Search Frictions and Idiosyncratic Price Dispersion in the US Housing Market∗

Nadia Kotova† and Anthony Lee Zhang‡

May 2019

Abstract

This paper studies the sources of idiosyncratic house price dispersion. We construct a search-and-bargaining model of the housing market, which predicts that idiosyncratic price dispersion should be positively correlated with time-on-market and negatively correlated with house prices and sales volume. Using a methodology which combines repeat-sales and hedonic approaches, we measure idiosyncratic price dispersion across locations and over time. We show that idiosyncratic price dispersion is countercyclical and seasonal, and that it is associated with prices, sales volume, and time-on-market in the directions predicted by our model, cross-sectionally as well as in panel regressions.

∗We appreciate comments from Mohammad Akbarpour, Sam Antill, Dmitry Arkhangelsky, Adrien Auclert, Lanier Benkard, Tim Bresnahan, Scarlet Chen, Daniel Chen, John Cochrane, Cody Cook, Rebecca Diamond, Evgeni Drynkim, Darrell Duffie, Liran Einav, Matthew Gentzkow, Steven Grenadier, Guido Imbens, Chad Jones, Eddie Lazear, Paul Milgrom, Jonas Mueller-Gastell, Michael Ostrovsky, Monika Piazzesi, Franklin Qian, Peter Reiss, Al Roth, Martin Schneider, Jesse Shapiro, Erling Skancke, Paulo Somaini, Rose Tan, Zach Taylor, Chris Tonetti, Bob Wilson, and Ali Yurukoglu, as well as seminar participants at Stanford.

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

‡Stanford Graduate School of Business, 655 Knight Way, Stanford, CA 94305; anthonyz@stanford.edu.
1 Introduction

Real estate represents a large component of household wealth in the US. The representative household in the US does not hold a diversified real estate portfolio – most households only own a single house, their primary place of residence. The literature has shown that a large fraction of the total price risk associated with owning houses is idiosyncratic to individual house purchases. However, the sources of idiosyncratic house price dispersion are not yet well understood.

This paper argues that search frictions are an important determinant of idiosyncratic house price volatility. We build a simple search-and-bargaining model of the housing market, which predicts that idiosyncratic price dispersion should be lower when sales volume and prices are high and time-on-market is low. Bringing the model to data, we show that price dispersion is robustly correlated with prices, volume, and time-on-market in the directions predicted by the model, over the business cycle, seasonally, cross-sectionally, and in panel regressions.

In our model, there is an exogenously specified mass of buyers who wish to buy a house and become a homeowner in a given market. Matched homeowners periodically receive separation shocks, which allow them to list their houses on the market and become sellers. Sellers have heterogeneous holding costs which they bear per unit time that their houses are on the market. Since prices are set through Nash bargaining, sellers’ holding costs affect trade prices: sellers with higher holding costs receive lower trade prices from identical buyers, generating dispersion in the prices of identical houses in equilibrium. The model predicts that there is a tight relationship between price dispersion and time-on-market: fixing the distribution of sellers’ holding costs, increasing time-on-market causes sellers’ outside options to become more disperse, increasing equilibrium price dispersion. Thus, the model predicts that price dispersion is lower in tight markets, when the mass of buyers is large, volume and prices are high, and time-on-market is low.

We measure idiosyncratic price dispersion using microdata from CoreLogic on 36 million house sales over the period 2000-2017. We use a novel measurement strategy, which combines repeat-sales and hedonic methodologies: we regress sale prices on zipcode-month fixed effects and house fixed effects, as well as a smooth function of house characteristics and time. This specification allows both observed and unobserved attributes of houses to have time-invariant effects on the level of a house’s price, while also allowing observable house characteristics to affect a house’s price path over time. We
measure idiosyncratic price dispersion using the residuals from this regression, which we can aggregate flexibly over locations and over time.

We proceed by demonstrating four stylized facts about idiosyncratic price dispersion. First, idiosyncratic price dispersion is countercyclical, decreasing during the housing boom of 2000-2004, increasing during the 2008 bust, and decreasing during the subsequent recovery. Second, idiosyncratic price dispersion is seasonal at the yearly level: it is lower in summer than winter, when prices and volume are highest and time-on-market is lowest. Third, in the cross-section of zip codes, idiosyncratic price dispersion is robustly associated with time-on-market over the period 2012-2016. Fourth, in panel regressions, zipcode-years that experience unusual increases in volume or prices, or decreases in time-on-market, experience decreases in price dispersion. All these empirical facts support the predictions of our model, tying together the dynamics of prices, volume, time-on-market and price dispersion.

Our results have implications for research on household portfolio choice. Since idiosyncratic price dispersion is countercyclical and seasonal; thus, households who are selling houses during winter, or during recessions, are exposed not only to lower prices but also to increased idiosyncratic volatility. In the seasonal case, we find that the size of the increase in idiosyncratic volatility is approximately half the magnitude of the decrease in average price. Thus, variation in idiosyncratic house price variance over time contributes nontrivially to risk in households’ portfolios over time. Our results also provide a counterpoint to “irrational exuberance” theories of housing booms. Housing booms may destabilize house prices on average, but idiosyncratic price dispersion actually decreases during boom periods. This does not imply that bounded rationality plays no role in housing booms, but suggests that, even if booms are generated by some form of irrational exuberance, market forces appear to discipline the relative prices of houses fairly well during boom periods.

1.1 Related literature

This paper is related to a number of strands of literature. Most directly, this paper is related to a literature studying idiosyncratic house price volatility. Case and Shiller (1989) is one of the first papers to show that prices of individual houses are much more volatile than city-wide average prices. Sagi (2015) studies idiosyncratic risk in commercial real estate, showing that the idiosyncratic component of house price risk does not scale
with holding period, and calibrates a model to fit the data. Giacoletti (2017) studies the residential real estate market in California, showing that idiosyncratic house price risk does not scale with holding period, suggesting that much of idiosyncratic house price volatility is due to liquidity risk; in addition, using an instrument for zipcode-level shocks to mortgage credit, Giacoletti shows that decreasing local credit availability increases idiosyncratic risk. Peng and Thibodeau (2017) calculates price dispersion at the zipcode level using a hedonic regression specification, and documents relationships between idiosyncratic price dispersion and characteristics of zipcodes such as average income.

Our work builds on the results of Sagi (2015) and Giacoletti (2017). Our measurement strategy, which is novel to the literature, builds on the observation that the holding-period structure of idiosyncratic risk is flat, and allows us to study how idiosyncratic risk varies over time, accommodating time-invariant house fixed effects while also flexibly controlling for time-varying effects of observable characteristics on prices. This methodology allows us to discover new features of the behavior of idiosyncratic house price risk seasonally and over the business cycle. We discuss in detail the relationship between our method for measuring idiosyncratic house price dispersion, and existing methods in the literature, in subsection B.4.

More broadly, our work fits into the literature applying search models to housing markets\textsuperscript{1} and in financial markets more generally\textsuperscript{2}. To our knowledge, with the exception of Sagi (2015), we are the first paper to attempt to use search models to understand idiosyncratic house price dispersion. Our model accomplishes this by allowing sellers to have heterogeneous and persistent holding costs, which determine trade prices together with the standard match quality shock. This generates the main comparative static of our model: the effect of sellers’ holding costs on equilibrium price variance depends on equilibrium time-on-market. Relative to Sagi (2015), our model differs technically in that it is closer to the Diamond-Mortensen-Pissarides labor search framework\textsuperscript{3} and practically in that the model is aimed at highlighting the connections between sales, volume, time-on-market, and idiosyncratic price dispersion.

The fact that volume, prices, and time-on-market are correlated in housing markets

\textsuperscript{1}See, for example, Wheaton (1990), Piazzesi and Schneider (2009), Genesove and Han (2012), Ngai and Tenreyro (2014), Head, Lloyd-Ellis and Sun (2014), Piazzesi, Schneider and Stroebel (2015), Sagi (2015), Albrecht, Gautier and Vroman (2016).

\textsuperscript{2}See, for example, Duffie, Gärleanu and Pedersen (2005). Our model is a continuous-time random matching model, as described in Duffie, Qiao and Sun (2017).

\textsuperscript{3}See, for example, Mortensen and Pissarides (1994), or the survey article Rogerson, Shimer and Wright (2005).
is the subject of a fairly large body of research; see, for example, Stein (1995), Krainer (2001), Genesove and Mayer (2001), Leung, Lau and Leong (2002), Clayton, Miller and Peng (2010), Diaz and Jerez (2013), and DeFusco, Nathanson and Zwick (2017). Our contribution to this literature is to show how idiosyncratic price dispersion co-moves with these variables.

1.2 Outline

The rest of the paper proceeds as follows. Section 2 constructs a search-and-bargaining model of the housing market and describes the theoretical predictions. Section 3 describes the data and how we construct our measure of price dispersion. Section 4 contains our empirical results. Section 5 discusses various implications of our findings.

2 Model

2.1 Primitives

We model a housing market in a single, isolated geographical location, with 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.

2.1.1 Sellers

There is some mass $M_S$ of sellers in the market, who are unmatched with their houses and are waiting to sell their houses to buyers. Sellers who sell their house permanently leave the market thereafter, with continuation utility normalized to 0. Sellers are heterogeneous, and must pay some holding cost, $c > 0$, per unit time that they keep their house on the market. As we will describe in subsection 2.1.3, holding costs $c$ are drawn from a distribution $F(\cdot)$, at the time that a matched owner chooses to unmatch with her house and become a seller, and do not vary over time. Holding costs reflect differences in sellers’ urgency to sell. Sellers with high values of $c$ may be credit-constrained, perhaps because they need to sell their existing house before they can purchase a new house; sellers with low $c$ values may be willing to wait longer to sell at higher prices, perhaps because they
are less capital-constrained, or because they are able to rent their house out while it is in the market.

