SPECULATIVE DYNAMICS OF PRICES AND VOLUME

Anthony A. DeFusco
Charles G. Nathanson
Eric Zwick

Speculative Dynamics of Prices and Volume
Anthony A. DeFusco, Charles G. Nathanson, and Eric Zwick
NBER Working Paper No. 23449
May 2017, Revised July 2020
JEL No. E32,G02,G12,R3

ABSTRACT

Using data on 50 million home sales from the recent U.S. housing cycle, we document that much of the variation in volume came from the rise and fall in short-term speculation. Cities with larger speculative booms have larger price cycles, sharper increases in unsold listings as the market turns, and more eventual foreclosures. We present a model in which predictable price increases endogenously attract short-term buyers more than long-term buyers. Short-term buyers amplify volume by selling faster and destabilize prices through positive feedback. Our model matches key aggregate patterns, including the lead{lag price{volume relation and a sharp rise in inventories.

Anthony A. DeFusco
Northwestern University
Kellogg School of Management
2211 Campus Drive
Office 4463
Evanston, Illinois 60208
anthonydefusco@gmail.com

Charles G. Nathanson
Kellogg School of Management
Northwestern University
2211 Campus Drive, Room 4479
Evanston, IL 60208-0898
nathanson@kellogg.northwestern.edu

Eric Zwick
Booth School of Business
University of Chicago
5807 South Woodlawn Avenue
Chicago, IL 60637
and NBER
ezwick@chicagobooth.edu

The United States underwent an enormous housing market cycle between 2000 and 2011 (Figure 1). The rise and fall in house prices caused several problems for the U.S. economy. During the boom, a surge in housing investment drew resources into construction from other sectors (Charles, Hurst and Notowidigdo, 2018) and contributed to a capital overhang that slowed the economic recovery from the subsequent recession (Rognlie, Shleifer and Simsek, 2017). During the bust, millions of households lost their homes in foreclosure, and falling house prices led many others to cut consumption (Mayer, Pence and Sherlund, 2009; Mian, Rao and Sufi, 2013; Mian, Sufi and Trebbi, 2015; Guren and McQuade, 2020). Large real estate cycles are not unique to the U.S. (Mayer, 2011) or to this time period (Case, 2008; Glaeser, 2013). Given the economic costs of these recurring episodes, understanding their cause is critical for economists and policymakers.

This paper presents evidence that speculation was a key driver of this real estate cycle.[1] Three stylized facts from the cycle guide our analysis. First, prices and volume jointly rise and fall through the cycle. Second, volume falls before prices, resulting in a pronounced lead–lag relation between prices and volume. Third, the period during which prices continue to rise despite falling volume coincides with rapidly accumulating unsold listings. We refer to this period as the *quiet*, which is preceded by the *boom* and followed by the *bust*. These stylized facts hold on average across cities and are especially pronounced in cities with larger cycles. They suggest that focusing on *who* was most active during each phase of the cycle can shed light on the underlying mechanisms.

We study the behavior of speculative homebuyers during each phase of the housing cycle using transaction-level data collected by CoreLogic on 50 million home sales between 1995 and 2011. We measure speculative buying and selling over time and across 115 metropolitan statistical areas (MSAs), which together represent 48% of the U.S. housing stock. We pursue two complementary approaches to identifying speculative activity. First, following Bayer, Geissler, Mangum and Roberts (2020), we classify transactions based on their realized holding periods, denoting those buyers who resell the property within three years as short-term buyers. Second, following Chinco and Mayer (2015), we classify transactions based on the inferred occupancy status of the property, denoting buyers who list a mailing address distinct from the property address as non-occupant buyers. We supplement our transaction data

---

[1]Harrison and Kreps (1978, p. 323) define speculation as follows: "Investors exhibit speculative behavior if the right to resell a stock makes them willing to pay more for it than they would pay if obliged to hold it forever."

with a separate CoreLogic data set on homes listed for sale, sourced from a consortium of local MLS boards. We link these data to transaction records to study the role of speculative buyers for inventory dynamics across MSAs.

While overall volume increases substantially during the boom years of 2000–2005, both short-term and non-occupant volume rise dramatically more. In an accounting sense, growth in speculative volume explains between 40% and 50% of total volume growth. This relation is also strong in the cross-section, as speculative volume growth can account for 30% to 50% of total volume growth across MSAs. Cities with stronger speculative volume booms also experience larger house price booms: MSAs with a one standard deviation larger short-volume and non-occupant boom see 25 and 15 percentage point larger cumulative price increases, respectively.

As the volume boom ends, price growth remains positive but slows, and unsold listings accumulate. Across MSAs, these patterns are more pronounced in cities with larger speculative volume booms. Our linked listing-transaction data further reveal that short-term buyers disproportionately contribute to the surge in aggregate inventories. MSAs with larger speculative volume booms also see substantially larger price busts, volume busts, and total foreclosures in the final phase of the cycle.

Our results suggest that the differential entry of speculative buyers plays a central amplifying role in the cycle. In robustness analysis, we consider and rule out several alternative explanations of the rise in speculative volume. These alternatives include move-up purchases due to rising home equity, the entry of professional real estate arbitrageurs, and various mechanical concerns arising from the way we measure short-term volume. Together, these analyses lend support to our preferred interpretation, which focuses on a class of inexperienced speculative entrants into the housing market during the boom.

Consistent with our interpretation of the data, a National Association of Realtors survey reveals wide variation in expected holding times, shorter expected holding times among investors, and increases in the short-term buyer share following recent price gains. We confirm the statistical link between house price changes and speculative buyer entry using monthly data and a panel VAR specification as in Chinco and Mayer (2015). Short-buyer entry is strongly predictive of subsequent house price growth and predicted by recent past price growth, whereas non-occupant entry can be predicted by past price growth but is less informative for predicting subsequent prices. In lower-frequency horserace specifications,

short-term volume also tends to be a stronger predictor of cycle dynamics than the non-occupant boom. At the same time, non-occupant buyers disproportionately contribute to the growth in short-term volume, indicating significant overlap between these categories. One interpretation of these results is that short-term volume is a more precise measure of speculative activity than non-occupant volume, perhaps because non-occupant volume also includes longer-term "cash flow" investors and vacation homebuyers.

The last part of the paper presents a model that accounts for the evidence and allows us to quantify the relative contribution of different types of speculators to the cycle. Our approach adapts core insights from Cutler, Poterba and Summers (1990), De Long, Shleifer, Summers and Waldmann (1990), and Hong and Stein (1999) to study the housing market.[2] As in these papers, extrapolation—the belief that prices continue to rise after recent gains— causes a predictable boom and bust in house prices after a positive demand shock. In contrast to those papers, we relax the assumption of Walrasian market clearing, so that homes listed for sale may not sell immediately. To do so, we microfound extrapolation using the approach in Glaeser and Nathanson (2017) and then extend their framework to a non-Walrasian setting.

In our model, a mover attempts to sell her house by posting a list price. A potential buyer arrives and decides whether to purchase the house at that price. Potential buyers differ in the benefits they derive from owning a house; non-occupants benefit less than occupants. Buyers also differ in the expected amount of time until becoming a mover; short-term buyers have shorter horizons ex ante. The average flow benefit of potential buyers fluctuates randomly over time. Agents cannot observe this demand process, but they can observe the history of price growth and the share of listings that sell each period. Using this market data, agents infer the current level and growth rate of the demand process and optimally make decisions in light of these beliefs—the choice of list price for movers, and whether or not to purchase for potential buyers. As in Glaeser and Nathanson (2017), agents mistakenly believe that potential buyers neglect time-variation in the growth rate when deciding whether to buy.

We study how our housing market responds to a large, unexpected increase to the growth rate of the demand process. We choose parameter values to match facts about the housing market, including the boom–bust cycle in prices and volume and the baseline speculative

---

[2]Section 6 motivates our model by reviewing related theoretical work. We also highlight the aspects of our empirical results that prior work can and cannot explain.

share of volume. In the model, the quiet occurs when agents overestimate the level of the demand process and believe it continues to grow. This mistaken belief causes movers to increase their list prices despite falling transaction volume.

We then use this setting to explore the relative importance of short-term and non-occupant volume, because the model allows us to separate a buyer's horizon from the utility she receives from buying a house. Much of the rise in volume comes from non-occupant purchases and short-term sales because speculators disproportionately buy housing as prices rise. In a counterfactual without short-term potential buyers, the price bust nearly disappears. The same holds in a counterfactual without non-occupants, but only because many non-occupants have short horizons. Eliminating non-occupants while keeping the horizon distribution constant fails to attenuate the housing cycle. These results suggest that short-term speculation causes the house price cycle in the model.

