

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

Nadia Kotova[†] and Anthony Lee Zhang[‡]

PRELIMINARY AND INCOMPLETE, PLEASE DO NOT CITE OR
CIRCULATE WITHOUT PERMISSION

Abstract

This paper studies the sources of idiosyncratic price dispersion in the US housing market. 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 measures of market tightness in panel and cross-sectional regressions. 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.

*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 Drynkin, Darrell Duffie, Liran Einav, Matthew Gentzkow, Guido Imbens, Chad Jones, Eddie Lazear, Paul Milgrom, Sean Myers, Jonas Mueller-Gastell, Michael Ostrovsky, Monika Piazzesi, Peter Reiss, Al Roth, Martin Schneider, Jesse Shapiro, Erling Skancke, Paulo Somaini, Rose Tan, Zach Taylor, Chris Tonetti, Robert Wilson, and Ali Yurukoglu, as well as seminar participants at Stanford, UChicago, the Trans-Atlantic Doctoral Conference, the Young Economists Symposium, and the XX April International Academic Conference.

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

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

1 Introduction

Real estate represents a large component of household wealth in the US. Real estate is not necessarily a poor investment choice; Jordà, Schularick and Taylor (2019) argue that the risk-adjusted total rate of return on house price indices is higher than that of stock market indices in many countries. But 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. Since many households also hold substantial mortgage debt, the value of a household’s primary residence often exceeds the household’s net worth. The literature has shown that a large fraction of the total risk associated with homeownership is idiosyncratic to individual house purchases. However, the sources of idiosyncratic house price dispersion are not yet well understood.

The mechanisms of price formation in the housing market, and the sources of housing booms and busts, are also not well understood. Some authors have suggested that housing booms and busts are partially caused by speculation by boundedly rational agents, whose trading activity is not linked to fundamentals, and thus destabilizes prices.¹ While many papers study trends in average house prices, there is very little work studying the behavior of idiosyncratic house price dispersion over boom and bust cycles. The extent to which housing booms destabilize the relative prices of individual houses can potentially shed some light on the extent to which irrational or boundedly rational speculation destabilizes housing markets in boom periods.

This paper argues that search frictions are an important determinant of idiosyncratic house price volatility. We measure idiosyncratic price dispersion at the level of individual house sales, combining repeat-sales and hedonic methodologies, which allows us to flexibly study price dispersion across locations and over time. We demonstrate a number of new stylized facts: idiosyncratic 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. To rationalize these results, 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.

We measure idiosyncratic price dispersion (IPD) using microdata from CoreLogic on

¹See, for example, Bayer, Mangum and Roberts (2016), Glaeser and Nathanson (2017), DeFusco, Nathanson and Zwick (2017), Mian and Sufi (2018).

36 million house sales over the period 2000-2017. Our measurement strategy combines repeat-sales and hedonic methodologies: we regress log 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 house characteristics to affect the average price of a house, while also allowing observable characteristics to affect a house's price path over time. We measure IPD using the residuals from this regression, which we can aggregate flexibly over locations and time periods. We find that the average of the estimated residual across zipcodes is 16.8% of house prices, with a standard deviation of 4.6%.

We demonstrate a number of new stylized facts about idiosyncratic price dispersion. First, IPD is countercyclical, decreasing during the housing boom, decreasing during the bust, and increasing during the subsequent recovery. Second, IPD is also seasonal, systematically decreasing during the summer hot season of the housing market. Third, IPD is negatively correlated with prices and sales volume, and positively correlated with time-on-market, vacancy rates, and other measures of inverse market tightness.

These stylized facts suggest that an important driver of idiosyncratic house price dispersion is market liquidity or tightness. We formalize this intuition by building a search-and-bargaining model of the housing market. 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 shocks, which allow them to list their houses on the market and become sellers. Sellers have heterogeneous costs 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 prices of identical houses in equilibrium. We show 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. As a result, 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 calibrate the model to aggregate moments of the housing market, and show that the model produces plausible predictions about the differences in sale prices attained by sellers with higher and lower holding costs.

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 housing booms are generated by some form of irrational exuberance, market forces appear to discipline the relative prices of houses fairly well even 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 number of papers studying idiosyncratic house price volatility. Case and Shiller (1988) 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. Our cross-sectional and panel results complement Giacoletti, showing that idiosyncratic price dispersion is correlated not only with sales volume, but with many other measures of market liquidity and tightness. 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 2.4.

More broadly, our work fits into the literature applying search models to housing markets² and in financial markets more generally.³ 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,⁴ 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 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 4 constructs a search-and-bargaining model of the housing market and describes the theoretical predictions. Section 2 describes the data and how we construct our measure of price dispersion. Section 3 contains our empirical results. Section 4 describes our model and theoretical results. Section 5 describes

²See, 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).

³See, for example, Duffie, Gârleanu and Pedersen (2005). Our model is a continuous-time random matching model, as described in Duffie, Qiao and Sun (2017).

⁴See, for example, Mortensen and Pissarides (1994), or the survey article Rogerson, Shimer and Wright (2005)

Table 1: Descriptive statistics

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

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

our calibration exercise. Section 6 discusses implications of our findings and concludes.

2 Data and measurement

2.1 Data sources

We use a number of data sources for our analysis. We use microdata on house sales and characteristics from the CoreLogic tax and deed database, spanning the time period 2000-2017. Other datasets we use are Realtor.com data for listings and time-on-market at the zipcode level; Zillow Research data for total listings and time-on-market at the county level, as well as the Zillow Home Value Index (ZHVI); and the ACS for demographic characteristics of zipcodes and counties. Further details of the steps we take to clean and merge the datasets are described in appendix A.

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 A.1, and descriptive statistics for zipcodes in our primary estimation sample dataset are shown in Table 1. Our primary dataset comprises 36 million house sales within 3,870 zipcodes, over the period 2000-2017. As appendix table 6 shows, our sample contains 11.7% of all zipcodes, but covers around 42.7% of the US population. Our sample is concentrated in relatively large, dense, and high-income zipcodes, but is relatively representative in terms of other demographic characteristics.

2.2 Measuring idiosyncratic price dispersion

Let i index properties, as determined by assessor parcel numbers (APNs). Let x_i represent a vector of time-invariant characteristics of house i ; we use the house's geographic location, square footage, age, and numbers of bedrooms and bathrooms. Let t denote months, z denote zipcodes, and let p_{it} denote the log price of house i sold in month t . We assume the following specification for log house prices:

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

In words, (1) says that prices are determined by a time-invariant house fixed effect, γ_i , a zip-month fixed effect, η_{zt} , a smooth function $f_z(x_i, t)$ of observable house characteristics x_i and time t , and a mean-0 error term ϵ_{it} .

Specification (1) combines repeat-sales and hedonic methodologies. The γ_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 effects on the log price of a given house. The η_{zt} term absorbs parallel shifts in log prices within a zipcode over time. Thus, absent the $f_z(x_i, t)$ term, specification (1) behaves like a repeat-sales specification, implying that the conditional expectations of house prices within a zipcode should follow parallel trends.

The $f_z(x_i, t)$ term relaxes the parallel trends assumption by allowing houses with different observable characteristics x_i to appreciate at different rates.⁵ Houses at different geographic locations may appreciate at different rates, for example if a certain segment of a zipcode begins gentrifying over time; larger or older houses may also appreciate at different rates, since houses with different characteristics are not necessarily perfect substitutes to home buyers. Since we are interested in idiosyncratic house price variation, we include the $f_z(x_i, t)$ term to filter out intertemporal variation in house prices which is correlated with observable house characteristics. Effectively, specification (1) is a hedonic model for changes in relative prices of houses with different characteristics x_i .

⁵Landvoigt, Piazzesi and Schneider (2015) show that returns for houses at different price points in San Diego differ substantially, implying that the $f_z(x_i, t)$ term is potentially important. We do not allow the level of log prices to affect returns directly; this is econometrically difficult since log prices are themselves affected by idiosyncratic price dispersion ϵ_{it} . However, we can capture the differential appreciation trends noted by Landvoigt, Piazzesi and Schneider (2015) as long as most factors which affect average house prices are included in our vector x_i of observable characteristics. x_i contains location, square footage, bedrooms, bathrooms, and year built, so we believe x_i should capture most first-order features which affect the relative prices of houses within a zipcode.

Thus, (1) filters out zipcode-month trends η_{zt} , time-invariant house fixed effects γ_i , and smooth time-varying effects of observables x_i from observed prices; the error term, ϵ_{it} , is our measure of idiosyncratic house price variation.

Before describing how we estimate specification (1), we note two conceptual problems with using specification (1). First, (1) is vulnerable to unobserved characteristics with time-varying effects on prices. Suppose there are some characteristics y_i of houses, which are observed by market participants, but which are not in the vector x_i of characteristics that we observe. Any time-invariant effects of y_i on log house prices will be absorbed into our house fixed effects γ_i ; however, if y_i has time-varying effects on log prices, specification (1) will incorrectly attribute these effects to the idiosyncratic error term ϵ_{it} .

Second, we observe characteristics at a single point in time, so we cannot account for time-varying house characteristics; this means, for example, that we cannot account for the effects of house renovations or improvements on house prices. Again, these effects may have predictable effects on house prices from the perspective of market participants, but we will incorrectly attribute these effects to ϵ_{it} .

Both of these issues will tend to bias our estimates of idiosyncratic price dispersion upwards. However, most of our analysis in section 3 consists of regressing ϵ_{it} against various covariates; unobserved heterogeneity and time-varying characteristics introduce measurement error in ϵ_{it} , which does not affect the interpretation of our regression results as long as they are not correlated with our regression covariates.

Moreover, the magnitude of both issues seems to be fairly small. In appendix B.2, we estimate specification (1) excluding the $f_z(x_i, t)$ term. This decreases the estimated standard deviation of ϵ_{it} only approximately 0.6% of house prices on average, implying that observable characteristics have relatively small time-varying effects on house prices. Assuming x_i contains most characteristics of houses relevant for valuation, unobservable characteristics should also have relatively small time-varying effects on prices. Giacoletti (2017) observes data on remodelling expenditures for houses in California, and performs a similar exercise: comparing figure E.vi to figure 2 in Giacoletti (2017), excluding remodelled properties decreases Giacoletti's estimated standard deviation of house returns by at most around 2% of house prices, which is also a relatively small fraction of total idiosyncratic house price variation.

2.3 Implementation

In order to implement specification (1), we need to choose a functional form for the $f_z(x_i, t)$ functions. Since houses have a large number of characteristics, we cannot estimate a fully interacted polynomial in all characteristics, so we use an additive functional form:

$$f_z(x_i, t) = g_z^{\text{latlong}}(t, \text{lat}_i, \text{long}_i) + g_z^{\text{sqft}}(t, \text{sqft}_i) + g_z^{\text{yrbuilt}}(t, \text{yrbuilt}_i) + g_z^{\text{bedrooms}}(t, \text{bedrooms}_i) + g_z^{\text{bathrooms}}(t, \text{bathrooms}_i) \quad (2)$$

The functions g_z^{latlong} , g_z^{sqft} and g_z^{yrbuilt} are fully-interacted third-order polynomial in their constituent components, and the functions g_z^{bedrooms} and $g_z^{\text{bathrooms}}$ interact dummies for a given house having 1, 2, or 3 or more bedrooms with third-order polynomials in time. Given this functional form for $f_z(x_i, t)$, specification (1) is a standard fixed effects regression, and we estimate specification (1) using OLS separately for each zipcode in our sample.⁶ Once we have estimated specification (1), we estimate squared residuals $\hat{\epsilon}_{it}^2$ for each house sale as:

$$\hat{\epsilon}_{it}^2 = \frac{N_z}{N_z - K_z} (p_{it} - \hat{p}_{it})^2 \quad (3)$$

where N_z is the number of house sales in zipcode z , and K_z is the number of parameters estimated from specification (1). The term $\frac{N_z}{N_z - K_z}$ is a degrees-of-freedom correction, which causes variance estimates to be unbiased at the zipcode level; this is important to include because most houses are sold relatively few times, so the number of parameters K_z is nontrivially large relative to the number of house sales N_z in our dataset.⁷ Equation (3) thus gives us a measure of idiosyncratic price dispersion, $\hat{\epsilon}_{it}^2$, at the level of each individual house sale. We can then average the estimated errors $\hat{\epsilon}_{it}^2$ at various levels of

⁶Intuitively, additivity in specification (2) rules out many interaction effects between characteristics. Older or larger houses can appreciate faster or slower than younger or smaller houses; however, houses which are both large and old are constrained to appreciate at a rate equal to that of houses which are large and houses which are old. The only interaction term we include is the $g_z^{\text{latlong}}(t, \text{lat}_i, \text{long}_i)$ function, which interacts latitude and longitude. This is important because it is implausible that latitude and longitude have additive effects on prices; effectively, this specification allows house prices to vary smoothly as a function of a house's geographic location over time.

⁷More formally, assuming homoskedasticity within zipcodes, $\sigma_{izt}^2 = \sigma_z^2$, the degrees-of-freedom correction in expression (3) causes the expectation of $\hat{\epsilon}_{izt}^2$ to be equal to the true variance, σ_z^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 B.3, 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.

aggregation, and analyze the factors that correlate with idiosyncratic price dispersion.

For example, we can estimate the standard deviation of ϵ_{it} at the zipcode level, over the period 2012-2016, by taking:

$$\hat{\sigma}_z = \sqrt{\frac{\sum_i \hat{\epsilon}_{it}^2}{N_{z,2012-2016}}} \quad (4)$$

where (4) sums over all house sales that happened in zipcode z over the time period 2012-2016, and $N_{z,2012-2016}$ is the number of observations in zipcode z over this period. Figure 1 shows a density plot of $\hat{\sigma}_z$; the mean of $\hat{\sigma}_z$ across zipcodes is 16.8%, and standard deviation 4.6%. The distribution of $\hat{\sigma}_z$ is slightly right-skewed: the 10th percentile is 11.2% and the 90th percentile is 22.6%. In order to calculate how much this affects the dispersion of house returns, note that an agent who buys and sells a house incurs the ϵ_{it} 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 will face a return around 15.9% more variable than the zipcode-month mean price, and an agent in a 90th percentile zipcode incurs 31.9% additional tracking error.