In equilibrium, a seller of holding cost \( c \) will have some expected value \( V_S(c) \). A seller who matches with a buyer and sells her house for some price \( P \) receives utility \( P \) and continuation utility 0, so her net utility gain is \( P - V_S(c) \). We will characterize \( V_S(c) \) in subsection 2.2.2 below.

### 2.1.2 Buyers

There is a stationary mass \( M_B \) of potential buyers who wish to purchase houses from sellers and become matched homeowners. Buyers are ex-ante identical and receive flow value normalized to 0 while waiting to purchase a house. When a buyer is matched with a house, she draws a match value \( \epsilon \sim G(\cdot) \), which represents the utility she receives if she purchases the house. For simplicity, we treat \( \epsilon \) as a stock value which is attained the instant that the buyer purchases a house, rather than a flow value attained over time. We will assume that \( G(\cdot) \) is an exponential distribution, so \( \epsilon \) has some mean \( \sigma_\epsilon \) and variance \( \sigma_\epsilon^2 \). We assume that match value draws are independent across matches.

In equilibrium, unmatched buyers have some expected value \( V_B \). A buyer who matches with a house of quality \( \epsilon \) and pays price \( P \) to purchase the house receives stock utility \( \epsilon \), and then becomes a matched homeowner, attaining expected utility \( V_M \). Thus, the net utility gain for the buyer is:

\[
\epsilon + V_M - P - V_B
\]

We will characterize buyers’ expected value \( V_B \) in subsection 2.2.2 below.

### 2.1.3 Matched homeowners

Matched homeowners are buyers who have purchased houses, and who have not yet decided to become sellers. Since we have assumed that buyers receive their match quality draw \( \epsilon \) as a stock quantity at the moment of a home purchase, matched homeowners are identical; we normalize the flow utility that they receive from their houses to 0. Since there is a unit mass of houses in the market, the masses of matched homeowners and unmatched sellers must always sum to 1, so the mass of matched homeowners at any given moment is \( 1 - M_S \).

A matched homeowner receives separation shocks at Poisson rate \( \lambda_M \), at which point
the matched homeowner can decide to become a seller. Each time a matched homeowner receives a separation shock, she draws some separation value \( u_{sep} \sim H(\cdot) \), which determines her stock value from unmatching with her house and becoming a seller. A separation shock can be thought of as, for example, a job offer in a different city, which causes the homeowner to consider moving out of the market; \( u_{sep} \) can be thought of as the attractiveness of the job offer.

A matched homeowner will choose to unmatch with her house and become a seller if \( u_{sep} \) is sufficiently high, relative to the expected value from becoming a house seller. Suppose that a matched owner receives a separation shock with separation utility \( u_{sep} \). If she chooses to remain matched, her continuation utility is the equilibrium expected value \( V_M \) of matched owners. If she unmatches and becomes a seller, she attains her separation utility \( u_{sep} \), then receives the expectation of the seller value function \( V_S(c) \) over the distribution of seller costs \( c \). Matched owners will separate if the expected value of becoming a seller exceeds the expected value of remaining matched; thus, there is a cutoff utility \( u^*_{sep} \), satisfying

\[
 u^*_{sep} = V_M - \int V_S(c) \, dF(c)
\]

such that matched owners who receive moving shock with \( u_{sep} > u^*_{sep} \) will choose to unmatch and become sellers.

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

\[
 rV_M = \lambda_M \int_{u_{sep} > u^*_{sep}} [u_{sep} + V_S(c) - V_M] \, dF(c) \, dH(u_{sep})
\]

Intuitively, expression (1) states that matched owners receive separation shocks at rate \( \lambda_M \), at which point they draw \( u_{sep} \) from the distribution \( H(\cdot) \). If \( u_{sep} > u^*_{sep} \), the matched owner becomes a seller, attaining her separation utility \( u_{sep} \), plus the expectation of the seller value function \( V_S(c) \) over the seller cost distribution \( F(c) \), net of the matched owner value \( V_M \).

The fact that \( u_{sep} \) is a nondegenerate random variable implies that matched owners’ decisions to put their houses on the market are sensitive to market conditions. If the expectation of \( V_S(c) \) is high, and matched homeowners expect that they will achieve high surplus on the market when they become sellers, matched owners are more likely to accept separation shocks and become sellers. This causes sales volume to be higher in
markets where the mass of buyers, \( M_B \), is large, and thus the expected value of being a seller, \( V_S(c) \), is high.

### 2.1.4 Matching

Suppose that \( M_B \) unmatched buyers and \( M_S \) unmatched sellers are present in the market. We assume that matches between buyers and sellers are generated at a flow rate, which depends on \( M_B \) and \( M_S \) through a *matching function*, \( m(M_B, M_S) \). We will assume that the matching function 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 happens at Poisson rates \( \lambda_B \) and \( \lambda_S \) respectively; these are given by:

\[
\lambda_B = \frac{m(M_B, M_S)}{M_B}, \quad \lambda_S = \frac{m(M_B, M_S)}{M_S}
\]

The distribution of flow utilities \( c \) of the stationary mass \( M_S \) of sellers will, in general, differ from the distribution of values \( F(\cdot) \) among sellers entering the market, since sellers with lower holding costs will hold on to their houses for longer. We use \( F_{eq}(c) \) to denote the distribution of seller values in stationary equilibrium; we describe how \( F_{eq}(c) \) is determined in subsection 2.2.1 below. Thus, in stationary equilibrium, a matched buyer meets a seller with value \( c \) randomly drawn from \( F_{eq}(c) \).

### 2.1.5 Price determination

We assume that prices are set using Nash bargaining. Suppose that a buyer is matched with a seller with cost \( c \), and the buyer draws match-specific value \( \epsilon \). If the buyer and seller trade, the sum of their utilities is the buyer’s continuation utility \( \epsilon + V_M \) since we have normalized the continuation utility of sellers who leave the market to 0. If they do not trade, the buyer receives \( V_B \) and the seller receives \( V_S(c) \). Thus, the bilateral match surplus from trade is

\[
\epsilon + V_M - V_B - V_S(c)
\]

Trade occurs if the bilateral match surplus is nonnegative; that is, a buyer who draws match quality \( \epsilon \) will purchase a house from a seller with type \( c \) if:
\[ \epsilon + V_M - V_S(c) - V_B > 0 \]

This implies that, for each seller type \( c \), there is a cutoff match quality, \( \epsilon^*(c) \), such that the seller trades with all buyers with match quality draws higher than \( \epsilon^*(c) \), where:

\[ \epsilon^*(c) = V_B + V_S(c) - V_M \]

We assume that, when trade occurs, prices are set through Nash bargaining: the price that the seller receives is equal to her outside option, \( V_S(c) \), plus a share \( \theta \) of the bilateral match surplus. That is, the price \( P(\epsilon, c) \) from a match between a seller of type \( c \) and a buyer with match value \( \epsilon \) is:

\[ P(\epsilon, c) = \theta (\epsilon + V_M - V_B - V_S(c)) + V_S(c) \quad (2) \]

2.1.6 Discussion of assumptions

Our model for agents’ entry rates and utility functions contains a number of simplifying assumptions. We assume that buyers and matched owners are ex-ante identical, so only sellers have persistent holding costs, which are drawn after they have decided to unmatch from their houses. This is largely for analytical convenience, as it implies that buyers’ and matched owners’ expected values are scalars, so only sellers have nonconstant value functions that depend on \( c \) in equilibrium. We also assume, for simplicity, that sellers’ values are completely persistent over time, ruling out the possibility that sellers become more impatient over time. We assume that there is an exogenous mass \( M_B \) of buyers. We could alternatively specify some rate \( b \) at which buyers enter the market and some rate \( \eta_B \) at which unmatched buyers exit; \( M_B \) can be thought of as the equilibrium mass of buyers resulting from such a process.

2.2 Equilibrium

2.2.1 Flow equality

We analyze the model in a stationary equilibrium, requiring the inflow and outflow rates of all types of agents to be equal. The equilibrium rate at which matched homeowners
separate and becoming sellers of type $c$, per unit time, is:

$$(1 - M_S) \lambda_M (1 - H(u_{sep}^*)) f(c)$$

In words, this is the product of the total mass of matched homeowners, $1 - M_S$; the rate at which shocks homeowners receive separation shocks, $\lambda_M$; the fraction of separation shocks which result in unmatching, $1 - H(u_{sep}^*)$; and the density $f(c)$ of entering sellers with value $c$.

In stationary equilibrium, this inflow rate must be equal to the rate at which sellers of type $c$ sell their houses and leave the market. The equilibrium rate at which sellers of type $c$ leave is:

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

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

Equating these two rates, we require:

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

Expression (3) automatically implies that the rate at which buyers turn into matched owners is equal to the rate at which matched owners become sellers. To see this, note that, integrating (3) over $c$, we have:

$$(1 - M_S) \lambda_M (1 - H(u_{sep}^*)) = \lambda_S M_S \int (1 - G(\epsilon^*(c))) f_{eq}(c) \, dc$$

The left hand side of expression (4) is the total rate at which matched owners become sellers. The right hand side is the rate at which sellers match with buyers multiplied by the probability of trade $1 - G(\epsilon^*(c))$ integrated over the equilibrium distribution of $c$, $f_{eq}(c)$, on the market. However, since each successful sale turns a buyer into a matched owner, the RHS of is also expression (4) is also equal to the flow rate of buyers turning into matched owners. Thus, (4) implies that the flows into and out of matched ownership are also equal.
2.2.2 Value functions

Given expression (2) for prices, we can write sellers’ and buyers’ continuous-time Bellman equations, which pin down $V_B$ and $V_S(c)$. Given the buyer match rate $\lambda_B$, trade cutoffs $\epsilon^*(c)$, the equilibrium distribution of seller values $F_{eq}(c)$, and the seller value function $V_S(c)$, the equilibrium value of buyers, $V_B$, must satisfy:

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

In words, expression (5) can be interpreted as follows. At rate $\lambda_B$, the buyer is matched to a seller with type randomly drawn from $F_{eq}(\cdot)$, and the buyer draws match quality $\epsilon$ from $G(\cdot)$. If the buyer’s match quality draw, $\epsilon$, is higher than the seller’s match quality cutoff, $\epsilon^*(c)$, trade occurs, and the buyer receives a share $(1 - \theta)$ of the bilateral match surplus.