Previous work has examined short-term buyers (Adelino, Schoar and Severino, 2016; Bayer, Geissler, Mangum and Roberts, 2020; Bayer, Mangum and Roberts, 2016) and non-occupant buyers (Haughwout, Lee, Tracy and van der Klaauw, 2011; Bhutta, 2015; Gao, Sockin and Xiong, 2019; Chinco and Mayer, 2015) during this cycle. We contribute to this empirical literature in four ways. First, unlike many studies, we use deeds records instead of mortgage records, allowing us to observe speculation among all-cash buyers. Because all-cash purchases disproportionately come from speculators and constitute a large share of total sales, research relying on mortgage records likely undercounts speculation. Second, the number of MSAs in our sample—115—is considerably larger than in some other studies, allowing us to establish cross-MSA relations between speculation and other aspects of the housing cycle. Third, we introduce new microdata on homes listed for sale that allow us to study the joint dynamics of prices, volume, and inventories in the cross-section of cities, and document the role of recent buyers in driving the surge of listings during the quiet. Finally, we relate the price cycle to both types of speculation simultaneously, whereas the previous literature has tended to look at only one type. We find substantial overlap between the two types and, interestingly, a more robust relation of the price cycle to short-term than non-occupant buying. Our model sheds light on this result.

4

1 Dynamics of Prices, Volume, and Inventory

In this paper, we present evidence and a model showing that short-term speculation was a central amplifying force of the last U.S. housing cycle. This section presents three stylized facts from that cycle that guide our analysis. First, prices and volume jointly rise and fall through the cycle. Second, volume falls before prices, resulting in a pronounced lead–lag relation between prices and volume. Third, the period during which prices continue to rise despite falling volume coincides with rapidly accumulating unsold listings. We refer to this period as the *quiet*, which is preceded by the *boom* and followed by the *bust*.

Figure 1, Panel A, plots aggregate trends in prices and transaction volume between 2000 and 2011. Panels B through E plot analogous series for four cities that represent regions with the largest boom–bust cycles during this time: Phoenix, AZ; Las Vegas, NV; Orlando, FL; and Bakersfield, CA. During the 2000s housing cycle, volume peaks before prices, and there is a sustained period during which volume is falling rapidly on high prices. This dynamic holds consistently across regions that experienced large price cycles. At the aggregate level, volume rises to 150% of its level in 2000 and then falls back to this level before prices begin to fall. In the four cities in Panels B through E, volume more than doubles during the boom. Prices subsequently peak between 200% and 300% of their 2000 levels.

Figure 2 shows that this lead–lag relation between prices and volume also holds on average across all MSAs in our sample. We search for the horizon over which a given change in volume has the most predictive power for the contemporaneous change in prices at the MSA level. Changes in volume generally lead changes in prices if the correlation between prices and volume is maximized at a positive lag.

To implement this search, we build a monthly panel of log house prices and transaction volume at the MSA level running from January 2000 to December 2011.[3] We normalize transaction volume in each MSA-month by dividing by the total housing stock for the MSA recorded in the 2000 Census. Using this panel, we run a series of simple regressions of the form:

$$p_{i,t} = \beta_\tau v_{i,t-\tau} + \eta_{i,t}, \tag{1}$$

where p is log price, v is volume, i indexes MSAs, and time is measured in months. To

---

[3]The data used to construct this panel and the sample restrictions we impose are discussed in detail in Section 2.1 below.

account for the seasonal adjustment in the CoreLogic price indices, for each regression we demean prices at the MSA level and demean volume at the MSA–calendar month level.[4]

The coefficient β_τ provides an estimate of how movements in volume around MSA–calendar month averages at a τ-month lag are correlated with contemporaneous movements in prices around MSA averages. We run these regressions separately for up to 4 years of lags ($\tau = 48$) and one year of leads ($\tau = -12$). Figure 2, Panel A, plots the implied correlation from each regression along with its 95% confidence interval. The correlation is positive at most leads and lags but reaches its maximum at a positive lag of 24 months. Thus, changes in volume generally lead changes in prices by about two years.

Figure 3, Panel A, plots aggregate trends in prices and inventories of homes listed for sale between 2000 and 2011. Panels B through E plot analogous series for four cities that represent the same regions as in Figure 1.[5] During the period when the relation between volume and prices reverses, aggregate inventories rise dramatically to nearly double their level during the early years of the cycle. This pattern also characterizes the joint dynamic of prices and inventories across cities in Panels B through E. In Phoenix, Reno, and Bakersfield, inventories rise during the quiet to between double and triple their levels during the boom. In Daytona Beach, inventories rise to 450% of their pre-quiet levels.

These stylized facts suggest that focusing on the dynamic of quantities—both volume and inventories—can shed light on the drivers of the cycle. In particular, determining *who* was most heavily participating in the housing market during each phase of the cycle may differentiate between various explanations for that cycle.

2   Data

The primary goal of our empirical analysis is to study the behavior of speculative home buyers during each phase of the housing cycle. This section describes our data and how we identify speculative buyers. Further information is in Internet Appendix I.

---

[4]For other work regressing house prices on lagged transaction volume, see Leung, Lau and Leong (2002), Clayton, Miller and Peng (2010), and Head, Lloyd-Ellis and Sun (2014).

[5]Data on unsold inventory is unavailable for Las Vegas, NV and Orlando, FL. Because of this, Figure 3, Panels C and D, use data from Reno, NV and Daytona Beach, FL instead. We plot aggregate inventories from the NAR, which are available starting in 2000. Our MSA-level inventory data are available for these cities starting in 2001.

2.1 Data Sources and Sample Selection

We use data on individual housing transactions provided by CoreLogic, a private vendor that collects and standardizes publicly available tax assessments and deeds records from municipalities across the U.S. Our main analysis data span the years 1995 through 2014 and include observations from 115 MSAs, which together represent 48% of the U.S. housing stock. In analyses that require us to identify an owner's occupancy status we use a smaller subset of 102 MSAs for which we can be sure that there were no major changes in the way that mailing addresses were coded during our sample period. Appendix I describes the approach we use to select these MSAs. Our analysis of the housing cycle covers the time period 2000 through 2011 because measuring realized holding periods requires observing consecutive transactions.

We include all transactions of single-family homes, condos, or duplexes that satisfy the following filters: (a) the transaction is categorized by CoreLogic as occurring at arm's length, (b) there is a nonzero transaction price, and (c) the transaction is not coded by CoreLogic as being a nominal transfer of title between lenders following a foreclosure. We then drop a small number of duplicate transactions where the same property is observed to sell multiple times at the same price on the same day or where multiple transactions occur between the same buyer and seller at the same price on the same day. Internet Appendix I specifies the exact steps followed to arrive at a final sample of 51,080,640 transactions. Given the geographic coverage of these data and their source in administrative records, our analysis sample serves as a proxy for the population of transactions in the U.S. during the sample period.

In addition to this transaction-level data, we use data on the listing behavior of individual homeowners. Our listings data is also provided by CoreLogic and is sourced from a consortium of local Multiple Listing Service (MLS) boards throughout the country. For each listing, we observe the date the home was originally offered for sale, an indicator for whether the listing ever sold, and the date of sale for those that did. We link these data to the deeds data using the assessor's parcel number (APN) for the property. When analyzing listings, we focus our attention on a subset of the 115 MSAs for which we can be relatively certain that the listings data is representative of the majority of owner-occupied home sales in the area. Internet Appendix I describes in detail the approach we use to select these MSAs, leaving us with a final sample of 57 MSAs for our listings analysis.

We supplement these transaction- and listing-level data with national and MSA-level

housing stock counts from the U.S. Census, national counts of sales and listings of existing homes from the National Association of Realtors (NAR), and national, MSA, and ZIP code-level nominal house-price indices from CoreLogic.

We also use survey data to study heterogeneity in expected holding horizons in the cross-section and over time. Each March, as part of the Investment and Vacation Home Buyers Survey, the NAR surveys a nationally representative sample of around 2,000 individuals who purchased a home in the previous year. The survey asks respondents to report the type of home purchased (investment property, primary residence, or vacation property) as well as the "length of time [the] buyer plans to own [the] property." Data on expected holding times and the share of purchases of each type are available between 2008 and 2015.

