Liquidity in Residential Real Estate Markets*

Erica Xuewei Jiang[†], Nadia Kotova[‡], and Anthony Lee Zhang[§]

March 2024

Abstract

What is the right metric, or combination of metrics, to measure liquidity in housing markets? We show that time-on-market and price dispersion, two potential liquidity metrics, can be thought of like quantity and price measures in a supply and demand system for liquidity. The two metrics should thus be jointly used to measure market liquidity, because either metric alone can produce misleading implications for outcomes such as house prices. We provide empirical evidence for the framework's predictions using data on the US housing market.

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, 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.

†University of Southern California Marshall School of Business; erica.jiang@marshall.usc.edu

‡Linkedin; nkotova@linkedin.com.

§University of Chicago Booth School of Business; anthony.zhang@chicagobooth.edu.

1 Introduction

Residential real estate is a notoriously illiquid asset class; liquidity conditions in real estate markets can affect household mobility, and thus is an important factor monitored by policymakers.[1] Intuitively, a house is liquid if it can be sold quickly and at a stable price. Yet, while intuitive, it is surprisingly difficult to define liquidity in a way that allows it to be precisely measured and quantified. The academic literature has adopted many different measures of liquidity, but the relationships between different measures are not well understood. Empirically, do different measures of market liquidity always co-move together? If not, what do independent movements in different measures of market liquidity tell us about what primitives are changing in housing markets?

We construct a stylized random search model to show that two measures of housing market liquidity – time-on-market (TOM) and price dispersion (PD) – can be thought of like quantity and price measures in a *supply and demand system for liquidity.* Sellers, who demand liquidity, have different utility costs of keeping their houses on the market per unit time. Buyers, who supply liquidity, receive an independent match quality shock every time they match with a house. The key tradeoff sellers face is that selling quickly is costly: if sellers choose to sell faster, they may not match with high-valued buyers, and must accept lower prices as a result. In the model, when liquidity *supply* increases – that is, there are more buyers – sellers can sell more quickly, and at higher and more stable prices. In this case, TOM and PD both decrease while price increases. However, when liquidity *demand* increases – that is, sellers become more impatient in the aggregate – sellers will choose to sell faster but may incur lower and less stable prices as a result. In this case, TOM decreases, but PD may increase, and prices will decrease.

In academic and industry settings, TOM is often used as a measure of housing market liquidity. Our theory suggests that TOM alone can be a misleading measure of housing market liquidity: a TOM decrease is associated with an *increase* in house prices if it is driven by liquidity supply but is associated with a *decrease* in house prices if it is driven by liquidity demand. Measuring PD, together with TOM, helps more accurately summarize liquidity conditions in housing markets. We empirically show that proxies for liquidity supply

[1] For example, see the Federal Housing Finance Agency (FHFA) discussion of housing supply and liquidity during the COVID-19 era (Contat and Rogers, 2022).

and demand move TOM, PD, and prices in the directions predicted by the model.

We measure TOM and PD using house transaction records in the US. TOM is measured as the time between the listing date and the transaction date of a property. PD is defined as the extent to which similar houses sold at similar points in time sell for different prices. To measure price dispersion, 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 squared residuals from this regression as our measure of PD.

We begin by presenting stylized facts about the relationship between TOM and PD. In the time series, TOM and PD appear to co-move mostly positively. Both metrics are countercyclical and seasonal and co-move with other measures of housing market "hotness", such as sales volume and average prices. However, in the cross section, TOM and PD are only weakly positively correlated, and there is substantial idiosyncratic variation in both measures. One interpretation of these stylized facts is that housing market liquidity may be a multi-dimensional object, which TOM and PD measure different aspects of.

We then build a standard random search model to understand the drivers of shifts in TOM and PD. In the model, a mass of available buyers supply liquidity. Sellers demand liquidity: they wish to sell their houses quickly and are willing to accept price discounts to do so. Thus, sellers trade off time-on-market with average prices: more patient sellers wait longer for high-valued buyers, in order to sell for higher prices.

In a stylized manner, TOM can be thought of as the "quantity" of liquidity that results in equilibrium, and PD can be thought of as the "price" of liquidity. When there are more available buyers, liquidity supply increases. Sellers can sell more quickly, and at higher and more stable prices. Thus, both TOM and PD decrease, and housing markets look better on both dimensions. However, this is not the only way that market outcomes can vary. When sellers' holding costs increase, so sellers are on average less patient, liquidity demand increases. Sellers will then choose to sell faster, but may incur lower and less stable prices as a result. In this case, TOM decreases, but prices decrease; moreover, PD may in fact increase, if the decrease in prices is large enough. Monitoring TOM alone is thus insufficient to summarize liquidity conditions in housing markets, since TOM decreases can be driven by liquidity supply or demand; observing PD can help, since liquidity demand increases can cause PD to increase, whereas liquidity supply increases always cause PD to decrease.

Our model thus predicts that increases in liquidity supply should be associated with lower TOM and PD, and higher prices; and increases in liquidity demand should be associated with lower TOM, but *lower* prices, and possibly higher PD. We proceed to test these predictions empirically.

Liquidity supply in our model is determined by the number of potential buyers in a market. Population growth rates in an area are thus plausibly related to liquidity supply. However, population growth is affected by a large number of variables which may have other effects on outcomes. Thus, we use an Bartik-IV approach, following Schubert (2021), to construct a plausibly exogenous shock to liquidity supply based on *migration spillovers* from high-productivity areas. Intuitively, suppose, for example, Cook County experiences large productivity shocks, which increase house prices; Cook County may thus experience outmigration, creating net inmigration flows to areas with strong historical migration links to Cook County. Predicted migration flows based on historical migration patterns are thus plausibly exogenous shocks to housing market tightness in these areas with strong migration links. Consistent with this narrative, an IV based on interacting of county productivity shocks and historical migration links indeed predicts stronger migration flows, and appears to be uncorrelated with other local economic conditions that may affect housing market tightness.

As our model predicts, counties with high values of our liquidity supply proxy tend to have lower TOM, lower PD, and higher price growth. Quantitatively, a 1% increase in migration spillover induced population growth is associated with a 0.39 month (about 12 days) decrease in time-on-market, 1% lower price dispersion, and 2% higher house price growth.

As an empirical proxy for liquidity demand, we use average seller home equity at zipcode level. A number of papers have documented that sellers with less home equity tend to set higher prices, waiting longer to sell as a result (Genesove and Mayer, 1997, 2001; Guren, 2018). In the context of our model, we interpret sellers with less home equity as having lower holding costs; zipcodes with more of these sellers thus can be thought of as having lower liquidity demand. To obtain plausibly exogenous variation in zipcode seller home equity, we follow Guren (2018) and Huang, Nelson and Ross (2021) to find county-level house price growth since the last house transaction for every listed property and aggregate it to zipcode-month level. We use this measure to instrument for zipcode seller home equity and include county-year fixed effects in the regressions to absorb variation in yearly house price growth. Our design thus compares outcomes in zipcodes with higher versus lower values of instrumented home equity, due to differences in average historical transaction timing, within the same county.

As our model predicts, zipcodes with higher instrumented home equity, and thus higher liquidity demand, tend to have lower TOM, but in fact lower prices and higher PD. A 1% increase in home equity ratio is associated with approximately a 1% increase in price dispersion, a 0.9 month decrease in time-on-market, and a 10% decline in house price. Our empirical results thus support our model's conclusion that a decrease in TOM can be associated with either an increase or decrease in prices, depending on whether the driver is liquidity supply or demand; moreover, PD provides some information helping to distinguish these two cases.

Quantitatively, how much do liquidity supply and demand shocks move prices, and how large are the mistakes one could make if one forecasted price changes using only a single liquidity measure? We address these questions by calibrating our model to the data, matching various aggregate moments from our data and the housing microstructure literature. Our calibrated model produces average "liquidity discounts" – price discounts that sellers incur for selling faster – in line with past work, though the range of estimates for this quantity in the literature is fairly wide.

In our calibrated model, if we observe TOM decrease by 5%, and PD decrease by 5%, our model rationalizes this as the outcome from an increase in liquidity supply, and a smaller increase in liquidity demand, which causes prices to rise by 3.37%. Suppose instead we see a TOM decrease, but a PD *increase* of 5%: we infer that this is driven by a decrease in liquidity supply and an increase in liquidity demand, which causes prices to fall by 6.48%. Thus, our calibrated model suggests that a TOM increase can be associated with an increase or decrease in prices, depending on whether PD increases or decreases, illustrating the importance of using both quantities to measure housing market liquidity.

In the academic and industry literature, time-on-market is widely used as an indicator for housing market liquidity.[2] Price dispersion has received comparatively less attention;

[2]On the industry side, three measures of market "hotness" that use TOM as an input are Realtor.com's Market

the academic literature on house price dispersion is younger, and we are not aware of any industry liquidity measures which utilize price dispersion. Our work builds on two recent papers showing that idiosyncratic house-level risk in real estate markets – what we call price dispersion – is basically unrelated to holding periods, and thus is likely driven by transaction-level illiquidity rather than value "drift" (Giacoletti, 2021; Sagi, 2021).[3]

A natural implication of these papers is that price dispersion could be used as a measure of housing market liquidity at an aggregated level. Our paper evaluates this implication, analyzing theoretically and empirically how PD and TOM are related to each other and to market primitives. PD is informative in the same way as TOM when outcomes are driven by liquidity supply: we showed that PD co-moves seasonally and cyclically with TOM, and also decreases in response to increased buying pressure in individual counties. PD may also provide additional information in settings where market outcomes are driven by increases in liquidity demand, where observing TOM alone would give misleading implications about how market liquidity would affect prices. Our work suggests that price dispersion could be calculated and monitored as an additional indicator for housing market liquidity. Price dispersion is fairly straightforward to measure, so it is not logistically difficult to calculate price dispersion at a granular geographical level on an ongoing basis.

Besides the literature on price dispersion and time-on-market, our paper builds on a set of papers arguing that sellers in housing markets face a tradeoff between TOM and prices: sellers can choose to sell faster, but incur a "liquidity discount" as a result. There is empirical support for the existence of this menu in a number of papers (Genesove and Mayer, 1997, 2001; Levitt and Syverson, 2008; Hendel, Nevo and Ortalo-Magné, 2009; Guren, 2018; Buchak et al., 2020), and a related set of papers shows that foreclosure sales tend to have significantly

Hotness Index, which uses is based on time-on-market and average listing views; Zillow's older Market Health Index, which is based on price trends, homeowners' financial health, and time-on-market; and John Burns Research & Consulting's Burns Housing Market Hotness Index, which uses data on sales and prices, job growth, new constructions, and months of resale supply, a metric related to time-on-market. On the academic side, Lippman and McCall (1986), in general asset markets, suggests time-to-sale as a measure of liquidity. Kluger and Miller (1990) proposes a relative measure of house liquidity based on relative sale probabilities. Lin and Vandell (2007) discusses how to account for TOM risk in measuring housing returns. Carrillo (2013) constructs a housing market "hotness" index based on list prices, sale prices and time-on-market, and Carrillo and Williams (2015) constructs a "repeat-time-on-market index" to measure housing market liquidity.

[3]There are a few other related papers analyzing house price dispersion. Case and Shiller (1988), an early paper showing that prices of individual houses are much more volatile than city-wide average prices. Ben-Shahar and Golan (2019) show that improved disclosure of transaction prices reduces price dispersion in the Israeli housing market. Rekkas, Wright and Zhu (2020) suggests that it is difficult to quantitatively rationalize house price dispersion from hedonic models entirely through search frictions. Amaral (2024) shows that properties with higher price dispersion have higher rental yields and total returns.

reduced prices (Pennington-Cross, 2006; Clauretie and Daneshvary, 2009; Campbell, Giglio and Pathak, 2011; Harding, Rosenblatt and Yao, 2012; Zhou et al., 2015). Many theoretical papers in the search-and-matching literature, such as Guren and McQuade (2020), Andersen et al. (2022), Rekkas, Wright and Zhu (2020), and others, analyze determinants and consequences of this menu. The TOM-price tradeoff plays an important role in our model: our core results about liquidity supply and demand can be thought of, in a stylized way, as describing whether market outcomes are driven by changes in buyer availability, which shifts the set of possible TOM-price combinations, or seller urgency, which shifts sellers' aggregate choices this tradeoff. Our calibrated model is also able to quantitatively match existing estimates of the TOM-PD tradeoff, though there is a fairly wide range of estimates of this quantity in the literature.

More broadly, our paper fits into a literature on housing market microstructure. We use a random-search and Nash bargaining model of house trading and price determination. Wheaton (1990) is an early paper in this literature, and other papers using this class of approaches include Krainer (2001), Albrecht et al. (2007), Novy-Marx (2009), Piazzesi and Schneider (2009), Genesove and Han (2012), Ngai and Tenreyro (2014), Head, Lloyd-Ellis and Sun (2014), Piazzesi, Schneider and Stroebel (2020), Gabrovski and Ortego-Marti (2019), Gabrovski and Ortego-Marti (2021), Ouazad and Rancière (2019), Anenberg and Bayer (2020), Guren and McQuade (2020), Buchak et al. (2020), and Sagi (2021).[4] Our model does not make a substantial technical contribution relative to this literature: differences in sellers' holding costs are used in many papers, such as Anenberg and Bayer, Guren and McQuade, Buchak et al., and Sagi, and dispersion in buyer values is also commonly used, with Guren and McQuade using the same exponential buyer value distribution we use. Our main contribution relative to this literature is to conduct a new comparative static the literature has not analyzed: to show how liquidity "supply", buying pressure, and liquidity "demand", holding costs, affect the liquidity measures of TOM and PD, and what this tells us about the direction that house prices move.

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

[4]Other modelling approaches, which are less related technically to our model, are price posting models, such as Andersen et al. (2022); auction-based models, such as Arefeva (2019) and Anundsen et al. (2023); and directed-search models, such as Albrecht, Gautier and Vroman (2016), Guren (2018), and Rekkas, Wright and Zhu (2020).

time-series. Section 4 describes our model, theoretical results, and predictions. Section 5 tests our empirical predictions about liquidity supply, and Section 6 tests predictions about liquidity demand. 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. 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: 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 for 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:

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

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, γ_i, a county-month fixed effect, η_{ct}, a smooth function $f_c(x_i, t)$ of observable house characteristics x_i and time t, and a mean-0 error term ϵ_{it}.

Specification (1) combines elements of repeat-sales and hedonic models of house prices. The house fixed effect term, γ_i, absorbs all features of a house, observed and unobserved, which have time-invariant effects on the price of house i. The η_{ct} term absorbs parallel shifts in log house prices in a county over time. The $f_c(x_i, t)$ 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.

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 1a 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 A1 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 1b 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 A2 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 1a, 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 suggests 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 2, in which we plot the cross-section of time-on-market and logSD across counties. There is substantial dispersion in both liquidity measures. The mean of price dispersion is 18% of house prices, and the standard deviation is about 4%. For time-on-market, the mean is 3.2 months, and the standard deviation is 0.7.

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 measures are correlated with prices: high-price counties tend to be in the lower left quadrant of Figure 2, 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.

One interpretation of these facts is that liquidity in housing markets may not be a simple one-dimensional object; time-on-market and price dispersion may measure different aspects of price dispersion. To better understand these facts, we proceed to construct a search-and-bargaining model of the housing market.

4 Model

We now build a model, which combines a set of standard elements used in the housing search literature, and show how the model can be used to interpret the drivers of 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. Our core focus is comparing the effects of two different shifts in market primitives. When there are more buyers in the market, all sellers sell more quickly, and seller preferences affect prices less, so time-on-market decreases and relative price dispersion decreases; we call this *liquidity supply*. When sellers' holding costs are larger on average, time-on-market decreases, but relative price dispersion tends to increase: we call this *liquidity demand*.

The main contribution of the model is not technical; we use a basically standard random search model, with noncentered-exponentially distributed buyer value heterogeneity as in Guren and McQuade (2020), and persistent differences in sellers' types, as in for example Anenberg and Bayer (2020), Guren and McQuade (2020), Buchak et al. (2020), and Sagi (2021). The main technical departure from the literature is that we assume a continuous distribution of seller types; this does not qualitatively change the model outcomes, but allows us to cleanly characterize the derivative of the seller value function with respect to holding cost, showing how it depends on time-on-market and other primitives.

4.1 Setup

There is a unit mass of identical houses. Time is continuous, and all agents discount the future at rate r. There are three kinds of agents: sellers, buyers, and matched homeowners.

The lifecycle of agents is as follows: buyers exogenously choose whether to enter the market, purchase houses and become matched homeowners; matched homeowners receive separation shocks to become sellers; and sellers who successfully sell their houses leave the market.

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 normalized 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 $F(\cdot)$, which is a uniform distribution with support $[(\bar{c}-\Delta_c),(\bar{c}+\Delta_c)]$. Sellers with higher holding costs will tend to sell faster but at lower prices. The mean cost $\bar{c}$ plays an important role in our model, capturing what we call aggregate liquidity "demand" from sellers. When $\bar{c}$ is higher, holding costs are on average higher, implying that sellers will on average sell faster at the cost of incurring lower average prices. 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 $F(c)$.

There are many possible sources of heterogeneity in c. For example, 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). In the context of our model, we think of such sellers as having lower holding costs.[5]

Buyers. There is a mass M_B of potential buyers who are present in the market. We take M_B as exogenous for simplicity, and show how M_B could be endogenized in Appendix B.6.[6] M_B can be thought of as a measure of liquidity "supply": when M_B is high, sellers will match to prospective buyers quickly, allowing them to sell quickly and at stable prices.

All active buyers are identical, and receive flow utility normalized to 0 while waiting to

[5] Other papers have also implicitly or explicitly modelled persistent heterogeneity in sellers' effective holding costs: in Anenberg and Bayer (2020), agents who wish to sell and then buy in a city have different effective costs than sellers who wish to sell and exit the market entirely, and in Guren and McQuade (2020), lenders selling foreclosed houses have higher holding costs than market makers selling non-foreclosed houses. There are other possible sources. For example, since selling houses involves significant time and effort for owners, sellers with higher incomes may have higher values of c; also, 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, 2020).

[6] In Appendix B.6, M_B is the equilibrium outcome from a process in which potential buyers enter the market at some exogenous flow rate η_B and choose to remain active if the expected value from market participation is sufficiently high.

buy a house.[7] Buyers meet sellers through a matching process described later. When a buyer meets a seller, he draws, independently across matches, some idiosyncratic *match utility* ϵ from $G(\cdot)$, which is non-centered exponential, with lower bound ϵ_0 and standard deviation equal to σ_ϵ.[8] If the buyer buys the house, he becomes a matched homeowner, receiving ϵ from the house per unit time, until he receives a separation shock and becomes a seller.

Technically, the assumption about idiosyncratic match utility is a common model element in the literature (Krainer, 2001; Novy-Marx, 2009; Ngai and Tenreyro, 2014; Anenberg and Bayer, 2020; Guren and McQuade, 2020; Anundsen et al., 2023), and is necessary for the result that sellers with different holding costs have different average time-on-market. In particular, sellers with lower holding costs will tend to trade with a higher ϵ cutoff, waiting for higher-valued buyers, whereas sellers with higher holding costs will tend to sell to lower-valued buyers.