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

$$rV_S(c) = -c + \lambda_S \int_{\epsilon > \epsilon^*(c)} \theta (\epsilon + V_M - V_B - V_S(c)) \ dG(\epsilon)$$  \hspace{1cm} (6)

In words, expression (6) states that a seller of type $c$ receives flow value $-c$ from their house while they are waiting for buyers. At rate $\lambda_S$, the seller meets a buyer with match value $\epsilon$ randomly drawn from $G(\cdot)$. If $\epsilon > \epsilon^*(c)$, trade occurs, and the seller receives a share $\theta$ of the bilateral match surplus.

2.2.3 Equilibrium conditions

Collecting equilibrium conditions from earlier subsections, in our model, stationary equilibrium is described by a set of equations listed in the following proposition.

**Proposition 1.** Given primitives $M_B, F(c), G(\epsilon), H(u_{sep}), \lambda_M, \alpha, \phi, \theta, r$, a stationary equilibrium of the model is described by a seller mass $M_S$, matching rates $\lambda_S, \lambda_B$, value functions $V_S(c), V_B, V_M$, and cutoffs $\epsilon^*(c), u_{sep}^*$, which satisfy the following conditions:

Buyer, seller, and matched owner Bellman equations:

$$rV_B = \lambda_B \int \int_{\epsilon > \epsilon^*(c)} [(1 - \theta) (\epsilon + V_M - V_B - V_S(c))] \ dG(\epsilon) \ dF_{eq}(c)$$  \hspace{1cm} (7)
\[ rV_S(c) = -c + \lambda_S \int_{\epsilon > \epsilon^*(c)} \theta (\epsilon + V_M - V_B - V_S(c)) \, dG(\epsilon) \]  
(8)

\[ rV_M = \lambda_M \int_{u_{sep} > u_{sep}^*} \int [u_{sep} + V_S(c) - V_M] \, dF(c) \, dH(u_{sep}) \]  
(9)

\(\epsilon\) and \(u_{sep}\) cutoffs:

\[ \epsilon^*(c) = V_S(c) + V_B - V_M \]  
(10)

\[ u_{sep}^* = V_M - \int V_S(c) \, dF(c) \]  
(11)

Matching rates:

\[ M_S\lambda_S = M_B\lambda_B = \alpha M_B^\phi M_S^{1-\phi} \]  
(12)

Flow equality:

\[ (1 - M_S) \lambda_M (1 - H(u_{sep}^*)) f(c) = \lambda_S M_S f_{eq}(c) (1 - G(\epsilon^*)) \]  
(13)

### 2.3 Price dispersion in equilibrium

The primary outcome of the model we are interested in is price dispersion. In the model, this corresponds to the variance of transaction prices, \( P(\epsilon, c) \), with respect to the joint distribution of \( c \) and \( \epsilon \) among buyers and sellers whose meetings result in trade in any given instant of time. Appendix A.1 shows that there is a simple expression for price variance, given in the following claim.

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

\[ \text{Var}_{c \sim F(\cdot)} (V_S(c)) + \theta^2 \sigma_\epsilon^2 \]  
(14)

Expression (14) shows that equilibrium transaction price vary because trading sellers have different value functions \( V_S(c) \), due to differences in their persistent holding costs \( c \), and because trading buyers have different idiosyncratic match quality draws \( \epsilon \). The buyer value term only depends on the bargaining parameter \( \theta \) and the match quality variance \( \sigma_\epsilon^2 \), which we have assumed are primitives of the model, so we will focus on the seller value term \( \text{Var}_{c \sim F(\cdot)} (V_S(c)) \). We cannot directly characterize \( V_S(c) \), but there is a simple analytical expression for its derivative. First, define the expected time-on-market
in equilibrium for a seller of type \(c\), \(\text{TOM}(c)\), as:

\[
\text{TOM}(c) = \frac{1}{\lambda_S (1 - G(e^*(c)))}
\]  

(15)

The term \(\lambda_S (1 - G(e^*(c)))\) is the product of \(\lambda_S\), the equilibrium rate at which sellers meet buyers, and \(1 - G(e^*(c))\), the fraction of meetings for a seller of type \(c\) that result in trade. It is thus the net rate at which type \(c\) sellers successfully sell their houses and leave the market, and therefore, its inverse is the expected time-on-market for a seller of type \(c\). Claim 2, proved in appendix A.2, characterizes \(V_S'(c)\).

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

\[
V_S'(c) = \frac{-\text{TOM}(c)}{r\text{TOM}(c) + \theta}
\]  

(16)

Claim 2 implies that \(V_S'(c)\) is negative – sellers with higher holding costs have lower expected values – and that the magnitude of \(V_S'(c)\) is strictly increasing in equilibrium time-on-market, \(\text{TOM}(c)\). The intuition behind claim 2 is that sellers’ holding costs \(c\) are incurred per unit time sellers spend on the market. In a market in which sellers are quickly matched to high-valued buyers and equilibrium time-on-market is low, sellers’ total expected holding costs are low, regardless of sellers’ holding costs per unit time, \(c\). Thus, dispersion in holding costs does not translate into substantial dispersion in sellers’ expected values, \(V_S(c)\). Conversely, if equilibrium time-on-market is high, sellers’ total holding costs are much higher, and the difference between the expected total holding costs borne by sellers with different values of \(c\) increases. As a result, dispersion in holding costs translates into more dispersion in sellers’ expected values. Thus, the pass-through of holding costs into sellers’ value functions, and thus price dispersion, is tightly linked to the equilibrium time-on-market function, \(\text{TOM}(c)\).

To formalize the link between time-on-market and price dispersion, appendix A.3 shows that, if time-on-market \(\text{TOM}(c)\) increases, pointwise for every \(c\), then equilibrium price dispersion will also increase.

Claim 3. Fix \(F(e), G(e), \theta, r\). Consider two sets of model parameters,

\[
\Theta_1 = (M^1_B, H^1(u_{sep}), \lambda^1_M, \alpha^1, \phi^1), \quad \Theta_2 = (M^2_B, H^2(u_{sep}), \lambda^2_M, \alpha^2, \phi^2)
\]

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

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

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

### 2.4 Comparative statics

We are interested in the predictions the model makes about how volume, prices, time-on-market, and price dispersion co-vary. We derive expressions for all four variables in the context of our model in appendix A.4. We numerically solve the model, setting primitives as follows: distributions of costs, match qualities and separation utilities are \( F(c) \sim U[-1, 0] \), \( G(\epsilon) \sim \exp(1) \), \( H(u_{sep}) \sim \exp(1) \), the separation rate is \( \lambda_M = 0.01 \), parameters of the matching function are \( \alpha = 1 \) and \( \phi = 0.5 \), the bargaining parameter is \( \theta = 0.5 \), and the discount rate is \( r = 1 \). We assume that variation in outcomes is driven by variation in \( M_B \); that is, we are assuming that variation in outcomes is driven by variation in buying pressure, represented by the mass of buyers who wish to enter the market and buy houses. We solve the model for a grid of \( M_B \) values, and we show how the four outcome variables of interest vary in figure 1. Appendix A.4 derives analytical expressions for all quantities plotted in figure 1.

Figure 1 demonstrates that, when \( M_B \) is increased holding all other model parameters fixed, total sales volume and average prices increase, while average time-on-market and price dispersion decrease. Intuitively, an increase in the mass of buyers \( M_B \) directly affects the rate at which sellers are matched to potential buyers, \( \lambda_S \). Sellers’ continuation payoff from searching improves, and sellers become more selective, increasing the realizations of \( \epsilon \) conditional on trade; both effects increase prices. Since sellers’ value functions are higher, matched sellers are more willing to unmatch and become sellers, so \( u_{sep}^* \) decreases. Equilibrium sales volume increases because both \( M_B \) and \( M_S \) increase. Since the mass of buyers \( M_B \) is larger, the seller match rate \( \lambda_S \) increases, so time-on-market decreases. Thus, due to claim 1, equilibrium price dispersion also decreases.

Figure 1 thus captures the core prediction of our model: idiosyncratic price dispersion should be low when time-on-market is low, and sales volume and prices are high. The

\[^4\text{However, this is counteracted by an increase in seller selectivity, that is, an increase in the cutoffs } c^{*} (c). \text{ In principle, this effect could be strong enough that time-on-market actually increases when } M_B \text{ increases. However, this does not occur in our simulation.}\]
Table 1: Descriptive statistics

<table>
<thead>
<tr>
<th>Name</th>
<th>Mean</th>
<th>SD</th>
<th>P10</th>
<th>P90</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sales</td>
<td>9314</td>
<td>4829</td>
<td>4969</td>
<td>15653</td>
</tr>
<tr>
<td>Average monthly sales</td>
<td>46</td>
<td>23</td>
<td>25</td>
<td>76</td>
</tr>
<tr>
<td>Mean price (x1000 USD)</td>
<td>425.3</td>
<td>742.8</td>
<td>140.3</td>
<td>691.9</td>
</tr>
<tr>
<td>Mean TOM (Months), 2012-2016</td>
<td>2.27</td>
<td>0.66</td>
<td>1.51</td>
<td>3.10</td>
</tr>
<tr>
<td>Total zipcodes</td>
<td>3870</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total sales (mil)</td>
<td>36.05</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes. Monthly sales and mean price data is from CoreLogic Deed data and mean time-on-market data is from Realtor.com.

following sections empirically test these predictions.

