_B^1, \epsilon_0^1), \Theta_2 = (\alpha^2, \phi^2, \lambda_M^2, \eta_B^2, \epsilon_0^2)$$

such that time-on-market is uniformly higher in stationary equilibrium under  $\Theta_1$ ; that is, letting  $\text{TOM}_{\Theta_1}(c)$  denote the equilibrium time-on-market function under  $\Theta_1$ ,

$$\text{TOM}_{\Theta_1}(c) > \text{TOM}_{\Theta_2}(c) \quad \forall c$$

Then equilibrium price dispersion will also be higher under  $\Theta_1$  than  $\Theta_2$ .

### 4.3 Comparative statics

To illustrate how our model maps primitives to outcomes, we solve the model computationally, and show comparative statics with respect to buyers inflow rate  $\eta_B$  and sellers

average holding costs  $\bar{c}$ .<sup>4</sup> Specifically, Figure 6 shows average prices,  $E(P)$ , average time-on-market, TOM, and price dispersion in levels,  $\text{Var}(P)$ , and relative price dispersion,  $\text{LogSD}(P)$ .<sup>5</sup>  $\text{LogSD}(P)$  is the model counterpart to our empirical measure of idiosyncratic price dispersion.

The left panel shows how outcomes vary as we change the buyer inflow rate,  $\eta_B$ . As  $\eta_B$  increases, the market is tighter, so sellers are able to sell quicker and at higher prices. Since time-on-market decreases, price dispersion in levels,  $\text{Var}(P)$ , also decreases. As a result,  $\text{LogSD}(P)$  unambiguously decreases, because the numerator, which depends on  $\text{Var}(P)$ , decreases, and the denominator,  $E(P)$ , increases.

The right panel demonstrates what happens when we change sellers' average holding costs,  $\bar{c}$ . As  $\bar{c}$  increases, sellers' costs of staying on the market increase, so they sell faster and at lower prices. Thus, by Claim 2, price dispersion in levels,  $\text{Var}(P)$ , also decreases. However,  $\text{LogSD}(P)$  still increases, because the average price  $E(P)$  falls at a faster rate than  $\text{Var}(P)$  in the nominator.

Thus, there are two main channels through which market liquidity affects relative price dispersion,  $\text{LogSD}(P)$ . The first channel is through dispersion in sellers' value functions,  $V_S(c)$ . Any change in liquidity supply or demand that results in higher  $\text{TOM}(c)$ , also leads to higher dispersion in  $V_S(c)$  and, therefore, higher  $\text{Var}(P)$ . The second channel is through changes in the average price  $E(P)$ . As the average price increases, the relative price dispersion  $\text{LogSD}(P)$  decreases.

#### 4.4 The supply and demand for liquidity

In Figure 7, we simulate our model. The top two panels show the time-on-market and expected price menu generated by individual sellers with different holding costs. We normalize expected prices such that the expected price of a seller with median holding cost is 1. Each line corresponds to one equilibrium. The dots represent the time-on-market and expected price pairs chosen by sellers with 0th, 30th, 60th, and 100th percentile holding costs  $c$ .

In our model, there are two forces that drive liquidity measures, which we call *liquidity supply* and *liquidity demand*. *Liquidity supply* is  $\eta_B$ , the inflow rate of buyers. Intuitively,

---

<sup>4</sup>We assume  $c \sim U[\bar{c} - \Delta_c, \bar{c} + \Delta_c]$ ,  $\epsilon_0 = 4$ ,  $\sigma_\epsilon = 0.33$ ,  $\Delta_c = 2$ ,  $\lambda_m = 0.1$ ,  $r = 0.05$ ,  $\alpha = 1$ ,  $\phi = 0.84$ ,  $\theta = 0.5$ ,  $\xi \sim N(0, 40)$ .

<sup>5</sup>For ease of computation, we approximate  $\text{LogSD}(P) \approx \frac{\sqrt{\text{Var}(P)}}{E(P)}$ .

when there are many buyers, the supply of liquidity to sellers is high. In equilibrium, 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 liquidity supply increases, both time-on-market and price dispersion decrease.

*Liquidity demand* is  $\bar{c}$ , the average urgency of sellers to sell. When  $\bar{c}$  increases, time-on-market decreases, as sellers try to sell faster, but price dispersion increases: 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 relative price dispersion increases.

Both effects are summarized in the bottom panel of Figure 7. This plot shows the two aggregate liquidity measures: 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  $\eta_B$ , and shows outcomes as we vary  $\bar{c}$ . We see that, holding fixed liquidity supply  $\eta_B$ , varying liquidity demand,  $\bar{c}$ , causes time-on-market and price dispersion to co-move negatively. Holding fixed liquidity demand  $\bar{c}$ , increasing liquidity supply  $\eta_B$  causes the entire frontier to shift: the market is able to attain both lower time-on-market and lower price dispersion.

The model sheds light on the stylized facts we documented in the previous section. Time-on-market and price dispersion can shift due to either liquidity supply or demand. In the time series, at both seasonal and business-cycle frequencies, both liquidity measures co-move positively with each other, and negatively with prices and volume. Through the lens of our model, this is consistent with time-series movements being driven primarily by liquidity supply shifts. Liquidity supply is higher in boom periods, and in the summer hot season, so prices and volume are higher, and time-on-market and price dispersion are both lower.

In contrast, to explain the independent variation in time-on-market and price dispersion in the cross-section of counties, both liquidity supply and demand must play a role. Counties towards the bottom left of Figure 5 likely have high liquidity supply, leading both time-on-market and price dispersion to be low; similarly, counties towards the top right have low liquidity supply. Counties towards the top left have low time-on-market, but high price dispersion: our model rationalizes this by saying *liquidity demand* is high. Sellers in these areas have high holding costs, so they choose to sell quickly, even though this leads to lower and less stable prices. Similarly, counties towards the bottom right

have low liquidity demand: sellers are willing to wait longer to sell at higher and more stable prices.

The idea that time-on-market and price dispersion are jointly determined by the supply and demand for liquidity generates two testable predictions to bring to the data: proxies for liquidity supply and liquidity demand should have different correlations with the two liquidity measures.

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

**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.*

## 5 Empirical results

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

$$\log SD_c = Z_c^S \alpha_1 + X_c \beta_1 + \epsilon_c \quad (17)$$

$$TOM_c = Z_c^S \alpha_2 + X_c \beta_2 + \epsilon_c \quad (18)$$

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, which account for the demographic composition of the county as well as characteristics of its housing stock: the average age of houses, average square footage, average bedroom and bathroom counts of sold 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.

Ideally, proxies for liquidity supply should be variables associated with net buying pressure: liquidity supply should be high in cities where there are many buyers, relative to the available housing stock. We use two proxies for liquidity supply. The first is the population growth rate in a county: high population growth should imply high inflow rates of buyers, which implies that housing markets should be tight. We show results from

these regressions in Table 2. Columns 1 and 3 show results from regressing logSD and time-on-market on population growth rates. Both coefficients are negative, supporting Prediction 1. Columns 2 and 4 add state fixed effects: both coefficients remain negative, though the time-on-market coefficient loses significance.

Finally, column 5 regresses logSD on time-on-market and controls, and column 6 regresses logSD on time-on-market, controls, and state fixed effects. Time-on-market and logSD are significantly positively associated in the cross-section in both specifications, suggesting that the positive association in Figure 5 survives controlling for various observables.

Migration rates are affected by a large number of variables, so they may not be an exogenous shifter of liquidity supply. We use an empirical approach based on Schubert (2021) to construct a plausibly exogenous shock to liquidity supply: *migration spillovers* from high-productivity areas. Intuitively, suppose, for example, Cook County experiences large productivity shocks, which increase house prices. This will tend to create outmigration from Cook County, which will create net immigration flows to counties with strong historical migration links to Cook County. These migration flows are thus plausibly exogenous shocks to housing market tightness in counties connected to Cook County by migration.

Formally, we construct plausibly exogenous local wage “Bartik” shocks  $B_{c,2012-2016}$  by combining local employment shares  $\omega_{c,i,2010}$  of workers in 2-digit NAICS industries indexed by  $i$  in county  $c$  in 2010 with national wage growth rate  $\Delta \ln W_{-c,i,2012-2016}$  in that industry over the period of interest 2012-2016.<sup>6</sup>

$$B_{c,2012-2016} = \sum_i \omega_{c,i,2010} \Delta \ln W_{-c,i,2012-2016} \quad (19)$$

We measure migration exposure  $\mu_{2008-2012}^{c \leftarrow c'}$  of county  $c$  to county  $c'$  economic shocks as a fraction of migrants to county  $c$  that came from county  $c'$  using ACS 2008-2012 5-year sample.<sup>7</sup> We then define our liquidity supply shifter as:

---

<sup>6</sup>We fix industry exposure shares at their 2010 level to reduce bias from endogeneity in the local industry exposure. We also compute national wage growth as leave-one-out measures to avoid mechanical correlation between the national trend estimate and county  $c$  wages.

<sup>7</sup>We calculate migration shares using ACS 2008-2012 5-year data sample to avoid concerns over endogenous changes in the migration weights. To minimize measurement error, we exclude such counties  $c'$  such that fewer than 150 people migrated to  $c$  from  $c'$ .

$$M_c = \sum_{c'} \mu_{2008-2012}^{c \leftarrow c'} B_{c', 2012-2016} \quad (20)$$

One concern for our identification assumptions is that, if Cook County and, for example, Orange County have large historical migration flows, they may also have similar industrial composition. This could cause our exclusion restriction to be violated, because productivity shocks in Cook County and Orange County could be correlated. To address this concern, we follow Chodorow-Reich and Wieland (2020) and include Orange County's own productivity shock,  $B_{c, 2012-2016}$ , as a control variable in the regression. Effectively, identification is then coming from variation in Cook County's productivity shocks that is orthogonal to productivity in Orange County.

In Table 3, we show estimates of specifications (17) and (18), where we include  $M_c$  as the instrumental variable, and control for  $B_{c, 2012-2016}$ . Columns 1 and 4 show results for logSD and TOM respectively on the instrument, without any controls: both coefficients are negative and significant, consistent with our theory's predictions. Columns 2 and 5 add demographic controls: coefficients remain negative and significant. Columns 3 and 6 add state fixed effects, and both coefficients lose significance, perhaps because the instrument does not have sufficient power within states. Together, the specifications in Table 3 lend additional support to Prediction 1, that increases in liquidity supply are associated with decreases in both time-on-market and price dispersion.

**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.

$$\log SD_c = Z_c^D \alpha_1 + X_c \beta_1 + \epsilon_c \quad (21)$$

$$TOM_c = Z_c^D \alpha_2 + X_c \beta_2 + \epsilon_c \quad (22)$$

We control for the same variables as in (17) and (18). Our liquidity demand shifter is household income, controlling for house prices. The reasoning behind this proxy is that income measures households' value of time. Higher income households should have a higher value of time, so they should have larger opportunity costs for keeping their houses on the market. Hence, higher income households should be willing to sell their houses faster, even if they have to sell for lower prices. Counties with high income-to-price ratios should thus have high liquidity demand: in the aggregate, home sellers in these

counties should shift along the liquidity supply curve, selling faster, despite incurring lower and less stable prices as a result.

Columns 1 and 2 of Table 4 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. Similarly, 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.

## 5.1 Alternative explanations

Here, we discuss some possible alternative explanations of our empirical results.

**Liquidity supply.** Several papers argue that 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, several studies suggest that dispersion in market participants' beliefs (Glaeser and Nathanson (2017), Nathanson and Zwick (2018)) and the market share of short-term speculators (Bayer et al. (2011), DeFusco, Nathanson and Zwick (2017)) both increase during housing booms. 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 Stroebe (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 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 liquidity statements deal with price *dispersion*, not price *levels*. 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 disconnected from fundamentals. Our results show that, while booms may well involve irrationality and a departure of price levels from fundamentals, booms also improve liquidity and stabilize relative prices.

**Liquidity demand.** 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 it is indeed the case (Kurlat and Stroebel (2015), Stroebel (2016)). This mechanism could affect both 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 are 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.

Another explanation for the positive association between income and price dispersion is that high-income counties could have 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 inflow rate,  $\eta_B$ , could be lower when houses are more heterogeneous, since any individual house has fewer interested buyers. Second, the variance of buyer values,  $\sigma_\epsilon$ , could be larger, if buyers have more specific preferences over different house characteristics.

However, comparative statics results in Appendix B.8 show that, within our model, both of these parameter changes would increase price dispersion as well as time-on-market. Intuitively, decreasing  $\eta_B$  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.