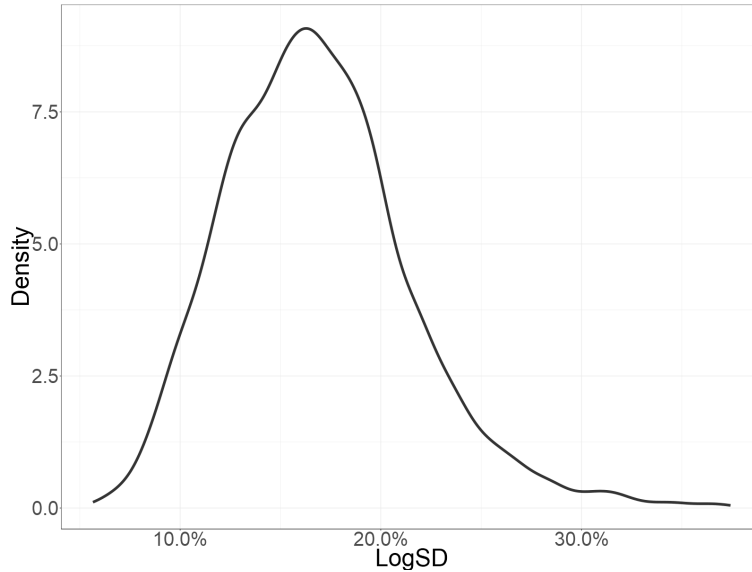
In the following section, we will construct analogs of (4) at different levels of geographical and temporal aggregation. Since aggregates such as $\hat{\sigma}_z$ can be interpreted as the standard deviation of ϵ_{it} , the idiosyncratic component of log house prices, we will often refer to these aggregates as the “logSD” of house prices in the following subsections.

2.4 Literature

A number of other papers have attempted to measure idiosyncratic house price dispersion. Giacoletti (2017) uses the same Corelogic data that we use to measure idiosyncratic price dispersion in the metropolitan areas of San Francisco, San Deigo, and Los Angeles. Unlike our specification (1), Giacoletti uses returns, rather than individual house sales, as the primary unit of analysis.⁸ Similarly, Sagi (2015) also shows that return variances are very flat with respect to holding-period length, using data on commercial real estate sales.

⁸There 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.

Figure 1: Distribution of $\hat{\sigma}_{zt}$



Notes: Distribution of $\hat{\sigma}_z$. One data point is a zipcode.

Relative to these papers, the benefit of our specification is that it allows us to trends in idiosyncratic price dispersion over time. In section 3, 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 $f_z(x_i, t)$ term in specification (1), which effectively allows the hedonic coefficients on

different characteristics to change continuously over time.

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. 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, as we discuss in footnote 5, to the extent that the factors which affect prices are summarized by our house characteristics x_t , our specification will also be able to capture these trends.

Quantitatively, Giacoletti (2017), using data from 1989 to 2013, finds that the standard deviation of idiosyncratic component of returns is approximately 9.6%-11.8% in San Diego, 13.9%-16.5% in Los Angeles, and 13.7-17.6% in San Francisco. Landvoigt, Piazzesi and Schneider (2015) finds a similar SD of 8.8%-13.8% for San Diego over the time horizon 1999-2007. In our sample, over the time period 2000-2017, we estimate return standard deviations of 15.7% for San Diego, 16.8% for Los Angeles, and 19.1% for San Francisco.⁹ Our estimates are thus roughly in line with the estimates from Giacoletti (2017) and Landvoigt, Piazzesi and Schneider (2015), preserving the ordering of idiosyncratic price dispersion between the three regions, although our estimates are somewhat higher than theirs. Moreover, similar to our findings in subsection 3.1 below, Landvoigt, Piazzesi and Schneider (2015) finds that price dispersion increased during the 2008 housing bust, though their sample does not include the subsequent recovery. Thus, our paper, Giacoletti (2017), and Landvoigt, Piazzesi and Schneider (2015) arrive at similar estimates using different methodologies, datasets, time horizons, and geographic definitions, suggesting that idiosyncratic price dispersion can be measured fairly robustly across locations and over time.

⁹To calculate these quantities, we take sales-weighted averages of $\hat{\sigma}_z^2$ for all zipcodes within the San Diego-Carlsbad-San Marcos, San Francisco-Oakland-Fremont, and Los Angeles-Long Beach-Anaheim CBSAs. We then multiply $\hat{\sigma}_z$ by a factor of $\sqrt{2}$, to convert standard deviations of prices at each sale to standard deviations of returns, for comparability to Giacoletti (2017) and Landvoigt, Piazzesi and Schneider (2015).

3 Empirical results

In this section, we document new stylized facts about idiosyncratic house price dispersion: we show that price dispersion is countercyclical, seasonal, and correlated with time-on-market and other measures of market tightness in panel and cross-sectional regressions. Details of the construction of the datasets used in this section are discussed in appendices A.6 and A.7.

3.1 Idiosyncratic price dispersion over the business cycle

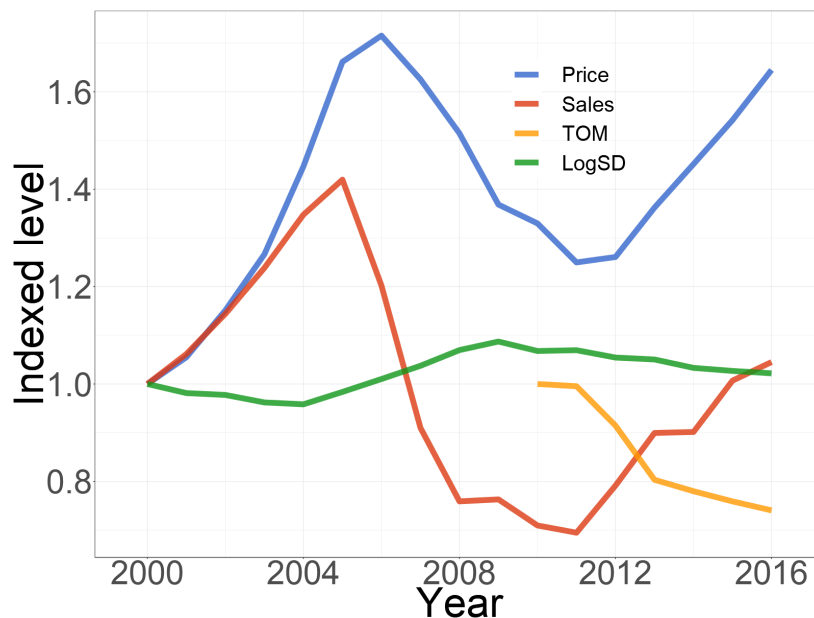
In figure 2, we show the behavior of indexed total sales, logSD, prices, and time-on-market at the yearly level. Figure 2 shows that idiosyncratic house price dispersion is countercyclical. Price dispersion is decreasing from 2000-2004, as the housing market is booming and volume and prices are increasing; from 2004 onwards, idiosyncratic price dispersion reverses direction and starts increasing, peaking in 2010, in the midst of the housing bust, as 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 we only observe time-on-market from 2010 onwards, time-on-market is also smoothly decreasing throughout the 2010-2016 recovery.

Quantitatively, total housing sales in our sample increased 41% from 2000 to 2005, dropped to a trough of 69% of its 2000 value in 2011, and recovered to 4.5% above its 2000 value in 2016. Average prices increase 72% from 2000 to 2007, drop to 25% 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 perfectly with the time series for sales or price series, logSD drops to 4.1% (in units of $\hat{\sigma}$, 0.7%) below its 2000 value near the peak of the boom in 2004, increases to 8.7% (1.4%) above its 2000 value in 2009, and decreases to 2.2% (0.3%) above its 2000 value in 2016. From 2010 to 2016, time-on-market decreases 26%.

We believe that we the first paper to empirically document the countercyclicality of idiosyncratic house price dispersion.¹⁰ The cyclical changes in price dispersion are 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

¹⁰An 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 2: 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) 2010-2016. All variables are indexed so that they are equal to 1 in the first year that we observe them.

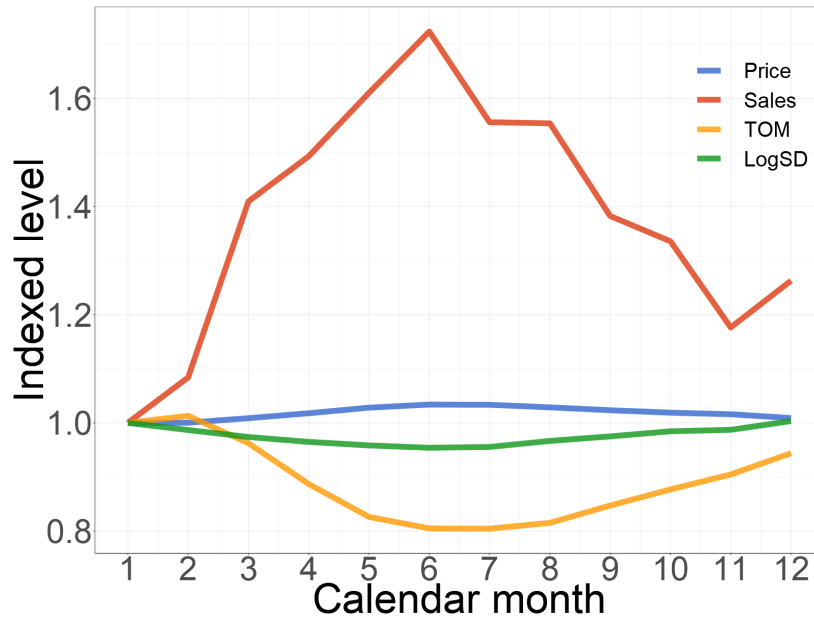
logSD across zipcodes.

3.2 Seasonality

Figure 3 shows the seasonal behavior of prices, total sales, time-on-market, and logSD, aggregated to the level of calendar months over the period 2010-2016. Total sales, time-on-market and logSD are from our data. Prices are from the FHFA US-wide monthly house price index; we use the FHFA index rather than the ZHVI because the ZHVI is seasonally adjusted, so cannot capture price variation over calendar months. We filter out low-frequency trends in all four time series, through a procedure we describe in appendix A.5.

Figure 3 shows that all four variables are seasonal. Quantitatively, there are on average 82% more house sales in June than in January. Prices are around 3.4% higher in June, time-on-market is around 35% lower, and logSD is around 7.2% (in terms of house prices, 1.2%) lower. The seasonality of sales, prices, and time-on-market is known in the literature

Figure 3: 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 2010-2016. All variables are indexed by dividing by their January level.

(Ngai and Tenreyro, 2014), but we believe we are the first to show that idiosyncratic house price dispersion is also seasonal. The magnitude of the seasonal variation is also nontrivially large, at around 35% of the seasonal effect on average house prices.

3.3 Panel regressions

Table 2 shows county-year panel regressions of logSD on various covariates. We run panel regressions at the year level instead of the month level to avoid picking up seasonal variation in our regressions. Similarly to subsection 3.1 above, we use Zillow’s single-family residence ZHVI’s to measure prices at the county-year level.

Columns 1 and 2 show that logSD is negatively correlated with prices and sales volume, and column 3 shows that price dispersion is negatively correlated with time-on-market; thus, the panel regressions support the findings in figures 2 and 3. Note that the panel variation is formally distinct from the variation shown in figures 2 and 3: figure 2 focuses on aggregate trends in outcome variables, on average across counties within a given

year, and figure 3 takes average within months; in contrast, the estimates in table 2 absorb correlated variation into year fixed-effects. Thus, for example, the interpretation of columns 1-2 is that, when a county-year experiences an unusually large increase change in prices or volume, relative to other counties in the same year, it also experiences an unusual decline in price dispersion.

Columns 4-6 regress price dispersion on three measures of market tightness: column 4 uses the fraction of the total housing stock which is vacant, column 5 uses the yearly population growth rate, and column 6 uses the fraction of the housing stock which is listed for sale. Price dispersion is positively correlated with vacancy rates and the fraction of the housing stock which is listed, though the latter coefficient is insignificant, and is negatively correlated with population growth rates. Together, these three specifications suggest that price dispersion is lower in tighter markets with more buying pressure.

Column 7 includes all variables in a single specification. The coefficients on vacancy rates and population growth rates remain significant with unchanged signs, whereas other coefficients lose significance.

3.4 Zipcode cross-sectional regressions

In this section, we study variation in idiosyncratic price dispersion in the cross-section of zipcodes. As described in appendix A.7, we aggregate price dispersion and a variety of dependent variables to the zipcode level over the period 2012-2016. We regress price dispersion on various dependent variables, controlling for third-order polynomials in a number of control variables: the average age of sold houses, average 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. Results are shown in table 3.

Columns 1-4 of table 3 show that, analogous to the time-series and panel specifications in previous subsections, price dispersion is correlated with time-on-market and various measures of market tightness: vacancy rates, the fraction of the housing stock which is listed for sale, and the population growth rate, measured as the percentage change in each zipcode's population between the 2008-2012 and 2012-2016 ACS 5-year samples.

Column 5 shows that logSD is also correlated cross-sectionally with average prices in the cross-section of zipcodes. Column 6 includes all variables together; the signs of all coefficients are unchanged, though the magnitudes change somewhat. Columns 7 and 8

Table 2: County-year panel regressions

	(1)	(2)	(3)	(4)	(5)	(6)
			LogSD x 100			
Log ZHVI	-1.056*** (0.369)					-0.834 (0.768)
Log sales		-0.971*** (0.194)				-1.750*** (0.541)
Time on market (months)			0.521*** (0.162)			0.170 (0.175)
Vacancy rate				16.204*** (1.798)		10.804*** (2.407)
Pop growth rate					-8.726*** (2.057)	-4.758 (3.779)
County fixed effects	X	X	X	X	X	X
Year fixed effects	X	X	X	X	X	X
Sample period	2000-2016	2000-2016	2010-2016	2007-2016	2007-2016	2010-2016
N	10,366	10,366	2,516	5,807	5,284	2,492
Adjusted R ²	0.858	0.859	0.911	0.895	0.891	0.919

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

add CBSA and state fixed effects; the signs of all coefficients are unchanged, and most coefficients remain significant.

3.5 Robustness checks

We run a variety of robustness checks for our main empirical results. In our main specifications, we run cross-sectional regressions at the zipcode level and panel regressions at the county-year level; appendix B.1 shows that results from zipcode-year panel regressions and county cross-sectional regressions are quantitatively similar to those in the main text. In appendix B.2, we show results from estimating price dispersion without including the polynomial term $f_z(x_i, t)$ in house characteristics and time; this represents the results from a pure repeat-sales regression, with house and zipcode-month fixed effects. We show that not including the polynomial term has quantitatively small effects on results.

The method we use to estimate idiosyncratic price variance is potentially vulnerable to two flaws. First, if there is a random walk component of price dispersion, which depends on the time between sales, our specification will not capture this. Second, we will mechanically infer lower idiosyncratic price dispersion for houses which are sold more times in our dataset. In appendix B.3, we account for these effects by removing all components of price variance, at the zipcode level, which can be explained by a flexible function of average time-between-sales and the number of times a house is sold. We show that most of our empirical results continue to hold, though the magnitudes of our estimated coefficients decrease somewhat. Finally, in appendix B.4, we show that most of our county-level results are robust to using either Realtor.com or Zillow as a source for time-on-market, although the county panel coefficient on time-on-market loses significance.

4 Model

In this section, we construct a search model to rationalize the empirical results described in the preceding section.