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 ϵ receives flow utility ϵ from their house per unit time. We use $G_{eq}(\epsilon)$ to denote the distribution of match utilities among matched homeowners, which will in general differ from $G(\epsilon)$. At Poisson rate λ_M, homeowners receive separation shocks and become sellers. These shocks can be thought of as life events which cause homeowners to have to change location. Similar assumptions are made in many papers, including Krainer (2001), Ngai and Tenreyro (2014), and Guren and McQuade (2020).

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 ϵ. If the buyer purchases at price P, the buyer receives $V_M(\epsilon) - P$, and the seller leaves the market and receives P. If the buyer does not purchase, the buyer receives V_B and the seller

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

[8] The noncentered exponential distribution is also used in Guren and McQuade (2020); this is analytically convenient, since it makes the level of price variance induced by buyer value dispersion does not depend on seller patience and other market conditions, technically simplifying the model substantially.

receives $V_S(c)$. The bilateral match surplus from trade, as a function of ϵ and c, is thus:

$$V_M(\epsilon) - V_B - V_S(c) \tag{2}$$

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

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

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

$$P(c,\epsilon) = V_S(c) + \theta\left(V_M(\epsilon) - V_B - V_S(c)\right) \tag{3}$$

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

$$\epsilon^*(c) \geq \epsilon_0 \ \forall c \tag{4}$$

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

In our 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, 2023). We abstract away from these other factors for simplicity; they can be thought of as adding additional error terms to the price equation (3) 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 λ_B and λ_S, given by:

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

Our baseline model is thus a random search model; we show in Appendix B.8 that we are able to partially generalize some of our results to a broader class of models.

4.2 Equilibrium

In our model, stationary equilibrium requires two sets of conditions to be satisfied: the decisions of 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:*

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

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

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

$$rV_S\left(c\right) = -c + \lambda_S \int_{\epsilon > \epsilon^*(c)} \theta \big(V_M\left(\epsilon\right) - V_B - V_S\left(c\right)\big)\, dG\left(\epsilon\right) \tag{6}$$

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

Trade cutoffs:

$$V_M\big(\epsilon^*\left(c\right)\big) = V_S\left(c\right) + V_B \tag{8}$$

Matching rates:

$$M_S\lambda_S = M_B\lambda_B = \alpha M_B^{\phi} M_S^{1-\phi} \tag{9}$$

Flow equality:

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

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

The following claim characterizes time-on-market and price dispersion in the model.

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

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

Expected time-on-market across all seller types is thus the expectation of (12) with respect to the seller cost distribution $F(c)$. The dollar variance of prices $P(c,\epsilon)$ among trading sellers and buyers is:

$$Var\left(P(c,\epsilon)\right) = \underbrace{Var_{c\sim F(\cdot)}\left(V_S(c)\right)}_{Seller\ \ component} + \underbrace{\left(\frac{\theta\sigma_\epsilon}{r+\lambda_M}\right)^2}_{Buyer\ \ component} \tag{13}$$

The relative standard deviation of prices is:

$$\frac{\sqrt{Var_{c\sim F(\cdot)}\left(V_S(c)\right) + \left(\frac{\theta\sigma_\epsilon}{r+\lambda_M}\right)^2}}{E[P]} \tag{14}$$

Time-on-market. Expression (12) of Claim 1 states that TOM depends on two factors: λ_S, the rate at which sellers meet buyers; and $1-G\left(\epsilon^*(c)\right)$, the probability that a seller meets a buyer with match quality higher than the matching cutoff $\epsilon^*(c)$. The meeting rate λ_S depends on the ratio of buyers to sellers, through (9):

$$\lambda_S = \alpha\left(\frac{M_B}{M_S}\right)^\phi$$

Hence, when the mass of buyers M_B increases, TOM will tend to decrease. The match quality cutoff $\epsilon^*(c)$ is lower for sellers with higher costs c: hence, an overall increase in sellers' costs will cause a larger fraction of buyer meetings to result in trade, which tends to decrease TOM.

Price dispersion. Expression (13) shows that dollar price variance can be decomposed into terms attributable to buyers' and sellers' values: due to the Nash bargaining structure of our model, buyers with higher values will pay higher prices, and sellers with higher costs will attain lower prices. The buyer component of price dispersion, $\left(\frac{\theta\sigma_\epsilon}{r+\lambda_M}\right)^2$, depends on bargaining power θ, the dispersion of match values σ_ϵ, the discount rate r, and the rate at which matched owners receive separation shocks λ_M. We take all these parameters as exogenous and do not vary them in our comparative statics.[9]

The seller value component, $Var_{c\sim F(\cdot)}\left(V_S\left(c\right)\right)$, plays a major role in our analysis. It is driven by how much dispersion there is in sellers' flow holding costs c, as well as the slope $V_S'\left(c\right)$), which governs how much the values of sellers with different costs differ.[10] We characterize this slope in the following claim.

Claim 2. *We have:*

$$V_S'\left(c\right) = \frac{-TOM\left(c\right)}{rTOM\left(c\right)+\theta} \tag{15}$$

where $TOM\left(c\right)$ is expected time-on-market for a seller of type c. $\left|V_S'\left(c\right)\right|$ is strictly increasing in $TOM\left(c\right)$, holding fixed r and θ.

Claim 2 states that when TOM is higher, a given difference in holding costs c leads to a bigger difference in sellers' attained values $V_S'\left(c\right)$ and thus the seller component of dollar price variance in 13. Intuitively, this is because differences in holding costs matter more when TOM is higher: if TOM is low, all sellers sell quickly, so differences in holdings costs matter less for sellers' values and prices. Thus, Claim 2 and Expression (13) imply that dollar price dispersion will tend to be higher when time-on-market is higher, holding fixed buyer value dispersion and seller cost dispersion. Finally, note that we empirically measure *normalized*

[9] Technically, we need to include dispersion in buyers' ϵ values not because this contributes to price dispersion, but because this assumption creates a nontrivial time-on-market "menu" for sellers. If all buyers had the same value, then all types of sellers who trade at all will trade with the first buyer they meet, so time-on-market will not depend on holding costs. When buyers are differentiated, impatient sellers can adopt lower $\epsilon^*\left(c\right)$ thresholds, lowering their time-to-sale in exchange for lower prices.

[10] The assumption of a continuous distribution of seller values, which is not common in the literature, is helpful because it allows us to think about the derivative $V_S'\left(c\right)$, rather than thinking about differences of values between discrete groups of sellers with different holding costs, as in Anenberg and Bayer (2020) and Guren and McQuade (2020).

price dispersion, which is also affected by the price level: holding dollar price dispersion fixed, forces that tend to increase prices tend to decrease normalized price dispersion.

Having in hand the model variables which influence time-on-market and price dispersion, we proceed to analyze the core liquidity "supply" and "demand" counterfactuals.

4.3 The Supply and Demand for Liquidity

The core idea in our paper is that TOM and PD can be thought of like equilibrium "quantity" and "price" outcomes from a liquidity supply and demand system, where we think of liquidity as "supplied" by buyers flowing into the market, and "demanded" by sellers who wish to sell their houses. Liquidity supply corresponds to the buyer mass M_B. When M_B increases, there are two effects, as illustrated by Panel (a) of Figure 3. First, time-on-market decreases because the rate at which sellers meet buyers (λ_S) increases, and thus, sellers sell faster. All else fixed, the decrease in TOM tends to decrease dollar price dispersion, through Expression (13) and Claim 2. Second, average prices increase because when there are more buyers, sellers match faster and expect to find buyers with higher ϵ values. The price increase tends to decrease relative price dispersion, since the denominator of (14) decreases. These forces imply that, when M_B increases, time-on-market and price dispersion should both unambiguously decrease. An informal intuition for this result is that, when M_B increases, sellers can sell both faster and at higher prices; thus, in equilibrium, both TOM and PD will tend to decrease.

Liquidity demand corresponds to $\bar{c}$, the average seller holding cost. Panel (b) of Figure 3 illustrates the effects as liquidity demand increases. When $\bar{c}$ increases, time-on-market decreases, since sellers are more impatient on average and attempt to sell faster. The effects on price dispersion are more subtle, involving two channels. First, since time-on-market is lower, expression (13) and Claim 2 imply that price dispersion level will tend to decrease. Second, sellers' impatience tends to decrease the level of prices, which decreases the denominator in (14), pushing relative price dispersion to increase. As a result of these channels, when liquidity demand increases, time-on-market will tend to decrease, but relative price dispersion may in fact increase. An informal intuition is that, if sellers become more impatient and liquidity demand increases, sellers can be thought of as choosing to sell faster, but incurring lower prices as a result, thus lowering TOM but potentially increasing price dispersion. Price dispersion is more likely to increase with liquidity demand when prices decrease more; moreover, prices decrease more with liquidity demand when buyer value dispersion $\sigma_{epsilon}$ is larger.

The model sheds light on the stylized facts we documented in the previous section. In the time series, at both seasonal and business-cycle frequencies, both liquidity measures co-move positively with each other, and negatively with prices. This is consistent with time-series movements being driven primarily by liquidity supply shifts: aggregate buying pressure may be higher in summer and in boom times, driving both time-on-market and price dispersion to decrease. In contrast, counties may experience different liquidity supply and demand shocks, which may contribute to explaining why the correlation in time-on-market and price dispersion is weak in the cross-section.

The distinction between liquidity supply and demand-driven variation is important because the two kinds of shocks have distinct effects on *prices*. An informal intuition across asset classes is that liquidity is priced; therefore, an improvement to market liquidity should be associated with an increase in asset prices. Figure 3 shows that this intuition can fail if housing market liquidity is measured purely using time-on-market, ignoring price dispersion: if time-on-market decreases because liquidity *demand* increases, then prices will tend to decrease. In such cases, time-on-market alone is a misleading measure of liquidity: time-on-market is lower not because markets are more liquid, but because sellers have higher holding costs, and thus demand more of available liquidity. Price dispersion plays a potentially useful role in distinguishing these cases, since if time-on-market decreases are driven by liquidity demand increases, price dispersion may increase.

The model generates two testable predictions to bring to the data: proxies for liquidity supply and liquidity demand should have different correlation patterns with time-on-market, price dispersion, and the level of prices.

Prediction 1. *Suppose a variable Z_i is a proxy for liquidity supply. Z_i should be negatively correlated with price dispersion and time-on-market, and positively correlated with prices.*

Prediction 2. *Suppose a variable Z_i is a proxy for liquidity demand. Z_i should be negatively correlated with time-on-market and prices, but may be positively correlated with price dispersion.*

Note that the implications of Prediction 2 for price dispersion are somewhat subtle: price dispersion is an imperfect metric for liquidity demand, since increases in liquidity demand can cause price dispersion to decrease, because liquidity demand increases create two forces pushing PD in different directions. Essentially, if liquidity demand increases have a very large effect on prices, price dispersion will tend to increase as liquidity demand increases. The price effects of liquidity demand increases tend to be larger when σ_ϵ, the dispersion in buyer values, is larger. Note, however, Prediction 1 unambiguously states that liquidity supply measures should be negatively correlated with price dispersion in our model: both the dispersion and price-level channels push towards lower price dispersion. Thus, if we observe that a proxy for increased liquidity demand is associated with lower time-on-market and higher price dispersion, this cannot be driven by liquidity supply shifts in our model.

Finally, we briefly discussed trade volumes in Section 3 because volume plays a role in many papers in the housing microstructure literature (Clayton, Miller and Peng, 2010; Diaz and Jerez, 2013; Ngai and Tenreyro, 2014; Ngai and Sheedy, 2020; Anenberg and Bayer, 2020; Bayer et al., 2020; DeFusco, Nathanson and Zwick, 2022). We do not emphasize trade volumes as an output of our model, because volumes are not useful in distinguishing between supply and demand shocks in our setting. In our model, volumes tend to co-move in the same direction as time-on-market: volumes increase when either liquidity supply or demand increase. Our model is also technically not well-suited to analyzing trade volumes. Since we use a stationary-equilibrium model, we cannot capture changes in volume that come from inventory accumulation or depletion, which may be important factors driving volume fluctuation in reality. Our model also does not capture other important drivers of trade volumes, such as propensity of joint buyer-sellers (Anenberg and Bayer, 2020), or speculators (Bayer et al., 2020; DeFusco, Nathanson and Zwick, 2022).

5 Empirical Evidence for Liquidity Supply Effects

In this section, we provide empirical evidence for Prediction 1 in the cross-section by estimating the following specifications for county-level dependent variables Y_c equal to price dispersion (logSD), time-on-market (TOM), and house price growth (Log Price Growth):

$$Y_c = Z_c^S \alpha + X_c \beta + \epsilon_c \tag{16}$$

where Z_c^S is a county-level liquidity supply shifter, and X_c is a vector of controls.

We begin by discussing the forces that affect liquidity supply, and presenting the baseline results using population growth as an empirical liquidity supply shifter. We then introduce a novel instrumental variable (IV) that captures the exogenous variation in population migration. In presenting the IV results, we provide empirical evidence about migration flows and the demographic composition of migration to support IV relevance.

5.1 Baseline

In our model, liquidity supply represents buying pressure: we think of M_B as high in cities where there are many buyers, relative to the available housing stock. To capture this concept, we use the population growth rate in a county as a county-level liquidity supply shifter. The idea is that high population growth should imply a high inflow rate of buyers, which we interpret as liquidity supply M_B being high in our model.

Before presenting the estimation results, we make non-parametric plots of price dispersion (logSD), time-on-market (TOM), and transaction price against our liquidity supply shifter, population growth, in Figure 4. Consistent with our model prediction, both logSD and TOM are negatively correlated with liquidity supply, while price increases with liquidity supply.

Panel A of Table 2 shows the results of Equation 16, where Z_c^S is measured using population growth rates. In all columns, 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 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, and unemployed. Both logSD and time-on-market are negatively associated with local population growth, while house price growth is positively correlated with local population growth. A 1% population growth is associated with a 0.3% decrease in price dispersion, a 0.08 month

(2.4 day) decrease in time-on-market, and a 0.62% increase in house price growth. Thus, the results are consistent with Prediction 1 that higher liquidity supply leads to smaller price dispersion, shorter time-on-market, and higher price.

5.2 IV for Population Growth

Population growth is affected by a large number of variables, so it may not be an exogenous shifter of liquidity supply. To overcome this endogeneity issue, we use an empirical approach based on Schubert (2021) to find plausibly exogenous shocks to liquidity supply: productivity shock spillovers through historical migration links.

Intuitively, suppose, for example, Cook County experiences large productivity shocks, which increase house prices. The higher living expenses in Cook county may force high-skilled workers to move to other regions (Diamond and Gaubert, 2022). This will tend to create migration inflows to counties with strong historical migration links to Cook County. These migration flows are thus plausibly exogenous shocks in counties connected to Cook County by migration.

We construct this IV for any given county c in three steps. We first measure historical migration links between county c and any other counties in the nation c' using ACS 2008-2012 5-year data sample. We find the fraction of migrants to county c that came from county c' from 2008 to 2012 and denote it by $\mu^{c \leftarrow c'}_{2008-2012}$.[11]

We then find plausibly exogenous local wage "Bartik" shocks in all counties during the period of interest, 2012-2016. Denote the shock in county c during this period by $B_{c,2012-2016}$. We construct $B_{c,2012-2016}$ by combining local employment shares in every 2-digit NAICS industry in 2010 with national wage growth rate in that industry over 2012-2016:

$$B_{c,2012-2016} = \sum_i \omega_{c,i,2010} \Delta \ln W_{-c,i,2012-2016} \tag{17}$$

where $\omega_{c,i,2010}$ is the share of workers in industry i in county c in 2010, and $\Delta \ln W_{-c,i,2012-2016}$ is the national wage growth rate in industry i over 2012-2016. We fix industry exposure shares

[11] 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'.

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.

Finally, we construct our liquidity supply shifter for county c during the period of interest, $M_{c,2012-2016}$, by weighting the shock in every other county by its historical migration link with county c and summing over these shocks:

$$M_{c,2012-2016} = \sum_{c'} \mu_{2008-2012}^{c \leftarrow c'} B_{c',2012-2016} \tag{18}$$

Note that we choose to use migration spillovers ($M_{c,2012-2016}$) instead of local Bartik shocks as our instrument because local shocks are likely associated with or caused by changes in the underlying economic conditions in the local areas that simultaneously drive the changes in the housing market, whereas our instrument is less likely to be subject to such an issue, which we will discuss in detail later. Another reason is related to the demographic composition of migration flows between the shocked and the historically linked counties, and which type of population growth is more likely to affect the housing demand. According to Diamond and Gaubert (2022), population growth driven by local productivity shocks tends to concentrate among low-income, while high-income earners move from areas with increased living expenses to other regions, contributing to the population growth in those regions. Since high-income earners are more likely to be homeowners, the population growth driven by influx of high-income earners is more relevant for our analysis.[12]

IV Relevance. We provide empirical evidence in Table 3 to show migration flows between historically linked counties induced by the Bartik shocks. We construct a sample of county-to-county pairs and regress the migration flow to county c from any other county c' during 2012-2016 on shocks in county c' ($B_{c',2012-2016}$), the historical link between the two counties ($\mu_{2008-2012}^{c \leftarrow c'}$), and the interaction between $B_{c',2012-2016}$ and $\mu_{2008-2012}^{c \leftarrow c'}$. In column 1, we control for county c's local productivity shocks $B_{c,2012-2016}$. In column 2, we further control for local county c fixed effects, while in column 3, we add paired county c' fixed effects to column 2's

[12] Although we use migration spillovers as our main instrument, all the predictions of our model still hold if we simply use regular Bartik shocks, $B_{c',2012-2016}$, directly as proxies for liquidity supply; Table A2 repeats the specifications in Table 2, Panel B using $B_{c',2012-2016}$ as our IV instead of $M_{c,2012-2016}$, and all coefficients are statistically significant with the expected signs.

specification.

The estimated coefficient on paired county shock ($B_{c',2012-2016}$) is positive and significant, suggesting that a larger number of people migrate from county c' to other counties when county c' experiences productivity shocks. More importantly, the estimated coefficient on the interaction term is positive and statistically significant across all specifications. Thus, there are more migrants to its historically linked counties after local shocks. When a county experiences a one standard deviation increase in productivity, there are 0.02% local residents migrating to counties without any historical link. A one standard deviation increase in historical link between the shocked county and the other county is associated with a 1.83-2.99% higher number of migrants to the other county. Furthermore, Figure 5 shows that high-income households are more likely to migrate from a county that experiences productivity shocks to its historically linked counties, consistent with the migration sorting documented in Diamond and Gaubert (2022).[13]

In the aggregate, our IV appears relevant for county migration inflows: column 4 of Table 3 regresses total gross migration inflows, normalized by county population, on our migration spillover instrument $M_{c,2012-2016}$, and the coefficient is positive and significant. Finally, we confirm that our IV is indeed associated with net population growth, after accounting for migration outflows: column 1 of Table 2, Panel B, shows that a one standard deviation increase in migration spillovers is associated with a statistically significant 0.81% increase in population growth rate.