2.2   Identifying Speculators

We identify speculators using two complementary approaches, each of which has been used in prior work on the U.S. housing boom. In the first approach, we categorize transactions based on the their realized holding periods. We denote transactions held for less than 3 years as "short-term" sales and track the evolution of these sales over time. This approach follows Bayer, Geissler, Mangum and Roberts (2020) who classify speculators in a similar way based on the argument that those holding homes for short time periods are more likely to have purchased those homes for investment purposes.

One potential concern with this classification is that holding periods are not fixed at the time of purchase. Thus, changes in the distribution of realized holding periods over the course of a housing cycle could be driven not by differential entry and exit of speculative buyers, but rather by endogenous changes in holding periods at the individual level and mechanical changes in underlying market liquidity. We address this concern in several ways in our analysis. Our main strategy, however, simply uses an alternative approach to classifying speculators that does not suffer from this limitation.

Our second approach classifies homebuyers based on their occupancy status. Those who purchase a home without the intent to occupy it immediately are more "speculative" in the sense that a larger portion of their overall expected return is derived from capital gains rather than from the consumption value of living in the home. To identify these buyers, we follow Chinco and Mayer (2015) and mark buyers as non-occupants when the transaction lists the buyer's mailing address as distinct from the property address. While this proxy may

8

misclassify some non-occupants as living in the home if they choose to list the property's address for property-tax-collection purposes, we believe it to be a useful gauge of the level of non-occupant purchases. Moreover, this measure of speculation does not suffer from the same issue as our short-term buyer measure since it is based only on characteristics of the buyer that are fixed at the time of purchase.

One key advantage of both methods we use to identify speculators is that they are based on the full sample of housing transactions. Other work has identified speculators based on the presence of multiple first-lien mortgage records in credit reporting data or self-reported occupancy status on loan applications (Haughwout, Lee, Tracy and van der Klaauw, 2011; Gao, Sockin and Xiong, 2019; Mian and Sufi, 2019). While based on similar ideas, such approaches run the risk of omitting a substantial fraction of speculative activity.

Table 1 demonstrates this point using summary statistics on the proportion of all-cash purchases in our data. Column 1 shows that in our sample, 29 percent of short-term buyers and 38 percent of non-occupant buyers did not use a mortgage when purchasing their property. These shares exceed the all-cash share among all buyers, which is 20 percent, suggesting that mortgage-based measures of speculation may differentially underrepresent speculative activity. The remaining columns of the table, which report averages at the MSA-by-month level, show that the role of all-cash transactions among buyers we identify as speculative remains high at all points in the housing cycle.[6] The behavior of these buyers would go unobserved in any analysis of speculative activity based on mortgage data alone.

3 Speculators During the Boom

3.1 Quantities and Prices

Figure 4 presents a simple illustration of the quantitative importance of speculative activity during the 2000–2011 U.S. housing cycle. The figure plots monthly aggregate time series for total transaction volume (with and without new construction), short-holding-period volume, and non-occupant volume calculated using our underlying sample of CoreLogic deed transfers. Each series is separately normalized relative to its average value in the year 2000 and

---

[6]Studying the role of speculators during the recovery from the crash is not a central focus of our paper. Nevertheless, it is interesting to note that the all-cash share rises to 50 percent of speculative purchases during the bust.

seasonally adjusted by removing calendar-month fixed effects. For reference, the raw counts of each type of transaction in the years 2000, 2005, and 2010 are also reported in the upper right corner of the figure. To abstract from the effect of foreclosures on speculative volume during the bust, we exclude foreclosures from the series in this figure.

While overall volume increased by roughly 40% during the boom years of 2000–2005, speculative volume increased dramatically more. Both short-term sales and purchases by non-occupants approximately doubled between 2000 and 2005. Not only did these speculative components of volume increase more rapidly, but their increase also accounted for a non-trivial portion of the overall increase in volume during this period. For example, total volume increased from 2.73 million transactions in 2000 to 3.82 million in 2005. During the same time period, short-holding-period volume increased from 510 to 940 thousand transactions, which implies that volume growth in this category alone can account for 39%(= 0.43/1.09) of the *total* volume increase during the boom. A similar calculation for non-occupant volume (in the 102 MSAs with reliable non-occupant data) implies that this measure of speculative activity can account for 53%(= 0.52/0.98) of the volume increase in the boom. If we exclude new construction from the total volume statistics—because short-term sales can only involve homes previously sold—short-term volume accounts for 57%(= 0.43/0.75) of the aggregate increase in existing home sales. These calculations illustrate that speculators were, in an accounting sense, a key driver of the volume boom.

The shift in the composition of volume toward speculative buyers also correlates highly with changes in total volume across local markets. This correlation can be seen in the top two panels of Figure 5. Panel A presents scatter plots of the percent change in total volume at the MSA-level from 2000–2005 versus the percent change in volume for short holding periods and long holding periods separately. Not only does the growth in volume of short-holding-period transactions correlate strongly with the increase in total volume across MSAs, but the magnitude of this relation is also much stronger for short holding periods relative to long holding periods. A similar conclusion arises from Panel B, which presents analogous scatter plots grouping transactions according to the occupancy status of the buyer rather than the holding period of the seller. The relation between total volume growth and non-occupant volume growth across MSAs is strong, positive, and larger in magnitude than the corresponding relation with growth in sales to owner-occupants.

Panels C and D further show that these cross-sectional differences in the growth rate

of speculative volume explain a significant portion of the differences in the growth in total volume across MSAs. For each MSA, we plot the change in either short-holding-period volume (Panel C) or non-occupant volume (Panel D) divided by initial total volume on the y-axis against the percent change in total volume on the x-axis. The slope of this line provides an estimate of how much of a given increase in total volume during this period came in the form of short-holding-period or non-occupant volume. For short-holding-period volume, the answer is 30%.[7] For non-occupant volume the slope is even larger and implies that for the average MSA in our sample 54% of the increase in total volume between 2000 and 2005 came from non-occupant purchases. Thus, shifts in the composition of volume toward speculative buyers appear to have been a major determinant of changes in total transaction volume during the boom.

Table 2 shows how speculative volume relates to measures of the size of the price and quantity cycles in the cross-section of MSAs. We estimate the correlation between growth in each speculative measure and housing market outcomes and perform a horserace analysis that regresses these outcomes on both measures of speculative activity. To aid interpretation of these relations, we scale the change in outcomes for all quantity measures relative to total volume in 2003. The regressions do not annualize changes, so we report annualized coefficients separately in the table. Appendix Table IA.II reports summary statistics.

We focus here on the house price boom (Panel A, columns 1-3) and return to the other outcomes in Section 4. House price booms are strongly related to the size of speculative volume booms across cities. Cities with a one standard deviation larger short-volume boom (12.9%) see a 24.9 percentage point larger cumulative price increase during the boom.[8] Cities with a one standard deviation larger non-occupant boom (27.1%) see a 15.4 percentage point larger cumulative price increase during the boom. On average across cities, prices rise by 97% during the boom and quiet. Thus, the relation between speculative volume and prices is economically large in the cross-section. Notably, in the horserace specification, the short-volume boom retains a strong positive association while the non-occupant boom reverses sign. We return to this result below.

To further investigate the link between house price changes and speculative buyer entry,

---

[7]If we exclude new construction from total volume the 30% figure rises slightly to 36%. This relation is indicated in the figure by the hollow squares in Panel C.

[8]Adelino, Schoar and Severino (2016) and Bayer, Geissler, Mangum and Roberts (2020) also document positive relations between short-term buying and price booms across regions during this period.

we examine higher frequency data. Speculative buyers may both cause and respond to house price changes. Because of the potential for this type of strong feedback mechanism, we do not attempt to directly identify the "causal" effect of speculators on house prices.[9] Instead, we follow the approach in Chinco and Mayer (2015), who estimate predictive regressions that are flexible enough to allow for some types of feedback between speculative entry and house prices. In particular, we estimate a series of panel vector auto-regressions (pVARs) describing the relation between house price growth and the share of purchases made by non-occupant buyers and "short buyers" (i.e., those who will sell within three years of purchase) at a monthly frequency in each MSA between January 2000 and December 2006 (the year when prices peaked).

Table 3 reports results from three different pVAR specifications. In column 1, we estimate a simple two-equation model that jointly links both month-over-month house price growth to the lagged share of transactions by short-buyers (top panel) and the contemporaneous short-buyer share to lagged house price appreciation (middle panel). Both equations also include lags of the relevant dependent variable (house price appreciation in the top panel and the short-buyer share in the middle panel).