Table 3: Zipcode cross-sectional regressions

	LogSD x 100						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Time on market (months)	2.463*** (0.088)				1.910*** (0.095)	2.941*** (0.120)	2.345*** (0.093)
Vacancy rate		15.335*** (0.829)			7.486*** (0.852)	2.121*** (0.770)	4.412*** (0.757)
Pop growth			-1.729* (0.886)		0.401 (0.783)	-1.130** (0.572)	-1.095* (0.639)
Mean log price					-3.406*** (0.193)	-1.075*** (0.222)	-1.643*** (0.199)
Controls	X	X	X	X	X	X	X
Fixed effects						State	CBSA
N	4,109	4,109	4,109	4,109	4,109	4,109	4,109
Adjusted R ²	0.542	0.496	0.455	0.497	0.580	0.797	0.732

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

4.1 Primitives

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

4.1.1 Sellers

There is some mass M_S of sellers in the market who are waiting to sell their houses to buyers. As we describe in subsection 4.1.3, sellers are matched homeowners who have received separation shocks from their houses. Once a seller successfully sells her house, she permanently leaves the market, attaining a continuation value which we normalize to 0.

Sellers are heterogeneous, characterized by some time-invariant *holding utility*, $v > 0$, which she receives as a flow per unit time that she keeps her house on the market. As we will describe in subsection 4.1.3, holding utilities v are drawn independently from a distribution $F(v)$, at the time that a matched owner chooses to unmatched with her house and become a seller. Holding utilities v play an important role in our model: sellers with higher values of v are more patient, so they are more willing to wait longer to sell for better prices. This will generate price dispersion in our model. How much price dispersion this generates will depend on market tightness.

Differences in holding utilities v represent differences in sellers' urgency to sell, which may come from a number of sources. 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.¹¹ Certain sellers may be more credit constrained than others, for example if they need to sell their house before they can afford to buy another one. Certain sellers may be renting their houses out while attempting to sell, so they are more willing to wait for higher sale prices.

In equilibrium, a seller with holding utility v attains some expected value $V_S(v)$, which

¹¹This is related to Anenberg and Bayer (2013), who study a model in which sellers may buy houses before attempting to sell their houses; sellers who buy before selling and who sell before buying will thus endogenously have different flow utilities from their houses.

we will characterize in subsection 4.2.2 below. A seller who matches with a buyer and sells her house for some price P receives utility P and continuation value 0, so her net gain in expected value is:

$$P - V_S(v)$$

We use $F_{eq}(v)$ denote the distribution of holding utilities among sellers in stationary equilibrium. $F_{eq}(v)$ will in general differ from $F(v)$, the distribution of holding utilities among sellers entering the market, because sellers with higher holding utilities will wait longer to sell their houses. We describe how $F_{eq}(v)$ is determined in subsection 4.2.1 below.

4.1.2 Buyers

There is some mass M_B of buyers who are present in the market, who are trying to purchase a house. Buyers waiting to match with sellers are identical, and receive flow utility normalized to 0 while waiting. Buyers enter the market at some exogeneous flow rate η_B . We assume that $\eta_B < \lambda_M$, where λ_M is the rate at which matched owners receive separation shocks and become homeowners; we show in subsection 4.2.1 that this condition is necessary for the model to admit a stationary equilibrium. Buyers can only exit the market by purchasing houses from sellers.

Buyers meet sellers through a process we describe in subsection 4.1.4 below. When a buyer meets a seller, he draws, independently across matches, some *match utility* $\epsilon \sim G(\cdot)$ for the house. We will occasionally refer to ϵ also as representing *match quality*. If the buyer buys the house, he becomes a matched homeowner, receiving ϵ from the house per unit time, until he receives a separation shock and becomes a seller.

We assume that the distribution of match utilities $G(\cdot)$ is exponential, with standard deviation σ_ϵ , and lower bound ϵ_0 , which may differ from 0. The assumption of exponential values is required for proposition 2, which analytically decomposes price variance into components attributable to seller and buyer values; appendix C.3 shows how the results change when we allow for general match utility distributions.

Since unmatched buyers are undifferentiated, the expected value of an unmatched buyer in stationary equilibrium is some scalar V_B , which we will characterize in subsection 4.2.2 below. If a buyer matches with a seller, draws match utility ϵ , and purchases the house at price P , he become a matched homeowner with value $V_M(\epsilon)$, so he achieves a

net gain in expected value of:

$$V_M(\epsilon) - P - V_B$$

Our baseline model assumes buyers are undifferentiated for simplicity, as it implies that V_B is a scalar, so the model is easier to solve computationally. In practice, buyers may also have different urgencies to buy houses, so their flow utilities for remaining unmatched may differ. The model is robust to including buyer heterogeneity: appendix C.6 derives equilibrium conditions under persistent buyer heterogeneity, and shows that all of the main theoretical results of the model concerning price dispersion – propositions 2, 3, and 4 – continue to hold.

4.1.3 Matched homeowners

Matched homeowners are buyers who have purchased houses, and have not yet received separation shocks to become sellers. Since there is a unit mass of houses, and each house is owned either by a matched homeowner or a seller, the total mass of matched homeowners is $1 - M_S$. Matched homeowners with match utility ϵ receive flow utility ϵ from their house per unit time. Matched homeowners receive separation shocks at some exogenous Poisson rate λ_M ; upon receiving a separation shock, a matched homeowner draws some v from the exogenous distribution $F(\cdot)$, and becomes a seller with flow utility v . A separation shock can be thought of as, for example, a job offer in a different city, which forces the homeowner to decide to sell her house and move out of the market.

We use $G_{eq}(\epsilon)$ denote the distribution of match utilities among matched homeowners in stationary equilibrium. This will in general differ from $G(\epsilon)$, the distribution of match utilities that buyers draw when they meet sellers, because higher draws of ϵ are more likely to result in successful trade. We will describe how $G_{eq}(\epsilon)$ is determined in subsection 4.2.1 below.

4.1.4 Matching

Matches between buyers and sellers are generated at a flow rate $m(M_B, M_S)$, which depends on the masses of unmatched buyers M_B and unmatched sellers M_S present in the market. As is standard in the classical search-and-matching literature,¹² we assume

¹²See Rogerson, Shimer and Wright (2005) for a survey of this literature.

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 happens at Poisson rates λ_B and λ_S respectively, given by:

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

Since buyer match utilities are drawn randomly from $G(\epsilon)$ each time a match occurs, sellers meet buyers with match utilities $\epsilon \sim G(\cdot)$; buyers meet sellers with holding utilities drawn from $F_{eq}(v)$, the stationary distribution of seller holding utilities.

4.1.5 Price determination

Suppose that a buyer is matched with a seller with holding utility v , and the buyer draws match utility ϵ . If the buyer and seller trade, the sum of their value functions is the buyer's continuation value $V_M(\epsilon)$, plus the seller's continuation value, which we have normalized to 0. If they do not trade, the buyer receives the expected value from being an unmatched buyer, V_B , and the seller receives $V_S(v)$. Thus, the bilateral match surplus from trade when the buyer's match utility is ϵ and the seller's holding utility is v is:

$$V_M(\epsilon) - V_B - V_S(v) \tag{5}$$

We assume that trade occurs if the bilateral match surplus is nonnegative. This implies that a seller with holding utility v will trade with any buyer that draws a match utility higher than some *trade cutoff* $\epsilon^*(v)$. Rearranging expression (5), $\epsilon^*(v)$ must satisfy:

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

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

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

that is, no seller type wishes to trade with all buyer types. This is an assumption on equilibrium objects, so it is hard to verify; we do not impose it in the calibration, but we

impose it in this section because it is needed to prove our theoretical results.

When trade occurs, we assume that prices are set through Nash bargaining. When a seller with holding utility v meets a buyer with match utility ϵ , the price that the seller receives is equal to her outside option, $V_S(v)$, plus a share θ of the bilateral match surplus; that is,

$$P(v, \epsilon) = V_S(v) + \theta (V_M(\epsilon) - V_B - V_S(v)) \quad (6)$$

4.2 Stationary equilibrium

4.2.1 Flow equality

There are four flows in our model: buyers enter the market, buyers turn into matched owners, matched owners turn into sellers, and sellers sell and leave the market. We analyze the model in a stationary equilibrium, so all flows of all agent types must be equal.

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

$$(1 - M_S) \lambda_M f(v) \quad (7)$$

In words, this is the product of the total mass of matched homeowners, $1 - M_S$; the rate at which shocks homeowners receive separation shocks, λ_M ; and the density $f(v)$ of entering sellers with value v .¹³ The equilibrium rate at which sellers of type v sell their houses and leave the market is:

$$M_S f_{eq}(v) \lambda_S (1 - G(\epsilon^*(v))) \quad (8)$$

In words, this is the product of the mass of sellers, M_S ; the density of values among sellers in equilibrium, $f_{eq}(v)$; 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 $\epsilon^*(v)$ for a seller of type v , which is $1 - G(\epsilon^*(v))$. In stationary equilibrium, expressions (7) and (8) 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,

¹³Since the distribution $F(v)$ of holding utilities v does not depend on matched homeowners' match utility ϵ , we do not need to explicitly integrate over the distribution $G_{eq}(\epsilon)$ in expression (7).

integrating (7) and (8) over v , we have:

$$(1 - M_S) \lambda_M = \int_v \lambda_S M_S (1 - G(\epsilon^*(v))) f_{eq}(v) dv \quad (9)$$

Moreover, since each successful sale turns a buyer into a matched homeowner, the RHS of (9) 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:

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

In words, the numerator is the flow rate at which a seller of value v successfully trades with a buyer with match utility below ϵ , integrated over the equilibrium distribution $f_{eq}(v)$ of holding utilities v among sellers. The denominator is the RHS of (9), the total flow rate at which buyers become matched homeowners.

Finally, the rate at which buyers enter the market must be equal to all other flow rates; using (9), this condition can be written as:

$$(1 - M_S) \lambda_M = \eta_B \quad (10)$$

Since M_S cannot be negative, expression (10) shows why we require $\eta_B < \lambda_M$ in order for the model to admit a stationary equilibrium.

4.2.2 Value functions

Given expression (6) for prices, we can write the seller, buyer, and matched owner continuous-time Bellman equations. Given the buyer match rate λ_B , trade cutoffs $\epsilon^*(v)$, the equilibrium distribution of seller values $F_{eq}(v)$, and the seller value function $V_S(v)$, the equilibrium value of buyers, V_B , must satisfy:

$$rV_B = \lambda_B \int \int_{\epsilon > \epsilon^*(v)} [(1 - \theta)(V_M(\epsilon) - V_B - V_S(v))] dG(\epsilon) dF_{eq}(v) \quad (11)$$

In words, expression (11) 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, $\epsilon^*(v)$, 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 $\epsilon^*(v)$, and the buyer value V_B , the seller value function $V_S(v)$ satisfies:

$$rV_S(v) = v + \lambda_S \int_{\epsilon > \epsilon^*(v)} \theta (V_M(\epsilon) - V_B - V_S(v)) dG(\epsilon) \quad (12)$$

In words, expression (12) states that a seller of type v receives flow value v 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 $\epsilon > \epsilon^*(v)$, 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:

$$rV_M(\epsilon) = \epsilon + \lambda_M \left(\int V_S(v) dF(v) - V_M(\epsilon) \right) \quad (13)$$

In words, expression (13) 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 $V_S(v)$ over the seller holding utility distribution $F(v)$.

4.2.3 Equilibrium conditions

Collecting equilibrium conditions from earlier subsections, in our model, stationary equilibrium is characterized by the following set of equations.

Proposition 1. *Given primitives*

$$r, \lambda_M, \eta_B, \alpha, \phi, \theta, F(v), \epsilon_0, \sigma_\epsilon^2$$

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

Buyer, seller, and matched owner Bellman equations:

$$rV_B = \lambda_B \int \int_{\epsilon > \epsilon^*(v)} [(1 - \theta)(V_M(\epsilon) - V_B - V_S(v))] dG(\epsilon) dF_{eq}(v) \quad (14)$$

$$rV_S(v) = v + \lambda_S \int_{\epsilon > \epsilon^*(v)} \theta(V_M(\epsilon) - V_B - V_S(v)) dG(\epsilon) \quad (15)$$

$$rV_M(\epsilon) = \epsilon + \lambda_M \left(\int V_S(v) dF(v) - V_M(\epsilon) \right) \quad (16)$$

Trade cutoffs:

$$V_M(\epsilon^*(v)) = V_S(v) + V_B \quad (17)$$

Matching rates:

$$M_S \lambda_S = M_B \lambda_B = \alpha M_B^\phi M_S^{1-\phi} \quad (18)$$

Flow equality:

$$(1 - M_S) \lambda_M f(v) = \lambda_S M_S f_{eq}(v) (1 - G(\epsilon^*(v))) \quad (19)$$

$$G_{eq}(\epsilon) = \frac{\int_v \lambda_S M_S \left[\int_{\tilde{\epsilon}=\epsilon_0}^{\epsilon} 1(\tilde{\epsilon} > \epsilon^*(v)) dG(\tilde{\epsilon}) \right] dF_{eq}(v)}{\int_v \lambda_S M_S (1 - G(\epsilon^*(v))) dF_{eq}(v)} \quad (20)$$

$$(1 - M_S) \lambda_M = \eta_B \quad (21)$$

4.3 Price dispersion

Since houses in the model are identical, we measure price dispersion in the model as the variance of prices, $P(v, \epsilon)$, with respect to the joint distribution of v and ϵ among buyers and sellers whose meetings result in trade. Proposition C.2 shows that equilibrium price dispersion admits a simple analytical characterization.

Proposition 2. *In stationary equilibrium, the variance of $P(v, \epsilon)$ among trading sellers and buyers is:*

$$\underbrace{\text{Var}_{v \sim F(\cdot)}(V_S(v))}_{\text{Seller holding utility}} + \underbrace{\left(\frac{\theta}{r + \lambda_M} \right)^2 \sigma_\epsilon^2}_{\text{Buyer match utility}} \quad (22)$$

Expression (22) shows that equilibrium price dispersion arises from two sources: differences in sellers' value functions $V(v)$, caused by differences in sellers' holding utilities; and differences in buyers' match utilities ϵ . The buyer value term only depends on $\theta, r, \lambda_M, \sigma_\epsilon^2$,

which are primitives of the model. The seller value term $V_S(v)$ is an equilibrium object which we cannot analytically characterize, but there is a simple analytical expression for its derivative $V'_S(v)$. Define the expected time-on-market in equilibrium for a seller of type v , $\text{TOM}(v)$, as:

$$\text{TOM}(v) \equiv \frac{1}{\lambda_S (1 - G(\epsilon^*(v)))} \quad (23)$$

In words, time-on-market is the inverse of $\lambda_S (1 - G(\epsilon^*(v)))$, 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 v that result in trade, $(1 - G(\epsilon^*(v)))$.