Intuitively, a large class of theories can explain positive co-movement in time-on-



market and price dispersion, but it is harder to explain the negative co-movement we find. Our model rationalizes this by saying that, when sellers have higher values of time, they rationally trade off “dollar liquidity” for increased “time liquidity”, by choosing to sell faster but at less stable prices. We are not aware of other theoretical forces that would move time-on-market and price dispersion in opposite directions.

## **6 Robustness checks**

### **6.1 Panel regressions**

Most of our cross-sectional results also hold in panel regressions. In Appendix Tables A2 and A3, we run panel regressions of time-on-market and price dispersion on the same liquidity supply and demand shifters as in Tables 2 and 4. The results from panel regressions are mostly analogous to the cross-sectional results.

### **6.2 Measurement concerns**

Our results are robust to several different ways to measure price dispersion and time-on-market. In Appendix C.2, 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. The results are shown in Appendix Figures A2 to A4, and Appendix Tables A4 to A7. Appendix C.3 discusses how our estimation methodology for price dispersion relates to other papers in the literature on idiosyncratic house price dispersion, and demonstrates that our measures of price dispersion are in line with the literature. In Appendix C.4, we consider time-on-market measures from Zillow and Realtor.com, and results are shown in Appendix Figures A5 and A6, and Appendix Tables A8 to A10. In all cases, results are qualitatively unchanged.

Conceptually, specification (1) 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 (1) includes both house fixed effects and time-varying effects of observable characteristics, (1) 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 effects of house characteristics on prices which specification (1) cannot capture. First, our data only allow us to observe characteristics at a single point in time, so specification (1) 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 (1) absorbs time-invariant effects of unobservables into the house fixed effects,  $\gamma_i$ , (1) 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 (1).

Both effects are likely to be quantitatively small. First, Giacoletti (2017) observes data on remodeling expenditures for houses in California. Accounting for remodelling decreases the estimated standard deviation of returns by only around 2% of house prices. Second, in Appendix C.2, 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.

## 7 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.

### 7.1 Methodology

**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  $\phi$  to be 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  $\alpha$  and the inflow of buyers  $\eta_B$ , 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  $\lambda_M$ , the buyer entry rate  $\eta_B$ , the parameters of seller holding cost distribution, which we assume to be  $U[\bar{c} - \Delta_c, \bar{c} + \Delta_c]$ , and the parameters of the buyer match value exponential 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 5 shows the values of empirical moments that we target.

Next, we choose  $\Delta_c$  to match the empirical relationship between time-on-market and price dispersion. Claim 2 in Section 4 shows that time-on-market determines the extent to which dispersion in sellers’ holding costs,  $\Delta_c$ , translates into dispersion in sellers’ continuation values,  $\text{Var}(V_S(c))$ , which in turn translates into dispersion in prices,  $\text{Var}(P)$ . Hence, the empirical relationship between time-on-market and price dispersion is informative about  $\Delta_c$ . We use the coefficient from the panel regression of price dispersion ( $\log\text{SD}$ ) on time-on-market controlling for average prices to calibrate  $\Delta_c$ . Controlling for the direct effect of prices on  $\log\text{SD}$  allows us to isolate the effect of time-on-market on dispersion in sellers’ values. Since time-series changes in prices,  $\log\text{SD}$ , and time-on-market are likely to be driven by changes in liquidity supply, we perturb the liquidity supply parameter in our model,  $\eta_B$ , around the equilibrium to generate the model-implied relationship between time-on-market and price dispersion. Then, we calibrate  $\Delta_c$  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 D.1.

Table 5 shows our estimated parameter values that we estimate. We estimate that match quality  $\epsilon$  has a lower bound of \$564 monthly, or \$6,772 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  $\epsilon$  among successful buyers is \$1,534 monthly, and \$18,408 annually. We assumed that sellers’ values are uniformly

distributed on  $[\bar{c} + \Delta_c, \bar{c} - \Delta_c]$ . We estimate that the mean  $\bar{c}$  is equal to \$6,458 per month, and the range parameter  $\Delta_c$  is equal to \$4,208. We then calculate that buyers' average value from homeownership as the integrated flow value of match utility  $\epsilon$ , until buyers receive a separation shock, to be equal to \$176,271. Similarly, we calculate that sellers' average total loss from keeping their houses on the market is equal to \$14,349.

## 7.2 Results

Using the estimated parameter values for our model, we 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. Figure 8 shows the “menu” of time-on-markets and average prices attained by sellers with different values.

Sellers with 75th percentile holding cost  $c$  spend on average 3.13 months on the market, and attain expected sale prices of \$271,957. Sellers with 25th percentile holding cost spend 1.79 months, and attain expected sale prices of \$252,684. That is, 75th percentile sellers take 1.34 more months to sell, and achieve \$19,273 higher prices – in percentage terms, 7% higher prices. The implied effect of spending an extra month on the market is, therefore, 5% higher prices.

These estimates are useful, because the tradeoff between time-on-market and average sale prices has been extensively studied in the housing literature. In Table 6, we survey a number of estimates of the magnitude of this tradeoff from the housing literature. We divide papers into two groups, based on whether estimates use foreclosed houses or not, since price discounts in the foreclosure literature are systematically higher than estimates from other papers. The non-foreclosure shed light on tradeoffs between prices and time-on-market using various variables that shift sellers' urgency to sell: owners' equity position (Genesove and Mayer (1997), Guren (2018)), nominal losses (Genesove and Mayer (2001)), whether the homeowner is a realtor (Levitt and Syverson (2008)), whether the house is FSBO (Hendel, Nevo and Ortalo-Magné (2009)), and whether the seller uses an I-buyer (Buchak et al. (2020)). These papers find effects with consistent signs: forces that lead sellers to sell faster also lead to lower average sale prices. We can thus calculate liquidity discounts from each of these papers, by dividing the estimated price effect by the time-on-market effect, and scaling these estimates so that they represent the implied percentage price increase from spending an extra month on the market. Further details on how we arrive at these estimates are described in Appendix D.3.

The main finding from Table 6 is that the non-foreclosure estimates of liquidity discounts are mostly within the range of 1.9% to 11%.<sup>8</sup> That is, the literature estimates that, if a seller spent an extra month on the market, she would attain a sale price around 1.9% to 11% higher on average. Our estimated effect, of 5%, is right in the middle of this range. We note that we did 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.

Our calibration thus shows that our model can simultaneously rationalize the relationship between aggregate price dispersion and time-on-market across counties and over time, and micro-estimates of the tradeoff between time-on-market and average prices faced by individual house sellers. Thus, besides its qualitative role in rationalizing the behavior of time-on-market and price dispersion in the cross-section of counties, our model appears to be sufficiently realistic that it can be used for quantitatively studying the link between aggregate market liquidity measures, and what they imply about the decisions facing individual home sellers.

## 8 Conclusion

In this paper, we have constructed a rich panel tracking time-on-market and price dispersion across US counties over time. In the time series, at both seasonal and business cycle frequencies, time-on-market and price dispersion co-move closely. However, in the cross-section of counties, there is substantial independent variation: time-on-market and price dispersion are not well correlated with each other, or with the level of house prices.

We constructed a search-and-bargaining model to rationalize these findings. Time-on-market and price dispersion 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. We find support for the model's predictions 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.

---

<sup>8</sup>There is one outlier: the upper bound of the estimate from Genesove and Mayer (2001) is 24%.

Together, our findings suggest that time-on-market alone is not sufficient for measuring housing market liquidity: academics and policymakers who are interested in studying and monitoring housing market liquidity should calculate and track idiosyncratic price dispersion alongside time-on-market, as the two metrics contain partially independent information.

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,<sup>9</sup> but in some markets explicit auctions are used.<sup>10</sup> 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, Chincio 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. Future work could analyze how the composition of housing market participants affects liquidity measures in housing markets.

---

<sup>9</sup>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.

<sup>10</sup>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.** 2013. "To sell or not to sell: Measuring the heat of the housing market." *Real estate economics*, 41(2): 310–346.
- 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.
- Chodorow-Reich, Gabriel, and Johannes Wieland.** 2020. "Secular labor reallocation and business cycles." *Journal of Political Economy*, 128(6): 2245–2287.
- 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.



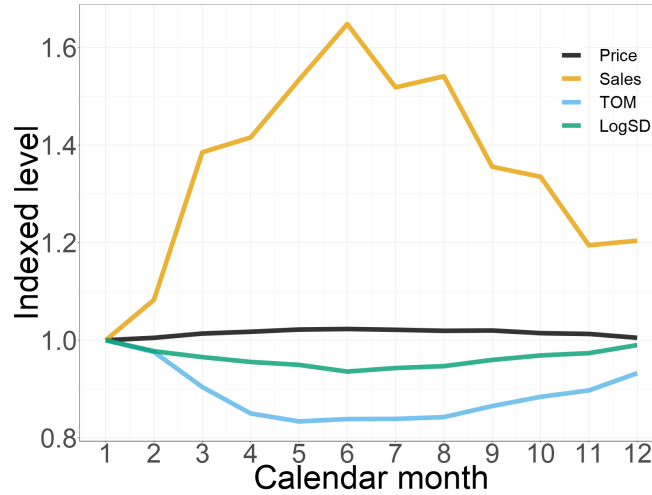
- Gabrovski, Miroslav, and Victor Ortego-Marti.** 2019a. "The cyclical behavior of the Beveridge Curve in the housing market." *Journal of Economic Theory*, 181: 361–381.
- Gabrovski, Miroslav, and Victor Ortego-Marti.** 2019b. "Search and Credit Frictions in the Housing Market." 27.
- Gavazza, Alessandro.** 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.
- Giaconetti, Marco.** 2017. "Idiosyncratic Risk in Housing Markets." Social Science Research Network SSRN Scholarly Paper ID 2995323, Rochester, NY.
- Giaconetti, 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.
- Kluger, Brian D, and Norman G Miller.** 1990. "Measuring residential real estate liquidity." *Real Estate Economics*, 18(2): 145–159.
- Krainer, John.** 2001a. "A Theory of Liquidity in Residential Real Estate Markets." *Journal of Urban Economics*, 49(1): 32 – 53.
- Krainer, John.** 2001b. "A theory of liquidity in residential real estate markets." *Journal of urban Economics*, 49(1): 32–53.
- Kurlat, Pablo, and Johannes Stroebe.** 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.
- Lin, Zhenguo, and Kerry D Vandell.** 2007. "Illiquidity and pricing biases in the real estate market." *Real Estate Economics*, 35(3): 291–330.
- Lippman, Steven A, and John J McCall.** 1986. "An operational measure of liquidity." *The American Economic Review*, 76(1): 43–55.
- 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.
- Novy-Marx, Robert.** 2009. "Hot and cold markets." *Real Estate Economics*, 37(1): 1–22.
- 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 Stroebe.** 2015. "Segmented Housing Search." National Bureau of Economic Research Working Paper 20823.
- Sagi, Jacob S.** 2015. "Asset-Level Risk and Return in Real Estate Investments." *SSRN Electronic Journal*.

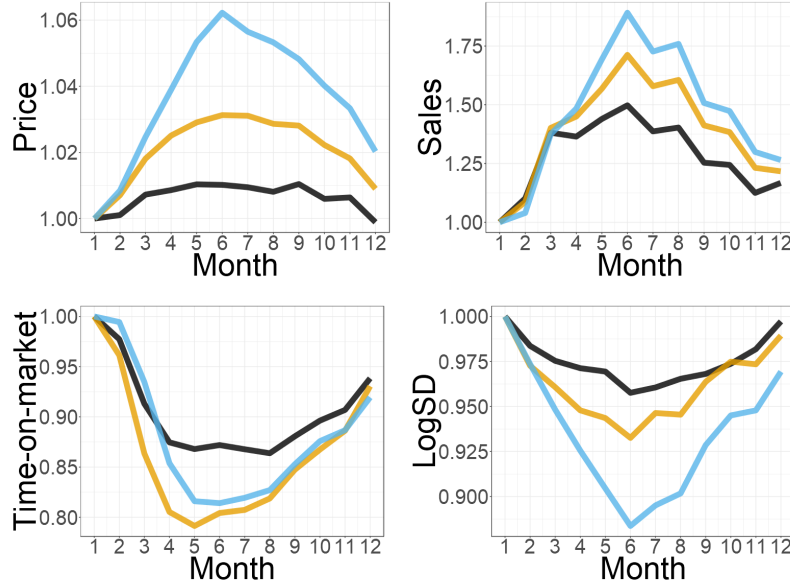
- Schubert, Gregor.** 2021. "House Price Contagion and U.S. City Migration Networks." Working Paper.
- Stein, Jeremy C.** 1995. "Prices and Trading Volume in the Housing Market: A Model with Down-Payment Effects." *The Quarterly Journal of Economics*, 110(2): 379–406.
- Stroebel, Johannes.** 2016. "Asymmetric 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.