The results indicate that a 1 percentage point increase in the fraction of purchases made by short-term buyers in a given month is associated with a 0.02 percentage point increase in the house-price appreciation rate in the following month. That is, short-buyer entry is predictive of subsequent house price growth, though we stress that these predictive regressions do not necessarily imply a causal relation. Interestingly, the results in the middle panel indicate that short-buyer entry can also be predicted by recent house price growth. A 1 percentage point increase in house price growth in the prior month is associated with a 0.16 percentage point increase in the short-buyer share of entrants.

In column 2, we estimate a similar model swapping out the short-buyer share for the share of purchases made by non-occupants. Unlike short-buyer entry, non-occupant entry does not appear to be predictive for subsequent house price growth. The coefficient on the lagged non-occupant share in the top panel is roughly half the magnitude of its short-buyer analog from column 1 and is not statistically significant. Non-occupants do, however, appear to respond similarly to past price growth. The estimate in the bottom panel indicates that

---

[9]Gao, Sockin and Xiong (2019) exploit state capital gains tax changes as an instrument for speculation and use this variation to measure the consequences of housing speculation for the real economy.

a 1 percentage point increase in house price growth in the prior month is associated with a 0.12 percentage point increase in the non-occupant share of entrants. This estimate is qualitatively similar to and statistically indistinguishable from the analogous coefficient for short-term buyers.

Finally, in column 3 of the table we estimate a three-equation pVAR that allows for joint relations between all three variables of interest. The results from this specification are both qualitatively and quantitatively similar to those from columns 1 and 2. Short-buyer entry is strongly predictive of subsequent house price growth and predicted by recent past price growth, whereas non-occupant entry can be predicted by past price growth but is less informative for predicting subsequent prices. Stronger predictive power for the short-buyer share is also consistent with the horserace specification in Table 2.

These results are similar both qualitatively and quantitatively to those in Chinco and Mayer (2015) (see their Table 7). They find coefficients for lagged out-of-town second-house buyers versus house price growth of 0.02 percentage points, which matches our short-buyer share coefficient. They find that local second-house buyers do not predict future house price growth. Combining their two groups of second-house buyers would deliver an estimate identical to our non-occupant coefficient. Relative to their specification, we consider a sample of MSAs that is five times as large and focus on the distinction between short-term buyers and non-occupants rather than differences within the group of non-occupants.

3.2 Characterizing Speculative Buyers

Our results thus far indicate that short-term buyers were a major driver of changes in transaction volume over time and across MSAs during the boom, and that more speculative entry is associated with more price growth. In this section, we use our detailed microdata to shed further light on the nature of these speculative short-term purchases.

First, we ask what share of short-term volume was from sellers who were non-occupant buyers. The results above indicate that both short-term and non-occupant buyers were disproportionately active during the run-up in house prices from 2000 to 2005, though with potentially different amplification effects on house prices. However, there may be overlap between these two groups. Focusing on the 102 MSAs with reliable non-occupant data, of the 2000–2005 short-term volume, we find that 800 thousand out of 3.00 million (27%) were non-occupant buyers (excluding developer buyers, defined below). Between 2000 and 2005,

the number of short-term-non-occupant-buyer transactions increases from 90 thousand to 230 thousand, or 39% of the overall growth in short-term transactions (which grew from 370 thousand to 730 thousand, excluding developer buyers). Non-occupant buyers thus account for an excess share of the growth in short-term buyers, further suggesting that speculative motives drive short-term trading behavior.

Second, we ask what share of short-term buyers were experienced investors versus inexperienced speculators in merely one or two homes. To answer this question, we count the total number of transactions for each unique buyer name in an MSA and then ask what share of total transactions in that MSA are associated with buyers with few purchases during the entire sample period versus buyers with many purchases. We classify buyers with one or two purchases as inexperienced and those with three or more as experienced. Of the 2000–2005 short-term volume, 2.52 million of 3.44 million (73%) were inexperienced buyers (excluding developer buyers). Between 2000 and 2005, the number of inexperienced short-term-buyer transactions increases from 320 thousand to 590 thousand, or 71% of the growth in short-term transactions.

Consistent with the evidence in Bayer, Geissler, Mangum and Roberts (2020), who use a similar methodology, entry of inexperienced buyers is critical for understanding the growth in aggregate volume. The relative lack of experience among this class of investors may also be relevant for understanding the contemporaneous patterns in prices. Bayer, Geissler, Mangum and Roberts (2020) and Bayer, Mangum and Roberts (2016) show that inexperienced short-term investors in Los Angeles and some other cities pursue a momentum-trading strategy and that their behavior is influenced by that of other nearby speculators, respectively. Both of these patterns are consistent with the notion of extrapolation-induced entry of short-term buyers we consider in our model.

Third, we ask what role credit played in enabling short-term volume. We evaluate this question by decomposing the increase in short-term selling into transactions based on how much leverage the buyer originally used. We focus on a low-leverage group (purchase loan-to-value (LTV) $< 60\%$), a medium-leverage group (purchase LTV $\in [60\%, 85\%)$), and a high-leverage group (purchase LTV $> 85\%$). Of the 2000–2005 short-term volume, 1.19 million (31%) were low-LTV buyers, 1.32 million (34%) were medium-LTV buyers, and 1.33 million (35%) were high-LTV buyers. In contrast, for the long-term volume transactions for which we observe purchase LTVs (i.e., with initial purchase during or after 1995), the

distribution skews more toward high-leverage buyers, with 22% in the low-LTV, 30% in the medium-LTV, and 48% in the high-LTV groups, respectively. Between 2000 and 2005, the number of low-LTV, medium-LTV, and high-LTV short-term-buyer transactions increases from 200 to 250 thousand, from 140 to 370 thousand, and from 170 to 280 thousand, or 13%, 59%, and 28% of the growth in short-term transactions, respectively.

As in our analysis of cash transactions among speculative buyers (Table 1), short-term volume is associated with lower use of leverage in the cross-section relative to the general population.[10] At the same time, the proportional growth in short-term buying is stronger among high-LTV sellers, making a larger relative contribution to the overall growth in short-term volume. This evidence is consistent with high credit growth among speculative buyers during the boom, as documented by Haughwout, Lee, Tracy and van der Klaauw (2011), Bhutta (2015), and Mian and Sufi (2019). While speculative buyers may not all have been credit-constrained, our results align with the idea that credit supply can enable speculative entry into the housing market. Thus, although our theoretical analysis abstracts from shifts in credit supply, we view our extrapolation-based story as complementary to credit-supply explanations of the boom.

Last, we ask what share of short-term volume was due to firms or developers rather than individuals. We mark transactions as developer purchases when the buyer name is both not parsed as a person by CoreLogic and contains strings reflecting developer names.[11] Across the cities in our sample, these transactions account for 6% of total volume and 10% of the growth in volume between 2000 and 2005. Of the 4.02 million transactions between 2000 and 2005 made by buyers with short-holding periods, 580 thousand (14%) were developer buyers. Between 2000 and 2005, the number of short-term-buyer transactions increases from 510 thousand to 950 thousand while the number of short-term-developer-buyer transactions increases from 80 thousand to 130 thousand, or 12% of the growth in short-term transactions. We conclude that, though developers were actively involved in the housing market, they did not contribute disproportionately to the growth in short-term volume during the boom. A possible explanation is that developers were more likely to engage in speculation in the raw land market (Nathanson and Zwick, 2018).

---

[10]Internet Appendix Table IA.I extends Table 1 to look at average purchase LTVs for short-term and non-occupant buyers. Both speculative buyer types have lower average LTVs, which is exclusively driven by their higher cash transaction shares.

[11]We identify developer names using CoreLogic's internal new construction flag, as Internet Appendix I describes.

Taken together, the results point to the importance of a class of inexperienced speculative entrants into the housing market during the cycle. These short-term speculators are increasingly likely to be non-occupant purchasers over the course of the boom, and they depend less on credit on average than the general population of homebuyers. These findings both suggest these buyers are not renters transitioning to homeownership. In Section 5.3, we also find that a relatively small share of the new buyers are existing homeowners trading up to a new house. The evidence is therefore most consistent with the interpretation that these buyers are amateur investors buying additional property in pursuit of capital gains.