Proposition 3. *The seller value function $V_S(v)$ satisfies:*

$$V'_S(v) = \frac{\text{TOM}(v)}{r\text{TOM}(v) + \theta} \quad (24)$$

Qualitatively, proposition 3 shows that $V'_S(v)$ is higher, so the dispersion in seller values is higher, when equilibrium time-on-market $\text{TOM}(v)$ is higher. Intuitively, variation in sellers' holding utilities v only have large effects on sellers' expected values if sellers have to spend a long time on the market. In a market where sellers are quickly matched to high-valued buyers and equilibrium time-on-market is low, total expected holding costs are low regardless of whether sellers' holding costs per unit time, v , are high or low; thus, dispersion in holding costs does not translate into substantial dispersion in sellers' value function $V_S(v)$. Conversely, if equilibrium time-on-market is high, total holding costs are much higher, and total holding costs increase more the larger sellers' per-unit-time holding costs v are. 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, $\text{TOM}(v)$.

To formalize the link between time-on-market and price dispersion, appendix C.5 shows that, if time-on-market $\text{TOM}(v)$ increases, pointwise for every v , then equilibrium price dispersion will also increase.

Proposition 4. *Fix $F(v)$, σ_ϵ^2 , r , θ . Consider two sets of model parameters,*

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

such that time-on-market is uniformly higher in stationary equilibrium under Θ_1 ; that is, letting

$TOM_{\Theta_1}(v)$ denote the equilibrium time-on-market function under Θ_1 ,

$$TOM_{\Theta_1}(v) > TOM_{\Theta_2}(v) \quad \forall v$$

Then equilibrium price dispersion will also be higher under Θ_1 than Θ_2 .

Together, propositions 3 and 4 show that there is a tight link between time-on-market and price dispersion: fixing the interest rate r , bargaining power θ , the distribution of seller holding utilities $F(v)$ and the variance of match quality σ_ϵ^2 , the equilibrium time-on-market function $TOM(v)$ is a sufficient statistic for equilibrium price dispersion; that is, all remaining parameters of the model only affect equilibrium price dispersion through their effect on the equilibrium time-on-market function. Any set of parameters which increases equilibrium time on market for all seller types will also increase equilibrium price dispersion.

4.4 Comparative statics

To illustrate how our model maps primitives to outcomes, we solve the model computationally, and show comparative statics with respect to various input parameters in figure 4. We parametrize $F(v)$ as uniform on the interval

$$[\bar{v} - \Delta_v, \bar{v} + \Delta_v]$$

so that \bar{v} is the mean of entering seller values, and Δ_v is a parameter governing the spread of seller values. Choices for other parameters are described in appendix C.8.

For ease of computation, in figure 4, we currently fix M_B as an exogenous parameter, instead of holding η_B fixed; we will fix this in a future version of the draft.

The top row of figure 4 shows the effect of varying M_B, ϵ_0, \bar{v} on outcomes. Intuitively, these parameters can be thought of as market tightness parameters, which shift tightness without affecting the dispersion of buyer match quality or seller holding utilities: increasing M_B increases the mass of buyers, increasing ϵ_0 increases the average level of buyers' match utilities without affecting their dispersion, and decreasing \bar{v} causes sellers to want to sell more urgently. The effects of these parameters on time-on-market, and price dispersion are thus quantitatively and qualitatively similar: increasing M_B or ϵ_0 , or decreasing \bar{v} , causes time-on-market and price dispersion to decrease. Moreover, the

relative slopes of the time-on-market and price dispersion lines is relatively similar for all graphs in the top row. This illustrates the result of propositions 2 and 3: holding fixed r, λ_M, θ , and the dispersions of buyer and seller utilities $\sigma_\epsilon, \Delta_v$, time-on-market is close to a sufficient statistic for price dispersion.¹⁴

The bottom row of figure 4 shows the effect of varying other parameters. Increasing the dispersion of buyer values, σ_ϵ , increases prices, time-on-market, and price dispersion. This is because increasing buyer match quality heterogeneity increases the option value of waiting for a better match, leading to longer equilibrium waiting times and higher prices. Relatively speaking, increasing σ_ϵ increases price dispersion more quickly, per unit change in time-on-market, compared to the variables M_B, ϵ_0, \bar{v} in the top row of figure 4. This is because increasing σ_ϵ has two effects on price dispersion: an indirect effect through increasing time-on-market, which affects the seller holding cost term in expression (22), and a direct effect through increasing the buyer match utility dispersion term in expression (22). Similarly to σ_ϵ , increasing Δ_v increases time-on-market and price dispersion, but the relative effect on price dispersion is much larger than the effect of the pure tightness variables M_B, ϵ_0, \bar{v} , because Δ_v has a direct effect on price dispersion through increasing the dispersion of sellers' per-unit-time holding costs.

4.4.1 Buyer value heterogeneity and price dispersion

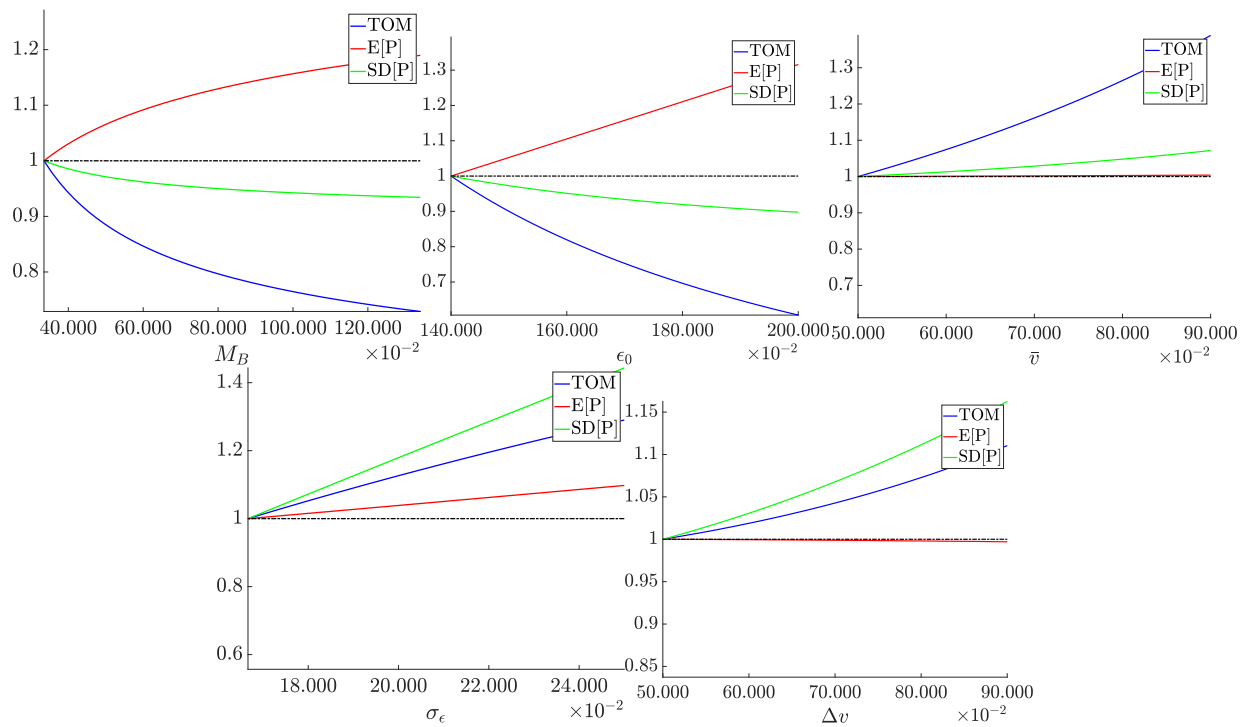
In section 3, we showed that time-on-market and price dispersion are correlated over years, calendar months, and in panel and cross-sectional regressions. Table 4 shows the correlation between TOM and PD across these specifications. To extract implied coefficients from the yearly and seasonal plots of figures 2 and 3, we run univariate regressions of price dispersion on time-on-market. The "county panel" row of table 4 reports the coefficient in column 3 of table 2,¹⁵ and the "zipcode cross-sectional" row reports the range of estimates of the time-on-market coefficient from tables 3 and 5

Table tab:coef table shows that the implied regression coefficients between price dispersion and time-on-market from the yearly, calendar month, and panel specifications are similar, but the zipcode cross-sectional regressions produce a much larger coefficient. The

¹⁴The equilibrium time-on-market function $TOM(v)$ is formally a sufficient statistic for price dispersion; average time-on-market is not sufficient. However, the top row of figure 4 illustrates that, in practice, the relationship between average time-on-market and price dispersion seem to be fairly robust to various ways to move time-on-market.

¹⁵We exclude the time-on-market coefficient in column 7 of table 4, as it is not significantly distinguishable from 0.

Figure 4: Comparative statics



Notes. Average prices, time-on-market, and the standard deviation of equilibrium prices, as model parameters vary. All variables are indexed so they are equal to 1 at the lowest value of the x variable.

model suggests a tentative explanation for this finding. Dispersion in buyer match quality, σ_ϵ , is likely to be fairly stable over time within a location, but may vary substantially across locations. In locations where the housing stock is fairly homogeneous, σ_ϵ will be low, as houses are relatively good substitutes; if the housing stock is more heterogeneous, match quality matters more and σ_ϵ should be higher.

Under this interpretation, the coefficient in a regression of price dispersion on time-on-market reflects variation in market tightness over time, that is, variables in the top row of figure 4. From 4, we thus expect the implied correlation between time-on-market and price dispersion is thus expected to be low in relative terms, and fairly consistent across specifications. The cross-sectional regression coefficient, in contrast, contains both variation in tightness and variation in match quality dispersion, σ_ϵ . The bottom-left panel of figure 4 shows that variation in σ_ϵ will tend to move price dispersion more, per unit that time-on-market moves.

We can test this hypothesis as follows. If σ_ϵ varies across zipcodes, figure 4 predicts that time-on-market should be higher in markets with higher match quality dispersion. Moreover, if we attempt to control for σ_ϵ , comparing zipcodes with more similar dispersion in match quality, the coefficient in a regression of price dispersion on time-on-market should decrease.

Table 5 tests these predictions. As proxies for σ_ϵ , we use the standard deviation in year built and square footage across houses sold in 2012-2016 for each zipcode. These are measures of house heterogeneity; we believe they are reasonable proxies for σ_ϵ because buyers' values will be less disperse if houses in a given market are more similar. In columns 1-3, we regress time-on-market on SD year built and SD square footage, separately and combined; in all specifications, the coefficient is negative, so zipcodes with more heterogeneous housing stocks have higher time-on-market, confirming our first prediction. In columns 4 and 5, we logSD on time-on-market, controlling for SD year built and SD square footage; as predicted, the coefficient on time-on-market decreases.

This suggests that buyer value heterogeneity may partially explain why the cross-sectional relationship between logSD and TOM is steeper than the panel and time-series relationships. The coefficient in column 5 is still around twice as large as the panel and time-series relationships; however, our proxies for σ_ϵ are far from perfect, so it is conceivable that the TOM coefficient could be reduced further with better proxies for σ_ϵ .

Table 4: Correlation between TOM and PD across different specifications

Type	Coef
Yearly	0.695
Seasonal	0.667
Panel	0.521
Cross-sectional	1.498 - 2.941

Notes. Estimates of the correlation between price dispersion and time-on-market from yearly, seasonal, panel and cross-sectional analyses.

5 Calibration (PRELIMINARY AND INCOMPLETE)

In order to quantify the tradeoffs agents face within the model, we calibrate the model to roughly match a representative city in the US, over the time period 2010-2016.

5.1 Parametric Assumptions

In order to bring the model to data, as in subsection 4.4 above, we parametrize $F(\cdot)$ as a uniform distribution, on the interval:

$$[\bar{v} - \Delta_v, \bar{v} + \Delta_v]$$

so that \bar{v} is the mean of entering seller values, and Δ_v is a parameter governing the spread of seller values.

5.1.1 Externally calibrated parameters

We assume that a unit of time in the model is equal to one year. We set the yearly discount rate $r = 0.052$, so that the annual discount factor is 0.95. We assume symmetric bargaining power, so $\theta = 0.5$. We choose the matching function elasticity ϕ to 0.84, based on Genesove and Han (2012), who estimate this elasticity using National Association of Realtors survey data, which captures both buyer and seller time-on-market; this estimate is also used by Anenberg and Bayer (2013).

Since we do not observe buyer time-on-market, we cannot separately identify the

Table 5: Zipcode cross-sectional regressions with house heterogeneity measures

	Time on market (months)			LogSD x 100	
	(1)	(2)	(3)	(4)	(5)
Time on market (months)				1.910*** (0.095)	1.498*** (0.101)
Norm SD yr built	0.214*** (0.010)		0.174*** (0.010)		0.625*** (0.069)
Norm SD sqft		0.188*** (0.011)	0.124*** (0.011)		0.309*** (0.073)
Vacancy rate	4.250*** (0.118)	3.936*** (0.120)	4.114*** (0.117)	7.486*** (0.852)	9.339*** (0.857)
Pop growth	-0.454*** (0.122)	-0.130 (0.124)	-0.345*** (0.121)	0.401 (0.783)	-0.176 (0.777)
Mean log price	-0.302*** (0.030)	-0.267*** (0.030)	-0.301*** (0.029)	-3.406*** (0.193)	-3.668*** (0.192)
Controls	X	X	X	X	X
N	4,109	4,109	4,109	4,109	4,109
Adjusted R ²	0.494	0.475	0.509	0.580	0.593

Notes. Each data point is a zipcode. Regressions are weighted by the average number of sales in a given zipcode over the time period 2012-2016. “Norm SD yr built” and “Norm SD sqft” are, respectively, the standard deviation of year built and square footage among houses sold during 2012-2016 within the given zipcode, normalized across all zipcodes in our sample. Column 4 is identical to column 5 of table 3

matching efficiency parameter α and the mass of buyers M_B , as they have identical effects on matching efficiency and thus seller time-on-market; thus, we normalize the match efficiency parameter to $\alpha = 1$.

5.1.2 Parameters Estimated by Method of Moments

The remaining unknown parameters are the parameters of buyer and seller flow value distributions, $\epsilon_0, \sigma_\epsilon, \bar{v}, \Delta_v$, the buyer inflow rate η_B , and the separation rate λ_M . We will calibrate the model to exactly match moments of each year in the data from 2010-2016, treating each year as a separate steady-state equilibrium. The parameter Δ_v , the dispersion of seller values, will be set to a constant across all years; we will describe how we set Δ_v in the following subsection.

For each year, we choose $\epsilon_0, \sigma_\epsilon, \bar{v}, \eta_B, \lambda_M$ to match the following moments:

- Average house prices (Zillow ZHVI)
- Average dollar idiosyncratic price dispersion (our calculations)
- Average time-on-market (Zillow Research)
- Turnover rates: Sales volume as a fraction of total housing stock (Corelogic deed + tax data)
- Average number of house visits by buyers (9.96, from Genesove and Han (2012))

Technically, fixing Δ_v , we match these 5 moments using GMM, running gradient descent over our 5 parameters. Within the model, all parameters essentially affect all moments, but the intuitive mapping between parameters and moments is as follows. Average house prices are matched by shifting the levels of \bar{v} and ϵ_0 upwards. The level of idiosyncratic price dispersion is matched by increasing σ_ϵ . Average time on market and sales volume are jointly matched by choosing η_B and λ_M , the buyer entry rate and the matched owner separation rate, which jointly contribute to pinning down the stationary equilibrium levels of M_B and M_S . Finally, the average number of house visits pins down the difference between ϵ_0 and \bar{v} ; if buyer values tend to be much higher than seller values, more meetings will result in trade, and vice versa.