Exclusion Restriction. For exclusion, we show that the instrument is uncorrelated with local economic conditions that could affect housing market tightness (Goldsmith-Pinkham, Sorkin and Swift, 2020). We perform a balance test in Figure A8. We regress the migration spillover instrument $M_{c,2012-2016}$ on a set of local characteristics, including housing stock conditions and local demographics as well as economic conditions. The instrument is not statistically significantly correlated with these characteristics.

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

[13]Since ACS 2012-2016 migration data no longer collects demographic information, we change to estimate productivity shock using years before 2007 and use 2007-2011 ACS data to analyze migration flows by demographics.

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.

5.3 IV Results

Panel B of Table 2 presents the results of the following IV specification:

$$\begin{aligned} \text{First Stage: } Z_c^S &= \alpha_1 M_c + B_c\beta_1 + X_c\gamma_1 + \epsilon_c \\ \text{Second Stage: } Y_c &= \alpha_2 \hat{Z}_c^S + B_c\beta_2 + X_c\gamma_2 + \epsilon_c \end{aligned} \tag{19}$$

where we use migration spillover (M_c) to instrument for population growth and control for $B_{c,2012-2016}$ and the same set of local controls. Column 1 shows the first stage result, suggesting that our instrument drives population growth in a statistically and economically significant manner. Columns 2-4 show the second stage results for logSD, TOM, and house price growth, respectively. The estimated effects of our liquidity supply shifter on both price dispersion and time-on-market are negative, supporting our model predictions. Also consistent with our theory's prediction, liquidity supply is positively correlated with house price growth. Quantitatively, a 1% increase in migration spillover induced population growth is associated with 1% lower price dispersion, 0.39 month (about 12 days) lower time-on-market, and 2% higher house price growth. In columns 5-7, we confirm the results in reduced-form IV specifications in the last three columns, where we regress logSD, TOM, and house price growth directly on M_c. Together, the results in Table 2 support Prediction 1: increases in liquidity supply are associated with increases in prices, and decreases in both time-on-market and price dispersion.

While we use prices as a dependent variable in Table 2 for completeness, the idea that these migration spillovers move house prices is not new to the literature, being the focus of Schubert (2021). Our contribution is to show that this migration shock also moves PD and TOM in the directions predicted by our model, a fact we believe is new to the literature.

Giacoletti (2021) has some results related to our liquidity supply analysis, finding that idiosyncratic price dispersion is associated with a few measures of general market "illiquidity": more atypical properties, and zipcodes with a more heterogeneous housing stock, have lower price dispersion, and also that expansions in credit availability are associated with decreases in price dispersion. Increases in buyer credit availability can be thought of as a liquidity supply shock in the context of our model: we might think of this as either an increase in the mass of willing buyers M_B, or an upwards shift in the distribution of buyers' WTP, which has similar effects. Changes in housing stock homogeneity are somewhat harder to map directly into our model; we can think of this as driven by "thick market effects" increasing matching rates, as in Ngai and Tenreyro (2014), or in reduced-form as a decrease in buyer value dispersion σ_ϵ. Both effects would decrease price dispersion in our model, but map somewhat less cleanly to our notion of liquidity supply.

6 Empirical Evidence for Liquidity Demand Effect

Prediction 2 states that increases in liquidity demand should always decrease TOM and prices, but may actually increase price dispersion. The magnitude of the effect on prices plays an important role: in settings where liquidity demand shocks decrease prices relatively little, price dispersion may tend to decrease with liquidity demand; if liquidity demand shocks decrease prices substantially, price dispersion will tend to increase with liquidity demand. To empirically test this prediction, we estimate the following specification separately for subsamples of transactions that are likely to have different levels of price discount induced by liquidity demand:

$$Y_{z,t} = Z^D_{z,t}\alpha + X_{z,t}\beta + \mu_{c,t} + \epsilon_{z,t} \tag{20}$$

where z indicates zipcodes, t indicates transaction month, and c indicates the county that zipcode z belongs to. $Y_{z,t}$ is our outcome variable of interest: price dispersion (logSD), time-on-market (TOM), and house price (log Price). $Z^D_{z,t}$ is our liquidity demand shifter, which we will introduce below. $X_{z,t}$ is a set of zipcode-month level controls. $\mu_{c,t}$ is county-year fixed effects.

In our model, increase in liquidity demand leads to a larger price discount when buyer valuations are more dispersed. To capture this idea, we construct subsamples based on whether a transaction is labeled as a distressed sale in the Corelogic Tax and Deed database.[14] The idea is that the pool of potential buyers in a distressed sale is small and buyers may have very different valuations due to information frictions, thereby liquidity demand presumably leads to larger price discounts for such transactions. We will verify this assumption later in the analysis.

We begin by introducing our liquidity demand shifter (Z_z^D). We then present the results and discuss identification assumptions.

6.1 Measurement and Instrument

We use zipcode seller home equity value as a liquidity demand shifter. The literature has documented that sellers with less home equity tend to set higher prices, waiting longer to sell as a result (Genesove and Mayer, 1997, 2001; Guren, 2018). In the context of our model, we interpret these sellers as having lower holding costs and thereby less liquidity demand.[15]

We construct seller's home equity ratio for every property that was transacted more than once in our sample period and aggregate it to zipcode-month level. We need two pieces of information to find seller's home equity ratio: the value of the property and seller's outstanding mortgage balance. For each property, we obtain the following pieces of information associated with the last transaction from Corelogic Tax and Deed: the date, the property price, the initial mortgage amount, and the mortgage term. To determine the value of the property, we multiply the property's last transaction price by its value appreciation since last transaction. We use zipcode price appreciation to proxy for the value appreciation of each individual property in that zipcode. Therefore, the value of property i located in zipcode z in year T is

$$\tilde{P}_{i,T} = P_{i,t} \times \Delta_{z,t,T}, \tag{21}$$

where $P_{i,t}$ is the previous transaction price, and $\Delta_{z,t,T}$ is zipcode price appreciation from year t to year T calculated using Zillow zipcode price index: $\Delta_{z,t,T} = \frac{P_{z,T}}{P_{z,t}}$.

[14] Note that most of these distressed sales are not foreclosures but shortsales, hence seller equity position still matters, which we will provide empirical evidence for later.

[15] We thank an anonymous referee for suggesting this measure.

To determine the seller's outstanding mortgage balance, we estimate the monthly payment and then find the outstanding balance in year T according to the amortization schedule. In particular, we use the initial balance, loan term, and the 30-year mortgage rate in the origination year to estimate the monthly payment:

$$payment_i = loan_{i,t} \times \frac{\bar{r}_t(1+\bar{r}_t)^{N_{i,t}}}{(1+\bar{r}_t)^{N_{i,t}-1}}, \tag{22}$$

where t denotes the origination year, $loan_{i,t}$ is the total loan amount, $\bar{r}_t$ is the 30-year mortgage rate in the origination year,[16] and $N_{i,t}$ is the number of remaining months to maturity. We estimate the remaining balance in year t using the following formula:[17]

$$loan_{i,T} = \frac{payment_i}{\bar{r}_t} \times (1 - \frac{1}{(1+\bar{r}_t)^{N_{i,T}}}) \tag{23}$$

For home purchases paid in cash, $loan_{i,T}$ is zero.

We find the home equity ratio by dividing the seller's equity stake of the property after paying off the debt by the previous transaction price.

$$HomeEquityRatio_{i,T} = \frac{\tilde{P}_{i,T} - loan_{i,T}}{P_{i,t}} \tag{24}$$

Finally, to obtain a liquidity demand shifter at zipcode level, we take the average home equity ratio of all individual properties that are transacted in a given month.

Instrument. There are three sources of variation in our liquidity demand shifter: the timing of the previous transactions of listed houses in a given zipcode-month, zipcode house price growth during the holding period, and the initial loan-to-value ratios of the listed houses. While we can control for zipcode characteristics and include county-year fixed effects, which absorb variation in housing market conditions across counties over time, there may still be concerns surrounding endogeneity due to local zipcode price changes. Such endogeneity could downward bias our estimation, especially in the analysis regarding prices.

[16] The data can be downloaded from here: https:www.freddiemac.compmmspmms30

[17] This approximation does not account for accrued interests due to late payments.

As an attempt to address this concern, in a manner similar to papers such as Guren (2018) and Huang, Nelson and Ross (2021), we find county-level house price growth from the previous transaction date to the current transaction date of every listed property and aggregate it to zipcode-month level to instrument for zipcode home equity:

$$IV^D_{z,t} = \sum_{i \in (z,t)} \frac{P_{c,t}}{P_{c,\tau(i)}} \tag{25}$$

where z is zipcode, t is month, i is property, c is the county that zipcode z belongs to, and $\tau(i)$ is the previous transaction month of property i. The inclusion of county-year fixed effects absorbs variation in yearly county house price growth; identification thus comes from differences in average historical transaction timing between zipcodes within a county. Intuitively, suppose that a county experienced a house price boom in 2011; suppose zipcode A happened to have many trades occur in 2010, whereas zipcode B had more trades in 2012. Current house sellers in zipcode A will then on average have more home equity, having more exposure to the house price boom. Since the driver of these differences is the historical timing of transactions, and houses transact fairly infrequently, this IV is also plausibly less confounded by recent market conditions than if we used home equity directly.

In the following section, we estimate the following specifications for price dispersion, time-on-market, and transaction prices:

$$\begin{aligned} \text{First Stage: } Z^D_{z,t} &= \alpha_1 IV^D_{z,t} + X_{z,t}\beta_1 + \mu_{c,t} + \epsilon_{z,t} \\ \text{Second Stage: } Y_{z,t} &= \alpha_2 \hat{Z}^D_{z,t} + X_{z,t}\beta_2 + \mu_{c,t} + \epsilon_{z,t} \end{aligned} \tag{26}$$

Table A3 shows the first stage estimation results, suggesting that county house price growth is a strong predictor of zipcode seller home equity. Moreover, we present the correlation between our transaction-level liquidity demand shifters and property and zipcode characteristics in Figure A9. Not surprisingly, properties with higher seller home equity and properties with lower seller home equity tend to locate in zipcodes with different economic conditions. But our instrument is not correlated with property and local characteristics, with property age being the only exception. In the second stage estimation, we include property age group fixed effects to account for this correlation.

6.2 Results

Before presenting the estimation results, we make non-parametric plots of price dispersion (logSD), time-on-market (TOM), and transaction price against our liquidity demand shifter in Figure 6. Consistent with our model prediction, logSD is positively correlated with liquidity demand, while TOM is negatively correlated with liquidity demand. Moreover, prices decline with our liquidity demand shifter.

Table 4 presents the estimation results. Columns 1-3 focus on distressed sales. Columns 4-6 focus on non-distressed sales. Columns 7-9 use the full sample while we add an indicator for whether the observation is from the sample of distressed sales in the regression.

Comparing two zipcodes with similar characteristics transacted in the same county-year, the one with more sellers that have higher average home equity, and thus higher liquidity demand, tends to have larger price dispersion (column 7), shorter time on market (column 8), and lower transaction prices (column 9). The magnitudes of these correlations are nontrivial: a 1% increase in home equity ratio is associated with approximately a 1% increase in price dispersion, a 0.9 month decrease in time-on-market, and a 9.85% decrease in transaction prices.

The magnitude of the liquidity demand effect on prices plays an important role. In the sample of distressed sales, higher liquidity demand significantly lowers prices (column 3), the magnitude of which is more than twice the effect in the sample of non-distressed sales (column 6). Consistent with Prediction 2 of the model, when liquidity demand shocks decrease prices relatively little, price dispersion decreases with liquidity demand (column 4); and when liquidity demand shocks decrease prices substantially, price dispersion increases with liquidity demand (column 1). In contrast, liquidity demand always decrease time-on-market (columns 2 and 5).

Our results in this section thus suggest that higher home equity is associated with lower TOM, but actually lower prices. Price dispersion, while also an imperfect metric, conveys some incremental information. We show that on average in our sample, zipcodes with higher liquidity demand proxy in fact have higher price dispersion, and this is driven primarily by distressed sales, where the price impact of liquidity demand appears higher. While we do not have enough evidence to take a strong stance as to why liquidity demand affects prices more for distressed sales relative to non-distressed sales, one possibility is that the dispersion in buyer values – σ_ϵ in the model – is larger for distressed sales: non-distressed sales may tend to have at least a few interested buyers with similar values, whereas distressed sales may happen at times and market conditions in which it is more difficult to find buyers with reasonable values in a short period of time. This provides some evidence for the idea that PD may convey incremental information over TOM in some cases, helping to identify markets in which TOM decreases are driven by liquidity demand rather than supply.

Note that the empirical associations between home equity, time-on-market, and house prices have also been documented in the literature (Genesove and Mayer, 1997, 2001; Guren, 2018). Our key new empirical finding is the relation between home equity and price dispersion, and how this relation varies by distressed and non-distressed sales. We still demonstrate all these facts using our sample to show that, thinking of home equity as a proxy for liquidity demand, home equity associates with price dispersion, time-on-market, and prices in the directions predicted by our model.

7 Quantification and Implications

In this section, we calibrate our model, to show how movements in TOM can be misleading about the direction prices move, if TOM is used as a single-dimensional measure of liquidity.

7.1 Calibration Methodology

We calibrate the model to average US-level data in 2016. We externally calibrate the discount rate r, bargaining parameter θ, matching function elasticity ϕ, and the match efficiency parameter α. We then calibrate the remaining parameters to match moments: the rate at which matched homeowners become sellers λ_M, the buyer mass M_B, the parameters of seller holding cost distribution, and the parameters of the buyer match value exponential distribution $\epsilon_0, \sigma_\epsilon$.

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 $\theta = 0.5$, following Anenberg and Bayer (2020) and Arefeva (2019). We choose the matching function elasticity ϕ to be 0.84, which is estimated by Genesove and Han (2012) and also used in Anenberg and Bayer (2020). Since we do not observe buyer time-on-market, we cannot separately identify the matching efficiency parameter α and the inflow rate of buyers M_B, so we normalize the match efficiency parameter to $\alpha = 1$.

Moment matching. The remaining parameters in our model are the rate at which matched homeowners become sellers λ_M, the buyer mass M_B, the seller holding cost distribution $U[(\bar{c} - \Delta_c), (\bar{c} + \Delta_c)]$, and the parameters of the buyer match value distribution $(\epsilon_0, \sigma_\epsilon)$. First, we choose $\lambda_M, M_B, \bar{c}, \epsilon_0, \sigma_\epsilon$ to match target moments. From our data we calculate the sales-weighted sample averages of house price (Zillow's ZHVI) and time-on-market (Zillow), and the turnover rate, which is calculated as the total house sales as a fraction of the total housing stock (Corelogic deed and tax). We also target the average number of houses that buyers visit before buying, estimated to be 9.96 by Genesove and Han (2012). Finally, we target the buyer component of price dispersion. Anundsen et al. (2023) estimate a model of house price dispersion using bidding data from Norwegian housing auctions, and find that the log standard deviation of buyers' willingness-to-pay for houses in their estimated model is 4.65%.[18] We use this as a target moment, requiring the square root of the buyer component of price variance – that is, the square root of the right term in (13) – to be 4.65% of expected prices.[19] While it is not ideal to use a dispersion estimate from Norwegian data, we are unaware of publicly available bid data on housing auctions in the US, making it much more difficult to estimate buyer willingness-to-pay dispersion in the US setting.

Next, we choose Δ_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, Δ_c, translates into dispersion in sellers' continuation values, $Var(V_S(c))$, which in turn translates into dispersion in prices, $Var(P)$. Hence, the empirical relationship between time-on-market and price dispersion is informative about Δ_c. Since time-series changes in prices, logSD, and time-on-market are likely to be driven

[18] See Table 3 of Anundsen et al. (2023).

[19] We show this in expression (71) of Appendix D.1.1.

by changes in liquidity supply, we perturb the liquidity supply parameter in our model, M_B, around the equilibrium to generate the model-implied relationship between time-on-market and price dispersion. Then, we calibrate Δ_c such that the model-implied relationship between price dispersion and time-on-market matches the empirical coefficient. We describe our moment matching procedure in detail in Appendix D.1.

Table 5 shows our estimated parameter values. We estimate that match quality ϵ has a lower bound of \$739 monthly and a standard deviation of \$216 monthly. The average value of ϵ among successful buyers is \$1,431 monthly. The mean value $\bar{c}$ is equal to \$4,700 per month, and the range parameter Δ_c is equal to \$3,864. We then calculate that buyers' average value from homeownership as the integrated flow value of match utility ϵ, until buyers receive a separation shock, to be \$164,379. Similarly, we calculate that sellers' average total loss from keeping their houses on the market is equal to \$9,510.

Liquidity discounts. As an untargeted moment to evaluate model fit, we estimate *liquidity discounts* in our model: the time-on-market difference between impatient sellers and patient sellers and the resulting price difference. Comparing sellers with 75th and 25th percentiles of holding costs, our calibrated model suggests that the 75th percentile sellers take 1.56 more months to sell and achieve \$16,843 higher prices – or 6.31pp higher prices, as a percentage of average prices. Diving by the time-on-market gap, the implied effect of spending an extra month on the market is, therefore, 4.03pp higher prices.

These estimates are roughly consistent with liquidity discount estimates from the housing microstructure literature, as shown in Table A4 which collects a number of papers analyzing the relationship between time-on-market and average prices, using various shifters of sellers' patience.[20]

Robustness. Our calibration takes a number of strong stances on input parameters. In Appendix D.4, we analyze how our results differ when we vary three input parameters which

[20]We analyze two groups of papers. First, we look at a set of papers which analyze how various shifters of sellers' patience affect time-on-market and average prices; we calculate implied 1-month liquidity discounts using the estimates from each of these papers. Second, we look at studies analyzing discounts on foreclosed houses; these studies do not report the time-on-market difference between foreclosed and non-foreclosed houses, so we can only report the average foreclosure discount, and are unable to estimate one-month effects. Further details on how we arrive at these estimates are described in Appendix D.3.

are relatively difficult to take strong stances on: the bargaining power parameter θ; the target dispersion in buyers' values for houses; and the target average number of houses that buyers visit before buying. We show that liquidity discounts, and the effects of liquidity supply and demand shifts on outcomes, are moderately sensitive to variation in input parameters, meaning there is substantial quantitative uncertainty about these outcomes; however, in all our sensitivity checks, the signs of all effects are preserved. Moreover, our later results in this section are surprisingly insensitive to different alternative assumptions for inputs.

7.2 Prices, Time-on-Market, and Price Dispersion

Our model implies that TOM is an incomplete liquidity measure, since TOM decreases can be driven by either liquidity supply or demand variation, depending on whether price dispersion increases or decreases, and liquidity demand-driven decreases in TOM are associated with price *declines*. We now quantitatively demonstrate this using the calibrated model. Suppose a market experiences a certain-sized change in TOM and PD; we can uniquely back out the changes in liquidity supply and demand – M_B and $\bar{c}$ respectively – which rationalize this shift, holding fixed all other estimated parameters at their baseline values. We can then calculate the implications of these primitive shifts for prices.