Internet Appendix Figure IA.I presents additional evidence for this interpretation based on the Federal Reserve Survey of Consumer Expectations.[12] First, consistent with extrapolation, the share of respondents reporting that housing is a good or very good investment is strongly increasing in recent local house price appreciation. Second, those who view housing as a good investment also state a higher probability of buying a non-primary home. Third, there is a significant positive relation between recent house price appreciation and the probability of buying a non-primary home, which is driven by those with high liquid savings. This last result suggests the speculative behavior we document is not only due to a home equity effect.

4 Speculators During the Quiet and Bust

The previous section documented that speculative buyers played an outsized role in driving the increase in transaction volume during the boom and that their entry was strongly correlated with price changes across local markets. We now turn to studying the behavior of these investors during the later stages of the cycle.

One of the key stylized facts about the aggregate housing cycle is the existence of a long "quiet" period during which prices rise while transaction volumes rapidly fall. This period is also accompanied by a large increase in unsold listings. Table 2, Panel B, columns 4–6, shows that the rise in listings during the quiet correlates strongly with the run-up of speculative volume during the boom across MSAs. Cities with a one standard deviation larger short-volume boom (12.9%) see a larger cumulative increase in listings during the

---

[12]We thank Andreas Fuster for sharing this evidence with us.

quiet of 76.9 percentage points relative to 2003 total volume.[13] Cities with a one standard deviation larger non-occupant boom (27.1%) see a cumulative increase in listings during the quiet of 71.7 percentage points relative to 2003 total volume. Across cities, the mean increase in inventories during the quiet is 178% of 2003 total volume with a standard deviation of 144%. Thus, the relation between speculative booms and the rise of listings is quantitatively important in accounting for the cross-section of inventories.

Consistent with the aggregate evidence in Figure 3, which shows only a modest increase in listings during the boom, we find a small and statistically insignificant relation in the cross-section between speculative booms and the change in listings during the boom.[14] Given the strong cross-sectional relation during the boom between short-term volume and total volume, this fact suggests that the increase in demand during the boom was sufficient to absorb the increasing flow of listings from short-term buyers. As demand slowed, the continuing flow of listings from recent buyers saturated the market, resulting in accumulating unsold inventories during the quiet.

Figure 6 demonstrates this point with listings data linked to transaction data at the property level. The link to past transactions allows us to ask directly whether recent purchases disproportionately contribute to the surge of listings during the quiet. We plot monthly aggregate time series for total listings and short holding-period listings, defined as a listing where the previous sale occurred within the past three years. These data only count a home listed for sale the first time it appears during a listing spell, thereby measuring the flow of short-holding-period listings without double counting unsold listings. Each series is separately normalized relative to its average value in the year 2003 and seasonally adjusted by removing calendar-month fixed-effects.

The increase in listings during the quiet comes largely from recent purchases. While total listings rise to 150% of their 2003 average at the peak of the quiet, short-holding-period listings rise to 250% of their 2003 average and remain above 200% well into the bust. From

---

[13]Table 2 reports the change in the inventory of unsold listings. Internet Appendix Table IA.III reports analogous results using the change in the flow of new listings and shows qualitatively similar results. The rise in unsold listings during the quiet was driven both by an increase in the rate at which homes were listed for sale and a reduction in the probability of sale conditional on listing.

[14]Internet Appendix Table IA.II shows that the mean cumulative increase in listings from 2003 to 2005 is 92% relative to 2003 total volume with a standard deviation across cities of 95%. Of 57 MSAs in the sample, 12 see a decline in listings during this time. In terms of percentage changes, the mean cumulative increase is equivalent to a 25% (s.d.=33%) increase in accumulated listings between 2003 and 2005. This increase is modest compared with the mean price boom across MSAs of 98% (s.d.=48%) and the mean volume boom across MSAs of 48% (s.d.=43%).

2003 to 2007, we see an aggregate rise of listings within sample from 1.17 million in 2003 to 1.73 million in 2007. Short-holding-period listings rise from 280 thousand to 590 thousand, thus accounting for 55% of the rise in aggregate listings during this time. In later stages of the bust, short-holding-period listings fall well below their 2003 level, consistent with the idea that purchases during this phase of the cycle are more likely to include fundamental buyers and longer-term investors.

This evidence complements Genesove and Mayer (1997) and Genesove and Mayer (2001), who document the role of home equity and loss aversion, respectively, in preventing list prices from adjusting downward during a market downturn in Boston. Short-holding-period buyers are more likely to maintain high list prices because—in the home equity view—they will have paid down less of their mortgages when they turn to sell and because—in the loss aversion view—they will have paid higher initial prices than long-holding-period buyers. In our model, extrapolation creates another force causing recent buyers to set overly optimistic list prices, the same force that helps explain their initial entry into the market. Each of these forces likely plays a role in accounting for the facts.

Table 2, Panel C, considers how the size of the speculative boom is associated with the severity of the bust. Both total volume and prices fall substantially more after their respective peaks in cities with larger speculative booms. Cities with a one standard deviation larger short-volume boom and non-occupant boom respectively see cumulative declines in total volume (relative to 2003 volume) 13.5 and 13.9 percentage points larger. The results correspond to 7.4 and 4.5 percentage point larger cumulative price declines during the bust.

Total volume falls on average across cities by 63% in the quiet and bust relative to 2003 volume. Prices fall on average across cities by 28% during the bust. Thus, the size of the speculative volume boom is associated with larger busts in both volume and prices. These facts are consistent with the aggregate pattern in Figure 4, in which speculative volume declines more sharply during the quiet and bust than does total volume. Turning points in both short-holding-period volume and non-occupant volume exactly coincide with the turning point in aggregate volume, the sharp rise in listings during the quiet, and the slowing of price growth before its reversal.

Finally, we look at whether speculative booms are associated with higher foreclosures during the bust. Aside from the policy interest in foreclosures, this outcome is relevant for three reasons. First, in Section 2.2, we note that a significant share of speculative purchases

are all-cash purchases. However, more than half of speculative purchases involve mortgages. To the extent that speculators derive lower use benefits from housing, they may be more likely to default when underwater. Second, as speculative entry increases following past price growth, speculators are increasingly likely to buy when prices are higher, including later in the cycle. Third, to the extent the speculative boom amplifies the price cycle, it may cause more non-speculative buyers to become underwater in the bust, as these buyers happened to be unlucky in market timing.

We find that the short-term speculative boom coincides with a larger number of foreclosures during the bust, while the non-occupant boom does not. A one standard deviation increase in the short-volume boom is associated with 11.5 percentage points more foreclosures (relative to 2003 volume) during the bust, equal to 370 thousand more foreclosures. During this time, there were 2.68 million foreclosures across the 115 cities in our data. Cities with larger short-term speculative booms therefore experienced more severe foreclosure crises. In contrast, the relation between foreclosures and the non-occupant boom is insignificant and small in the pairwise specification and becomes meaningfully negative in the horserace specification.[15]

One interpretation of this result is that short-holding-period volume is a more precise measure of speculative activity than non-occupant volume, perhaps because non-occupant volume also includes longer-term "cash flow" investors and vacation homebuyers. Given significant overlap between the short-holding-period and non-occupant category, conditioning on the level of short-term volume would leave these latter types of non-occupants in the residual variation. This residual activity might actually mitigate the speculative cycle because these buyers are less likely to enter and exit the market concurrently with the short-term buyers. We explore this idea in the model, which allows us to separate a buyer's horizon from the utility she receives from buying a house.

---

[15]Related work documents a disproportionate share of investors among delinquencies and foreclosures. See, e.g., Haughwout, Lee, Tracy and van der Klaauw (2011) and Piskorski and Seru (2018). Because this work relies on mortgage data sets, it does not consider the significant number of all-cash investors, which may explain our different results for non-occupants relative to these papers. Guren and McQuade (2020) also relate the extent of foreclosures to the size of the boom in the cross-section.

5 Robustness and Alternative Explanations

5.1 Endogenous Holding Periods

The evidence above indicates that the differential entry of speculative buyers played a major role in driving the volume boom. However, the results for short-term volume growth are based on realized rather than expected holding periods. This way of measuring short-term speculation may complicate the interpretation of our results if buyers' intended holding periods endogenously respond to changes in economic conditions during the boom. The results on non-occupant buyers partially address this concern as they are based on a measure of speculative entry that does not suffer from the same issue. However, to address this issue further, we provide two direct pieces of evidence suggesting that the results for short-term volume are not just driven by endogenous changes in holding periods.