Figure 1: Seasonal variation in sales, prices, logSD, and time-on-market



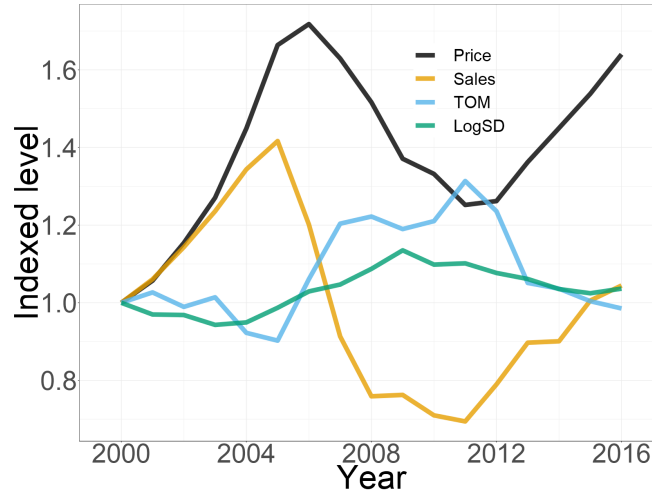
*Notes.* Total sales, average prices, logSD, and time-on-market (TOM) by calendar month. The time period of the data is 2000 to 2016. All variables are indexed by dividing by their January level. LogSD, our measure of idiosyncratic price dispersion, is calculated according to specification (1). TOM is from the Corelogic MLS data. Sales is calculated using the Corelogic data. Price comes from a repeat-sales monthly price index: we regress log sale prices on county-month and house fixed effects, and take the county-month fixed effects as a price index. For all variables, we filter out low-frequency trends by fitting a piece-wise linear trend with break points every 3 years, subtracting away the predicted values, and adding back the mean. The price, TOM, and LogSD lines are constructed as sales-weighted averages across counties to the calendar-month level, and then indexed so that the series is equal to 1 in January. Further details of data construction are described in Appendix A.5.

Figure 2: Seasonal variation in sales, prices, logSD, and time-on-market, heterogeneity



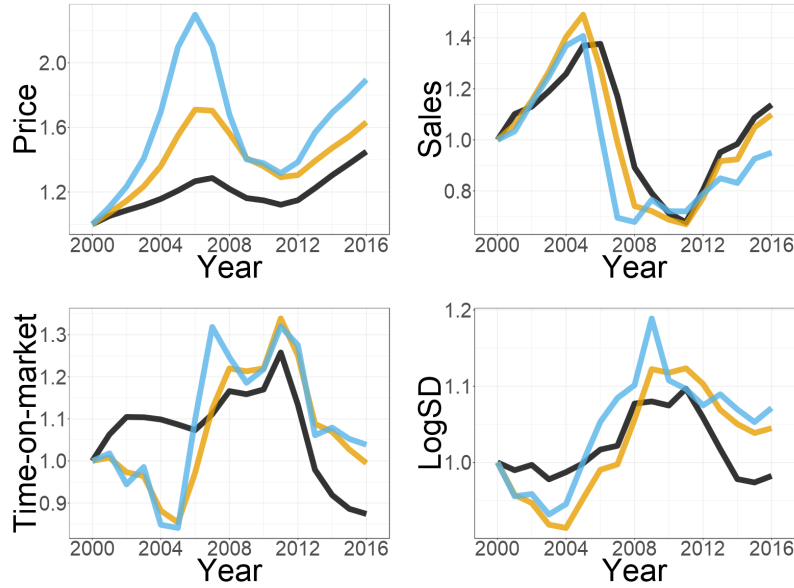
*Notes.* Prices, sales, time-on-market, and logSD for three quantile buckets of counties, divided based on the ratio between summer and winter prices. Each color represents one quantile bucket of counties. The time period of the data is 2000 to 2016. All variables are indexed by dividing by their January level. LogSD, our measure of idiosyncratic price dispersion, is calculated according to specification (1). TOM is from the Corelogic MLS data. Sales is calculated using the Corelogic data. Price comes from a repeat-sales monthly price index: we regress log sale prices on county-month and house fixed effects, and take the county-month fixed effects as a price index. For all variables, we filter out low-frequency trends by fitting a piece-wise linear trend with break points every 3 years, subtracting away the predicted values, and adding back the mean. The price, TOM, and LogSD lines are constructed as sales-weighted averages across counties to the calendar-month level, and then all series are indexed to equal 1 in January. Further details of data construction are described in Appendix A.5.

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



*Notes.* Yearly prices, sales, time-on-market, and logSD. The time period of the data is 2000 to 2016. LogSD, our measure of idiosyncratic price dispersion, is calculated according to specification (1). TOM is from the Corelogic MLS data. Price is the Zillow home value index. The price, TOM, and LogSD lines are constructed as sales-weighted averages across counties for each year, and then all four series are indexed to equal 1 in 2000. Further details of data construction are described in Appendix A.5.

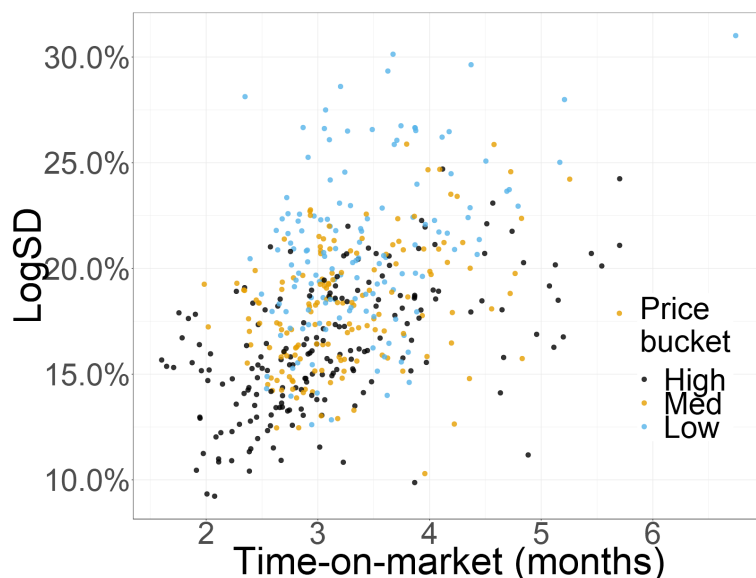
Figure 4: Business cycle variation in sales, prices, logSD, and time-on-market, heterogeneity



*Notes.* Yearly 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. Each color represents one quantile bucket of counties. The time period of the data is 2000 to 2016. LogSD, our measure of idiosyncratic price dispersion, is calculated according to specification (1). TOM is from the Corelogic MLS data. Sales is calculated as the sum of all sales in the Corelogic data. Price is the Zillow home value index. The price, TOM, and LogSD lines are constructed as sales-weighted averages across counties for each year, and then all four series are indexed to equal 1 in 2000. Further details of data construction are described in Appendix A.5.

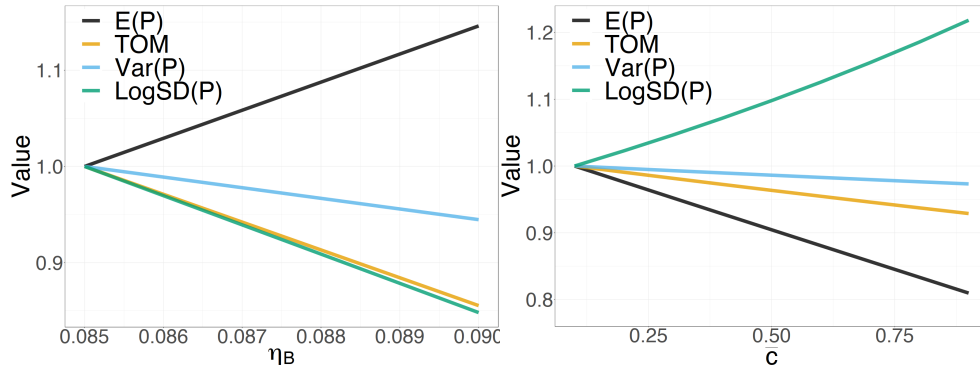


Figure 5: Cross-sectional distribution of LogSD and TOM



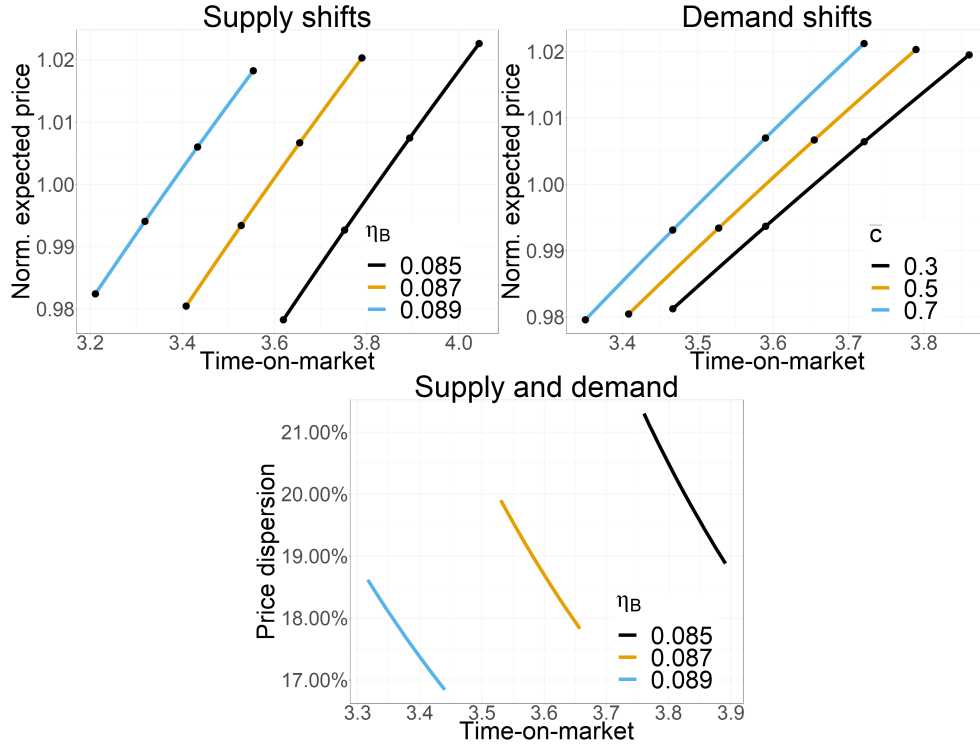
*Notes.* Distribution of time-on-market and price dispersion across counties. The data period is 2012-2016. Each data point is a county. LogSD, our measure of idiosyncratic price dispersion, is calculated according to specification (1). Time-on-market is from the Corelogic MLS data. We divide houses into three quantile buckets, high, medium, and low, according to median sale prices.

Figure 6: Model comparative statics



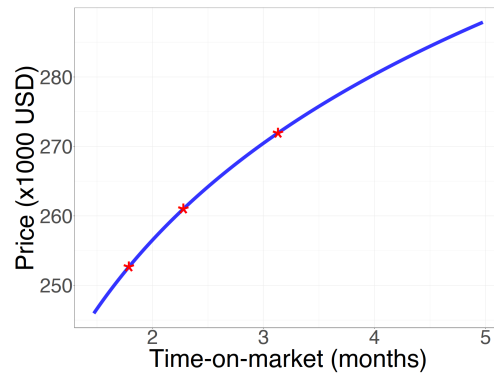
*Notes.* The two panels show how average price,  $E(P)$ , average time-on-market, TOM, price dispersion in levels,  $\text{Var}(P)$ , relative price dispersion,  $\text{LogSD}(P)$ , change as we vary buyers inflow rate  $\eta_B$  (left panel) and average holding cost  $\bar{c}$  (right panel).

Figure 7: Liquidity supply and demand



*Notes.* The top two panels show the expected price and time-on-market schedules chosen by sellers with different holding costs. The black dots represent the expected price and time-on-market pairs chosen by sellers with 0th, 30th, 60th, and 100th percentile holding cost  $c$ . The price is normalized by the price attained by a seller with a median holding cost. The top left panel plots price and time-on-market schedules for three different values of buyers inflow rate  $\eta_B = 0.085, 0.087, 0.089$ . The top right panel plots price and time-on-market schedules for three different values of sellers average holding costs  $\bar{c} = 0.3, 0.5, 0.7$ . The bottom panel plots equilibrium average time-on-market on the x-axis, and equilibrium market-level price dispersion as a percentage of average price on the y-axis. Each curve is a fixed value of  $\eta_B$ , and shows outcomes as we vary  $\bar{c}$ .