We set the Δ_v parameter to be constant across years, and we choose it to match our estimated regression coefficient of price dispersion on time-on-market. As table 4 shows,

we have a number of different estimates for this coefficient; lower values of this correlation are more conservative, causing us to infer lower Δ_v values, and thus lower effects of liquidity on seller outcomes. We thus attempt to match the panel coefficient of 0.521, choosing parameters so that, when tightness varies in the model, the regression coefficient of price dispersion on time-on-market is 0.521.

Matching this coefficient in the model is nontrivial, because the relationship between time-on-market and price dispersion in the model is nonlinear. Our model is not well-suited for simulating a panel regression, because our model matches an average county in the data, and does not have county-level persistent heterogeneity; thus, we simulate a cross-sectional regression in the model, and match the regression coefficient to our estimated panel coefficient from the data. Precisely, for each year, we create a grid of TOM's, by varying tightness through the buyer entry rate η_B , which matches the cross-sectional distribution of county-level TOMs in the data. We then pool all simulated values of TOM and PD across years, and regress price dispersion on time-on-market using the simulated data across all years; we choose Δ_v so that this correlation coefficient is equal to 0.521.

NOTE: The calibration results of this section are very preliminary, and will move around a lot as we vary the calibration strategy in future drafts.

5.2 Results

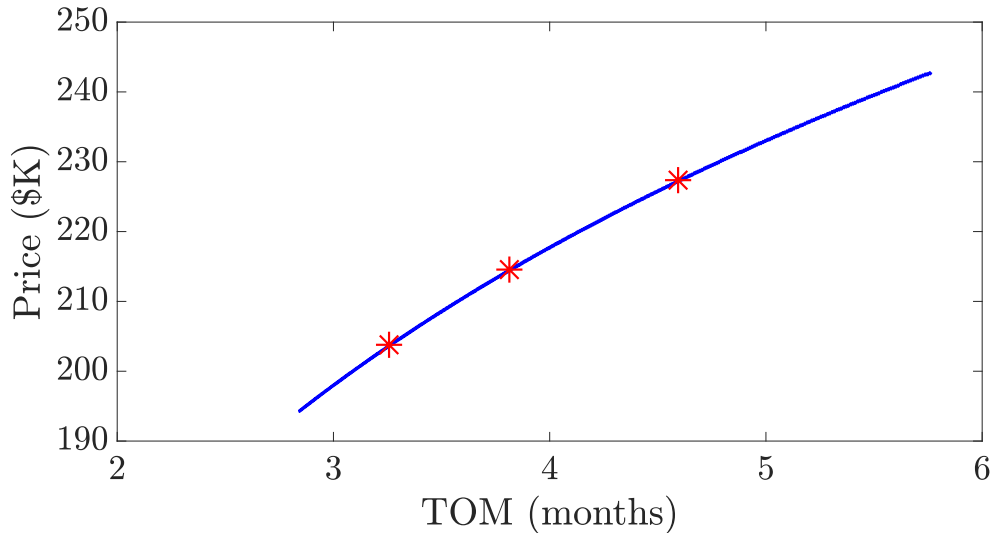
We have currently calibrated the model only to the year 2010. Figures 5 and 6 show two simple counterfactuals that our model generates.

Figure 5 shows, for a market with time-on-market set to approximately 4 months, the tradeoff¹⁶ that sellers with different holding utilities face between time-on-market and sale prices. Sellers with higher holding utilities wait longer to sell their houses, but achieve higher prices as a result. Under our estimates, waiting an extra month on the market increases the price that a seller attains by just under \$20,000, or just under 10% of average house prices.

Figure 6 shows the difference in expected sale price for agents with high (75th percentile) and low (25th percentile) values of v , as we vary tightness; the x-axis shows average

¹⁶This is somewhat an abuse of terminology; figure 5 does not truly represent a “tradeoff”, because in the context of Nash bargaining models, agents’ types are assumed to be common knowledge, so agents cannot simply replicate the outcomes attained by other agents.

Figure 5: Time-on-market and price tradeoff



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

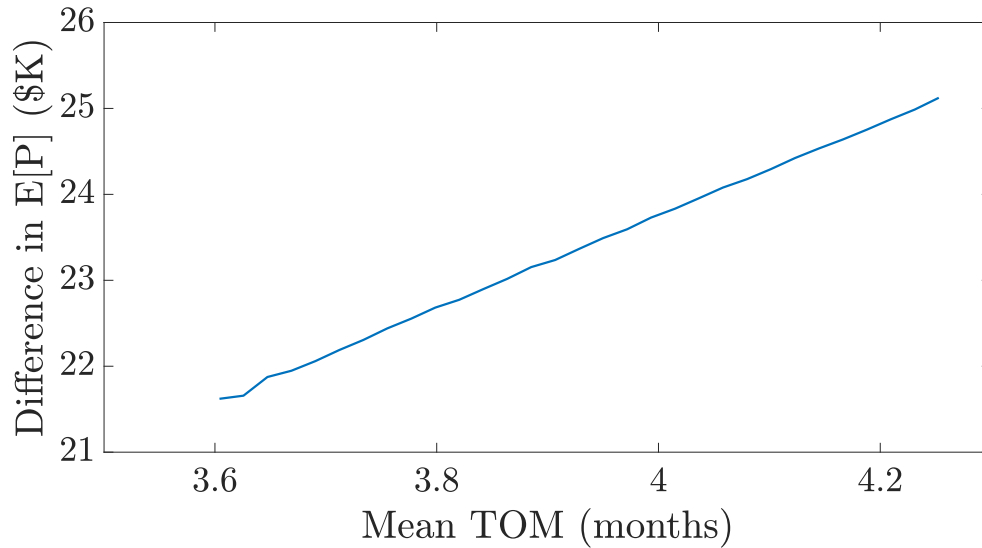
time-on-market. In tighter markets, time-on-market is lower, so differences in prices sellers with different values attain are smaller. Quantitatively, decreasing time-on-market from 4.2 to 3.6 months decreases the 75-25 price gap by around \$3000.

6 Discussion (PRELIMINARY AND INCOMPLETE)

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

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. Informally, one might summarize our results as saying that liquidity is an important driver of idiosyncratic house price dispersion. Many recent theories of housing market booms and busts

Figure 6: Liquidity discounts



Notes. Difference in price attained by sellers with 75th and 25th percentile values of v , vs average time-on-market.

emphasize some degree of belief heterogeneity or imperfect rationality.¹⁷ Under these theories, extrapolative or boundedly 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 is destabilizing 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.

6.2 Idiosyncratic risk and household welfare

The evidence in this paper adds to a few recent papers suggesting that liquidity is an important driver of idiosyncratic house price volatility (Sagi, Giacoletti (2017)). A common argument in the literature is that house price volatility is misleading as a measure of risk because housing is a hedge against increasing local costs of living.¹⁸ This may be true

¹⁷See, for example, Glaeser and Nathanson (2017), DeFusco, Nathanson and Zwick (2017), Bayer, Mangum and Roberts (2016), and Mian and Sufi (2018).

¹⁸See, for example, Han (2010) and Han (2013).

about *average* house prices, but the *idiosyncratic* component of house price volatility is a pure risk, that a single household is unable to diversify, and does not hedge against anything from the perspective of the household.

Our results showing that idiosyncratic volatility varies over calendar months, and over the business cycle, also have welfare implications. 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, but in the seasonal case, the average and idiosyncratic components have roughly the same order of magnitude. Thus, changes in expected idiosyncratic risk may be important to account for in measuring the effects of a business cycle and seasonal fluctuations on agents' welfare.

6.3 Mechanisms

We have shown that time-on-market is correlated with price dispersion; our model rationalizes this using a stylized search-and-bargaining framework, in which low time-on-market causes sellers' outside options to converge towards each other. This is obviously an unrealistic model in practice, and a question left unanswered is how price convergence occurs in the housing market. For example, prices could be stabilized by the presence of a relatively small number of arbitrageurs; by fire sales happening at more stable prices; or other reasons. Future work should analyze the mechanism by which/who is responsible for stabilizing housing prices and decreasing idiosyncratic price volatility.

References

- Albrecht, James, Pieter A. Gautier, and Susan Vroman.** 2016. "Directed search in the housing market." *Review of Economic Dynamics*, 19: 218 – 231. Special Issue in Honor of Dale Mortensen.
- Anenberg, Elliot, and Patrick Bayer.** 2013. "Endogenous Sources of Volatility in Housing Markets: The Joint Buyer-Seller Problem." National Bureau of Economic Research Working Paper 18980.

- Bayer, Patrick, Christopher Geissler, Kyle Mangum, and James Roberts.** 2011. "Speculators and Middlemen: The Strategy and Performance of Investors in the Housing Market." National Bureau of Economic Research w16784, Cambridge, MA.
- Bayer, Patrick, Kyle Mangum, and James W Roberts.** 2016. "Speculative Fever: Investor Contagion in the Housing Bubble." National Bureau of Economic Research Working Paper 22065.
- Case, Karl, and Robert Shiller.** 1988. "The Efficiency of the Market for Single-Family Homes." National Bureau of Economic Research w2506, Cambridge, MA.
- Clayton, Jim, Norman Miller, and Liang Peng.** 2010. "Price-volume Correlation in the Housing Market: Causality and Co-movements." *The Journal of Real Estate Finance and Economics*, 40(1): 14–40.
- DeFusco, Anthony A, Charles G Nathanson, and Eric Zwick.** 2017. "Speculative Dynamics of Prices and Volume." National Bureau of Economic Research Working Paper 23449.
- Diaz, Antonia, and Belen Jerez.** 2013. "House Prices, Sales, and Time on the Market: A Search-Theoretic Framework." *International Economic Review*, 54(3): 837–872.
- Duffie, Darrell, Lei Qiao, and Yeneng Sun.** 2017. "Continuous time random matching."
- Duffie, Darrell, Nicolae Gârleanu, and Lasse Heje Pedersen.** 2005. "Over-the-counter markets." *Econometrica*, 73(6): 1815–1847.
- Genesove, David, and Christopher Mayer.** 2001. "Loss Aversion and Seller Behavior: Evidence from the Housing Market." *The Quarterly Journal of Economics*, 116(4): 1233–1260.
- Genesove, David, and Lu Han.** 2012. "Search and matching in the housing market." *Journal of Urban Economics*, 72(1): 31–45.
- Giacoletti, Marco.** 2017. "Idiosyncratic Risk in Housing Markets." Social Science Research Network SSRN Scholarly Paper ID 2995323, Rochester, NY.
- Giacoletti, Marco, and Victor Westrupp.** 2017. "The Risk-Adjusted Performance of Asset Flippers." Social Science Research Network SSRN Scholarly Paper ID 2892439, Rochester, NY.

- Glaeser, Edward L, and Charles G Nathanson.** 2017. "An extrapolative model of house price dynamics." *Journal of Financial Economics*, 126(1): 147–170.
- Han, Lu.** 2010. "The Effects of Price Risk on Housing Demand: Empirical Evidence from U.S. Markets." *The Review of Financial Studies*, 23(11): 3889–3928.
- Han, Lu.** 2013. "Understanding the Puzzling Risk-Return Relationship for Housing." *The Review of Financial Studies*, 26(4): 877–928.
- 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.
- Jordà, Òscar, Moritz Schularick, and Alan M Taylor.** 2019. "The Total Risk Premium Puzzle." National Bureau of Economic Research Working Paper 25653.
- 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.
- Leung, Charles, Garion Lau, and Youngman Leong.** 2002. "Testing Alternative Theories of the Property Price-Trading Volume Correlation." *Journal of Real Estate Research*, 23(3): 253–264.
- Mian, Atif, and Amir Sufi.** 2018. "Credit supply and housing speculation." National Bureau of Economic Research.
- Mortensen, Dale T, and Christopher A Pissarides.** 1994. "Job creation and job destruction in the theory of unemployment." *The review of economic studies*, 61(3): 397–415.
- Ngai, L. Rachel, and Silvana Tenreyro.** 2014. "Hot and Cold Seasons in the Housing Market." *American Economic Review*, 104(12): 3991–4026.
- 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.
- Piazzesi, Monika, and Martin Schneider.** 2009. "Momentum traders in the housing market: survey evidence and a search model." *American Economic Review*, 99(2): 406–11.

- Piazzesi, Monika, Martin Schneider, and Johannes Stroebel.** 2015. "Segmented Housing Search." National Bureau of Economic Research Working Paper 20823.
- Rogerson, Richard, Robert Shimer, and Randall Wright.** 2005. "Search-Theoretic Models of the Labor Market: A Survey." *Journal of Economic Literature*, 43(4): 959–988.
- Sagi, Jacob S.** 2015. "Asset-Level Risk and Return in Real Estate Investments." *SSRN Electronic Journal*.
- Stein, Jeremy C.** 1995. "Prices and Trading Volume in the Housing Market: A Model with Down-Payment Effects." *The Quarterly Journal of Economics*, 110(2): 379–406.
- Wheaton, William C.** 1990. "Vacancy, Search, and Prices in a Housing Market Matching Model." *Journal of Political Economy*, 98(6): 1270–1292.

Appendix

A Data description and cleaning

A.1 CoreLogic tax and deed data

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

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

We use the dataset that results from these cleaning steps to measure monthly sales by 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 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 for "house flips", as well as instances where reported sale price seems anomalous. To filter for flips, 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)). To filter for potentially anomalous prices,

Table 6: Characteristics of zipcodes in the primary dataset

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

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

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,

Second, specification (1) 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 (1), and thus we filter to zipcodes 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 6 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.

A.2 ACS

We use demographic information about zipcodes and counties from the ACS. For county-years, we use ACS 1-year samples spanning the years 2006-2017; for zipcodes, we use

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. The demographic 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.