3 Data and measurement

3.1 Data sources

Our empirical analysis uses a number of data sources. We use microdata on house sales and characteristics from the CoreLogic tax and deed database, spanning the time period 2000-2017. We use Realtor.com data on median time-on-market at the zipcode-month level from 2012-2017. We also use the Zillow Home Value Index (ZHVI) as a zipcode-month level price index, and we use the ACS 2012-2016 5-year sample for data on various demographic characteristics of zipcodes. Further details of the steps we take to clean and merge the datasets are described in appendix B.

Since we estimate price dispersion using a repeat-sales specification, we only use zipcodes with a fairly large number of sales; details of how we select zipcodes are described in appendix B.1, and descriptive statistics for zipcodes in our primary estimation sample dataset are shown in Table 1. The dataset we use to estimate price dispersion consists of 36 million house sales within 3,870 zipcodes, over the period 2000-2017. As appendix table 4, our sample covers around 42.7% of the US population despite only containing 11.7% of all zipcodes. Our sample is concentrated in relatively large, dense, and high-income zipcodes, but is relatively representative in terms of other demographic characteristics.
Figure 1: Comparative statics

Notes. Average prices, volume, time-on-market, and price variance, as $M_B$ is increased, holding fixed all other model parameters.
3.2 Measuring idiosyncratic price dispersion

Let $i$ index properties, as determined by CoreLogic APNs. Let $x_i$ represent a vector of time-invariant characteristics of house $i$. Let $t$ denote months, $z$ denote zipcodes, and let $p_{izt}$ denote the log price of house $i$ in zipcode $z$ traded in period $t$. We assume the following specification:

$$ p_{izt} = \gamma_i + \eta_{zt} + f_z(x_i, t) + \epsilon_{izt} $$

In words, we assume that prices are determined by a time-invariant house fixed effect, $\gamma_i$, a zip-month fixed effect, $\eta_{zt}$, a smooth function $f_z(x_i, t)$ of characteristics and time, and a mean-0 error term $\epsilon_{izt}$. This specification is a combination of repeat-sales and hedonic models of house prices. The $\gamma_i$ term absorbs all variation in the level of sale prices for a given house, capturing both observed and unobserved features of houses which have a time-invariant effect on the log price of a house relative to other houses in its zipcode. The $\eta_{zt}$ term absorbs uniform changes of log prices of houses within a zipcode over time. Thus, ignoring the $f_z(x_i, t)$ term, specification (17) is a repeat-sales specification; it implies that, ignoring the error term $\epsilon_{izt}$, house prices within a zipcode should follow parallel trends. If a given house sells for a price 30% above another house in a given month, it should sell for 30% above the price at any future month during which both houses are sold. Any violation of parallel price trends is attributed to idiosyncratic price volatility, captured by the error term $\epsilon_{izt}$.

The benefit of repeat-sales specifications is that both observed and unobserved time-invariant features of house prices are absorbed into house fixed effects, $\gamma_i$. A weakness of a pure repeat-sales specification is that the assumption that house prices follow parallel trends assumption may be unrealistic. Suppose that new amenities are built in certain parts of a zipcode, or that migrants to a given zipcode prefer one-bedroom houses over two-bedroom houses. Houses near the new amenities, or one-bedroom houses, would appreciate more than other houses in the zipcode, thus violating the assumption of parallel trends embedded in pure repeat-sales specifications. However, these violations are not due to idiosyncratic volatility, particular to the sale of an individual house; they are due to systematic changes in the relative prices of houses with certain characteristics, at certain locations, over time.

To the extent that changes in relative values of houses over time are related to characteristics of houses which are observed by the econometrician, as part of the vector $x_i$, we can accommodate these shifts in relative prices, by allowing the conditional mean of...
$p_{izt}$ to vary as a function of observable characteristics over time. This is what the $f_z(x_i, t)$ term does: it allows the price paths of houses with different characteristics to differ, but restricts these differences to be smooth functions of characteristics $x_i$ and time $t$. Once we have accounted for price changes for a given house over time which are attributable to observable characteristics, we attribute any residual deviations of house prices from their conditional mean to idiosyncratic variance. Thus, specification (17) combines repeat-sales and hedonic approaches by absorbing all time-invariant price deviations of a given house using a house fixed effect, but then allowing observable house characteristics to have time-varying effects on a house’s price path.

There are a number of problems with using specification (17) to measure price dispersion. First, there may be unobserved heterogeneity in characteristics that have time-varying effects on characteristics, which we are not able to capture using the $f_z(x_i, t)$ term. Second, the $f_z(x_i, t)$ term only captures effects of house characteristics that are fixed for a given house over time; we are not able to deal with changes in house characteristics $x_i$ over time, which we do not observe. Both of these issues will tend to bias our estimates of $\sigma_{izt}^2$ upwards, as they are sources of predictable variation in house prices that are observable to agents in the model, but that we attribute to idiosyncratic variation.

### 3.3 Implementation

Practically, in order to implement specification (17), we need to choose a functional form for the hedonic term, $f_z(x_i, t)$. Since houses have a large number of characteristics, we cannot generate interaction effects for all characteristic; the number of parameters would increase too rapidly due to the curse of dimensionality. Instead, we use an additive functional form for $f_z(x_i, t)$:

$$f_z(x_i, t) = g_{z \text{laglong}}(t, \text{lat}_i, \text{long}_i) + g_{z \text{sqft}}(t, \text{sqft}_i) + g_{z \text{yrbuilt}}(t, \text{yrbuilt}_i) +$$
$$g_{z \text{bedrooms}}(t, \text{bedrooms}_i) + g_{z \text{bathrooms}}(t, \text{bathrooms}_i) \quad (18)$$

The functions $g_{z \text{laglong}}$, $g_{z \text{sqft}}$ and $g_{z \text{yrbuilt}}$ are fully interacted third-order polynomials in their constituent components. The functions $g_{z \text{bedrooms}}$ and $g_{z \text{bathrooms}}$ interact dummies for a given house having 1, 2, or more than 3 bedrooms or bathrooms, respectively, with third-order polynomials in time. In all cases, we estimate polynomial coefficients jointly with the fixed effects in specification (17). We estimate all polynomial coefficients
separately for each zipcode in our sample.

Intuitively, additivity in specification (18) imposes the following restrictions. Older or larger houses can appreciate faster or slower than younger or smaller houses, in a manner which varies flexibly with house size or age and time, captured by the functions $g^\text{sqft}_z$ and $g^\text{yrbuilt}_z$. However, we rule out most interaction effects: houses which are both large and old are constrained to appreciate at a rate determined by the sum of the functions $g^\text{sqft}_z$ and $g^\text{yrbuilt}_z$. The exception to this is that the $g^{\text{latlong}}_z$ function interacts latitude and longitude; this is important because it is implausible that latitude and longitude have additive effects on prices, and effectively this allows house prices to vary smoothly with respect to a house’s geographic location over time.

Once we have estimated (17), we estimate squared residuals $\hat{\epsilon}^2_{izt}$ for each house sale as:

$$\hat{\epsilon}^2_{izt} = \frac{N_z}{N_z - K_z} (p_{izt} - \hat{p}_{izt})^2$$

(19)

where $N_z$ is the number of house sales in zipcode $z$, and $K_z$ is the number of parameters estimated from specification (17). The term $\frac{N_z}{N_z - K_z}$ is a degrees-of-freedom correction; this is important to include because, since most houses are sold relatively few times, the number of parameters, $K_z$, is nontrivially large relative to the number of data points, $N_z$.

Equation (19) thus gives us a measure of idiosyncratic price dispersion, $\hat{\epsilon}^2_{izt}$, at the level of each individual house sale. We will aggregate $\hat{\epsilon}^2_{izt}$ to various levels in following sections, and analyze the factors that affect its magnitude.

For example, for the cross-sectional analysis of Section 4.3, we wish to analyze the behavior of $\sigma^2_{izt}$ across zipcodes, for the period 2012-2016. We construct a measure of price dispersion at the zipcode level, $\hat{\sigma}_z$, by averaging $\hat{\epsilon}^2_{izt}$ for all sales within a zipcode, then taking the square root; that is,

$$\hat{\sigma}_z = \sqrt{\frac{\sum_t \hat{\epsilon}^2_{izt}}{N_{z,2012-2016}}}$$

where the sum is over all house sales that happened in the time period 2012-2016, and $N_{z,2012-2016}$ is the number of observations in zipcode $z$ over this period. The constructed

\[5\] Assuming homoskedasticity within zipcodes, $\sigma^2_{izt} = \sigma_z^2$, the degrees-of-freedom correction in expression (19) causes the expectation of $\hat{\epsilon}^2_{izt}$ to be equal to the true variance, $\sigma^2_z$. 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.
\( \hat{\sigma}_z \) can be interpreted as an estimate of the log standard deviation of idiosyncratic house prices, after we have removed time-invariant quality and the time-varying effect of observable characteristics. We can similarly construct measures of price dispersion at other levels of aggregation: in section 4, we will analyze the behavior of price dispersion over calendar years, calendar months, and zipcode-years. We will often refer to these price dispersion estimates as “logSD”, meaning the log standard deviation of idiosyncratic price variance.

Figure 2 shows the distribution of \( \hat{\sigma}_z \) across zipcodes in our sample. \( \hat{\sigma}_z \) has mean 16.8%, and standard deviation 4.6%; it is somewhat right-skewed, so the 10th percentile is 11.2% and the 90th percentile is 22.6%. In order to calculate how much this affects returns, note that an agent who buys and sells a house incurs the \( \epsilon_{izt} \) error twice, once upon purchase and once upon sale; thus, multiplying these estimates by a factor \( \sqrt{2} \), an agent in a 10th percentile zipcode who buys and sells a house incurs 15.9% additional error beyond the riskiness of the zip-month mean price, and an agent in a 90th percentile zipcode incurs 31.9% additional error.