Figure 8: Calibrated liquidity discounts



*Notes.* This plot shows the average price and time-on-market achieved by sellers with different holding utilities  $c$ , under our calibrated model. The x-axis is time-on-market, and y-axis is price, in USD thousands. The red stars represent, from left to right, the 25th, 50th, and 75th percentiles of seller holding utilities.

Table 1: Descriptive statistics, county sample

Name	Mean	SD	P10	P90
Average monthly sales	418	597	72	919
Mean price (x1000 USD)	491.2	1462.0	143.6	581.7
Mean TOM (Months)	3.25	0.75	2.43	4.20
Total counties	472			
Total sales	11,807,040			

*Notes.* Summary statistics of average monthly sales, average prices, and average time-on-market, across counties in our sample. Monthly sales and mean price data are from the Corelogic Deed dataset. Time-on-market is calculated from the Corelogic MLS dataset at the house sale level as the average difference between closing date and original listing date.

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

	<i>Dependent variable:</i>					
	LogSD (1)	LogSD (2)	TOM (3)	TOM (4)	LogSD (5)	LogSD (6)
Pop growth	-0.212*** (0.076)	-0.224*** (0.069)	-0.035** (0.015)	-0.025* (0.013)		
TOM					1.093*** (0.256)	1.705*** (0.275)
Controls	X	X	X	X	X	X
Fixed effects		State		State		State
Observations	400	400	400	400	400	400
Adjusted R <sup>2</sup>	0.609	0.786	0.532	0.765	0.620	0.803

*Notes.* Each data point is a county. The sample time period is 2012-2016. Regressions are weighted by the total number of sales within the county. The dependent variable in columns 1-2 and 5-6 is LogSD, our measure of idiosyncratic price dispersion, which is calculated according to specification (1). The dependent variable in columns 3-4 is time-on-market, from the Corelogic MLS data. The independent variable, population growth rates, are from the ACS. We control for third-order polynomials in the average age of houses, average square footage, average bedroom and bathroom counts of sold 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. \*  $p < .1$ , \*\*  $p < .05$ , \*\*\*  $p < .01$ .

Table 3: County cross-sectional regressions, liquidity supply shifters, Bartik shocks

	<i>Dependent variable:</i>					
	LogSD (1)	LogSD (2)	LogSD (3)	TOM (4)	TOM (5)	TOM (6)
Mig spillover	-3.782*** (0.805)	-2.012** (0.810)	0.014 (0.834)	-0.924*** (0.159)	-0.520*** (0.163)	-0.048 (0.160)
Controls		X	X		X	X
Fixed effects			State			State
Observations	400	400	400	400	400	400
Adjusted R <sup>2</sup>	0.225	0.611	0.779	0.109	0.538	0.762

*Notes.* Each data point is a county. The sample time period is 2012-2016. Regressions are weighted by the total number of sales within the county. The dependent variable in columns 1-2 is LogSD, our measure of idiosyncratic price dispersion, which is calculated according to specification (1). The dependent variable in columns 3-4 is time-on-market, from the Corelogic MLS data. "Mig spillover", the independent variable, is a cross-sectional IV, measuring migration spillovers from high-productivity areas. These are defined formally in (19) and (20) of Section 5. In all specifications, we control for each county's own Bartik productivity shock,  $B_{c,2012-2016}$ . In columns 2-3 and 5-6, we control for third-order polynomials in the average age of houses, average square footage, average bedroom and bathroom counts of sold 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.\* $p < .1$ , \*\*  $p < .05$ , \*\*\*  $p < .01$ .

Table 4: County cross-sectional regressions, liquidity demand shifters

	<i>Dependent variable:</i>			
	LogSD (1)	LogSD (2)	TOM (3)	TOM (4)
Log price	-4.093*** (0.661)	-2.150*** (0.782)	0.500*** (0.137)	0.771*** (0.146)
Log income	5.551*** (2.005)	8.182*** (2.106)	-1.580*** (0.417)	-0.874** (0.394)
Controls	X	X	X	X
Fixed effects		State		State
Observations	400	400	400	400
Adjusted R <sup>2</sup>	0.642	0.789	0.545	0.782

*Notes.* Each data point is a county. The sample time period is 2012-2016. Regressions are weighted by the total number of sales within the county. The dependent variable in columns 1-2 is LogSD, our measure of idiosyncratic price dispersion, which is calculated according to specification (1). The dependent variable in columns 3-4 is time-on-market, from the Corelogic MLS data. We control for third-order polynomials in the average age of houses, average square footage, average bedroom and bathroom counts of sold 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. \*p < .1, \*\* p < .05, \*\*\* p < .01.



Table 5: Moment and parameter values

Moment	Value	Parameter	Value
Price	266.72	$\lambda_M$	0.052
LogSD	16.84%	$\epsilon_0$	0.5644
TOM (Months)	2.554	$\sigma_\epsilon$	0.2994
Turnover rate	0.051	$\eta_B$	0.05124
Num. visits	9.960	$\bar{c}$	6.458
PD-TOM Corr	0.876	$\Delta_c$	4.208

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

Table 6: 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 quantities are described in Appendix D.3.

# Appendix

## A Supplementary material for Section 2

### A.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), and APN sequence number, 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 property, 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 sales 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 low prior to this date. 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 manually filter out some early county-years for which the total number of sales is low.

We use the dataset that results from these cleaning steps to measure sales and average prices 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 2.2.

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 out “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, 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. Finally, if a house is ever sold at a price which is more than 5 times higher or lower than the median house price in the same county-year, we drop all observations of the house from our dataset.

Specification (1) 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 (1), so 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 from the ACS. Our dataset constitutes approximately 14.7% of all counties. Counties in our sample are larger and denser than average, so our sample constitutes around 61.7% of the total US population. The average income of counties in our sample is somewhat higher than average, but our sample is quite representative of all counties in terms of age, race, and the fraction of the population that is married.

We measure time-on-market using Corelogic MLS dataset, which contains data on individual house listings. As in the deed and tax data, 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 were sold eventually 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 time-on-market as missing.

## A.2 ACS

We use county-level demographic information 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 2006-2016. 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. To construct migration shares, we use county-to-county migration flows from the ACS 2008-2012 5-year sample. To minimize measurement error when computing in-migration exposure of county  $c$  to county  $c'$ , we drop all origin-destination pairs such that fewer than 150 people migrated from the origin to the destination.

## A.3 Quarterly Census of Employment and Wages

We use 2010, 2012, and 2016 data files from Quarterly Census of Employment and Wages (QCEW) to get data on industry-specific wages and employment for each county.

## A.4 Other time-on-market data sources

We use two other alternative data sources for time-on-market data: Zillow Research time-on-market, which is available at the county-month level from 2010-2016, and Realtor.com time-on-market, which is available at the county-month level from 2012-2016.

## A.5 Yearly and seasonal data construction

To construct the dataset used in Figures 3 and 4, we first filter to counties which we observe every year from 2000 to 2016. This leaves us with 447 counties, comprising approximately 40.17 million home sales. To construct the LogSD line in Figures 3 and 4, 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. 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 seasonal dataset, we filter to counties in which we observe positive sales for every month from 2000 to 2016. This is stricter than our filter for the yearly plots, so we get somewhat fewer counties: we are left with 162 counties, comprising approximately 21.19 million home sales over this time period. We first collapse the data to year-month level, taking 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 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.

Since all four variables – prices, price dispersion, sales, and time-on-market – have low-frequency trends over time, for the seasonal dataset, we detrend the data by fitting a piece-wise linear trend with break points every 3 years, 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 1 and 2.

## A.6 Cross-sectional data construction

To construct the county-level dataset, we take the average of the estimated residuals  $\hat{\epsilon}_{it}^2$  for each county in our sample for the time period 2012-2016. We use this time period to match the time horizon of the 5-year ACS sample. We measure total housing units and other demographic covariates for counties using the 2012-2016 ACS 5-year sample, as described in Appendix A.2 above.

Note that while our cross-sectional regressions only use estimates of  $\hat{\epsilon}_{it}^2$  from 2012-2016, these estimates are calculated based on data from the entire sample period 2000 to 2016. In other words, we estimate house fixed effects and error terms  $\hat{\epsilon}_{it}^2$  using a 17-year period, but only use error estimates from the 5-year period 2012-2016 for our cross-sectional regressions. Using the full sample period for the baseline regression is important, since we could not estimate house fixed effects without a fairly long sample period, in which

many houses are sold twice. Using the restricted sample for the cross-sectional regressions allows us to match the time period of the ACS 5-year sample that we use for county demographics.

## A.7 Implementation of specification (1)

When we estimate specification (1), 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:

$$f_c(x_i, t) = g_c^{\text{latlong}}(t, \text{lat}_i, \text{long}_i) + g_c^{\text{sqft}}(t, \text{sqft}_i) + g_c^{\text{yrbuilt}}(t, \text{yrbuilt}_i) + g_c^{\text{bedrooms}}(t, \text{bedrooms}_i) + g_c^{\text{bathrooms}}(t, \text{bathrooms}_i) \quad (23)$$

The functions  $g_c^{\text{latlong}}$ ,  $g_c^{\text{sqft}}$ , and  $g_c^{\text{yrbuilt}}$  are interacted third-order polynomials in their constituent components, and the functions  $g_c^{\text{bedrooms}}$  and  $g_c^{\text{bathrooms}}$  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 (23) 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  $g_c^{\text{latlong}}(t, \text{lat}_i, \text{long}_i)$  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 (1) is a standard fixed effects regression, and we estimate specification (1) using OLS separately for each county in our sample. Once we have estimated specification (1), we estimate squared residuals  $\hat{\epsilon}_{it}^2$  for each house sale as:

$$\hat{\epsilon}_{it}^2 = \frac{N_c}{N_c - K_c} (p_{it} - \hat{p}_{it})^2 \quad (24)$$

where  $N_c$  is the number of house sales in county  $c$ , and  $K_c$  is the number of parameters estimated from specification (1). 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 (24) causes the expectation of  $\hat{\epsilon}_{it}^2$  to be equal to the true variance,  $\sigma_c^2$ . 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 C.2, we further adjust the estimated residuals  $\hat{\epsilon}_{it}$  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.

## B Supplementary material for Section 4

### B.1 Stationary equilibrium conditions

#### B.1.1 Bellman equations

Given the buyer match rate  $\lambda_B$ , 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:

$$rV_B = \lambda_B \int \int_{\epsilon > \epsilon^*(c)} [(1 - \theta)(V_M(\epsilon) - V_B - V_S(c))] dG(\epsilon) dF_{eq}(c) \quad (25)$$

In words, expression (25) can be interpreted as follows. At rate  $\lambda_B$ , the buyer is matched to a seller with type randomly drawn from  $F_{eq}(\cdot)$ , and the buyer draws match quality  $\epsilon$  from  $G(\cdot)$ . If the buyer's match quality draw,  $\epsilon$ , 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  $\lambda_S$ , trade cutoffs  $\epsilon^*(c)$ , and the buyer value  $V_B$ , the seller value function  $V_S(c)$  satisfies:

$$rV_S(c) = -c + \lambda_S \int_{\epsilon > \epsilon^*(c)} \theta(V_M(\epsilon) - V_B - V_S(c)) dG(\epsilon) \quad (26)$$

In words, expression (26) states that a seller of type  $c$  receives flow value  $-c$  from their house while they are waiting for buyers. At rate  $\lambda_S$ , the seller meets a buyer with match value  $\epsilon$  randomly drawn from  $G(\cdot)$ . If  $\epsilon > \epsilon^*(c)$ , trade occurs, and the seller receives a



share  $\theta$  of the bilateral match surplus.

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

$$rV_M(\epsilon) = \epsilon + \lambda_M \left( \int V_S(c) dF(c) - V_M(\epsilon) \right) \quad (27)$$

In words, expression (27) states that matched owners get flow value  $\epsilon$  while matched to their house and receive separation shocks at rate  $\lambda_M$ , at which point they become sellers and attain the expectation of the seller value function  $V_S(c)$  over the seller holding cost distribution  $F(c)$ .

### B.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:

$$(1 - M_S) \lambda_M f(c) \quad (28)$$

In words, this is the product of the total mass of matched homeowners,  $1 - M_S$ ; the rate at which homeowners receive separation shocks,  $\lambda_M$ ; and the density  $f(c)$  of entering sellers with value  $c$ .<sup>11</sup> The equilibrium rate at which sellers of type  $c$  sell their houses and leave the market is:

$$M_S f_{eq}(c) \lambda_S (1 - G(\epsilon^*(c))) \quad (29)$$

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,  $\lambda_S$ ; and the probability that the match utility draw  $\epsilon$  exceeds the trade cutoff  $\epsilon^*(c)$  for a seller of type  $c$ , which is  $1 - G(\epsilon^*(c))$ . In stationary equilibrium, expressions (28) and (29) 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 (28) and (29) over  $c$ , we have:

$$(1 - M_S) \lambda_M = \int_c \lambda_S M_S (1 - G(\epsilon^*(c))) f_{eq}(c) dc \quad (30)$$

---

<sup>11</sup>Since the distribution  $F(c)$  of holding costs does not depend on matched homeowners' match utility  $\epsilon$ , we do not need to explicitly integrate over the distribution  $G_{eq}(\epsilon)$  in expression (28).