Our first approach instruments for realized short-term volume growth using ex-ante demographic characteristics of an area that are likely to be correlated with intended short holding periods among the population of potential homebuyers. In particular, we use the 2000 Census 5% microdata to calculate the share of recent homebuyers (within the last 5 years) in each MSA that were either younger than 35 or aged 65 and older at the time of questioning and include both shares as instruments for 2000–2005 short-term volume growth. This approach follows Edelstein and Qian (2014), who use data from the American Housing Survey to study demographic and mortgage characteristics as predictors of ex-ante investment horizon. Both older and younger buyers tend to have shorter horizons than middle-aged buyers, likely due to life cycle forces that affect the propensity to move, which gives the instrument its relevance.[16]

The strength of this instrument is that it is predetermined relative to the realized holding periods for sellers during the boom and may therefore help purge our estimates of mechanical bias arising from endogenous changes in holding periods over the course of an ownership spell. We stress that this instrument does not remove the potential influence of age-specific shocks, so we do not interpret the IV regressions as demonstrating a causal relation. Rather our goal with this exercise is to mitigate potential mechanical feedback effects between total volume and short-term volume.

---

[16]Internet Appendix Table IA.IV reports the first stage regressions of the short volume boom on the old and young shares.

Table 4 presents the results. As a baseline, we first show that a basic OLS regression of the 2000–2005 percent change in total volume on the 2000–2005 change in short-term volume divided by year-2000 total volume replicates the conclusion from Figure 5, Panel C. Column 1 presents this result. Because we are interested in instrumenting for short-volume growth, the left- and right-hand-side variables in this regression are swapped relative to their analogs in Figure 5. Thus, the coefficient estimate of 2.3 reported in Panel A is not directly comparable to the 0.3 number from Figure 5, Panel C. Panel B of the table, however, reports a variance decomposition indicating that 33 percent of the variation in total volume growth across MSAs can be explained by changes in short-term volume, which matches the short-term volume result from Figure 5. Column 2 shows that the same regression using non-occupant volume on the right-hand-side replicates the corresponding Figure 5 result for that measure of speculation. Columns 3 and 4 report quantitatively similar relations in ZIP-code level regressions with MSA fixed effects.[17]

Table 4, column 5, shows that the short-term volume coefficient does not fall when we instrument using year-2000 homebuyer age. If a mechanical relation were driving this correlation, we would expect the IV coefficient to fall relative to the OLS. Instead, the coefficient modestly (and insignificantly) increases from 2.30 to 2.85. Thus, the change in realized short-term volume is quantitatively important for determining overall volume growth even when using only the portion of short-term volume growth predicted by pre-existing buyer characteristics.

Our second approach to addressing the measurement issues associated with studying realized rather than expected holding periods leverages survey data from the National Association of Realtors (NAR) that asks recent buyers about their intended holding period. Unfortunately these data are only available at the national level from 2008 onward. However, the data that are available suggest that expected investment horizons vary considerably across individuals and commove strongly with recent price changes.

---

[17]Throughout the paper, we focus our empirical analysis on MSA-level outcomes for two reasons. First, while there is independent and interesting variation across ZIP codes within cities, the variation across cities is likely more informative for the aggregate housing cycle. Focusing on ZIP-level analysis would effectively place much of the interesting variation into MSA-by-time fixed effects. Second, and related to the first, spatial correlation across ZIP codes within cities hinders interpretation of cross-sectional results for some housing market outcomes. For example, MSA fixed effects account for 86% of the variation in house price booms across ZIP codes, but only 16% of the variation in volume booms across ZIP codes. It is likely this difference is due to data limitations in house price index estimation, with local price indices often derived from spatial interpolation. This issue may help explain differences in results in cross-MSA analyses, as in our paper, and cross-ZIP, within-MSA analyses, as in Griffin, Kruger and Maturana (2020).

Figure 7, Panel A, documents the substantial cross-sectional heterogeneity in expected holding times among respondents to the NAR's Investment and Vacation Home Buyers Survey. Each bar reports an equal-weighted average of the share of recent buyers who report a given expected holding time across survey years. Averages are reported separately by property type.

We emphasize three facts from this data. First, the vast majority of recent homebuyers (roughly 80%) report knowing what their expected holding time will be. Second, there is wide variation in expected holding times among those who report. About half of the expected holding times are between 0 and 11 years and are distributed somewhat uniformly over that range. The survey question groups the remaining half of the responses into a single expected holding time of greater than or equal to 11 years; however, there may be substantial variation within that group as well. Third, expected holding times also vary in an intuitive way across property types. Recent buyers of investment properties report substantially shorter expected holding periods than recent buyers of primary residences.

This baseline heterogeneity in expected holding periods correlates strongly in the time series with recent house price changes. To demonstrate this, we separately calculate for each year of the survey the fraction of respondents (except those reporting "don't know") who report an expected holding time of less than 3 years or had already sold their property by the time of the survey. Figure 7, Panel B, plots this short-term buyer share against annual house price appreciation at the national level. A simple regression of the short-term buyer share on the equal-weighted average year-over-year change in the nominal quarterly FHFA U.S. house price index during the survey year yields a statistically significant coefficient estimate of 0.82. This coefficient implies that a recent nominal gain of 10% in house prices is associated with an increase in the short-term buyer share of 8.2 percentage points. For reference, nominal house price appreciation was 11% in the U.S. in 2005 and much larger in some metropolitan areas. Thus, changes in house prices during the 2000–2005 boom period may have induced significant shifts in the distribution of expected holding times among homebuyers entering the market at that time.

5.2 Mechanical Short-Term Volume

In Figure 4 we document a rise in the share of volume coming from short-term sales during the boom. Our interpretation of this pattern is that short-term volume rises due to a shift

in the composition of buyers toward those with shorter intended holding periods. However, even in the absence of such a shift, any increase in total volume during the early part of the boom will generate a mechanical increase in the share of late-boom volume coming from short-term sales. The richness of our data allows us to quantify the contribution of this mechanical force relative to changes in the composition of buyers.

For each pair of distinct months between 1995 and 2005, we compute a conditional selling hazard $\pi_{t',t}$. This hazard is the share of homes purchased in month t'—and that have not yet sold by month t—that sell in month t. By focusing on selling hazards instead of total volume, we remove the mechanical force that comes from volume increasing over the cycle.

We estimate the following regression at the month-pair level:

$$\pi_{t',t} = \alpha^{buy}_{y(t')} + \alpha^{sell}_{y(t)} + \alpha^{duration}_{t-t'} + \epsilon_{t',t},$$

where $y(\cdot)$ gives the year of the month. The first set of fixed effects, $\alpha^{buy}_{y(t')}$, captures the average propensity of buyer cohorts from year $y(t')$ to sell in any future year. The second set of fixed effects, $\alpha^{sell}_{y(t)}$, captures the average propensity of all owners to sell in year $y(t)$. The third set of fixed effects, $\alpha^{duration}_{t-t'}$, measures time-invariant selling hazard profiles as a function of time elapsed since purchase $t - t'$. We interpret year-to-year movements in $\alpha^{buy}_{y(t')}$ as changes in the composition of buyers across those years, holding fixed both year-specific shocks to selling hazards that affect all cohorts equally and duration-specific drivers of selling hazards that do not vary over the cycle.

Internet Appendix II reports a sharp increase in $\hat{\alpha}^{buy}_{y(t')}$ from $y(t') = 2000$ to $y(t') = 2005$. The magnitude implies a 3.2 percentage point larger annual selling hazard of buyers later in the boom. Using these estimates, we perform a counterfactual in which $\alpha^{buy}_{y(t')}$ remains constant at its estimate in 2000 throughout the boom, representing a situation in which the composition of buyers remains constant. In this counterfactual, the disproportionate rise in short-term volume falls by 88%. Therefore, the changing composition of buyers during the boom can explain almost all of the disproportionate rise in short-term volume.

5.3   Repeat Buyers

The patterns we document are consistent with speculative motives leading short-term buyers to enter and exit the local housing market in response to expected capital gains. But some

short-term sellers likely do not exit the market and instead choose to buy another house within the same MSA. Such a pattern may reflect move-up purchases enabled by higher home equity during the boom, as in Stein (1995) and Ortalo-Magné and Rady (2006) or the repeated buying and selling of homes within the same market by experienced "flippers" as in Choi, Hong and Scheinkman (2014) and Bayer, Geissler, Mangum and Roberts (2020).