Appendix C.1 runs specification (17) without the hedonic \( f_z(x_i, t) \) term, using a pure repeat-sales specification. The resultant residuals are approximately 3.8% (in logSD units, 0.61%) larger than our baseline specification, but the two specifications produce highly correlated estimates. Thus, in practice, the effect of the hedonic \( f_z(x_i, t) \) appears to be quantitatively modest.

### 3.4 Literature comparison

A number of other papers have attempted to measure idiosyncratic house price dispersion; we briefly discuss these, and how they relate to our methodology in this paper.

Giacoletti (2017) 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 (17), Giacoletti uses returns, rather than individual house sales, as the primary unit of analysis. Similarly, Sagi (2015) also shows that return variances

\footnote{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 zipcode-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.}
Notes: Distribution of $\hat{\sigma}_{zt}$.

are very flat with respect to holding-period length, using data on commercial real estate sales. Relative to these papers, the benefit of our specification is that it allows us to trends in idiosyncratic price dispersion over time. In section 4, we will show that our measure of idiosyncratic price dispersion moves over time, in aggregate and in panel specifications, in the directions predicted by our model. Moreover, these papers essentially use repeat-sales specifications; our partially hedonic specification relaxes the assumption of parallel trends, allowing observable house characteristics to affect the price path of a given house.

Another approach to measuring price dispersion is to use a purely hedonic model. Peng and Thibodeau (2017) uses a hedonic specification to measure price dispersion, analyzing the relationship between idiosyncratic price dispersion and various other variables in the cross-section of zipcodes. Our specification has a number of advantages over purely hedonic specifications. First, in our specification, all characteristics of houses which may affect time-invariant house quality, observed and unobserved, are absorbed into the house fixed effect; hence, our specification is more robust to unobservables which have time-invariant effects on prices. Second, 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_2(x_t, t) term in specification (17), which effectively allows the hedonic coefficients on different characteristics to change continuously over time. The focus of our paper is on the relationship of liquidity to price dispersion, so we do not analyze other factors that affect idiosyncratic price dispersion in the main text of the paper; however, in appendix C.2 we analyze some cross-sectional correlations between other variables and price dispersion.

Two other papers with other measurement strategies for idiosyncratic price dispersion are Anenberg and Bayer (2013) and Landvoigt, Piazzesi and Schneider (2015). Anenberg and Bayer (2013), 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. This is not strictly more restrictive than our specification, since we do not include previous sale prices in specification (17) however, to the extent that the factors which affect prices are summarized by our house characteristics x_t, location, size, year built, and number of bedrooms and bathrooms, our specification will also be able to capture these trends.

4 Results

4.1 Idiosyncratic price dispersion over the business cycle

In this subsection, we show that idiosyncratic price dispersion varies over the business cycle, in the directions predicted by our model. Figure 3 shows the behavior of indexed total sales, logSD, prices, and time-on-market at the yearly level. We filter to zipcodes which we observe every year from 2000-2016; this leaves us with 3,078 zipcodes, comprising approximately 29 million home sales. To construct the LogSD line in figure 4, we average \hat{\epsilon}_{izt}^2 over zip-months within a given 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 zip-months in a given year; we show data for 2013-2016 because these are the only years in which we observe time-on-market for all calendar months of the year. The price line represents the sales-weighted average of the Zillow Home Value Index for single-family residences across zip-months in a given year. We show

Figure 3 shows that idiosyncratic house price dispersion is countercyclical. Price
dispersion is decreasing from 2000-2004, as the housing market is booming and volume and prices are increasing. Idiosyncratic price dispersion reverses direction and starts increasing, peaking in 2010, when sales and prices are both low due to the housing bust. Price dispersion then starts increasing once again as sales volume and average prices recover; while our data on time-on-market only spans four years, time-on-market is also decreasing over much of this time period.

Quantitatively, total housing sales within zipcodes in our sample increased 45% from 2000 to 2005, dropped to a trough of 70% of its 2000 value in 2011, and recovered to 5.9% above its 2000 value in 2016. Prices increase 66% from 2000 to 2007, drop to 22% above their 2000 value in 2011, and increase to 64% above their 2000 value in 2016. While the peaks and troughs of the logSD series do not align with the total sales or price series, logSD drops to 3.8% (in units of $\hat{\sigma}$, 0.6%) below its 2000 value near the peak of the boom in 2004, increases to 9.0% (1.4%) above its 2000 value in 2009, and decreases to 1.8% (0.3%) above its 2000 value in 2016.

Our theory is thus consistent with movements in sales, prices, time-on-market, and idiosyncratic price dispersion over the business cycle. We believe that we the first paper to empirically document the countercyclicality of idiosyncratic house price dispersion across the US. The change in price dispersion over the business cycle is nontrivially large – the logSD line moves 2%, in units of $\hat{\sigma}$, from its trough in 2004 to its peak in 2009, which is approximately 43% of the cross-sectional standard deviation in logSD across zipcodes.

### 4.2 Seasonality

Idiosyncratic price dispersion also varies seasonally, in the directions predicted by our model. In figure 4, we show indexed prices, sales, logSD (that is, $\hat{\sigma}_t$), and time-on-market at the monthly level over 2013-2016, for all zipcodes in which we observe positive sales in each of the 48 months within this time period. All variables are constructed analogously to figure 3, by taking sales-weighted averages of all zip-months associated with a given calendar month.

Figure 4 shows that prices, total sales, time-on-market, and LogSD are all highly seasonal. Within our sample, there are on average 82% more house sales in June than in

---

7One exception is Landvoigt, Piazzesi and Schneider (2015), who use a different methodology to demonstrate that idiosyncratic price dispersion increased in San Diego from 2005-2007, but does not note the decrease in price dispersion from 2009 onwards.
Figure 3: Time-series variation in prices, sales, logSD, and time-on-market

Notes. Indexed total sales, prices, and logSD, 2000-2016, and time-on-market (TOM) 2013-2016. To calculate the logSD and time-on-market series, $\hat{\sigma}^2_{zt}$ and time-on-market respectively are averaged over zip-months, weighted by sales; the logSD line is the square root of the weighted average of $\hat{\sigma}^2_{zt}$ by year. All variables are indexed, i.e. divided by their level in the first year we observe them.
Figure 4: Monthly variation in sales, prices, logSD, and time-on-market

Notes. Indexed total sales, average prices, log standard deviation, and time-on-market (TOM) by calendar month, over the time period 2013-2016. To calculate the logSD and time-on-market series, $\sigma^2_{zt}$ and time-on-market respectively are averaged over zip-months, weighted by sales; the logSD line is the square root of the weighted average of $\sigma^2_{zt}$ by month. All variables are indexed, i.e. divided by their January level.

January; prices are around 5.8% higher in June, time-on-market is 35% lower, and our logSD measure of price dispersion is around 7.2% (in units of $\sigma_t$, 1.2%) lower.

It is known in the literature that sales, price, and time-on-market are seasonal – prices and sales systematically increase and time-on-market systematically decrease in summer (Ngai and Tenreyro, 2014). However, we believe we are the first to show that idiosyncratic house price dispersion is also seasonal, systematically decreasing in summer and increasing in winter, as predicted by our model. An agent who sells her house in winter thus not only attains lower sale prices on average but also more variable prices; quantitatively, both effects are nontrivially large.
4.3 Evidence from the cross-section of zipcodes

Next, we test whether the predictions of our model hold cross-sectionally across zipcodes. We do not test the relationship between sales, average prices and price dispersion in the cross-section of zipcodes; variation in total sales and average prices across zipcodes is likely driven by population density, average incomes, and other factors outside of our model, in addition to differences in market tightness. However, claim 3 of our model shows that time-on-market has a tight relationship with price dispersion: if the distributions of $c$ and $\epsilon$, the discount rate $r$, and the bargaining parameter $\theta$ are the same in two zipcodes, and time-on-market is uniformly higher for all seller types in one zipcode, then idiosyncratic price dispersion will also be higher. Thus, we will test whether time-on-market is correlated with price dispersion in the cross-section of zipcodes.

Using our dataset spanning 3,870 zipcodes over 2012-2016, we aggregate time-on-market and price dispersion to the zipcode level. We test the following specification:

$$\hat{\sigma}_z = \beta_1 T_{OMz} + \gamma X_z + \xi_z$$  \hspace{1cm} (20)

$X_z$ represents a vector of zipcode-level controls, and $\xi_z$ is a mean-0 noise term. The model predicts that $\beta_1$ should be positive.

Table 2 shows the results from these regressions. In column 1, we regress $\hat{\sigma}_z$ on a variety of zipcode-level controls without including time-on-market: zipcode averages of year built of sold houses, log price, months from previous sale, log income, and the fractions of the zipcode’s population which are aged 18-35, 35-64, black, high school and college graduates respectively, married, unemployed, and homeowners. We include third-order polynomials in each of these control variables in all specifications. The $R^2$ in column 1 is 0.503, so approximately half of the cross-sectional variation across zipcodes in idiosyncratic price dispersion can be explained by our controls. We further analyze how various demographic characteristics of zipcodes are correlated with price dispersion in appendix C.2.

In column 2, we add time-on-market in months, $T_{OMz}$, to the regression. The coefficient is positive and significant: as predicted by theory, zipcodes with higher time-on-market have higher price dispersion. A one standard deviation increase in time-on-market (0.66 months) is associated with a 1.5% increase in idiosyncratic price dispersion. As discussed in subsection 3.3, idiosyncratic price dispersion has a mean of 16.7% and a standard deviation of 4.6% across zipcodes in our sample, so the size of the coefficient
on time-on-market is nontrivially large. Moreover, the $R^2$ in column 2 is 0.579, indicating that time-on-market has substantial explanatory power for idiosyncratic price dispersion which is not captured in our control variables. Columns 4 and 5 add state and CBSA fixed effects, the time-on-market coefficient remains significant, with similar magnitudes, in both specifications.