Moreover, since each successful sale turns a buyer into a matched homeowner, the RHS of (30) is also equal to the rate at which buyers turn into matched homeowners.

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

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

In words, the numerator is the flow rate at which a seller of value  $c$  successfully trades with a buyer with match utility below  $\epsilon$ , integrated over the equilibrium distribution  $f_{eq}(c)$  of holding costs  $c$  among sellers. The denominator is the RHS of (30), 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  $\xi^*$  enter, the inflow rate of buyers is equal to:

$$\eta_B (1 - F_{\xi}(\xi^*))$$

Hence, we must have:

$$(1 - M_S) \lambda_M = \eta_B (1 - F_{\xi}(\xi^*))$$

Substituting for the cutoff  $\xi^*$  using (2), this simplifies to:

$$(1 - M_S) \lambda_M = \eta_B (1 - F_{\xi}(-V_B)) \quad (31)$$

## B.2 Proof of Claim 1

### B.2.1 Time-on-market

Time-on-market for a seller of type  $c$  is the inverse of  $\lambda_S (1 - G(\epsilon^*(c)))$ , which is the product of the equilibrium rate at which sellers meet buyers,  $\lambda_S$ , and the fraction of meetings for a seller of type  $c$  that result in trade,  $(1 - G(\epsilon^*(c)))$ . Average time-on-market, (14), is the expectation of this.

### B.2.2 Price dispersion

From (4), prices are:

$$P(\epsilon, c) = \theta (V_M(\epsilon) - V_B - V_S(c)) + V_S(c) \quad (32)$$

We wish to take the variance of expression (32) with respect to the joint distribution of holding costs  $c$  and match utilities  $\epsilon$  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 (10) 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:

$$F_{tr}(c) = F(c) \quad (33)$$

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

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

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,

$$F_{tr}(\epsilon | c) = G(\epsilon | \epsilon > \epsilon^*(c)) \quad (34)$$

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

Having characterized  $F_{tr}(c, \epsilon)$ , we can now take the variance of expression (32) for prices. Applying the law of iterated expectations, price variance can be written as:

$$\text{Var}(P(\epsilon, c)) = E_{c \sim F_{tr}(c)} \left[ \text{Var}_{\epsilon \sim F_{tr}(\epsilon|c)} (P(\epsilon, c) | c) \right] + \text{Var}_{c \sim F_{tr}(c)} \left( E_{\epsilon \sim F_{tr}(\epsilon|c)} [P(\epsilon, c) | c] \right) \quad (35)$$

Substituting (33) and (34), we can write this as:

$$\text{Var} (P (\epsilon, c)) = E_{c \sim F(c)} \left[ \text{Var}_{\epsilon \sim G(\epsilon | \epsilon > \epsilon^*(c))} (P (\epsilon, c) | c) \right] + \text{Var}_{c \sim F(c)} \left( E_{\epsilon \sim G(\epsilon | \epsilon > \epsilon^*(c))} [P (\epsilon, c) | c] \right) \quad (36)$$

First, we characterize the left term on the RHS of (35). Conditional on  $c$ , the only random term in  $P (\epsilon, c)$  conditional on  $c$  is the buyer's match utility  $\epsilon$ ; thus, substituting expression (49) for  $P (\epsilon, c)$  and ignoring constant terms, we have:

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

In words,

$$\text{Var}_{\epsilon \sim G(\epsilon | \epsilon > \epsilon^*(c))} (\epsilon)$$

is the variance of an exponential random variable  $\epsilon$ , conditional on  $\epsilon$  being above some cutoff  $\epsilon^*(c)$ , which is greater than its lower bound  $\epsilon_0$ . This conditional distribution has variance equal to the unconditional variance of  $\epsilon$ ,  $\sigma_\epsilon^2$ , for any cutoff  $\epsilon^*(c)$ ; thus, we have:

$$\text{Var}_{\epsilon \sim G(\epsilon | \epsilon > \epsilon^*(c))} (P (\epsilon, c) | c) = \left( \frac{\theta}{r + \lambda_M} \right)^2 \sigma_\epsilon^2 \quad (37)$$

Since expression (37) is independent of  $c$ , we also have:

$$E_{c \sim F(c)} \left[ \text{Var}_{\epsilon \sim G(\epsilon | \epsilon > \epsilon^*(c))} (P (\epsilon, c) | c) \right] = \left( \frac{\theta}{r + \lambda_M} \right)^2 \sigma_\epsilon^2 \quad (38)$$

Now we move to the right term in expression (35). Substituting expression (49) for prices, we have:

$$\begin{aligned} \text{Var}_{c \sim F(c)} \left( E_{\epsilon \sim G(\epsilon | \epsilon > \epsilon^*(c))} [P (\epsilon, c) | c] \right) = \\ \text{Var}_{c \sim F(c)} \left( E_{\epsilon \sim G(\epsilon | \epsilon > \epsilon^*(c))} \left[ V_S (c) + \theta \left( \frac{\epsilon - \epsilon^*(c)}{r + \lambda_M} \right) \middle| c \right] \right) \end{aligned} \quad (39)$$

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

$$\text{Var}_{c \sim F(c)} \left( V_S (c) + \frac{\theta}{r + \lambda_M} E_{\epsilon \sim G(\epsilon | \epsilon > \epsilon^*(c))} [\epsilon - \epsilon^*(c) | c] \right) \quad (40)$$

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

$$\mathbb{E}_{\epsilon \sim F_{tr}(\epsilon|c)} [\epsilon - \epsilon^*(c) | c]$$

is equal to  $\sigma_\epsilon$ , the standard deviation of  $\epsilon$ . It is thus constant with respect to  $\epsilon^*(c)$  and thus  $c$ , and can be ignored when calculating the variance in (40). Hence,

$$\text{Var}_{c \sim F(c)} \left( \mathbb{E}_{\epsilon \sim G(\epsilon|\epsilon > \epsilon^*(c))} [P(\epsilon, c) | c] \right) = \text{Var}_{c \sim F(c)} (V_S(c)) \quad (41)$$

Substituting (38) and (41) into (36), we have

$$\text{Var}(P(\epsilon, c)) = \text{Var}_{c \sim F(\cdot)} (V_S(c)) + \left( \frac{\theta}{r + \lambda_M} \right)^2 \sigma_\epsilon^2 \quad (42)$$

Now, taking the expectation of prices from (49) below, we have:

$$\mathbb{E}[P(\epsilon, c)] = \mathbb{E}[V_S(c)] + \frac{\theta \sigma_\epsilon}{r + \lambda_M} \quad (43)$$

Using (42) and (43), we get (15).

### B.3 Expressions for $V_M(\epsilon)$ , $\epsilon^*(c)$ , $P(\epsilon, c)$

To begin with, we analytically characterize  $V_M(\epsilon)$ . From expression (27), we have:

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

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

$$V_M(\epsilon) = \frac{\epsilon}{r + \lambda_M} + \frac{\lambda_M}{r + \lambda_M} \int V_S(c) dF(c) \quad (44)$$

Using expression (44), we can also characterize the trade cutoff function  $\epsilon^*(c)$ . Trade occurs if:

$$\begin{aligned} V_M(\epsilon) &\geq V_B + V_S(c) \\ \implies \frac{\lambda_M}{r + \lambda_M} \int V_S(c) dF(c) + \frac{\epsilon}{r + \lambda_M} &\geq V_B + V_S(c) \end{aligned} \quad (45)$$

Since we have assumed that  $\epsilon^*(c)$  is greater than  $\epsilon_0$ , the lower bound of  $G(\cdot)$ , we can treat expression (45) as an equality. Solving for  $\epsilon^*(c)$ , we have:

$$\epsilon^*(c) = (r + \lambda_M) [V_B + V_S(c)] - \lambda_M \int V_S(c) dF(c) \quad (46)$$

Using (44) and (46) we can also characterize equilibrium prices. From (4), we have:

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

Substituting for  $V_M(\epsilon)$  using (44), we have:

$$P(\epsilon, c) = V_S(c) + \theta \left( \frac{\epsilon}{r + \lambda_M} + \frac{\lambda_M}{r + \lambda_M} \int V_S(c) dF(c) - V_B - V_S(c) \right) \quad (47)$$

Now, we can write (46) as:

$$\frac{\lambda_M}{r + \lambda_M} \int V_S(c) dF(c) - V_B - V_S(c) = -\frac{\epsilon^*(c)}{r + \lambda_M} \quad (48)$$

Hence, substituting (48) into (47), we get:

$$P(\epsilon, c) = V_S(c) + \theta \left( \frac{\epsilon - \epsilon^*(c)}{r + \lambda_M} \right) \quad (49)$$

## B.4 Proof of Claim 2

From expression (6) in proposition 1, the seller value function  $V_S(c)$  is:

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

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

$$rV'_S(c) = -1 - \lambda_S \theta(V_M(\epsilon^*(c)) - V_B - V_S(c)) g(\epsilon^*(c)) \frac{d\epsilon^*(c)}{dc} + \lambda_S \int \theta(-V'_S(c)) 1(\epsilon > \epsilon^*(c)) dG(\epsilon) \quad (50)$$

By definition of  $\epsilon^*(c)$  in (8):

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

so the middle term is 0. Hence, (50) becomes:

$$rV'_S(c) = -1 + \lambda_S \theta (-V'_S(c)) (1 - G(\epsilon^*(c)))$$

Solving for  $V'_S(c)$ , we have:

$$V'_S(c) = \frac{-1}{r + \lambda_S \theta (1 - G(\epsilon^*(c)))} \quad (51)$$

Substituting expression (14) for  $TOM(c)$  in the denominator of (51), we have:

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

Rearranging, we have (16).

## B.5 Proof of Claim 3

From expression (16), if  $TOM_{\Theta_1}(c) > TOM_{\Theta_2}(c)$  for all  $c$ , and if  $r, \theta$  are the same in the two sets of primitives, then  $V'_S(c)$  must also be strictly larger in absolute value, pointwise in  $c$ , under parameters  $\Theta_1$  than  $\Theta_2$ , for all  $c$ . From expression (15) in Claim 1, we have:

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

Holding fixed  $G(\epsilon)$ , the buyer value term in  $\text{Var}(P(\epsilon, c))$  is also the same under the two sets of primitives. Hence, we must prove that, if  $V'_S(c)$  is strictly increased pointwise in  $c$ , then  $\text{Var}_{c \sim F(\cdot)}(V_S(c))$  must also strictly increase.

To prove this, suppose a random variable  $X$  has some distribution function  $G(\cdot)$ . Its variance can be written as:

$$\text{Var}(X) = \min_{\bar{x}} \int (x - \bar{x})^2 dG(x) \quad (52)$$

To prove expression (52), note that:

$$\begin{aligned}
\int (x - \bar{x})^2 dG(x) &= \int (x - E(x) + E(x) - \bar{x})^2 dG(x) \\
&= \int (x - E(x))^2 + 2(x - E(x))(E(x) - \bar{x}) + (E(x) - \bar{x})^2 dG(x) \\
&= \int (x - E(x))^2 + (E(x) - \bar{x})^2 dG(x)
\end{aligned}$$

Thus,

$$\begin{aligned}
\min_{\bar{x}} \int (x - \bar{x})^2 dG(x) &= \min_{\bar{x}} \int (x - E(x))^2 dG(x) + \int (E(x) - \bar{x})^2 dG(x) \\
&= \int (x - E(x))^2 dG(x) = \text{Var}(X)
\end{aligned}$$

Now, call the distribution of  $V_S(c)$  among trading sellers  $F_{V_S}(V)$ . Using expression (52), we can write the variance of  $V_S(c)$  as:

$$\text{Var}(V_S(c)) = \min_{\bar{V}} \int (V - \bar{V})^2 dF_{V_S}(V) \quad (53)$$

Since the distribution of  $c$  among trading sellers is  $F(c)$ , and  $V_S(c)$  is a function of  $c$ , by changing variables to integrate over  $c$ , we have:

$$\text{Var}(V_S(c)) = \min_{\bar{c}} \int (V_S(c) - V_S(\bar{c}))^2 dF(c)$$

Hence, (53) becomes:

$$= \min_{\bar{c}} \int \left( \int_{\bar{c}}^c V'_S(c) dc \right)^2 dF(c) \quad (54)$$

A uniform increase in  $V'_S(c)$  causes the integral in (54) to strictly increase for any  $\bar{c}$ . Thus, if  $V'_S(c)$  is uniformly larger in absolute value under  $\Theta_1$  than  $\Theta_2$ , then  $\text{Var}(V_S(c))$  must also increase, and thus  $\text{Var}(P(\epsilon, c))$  must also increase.