To explore this alternative explanation, we follow the methodology of Anenberg and Bayer (2013) and construct a direct measure of repeated within-MSA purchases. We use the names of buyers and sellers to match transactions as being possibly linked in a joint buyer-seller event. For each sale transaction, we attempt to identify a purchase transaction in which the seller from the sale matches the buyer from the purchase. To allow the possibility that a purchase occurs before a sale or with a lag, we look for matches in a window of plus or minus one quarter around the quarter of the sale transaction. We only look for within-MSA matches, as purchases associated with cross-city moves are similar in spirit to our model.

Our match accounts for several anomalies that would lead a naive match strategy to understate the match rate.[18] Our approach is likely to overstate the number of true matches, because it does not use address information to restrict matches, and it allows common names to match even if they represent different people. Because we find a low match rate even with this aggressive strategy, we do not make use of address information in our algorithm or otherwise attempt to refine matches.

We focus on transactions between 2002 and 2011 because the seller name fields are incomplete in prior years for several cities. We also restrict sales transactions to those with human sellers, as indicated by the name being parsed and separated into first and last name fields by CoreLogic. The sample includes 16.3 million sales transactions. Of these, we are able to match 3.9 million to a linked buyer transaction, or 24%. Thus, three-quarters of transactions do not appear to be associated with joint buyer-seller decisions. Among sellers who had bought within the last three years, the match rate is slightly higher, equal to 31%, consistent with move-up purchase or flipper behavior. In addition, the match rates peak in 2005 at 29% and 38% for all transactions and short-term transactions, respectively.[19] These

---

[18]These include: inconsistent use of nicknames (e.g., Charles versus Charlie), initials in place of first names, the presence or absence of middle initials, transitions from a couples buyer to a single buyer via divorce, transitions from a single buyer to a couples buyer via cohabitation, and reversal of order in couples purchases.

[19]In terms of growth between 2002 and 2005, internal moves account for approximately 40% of the growth in aggregate volume in our data, and the growth in internal short-volume accounts for 46% of total short

patterns confirm and extend the findings in Anenberg and Bayer (2013), who conduct a similar match for the Los Angeles metro area and show that internal moves account for a substantial share of the volatility of transaction volume in that city. However, the evidence supports the notion that sellers not engaging in repeat purchases account for most of the short-term volume and its growth, even during the cycle's peak.

6 Summary of Empirical Findings and Theoretical Motivation

Our findings support a narrative in which short-term speculation amplifies the housing cycle. Moreover, short-term speculation is quantitatively first order in the following senses. First, at the aggregate level, short-term speculation accounts for a large share of transactions during the cycle. Second, across cities, those with larger speculative booms experience much larger overall cycles, both in terms of a larger boom and a more severe bust. We also find evidence consistent with extrapolative expectations driving the differential entry of short-term speculators across cities and with that entry amplifying the price cycle.

While this evidence is consistent with the short-term speculation narrative, we are not documenting a sharply identified causal link between speculators and the cycle. For example, our instrumental variables analysis primarily addresses concerns with simultaneity but not more general endogeneity concerns. The analysis of predictability running from the size of the speculative boom or the share of speculative activity to subsequent housing market outcomes similarly does not permit strong causal statements. The next section presents a model consistent with the empirical evidence, which also permits stronger causal statements within the model's framework and allows us to study the speculative mechanism in further detail.

Three strands of the literature theoretically explain the comovement of prices and volume in the housing market and asset markets more generally. The first consists of models in which investors disagree about asset values, such as Scheinkman and Xiong (2003). The second exploits features specific to the housing market, such as credit constraints (Stein,

---

volume growth. The importance of internal volume varies across cities and years during the boom, with the internal move share of MSA-level short-volume growth ranging from 35% to 46% on average. On average across MSAs, growth in internal short-volume accounts for 35% of the growth in total short volume in 2005, the peak year in total volume.

1995; Ortalo-Magné and Rady, 2006) or search and matching frictions (see the review in Han and Strange (2015)). Finally, two recent papers incorporate insights from psychology into models with extrapolative expectations to generate trade (Barberis, Greenwood, Jin and Shleifer, 2018; Liao and Peng, 2018). Some papers straddle multiple categories. Guren (2014) incorporates extrapolation into a search model of the housing market, while Piazzesi and Schneider (2009) and Burnside, Eichenbaum and Rebelo (2016) incorporate disagreement into the same. While all of these papers can explain the comovement of prices and volume during the boom and bust, there are three additional results from our empirical work that no prior model seems able to explain simultaneously.

First, the increase in volume during the boom, and listings during the boom and quiet, come disproportionately from short-term sales (Figures 4 and 6). Search-and-matching models struggle to generate this pattern if the decision to list is independent of homeowner characteristics, as in Wheaton (1990), Piazzesi and Schneider (2009), Díaz and Jerez (2013), Guren (2014), Head, Lloyd-Ellis and Sun (2014) and Anenberg and Bayer (2020).[20] These models may generate a mechanical increase in short-term volume, but they cannot explain the result in Section 5.2 that homeowners who bought later in the boom were more likely to resell than homeowners who bought earlier. In contrast, the disagreement and extrapolation–psychology papers seem able to generate a disproportionate short-term volume boom, as long as rising prices generate more disagreement or psychological urge to both buy and sell the asset.

Second, non-occupants constitute a disproportionate share of the increase in buying activity during the boom (Figure 4). Non-occupant purchasing is absent from many search-and-matching models, either because the owner-occupied and rental markets are separate (Guren, 2014), or because all non-occupant owners are previous occupants of the same house (Head, Lloyd-Ellis and Sun, 2014; Burnside, Eichenbaum and Rebelo, 2016). The extrapolation–psychology papers also provide no role for non-occupants, as they model more general asset markets where all owners receive the same flow benefits from the asset. Nathanson and Zwick (2018) present a disagreement model in which non-occupants disproportionately buy housing during a boom, but their model is static and is therefore not suited to explain the

---

[20]Two exceptions are Hedlund (2016) and Ngai and Sheedy (2016), who respectively focus on credit constraints and within-market moves. As we explain in Section 5.3, short-term volume increases significantly among low-LTV sellers, and most short-term sellers do not relocate within the same MSA. Therefore, these two papers do not explain all of the disproportionate rise in short-term volume during the boom.

dynamics at the heart of this paper.

The third result is the existence of the quiet, during which prices and volume diverge while listings accumulate (Figures 1 and 3). Disagreement papers and credit-constraint housing models predict a monotonic relation between prices and volume, and therefore do not explain a period when these outcomes move in opposite directions.[21] Barberis, Greenwood, Jin and Shleifer (2018) and Liao and Peng (2018) can generate a divergence of prices and volume, but listings fall with volume because of Walrasian market clearing. A similar pattern of prices, volume, and listings appears in Burnside, Eichenbaum and Rebelo (2016). In contrast, Guren (2014) matches all three variables. However, listings sharply decline during his boom (more than one-for-one with respect to prices), and they never rise above their pre-shock level in his impulse response. Empirically, we find that listings modestly rise during the boom in aggregate and in most MSAs (Section 4). The sharp rise in listings during the quiet, far above their 2000 level, is perhaps the most salient aspect of Figure 3.

In light of these papers, the goal of our model is to match the joint dynamics of prices, volume, and listings in a way that matches the disproportionate role of non-occupants and short-term sales in driving up volume during the boom and listings during the boom and quiet. Additionally, the model should explain the cross-sectional and time-series relations between speculative volume and other cycle outcomes in Tables 2 and 3. Finally, the model should clarify the differences between short-term and non-occupant volume: the short-holding-period boom tends to be a stronger predictor of quiet and bust dynamics than the non-occupant boom, and short-term volume is associated with stronger house price predictability in the pVARs.

7 The Model

7.1 Environment and Preferences

We present a discrete-time model of a city with a fixed amount of perfectly durable housing, normalized to have measure one. Agents go through a life cycle with three possible phases:

---

[21]Disagreement also struggles to explain the widespread optimism about house price growth during the boom we study (Case, Shiller and Thompson, 2012; Foote, Gerardi and Willen, 2012; Cheng, Raina and Xiong, 2014), although it can generate the dispersion in these beliefs (Piazzesi and Schneider, 2009; Burnside, Eichenbaum and Rebelo, 2016) and surely accounts for some of the average prices and volume in the housing market (Bailey, Cao, Kuchler and Stroebel, 2016).

potential buyer, stayer, and mover.