We do not include prices or sales in the cross-sectional regressions. The core assumption driving the comparative statics of subsection 2.4 is that variation in all outcome parameters is primarily driven by variation in buying pressure, represented by $M_B$ in our model; but differences in buying pressure are clearly not the main driver of sales and price variation cross-sectionally across zipcodes. Average prices and sales vary across zipcodes for various reasons; some zipcodes are larger or denser than others, and demographic characteristics such as income can vary significantly across zipcodes. We focus on time-on-market because claim 3 of our model shows that the relationship between time-on-market and price dispersion is quite robust: as long as $F(c), G(\epsilon), \tau, \theta$ are held constant, any changes in model parameters which decrease time-on-market will also decrease price dispersion.

Table 2: Price dispersion in the cross-section of zipcodes

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LogSD x 100</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time on market (months)</td>
<td>2.342</td>
<td>2.595</td>
<td>3.003</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.090)</td>
<td>(0.083)</td>
<td>(0.107)</td>
<td></td>
</tr>
<tr>
<td>Controls</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Fixed effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>State</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CBSA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$N$</td>
<td>3,870</td>
<td>3,870</td>
<td>3,870</td>
<td>3,870</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.503</td>
<td>0.579</td>
<td>0.747</td>
<td>0.832</td>
</tr>
</tbody>
</table>

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

---

8Since zip codes do not necessarily map one-to-one to CBSAs, or even to states, we first map zipcodes to the county which contains the largest fraction of the zipcode’s population, using the Census’s 2010 ZCTA-to-county relationship file, then map the county to its unique CBSA and state.
4.4 Panel regressions

Table 3 shows zipcode-year panel regressions of logSD on various variables. We run panel regressions at the year level instead of the month level because sales, time-on-market, prices, and idiosyncratic price dispersion are all seasonal at the monthly level. Similarly to subsections 4.1 and 4.2 above, we use Zillow’s single-family residence ZHVIs to measure prices at the zipcode-month level.

The panel regressions in table 3 support most of the predictions of our model. Columns 1 and 2 show that logSD is negatively correlated with prices and log sales, and column 3 shows that price dispersion is negatively correlated with time-on-market, in the 4 years in which we observe time-on-market. Column 4 includes all variables together, using data from 2013-2016; the coefficient on log sales loses significance, but the coefficients on log price and time-on-market remain significant, with unchanged signs.

The interpretation of these panel regressions is that, when a given zipcode-year experiences, relative to other zipcodes in the same year, a change in log sales or prices which is unusually high, or a change in time-on-market which is unusually low, price dispersion tends to also be unusually low in the zipcode-year. The panel regressions thus present substantively distinct evidence from the aggregate time series shown in subsection 4.1 above, further validating the predictions of our model.

5 Discussion

In this paper, we have built a model predicting that idiosyncratic price dispersion should be correlated positively with time-on-market and negatively with prices and sales volume. We have shown that these correlations hold robustly, cross-sectionally, in panel specifications, and in the aggregate seasonally and over the business cycle. These results suggest that liquidity is an important driver of idiosyncratic house price dispersion.

5.1 Rationality, arbitrage, and theories of housing booms and busts

Many recent theories of housing market booms and busts emphasize agents’ belief heterogeneity or imperfect rationality. Under these theories, extrapolative or boundedly
Table 3: Zipcode-year panel regressions

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log ZHVI</td>
<td>−1.355***</td>
<td>−1.540***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.092)</td>
<td>(0.327)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log sales</td>
<td>−0.726***</td>
<td>−0.090</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.104)</td>
<td>(0.154)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time on market (months)</td>
<td></td>
<td></td>
<td>0.285***</td>
<td>0.279***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.065)</td>
<td>(0.067)</td>
</tr>
<tr>
<td>Zip fixed effects</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Year fixed effects</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>N</td>
<td>61,135</td>
<td>61,135</td>
<td>14,882</td>
<td>14,882</td>
</tr>
<tr>
<td>R²</td>
<td>0.860</td>
<td>0.860</td>
<td>0.943</td>
<td>0.943</td>
</tr>
</tbody>
</table>

Notes. Each data point is a zipcode-month. Regressions are weighted by the number of sales within a given zip-month.

Rational buyers contribute to increasing volume and destabilizing prices. However, our results show that housing booms – both local and aggregate – systematically correlate with lower idiosyncratic price volatility. This does not rule out the possibility that housing booms are partially driven by irrational expectations; however, even if irrational exuberance destabilizes average prices, the forces of arbitrage seem to work well enough in housing booms that relative house prices are actually more stable during boom periods than bust periods.

5.2 Idiosyncratic risk and household welfare

A large fraction of household wealth in the US consists of real estate, so house price volatility is an important contributor to household portfolio risk. Not all house price movements represent pure risk -- a common argument in the literature is that movements in average house prices partially hedge households against changes in local costs of liv-
However, the idiosyncratic component of house price volatility is an undiversified source of uncertainty which does not hedge any other risks faced by the household.

Our results showing that idiosyncratic volatility varies over calendar months, and over the business cycle, have implications for household welfare. Agents who buy or sell houses in the housing bust are exposed to additional idiosyncratic risk. Similarly, agents who buy and sell houses in winter are exposed to more idiosyncratic risk than those who trade in summer. In the business cycle case, the change in price dispersion is small relative to the change in average prices; however, in the seasonal case, the average and idiosyncratic components are comparable in magnitude.

In our model, search frictions are particularly harmful to sellers with high holding costs; in practice, this corresponds to sellers who are credit-constrained or otherwise financially vulnerable. This implies that the costs of housing market frictions may disproportionately fall on the poor, exacerbating existing inequality.

\textsuperscript{10}See, for example, Han (2010) and Han (2013).
References


Appendix

A Supplementary material for section 2

A.1 Proof of claim 1

From (2), prices are:

\[ P(\epsilon, c) = \theta(\epsilon + V_M - V_B - V_S(c)) + V_S(c) \]  (21)

We wish to take the variance of expression (21) with respect to the joint distribution of costs \( c \) and match qualities \( \epsilon \) within the set of pairs of buyers and sellers that match and trade in any given moment; call this joint distribution \( F_{tr}(c, \epsilon) \).

First, let \( F_{tr}(c) \) be the marginal distribution of seller costs \( c \), given that trade occurs. By flow equality in expression (13) of proposition 1, the marginal distribution of costs among sellers who trade and exit the market at any moment must be the same as the distribution of costs among sellers that enter the platform; thus,

\[ F_{tr}(c) = F(c) \]  (22)

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

\[ F_{tr}(\epsilon | c) \]

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

\[ F_{tr}(\epsilon | c) = G(\epsilon | \epsilon > \epsilon^*(c)) \]  (23)

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

Having characterized \( F_{tr}(c, \epsilon) \), we can now take the variance of expression (21).
Applying the law of iterated expectations, price variance can be written as:

\[
\text{Var}(P(\epsilon,c)) = \mathbb{E}_{c \sim F_{tr}(c)} \left[ \text{Var}_{\epsilon \sim F_{tr}(\epsilon \mid c)}(P(\epsilon,c) \mid c) \right] + \text{Var}_{c \sim F_{tr}(c)} \left( \mathbb{E}_{\epsilon \sim F_{tr}(\epsilon \mid c)}[P(\epsilon,c) \mid c] \right)
\] (24)

First, we characterize the left term on the RHS of (24). Conditional on \(c\), the only random term in \(\text{Var}_{\epsilon \sim F_{tr}(\epsilon \mid c)}(P(\epsilon,c) \mid c)\) is \(\epsilon\); thus, substituting expression (21) into the variance and ignoring constant terms, we have:

\[
\text{Var}_{\epsilon \sim F_{tr}(\epsilon \mid c)}(P(\epsilon,c) \mid c) = \theta^2 \text{Var}_{\epsilon \sim G(\cdot)}(\epsilon \mid \epsilon > \epsilon^*(c))
\]

From the characterization of \(F_{tr}(\epsilon \mid c)\) in expression (23), we have:

\[
\text{Var}_{\epsilon \sim F_{tr}(\epsilon \mid c)}(\theta \epsilon \mid c) = \theta^2 \text{Var}_{\epsilon \sim G(\cdot)}(\epsilon \mid \epsilon > \epsilon^*(c))
\]

In words, \(\text{Var}_{\epsilon \sim G(\cdot)}(\epsilon \mid \epsilon > \epsilon^*(c))\) is the variance of an exponential random variable \(\epsilon\), conditional on it being above some cutoff \(\epsilon^*(c)\). This conditional distribution has variance equal to the unconditional variance of \(\epsilon\), \(\sigma^2_{\epsilon}\), for any cutoff \(\epsilon^*(c)\). Thus,

\[
\text{Var}_{\epsilon \sim F_{tr}(\epsilon \mid c)}(\epsilon \mid \epsilon > \epsilon^*(c)) = \sigma^2_{\epsilon}
\]

for any \(c\); thus, we have:

\[
\text{Var}_{\epsilon \sim F_{tr}(\epsilon \mid c)}(P(\epsilon,c) \mid c) = \theta^2 \sigma^2_{\epsilon}
\]

\[
\mathbb{E}_{c \sim F_{tr}(c)} \left[ \text{Var}_{\epsilon \sim F_{tr}(\epsilon \mid c)}(P(\epsilon,c) \mid c) \right] = \theta^2 \sigma^2_{\epsilon}
\]