Suppose we observe a market experiencing a 5% time-on-market decrease, and a 5% price dispersion decrease. Our model qualitatively predicts that this change should be mainly driven by liquidity supply; quantitatively, we find that this movement is explained by 15.53% increase in liquidity supply M_B, and a smaller 3.73% increase in liquidity demand $\bar{c}$. Since the change is driven mainly by liquidity supply increases, prices should rise: in the model, we find this shift is associated with a price increase of 3.37%. Conversely, suppose we see a market experiencing a time-on-market decrease of 5%, but a price dispersion *increase* of 5%. Qualitatively, this should be driven by an increase in liquidity *demand*; indeed, we find that this change is rationalized by a decrease in M_B of 22.69%, and an increase of 6.18% in $\bar{c}$. This has very different implications for prices: we find that this shift is associated with a price decrease of 6.48%. Thus, in our calibrated model, identical shifts in time-on-market can have very different implications for prices, and the direction that price dispersion moves is informative about this difference.

8 Conclusion

Liquidity is a seemingly intuitive concept which is notoriously hard to precisely define. Makower and Marschak (1938) suggest that liquidity is, "like the price level, a bundle of measurable properties." This paper formalizes this idea: in the context of housing markets, we propose that two metrics – time-on-market and price dispersion – can be thought of like "quantity" and "price" outcomes respectively in a supply and demand system for liquidity. Time-on-market is already commonly used as a component of housing liquidity measures, in industry and academic work; price dispersion has received comparatively little attention. We have shown that price dispersion behaves similarly to time-on-market in the relatively straightforward case where outcomes are driven by shifts in liquidity supply: TOM and PD move together seasonally and over the business cycle, and respond in the same direction to increased buying pressure from migration shocks. However, PD can in some cases increase, when outcomes are driven by proxies for increased liquidity demand. Empirically, and in our calibrated model, we find that prices also move in different directions, depending on whether a TOM increase is associated with an increase or decrease in PD.

Our work thus contributes to the academic understanding of what different metrics tell us about underlying housing market primitives, as well as applied efforts to construct simple and informative measures of housing market liquidity. More broadly, while we have focused on housing markets in the current paper, our insights about the equilibrium relationships between different liquidity measures may be applicable to other kinds of markets for illiquid assets.

References

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.

Amaral, Francisco. 2024. "Price Uncertainty and Returns to Housing." Working Paper.

Andersen, Steffen, Cristian Badarinza, Lu Liu, Julie Marx, and Tarun Ramadorai. 2022. "Reference dependence in the housing market." *American Economic Review*, 112(10): 3398–3440.

Anenberg, Elliot, and Patrick Bayer. 2020. "Endogenous sources of volatility in housing markets: The joint buyer–seller problem." *International Economic Review*, 61(3): 1195–1228.

Anundsen, André Kallak, Arne Lyshol, Plamen Nenov, and Erling Røed Larsen. 2023. "Match Quality and House Price Dispersion: Evidence from Norwegian Housing Auctions." Working Paper.

Anundsen, Andre K., Erling Roed Larsen, and Dag Einar Sommervoll. 2019. "Strategic 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 W Roberts. 2020. "Speculators and middlemen: The strategy and performance of investors in the housing market." *The Review of Financial Studies*, 33(11): 5212–5247.

Bayer, Patrick, Kyle Mangum, and James W Roberts. 2021. "Speculative fever: Investor contagion in the housing bubble." *American Economic Review*, 111(2): 609–651.

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. "Why is intermediating houses so difficult? evidence from ibuyers." National Bureau of Economic Research.

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, and Benjamin Williams. 2015. "The repeat time-on-the-market index."

Carrillo, Paul E, William M Doerner, and William D Larson. 2023. "House Price Markups and Mortgage Defaults." *Journal of Money, Credit and Banking*, 55(4): 747–782.

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

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.

Contat, Justin, and Malcolm Rogers. 2022. "Housing Supply and Liquidity in the COVID-19 Era." *Cityscape*, 24(3): 123–152.

DeFusco, Anthony A, Charles G Nathanson, and Eric Zwick. 2022. "Speculative dynamics of prices and volume." *Journal of Financial Economics*, 146(1): 205–229.

Diamond, Rebecca, and Cecile Gaubert. 2022. "Spatial sorting and inequality." *Annual Review of Economics*, 14: 795–819.

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.

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

Gabrovski, Miroslav, and Victor Ortego-Marti. 2021. "Search and credit frictions in the housing market." *European Economic Review*, 134: 103699.

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

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

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

Giacoletti, Marco. 2021. "Idiosyncratic risk in housing markets." *The Review of Financial Studies*, 34(8): 3695–3741.

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

Gilbukh, Sonia, and Paul S Goldsmith-Pinkham. 2023. "Heterogeneous real estate agents and the housing cycle." *Available at SSRN 3436797*.

Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift. 2020. "Bartik instruments: What, when, why, and how." *American Economic Review*, 110(8): 2586–2624.

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. 2020. "How do foreclosures exacerbate housing downturns?" *The Review of Economic Studies*, 87(3): 1331–1364.

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.

Huang, Weiran, Ashlyn Nelson, and Stephen Ross. 2021. "Foreclosure spillovers within broad neighborhoods." National Bureau of Economic Research.

Kluger, Brian D, and Norman G Miller. 1990. "Measuring residential real estate liquidity." *Real Estate Economics*, 18(2): 145–159.

Krainer, John. 2001. "A theory of liquidity in residential real estate markets." *Journal of urban Economics*, 49(1): 32–53.

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

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.

Makower, Helen, and Jacob Marschak. 1938. "Assets, prices and monetary theory." *Economica*, 5(19): 261–288.

Ngai, L Rachel, and Kevin D Sheedy. 2020. "The decision to move house and aggregate housing-market dynamics." *Journal of the European Economic Association*, 18(5): 2487–2531.

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 Stroebel. 2020. "Segmented housing search." *American Economic Review*, 110(3): 720–759.

Rekkas, Marie, Randall Wright, and Yu Zhu. 2020. "How Well Does Search Theory Explain Housing Prices?" *Available at SSRN 3706329*.

Sagi, Jacob S. 2021. "Asset-level risk and return in real estate investments." *The Review of Financial Studies*, 34(8): 3647–3694.

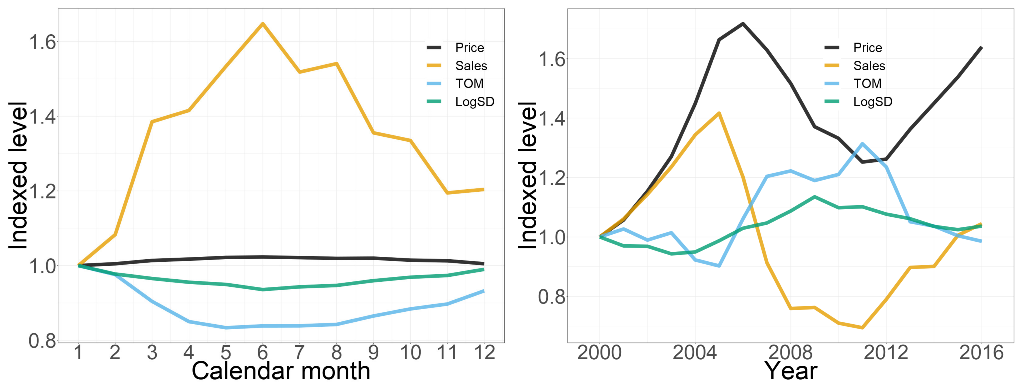
Schubert, Gregor. 2021. "House Price Contagion and U.S. City Migration Networks." Working Paper.

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.

Figures

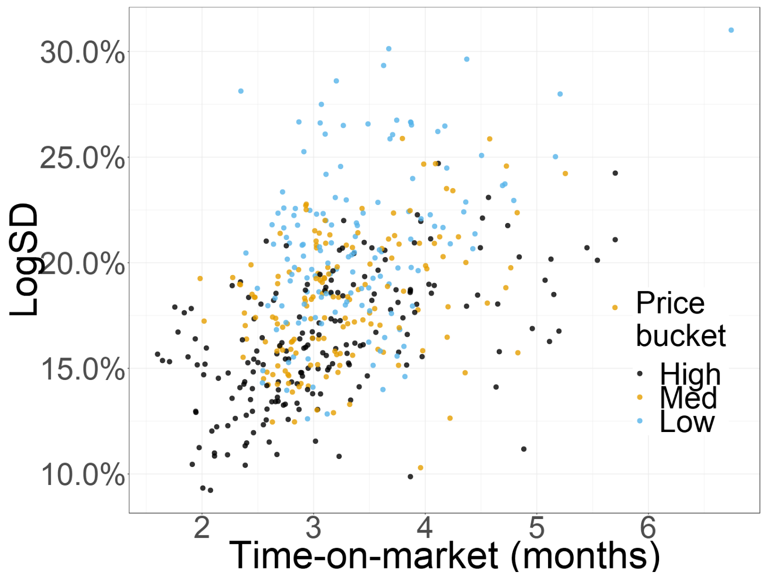
Figure 1. Seasonal and Business Cycle Variation in Sales, Prices, LogSD, and TOM



(a) Seasonality **(b)** Business Cycles

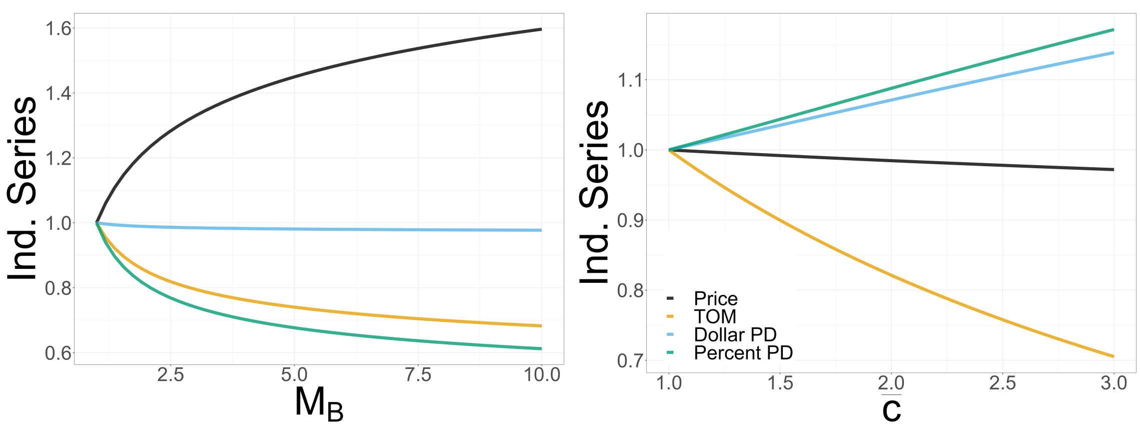
Notes. Panel (a) plots total sales, average prices, logSD, and time-on-market (TOM) by calendar month. The time period of the data is 2000 to 2016. All variables are indexed by dividing by their January level. *LogSD*, our measure of idiosyncratic price dispersion, is calculated according to specification (1). *TOM* is from the Corelogic MLS data. *Sales* is total sales, 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. Panel (b) plots yearly prices, sales, time-on-market, and logSD. 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. *Price* is the Zillow home value index. Further details of data construction are described in Appendix A.5.

Figure 2. 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 counties into three quantile buckets – high, medium, and low – according to their median sale prices; the price bucket is indicated by the color of each point.

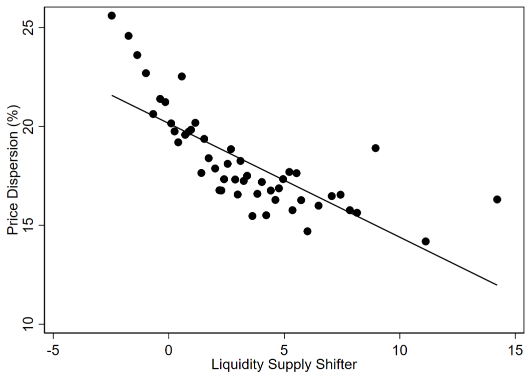
Figure 3. Model Comparative Statics



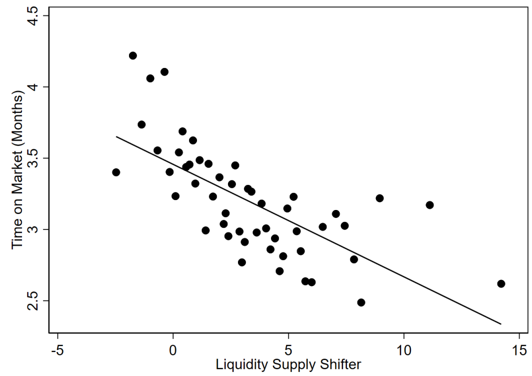
(a) Liquidity Supply **(b)** Liquidity Demand

Notes. Comparative statics of model outcomes with respect to liquidity supply M_B (left) and liquidity demand $\bar{c}$ (right). "Price" is average sale prices. "TOM" is average time-on-market. "Dollar PD" is the standard deviation of dollar prices, and "percent PD" is percentage price dispersion, that is, dollar PD over average prices. In the leftmost point of each plot, we set $r = 0.05263$, $\epsilon_0 = 1$, $\sigma_\epsilon = 0.3$, $M_B = 1$, $\lambda_M = 0.1$, $\bar{c} = 0.5$, $\Delta_c = 0.5$, $A = 1$, $\alpha = 0.84$, $\theta = 0.5$. In the left plot, we hold $\bar{c} = 0.5$, and increase M_B. In the right plot, we hold $M_B = 1$ and increase $\bar{c}$.

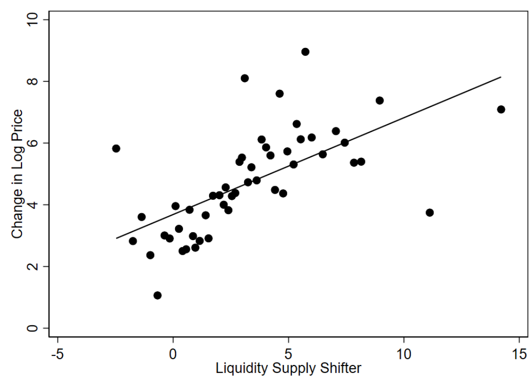
Figure 4. Liquidity Supply, LogSD, TOM, and Prices



(a) LogSD



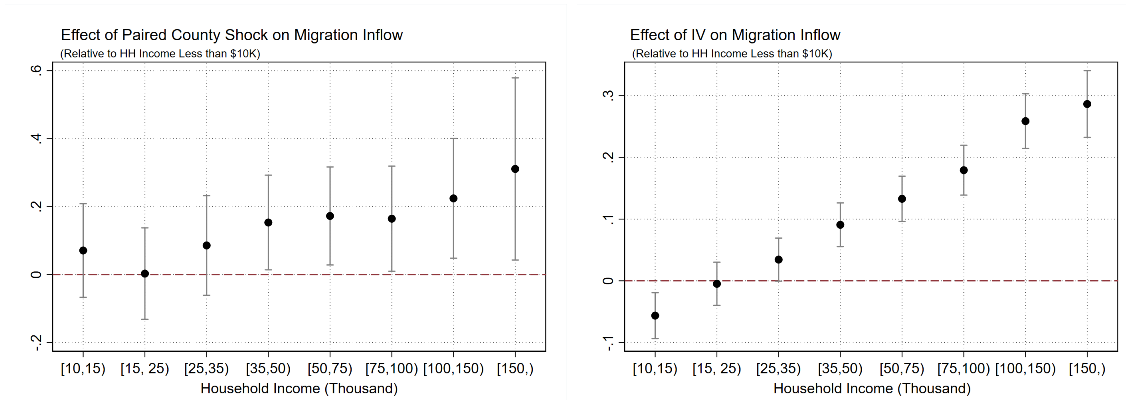
(b) TOM



(c) Price Growth

Notes. This figure shows a binned scatter plot of $LogSD$ (panel a), TOM (panel b), and price growth (panel c) against our liquidity supply shifter, population growth. In all panels, observations are at county level, the sample time period is during 2012-2016. We divide all observations into 50 equal-sized buckets based on population growth. Each dot in the figure indicates the average value of the observations in the bucket. The line in the figure is the linear best fit line. $LogSD$ is our measure of idiosyncratic price dispersion, calculated according to specification (1) and aggregated to county level. TOM is time-on-market, expressed in months, from the Corelogic MLS data. Population growth rates are obtained from the ACS. Log Price Growth is the average annual house price change during 2012-2016 for the county, from Zillow. *Source*: CoreLogic MLS, CoreLogic Deeds and Tax, ACS, and Zillow Price Index.

Figure 5. Migration Flows and Shocks to Liquidity Supply



(a) County-County Pair Level **(b)** County Level

Notes. Panel (a) then shows how the shifter $(B_{c'} \times \mu^{c \leftarrow c'})$ is associated with migration flows at the county-pair level, by plotting the estimated β_g from the following county-county pair level regression:

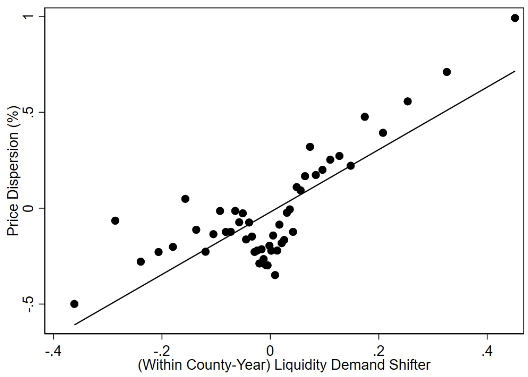
$$\begin{aligned} \text{Log Migration Flow}_{c,c',g} = & \sum_g \beta_g I(\text{Income}_{c,g} \in g) \times (B_{c'} \times \mu^{c \leftarrow c'}) + \gamma_g I(\text{Income}_{c,g} \in g) \times B_{c'} \\ & + \delta_g I(\text{Income}_{c,g} \in g) \times \mu^{c \leftarrow c'} + \eta_g I(\text{Income}_{c,g} \in g) \times B_c \\ & + \alpha_1 (B_{c'} \times \mu^{c \leftarrow c'}) + \alpha_2 (B_c \times \mu^{c \leftarrow c'}) + \epsilon_{c,c',g} \end{aligned}$$

Panel (b) tests how our liquidity supply shifter M_c is associated with migration inflows from households with different income levels at the county level, by plotting the estimated β_g from the following county-times-income-group level regression:

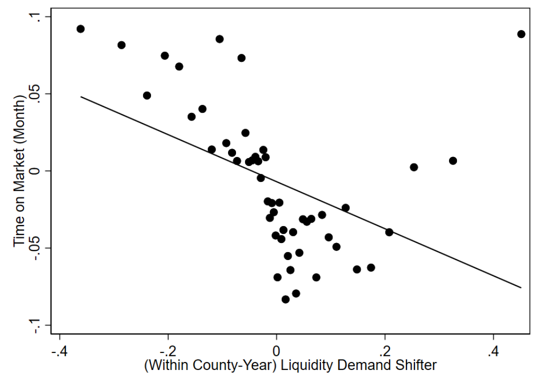
$$\text{Log Migration Inflow}_{c,g} = \sum_g \beta_g I(\text{Income}_{c,g} \in g) \times M_c + \gamma_g I(\text{Income}_{c,g} \in g) \times B_c + \delta_g I(\text{Income}_{c,g} \in g) + \alpha_1 M_c + \alpha_2 B_c + \epsilon_{c,g}$$

In Panel (a), observations are at the county pair-income group level. $\text{Log Migration Flow}_{c,c',g}$ is the log of total migrants in income group g who migrate from county c' to county c , during the period 2007-2011. B_c and $B_{c'}$ are the "Bartik" shocks in county c and county c' during 2007-2011, respectively. $\mu^{c \leftarrow c'}$ is the migration link between county c and county c', measured as the fraction of all migrants moving to county c that originate from county c', from 2005-2009. In Panel (b), observations are at county-income group level. $\text{Log Migration Inflow}_{c,g}$ is the log of migration inflows from households in income group g into county c during 2007-2011. M_c is our liquidity supply shifter for county c during 2007-2011, defined in Section 5.2, in which migration links are constructed using data during 2005-2009. $I(\text{Income}_{c,g} \in g)$ is an indicator for whether the income group is g. B_c is our wage "Bartik" shock in county c during 2007-2011, described in Subsection 5.2. Note that since ACS 2012-2016 migration data no longer collects demographic information, we change to estimate productivity shock using years before 2007 and use 2007-2011 ACS data to analyze migration flows by demographics. *Source*: 2005-2009 and 2007-2011 ACS County-to-County Migration files, and 2005, 2007, and 2011 Wage and Employment statistics.