## B.6 Derivation of model quantities

In this Appendix, we derive expressions for average prices, time-on-market, and price dispersion, which are plotted in Figure A1. 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  $\epsilon, c$  among successfully trading buyers and sellers.

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

$$\int \text{TOM}_S(c) dF(c)$$

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 (33) above that the distribution of holding costs among trading sellers is simply  $F_{\text{tr}}(c) = F(c)$ .

From Claim 1, equilibrium price variance as:

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

## B.7 Heterogeneous buyer urgency

We can extend the main model to accomodate persistent buyer heterogeneity. Suppose that buyers have some persistent type  $u \sim H(u)$ , 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  $\epsilon$ :

$$P(c, u, \epsilon) = V_S(c) + \theta (V_M(\epsilon) - V_B(u) - V_S(c)) \quad (55)$$

Thus, the match quality cutoff condition becomes:

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

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

$$\epsilon^*(c, u) \geq \epsilon_0 \quad \forall c, u$$

Buyers' and sellers' value functions become, respectively:

$$\begin{aligned} rV_B(u) &= u + \lambda_B \int_c \int_{\epsilon \geq \epsilon^*(c,u)} [(1 - \theta)(V_M(\epsilon) - V_B(u) - V_S(c))] dG(\epsilon) dF_{eq}(c) \\ rV_S(c) &= -c + \lambda_S \int_u \int_{\epsilon > \epsilon^*(c,u)} [\theta(V_M(\epsilon) - V_B(u) - V_S(c))] dG(\epsilon) dH_{eq}(u) \end{aligned} \quad (56)$$

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

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

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

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

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:

$$V_M(\epsilon) = \frac{\epsilon}{r + \lambda_M} + \frac{\lambda_M}{r + \lambda_M} \int_c V_S(c) dF(c) \quad (57)$$

The derivations in Appendix B.3 thus imply that:

$$P(c, u, \epsilon) = V_S(c) + \theta \left( \frac{\epsilon - \epsilon^*(c, u)}{r + \lambda_M} \right) \quad (58)$$

Now, similar to (35), we take the variance of prices, applying the law of iterated expectations with respect to  $c$  and  $u$ , to get:

$$\begin{aligned} \text{Var}(P(c, u, \epsilon)) &= E_{c, u \sim F_{tr}(c, u)} \left[ \text{Var}_{\epsilon \sim G_{tr}(\epsilon|c, u)} (P(c, u, \epsilon) | c, u) \right] + \\ &\quad \text{Var}_{c, u \sim F_{tr}(c, u)} \left( E_{\epsilon \sim G_{tr}(\epsilon|c, u)} [P(c, u, \epsilon) | c, u] \right) \end{aligned} \quad (59)$$

Where, analogously to expression (35),  $F_{tr}(c, u)$  is the joint distribution of  $c$  and  $u$  among trading buyers and sellers, and  $F_{tr}(\epsilon | c, u)$  is the conditional distribution of  $\epsilon$  given  $c, u$  among trading buyers and sellers. Analogously to the argument to Appendix B.2, we have:

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

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 B.2 go through essentially unchanged. Going through the steps, for the top term of (59), we have:

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

For the bottom term, substituting expression (58) for prices, we have:

$$\begin{aligned} \text{Var}_{c, u \sim F_{tr}(c, u)} \left( E_{\epsilon \sim G(\epsilon | \epsilon > \epsilon^*(c, u))} [P(c, u, \epsilon) | c, u] \right) &= \\ \text{Var}_{c, u \sim F_{tr}(c, u)} \left( E_{\epsilon \sim G(\epsilon | \epsilon > \epsilon^*(c, u))} \left[ V_S(c) + \theta \left( \frac{\epsilon - \epsilon^*(c, u)}{r + \lambda_M} \right) | c, u \right] \right) &= \\ \text{Var}_{c, u \sim F_{tr}(c, u)} \left( V_S(c) + \frac{\theta}{r + \lambda_M} E_{\epsilon \sim G(\epsilon | \epsilon > \epsilon^*(c, u))} [\epsilon - \epsilon^*(c, u) | c, u] \right) & \quad (60) \end{aligned}$$

Again, the left term of (60),

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

is equal to  $\sigma_\epsilon$ , which is independent of  $c, u$ , so we can ignore it in the variance calculation; (60) thus simplifies to

$$\text{Var}_{c, u \sim F_{tr}(c, u)} (V_S(c))$$

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

$$\text{Var}_{c \sim F(c)} (V_S(c))$$

This proves (15). The proof of (14) is identical to the baseline model. Finally, differentiating (56), we have:

$$rV'_S(c) = 1 + \lambda_S \theta \int_u \int_{\epsilon > \epsilon^*(c,u)} (-V'_S(c)) 1(\epsilon > \epsilon^*(c,u)) dG(\epsilon) dH_{eq}(u) \quad (61)$$

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

$$TOM(c) = \frac{1}{\lambda_S \int_u \int_{\epsilon > \epsilon^*(c,u)} 1(\epsilon > \epsilon^*(c,u)) dG(\epsilon) dH_{eq}(u)} \quad (62)$$

Combining (61) and (62), we have:

$$V'_S(c) = \frac{1}{r + \frac{\theta}{TOM(c)}} = \frac{TOM(c)}{rTOM(c) + \theta}$$

proving claim 2.

## B.8 Comparative statics

To illustrate how our model maps primitives to outcomes, we solve the model computationally, and supplement comparative statics results in the main text. Specifically, figure 6 shows how average prices,  $E(P)$ , average time-on-market,  $TOM$ , and price dispersion in levels,  $Var(P)$ , and relative price dispersion,  $LogSD(P)$ , change with respect to buyers' inflow rate  $\eta_B$ , homeowners' separation rate  $\lambda_m$ , parameters of buyers' match utility distribution  $\epsilon_0, \sigma_\epsilon$ , and parameters of sellers' holding cost distribution  $\bar{c}, \Delta_c$ .

# C Supplementary material for Section 6

## C.1 Panel regressions

Table A2 shows county-year panel regressions of  $\log SD_{ct}$  and  $TOM_{ct}$  on various liquidity supply measures, where a data point is a county-year. We use the time horizon 2007-2016, as the ACS 1-year samples with coverage of our variables are only available from 2007 onwards. This largely confirms findings from table 2 in the main text. Column 1 shows that time-on-market is positively correlated with  $\log SD_{ct}$  in the panel specification. Columns

2 and 3 show that immigration rates are negatively correlated with price dispersion and time-on-market in panel specifications. All coefficients are statistically significant.

Table A3 shows results from regressing  $\log SD_{ct}$  and  $TOM_{ct}$  on income and prices. The results are mostly analogous to those of Table 4: price increases predict decreases in  $\log SD_{ct}$ , and income increases predict increases in  $\log SD_{ct}$ . Price increases predict increases in  $TOM_{ct}$ , but the coefficient of  $TOM_{ct}$  on income is not significant.

## C.2 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 (1), estimating residuals using the specification:

$$p_{it} = \gamma_i + \eta_{ct} + \epsilon_{it} \quad (63)$$

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

**Pure hedonic specification:** Second, we omit house fixed effects from (1), estimating residuals using the following specification:

$$p_{it} = \eta_{ct} + f_c(x_i, t) + \epsilon_{it} \quad (64)$$

**Adjusting for time-between-sales and times sold:** Specification (1) implies that idiosyncratic price variance does not depend on the holding period. Also, when estimating (1),  $\hat{\epsilon}_{it}^2$  will tend to be larger for houses which are sold more times, because the house fixed effect  $\gamma_i$  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 A2 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  $\log SD$ ,  $|\hat{\epsilon}_{it}|$ , 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:

$$\hat{\epsilon}_{it}^2 = g_c(\text{sales}_i, \text{tbs}_i) + \zeta_{it} \quad (65)$$

Where,  $g_c(\text{sales}_i, \text{tbs}_i)$  interacts a vector of  $\text{sales}_i$  dummies with a fifth-order polynomial in  $\text{tbs}_i$ . 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  $\text{sales}_i$  and  $\text{tbs}_i$ . We then add back the mean of  $\hat{\epsilon}_{it}^2$  within county  $c$ :

$$\hat{\epsilon}_{\text{TBSadj},it}^2 = \hat{\zeta}_{it} + E_c \left[ \hat{\epsilon}_{it}^2 \right] \quad (66)$$

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

**Qualitative results:** Figure A3 compares residuals from the pure repeat-sales and pure hedonic specifications, (63) and (64) 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.

Figure A3 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, Figure A3 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 (66).

Figure A4 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, 3, and 4 in the main text. 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 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.

### **C.3 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 (1), Sagi (2015) and Giacoletti use returns, rather than individual house sales, as the primary unit of analysis.<sup>12</sup> 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 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 (1) to measure price dispersion. Specification (1) goes further, assuming that idiosyncratic variance has no relationship with holding period; this is violated in the data, but we relax this in Appendix C.2. 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

---

<sup>12</sup>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.

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 (1), which effectively allows the hedonic coefficients on different characteristics to change continuously over time. In Appendix C.2, 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 (1) does not nest that of Landvoigt, Piazzesi and Schneider (2015), since we do not include previous sale prices in specification (1); 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.<sup>13</sup>

---

<sup>13</sup>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).



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.

## **C.4 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 Figure A5, 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.

Figure A6 shows how Zillow and Realtor.com time-on-market behave seasonally and over the business cycle. All three data sources display seasonality, 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 time-on-market as the dependent variable from Tables 2 and 4 in the main text. 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.

## **D Supplementary material for Section 7**

## D.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  $\Delta_c$ , and in the inner loop we solve for all other parameters conditional on  $\Delta_c$ . In the inner loop, for any guess for  $\Delta_c$ , 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 D.1.2, and we choose  $\Delta_c$  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 D.1.3.

### D.1.1 Inner loop

Given any value of  $\Delta_c$ , 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 estate auctions,<sup>14</sup> 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  $\sigma_\epsilon$  such that:

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

While the mapping between moments and parameters is complex, roughly speaking, the input parameters determine the output moments as follows. The lower bound,  $\epsilon_0$ , and dispersion,  $\sigma_\epsilon^2$ , of buyer utilities jointly determine the level of buyer-induced price

---

<sup>14</sup>We use these data because we are unaware of publically available bid data on housing auctions in the US.

dispersion and the overall level of house prices. The mean of seller holding utilities,  $\bar{c}$ , and the lower bound,  $\epsilon_0$ , affect average gains-from-trade and thus move prices and the average number of house visits by buyers before purchasing. The entry rate  $\eta_B$  of buyers determines market tightness, which determines average time-on-market. The separation rate  $\lambda_M$  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:

$$(0.168)^2 \approx \underbrace{\frac{\text{Var}_{c \sim F(\cdot)}(V_S(c))}{(E(P(c, \epsilon)))^2}}_{\text{Seller holding value}} + \underbrace{\frac{1}{(E(P(c, \epsilon)))^2} \left( \frac{\theta}{r + \lambda_M} \right)^2 \sigma_\epsilon^2}_{\text{Buyer match value}} + \sigma_{\text{residual}}^2 \quad (67)$$

In our baseline model estimates, the seller holding value component has a standard deviation of 4.39% of prices, the buyer component is 6.45% of house prices, and the unobserved heterogeneity component is 14.92% of house prices. In terms of variance fractions, 6.80% of total logSD is attributable to seller values, 14.67% to buyer values, and 78.53% 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.

### D.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  $\eta_B$  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  $\Delta_c$  to match this model-predicted TOM-logSD relationship to the

county panel coefficient.

### D.1.3 Computation

Computationally, we solve the model by varying  $M_B$  rather than  $\eta_B$ ; this is computationally simpler, and for any equilibrium in terms of  $M_B$ , we can back out an  $\eta_B$  which implements this equilibrium, through (12). 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  $\lambda_S$  and  $\lambda_B$ , and then numerically solve (5), (6), (7) for  $V_S(c), V_M(c), V_B, \epsilon^*(c)$ . Given guesses for the trade cutoff  $\epsilon^*(c)$ , we can then use (10) to solve for  $f_{eq}(c)$ . We iterate these equations, updating in a penalized manner; 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.