For the rightmost term in expression (24), we have:

\[
\text{Var}_{c \sim F_{tr}(c)} \left( \mathbb{E}_{\epsilon \sim F_{tr}(\epsilon \mid c)}[P(\epsilon,c) \mid c] \right) = \text{Var}_{c \sim F_{tr}(c)} \left( \theta \left( \mathbb{E}_{\epsilon \sim F_{tr}(\epsilon \mid c)}[\epsilon \mid \epsilon > \epsilon^*(c)] + V_M - V_S(c) - V_B \right) + V_S(c) \right)
\]

Using that \(\epsilon^*(c) = V_S(c) + V_B - V_M\), and using the characterization of \(F_{tr}(\epsilon \mid c)\) from
expression (23), we can write this as:

$$\text{Var}_{c \sim \mathcal{F} \mathcal{T}_c(c)} (\theta \mathbb{E}_{\epsilon \sim \mathcal{G} \mathcal{T}} [\epsilon - \epsilon^*(c) | \epsilon > \epsilon^*(c)] + \mathcal{V}_S(c)) \tag{26}$$

Since we have assumed $G(\cdot)$ is exponential, the term:

$$\theta \mathbb{E}_{\epsilon \sim \mathcal{G} \mathcal{T}} [\epsilon - \epsilon^*(c) | \epsilon > \epsilon^*(c)]$$

is equal to the unconditional mean of $\epsilon$, $\sigma_\epsilon$. It is thus constant with respect to $\epsilon^*(c)$ and thus $c$, and can be ignored when calculating the variance in (26). Hence, we have:

$$\text{Var}_{c \sim \mathcal{F} \mathcal{T}_c(c)} (\mathbb{E} [P(\epsilon, c) | c]) = \text{Var}_{c \sim \mathcal{F} \mathcal{T}_c(c)} (\mathcal{V}_S(c)) \tag{27}$$

Where, we have used expression (22), stating that $\mathcal{F} \mathcal{T}_c(c) = \mathcal{F}(c)$. Combining (25) and (26), we have:

$$\text{Var} (P(\epsilon, c)) = \text{Var}_{c \sim \mathcal{F} \mathcal{T}_c(c)} (\mathcal{V}_S(c)) + \theta^2 \sigma_\epsilon^2 \tag{28}$$

as desired.

A.2 Proof of claim

From expression (8) in proposition, the seller value function $\mathcal{V}_S(c)$ is:

$$r \mathcal{V}_S(c) = -c + \lambda_\mathcal{S} \int_{\epsilon > \epsilon^*(c)} \theta \ (\epsilon - \mathcal{V}_B - \mathcal{V}_S(c)) \ d\mathcal{G}(\epsilon)$$

Differentiating with respect to $c$, we have:

$$r \mathcal{V}_S'(c) = -1 + \lambda_\mathcal{S} \int \theta \ (-\mathcal{V}_S'(c)) \ 1 (\epsilon > \epsilon^*(c)) \ d\mathcal{G}(\epsilon)$$

Computing the integral, this becomes:

$$r \mathcal{V}_S'(c) = -1 + \lambda_\mathcal{S} \theta \ (-\mathcal{V}_S'(c)) \ (1 - \mathcal{G}(\epsilon^*(c)))$$

Solving for $\mathcal{V}_S'(c)$, we have:

$$\mathcal{V}_S'(c) = \frac{-1}{r + \theta \lambda_\mathcal{S} \ (1 - \mathcal{G}(\epsilon^*(c)))} \tag{29}$$
Substituting expression (15) for \(TOM(c)\) in the denominator of (29), we have:

\[
V'_S(c) = \frac{-1}{\frac{\theta}{TOM(c)}}
\]

Rearranging, we have (16).

A.3 Proof of claim 3

From expression (16), if \(TOM_{\Theta_1}(c) > TOM_{\Theta_2}(c)\) for all \(c\), and if \(F(c), G(\epsilon), \theta, r\) are the same in the two sets of primitives, then \(V'_S(c)\) must also be strictly larger in absolute value, pointwise in \(c\), under parameters \(\Theta_1\) than \(\Theta_2\), for all \(c\). From expression (14) in claim 1, we have:

\[
\text{Var}(P(\epsilon, c)) = \text{Var}_{c \sim F(\cdot)}(V_S(c)) + \theta^2 \sigma_\epsilon^2
\]

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

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

\[
\text{Var}(X) = \min_{\bar{x}} \int (x - \bar{x})^2 \text{d}G(x)
\]  \hspace{1cm} (30)

To prove expression (30), note that:

\[
\int (x - \bar{x})^2 \text{d}G(x) = \int (x - E(x) + E(x) - \bar{x})^2 \text{d}G(x)
\]

\[
= \int (x - E(x))^2 + 2(x - E(x))(E(x) - \bar{x}) + (E(x) - \bar{x})^2 \text{d}G(x)
\]

\[
= \int (x - E(x))^2 + (E(x) - \bar{x})^2 \text{d}G(x)
\]

Thus,

\[
\min_{\bar{x}} \int (x - \bar{x})^2 \text{d}G(x) = \min_{\bar{x}} \int (x - E(x))^2 \text{d}G(x) + \int (E(x) - \bar{x})^2 \text{d}G(x)
\]
\[
= \int (x - \operatorname{E}(x))^2 \, dG(x) = \operatorname{Var}(X)
\]

Now, call the distribution of \(V_S(c)\) among trading sellers \(F_{V_S}(\cdot)\). Using expression (30), we can write the variance of \(V_S(c)\) as:

\[
\operatorname{Var}(V_S(c)) = \min_x \int (x - \bar{x})^2 \, dF_{V_S}(x) \tag{31}
\]

Since the distribution of \(c\) among trading sellers is \(F(c)\), and \(V_S(c)\) is a function of \(c\), by changing variables to integrate over \(c\), the distribution \(F_{V_S}(\cdot)\) as a function of \(c\) among trading sellers can be written as:

\[
dF_{V_S}(c) = V'_S(c) \, dF(c)
\]

Hence, (31) becomes:

\[
= \min_x \int (x - \bar{x})^2 V'_S(x) \, dF(x) \tag{32}
\]

A uniform increase in \(V'_S(c)\) causes the integral in (32) to strictly increase for any \(\bar{x}\). Thus, if \(V'_S(c)\) is uniformly higher under \(\Theta_1\) than \(\Theta_2\), then \(\operatorname{Var}(V_S(c))\) must also increase, and thus \(\operatorname{Var}(P(\epsilon,c))\) must also increase.

### A.4 Derivation of model quantities

In this appendix, we derive expressions for sales volume, average prices, time-on-market, and price dispersion, which are plotted against \(M_B\) in figure 1. Claim 1 characterized price dispersion, that is, the variance of prices \(P(\epsilon,c)\) among trading buyers and sellers:

\[
\operatorname{Var}_{c \sim F(\cdot)}(V_S(c)) + \theta^2 \sigma^2_\epsilon
\]

Equilibrium sales volume per unit time is the mass of sellers that get matched to a buyer, sell their house and leave the market per unit time. This is:

\[
M_S \lambda_S \int (1 - G(\epsilon^*(c))) \, dF_{eq}(c)
\]

Intuitively, this is the equilibrium mass of sellers \(M_S\), multiplied by the equilibrium seller matching rate \(\lambda_S\), multiplied by the probability that meetings result in trade; this is the integral of the probability that a seller of type \(c\) trades, which is \((1 - G(\epsilon^*(c)))\), over the
equilibrium distribution of seller holding costs, $F_{eq}(c)$.

Average time-on-market, over realized sales, is:

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

This is simply the average of time on market for a seller of type $c$ over the distribution of costs with respect to $F(c)$; note that we showed in appendix A.1 above that the distribution of values among trading sellers is simply $F_{tr}(c) = F(c)$.

The average transaction price conditional on trade is:

$$\int \int_{\epsilon > \epsilon^*(c)} \frac{P(\epsilon, c)}{1 - G(\epsilon^*(c))} \ dG(\epsilon) \ dF_{eq}(c)$$

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

B Data appendix

B.1 CoreLogic tax and deed data

We use transaction data from the CoreLogic deed data files. For each sale, CoreLogic records the price and date of the sale; housing units in the data are identified, within a FIPS county code, by an Assessor Parcel Number (APN), which is assigned to each plot of land by tax assessors. We also use CoreLogic’s tax assessment data for the fiscal year 2016-2017, which contains data on house-level characteristics: latitude, longitude, year built, square footage, and numbers of bedrooms and bathrooms. We merge the tax data to the CoreLogic deed data by APN and FIPS county code. Using the merged data, we calculate the average and standard deviation of all characteristics at the zipcode level.

We clean the deed data in a number of steps. First, we use only arms-length new construction or resales of single-family residences, which are not foreclosures, which have non-missing sale price, date, APN, and zipcode 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 data coverage is inconsistent for many states prior to 2000.
CoreLogic data coverage for counties is inconsistent over time; some counties have fairly low sales for earlier time periods, suggesting that data quality in early years is inconsistent. Hence, we filter out all county-years for which the total number of sales we observe is below 20% of average annual sales over all county-years in our sample. At this stage, we filter at the county level because data is collected from county records, hence we believe that data quality is most likely to vary at the county rather than zipcode level.

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

First, our measurement of price dispersion uses a repeat-sales specification, so we can only use houses that were sold multiple times. Moreover, we wish to filter for “house flips”, as well as instances where reported sale price seems anomalous. Thus, similarly to Landvoigt, Piazzesi and Schneider (2015) and Giacoletti (2017), if we ever observe a property whose annualized appreciation or depreciation is above 50% for any given pair of sales, we drop all observations of the property. Similarly, if a house is ever sold twice within a year, we drop all observations of the house; these are likely flips, which are known to be a peculiar segment of the real estate market (Bayer et al. (2011), Giacoletti and Westrupp (2017)).