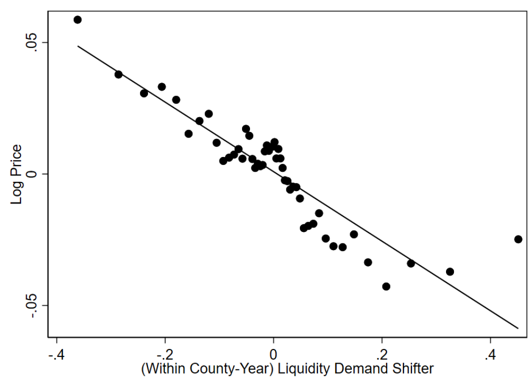
Figure 6. Liquidity Demand, LogSD, TOM, and Prices



(a) LogSD



(b) TOM



(c) Price

Notes. This figure shows a binned scatter plot of *LogSD* (panel a), TOM (panel b), and prices (panel c) against our liquidity demand shifter introduced in Section 6. Each observation in the underlying sample is a zipcode-month, and the sample period is from 2009 to 2017. We divide all observations into 50 equal-sized buckets based on values of the within-county-year liquidity demand shifter, which is constructed by purging out county-year variation in our liquidity demand shifter. Our liquidity demand shifter is average county level house price growth since the previous transaction date, averaged across all individual house transactions in a given zipcode-month. Each dot in the figure indicates the average value of the observations in the bucket. The line in the figure is the linear best fit line. *LogSD* is our measure of idiosyncratic price dispersion at zipcode-month level, which is calculated according to specification (1). TOM is time on market, expressed in months. *Source*: CoreLogic MLS, CoreLogic Deeds and Tax, ACS, and Zillow Price Index.

Tables

Table 1: Summary Statistics

This table reports summary statistics for the two main datasets: the county level sample, and the property transaction-level sample. The county level sample reports the values during 2012-2016. The transaction-level sample reports the values during 2000-2017.

	N	Mean	Stdev	P25	Median	P75
Panel A: County Level Sample						
LogSD	400	18.25	3.79	15.55	18.02	20.68
Time-on-Market	400	3.20	0.74	2.72	3.06	3.59
Log Price Growth	394	4.72	3.14	2.56	4.07	6.71
Population Growth	400	3.30	3.32	0.86	2.91	5.25
Migration Spillover $M_{c,2012-2016}$	400	33.39	1.00	32.71	33.27	34.01
Local Shock $B_{c,2012-2016}$	400	20.49	1.00	19.81	20.36	21.11
Panel B: Transaction Level Sample (Distressed Sales)						
LogSD	157,059	12.73	11.47	4.30	9.43	17.44
Time-on-Market	157,059	3.76	3.36	1.07	2.83	5.43
Log Price	157,059	12.05	0.56	11.72	12.04	12.41
Home Equity Ratio	157,059	-0.08	0.30	-0.22	-0.04	0.10
County Price Growth	155,817	0.90	0.22	0.77	0.90	1.01
Panel B: Transaction Level Sample (Non-Distressed Sales)						
LogSD	892,029	9.09	9.26	2.90	6.36	11.90
Time-on-Market	892,029	2.04	2.37	0.37	1.27	2.77
Log Price	892,029	12.48	0.56	12.10	12.47	12.85
Home Equity Ratio	892,029	0.22	0.23	0.09	0.22	0.35
County Price Growth	886,050	1.13	0.24	0.98	1.10	1.24

Table 2

Liquidity Supply Regressions

This table presents county level regression results about liquidity supply. Panel A reports the OLS results. Panel B reports the 2SLS results, where column 1 is the first stage, columns 2-4 are the second stage, and columns 5-7 are the reduced-form IV results. In both panels, observations are at county level, the sample time period is during 2012-2016, and regressions are weighted by the total number of sales within the county. *LogSD* is our measure of idiosyncratic price dispersion, calculated according to specification (1), and aggregated to county level. TOM is time-on-market, expressed in months, aggregated to county level from the Corelogic MLS data. Population growth rates are obtained from the ACS. Log Price Growth is the average annual house price change during 2012-2016, calculated using Zillow price index. Migration Spillover is a cross-sectional IV, measuring migration spillovers from historically connected high-productivity areas, which we describe in Subsection 5.2. We control for third-order polynomials in the average age of houses, 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, college graduates, and unemployed. In all specifications, we control for each county's own Bartik productivity shock (Local Shock). Robust standard errors are reported in parentheses. ***, **, * represent 1%, 5%, and 10% significance, respectively.

Panel A: OLS

	(1) LogSD	(2) TOM	(3) Log Price Growth
Population Growth	-0.30***	-0.08**	0.62***
	(0.102)	(0.032)	(0.113)
Controls	✓	✓	✓
Observations	400	400	394
Adjusted R^2	0.57	0.36	0.33

Panel B: 2SLS

	First Stage	Second Stage			Reduced-Form		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Pop Growth	LogSD	TOM	Log Price Growth	LogSD	TOM	Log Price Growth
$\overline{\text{Population Growth}}$		-0.98***	-0.39***	2.12***			
		(0.262)	(0.080)	(0.449)			
Migration Spillover	0.81***				-0.79***	-0.32***	1.71***
	(0.161)				(0.203)	(0.050)	(0.300)
Local Shock	0.22	-0.19	0.08	0.09	-0.41*	-0.00	0.58**
	(0.170)	(0.301)	(0.079)	(0.451)	(0.216)	(0.035)	(0.260)
Controls	✓	✓	✓	✓	✓	✓	✓
Observations	400	400	400	394	400	400	394
Adjusted R^2	0.79	-	-	-	0.58	0.42	0.37
Underidentification t-stat		19.84	19.84	19.57			
Underidentification p-value		0.00	0.00	0.00			
Weak identification t-stat		24.95	24.95	24.66			
Hansen J statistic		0.00	0.00	0.00			

Table 3
Productivity Shocks and Migration Flows

This table presents evidence about how local productivity shocks affect migration flows. Columns 1-3 are at the county-pair level: each observation is a pair of counties. In these three columns, the outcome variable is the log of county-to-county migration population. $B_{c',2012-2016}$ is our "Bartik" shock in county c' during 2012-2016, which is constructed by combining local employment shares in every 2-digit NAICS industry in 2010 with national wage growth rate in that industry over 2012-2016, as we describe in Subsection 5.2. $\mu^{c\leftarrow c'}_{2008-2012}$ is the migration link between county c and county c', measured as the fraction of all migrants moving to county c that originate from county c', from 2008 to 2012. Column 4 is at county level, in which each observation is a county. The outcome variable is migration inflow scaled by local population. $M_{c,2012-2016}$ is our liquidity supply shifter for county c during 2012-2016, defined in Subsection 5.2. We construct $M_{c,2012-2016}$ by weighting the shock in every other county by its historical migration link with county c and summing over these shocks. Standard errors reported in parentheses are double clustered by both counties in the county-pair in columns 1-3. Robust standard errors are reported in parentheses in column 4. ***, **, * represent 1%, 5%, and 10% significance, respectively.

	County-Pair Level			County Level
	(1)	(2)	(3)	(4)
	Ln(County-to-County Migration)			$\frac{\text{Migration Inflow}}{\text{Total Population}}$
$B_{c',2012-2016} \times \mu^{c\leftarrow c'}_{2008-2012}$	2.99***	2.84***	1.83**	
	(0.804)	(0.795)	(0.766)	
$B_{c',2012-2016}$	0.02***	0.02***		
	(0.002)	(0.002)		
$\mu^{c\leftarrow c'}_{2008-2012}$	12.64**	13.64***	18.75***	
	(5.029)	(4.986)	(4.830)	
$M_{c,2012-2016}$				0.75***
				(0.047)
$B_{c,2012-2016}$	0.03***			-0.01
	(0.002)			(0.056)
Local County FE		✓	✓	
Paired County FE			✓	
Observations	10.7M	10.7M	10.7M	3,213
Adjusted R^2	0.10	0.14	0.18	0.07

Table 4

Liquidity Demand Regressions

This table presents the regression results about liquidity demand. Each observation is a zipcode-month. Columns 1-3 focus on distressed sales. Columns 4-6 focus on non-distressed sales. Columns 7-9 uses the full sample combining distressed and non-distressed sales. The outcome variable in columns 1, 4, and 7 is $LogSD$, our measure of idiosyncratic price dispersion at zipcode-month level, which is calculated according to specification (1) and aggregated to the zipcode-month level. The outcome variable in columns 2, 5, and 8 is time-on-market, expressed in months, from the Corelogic MLS data and aggregated to the zipcode-month level. The outcome variable in columns 3, 6, and 9 is the log property transaction price, averaged at the zipcode-month level. The explanatory variable of interest is home equity ratio, instrumented using the average county level house price growth since the previous transaction date across all individual house transactions in the zipcode-distressed status-month, as we describe in Subsection 6.1. Zipcode controls include zipcode average square footage, number of bedrooms, number of bathrooms, their squared terms, share of minority, share of college graduates, median income, and their squared terms. The sample period is from 2009 to 2017. Standard errors are clustered at zipcode level. ***, **, * represent 1%, 5%, and 10% significance, respectively.

	Distressed Sales			Non-Distressed Sales			Full Sample		
	(1) LogSD	(2) TOM	(3) Log Price	(4) LogSD	(5) TOM	(6) Log Price	(7) LogSD	(8) TOM	(9) Log Price
$\overline{\text{Home Equity Ratio}}$	4.91***	-0.45***	-9.60***	-2.96***	-0.71***	-4.18***	1.06***	-0.93***	-9.85***
	(0.350)	(0.091)	(1.069)	(0.314)	(0.063)	(1.537)	(0.255)	(0.054)	(1.080)
Distressed Sale							2.66***	1.00***	-21.84***
							(0.110)	(0.018)	(0.392)
Zipcode Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Property Age Group FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
County-Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	98,628	98,628	98,628	238,512	238,512	238,512	337,479	337,479	337,479
Underidentification t-stat	868.44	731.44	868.44	1467.15	1467.15	1467.15	1295.76	1295.76	1295.76
Underidentification p-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Weak identification t-stat	8823.18	9661.99	8823.18	11279.37	11279.37	11279.37	10342.61	10342.61	10342.61
Hansen J Stat	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 5

Quantification: Moment and Parameter Values

This table presents the 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 λ_M are yearly turnover and separation rates respectively, and η_B is a fraction of the unit mass of houses per year.

(1) Moment	(2) Value	(3) Parameter	(4) Value
Price	266.72	λ_M	0.052
Buyer frac. LogSD	4.65%	ϵ_0	0.739
TOM (Months)	2.554	σ_ϵ	0.2159
Turnover rate	0.051	M_B	1.062
Num. visits	9.960	$\bar{c}$	4.70
PD-TOM Coef	0.876	Δ_c	3.864

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 number 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, Mangum and Roberts (2021), 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 codes. 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 seasonal dataset used in Figures 1a and A1, 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 piecewise linear trend with breakpoints 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 1a and A1.

To construct the time series dataset used in Figures 1b and A2, 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, 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.

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:

$$f_c(x_i, t) = g_c^{latlong}(t, lat_i, long_i) + g_c^{sqft}(t, sqft_i) + g_c^{yrbuilt}(t, yrbuilt_i) + g_c^{bedrooms}(t, bedrooms_i) + g_c^{bathrooms}(t, bathrooms_i) \quad (27)$$

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

Additivity in specification (27) rules out many interaction effects between characteristics. Older or larger houses can appreciate faster or slower than newer or smaller houses. However, houses which are both large and old are constrained to appreciate at a rate which is the sum of the "old house" and "large house" effects on prices. The only interaction term we include is the $g_c^{latlong}(t, lat_i, long_i)$ function, which interacts latitude and longitude. This is important because it is implausible that latitude and longitude have additive effects on prices; effectively, this specification allows house prices to vary smoothly as a function of a house's geographic location over time.

Given this functional form for $f_c(x_i, t)$, specification (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:

$$\hat{\epsilon}_{it}^2 = \frac{N_c}{N_c - K_c} \left(p_{it} - \hat{p}_{it}\right)^2 \tag{28}$$

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 non-trivially 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 (28) causes the expectation of $\hat{\epsilon}_{it}^2$ to be equal to the true variance, σ_c^2. We thus apply the homoskedastic variance adjustment term here, as we are not aware of any computationally tractable way to implement a degrees-of-freedom correction in the general heteroskedastic case. However, in Appendix 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 λ_B, trade cutoffs $\epsilon^*(c)$, the equilibrium distribution of seller values $F_{eq}(c)$, and the seller value function $V_S(c)$, the equilibrium value of buyers, V_B, must satisfy:

$$rV_B = \lambda_B \int \int_{\epsilon > \epsilon^*(c)} \Big[(1-\theta) \big(V_M(\epsilon) - V_B - V_S(c) \big) \Big] \, dG(\epsilon) \, dF_{eq}(c) \tag{29}$$

In words, expression (29) can be interpreted as follows. At rate λ_B, the buyer is matched to a seller with type randomly drawn from $F_{eq}(\cdot)$, and the buyer draws match quality ϵ from $G(\cdot)$. If the buyer's match quality draw, ϵ, is higher than the seller's match quality cutoff, $\epsilon^*(c)$, trade occurs, and the buyer receives a share $(1-\theta)$ of the bilateral match surplus.

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

$$rV_S(c) = v + \lambda_S \int_{\epsilon > \epsilon^*(c)} \theta \left(V_M(\epsilon) - V_B - V_S(c) \right) dG(\epsilon) \tag{30}$$

In words, expression (30) states that a seller of type c receives flow value $-c$ from their house while they are waiting for buyers. At rate λ_S, the seller meets a buyer with match value ϵ randomly drawn from $G(\cdot)$. If $\epsilon > \epsilon^*(c)$, trade occurs, and the seller receives a share θ of the bilateral match surplus.

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

$$rV_M(\epsilon) = \epsilon + \lambda_M \left(\int V_S(c)\, dF(c) - V_M(\epsilon) \right) \tag{31}$$

In words, expression (31) states that matched owners get flow value ϵ while matched to their house and receive separation shocks at rate λ_M, at which point they become sellers and attain the expectation of the seller value function $V_S(c)$ over the seller holding cost distribution $F(c)$.

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:

$$(1 - M_S)\, \lambda_M f(c) \tag{32}$$

In words, this is the product of the total mass of matched homeowners, $1 - M_S$; the rate at which homeowners receive separation shocks, λ_M; and the density $f(c)$ of entering sellers with value c.[21] The equilibrium rate at which sellers of type c sell their houses and leave the market is:

$$M_S f_{eq}(c)\, \lambda_S \left(1 - G\left(\epsilon^*(c) \right) \right) \tag{33}$$

[21] Since the distribution $F(c)$ of holding costs does not depend on matched homeowners' match utility ϵ, we do not need to explicitly integrate over the distribution $G_{eq}(\epsilon)$ in expression (32).

In words, this is the product of the mass of sellers, M_S; the density of values among sellers in equilibrium, $f_{eq}(c)$; the rate at which sellers are matched to buyers in equilibrium, λ_S; and the probability that the match utility draw ϵ exceeds the trade cutoff $\epsilon^*(c)$ for a seller of type c, which is $1 - G\left(\epsilon^*(c)\right)$. In stationary equilibrium, expressions (32) and (33) 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 (32) and (33) over c, we have:

$$(1 - M_S)\lambda_M = \int_c \lambda_S M_S \left(1 - G\left(\epsilon^*(c)\right)\right) f_{eq}(c)\, dc \tag{34}$$

Moreover, since each successful sale turns a buyer into a matched homeowner, the RHS of (34) is also equal to the rate at which buyers turn into matched homeowners.

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

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

In words, the numerator is the flow rate at which a seller of value c successfully trades with a buyer with match utility below ϵ, integrated over the equilibrium distribution $f_{eq}(c)$ of holding costs c among sellers. The denominator is the RHS of (34), the total flow rate at which buyers become matched homeowners.

B.2 Proof of Claim 1

Time-on-market. Time-on-market for a seller of type c is the inverse of $\lambda_S\left(1 - G\left(\epsilon^*(c)\right)\right)$, which is the product of the equilibrium rate at which sellers meet buyers, λ_S, and the fraction of meetings for a seller of type c that result in trade, $\left(1 - G\left(\epsilon^*(c)\right)\right)$. Average time-on-market is the expectation with respect to the distribution of seller costs among trading sellers in stationary equilibrium. The density of costs among trading (that is, outflowing) sellers is the RHS of (10) of Proposition 1, so it is proportional to

$$f_{eq}(c)\left(1-G\left(\epsilon^*(c)\right)\right)$$

By (10), this is proportional to the density $f(c)$ of entering sellers, hence average time-on-market can be calculated simply by taking the average of $\lambda_S\left(1-G\left(\epsilon^*(c)\right)\right)$ with respect to the entering cost distribution $f(c)$. In words, sellers can only exit the market by selling, hence the distribution of costs among exiting sellers must be the same as the distribution of costs among entering sellers, which is $f(c)$.

Price dispersion. It is in calculating price dispersion where the assumption that match quality is exponential, combined with the assumption in (4), helps most to simplify our results. Essentially, the derivation proceeds as follows. Exponential match quality implies that the distribution of $\left(\epsilon-\epsilon^*(c)\right)$ is independent of any $\epsilon^*(c)$ greater than ϵ_0. This basically allows the seller and buyer contributions to price variance to be analytically separated. The conditional variance of prices with respect to c, which depends on ϵ, is independent of $\epsilon^*(c)$ and can be thought of as buyers' contribution to price variance. The conditional expectation of prices with respect to c is just sellers' value $V_S(c)$ plus an expected markup which does not depend on c, implying that the variance of the conditional expectation is just equal to the variance of sellers' values $V_S(c)$, which can be thought of as sellers' contribution to price variance.