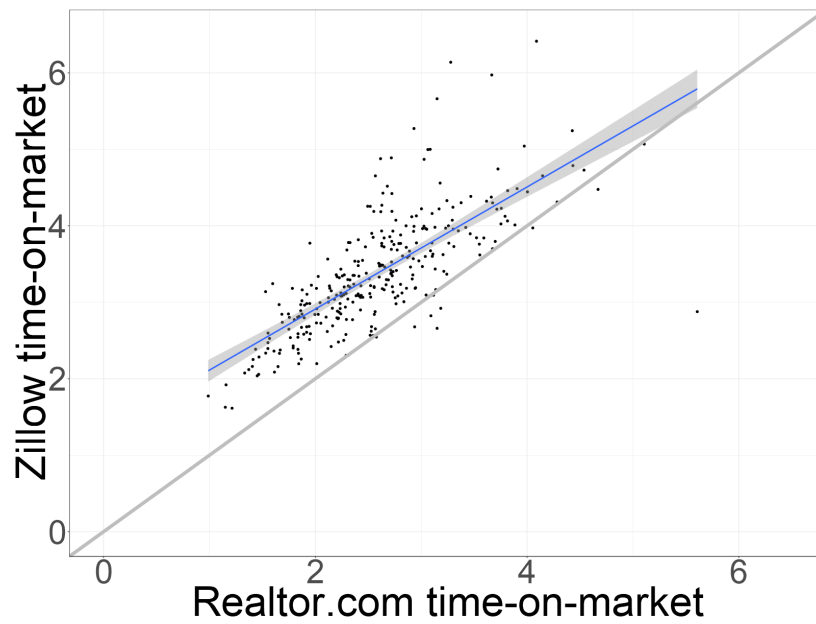
A.3 Zillow Research

At the county level, we use Zillow Research data on time-on-market, daily listings, and the Zillow Home Value Index (ZHVI) at the zipcode-month level. The ZHVI is available from 1997-2018, time-on-market is available from 2010-2018, and daily listings data is available from 2013-2018.

A.4 Realtor.com

For our zipcode-level results, we use data on time-on-market and total housing listings from Realtor.com. Realtor.com calculates median time-on-market and total listings at the zipcode-month level, using microdata from multiple listing services. The dataset we downloaded covers the time period May 2012 to April 2018. Zillow.com and Realtor.com time-on-market differ somewhat, because Zillow's measure is based on how long houses spend on the Zillow platform, whereas Realtor.com's data is based on multiple listing services. Figure 7 plots average time-on-market on Realtor.com and Zillow, respectively, for counties between the years 2012-2016. The time-on-numbers are strongly correlated, although Zillow's time-on-market numbers tend to be higher than Realtor.com's, and there is some independent variation. Appendix B.4 shows results from running our county-level regression specifications using time-on-market numbers from Realtor.com instead of Zillow Research.

Figure 7: Zillow vs Realtor.com time-on-market



Notes. Zillow vs Realtor.com time-on-market. Each data point is a county. Time-on-market is calculated as an average over county-months from 2012-2016.

A.5 Yearly and seasonal data construction

To construct the dataset used in figure 2, we first filter to counties which we observe every year from 2000-2016; this leaves us with 361 counties, comprising approximately 38.8 million home sales. To construct the LogSD line in figure 3, 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. We only have time-on-market data for seven complete years, from 2010-2016, but we are able to measure total sales, logSD, and prices for the longer time horizon 2000-2016.

To construct the seasonal dataset, we filter to the years 2010-2016, as we only observe time-on-market for these years. This leaves us with 262 counties, comprising approximately 13.8 million home sales over this time period. We first collapse the data to year-month level. To collapse, we take the sum over sales in all counties, the mean over all $\hat{\epsilon}_{it}^2$ terms that we estimate, and the sales-weighted average of time-on-market. For the price line, we take the monthly FHFA house price index for the entire US, also for the time horizon 2010-2016.

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

A.6 County-year and county-year-month data construction

To construct the county-year dataset, we first aggregate the estimated errors $\hat{\epsilon}_{it}$ from specification (1) to the county-year level. We merge to aggregated Zillow Research data to the county-year level, taking averages of time-on-market, ZHVI, and daily listings for all county-months within a county year, and we merge to ACS county-year data from 2007-2016. We construct the fraction of houses which are listed as Zillow’s daily listings divided by total housing units for a given county-year from the ACS. We use an

unbalanced panel for the panel regressions of subsection 3.3. For figure 3.1, we balance the panel, keeping only county-years that we see for the entire period 2000-2017. Balancing the panel decreases the number of counties represented from 733 to 361, but only drops around 18% of house sales from the dataset

To construct the monthly dataset, used in figure 3, we aggregate total house sales from Corelogic and estimated errors $\hat{\epsilon}_{it}$ from specification (1) to the county month level, for all months within 2010-2018. We then merge to Zillow Research data on county-month time-on-market and ZHVI. We balance the panel, keeping only county-months that we observe for all 84 months from January 2010 to December 2018. We then aggregate all variables to the calendar month level, weighting average time-on-market, prices, and the square of $\hat{\epsilon}_{it}$ by total sales, and we plot the resultant averages in figure 3.

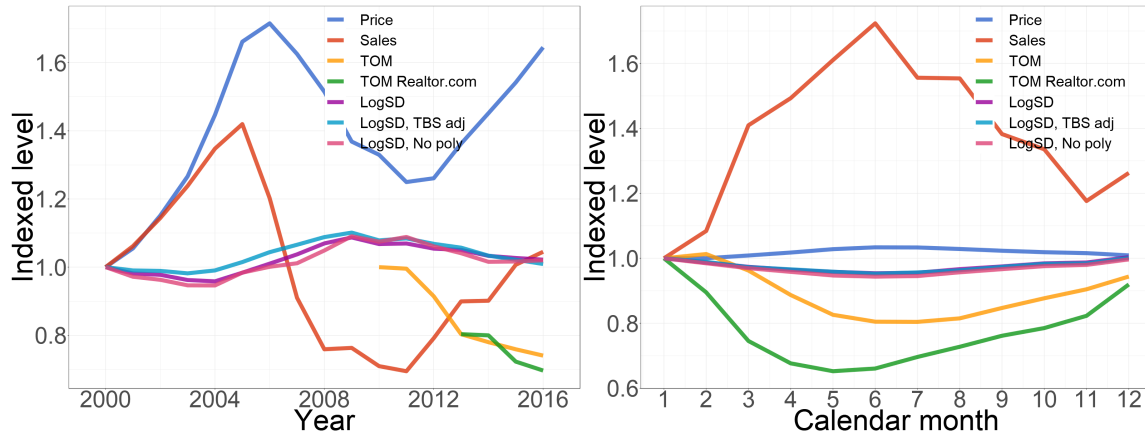
The zipcode-year level dataset used in appendix table 8 is constructed analogously, by aggregating price dispersion to the zipcode-year level, then merging to time-on-market from Realtor.com and the ZHVI from Zillow. Zillow's ZHVI is available at the zipcode-month level, but Zillow's time-on-market measure is not.

A.7 Zipcode cross-sectional data construction

To construct the zipcode-level dataset for the cross-sectional regressions of subsection 3.4, we take the average of the estimated residuals $\hat{\epsilon}_{it}^2$ for each zipcode in our sample for the time period 2012-2016. We then take listings and time-on-market data from Realtor.com, filter to the same period 2012-2016, and aggregate to the zipcode level; we calculate zipcode-level listings averaging over all zipcode-months we observe, and we calculate zipcode-level time-on-market by taking the listings-weighted average of median time-on-market for each zipcode-month. We measure total housing units and other demographic covariates for zipcodes using the 2012-2016 ACS 5-year sample, as described in appendix A.2 above.

We construct the county cross-sectional dataset used in appendix table 7 analogously, using Corelogic data for sales and price dispersion, Zillow and Realtor.com data for time-on-market, total listings, and the ZHVI, and the ACS 2012-2016 5-year sample for demographic covariates. We filter the Corelogic, Zillow, and Realtor.com datasets to the time period 2012-2016 to match the time period of the ACS dataset.

Figure 8: Yearly and monthly variation in price dispersion, robustness checks



Notes. Figures 3 and 2, where “LogSD” represents the original estimate $\sqrt{\hat{\sigma}_t^2}$, “LogSD, TBS adj” represents $\sqrt{\hat{\sigma}_{\text{TBSadj},t}^2}$, and “LogSD, No poly” represents $\sqrt{\hat{\sigma}_{\text{Nopoly},t}^2}$.

B Robustness checks

B.1 Zipcode-year and county cross-sectional regressions

Table 7 runs cross-sectional regressions at the county level, instead of the zipcode level as in table 3 in the main text. Most coefficients are significant, and the results are qualitatively similar: price dispersion is positively correlated with time-on-market, vacancy rate, and the mean fraction of the housing stock which is listed. Most results lose significance with state fixed effects, but this is likely because there are too few data points given the number of control coefficients we are estimating.

Similarly, table 8 reports results from zipcode-year panel regressions. Coefficients on prices, sales, and time-on-market are significant and have the same signs as those in table 2 in the main text, although the magnitudes of coefficients change somewhat. Since the publically available ACS 1-year samples do not include zipcode information, we are unable to measure vacancy rates, population growth rates, and listing fractions at the zipcode-year level.

Table 7: County cross-sectional regressions

	LogSD x 100				
	(1)	(2)	(3)	(4)	(5)
Time on market (months)	1.019*** (0.336)			0.694** (0.336)	1.481*** (0.409)
Vacancy rate		16.726*** (3.128)		16.820*** (3.899)	11.240*** (3.330)
Mean log price			-4.121*** (0.630)	-4.535*** (0.898)	-3.625*** (1.004)
Controls	X	X	X	X	X
Fixed effects					State
N	299	473	473	299	299
Adjusted R ²	0.443	0.461	0.477	0.510	0.717

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

Table 8: Zipcode-year panel regressions

	(1)	(2)	(3)	(4)
		LogSD x 100		
Log ZHVI	-1.201*** (0.098)			-1.703*** (0.352)
Log sales		-0.862*** (0.120)		-0.178 (0.169)
Time on market (months)			0.263*** (0.071)	0.243*** (0.073)
Zip fixed effects	X	X	X	X
Year fixed effects	X	X	X	X
Sample period	2000-2016	2000-2016	2013-2016	2013-2016
N	52,061	52,061	12,257	12,257
Adjusted R ²	0.848	0.850	0.924	0.924

Notes. Each data point is a zipcode-year. Regressions are weighted by the average number of sales in a given zipcode over all years we observe.

B.2 Excluding characteristics

As a robustness check, we estimate idiosyncratic price dispersion using an alternative specification in which we omit the polynomial $f_z(x_i, t)$ term, so that we do not allow characteristics to have time-varying effects on prices:

$$p_{izt} = \gamma_i + \eta_{zt} + \epsilon_{izt} \quad (25)$$

Call the resultant estimates $\hat{\epsilon}_{\text{Nopoly},it}^2$. Figure 9 plots $\hat{\epsilon}_{\text{Nopoly},it}^2$, aggregated to the zipcode level from 2012-2016, from specification (25) on the y-axis, against those from the baseline specification, (1) on the x-axis. As expected, most points in figure 9 lie above the line $y = x$, indicating that residuals from specification (25) are generally higher than those in the baseline specification, (1), as the baseline specification is strictly more flexible than (25). This is not always true – since the DF correction term

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

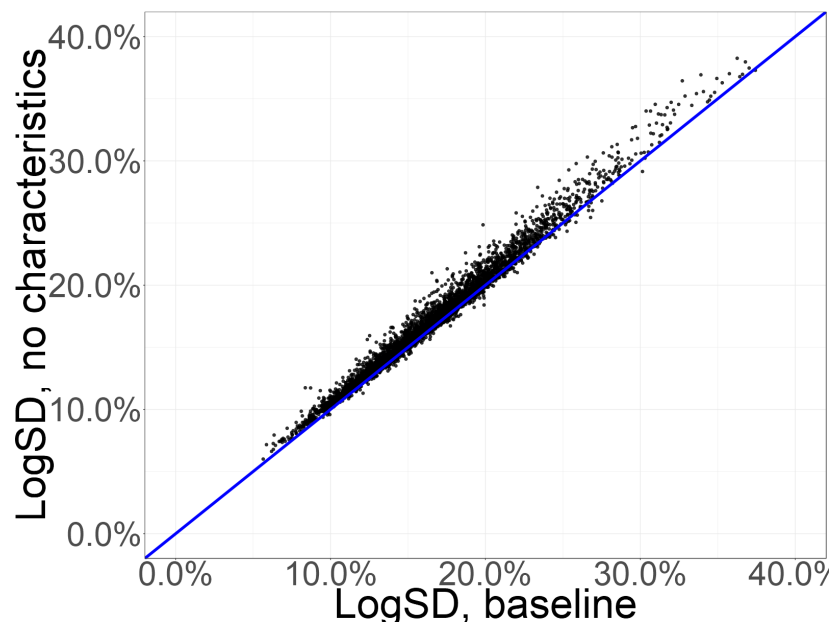
differs between the two specifications, it is possible for specification (25) to produce smaller error estimates than the baseline specification; thus, there are a few points in figure 9 which lie below the $y = x$ line. However, the difference between residual estimates is quantitatively small. The average ratio between residual standard deviations from (1) and (25) is 1.038, and figure 9 shows that the estimates from the two specifications are highly correlated. Thus, excluding $f_z(x_i, t)$ term does not appear to have a large quantitative effect on our estimates of $\hat{\epsilon}_{it}$.

To verify that the inclusion of the $f_z(x_i, t)$ term does not drive our results, we re-run the analyses of section 3 using $\hat{\epsilon}_{\text{Nopoly},it}^2$ in place of $\hat{\epsilon}_{it}^2$. The pink line in figure 8, labelled “LogSD, No poly”, shows the seasonal and time-series behavior of $\hat{\epsilon}_{\text{Nopoly},it}^2$. We find that $\hat{\epsilon}_{\text{Nopoly},it}^2$ tracks $\hat{\epsilon}_{it}^2$ closely, seasonally and over the business cycle. Tables 9, 10, 11 and 12 re-run all zipcode and county cross-sectional regressions using $\hat{\epsilon}_{\text{Nopoly},it}^2$ in place of $\hat{\epsilon}_{it}^2$; results are qualitatively and quantitatively similar to those in the main text.

B.3 Adjusting for time-between-sales and number of sales

Implicitly, our specification (1) assumes that idiosyncratic price variance does not depend on the holding period. This is rejected, for example, if house prices follow a random

Figure 9: 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, (1), and the x-axis shows estimates from specification (25), excluding controls for characteristics. Each data point represents a zipcode.

Table 9: Zipcode cross-sectional regressions, no polynomial term

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	LogSD x 100						
Time on market (months)	2.413*** (0.090)				1.779*** (0.097)	2.880*** (0.126)	2.229*** (0.097)
Vacancy rate		16.383*** (0.839)			9.066*** (0.869)	3.386*** (0.808)	5.474*** (0.787)
Pop growth			-1.340 (0.901)		0.757 (0.799)	-0.777 (0.600)	-0.774 (0.664)
Mean log price					-3.450*** (0.197)	-0.962*** (0.233)	-1.623*** (0.207)
Controls	X	X	X	X	X	X	X
Fixed effects						State	CBSA
N	4,109	4,109	4,109	4,109	4,109	4,109	4,109
Adjusted R ²	0.548	0.514	0.468	0.509	0.588	0.789	0.727

Notes. Re-running all specifications in table 3 without the $f_z(x_i, t)$ polynomial term. Each data point is a zipcode. Regressions are weighted by the number of sales within the zipcode over the time period 2012-2016.