Second, specification (17) involves a fairly large number of parameters: house and zipcode-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 (17), and thus we filter to zipcodes with at least 1000 house sales in total, and with at least 10 sales per month on average, after applying the filtering steps described above.

Appendix table 4 shows characteristics of the zipcodes in our estimation sample, compared to the universe of zipcodes in the 2012-2016 ACS 5-year dataset. Our dataset constitutes approximately 11.6% of all zipcodes. Zipcodes in our sample are larger and denser than average, so our sample constitutes around 48% of the total US population. In terms of demographics, the average income is somewhat higher than average for zipcodes in our sample, but our zipcodes are quite representative in terms of age, race, and the fraction of the population that is married.
Table 4: Characteristics of zipcodes in the primary dataset

<table>
<thead>
<tr>
<th></th>
<th>Sample mean</th>
<th>All zipcodes mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population</td>
<td>35551</td>
<td>9724</td>
</tr>
<tr>
<td>Pop / Sq mile</td>
<td>2615</td>
<td>1307</td>
</tr>
<tr>
<td>Housing units</td>
<td>14165</td>
<td>4095</td>
</tr>
<tr>
<td>Avg income</td>
<td>85426</td>
<td>69190</td>
</tr>
<tr>
<td>% Age 18-35</td>
<td>22.8%</td>
<td>20.3%</td>
</tr>
<tr>
<td>% Married</td>
<td>50.3%</td>
<td>52.1%</td>
</tr>
<tr>
<td>% Black</td>
<td>11.2%</td>
<td>7.69%</td>
</tr>
<tr>
<td>Total zipcodes</td>
<td>3870</td>
<td>33120</td>
</tr>
<tr>
<td>Total pop (1000’s)</td>
<td>137584</td>
<td>322072</td>
</tr>
</tbody>
</table>

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

B.2 Realtor.com time-on-market data

Our time-on-market data were downloaded from Realtor.com. Realtor.com calculates time-on-market at the zipcode-month level, using microdata from a variety of multiple listing services. Our dataset covers the time period of May 2012 to April 2018. Since the CoreLogic deed data we use end in August 2017, we only have merged data on price dispersion and time-on-market for the months of May 2012 to August 2017. For our cross-sectional regressions in subsection 4.3, we filter to time-on-market data from May 2012 to December 2016, as the ACS 5-year sample we use to measure zipcode demographics spans the five years from 2012-2016. For the time series analyses of subsections 4.1 and 4.2, as well as the panel analysis of subsection 4.4, we further filter to January 2013 to December 2016, so that we have four full years of data; we did not want to use partial years of data for these analyses because time-on-market fluctuates seasonally.

B.3 ACS zip code demographics

We take demographic information about zipcodes from the ACS 5-year sample spanning the years 2012-2016. We accessed the data using Social Explorer, a commercial provider of pre-aggregated ACS data.

For each zipcode, the demographic characteristics we use are housing unit density per square mile, log average income, unemployment rate (calculated as one minus the
fraction of population employed divided by the fraction of the population in the labor force), 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.

C Additional results and robustness checks

C.1 Excluding characteristics

We estimate residuals using an alternative specification in which we do not allow characteristics to have time-varying effects on prices:

$$p_{izt} = \gamma_i + \eta_{zt} + \epsilon_{izt}$$  \hspace{1cm} (33)

Figure 5 plots the estimated residual standard deviations, aggregated to the zipcode level from 2012-2016, from specification (33) on the y-axis, against those from the baseline specification, (17) on the x-axis. As expected, most points in figure 5 lie above the line $y = x$, indicating that residuals from specification (33) are generally higher than those in the baseline specification, (17), as the baseline specification is strictly more flexible than (33). This is not always true – since the DF correction term

$$\frac{N_z}{N_z - K_z}$$

differs between the two specifications, it is possible for specification (33) to produce smaller error estimates than the baseline specification; thus, there are a few points in figure 5 which lie below the $y = x$ line.

However, the difference between residual estimates is quantitatively small. The average ratio between residual standard deviations from (17) and (33) is 94%, and figure 5 shows that the estimates from the two specifications are highly correlated. Thus, in practice, including the hedonic $f_z(x_i, t)$ term does not appear to have a large quantitative effect on our estimates of idiosyncratic price dispersion.
Figure 5: Effect of including $f_z(x_i, t)$ on estimated price dispersion.

Notes. The y-axis shows estimates of $\hat{\sigma}_z$ using the baseline specification, (17), and the x-axis shows estimates from specification (33), excluding controls for characteristics. Each data point represents a zipcode.
C.2 Other factors correlated with price dispersion

Table 5 shows the coefficients from regressing price dispersion on various features of zipcodes. The explanatory variables are the same used as those in table 2 in the main text, but we allow them to enter specification linearly instead of as third-order polynomials, making their coefficients easier to interpret. First, we note that the coefficient on time-on-market retains its sign and approximate magnitude from table 2 in the main text, suggesting that our results are robust to relaxing functional form assumptions for the control variables.

Another feature of table 5 is that there are a number of variables that correlate robustly with idiosyncratic price dispersion, across different sets of fixed effects. Price dispersion is lower in zipcodes with higher average prices, and with newer houses on average. The coefficient on months from last sale is actually negative, so price dispersion is actually lower in zipcodes in which houses are sold less frequently. Price dispersion is higher in zipcodes with higher income, fewer individuals aged 18-35, higher fraction of the population which is black, more high school and college graduates, less married individuals, higher unemployment rate, and fewer homeowners.
Table 5: Other factors affecting price dispersion in the cross-section of zipcodes

<table>
<thead>
<tr>
<th></th>
<th>LogSD x 100</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>Time on market (months)</td>
<td>1.886***</td>
<td>2.245***</td>
<td>2.792***</td>
</tr>
<tr>
<td></td>
<td>(0.101)</td>
<td>(0.096)</td>
<td>(0.124)</td>
</tr>
<tr>
<td>Log unit density</td>
<td>−0.373***</td>
<td>−0.473***</td>
<td>−0.277***</td>
</tr>
<tr>
<td></td>
<td>(0.066)</td>
<td>(0.056)</td>
<td>(0.057)</td>
</tr>
<tr>
<td>Norm SD square footage</td>
<td>0.160**</td>
<td>0.171**</td>
<td>0.206***</td>
</tr>
<tr>
<td></td>
<td>(0.081)</td>
<td>(0.072)</td>
<td>(0.069)</td>
</tr>
<tr>
<td>Norm SD year built</td>
<td>0.147*</td>
<td>0.320***</td>
<td>0.259***</td>
</tr>
<tr>
<td></td>
<td>(0.076)</td>
<td>(0.067)</td>
<td>(0.062)</td>
</tr>
<tr>
<td>Mean log house price</td>
<td>−3.677***</td>
<td>−2.024***</td>
<td>−2.171***</td>
</tr>
<tr>
<td></td>
<td>(0.192)</td>
<td>(0.201)</td>
<td>(0.220)</td>
</tr>
<tr>
<td>Mean year built</td>
<td>−0.074***</td>
<td>−0.090***</td>
<td>−0.093***</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.004)</td>
<td>(0.004)</td>
</tr>
<tr>
<td>Mean months from last sale</td>
<td>−0.014***</td>
<td>−0.026***</td>
<td>−0.079***</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.005)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>Log mean income</td>
<td>3.933***</td>
<td>2.607**</td>
<td>2.772***</td>
</tr>
<tr>
<td></td>
<td>(0.560)</td>
<td>(0.520)</td>
<td>(0.525)</td>
</tr>
<tr>
<td></td>
<td>(1.540)</td>
<td>(1.351)</td>
<td>(1.334)</td>
</tr>
<tr>
<td>Fraction aged 35-64</td>
<td>−2.343</td>
<td>5.330***</td>
<td>2.753*</td>
</tr>
<tr>
<td></td>
<td>(1.738)</td>
<td>(1.459)</td>
<td>(1.421)</td>
</tr>
<tr>
<td>Fraction black</td>
<td>3.863***</td>
<td>3.756***</td>
<td>3.694***</td>
</tr>
<tr>
<td></td>
<td>(0.486)</td>
<td>(0.446)</td>
<td>(0.432)</td>
</tr>
<tr>
<td>Fraction HS grad</td>
<td>3.457*</td>
<td>13.733***</td>
<td>12.316***</td>
</tr>
<tr>
<td></td>
<td>(1.838)</td>
<td>(1.748)</td>
<td>(1.698)</td>
</tr>
<tr>
<td>Fraction college grad</td>
<td>−1.065</td>
<td>3.862***</td>
<td>3.108***</td>
</tr>
<tr>
<td></td>
<td>(1.035)</td>
<td>(1.022)</td>
<td>(1.017)</td>
</tr>
<tr>
<td>Fraction married</td>
<td>−9.456***</td>
<td>−11.433***</td>
<td>−9.236***</td>
</tr>
<tr>
<td></td>
<td>(1.580)</td>
<td>(1.398)</td>
<td>(1.369)</td>
</tr>
<tr>
<td>Unemployment rate</td>
<td>−2.482</td>
<td>10.772***</td>
<td>12.239***</td>
</tr>
<tr>
<td></td>
<td>(2.046)</td>
<td>(1.809)</td>
<td>(1.970)</td>
</tr>
<tr>
<td>Fraction homeowners</td>
<td>−2.868***</td>
<td>−2.705***</td>
<td>−2.229***</td>
</tr>
<tr>
<td></td>
<td>(0.807)</td>
<td>(0.673)</td>
<td>(0.634)</td>
</tr>
<tr>
<td>Controls</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Fixed effects</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>3,870</td>
<td>3,870</td>
<td>3,870</td>
</tr>
<tr>
<td>R²</td>
<td>0.520</td>
<td>0.711</td>
<td>0.801</td>
</tr>
</tbody>
</table>

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