To demonstrate this formally, from (3), prices are:

$$P(\epsilon,c)=\theta\left(V_M(\epsilon)-V_B-V_S(c)\right)+V_S(c) \tag{35}$$

We wish to take the variance of expression (35) with respect to the joint distribution of holding costs c and match utilities ϵ within the set of pairs of buyers and sellers that match and trade in any given moment; call this joint distribution $F_{tr}(c,\epsilon)$.

First, let $F_{tr}(c)$ be the marginal distribution of seller holding costs c, among the stationary mass of seller types that trade in any given time period. As we argued above, 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:

$$F_{tr}(c) = F(c) \tag{36}$$

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

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

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,

$$F_{tr}\left(\epsilon \mid c\right) = G\left(\epsilon \mid \epsilon > \epsilon^*(c)\right) \tag{37}$$

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

Having characterized $F_{tr}(c, \epsilon)$, we can now take the variance of expression (35) for prices. Applying the law of iterated expectations, price variance can be written as:

$$\begin{aligned} &Var\left(P(\epsilon, c)\right) = \\ &\quad E_{c \sim F_{tr}(c)}\left[Var_{\epsilon \sim F_{tr}\left(\epsilon \mid c\right)}\left(P(\epsilon, c) \mid c\right)\right] + Var_{c \sim F_{tr}(c)}\left(E_{\epsilon \sim F_{tr}\left(\epsilon \mid c\right)}\left[P(\epsilon, c) \mid c\right]\right) \end{aligned} \tag{38}$$

Substituting (36) and (37), we can write this as:

$$\begin{aligned} &Var\left(P(\epsilon, c)\right) = \\ &\quad E_{c \sim F(c)}\left[Var_{\epsilon \sim G\left(\epsilon \mid \epsilon > \epsilon^*(c)\right)}\left(P(\epsilon, c) \mid c\right)\right] + Var_{c \sim F(c)}\left(E_{\epsilon \sim G\left(\epsilon \mid \epsilon > \epsilon^*(c)\right)}\left[P(\epsilon, c) \mid c\right]\right) \end{aligned} \tag{39}$$

First, we characterize the left term on the RHS of (38). Conditional on c, the only random term in $P(\epsilon, c)$ conditional on c is the buyer's match utility ϵ; thus, substituting expression

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

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

In words,

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

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

$$Var_{\epsilon\sim G\left(\epsilon|\epsilon>\epsilon^*(c)\right)}\left(P(\epsilon,c)\mid c\right)=\left(\frac{\theta}{r+\lambda_M}\right)^2\sigma_\epsilon^2 \tag{40}$$

Since expression (40) is independent of c, we also have:

$$E_{c\sim F(c)}\left[Var_{\epsilon\sim G\left(\epsilon|\epsilon>\epsilon^*(c)\right)}\left(P(\epsilon,c)\mid c\right)\right]=\left(\frac{\theta}{r+\lambda_M}\right)^2\sigma_\epsilon^2 \tag{41}$$

Now we move to the right term in expression (38). Substituting expression (52) for prices, we have:

$$\begin{aligned}Var_{c\sim F(c)}&\left(E_{\epsilon\sim G\left(\epsilon|\epsilon>\epsilon^*(c)\right)}\left[P(\epsilon,c)\mid c\right]\right)=\\&Var_{c\sim F(c)}\left(E_{\epsilon\sim G\left(\epsilon|\epsilon>\epsilon^*(c)\right)}\left[V_S(c)+\theta\left(\frac{\epsilon-\epsilon^*(c)}{r+\lambda_M}\right)\Big| c\right]\right)\end{aligned} \tag{42}$$

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

$$Var_{c\sim F(c)}\left(V_S(c)+\frac{\theta}{r+\lambda_M}E_{\epsilon\sim G\left(\epsilon|\epsilon>\epsilon^*(c)\right)}\left[\epsilon-\epsilon^*(c)\mid c\right]\right) \tag{43}$$

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

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

is equal to σ_ϵ, the standard deviation of ϵ. It is thus constant with respect to $\epsilon^*(c)$ and thus c, and can be ignored when calculating the variance in (43). Hence,

$$Var_{c\sim F(c)}\left(E_{\epsilon\sim G\left(\epsilon|\epsilon>\epsilon^*(c)\right)}\left[P(\epsilon,c)\mid c\right]\right)=Var_{c\sim F(c)}\left(V_S(c)\right) \tag{44}$$

Substituting (41) and (44) into (39), we have

$$Var\left(P(\epsilon,c)\right)=Var_{c\sim F(\cdot)}\left(V_S(c)\right)+\left(\frac{\theta}{r+\lambda_M}\right)^2\sigma_\epsilon^2 \tag{45}$$

Now, taking the expectation of prices from (52) below, we have:

$$E\left[P(\epsilon,c)\right]=E\left[V_S(c)\right]+\frac{\theta\sigma_\epsilon}{r+\lambda_M} \tag{46}$$

Using (45) and (46), we get (13).

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

To begin with, we analytically characterize $V_M(\epsilon)$. From expression (31), 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:

$$V_M(\epsilon)=\frac{\epsilon}{r+\lambda_M}+\frac{\lambda_M}{r+\lambda_M}\int V_S(c)\,dF(c) \tag{47}$$

Using expression (47), we can also characterize the trade cutoff function $\epsilon^*(c)$. Trade occurs if:

$$\begin{aligned}V_M(\epsilon)&\geq V_B+V_S(c)\\ \implies \frac{\lambda_M}{r+\lambda_M}\int V_S(c)\,dF(c)+\frac{\epsilon}{r+\lambda_M}&\geq V_B+V_S(c)\end{aligned} \tag{48}$$

Since we have assumed that $\epsilon^*(c)$ is greater than ϵ_0, the lower bound of $G(\cdot)$, we can treat expression (48) as an equality. Solving for $\epsilon^*(c)$, we have:

$$\epsilon^*(c) = (r + \lambda_M)\left[V_B + V_S(c)\right] - \lambda_M \int V_S(c)\, dF(c) \tag{49}$$

Using (47) and (49) we can also characterize equilibrium prices. From (3), we have:

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

Substituting for $V_M(\epsilon)$ using (47), we have:

$$P(\epsilon, c) = V_S(c) + \theta\left(\frac{\epsilon}{r + \lambda_M} + \frac{\lambda_M}{r + \lambda_M}\int V_S(c)\, dF(c) - V_B - V_S(c)\right) \tag{50}$$

Now, we can write (49) as:

$$\frac{\lambda_M}{r + \lambda_M}\int V_S(c)\, dF(c) - V_B - V_S(c) = -\frac{\epsilon^*(c)}{r + \lambda_M} \tag{51}$$

Hence, substituting (51) into (50), we get:

$$P(\epsilon, c) = V_S(c) + \theta\left(\frac{\epsilon - \epsilon^*(c)}{r + \lambda_M}\right) \tag{52}$$

B.4 Proof of Claim 2

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

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

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

$$rV_S'(c) = -1 - \lambda_S\theta\left(V_M\left(\epsilon^*(c)\right) - V_B - V_S(c)\right) g\left(\epsilon^*(c)\right)\frac{d\epsilon^*(c)}{dc} + \lambda_S \int \theta\left(-V_S'(c)\right) 1\left(\epsilon > \epsilon^*(c)\right) dG(\epsilon) \tag{53}$$

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

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

so the middle term is 0. Hence, (53) becomes:

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

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

$$V_S' (c) = \frac{-1}{r + \lambda_S \theta \Big(1 - G \left(\epsilon^* (c)\right)\Big)} \tag{54}$$

Substituting expression (12) for $TOM (c)$ in the denominator of (54), we have:

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

Rearranging, we have (15).

B.5 Derivation of Model Quantities

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

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

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

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

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

This is simply the average of time-on-market for a seller of holding cost c over the distribution of holding costs $F(c)$; note that we showed in (36) above that the distribution of holding costs among trading sellers is simply $F_{tr}(c) = F(c)$.

From Claim 1, equilibrium price variance is:

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

The relative standard deviation of prices is simply the dollar standard deviation divided by the average price.

B.6 Microfounding the Buyer Mass M_B

In the main text, we take the mass of buyers waiting to buy houses, M_B, as exogenously given. This appendix shows how M_B can be microfounded through a model of buyer optimal entry.

Potential buyers enter the market at some exogenous flow rate η_B, and each buyer draws a value $\xi \sim H(\cdot)$ for entering the city. ξ can be thought of as representing the attractiveness of amenities and job opportunities in the county, which may vary idiosyncratically across buyers. After observing ξ, each potential buyer can choose to either enter the city, receiving utility ξ immediately and becoming an active homebuyer, or leave forever, receiving utility normalized to 0. 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:

$$\xi + V_B > 0 \tag{55}$$

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

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

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)).

Flow equality in the model then requires the entry rate of buyers to be equal to the matching rate of buyers. Since buyers transition to matched owners, who then transition to sellers, this is equivalent to requiring that the entry rate of buyers is equal to the rate at which matched owners turn into sellers:

$$(1 - M_S)\,\lambda_M = \eta_B\left(1 - H\left(-V_B\right)\right) \tag{56}$$

Note that η_B only enters the model through its effect on (56): all equations in Proposition 1 depend only on M_B. Thus, for any given M_B and equilibrium satisfying the conditions in Proposition 1, we can find an η_B which solves (56). Thus, we take M_B as given for simplicity in the main text; in the background, we can think of variation in the equilibrium M_B as being driven by variation in η_B.

B.7 Heterogeneous Buyer Urgency

We can extend the main model to accommodate 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 ϵ:

$$P(c, u, \epsilon) = V_S(c) + \theta\left(V_M(\epsilon) - V_B(u) - V_S(c)\right) \tag{57}$$

Thus, the match quality cutoff condition becomes:

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

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

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

Buyers' and sellers' value functions become, respectively:

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

$$rV_S(c) = -c + \lambda_S \int_u \int_{\epsilon > \epsilon^*(c,u)} \Big[\theta\left(V_M(\epsilon) - V_B(u) - V_S(c)\right)\Big]\, dG(\epsilon)\, dH_{eq}(u) \qquad (58)$$

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

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

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

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

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

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:

$$V_M(\epsilon) = \frac{\epsilon}{r + \lambda_M} + \frac{\lambda_M}{r + \lambda_M} \int_c V_S(c)\, dF(c) \qquad (59)$$

The derivations in Appendix B.3 thus imply that:

$$P(c, u, \epsilon) = V_S(c) + \theta\left(\frac{\epsilon - \epsilon^*(c,u)}{r + \lambda_M}\right) \qquad (60)$$

Now, similar to (38), we take the variance of prices, applying the law of iterated expectations with respect to c and u, to get:

$$Var\left(P\left(c,u,\epsilon\right)\right)=E_{c,u\sim F_{tr}(c,u)}\left[Var_{\epsilon\sim G_{tr}\left(\epsilon|c,u\right)}\left(P\left(c,u,\epsilon\right)\mid c,u\right)\right]+ \\ Var_{c,u\sim F_{tr}(c,u)}\left(E_{\epsilon\sim G_{tr}\left(\epsilon|c,u\right)}\left[P\left(c,u,\epsilon\right)\mid c,u\right]\right) \quad (61)$$

Where, analogously to expression (38), $F_{tr}\left(c,u\right)$ is the joint distribution of c and u among trading buyers and sellers, and $F_{tr}\left(\epsilon\mid c,u\right)$ is the conditional distribution of ϵ given c,u among trading buyers and sellers. Analogously to the argument to Appendix B.2, we have:

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

The joint distribution $F_{tr}\left(c,u\right)$ is more complicated to characterize; however, by flow equality, the marginal distributions of $F_{tr}\left(c,u\right)$ must be equal to the distributions of entering buyer and seller types, $F\left(c\right)$ and $H\left(u\right)$. This implies that the following steps in Appendix B.2 go through essentially unchanged. Going through the steps, for the top term of (61), we have:

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

For the bottom term, substituting expression (60) for prices, we have:

$$Var_{c,u\sim F_{tr}(c,u)}\left(E_{\epsilon\sim G\left(\epsilon|\epsilon>\epsilon^{*}(c,u)\right)}\left[P\left(c,u,\epsilon\right)\mid c,u\right]\right) \\ =Var_{c,u\sim F_{tr}(c,u)}\left(E_{\epsilon\sim G\left(\epsilon|\epsilon>\epsilon^{*}(c,u)\right)}\left[V_{S}\left(c\right)+\theta\left(\frac{\epsilon-\epsilon^{*}\left(c,u\right)}{r+\lambda_{M}}\right)\mid c,u\right]\right) \\ =Var_{c,u\sim F_{tr}(c,u)}\left(V_{S}\left(c\right)+\frac{\theta}{r+\lambda_{M}}E_{\epsilon\sim G\left(\epsilon|\epsilon>\epsilon^{*}(c,u)\right)}\left[\epsilon-\epsilon^{*}\left(c,u\right)\mid c,u\right]\right) \quad (62)$$

Again, the left term of (62),

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

is equal to σ_ϵ, which is independent of c, u, so we can ignore it in the variance calculation; (62) thus simplifies to

$$Var_{c,u\sim F_{tr}(c,u)}\left(V_S\left(c\right)\right)$$

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

$$Var_{c\sim F(c)}\left(V_S\left(c\right)\right)$$

This proves (13). The proof of (12) is identical to the baseline model. Finally, differentiating (58), we have:

$$rV_S'\left(c\right)=1+\lambda_S\theta\int_u\int_{\epsilon>\epsilon^{*}(c,u)}\left(-V_S'\left(c\right)\right)1\left(\epsilon>\epsilon^{*}\left(c,u\right)\right)dG\left(\epsilon\right)dH_{eq}\left(u\right) \tag{63}$$

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

$$TOM\left(c\right)=\frac{1}{\lambda_S\int_u\int_{\epsilon>\epsilon^{*}(c,u)}1\left(\epsilon>\epsilon^{*}\left(c,u\right)\right)dG\left(\epsilon\right)dH_{eq}\left(u\right)} \tag{64}$$

Combining (63) and (64), we have:

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

proving claim 2.

B.8 Alternative Models

A natural question is how sensitive our conclusions about liquidity supply and demand are to our specific modelling assumptions. The literature on search, and frictional markets more generally, is vast; our conclusions cannot generalize to all models. However, in this appendix, we highlight a natural class of models, which is overlapping but non-nested with our baseline model, in which a version of Claim 2 applies, regarding the relationship between time-on-market and sellers' values.

We consider a very general setting; there are atomistic sellers in stationary equilibrium. Sellers have some cost c per unit of time their house is on the market, which is continuously distributed. Sellers choose some selling strategy Γ. The strategy Γ is very general, and could represent, for example, a posted price; a state-contingent strategy for making offers and counter-offers to buyers in a bargaining game; or a set of buyer offers the seller is willing to consider. A strategy Γ leads to some stationary hazard rate of sale $\lambda(\Gamma)$, and some expected price conditional on sale $EP(\Gamma)$. The only assumption we make is that every seller can choose any value of Γ, and that any seller choosing strategy Γ achieves the same sale hazard rate and expected price.

Under these assumptions, the optimal payoff of a type c seller solves:

$$rV_S(c) = \max_{\Gamma} -c + \lambda(\Gamma)\big(EP(\Gamma) - V_S(c)\big) \tag{65}$$

As an example, in our baseline model where $\theta = 1$, the strategy Γ is fully described by the matching cutoff $\epsilon^*(c)$, and we have:

$$\Lambda(\Gamma) = \lambda_S \int_{\epsilon > \epsilon^*(c)} dG(\epsilon)$$

$$EP(\Gamma) = \frac{\int_{\epsilon > \epsilon^*(c)} \big(V_M(\epsilon) - V_B\big)\, dG(\epsilon)}{\int_{\epsilon > \epsilon^*(c)} dG(\epsilon)}$$

This setting is very general, capturing many selling games in which sellers can commit to selling strategies. The framework also encompasses a simple model in which sellers commit to pure price-posting. It also captures, for example, the directed search model of Albrecht, Gautier and Vroman (2016), which is a directed-search model of seller price posting, in which any seller type can replicate the strategy of any other seller type. In this setting, Γ represents sellers' posted price, as well as their state-contingent decisions on whether to accept any possible offer from buyers. This framework also nests our baseline model, when $\theta = 1$ and sellers have full bargaining power: in this setting, buyers can be thought of as offering their entire continuation buyers to sellers, and sellers simply select $\epsilon^* (c)$, the set of buyers whose offers they accept. Importantly, however, this setting does not nest the baseline model with general θ values: in Nash bargaining models with interior Nash bargaining weights, low-cost sellers cannot attain the same outcomes as high-cost sellers, since prices are assumed to be weighted averages of sellers' and buyers' values; another way to say this is that buyers can be thought of as fully knowing sellers' types, so a low-cost seller cannot credibly pretend to be a high-cost seller. This appendix should thus be thought of as showing that a similar logic to the baseline model applies to another set of models, which has some overlap with – but is not a superset of – the baseline model.

Since (65) is an optimization problem, we can apply the envelope theorem. Let $\Gamma^* (c)$ be the optimal strategy for a seller with cost c. To calculate $V_S' (c)$, we can differentiate (65), fixing the strategy Γ at $\Gamma^* (c)$:

$$rV_S' (c) = -1 + \lambda \left(\Gamma^* (c)\right) V_S' (c)$$

Rearranging, we have a generalization of Claim 3.

Claim 3. *We have:*

$$V_S' (c) = \frac{-1}{r + \lambda \left(\Gamma^* (c)\right)} \tag{66}$$

Note that (66) generalizes (15) in the sense that $\lambda \left(\Gamma^* (c)\right)$, the meeting rate under the optimal strategy $\Gamma^* (c)$, is the inverse of time-on-market. Thus, when time-on-market is lower, so $\lambda \left(\Gamma^* (c)\right)$ is higher, $V_S' (c)$ is lower and thus sellers' values are less sensitive to sellers' holding costs. We have thus shown that, in any model in which sellers have price-setting power, dispersion in sellers' values is closely related to time-on-market.

Compared to seller values, it is somewhat more difficult to characterize *price dispersion* in a model-independent way. For example, for sake of illustration, one could construct a model in which trade occurs at exogenously random prices: when a buyer and seller meet, a random price is drawn and buyers and sellers can accept or reject this price. There would be a large component of price dispersion in this model which is unrelated to buyer or seller preferences. However, we can define a generalized counterpart of the *seller component of price dispersion* as $Var\left(EP\left(\Gamma (c)\right)\right)$, the dispersion of expected prices across sellers. In any model where the seller markup $EP\left(\Gamma\left(c\right)\right)-V_S\left(c\right)$ is constant, or at least relatively stable, $Var\left(EP\left(\Gamma\left(c\right)\right)\right)$ will be close to $Var\left(V_S\left(c\right)\right)$. In such models, holding the distribution of seller costs fixed, when time-on-market is greater, $Var\left(V_S\left(c\right)\right)$ will tend to be greater, and thus the seller component of price variance $Var\left(EP\left(\Gamma\left(c\right)\right)\right)$ will also tend to be greater. Thus, holding the distribution of seller costs fixed, the seller component of price dispersion will tend to be greater when TOM is greater in a fairly broad class of models, so long as seller markups do not vary too much as model conditions change. This suggests that the "liquidity supply" effect that we discuss in the main text should generalize somewhat to other models, though this conclusion depends somewhat on the behavior of generalized markups.