Table 10: Zipcode-year panel regressions, no polynomial term

	LogSD x 100			
	(1)	(2)	(3)	(4)
Log ZHVI	-1.572*** (0.103)			-1.347*** (0.401)
Log sales		-0.929*** (0.115)		-0.438** (0.174)
Time on market (months)			0.534*** (0.075)	0.485*** (0.077)
Zip fixed effects	X	X	X	X
Year fixed effects	X	X	X	X
Sample period	2000-2016	2000-2016	2013-2016	2013-2016
N	52,036	52,036	12,257	12,257
Adjusted R ²	0.830	0.830	0.917	0.917

Notes. Re-running all specifications in table 8 without the $f_z(x_i, t)$ polynomial term. Each data point is a zipcode-year. Regressions are weighted by the average number of sales in a given zipcode over all years we observe.

Table 11: County cross-sectional regressions, no polynomial term

	LogSD x 100				
	(1)	(2)	(3)	(4)	(5)
Time on market (months)	0.945*** (0.346)			0.610* (0.344)	1.299*** (0.425)
Vacancy rate		16.931*** (3.186)		17.528*** (3.991)	11.275*** (3.456)
Mean log price			-4.329*** (0.639)	-4.809*** (0.919)	-3.824*** (1.042)
Controls	X	X	X	X	X
Fixed effects					State
N	299	473	473	299	299
Adjusted R ²	0.446	0.468	0.487	0.517	0.713

Notes. Re-running all specifications in table 7 without the $f_z(x_i, t)$ polynomial term. Each data point is a county. Regressions are weighted by the number of sales within the county over the time period 2012-2016.

Table 12: County-year panel regressions, no polynomial term

	(1)	(2)	(3)	(4)	(5)	(6)
			LogSD x 100			
Log ZHVI	-1.308*** (0.434)					-2.592*** (0.925)
Log sales		-1.010*** (0.226)				-1.476** (0.663)
Time on market (months)			0.339* (0.182)			0.082 (0.200)
Vacancy rate				19.166*** (1.942)		12.286*** (3.067)
Pop growth rate					-10.240*** (3.027)	-4.827 (5.387)
County fixed effects	X	X	X	X	X	X
Year fixed effects	X	X	X	X	X	X
Sample period	2000-2016	2000-2016	2010-2016	2007-2016	2007-2016	2010-2016
N	10,366	10,366	2,516	5,807	5,284	2,492
Adjusted R ²	0.819	0.820	0.894	0.855	0.849	0.902

Notes. Re-running all specifications in table 2 without the $f_z(x_i, t)$ polynomial term. Each data point is a county-year. Regressions are weighted by the average number of sales in a given county over all years we observe.

walk. However, Giacoletti (2017), using the same dataset as us, and Sagi (2015) using data on commercial real estate, reject the hypothesis that house prices follow a random walk; both papers show that the dispersion of returns increases with the holding period, but more slowly than the random walk model predicts. Another problem with specification (1) is that $\hat{\epsilon}_{it}^2$ will tend to be larger for houses which are sold more times, because the house fixed effect γ_i is estimated more precisely; we correct for this at the zipcode level, through the degrees-of-freedom correction term in (3), so that estimated residuals are unbiased once aggregated to the zipcode level, but residual estimates will still be biased for individual house sales.

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 10 plots kernel regressions fits of $\hat{\epsilon}_{it}^2$ 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 $sales_i$. Both distortions are visible in figure 10: the estimated logSD, $\hat{\epsilon}_{it}$, is on average higher for larger values of N , and for longer time-between-sales, although as shown in Giacoletti (2017) and Sagi (2015), the intercept of $\hat{\epsilon}_{it}$ at $tbs_i = 0$ is well above 0.

It is possible that the effects of $sales_i$ and tbs_i on $\hat{\epsilon}_{it}$ partially drive the results we observe in the main text. In order to rule out this possibility, we attempt to purge $\hat{\epsilon}_{it}^2$ of any variation which can be explained by tbs_i and $sales_i$. We do this as follows. 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 run the following regression, separately for each zipcode in our sample:

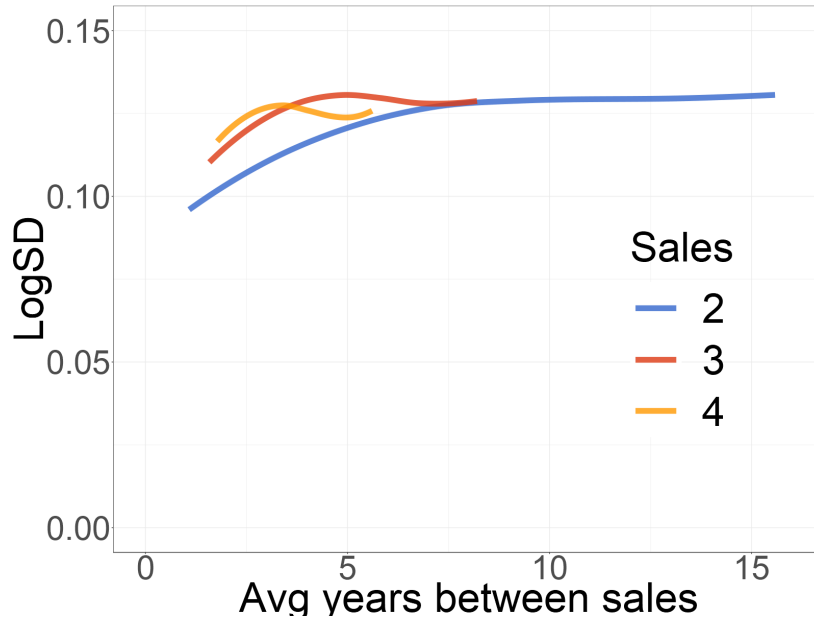
$$\hat{\epsilon}_{it}^2 = g_z(sales_i, tbs_i) + \hat{\zeta}_{it}$$

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

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

$\hat{\epsilon}_{TBSadj,it}^2$, where the subscript stands for "time-between-sales adjusted", can be interpreted as our estimate $\hat{\epsilon}_{it}^2$, nonparametrically purged of all variation explainable by a flexible but smooth function of $sales_i$ and tbs_i . We then re-run all analyses using $\hat{\epsilon}_{TBSadj,it}^2$ in place of $\hat{\epsilon}_{it}$.

Figure 10: Effect of sales_i and tbs_i on $\hat{\epsilon}_{it}^2$



Notes. Predictions from kernel regressions of $\hat{\epsilon}_{it}^2$ on tbs_i , separately for each value of sales_i.

The teal line in figure 8, labelled “LogSD, TBS adj”, shows the seasonal and time-series behavior of $\hat{\epsilon}_{TBSadj,it}^2$; we find that $\hat{\epsilon}_{TBSadj,it}^2$ tracks $\hat{\epsilon}_{it}^2$ closely seasonally and over the business cycle. Tables 13, 14, 15, and 16 re-run all zipcode and county cross-sectional regressions using $\hat{\epsilon}_{TBSadj,it}^2$ in place of $\hat{\epsilon}_{it}^2$. All coefficients have the same signs and remain significant, though most coefficients decline somewhat in magnitude.

B.4 Zillow vs Realtor.com time-on-market

For robustness, we repeat our primary county cross-sectional and panel regressions using Realtor.com time-on-market instead of Zillow. Time-on-market is significantly correlated with price dispersion in the county cross-sectional specification of column 1. However, the coefficient on time-on-market is not significant in the county panel specifications in columns 3 and 4.

Figure 8 shows yearly and monthly trends in time-on-market, measured using both Zillow and Realtor.com. The qualitative movements of time-on-market are similar, although the magnitudes and shapes of the curves differ somewhat. The difference is particularly

Table 13: Zipcode cross-sectional regressions, adjusted for number of sales and time-between-sales

	LogSD x 100						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Time on market (months)	1.727*** (0.070)				1.260*** (0.076)	2.155*** (0.091)	1.681*** (0.070)
Vacancy rate		11.785*** (0.645)			6.601*** (0.674)	2.277*** (0.585)	4.088*** (0.569)
Pop growth			-0.784 (0.690)		0.790 (0.620)	-0.377 (0.435)	-0.348 (0.480)
Mean log price					-2.950*** (0.163)	-0.337** (0.169)	-0.969*** (0.150)
Controls	X	X	X	X	X	X	X
Fixed effects						State	CBSA
N	4,109	4,109	4,109	4,109	4,109	4,109	4,109
Adjusted R ²	0.524	0.494	0.452	0.493	0.564	0.806	0.749

Notes. Re-running all specifications in table 3 using $\hat{\epsilon}_{i|zt}$ instead of $\hat{\epsilon}_{i|zt}$. Each data point is a zipcode. Regressions are weighted by the number of sales within the zipcode over the time period 2012-2016.

Table 14: Zipcode-year panel regressions, adjusted for number of sales and time-between-sales

	LogSD x 100			
	(1)	(2)	(3)	(4)
Log ZHVI	-0.751*** (0.063)			-1.695*** (0.290)
Log sales		-0.667*** (0.072)		-0.057 (0.135)
Time on market (months)			0.137** (0.059)	0.131** (0.061)
Zip fixed effects	X	X	X	X
Year fixed effects	X	X	X	X
Sample period	2000-2016	2000-2016	2013-2016	2013-2016
N	52,059	52,059	12,257	12,257
Adjusted R ²	0.882	0.884	0.926	0.926

Notes. Re-running all specifications in table 8 using $\hat{\xi}_{izt}$ instead of $\hat{\epsilon}_{izt}$. Each data point is a zipcode-year. Regressions are weighted by the average number of sales in a given zipcode over all years we observe.

Table 15: County cross-sectional regressions, adjusted for number of sales and time-between-sales

	LogSD x 100				
	(1)	(2)	(3)	(4)	(5)
Time on market (months)	0.626** (0.264)			0.385 (0.266)	1.024*** (0.305)
Vacancy rate		11.679*** (2.451)		12.414*** (3.090)	8.442*** (2.479)
Mean log price			-2.904*** (0.495)	-3.322*** (0.711)	-2.178*** (0.747)
Controls	X	X	X	X	X
Fixed effects					State
N	299	473	473	299	299
Adjusted R ²	0.418	0.426	0.440	0.479	0.734

Notes. Re-running all specifications in table 7 using $\hat{\xi}_{izt}$ instead of $\hat{\epsilon}_{izt}$. Each data point is a county. Regressions are weighted by the number of sales within the county over the time period 2012-2016.

Table 16: County-year panel regressions, adjusted for number of sales and time-between-sales

	(1)	(2)	(3)	(4)	(5)	(6)
Log ZHVI	-0.590*** (0.217)					-0.776* (0.467)
Log sales		-0.673*** (0.124)				-1.204*** (0.304)
Time on market (months)			0.285** (0.130)			0.056 (0.141)
Vacancy rate				9.333*** (1.115)		6.688*** (1.567)
Pop growth rate					-6.884*** (1.283)	-3.236 (2.472)
County fixed effects	X	X	X	X	X	X
Year fixed effects	X	X	X	X	X	X
Sample period	2000-2016	2000-2016	2010-2016	2007-2016	2007-2016	2010-2016
N	10,286	10,286	2,512	5,793	5,271	2,490
Adjusted R ²	0.886	0.888	0.923	0.912	0.909	0.927

Notes. Re-running all specifications in table 2 using $\hat{\xi}_{i,t}$ instead of $\hat{\epsilon}_{i,t}$. Each data point is a county-year. Regressions are weighted by the average number of sales in a given county over all years we observe.

noticeable at the seasonal level. Our hypothesis for these differences is that Zillow builds its dataset of house listings based on the MLS, which is the basis of Realtor.com's time-on-market measure, but updates its listings database with some lag. This would explain why Zillow time-on-market is longer than Realtor.com's, since houses disappear off Zillow

Zillow's time-on-market behaves like a lagged and smoothed version of Realtor.com time-on-market; we believe that

C Proofs and theoretical extensions

C.1 Expressions for $V_M(\epsilon)$, $\epsilon^*(v)$, $P(\epsilon, v)$

To begin with, we analytically characterize $V_M(\epsilon)$. From expression (13), we have:

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

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

$$V_M(\epsilon) = \frac{\epsilon}{r + \lambda_M} + \frac{\lambda_M}{r + \lambda_M} \int V_S(v) dF(v) \quad (26)$$

Using expression (26), we can also characterize the trade cutoff function $\epsilon^*(v)$. Trade occurs if:

$$\begin{aligned} V_M(\epsilon) &\geq V_B + V_S(v) \\ \implies \frac{\lambda_M}{r + \lambda_M} \int V_S(v) dF(v) + \frac{\epsilon}{r + \lambda_M} &\geq V_B + V_S(v) \end{aligned} \quad (27)$$

Since we have assumed in subsection 4.1.5 that $\epsilon^*(v)$ is greater than lb_ϵ , we can treat expression (27) as an equality. Solving for $\epsilon^*(v)$, we have:

$$\epsilon^*(v) = (r + \lambda_M) [V_B + V_S(v)] + \lambda_M \int V_S(v) dF(v) \quad (28)$$

Using (26) and (28) we can also characterize equilibrium prices. From (6), we have:

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

Table 17: Realtor.com vs Zillow time-on-market for counties

	LogSD x 100			
	(1)	(2)	(3)	(4)
Realtor.com time on market	-0.043 (0.193)	-0.075 (0.196)	1.372*** (0.271)	1.041 (0.633)
Log ZHVI		-2.470*** (0.772)		
Vacancy rate		5.661** (2.464)		-1.898 (6.424)
Daily list frac		31.003 (20.025)		196.011*** (68.795)
Log sales		0.526 (0.448)		
Pop growth rate		-11.302* (6.479)		
County fixed effects			X	X
Year fixed effects			X	X
Sample period	2012-2016	2012-2016	2013-2016	2013-2016
N	2,346	2,041	467	110
R ²	0.942	0.951	0.560	0.836

Notes. Each data point is a county or county-year or county.

Substituting for $V_M(\epsilon)$ using (26), we have:

$$P(\epsilon, v) = V_S(v) + \theta \left(\frac{\epsilon}{r + \lambda_M} + \frac{\lambda_M}{r + \lambda_M} \int V_S(v) dF(v) - V_B - V_S(v) \right) \quad (29)$$

Now, we can write (28) as:

$$\frac{\lambda_M}{r + \lambda_M} \int V_S(v) dF(v) - V_B - V_S(v) = \frac{\epsilon^*(v)}{r + \lambda_M} \quad (30)$$

Hence, substituting (30) into (29), we get:

$$P(\epsilon, v) = V_S(v) + \theta \left(\frac{\epsilon - \epsilon^*(v)}{r + \lambda_M} \right) \quad (31)$$

C.2 Proof of proposition 2

From (6), prices are:

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

We wish to take the variance of expression (32) with respect to the joint distribution of holding utilities v 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 $F_{tr}(v, \epsilon)$.