### D.1.4 Estimating surplus

We calculate buyers' total expected surplus as  $\int \frac{E[\epsilon | \epsilon > \epsilon^*(c)]}{r + \lambda_m} dF(c)$ , and sellers' total expected holding costs from staying on the market as  $\int c \cdot TOM(c) dF(c)$ .

## D.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  $\bar{P}$ :

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

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 (15) of Claim 1.

### D.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 6 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 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:

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

## Foreclosure discounts

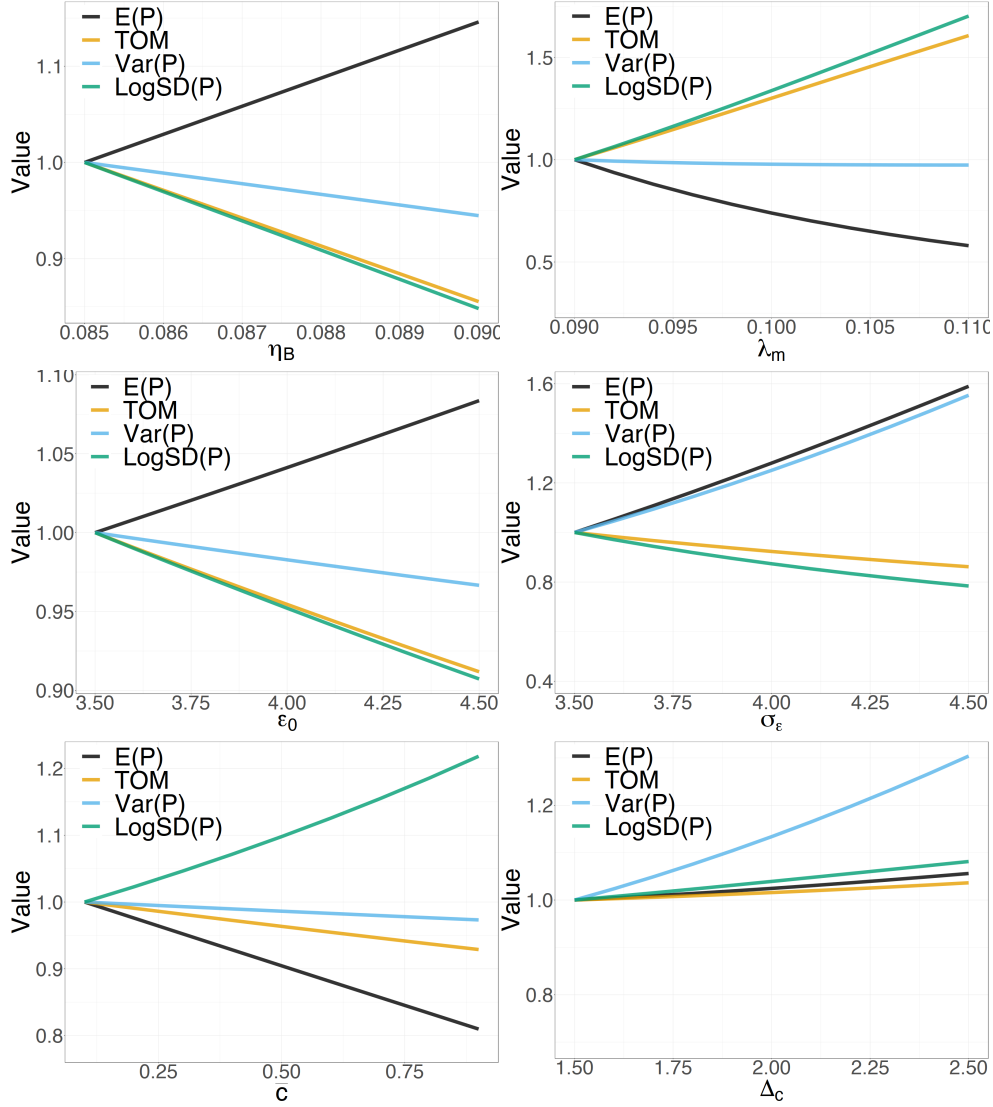
- Pennington-Cross (2006), using confidential data, finds foreclosure discounts or

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.

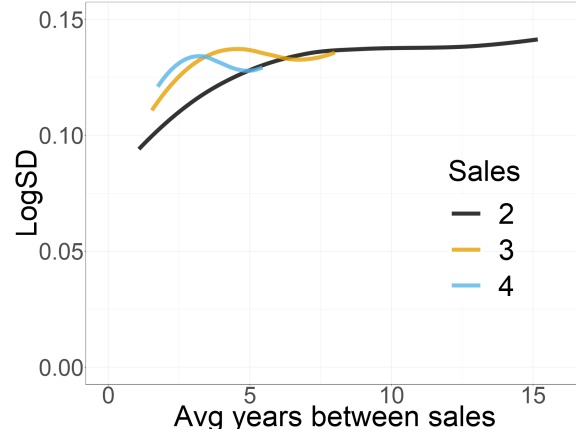


Figure A1: Model comparative statics



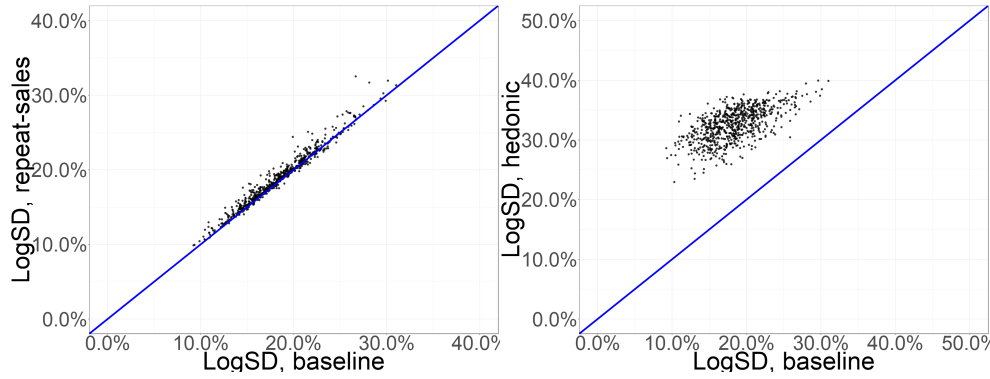
*Notes.* The six panels show how average price,  $E(P)$ , average time-on-market,  $TOM$ , price dispersion in levels,  $Var(P)$ , relative price dispersion,  $LogSD(P)$ , change, as we vary (from top to bottom, left to right) buyers inflow rate  $\eta_B$ , homeowner separation rate  $\lambda_m$ , the buyer match utility parameters  $\epsilon_0, \sigma_\epsilon$ , and the seller holding cost parameters  $\bar{c}, \Delta_c$ .

Figure A2: Effect of number of sales and time-between-sales on LogSD



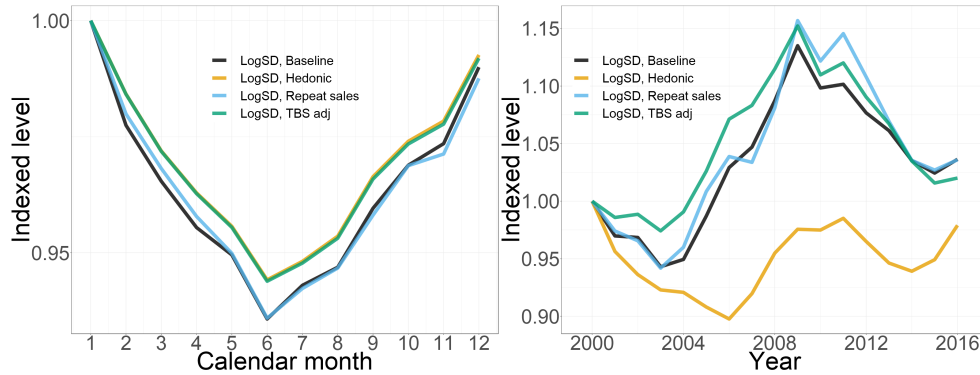
*Notes.* Variation of our estimated idiosyncratic house price residuals,  $\hat{\epsilon}_{it}^2$ , with respect to the average time between house sales, and the number of times a house is sold. We calculate  $\hat{\epsilon}_{it}^2$  for each house sale using specification (1), and then run a kernel regression of  $\hat{\epsilon}_{it}^2$  on  $tbs_i$ , the average time between house sales for house  $i$ . We run this regression separately for houses sold 2, 3, and 4 times, corresponding to the black, yellow, and blue lines. The figure shows the kernel regression estimates of conditional means of  $\hat{\epsilon}_{it}^2$ .

Figure A3: LogSD: alternative measurements



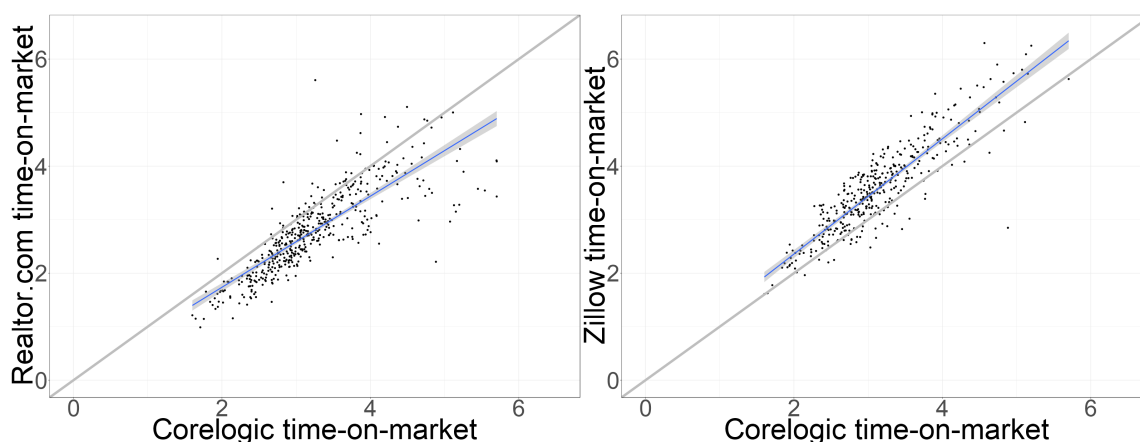
*Notes.* The left panel shows, on the x-axis, estimates of  $\hat{\sigma}_c$  from the baseline specification, (1), and on the y-axis, estimates of  $\hat{\sigma}_c$  from specification (63). Specification (63) is a pure repeat-sales model of house prices, omitting the  $f_c(x_i, t)$  term from specification (1), which is a flexible function of house characteristics and time. The right panel shows, on the x-axis, estimates of  $\hat{\sigma}_c$  from the baseline specification, and on the y-axis, estimates of  $\hat{\sigma}_c$  from specification (64). Specification (64) is a pure hedonic model of house prices, omitting the house fixed effect term  $\gamma_i$  from specification (1). In both plots, the sample period is 2012-2016. Each data point represents one county.

Figure A4: LogSD: Alternative measures in the time-series



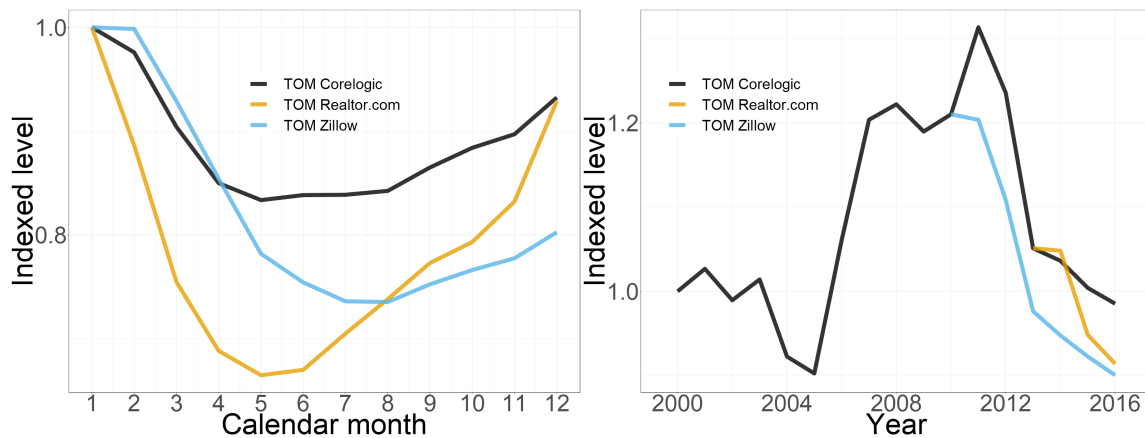
*Notes.* The left panel shows the four estimates of price dispersion – the baseline specification (1), the pure repeat-sales model (63), the pure hedonic model (64), and the time-between-sales adjusted estimates (66), over calendar months. For all variables, we filter out low-frequency trends by fitting a piece-wise linear trend with break points every 3 years, subtracting away the predicted values, and adding back the mean. We then index all lines to equal 1 in January. The right panel shows our four estimates of price dispersion over the business cycle. All four lines are constructed as sales-weighted averages across counties for each year, and then all series are indexed to equal 1 in 2000. For both plots, the time period is 2000 to 2016.