The other main force in our model is that, when liquidity demand increases, prices decrease, tending to increase relative price dispersion. This force is fairly robust across models: in most models, prices will tend to decline when sellers' holding costs increase, though as in the baseline model, the tendency for time-on-market to decrease produces a force that in principle could reverse this conclusion.

C Robustness Checks

C.1 Measurement Concerns

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 A3 to A5. 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 A6 and A7. In all cases, the alternative measures are closely correlated with our measures in the main text.

Conceptually, specification (1) in Section 2.2 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, γ_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 (2021) 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.

C.2 Price Dispersion: Validation of Methodology

In this appendix, we show that alternative methods for estimating price dispersion appear to produce correlated estimates to our baseline specification.

Pure repeat sales specification: First, we omit the polynomial $f_c(x_i, t)$ term from (1), estimating residuals using the specification:

$$p_{it} = \gamma_i + \eta_{ct} + \epsilon_{it} \tag{67}$$

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:

$$p_{it} = \eta_{ct} + f_c\left(x_i, t\right) + \epsilon_{it} \tag{68}$$

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 γ_i is estimated more precisely.

Let tbs_i be the average time-between-sales for house i, and let $sales_i$ be the total number of times we see house i being sold. Figure A3 plots a kernel regression fit of our estimated residuals, $|\hat{\epsilon}_{it}|$, against tbs_i, separately for $sales_i$ equal to 2, 3 and 4, for houses with tbs_i between the 1st and 99th percentiles for each value of $sales_i$. We see that the estimated logSD, $|\hat{\epsilon}_{it}|$, is on average higher when $sales_i$ and tbs_i are larger.

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

$$\hat{\epsilon}_{it}^2 = g_c\left(sales_i, tbs_i\right) + \zeta_{it} \tag{69}$$

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

$$\hat{\epsilon}_{TBSadj,it}^2 = \hat{\zeta}_{it} + E_c\left[\hat{\epsilon}_{it}^2\right] \tag{70}$$

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

Qualitative results: Figure A4 compares residuals from the pure repeat-sales and pure hedonic specifications, (67) and (68) 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 A4 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 A4 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 (70).

Figure A5 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.

C.3 Comparison of Price Dispersion Estimates to Literature

A number of other papers have attempted to measure idiosyncratic house price dispersion. Giacoletti (2021) uses the same Corelogic data that we use to measure idiosyncratic price dispersion in the metropolitan areas of San Francisco, San Diego, and Los Angeles. Unlike our specification (1), Sagi (2021) and Giacoletti use returns, rather than individual house sales, as the primary unit of analysis.[22] 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 (2021) 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

[22] 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.

paper is Carrillo, Doerner and Larson (2023), 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 complement Giacoletti (2021), 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) use 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 (2020) and Landvoigt, Piazzesi and Schneider (2015). Anenberg and Bayer (2020), 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 (2021), 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.[23] Our estimates are thus roughly in line with the estimates from Giacoletti (2021) 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 (2021), 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 Data Validation

Our time-on-market data agree fairly well with two other 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 A6, 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.

[23] 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 (2021) and Landvoigt, Piazzesi and Schneider (2015).

Figure A7 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.

D Supplementary Material for Section 7

D.1 Moment Matching

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

D.1.1 Inner Loop

Given any value of Δ_c, we choose $\lambda_M, M_B, \bar{c}, \epsilon_0, \sigma_\epsilon^2$ to match the average values of five moments, as we describe in the main text. Three of the target moments come from a sample of counties in 2016. The two target moments from the literature are the average number of houses that buyers visit before buying, which Genesove and Han (2012) find to be 9.96; and the dispersion in buyer values for houses, which Table 3 of Anundsen et al. (2023) estimates to be 0.0465 for houses. Estimated buyer dispersion is somewhat smaller for apartments than houses; we use houses because these are likely to be most comparable to the SFRs in our sample. The counterpart of this empirical moment in the model is the dispersion in buyers' WTP for houses, divided by expected prices; that is,

$$\frac{1}{E\left(P\left(c,\epsilon\right)\right)}\sqrt{\left(\frac{\theta}{r+\lambda_M}\right)^2\sigma_\epsilon^2}=0.0465 \tag{71}$$

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

Since buyer- and seller-induced price variance do not add up to total price variance, there is a residual term. This could be driven by a number of factors: such as unobserved house-level heterogeneity or renovations, or realtor bargaining frictions. Using our estimates, we can do a simple accounting of roughly how much of total price dispersion, under our estimates, is attributable to preferences. That is, we can decompose total logSD in the data – which on average across counties is 16.8% – into components attributable to buyer values, seller values, and residuals:

$$(0.168)^2\approx\underbrace{\frac{Var_{c\sim F(\cdot)}\left(V_S\left(c\right)\right)}{\left(E\left(P\left(c,\epsilon\right)\right)\right)^2}}_{Seller\ holding\ value}+\underbrace{\frac{1}{\left(E\left(P\left(c,\epsilon\right)\right)\right)^2}\left(\frac{\theta}{r+\lambda_M}\right)^2\sigma_\epsilon^2}_{Buyer\ match\ value}+\sigma_{residual}^2 \tag{72}$$

In our baseline model estimates, the seller holding value component has a standard deviation of 3.93% of prices, and the buyer component is 4.65% of house prices. In terms of variance fractions, 5.48% of total logSD is attributable to seller values, 5.48% to buyer values, and 86.86% 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 M_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 $\lambda_M, M_B, \bar{c}, \Delta_c, \epsilon_0, \sigma_\epsilon^2$ to obtain percentage price dispersion. We then regress simulated percentage price dispersion from our model on simulated time-on-market, to generate a model-predicted regression coefficient. We choose Δ_c to match this model-predicted TOM-logSD relationship to the county panel coefficient.

D.1.3 Computation

Given a vector of parameters $r, \alpha, \phi, \theta, \lambda_M, M_B, \bar{c}, \Delta_c, \epsilon_0, \sigma_\epsilon^2$ we solve the model by iteratively solving the Bellman equations and flow equality conditions until convergence. Given guesses for $f_{eq}(c), M_S$, we calculate λ_S and λ_B, and then numerically solve (5), (6), (7) for $V_S(c), V_M(\epsilon), V_B, \epsilon^*(c)$. 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 $\bar{P}$:

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

Hence, we generate the variance of prices as the variance of model-generated prices, divided by the squared mean of model-generated prices, where we calculate the variance of model-generated prices using expression (13) 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 A4 below. For the non-foreclosure papers, we estimate 1-month liquidity discounts from a number of papers which analyze various shifters of 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. The foreclosure papers do not report differences in time-on-market between foreclosed and non-foreclosed houses; thus, we simply report average foreclosure discounts for these papers.

Liquidity discounts

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

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

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

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

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

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

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

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

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

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

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

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

- Guren (2018) uses data from Los Angeles, San Diego, and San Francisco, from 1988-2013. Similar to Genesove and Mayer (1997), Guren uses price appreciation since purchase as an instrument for sellers' marginal utility for cash, and thus sellers' urgency. While Guren emphasizes the curvature of demand, we can use the slope of demand to estimate the 1-month price effect, using the IV estimates in his Figure 2. In this figure, a relative markup change of 4% – that is, from -0.02 to 0.02 – changes the 13-week sale probability of a house from 0.5 to 0.4. Assuming that house sales follow a Poisson process, and assuming 4 weeks in a month, the sale probabilities map to expected time-on-markets of 6.5 months and 8.125 months, respectively. Assuming list prices pass through perfectly to sale prices, this gives a lower estimate of the 1-month effect of:

$$\frac{4}{8.125 - 6.5} = 2.46\%$$

We note that the estimated time-on-markets, using this method, are much higher than our estimate of 2.6 months. If we instead assume time-on-market increases by 25%, using a base of 2.6 months, we can calculate an upper estimate of the 1-month effect as:

$$\frac{4}{\left(\frac{8.125}{6.5}-1\right)(2.6)} = 6.15\%$$

Another caveat to note is that our assumption that list prices pass through perfectly to sale prices is likely an overestimate. In support of this assumption, however, in Appendix D.1, Guren writes that "the modal house sells at its list price" in his sample.

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

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

Foreclosure discounts

- Pennington-Cross (2006), using confidential data, finds foreclosure discounts or around 22% of house prices.
- Clauretie and Daneshvary (2009), using data from Las Vegas from 2004-2007, find foreclosure discounts averaging around 10%.
- Campbell, Giglio and Pathak (2011), using data from Massachusetts from 1987-2009, find foreclosure discounts of around 27%. Note that Campbell, Giglio and Pathak also analyze discounts from "forced sales", driven by deaths or bankruptcies of sellers, but which are not foreclosures. These discounts are much smaller, at 3-7%.
- 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.

D.4 Sensitivity Analysis

Our calibration uses a number of moments and parameters from the literature; a natural question is how sensitive our quantitative conclusions are to these input values. To address this, we re-calibrate our model varying various inputs, and show how results change. We consider sensitivity to three parameters: θ, the bargaining power parameter; the percentage dispersion in buyers' values from houses, which we take from Anundsen, Larsen and Sommervoll (2019); and the number of times buyers view a house before buying, from Genesove and Han (2012). For θ, we test the results from setting $\theta = 0.75$ and 0.25. For the buyer value dispersion and average number of view target moments, we test sensitivity by re-estimating the model assuming the target moment is either 25% higher or lower than the baseline value. In each case, we re-estimate all other parameters to match target moments.

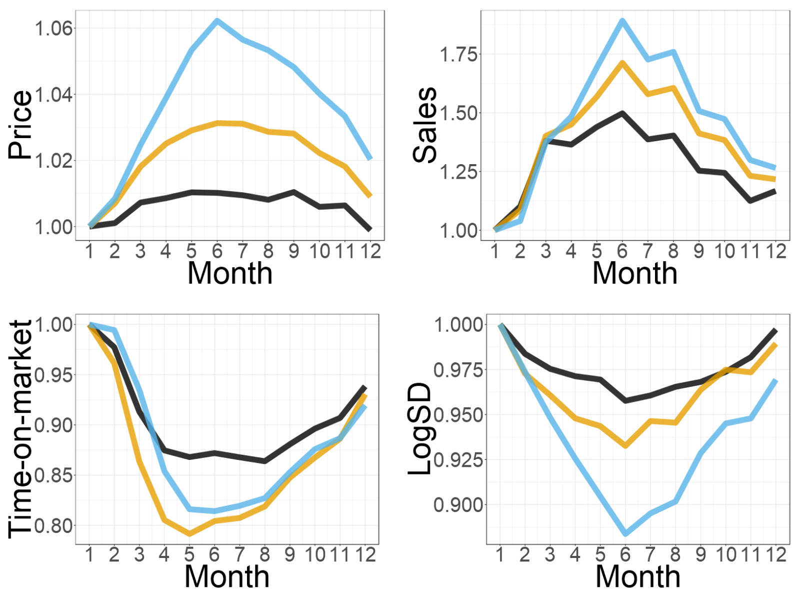
Table A5 shows how these alternative settings influence our parameter estimates. Varying θ substantially influences our estimate of σ_ϵ and Δ_c, both core parameters for our outcomes. Similarly, varying the buyer value dispersion moment influences our estimates of σ_ϵ substantially, though it has a smaller effect on Δ_c. The buyer number of views target moment influences M_B substantially, with lower effects on other parameters. λ_M is fairly stable and unaffected by any of these parameters.

Table A6 then shows how our core counterfactual outcomes vary across these alternative settings. The first column shows how liquidity discounts in the calibrated model, for selling 1 month faster, vary across the alternative settings; liquidity discounts are in the range of 3-8% for all settings.

The next 6 columns show how our results in Subsection 7.2 vary across the alternative settings; we show, for a 5% TOM decrease, paired with either a 5% decrease or 5% increase in PD, what we infer changes in liquidity supply and demand to be, and what we infer changes in prices to be. Comparing columns 2, 3, 5, and 6, we see that, while the signs of liquidity supply and demand shifts are preserved across specifications, the magnitudes vary somewhat. However, somewhat surprisingly, columns 4 and 7 show that we have reasonably robust estimates of the price changes associated with changes in TOM and PD: the scenario with a decrease in TOM and PD is associated with a roughly 3-4% increase in prices across all alternative specifications, and the scenario with a decrease in TOM and an increase in PD is associated with a roughly 6-7% decrease in prices in most scenarios. Together, these results suggest that our results are at least directionally robust to some alternative assumptions about calibrated inputs; in the case of price changes, the magnitudes are also surprisingly robust.

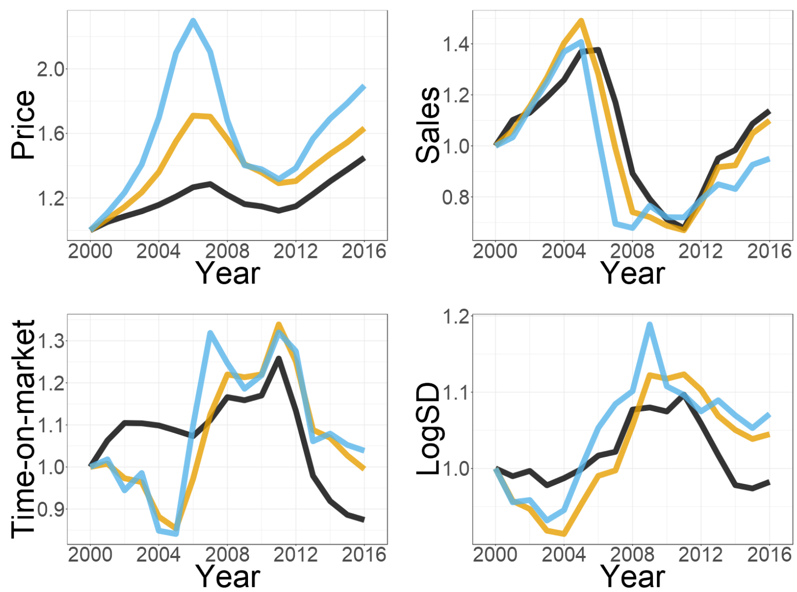
E Appendix Figures and Tables

Figure A1. Seasonal Variation in Sales, Prices, LogSD, and TOM, Heterogeneity



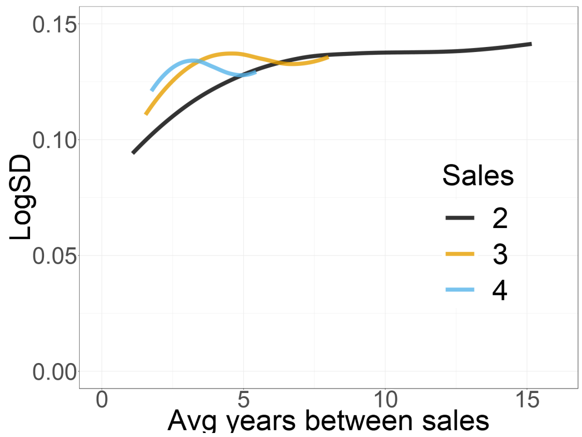
Notes. Prices, sales, time-on-market, and logSD for three quantile buckets of counties, divided based on the ratio between summer and winter prices. Each color represents one quantile bucket of counties. The time period of the data is 2000 to 2016. All variables are indexed by dividing by their January level. $LogSD$, our measure of idiosyncratic price dispersion, is calculated according to specification (1). Time-on-market 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*, Time-on-market, 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 on data construction are described in Appendix A.5.

Figure A2. Business Cycle Variation in Sales, Prices, LogSD, and TOM, Heterogeneity



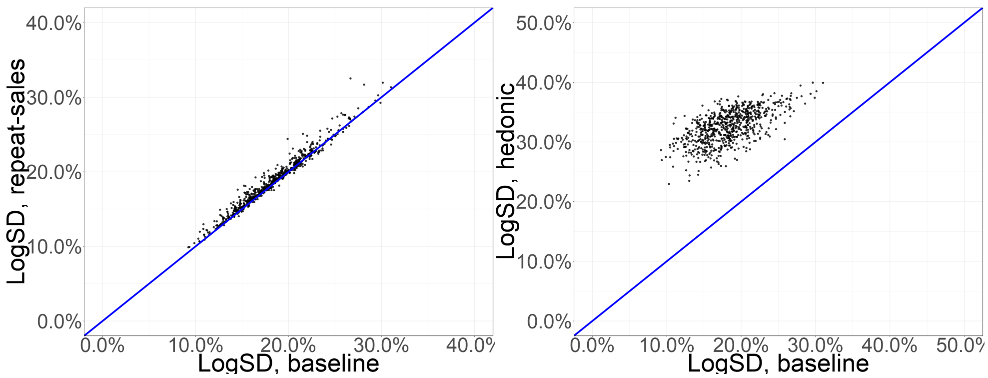
Notes. Yearly prices, sales, time-on-market, and logSD for three quantile buckets of counties, divided into three buckets based on the ratio between average prices in 2000 and 2005. Each color represents one quantile bucket of counties. The time period of the data is 2000 to 2016. $LogSD$, our measure 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 on data construction are described in Appendix A.5.

Figure A3. Effect of Number of Sales and Time-Between-Sales on LogSD



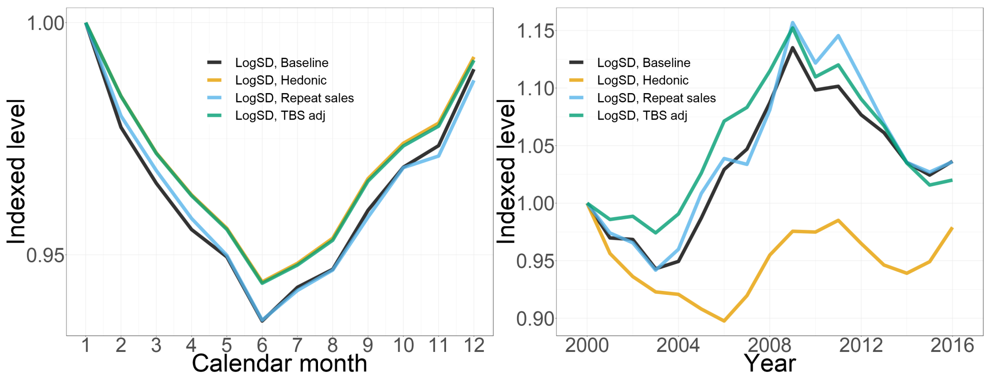
Notes. Variation of our estimated idiosyncratic house price residuals, $\hat{\epsilon}^2_{it}$, with respect to the average time between house sales, and the number of times a house is sold. We calculate $\hat{\epsilon}^2_{it}$ for each house sale using specification (1), and then run a kernel regression of $\hat{\epsilon}^2_{it}$ 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 $\hat{\epsilon}^2_{it}$.