First, let $F_{tr}(v)$ be the marginal distribution of seller holding utilities v , among the stationary mass of seller types that trade in any given time period. By flow equality in expression (19) of proposition 1, the marginal distribution of v among sellers who trade and exit the market at any moment must be the same as the distribution of v among sellers that enter the platform; thus, we simply have:

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

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

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

for all v ; that is, the distributions of buyer match utilities, conditional on trade occurring and conditional on a given seller holding utility v . Each time a seller of holding utility v

meets a buyer, a random match quality $\epsilon \sim G(\cdot)$ is drawn; trade occurs if $\epsilon > \epsilon^*(v)$. Thus,

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

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

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

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

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

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

First, we characterize the top term on the RHS of (35). Conditional on v , the only random term in $P(\epsilon, v)$ conditional on v is the buyer's match utility ϵ ; thus, substituting expression (31) for $P(\epsilon, v)$ and ignoring constant terms, we have:

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

In words,

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

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

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

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

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

Now we move to the bottom term in expression (35). Substituting expression (31) for prices, we have:

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

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

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

Since we have assumed $G(\cdot)$ is exponential, and $\epsilon^*(v) \geq \text{lb}_\epsilon$, the term:

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

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

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

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

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

as desired.

C.3 Non-exponential match quality

Proposition 2 does not hold if we relax the assumption that match quality ϵ is exponentially distributed. Under a general match quality distribution $G(\epsilon)$, expression (36) becomes:

$$\text{Var}(\mathbf{P}(\epsilon, \nu)) = \mathbb{E}_{\nu \sim F(\nu)} \left[\text{Var}_{\epsilon \sim G(\epsilon | \epsilon > \epsilon^*(\nu))} (\mathbf{P}(\epsilon, \nu) | \nu) \right] + \text{Var}_{\nu \sim F(\nu)} \left(\mathbb{E}_{\epsilon \sim G(\epsilon | \epsilon > \epsilon^*(\nu))} [\mathbf{P}(\epsilon, \nu) | \nu] \right)$$

For the left term, since ϵ is the only random term in prices conditional on the seller holding value ν , substituting expression (29) for prices, we have:

$$\mathbb{E}_{\nu \sim F(\nu)} \left[\text{Var}_{\epsilon \sim G(\epsilon | \epsilon > \epsilon^*(\nu))} (\mathbf{P}(\epsilon, \nu) | \nu) \right] = \left(\frac{\theta}{r + \lambda_M} \right)^2 \mathbb{E}_{\nu \sim F(\nu)} \left[\text{Var}_{\epsilon \sim G(\epsilon | \epsilon > \epsilon^*(\nu))} (\epsilon) \right]$$

That is, the left term is a constant times the expectation over $\nu \sim F(\nu)$ of the variance of match utility ϵ , conditional on match utility being above the trade cutoff, $\epsilon > \epsilon^*(\nu)$. This can still be interpreted as the average variance in buyer match utility, conditional on buyer match utility being above the trade cutoff $\epsilon^*(\nu)$; however, for general distributions, this expression does not simplify further.

For the right term, all algebra steps until (40) continue to hold, so we have:

$$\text{Var}_{\nu \sim F_{tr}(\nu)} \left(\mathbb{E}_{\epsilon \sim F_{tr}(\epsilon | \nu)} [\mathbf{P}(\epsilon, \nu) | \nu] \right) = \text{Var}_{\nu \sim F(\nu)} \left(V_S(\nu) + \frac{\theta}{r + \lambda_M} \mathbb{E}_{\epsilon \sim G(\epsilon | \epsilon > \epsilon^*(\nu))} [\epsilon - \epsilon^*(\nu)] \right) \quad (43)$$

This can be interpreted as the variance of seller values $V_S(\nu)$, plus a term

$$\frac{\theta}{r + \lambda_M} \mathbb{E}_{\epsilon \sim G(\epsilon | \epsilon > \epsilon^*(\nu))} [\epsilon - \epsilon^*(\nu)]$$

which can be thought of as the average of the “markup” term $\epsilon - \epsilon^*(\nu)$, conditional on ϵ being above the trade cutoff $\epsilon^*(\nu)$. For general distributions $G(\epsilon)$, the markup depends on $\epsilon^*(\nu)$ and thus ν , thus the markup term cannot be moved outside the variance expression, and (43) does not simplify further. Hence, an analog of proposition 2 which

holds for arbitrary $G(\epsilon)$ is:

$$\text{Var}(P(\epsilon, v)) = \underbrace{\left(\frac{\theta}{r + \lambda_M}\right)^2 \mathbb{E}_{v \sim F(v)} \left[\text{Var}_{\epsilon \sim G(\epsilon | \epsilon > \epsilon^*(v))}(\epsilon) \right]}_{\text{Buyer match utility}} + \underbrace{\text{Var}_{v \sim F(v)} \left(V_S(v) + \frac{\theta}{r + \lambda_M} \mathbb{E}_{\epsilon \sim G(\epsilon | \epsilon > \epsilon^*(v))}[\epsilon - \epsilon^*(v)] \right)}_{\text{Seller value plus markup}} \quad (44)$$

Expression (44) also shows why the assumptions of persistent heterogeneous seller values is needed for the results of the paper to hold. If sellers' holding values were homogeneous, price variance would only consist of the buyer match quality term in (44), which is the expectation over v of:

$$\text{Var}_{\epsilon \sim G(\epsilon | \epsilon > \epsilon^*(v))}(\epsilon) \quad (45)$$

This is the variance of ϵ , conditional on $\epsilon > \epsilon^*(v)$. For exponential distributions, (45) is constant in $\epsilon^*(v)$; for general distributions (45) can either increase or decrease as $\epsilon^*(v)$ increases.

C.4 Proof of proposition 3

From expression (15) in proposition 1, the seller value function $V_S(c)$ is:

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

Differentiating with respect to v , we have:

$$rV'_S(v) = 1 + \lambda_S \int \theta (-V'_S(v)) 1(\epsilon > \epsilon^*(v)) dG(\epsilon)$$

Computing the integral, this becomes:

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

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

$$V'_S(v) = \frac{1}{r + \theta \lambda_S (1 - G(\epsilon^*(v)))} \quad (46)$$

Substituting expression (23) for $\text{TOM}(v)$ in the denominator of (46), we have:

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

Rearranging, we have (24).

C.5 Proof of proposition 4

From expression (24), if $\text{TOM}_{\Theta_1}(v) > \text{TOM}_{\Theta_2}(v)$ for all v , and if r, θ are the same in the two sets of primitives, then $V'_S(v)$ must also be strictly larger in absolute value, pointwise in v , under parameters Θ_1 than Θ_2 , for all v . From expression (22) in claim 2, we have:

$$\text{Var}(P(\epsilon, v)) = \underbrace{\text{Var}_{v \sim F(\cdot)}(V_S(v))}_{\text{Sellers' value}} + \underbrace{\left(\frac{\theta}{r + \lambda_M}\right)^2 \sigma_\epsilon^2}_{\text{Buyers' value}}$$

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

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

$$\text{Var}(X) = \min_{\bar{x}} \int (x - \bar{x})^2 dG(x) \quad (47)$$

To prove expression (47), note that:

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

Thus,

$$\min_{\bar{x}} \int (x - \bar{x})^2 dG(x) = \min_{\bar{x}} \int (x - E(x))^2 dG(x) + \int (E(x) - \bar{x})^2 dG(x)$$

$$= \int (x - E(x))^2 dG(x) = \text{Var}(X)$$

Now, call the distribution of $V_S(v)$ among trading sellers $F_{V_S}(V)$. Using expression (47), we can write the variance of $V_S(v)$ as:

$$\text{Var}(V_S(v)) = \min_{\bar{x}} \int (x - \bar{x})^2 dF_{V_S}(V) \quad (48)$$

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

$$dF_{V_S}(V) = V_S'(v) dF(v)$$

Hence, (48) becomes:

$$= \min_{\bar{x}} \int (x - \bar{x})^2 V_S'(v) dF(v) \quad (49)$$

A uniform increase in $V_S'(c)$ causes the integral in (49) to strictly increase for any \bar{x} . Thus, if $V_S'(v)$ is uniformly higher under Θ_1 than Θ_2 , then $\text{Var}(V_S(v))$ must also increase, and thus $\text{Var}(P(\epsilon, v))$ must also increase.

C.6 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 utility v , buyers' urgency u , and buyers' match utility ϵ :

$$P(v, u, \epsilon) = V_S(v) + \theta (V_M(\epsilon) - V_B(u) - V_S(v)) \quad (50)$$

Thus, the match quality cutoff condition becomes:

$$\epsilon^*(v, u) = \{\epsilon : V_M(\epsilon) \geq V_B(u) + V_S(v)\}$$

Analogous to subsection 4.1.5, for our theoretical results, we will need to assume that

$$\epsilon^*(v, u) > \epsilon_0 \quad \forall v, u$$

Buyers' and sellers' value functions become, respectively:

$$\begin{aligned} rV_B(\mathbf{u}) &= \lambda_B \int_v \int_{\epsilon > \epsilon^*(v, \mathbf{u})} [(1 - \theta)(V_M(\epsilon) - V_B(\mathbf{u}) - V_S(v))] dG(\epsilon) dF_{eq}(v) \\ rV_S(v) &= v + \lambda_S \int_u \int_{\epsilon > \epsilon^*(v, \mathbf{u})} [\theta(V_M(\epsilon) - V_B(\mathbf{u}) - V_S(v))] dG(\epsilon) dH_{eq}(\mathbf{u}) \end{aligned} \quad (51)$$

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

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

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

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

Somewhat surprisingly, despite these changes to stationary equilibrium conditions, propositions 2, 3, and 4 continue to hold. To prove proposition 2, note that, when buyers have heterogeneous values, the matched owner value function is unchanged:

$$V_M(\epsilon) = \frac{\epsilon}{r + \lambda_M} + \frac{\lambda_M}{r + \lambda_M} \int V_S(v) dF(v) \quad (52)$$

The derivations in appendix C.1 thus imply that:

$$P(v, \mathbf{u}, \epsilon) = V_S(v) + \theta \left(\frac{\epsilon - \epsilon^*(v, \mathbf{u})}{r + \lambda_M} \right) \quad (53)$$

Now, similar to (35), we take the variance of prices, applying the law of iterated expectations with respect to v and \mathbf{u} , to get:

$$\begin{aligned} \text{Var}(P(v, \mathbf{u}, \epsilon)) &= E_{v, \mathbf{u} \sim F_{tr}(v, \mathbf{u})} \left[\text{Var}_{\epsilon \sim G_{tr}(\epsilon|v, \mathbf{u})} (P(v, \mathbf{u}, \epsilon) | v, \mathbf{u}) \right] + \\ &\quad \text{Var}_{v, \mathbf{u} \sim F_{tr}(v, \mathbf{u})} \left(E_{\epsilon \sim G_{tr}(\epsilon|v, \mathbf{u})} [P(v, \mathbf{u}, \epsilon) | v, \mathbf{u}] \right) \end{aligned} \quad (54)$$

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

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

The joint distribution $F_{tr}(v, u)$ is more complicated to characterize; however, by flow equality, the marginal distributions of $F_{tr}(v, u)$ must be equal to the distributions of entering buyer and seller types, $F(v)$ and $H(u)$. This implies that the following steps in appendix C.2 go through essentially unchanged. Going through the steps, for the top term of (54), we have:

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

For the bottom term, substituting expression (53) for prices, we have:

$$\begin{aligned} \text{Var}_{v, u \sim F_{tr}(v, u)} \left(\mathbb{E}_{\epsilon \sim G(\epsilon | \epsilon > \epsilon^*(v, u))} [P(v, u, \epsilon) | v, u] \right) &= \\ = \text{Var}_{v, u \sim F_{tr}(v, u)} \left(\mathbb{E}_{\epsilon \sim G(\epsilon | \epsilon > \epsilon^*(v, u))} \left[V_S(v) + \theta \left(\frac{\epsilon - \epsilon^*(v, u)}{r + \lambda_M} \right) | v, u \right] \right) &= \\ = \text{Var}_{v, u \sim F_{tr}(v, u)} \left(V_S(v) + \frac{\theta}{r + \lambda_M} \mathbb{E}_{\epsilon \sim G(\epsilon | \epsilon > \epsilon^*(v, u))} [\epsilon - \epsilon^*(v, u) | v, u] \right) & \quad (55) \end{aligned}$$

Again, the left term of (55),

$$\mathbb{E}_{\epsilon \sim G(\epsilon | \epsilon > \epsilon^*(v, u))} [\epsilon - \epsilon^*(v, u) | v, u]$$

is equal to σ_ϵ , which is independent of v, u , so we can ignore it in the variance calculation; (55) thus simplifies to

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

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

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

This proves proposition 2. Similarly, differentiating (51), we have:

$$rV'_S(v) = 1 + \lambda_S \int_u \int_{\epsilon > \epsilon^*(v,u)} (-V'_S(v)) \mathbf{1}(\epsilon > \epsilon^*(v)) dG(\epsilon) \quad (56)$$

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

$$\text{TOM}(v) = \frac{1}{\lambda_S \int_u \int_{\epsilon > \epsilon^*(v,u)} \theta \mathbf{1}(\epsilon > \epsilon^*(v)) dG(\epsilon)} \quad (57)$$

Combining (56) and (57), we have:

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

proving proposition 3. Finally, the proof of proposition 4 follows unchanged from the argument of appendix C.5.

C.7 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 4. Claim 2 characterized price dispersion, that is, the variance of prices $P(\epsilon, v)$ among trading buyers and sellers:

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

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:

$$M_S \lambda_S \int [1 - G(\epsilon^*(v))] dF_{\text{eq}}(v)$$

Intuitively, this is the equilibrium mass of sellers M_S , multiplied by the equilibrium seller matching rate λ_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 $1 - G(\epsilon^*(v))$, over the equilibrium distribution of seller holding utilities, $F_{\text{eq}}(v)$.

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

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

This is simply the average of time on market for a seller of holding utility v over the distribution of holding utilities $F(c)$; note that we showed in appendix C.2 above that the distribution of holding utilities among trading sellers is simply $F_{\text{tr}}(v) = F(v)$.

The average transaction price conditional on trade is:

$$\int \int P(\epsilon, v) dG(\epsilon | \epsilon > \epsilon^*(v)) dF(v)$$

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

C.8 Comparative statics

To create figure 4, we solve the model under the following assumptions on parameters:

$$r = 0.0526, \phi = 0.84, \theta = 0.5, \epsilon_0 = 1.7, \sigma_\epsilon = \frac{1}{5}, \alpha = 1$$

$$M_B = 0.67, \lambda_M = 0.035, \bar{v} = 0.7, \Delta_v = 0.7$$

Parameters are chosen to loosely match our calibration parameters, except that we have increased Δ_v somewhat so that the effect of sellers' holding costs on price dispersion in figure 4 is more clear. For the comparative statics of figure 4, we vary each parameter around its target value while holding all other variables fixed.