Figure A5: Time-on-market: alternative measurements



*Notes.* The left panel shows Corelogic MLS time-on-market on the x-axis, against Realtor.com time-on-market on the y-axis. The right panel shows Corelogic MLS time-on-market on the x-axis, against Zillow time-on-market on the y-axis. In both plots, the sample period is 2012-2016. Each data point represents one county.

Figure A6: Time-on-market: alternative measurements time series



*Notes.* The left panel shows our three estimates of time-on-market – from Corelogic, Realtor.com, and Zillow – over calendar months. For all variables, we filter out low-frequency trends by fitting a piece-wise linear trend with break points every 3 years, subtracting away the predicted values, and adding back the mean. We then index all lines to equal 1 in January. The right panel shows our four estimates of price dispersion over the business cycle. All four lines are constructed as sales-weighted averages across counties for each year. We index Corelogic time-on-market to equal 1 in 2000. We index Realtor.com and Zillow time-on-market so that they are equal to the indexed Corelogic time-on-market in the first year that we observe them.

Table A1: Characteristics of counties in our dataset

	Sample mean	All counties mean
Population	420,717	100,027
Pop / Sq mile	992	290
Housing units	171,513	42,120
Avg income	\$77,984	\$61,995
% Age 18-35	22.4%	20.7%
% Married	49.9%	51.3%
% Black	11.1%	9.00%
Total counties	472	3,220
Total pop (1000's)	198,578	322,088

*Notes.* Characteristics of counties in our primary estimation sample, compared to all counties in ACS 2012-2016 5-year sample. All variables – population, population density, the number of housing units, age, fraction of population which is married, and fraction of population which is Black – are from the ACS 2012-2016 5-year sample. “Sample mean” shows the mean of the variable within our main sample of counties. “All counties mean” shows the mean within all counties in the ACS 2012-2016 sample.

Table A2: County-year panel regressions, liquidity supply shifters

	<i>Dependent variable:</i>		
	LogSD (1)	LogSD (2)	TOM (3)
TOM	0.678*** (0.147)		
Pop growth		-12.130*** (3.374)	-2.002*** (0.604)
County fixed effects	X	X	X
Year fixed effects	X	X	X
Sample period	2007-2016	2007-2016	2007-2016
Observations	3,360	3,360	3,360
Adjusted R <sup>2</sup>	0.901	0.901	0.871

*Notes.* Each data point is a county-year. The sample period is 2007-2016. Regressions are weighted by the average number of sales in a given county over all years. Standard errors are clustered at the county level. All columns include county and year fixed effects. The dependent variable in columns 1-2 is LogSD, our measure of idiosyncratic price dispersion, which is calculated according to specification (1). The dependent variable in column 3 is time-on-market, which is calculated at the house level as the average difference between closing date and original listing date, and then averaged to the county-year level. Vacancy rates and population growth rates are from the ACS 1-year samples. \*p < .1, \*\* p < .05, \*\*\* p < .01.



Table A3: County-year panel regressions, liquidity demand shifters

	<i>Dependent variable:</i>	
	LogSD (1)	TOM (2)
Log price	-3.273*** (0.653)	0.266*** (0.097)
Log income	4.048*** (1.169)	0.033 (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 <sup>2</sup>	0.909	0.872

*Notes.* Each data point is a county-year. The sample period is 2007-2016. Regressions are weighted by the average number of sales in a given county over all years. Standard errors are clustered at the county level. All columns include county and year fixed effects. The dependent variable in column 1 is LogSD, our measure of idiosyncratic price dispersion, which is calculated according to specification (1). The dependent variable in column 2 is time-on-market, from the Corelogic MLS data. We calculate log median house prices at the county-year level using the Corelogic deed data, and mean log income is from the ACS 1-year samples. \*p < .1, \*\* p < .05, \*\*\* p < .01.

Table A4: County cross-sectional regressions, price dispersion

	<i>Dependent variable:</i>				
	LogSD				
	(1)	(2)	(3)	(4)	(5)
Log price	−4.093*** (0.661)				
Log income	5.551*** (2.005)				
TOM		1.093*** (0.256)			
Pop growth			−0.212*** (0.076)		
Mig spillover				−3.782*** (0.805)	−2.012** (0.810)
Controls	X	X	X	X	X
Fixed effects					
Observations	400	400	400	400	400
Adjusted R <sup>2</sup>	0.642	0.620	0.609	0.225	0.611

*Notes.* The dependent variable is LogSD, our measure of idiosyncratic price dispersion, which is calculated according to specification (1). Time-on-market is from the Corelogic MLS data. Mean log income and population growth rates are from the ACS 2012-2016 5-year sample. The independent variable in columns 4 and 5, “Mig spillover”, is a cross-sectional IV which measures migration spillovers from high-productivity areas, described in Section 5. In all specifications with the instrument, we control for each county’s own Bartik productivity shock,  $B_{c,2012-2016}$ . Each data point is a county. The sample time period is 2012-2016. Regressions are weighted by the total number of sales within the county. \* $p < .1$ , \*\*  $p < .05$ , \*\*\*  $p < .01$ .

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

	<i>Dependent variable:</i>				
	LogSD				
	(1)	(2)	(3)	(4)	(5)
Log price	-4.661*** (0.697)				
Log income	8.338*** (2.115)				
TOM		0.771*** (0.275)			
Pop growth			-0.164** (0.080)		
Mig spillover				-3.386*** (0.878)	-1.539* (0.863)
Controls	X	X	X	X	X
Fixed effects					
Observations	400	400	400	400	400
Adjusted R <sup>2</sup>	0.645	0.610	0.606	0.179	0.607

*Notes.* The dependent variable is idiosyncratic price dispersion, calculated using a repeat sales specification for house prices, (63). Time-on-market is from the Corelogic MLS data. Mean log income and population growth rates are from the ACS 2012-2016 5-year sample. The independent variable in columns 4 and 5, “Mig spillover”, is a cross-sectional IV which measures migration spillovers from high-productivity areas, described in Section 5. In all specifications with the instrument, we control for each county’s own Bartik productivity shock,  $B_{c,2012-2016}$ . Each data point is a county. The sample time period is 2012-2016. Regressions are weighted by the total number of sales within the county. \* $p < .1$ , \*\*  $p < .05$ , \*\*\*  $p < .01$ .

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

	<i>Dependent variable:</i>				
	(1)	(2)	(3)	(4)	(5)
Log price	−3.301*** (0.499)				
Log income	2.267 (1.513)				
TOM		0.700*** (0.199)			
Pop growth			−0.202*** (0.058)		
Mig spillover				−4.339*** (0.618)	−2.333*** (0.620)
Controls	X	X	X	X	X
Fixed effects					
Observations	400	400	400	400	400
Adjusted R <sup>2</sup>	0.643	0.597	0.597	0.201	0.600

*Notes.* The dependent variable is idiosyncratic price dispersion, calculated using a hedonic specification for house prices, (64). Time-on-market is from the Corelogic MLS data. Mean log income and population growth rates are from the ACS 2012-2016 5-year sample. The independent variable in columns 4 and 5, “Mig spillover”, is a cross-sectional IV which measures migration spillovers from high-productivity areas, described in Section 5. In all specifications with the instrument, we control for each county’s own Bartik productivity shock,  $B_{c,2012-2016}$ . Each data point is a county. The sample time period is 2012-2016. Regressions are weighted by the total number of sales within the county. \* $p < .1$ , \*\*  $p < .05$ , \*\*\*  $p < .01$ .

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

	<i>Dependent variable:</i>				
	LogSD				
	(1)	(2)	(3)	(4)	(5)
Log price	−3.210*** (0.487)				
Log income	5.278*** (1.476)				
TOM		0.462** (0.192)			
Pop growth			−0.132** (0.056)		
Mig spillover				−2.312*** (0.614)	−1.198** (0.601)
Controls	X	X	X	X	X
Fixed effects					
Observations	400	400	400	400	400
Adjusted R <sup>2</sup>	0.652	0.615	0.615	0.191	0.616

*Notes.* The dependent variable is LogSD, our measure of idiosyncratic price dispersion, calculated according to specification (1): we regress house sale prices on house fixed effects, county-year fixed effects, and a smooth function of house characteristics and time. We then nonparametrically purge the estimated residuals of any variation which is predictable based on the time-between-sales of a given house, and the number of times the house is sold, using equations (65) and (66). We calculate log median house prices using the Corelogic deed data. Time-on-market is from the Corelogic MLS data. Mean log income and population growth rates are from the ACS 2012-2016 5-year sample. Each data point is a county. The sample time period is 2012-2016. Regressions are weighted by the total number of sales within the county. The independent variable in columns 4 and 5, “Mig spillover”, is a cross-sectional IV which measures migration spillovers from high-productivity areas, described in Section 5. In all specifications with the instrument, we control for each county’s own Bartik productivity shock,  $B_{c,2012-2016}$ . \*  $p < .1$ , \*\*  $p < .05$ , \*\*\*  $p < .01$ .

Table A8: County cross-sectional regressions, time-on-market

	<i>Dependent variable:</i>			
	Time-on-market			
	(1)	(2)	(3)	(4)
Log price	0.500*** (0.137)			
Log income	-1.580*** (0.417)			
Pop growth		-0.035** (0.015)		
Mig spillover			-0.924*** (0.159)	-0.520*** (0.163)
Controls	X	X		X
Fixed effects				
Observations	400	400	400	400
Adjusted R <sup>2</sup>	0.545	0.532	0.109	0.538

*Notes.* The dependent variable is time-on-market from the Corelogic MLS dataset, which we calculate for each house sale as the difference between closing date and original listing date, and then averaged to the county level. We calculate log median house prices using the Corelogic deed data. Mean log income and population growth rates are from the ACS 2012-2016 5-year sample. The independent variable in columns 3 and 4, "Mig spillover", is a cross-sectional IV which measures migration spillovers from high-productivity areas, described in Section 5. In all specifications with the instrument, we control for each county's own Bartik productivity shock,  $B_{c,2012-2016}$ . Each data point is a county. The sample time period is 2012-2016. Regressions are weighted by the total number of sales within the county. \*  $p < .1$ , \*\*  $p < .05$ , \*\*\*  $p < .01$ .

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

	<i>Dependent variable:</i>			
	(1)	(2)	(3)	(4)
Log price	0.269 (0.179)			
Log income	-1.322** (0.544)			
Pop growth		-0.064*** (0.019)		
Mig spillover			-1.751*** (0.205)	-0.907*** (0.216)
Controls	X	X		X
Fixed effects				
Observations	321	321	321	321
Adjusted R <sup>2</sup>	0.601	0.609	0.277	0.618

*Notes.* The dependent variable is time-on-market from Zillow Research. We calculate log median house prices using the Corelogic deed data. Mean log income and population growth rates are from the ACS 2012-2016 5-year sample. The independent variable in columns 3 and 4, “Mig spillover”, is a cross-sectional IV which measures migration spillovers from high-productivity areas, described in Section 5. In all specifications with the instrument, we control for each county’s own Bartik productivity shock,  $B_{c,2012-2016}$ . Each data point is a county. The sample time period is 2012-2016. Regressions are weighted by the total number of sales within the county. \* $p < .1$ , \*\* $p < .05$ , \*\*\* $p < .01$ .

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

	<i>Dependent variable:</i>			
	Time-on-market			
	(1)	(2)	(3)	(4)
Log price	0.386*** (0.134)			
Log income	−2.008*** (0.408)			
Pop growth		−0.058*** (0.015)		
Mig spillover			−1.008*** (0.169)	−0.759*** (0.160)
Controls	X	X		X
Fixed effects				
Observations	382	382	382	382
Adjusted R <sup>2</sup>	0.613	0.604	0.118	0.612

*Notes.* The dependent variable is time-on-market from Realtor.com. We calculate log median house prices using the Corelogic deed data. Mean log income and population growth rates are from the ACS 2012-2016 5-year sample. The independent variable in columns 3 and 4, “Mig spillover”, is a cross-sectional IV which measures migration spillovers from high-productivity areas, described in Section 5. In all specifications with the instrument, we control for each county’s own Bartik productivity shock,  $B_{c,2012-2016}$ . Each data point is a county. The sample time period is 2012-2016. Regressions are weighted by the total number of sales within the county. \* $p < .1$ , \*\* $p < .05$ , \*\*\* $p < .01$ .