Figure A4. LogSD: Alternative Measurements



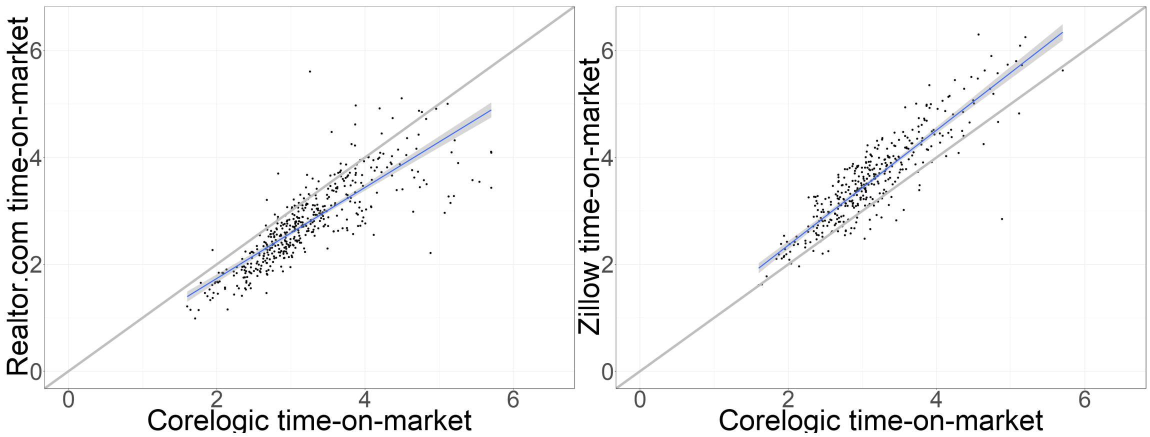
Notes. The left panel shows, on the x-axis, estimates of $\hat{\sigma}_c$ from the baseline specification, (1), and on the y-axis, estimates of $\hat{\sigma}_c$ from specification (67). Specification (67) is a pure repeat-sales model of house prices, omitting the $f_c(x_i, t)$ term from specification (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 (68). Specification (68) is a pure hedonic model of house prices, omitting the house fixed effect term γ_i from specification (1). In both plots, the sample period is 2012-2016. Each data point represents one county.

Figure A5. LogSD: Alternative Measures in the Time-Series



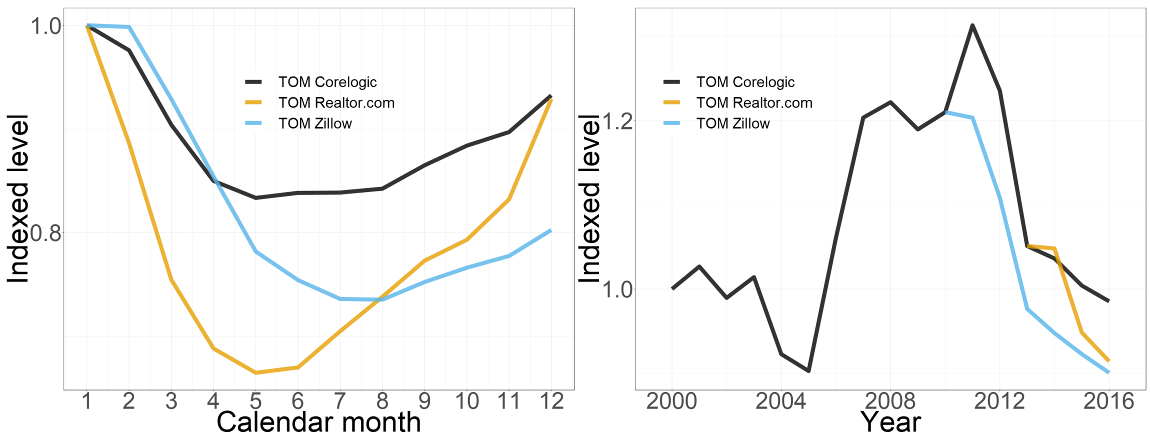
Notes. The left panel shows the four estimates of price dispersion – the baseline specification (1), the pure repeat-sales model (67), the pure hedonic model (68), and the time-between-sales adjusted estimates (70), over calendar months. For all variables, we filter out low-frequency trends by fitting 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 A6. Time-on-Market: Alternative Measurements



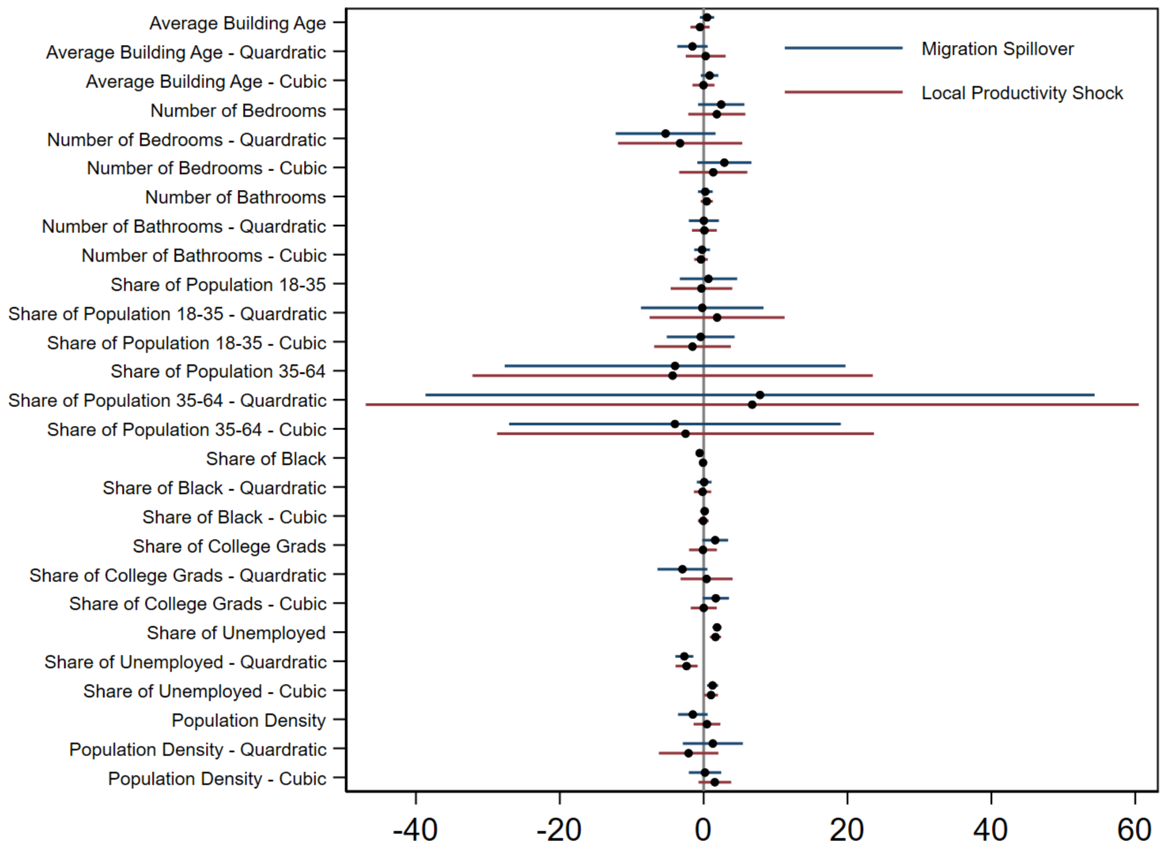
Notes. The left panel shows Corelogic MLS time-on-market on the x-axis, against Realtor.com time-on-market on the y-axis. The right panel shows Corelogic MLS time-on-market on the x-axis, against Zillow time-on-market on the y-axis. In both plots, the sample period is 2012-2016. Each data point represents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

Figure A7. Time-on-Market: Alternative Measurements in the Time Series



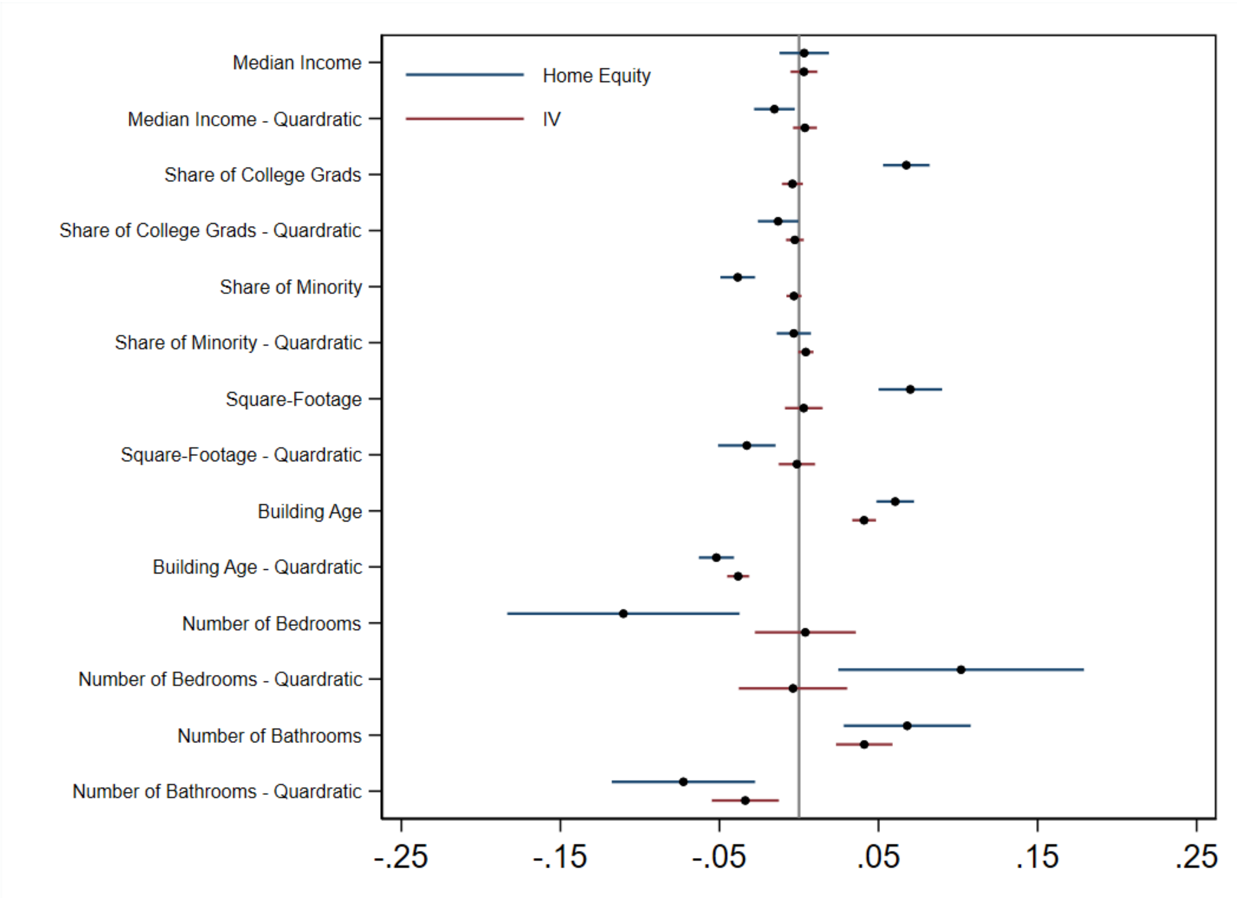
Notes. The left panel shows our three estimates of time-on-market – from Corelogic, Realtor.com, and Zillow – over calendar months. For all variables, we filter out low-frequency trends by fitting a piece-wise linear trend with break points every 3 years, subtracting away the predicted values, and adding 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.

Figure A8. Migration Spillover Instrument — Balance Test



Notes. This figure shows how the migration spillover instrument, $M_{c,2012-2016}$, is correlated with local county characteristics. The data period is 2012-2016. Each data point is a county. The figure plots the coefficient estimates, and the bars indicate 95% confidence intervals.

Figure A9. Liquidity Demand — Balance Test



Notes. This figure shows how our liquidity demand shifter, constructed based on home equity ratios and zipcode price growth as described in Subsection 6.1, is correlated with property characteristics. The data period is 2009-2017. Each data point is a distressed property. The figure plots the coefficient estimates, and the bars indicate 95% confidence intervals.

Table A1

Characteristics of Counties in our Dataset

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.

	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

Table A2

Liquidity Supply Regressions — Alternative IV

This table presents county level regression results about liquidity supply using local productivity shocks as an alternative IV. Column 1 is the first stage, columns 2-4 are the second stage, and columns 5-7 are the reduced IV results. In all columns, each data point is a county, the sample time period is 2012-2016, and regressions are weighted by the total number of sales within the county. The dependent variable in column 1 is population growth rate, obtained from the ACS. The dependent variable in columns 2 and 5 is $LogSD$, our measure of idiosyncratic price dispersion calculated according to specification (1). The dependent variable in columns 3 and 6 is time-on-market, expressed in month, from the Corelogic MLS data. The dependent variable in column 4 and 7 is log price growth, from Zillow price index. Local shock is each county's own Bartik productivity shock, $B_{c,2012-2016}$. 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. Robust standard errors are reported in parentheses. ***, **, * represent 1%, 5%, and 10% significance, respectively.

	2SLS						
	First Stage	Second Stage			Reduced-Form		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Pop Growth	LogSD	TOM	Log Price Growth	LogSD	TOM	Log Price Growth
Local Shock	0.46***				-0.64***	-0.10**	1.09***
	(0.174)				(0.218)	(0.038)	(0.265)
Population Growth		-1.40**	-0.21**	2.32***			
		(0.653)	(0.096)	(0.797)			
Controls	✓	✓	✓	✓	✓	✓	✓
Observations	400	400	400	394	400	400 394	
Adjusted R^2	0.77	-	-	-	0.56	0.33	0.28
Underidentification t-stat		7.57	7.57	7.71			
Underidentification p-value		0.01	0.01	0.01			
Weak identification t-stat		6.96	6.96	7.01			
Hansen J Stat		0.00	0.00	0.00			

Table A3

Liquidity Demand — IV First Stage

This table presents the IV first stage results for our liquidity demand analysis. Each observation is a zipcode-month. Column 1 focuses on distressed sales. Column 2 focuses on non-distressed sales. Column 3 uses the sample combining distressed and non-distressed sales. The outcome variable is zipcode home equity. The variable of interest is our instrument for liquidity demand, which is the average county-level house price growth since the previous transaction date across all individual house transactions in the zipcode-distressed status-month. Zipcode controls include zipcode average square footage, number of bedrooms, number of bathrooms, their squared terms, share of minority, share of college graduates, median income, and their squared terms. The sample period is from 2009 to 2017. Standard errors are clustered at zipcode level. ***, **, * represent 1%, 5%, and 10% significance, respectively.

	Home Equity		
	(1)	(2)	(3)
	Distressed Sales	Non-Distressed Sales	Full Sample
County House Price Growth	0.85***	0.53***	0.61***
	(0.009)	(0.005)	(0.006)
Property and Zipcode Controls	✓	✓	✓
Property Age Group FE	✓	✓	✓
County-Year FE	✓	✓	✓
Adjusted R^2	0.72	0.52	0.70
Observations	98,628	238,512	337,479

Table A4

Quantification: Liquidity Discounts in the Literature

This table presents the estimates of 1-month price effects and foreclosure discounts from the literature. Panel A shows estimates from non-foreclosure shifters of sellers' values; for these papers, the estimates correspond to how much prices would increase if time-on-market increased by a month. Panel B shows estimates of foreclosure discounts; these estimates compare foreclosure prices or returns, to prices on comparable houses which were not foreclosed on. These papers largely do not report the time-on-market difference between foreclosure and non-foreclosure houses, so these are not one-month effects. Further details of how we calculated these quantities are described in Appendix D.3.

Panel A: Non-Foreclosure

Paper	**1-month effect**
Genesove and Mayer (1997)	11.02pp
Genesove and Mayer (2001)	1.92pp-24pp
Levitt and Syverson (2008)	11.68pp
Hendel, Nevo and Ortalo-Magné (2009)	4.50pp
Guren (2018)	2.46pp-6.15pp
Buchak et al. (2020)	1.92pp-4.46pp

Panel B: Foreclosure

Paper	**1-month effect**
Pennington-Cross (2006)	22pp
Clauretie and Daneshvary (2009)	10pp
Campbell, Giglio and Pathak (2011)	27pp
Harding, Rosenblatt and Yao (2012)	5pp
Zhou et al. (2015)	11pp-26pp

Table A5

Quantification: Parameter Estimate Sensitivity to Input Moments

This table shows estimated parameters for the baseline model, and 6 different alternative specifications: high and low θ; high and low assumptions for buyer value dispersion as a fraction of prices; and high and low assumptions for the number of times buyers view houses on average before purchasing.

Case	λ_M	ϵ_0	σ_ϵ	M_B	$\bar{c}$	Δ_c
Baseline	0.052	0.7390	0.2159	1.062	4.700	3.864
High Theta	0.052	0.7753	0.1316	1.071	5.481	5.530
Low Theta	0.056	0.3928	0.4484	1.162	3.982	1.944
High Rel Buyer SD	0.053	0.6159	0.2731	1.095	5.918	4.066
Low Rel Buyer SD	0.052	0.8635	0.1619	1.039	3.814	3.965
High N Views	0.052	0.6677	0.2159	1.384	4.701	3.864
Low N Views	0.052	0.8421	0.2158	0.752	4.651	3.864

Table A6

Quantification: Counterfactual Estimate Sensitivity to Input Moments

This table shows estimated counterfactual magnitudes for the baseline model (row 1) along with 6 alternative specifications in which we consider high and low θ (rows 2 and 3), high and low assumptions for buyer value dispersion as a fraction of prices (rows 4 and 5), and high and low assumptions for the number of times buyers view houses on average before purchasing (rows 6 and 7). Column 1 shows liquidity discounts, i.e., the implied price discount for selling 1 month faster, calculated as described in Subsection 7.1 in the main text. Columns 2 and 3 show, if we observe TOM and PD both decrease by 5%, what we infer the changes in M_B and μ_c to be which rationalize this change, in each specification. Column 4 shows what we then infer the price change to be in each specification. Columns 5-7 show the inferred changes in M_B, μ_c, and prices when TOM decreases by 5%, but PD increases by 5%.

		TOM ↓ & PD ↓			TOM ↓ & PD ↑		
	(1) Liquidity Discount	(2) M_B Change	(3) μ_c Change	(4) Price Change	(5) M_B Change	(6) μ_c Change	(7) Price Change
Baseline	4.03%	15.53%	3.73%	3.37%	-22.69%	6.18%	-6.48%
High Theta	3.11%	32.30%	2.08%	3.69%	-36.48%	4.00%	-6.19%
Low Theta	7.46%	6.16%	5.23%	3.17%	-10.17%	8.11%	-6.66%
High Rel Buyer SD	4.97%	13.48%	4.14%	3.68%	-17.76%	6.09%	-6.20%
Low Rel Buyer SD	3.30%	17.97%	2.75%	2.98%	-29.92%	5.76%	-6.82%
High N Views	4.04%	16.98%	3.73%	3.37%	-24.53%	6.18%	-6.48%
Low N Views	4.00%	13.37%	3.70%	3.35%	-20.12%	6.18%	-6.50%