Each period, movers list their houses for sale. After posting a list price, each mover matches to a randomly selected potential buyer, who decides whether to purchase at the listed price or exit the housing market permanently.[22] If the potential buyer chooses to purchase then the mover receives the list price and exits the market. A purchasing potential buyer becomes a stayer and receives flow utility e^{d_i} at the beginning of each future period until he randomly becomes a mover, which happens with Poisson hazard λ_i.

At t, potential buyer flow utility satisfies

$$d_i = d_t + a_i, \tag{2}$$

where d_t is a time-varying demand shifter, and a_i varies across potential buyers at a given time. Each potential buyer has one of two occupancy types, $n_i \in \{0, 1\}$. The distribution of a_i across potential buyers of type n is $\mathcal{N}(\mu_n, \sigma_a^2)$. We normalize $\mu_0 = 0$ so that μ_1 gives the average log difference in flow utility between occupants ($n_i = 1$) and non-occupants ($n_i = 0$). The demand shifter, d_t, is a difference-stationary process with a persistent growth rate:

$$d_t = d_{t-1} + g_t + \epsilon_t^d$$
$$g_t = (1 - \rho)\mu + \rho g_{t-1} + \epsilon_t^g,$$

where ϵ_t^d and ϵ_t^g are mean-zero independent normals with variances $(1-\gamma)\sigma^2$ and $\gamma(1-\rho^2)\sigma^2$, so that σ^2 is the variance of Δd and $\gamma \in (0, 1)$ is the share of that variance coming from g.

Potential buyers vary in λ_i, a_i, and n_i. The mover hazard, λ, follows a discrete distribution β^λ. The share of each occupancy type is β_n, and β_n^λ is the share of each (n, λ) pair. To match the data on expected holding times in Figure 7, Panel A, we allow non-zero correlation between λ_i and n_i. We denote the CDF of a_i across all potential buyers by F, a mixture of two normals.[23]

---

[22]In other models, some movers fail to match to a potential buyer due to search frictions (Head, Lloyd-Ellis and Sun, 2014; Guren, 2018). We abstract from this possibility.

[23]Potential buyer types in our model bear some similarities to the taxonomies in Frankel and Froot (1986), Cutler, Poterba and Summers (1990) and De Long, Shleifer, Summers and Waldmann (1990), which feature positive feedback traders, fundamentalists, and rational arbitrageurs. Whereas those papers assume different objectives or beliefs across agents, we derive heterogeneous investment behavior arising from exogenous differences in horizons. Hong and Stein (1999) also connect investment to horizons, and we differ from that paper primarily by departing from Walrasian market clearing.

Agents are risk-neutral and act to maximize their expectation of the net present value of their utility. The flow utility of living outside the city equals zero, a normalization constant. Perfect credit markets exist with a constant interest rate equal to r. Potential buyers discount the time until becoming a mover at r. Movers discount time while being a mover at the rate $r_m \geq r$, which captures possible costs of moving. To rule out rational bubbles, we assume that $1 + r > e^{\mu + \sigma^2/2}$, the unconditional expected growth of demand, and guarantee that this inequality holds by setting $\mu = -\sigma^2/2$ in the quantitative exercise so that the unconditional expected growth rate of e^{d_t} is 0.

7.2 Information and Beliefs

We denote the average list price at t by P_t, and the share of those listings that sell by π_t. At t, agents observe the history of price changes and sales shares, $P_{t'}/P_{t'-1}$ and $\pi_{t'}$ for $t' < t$. Potential buyer i also observes her flow utility, d_i, occupancy type, n_i, horizon type, λ_i, and the list price to which he matches, $P_{i,t}$. Agents cannot observe the demand shifter, d, or its growth rate, g, and must infer current values of these latent demand variables using historical market data and their private information.

Glaeser and Nathanson (2017) propose a behavioral approximation called the *cap rate error* that agents use to solve this inference problem. The cap rate error is the belief that another potential buyer i decides to purchase a listing if and only if

$$e^{d_i} \geq \kappa P_{i,t}, \tag{3}$$

where κ is a time-invariant constant. By employing the cap rate error, agents infer demand growth from market data without taking a stand on the evolution of the beliefs of other market participants. Because agents neglect the sensitivity of market outcomes to others' beliefs, the cap rate error endogenously leads to extrapolative beliefs about house price growth as well as predictable booms and busts in house prices. We follow Glaeser and Nathanson (2017) in assuming that the cap rate error characterizes the beliefs of agents in our model. Our contribution is analyzing the implications for quantity dynamics. In Glaeser and Nathanson (2017), volume is constant and listings sell immediately.

We focus on equilibria in which all movers at a given time post the same list price (and discuss conditions for this outcome below). In this case, substituting (2) into (3) and then

taking logs yields

$$\tilde{d}_t = \log P_t - F^{-1}(1 - \pi_t) + \log \kappa, \tag{4}$$

where the tilde denotes an inference true under the cap rate error (but not necessarily in reality). Movers at t deduce the history of price levels, $P_{t'}$ for $t' < t$, from the history of price changes as well as the price they faced when they purchased their house. They directly observe $\pi_{t'}$ for $t' < t$. Therefore, using (4), they infer the full history of demand before time t as $\tilde{d}_{t'}$ for $t' < t$. Kalman filtering produces the following posterior beliefs about d_t and g_t:

Lemma 1. *Movers at t have a normal posterior on g_t and d_t with means*

$$\hat{g}_t = \mu + (1 - \alpha)\rho \sum_{j=1}^{\infty} (\alpha\rho)^{j-1} \left(\Delta\tilde{d}_{t-j} - \mu \right)$$

and $\hat{d}_t = \tilde{d}_{t-1} + \hat{g}_t$, where $\alpha \in (0, 1)$ is a constant depending on σ, γ, and ρ.

Proof. Appendix A.1. □

We denote the perceived posterior variance on d_t by $\tilde{\sigma}^2$. In the quantitative exercise, we choose κ so that the average value of $d_t - \hat{d}_t$ equals zero, as in Glaeser and Nathanson (2017). Lemma 1 implies the recursions

$$\hat{g}_{t+1} = (1 - \rho)\mu + \rho\hat{g}_t + \rho(1 - \alpha)\left(\tilde{d}_t - \hat{d}_t\right) \tag{5}$$

$$\hat{d}_{t+1} = \hat{d}_t + \hat{g}_{t+1} + \left(\tilde{d}_t - \hat{d}_t\right), \tag{6}$$

which are useful for defining value functions below. Intuitively, due to (4), $\hat{g}_t$ rises with past price growth, as in Glaeser and Nathanson (2017), and also with the growth of π, so that movers infer a higher demand growth rate when the speed at which listings sell is increasing.

7.3 Prices

Movers choose prices optimally given their mistaken belief about potential buyer demand. The demand curve that movers believe they face is

$$\tilde{\pi}(P, d_t) = 1 - F(\log P + \log \kappa - d_t).$$

30

The mover value function satisfies the recursion

$$V^m(\hat{d}_t, \hat{g}_t) = \sup_P E\left(\tilde{\pi}(P, d_t)P + (1 + r_m)^{-1}(1 - \tilde{\pi}(P, d_t))V^m(\hat{d}_{t+1}, \hat{g}_{t+1})\right), \tag{7}$$

where the expectation is over $d_t \sim \mathcal{N}(\hat{d}_t, \tilde{\sigma}^2)$. Because movers believe that $\tilde{d}_t = d_t$, each potential realization of d_t determines $\hat{d}_{t+1}$ and $\hat{g}_{t+1}$ via (5) and (6), so (7) is well-defined. Movers at a given time post the same list price when a unique solution to (7) exists, which we verify at each point of the state space of our quantitative exercise.[24] The following lemma clarifies how this list price depends on mover beliefs, $\hat{d}_t$ and $\hat{g}_t$.

Lemma 2. *The optimal list price takes the form $e^{\hat{d}_t}p(\hat{g}_t)$ for some function $p(\cdot)$.*

Proof. Appendix A.2. □

The log list price scales one-for-one with the current belief about the level of demand $\hat{d}_t$. It also depends on the belief about the demand growth rate $\hat{g}_t$ because the option of selling next period becomes more valuable when movers expect faster demand growth. In the limit of infinite mover impatience ($r_m \to \infty$), this option is irrelevant, so $p(\cdot)$ is constant. In this case, price setting closely resembles the extrapolative rule of thumb that Guren (2018) assumes, and price growth expectations satisfy a condition analogous to the reduced form extrapolation formulas that Barberis, Greenwood, Jin and Shleifer (2015, 2018) and Liao and Peng (2018) assume (see Appendix A.3).[25] In our quantitative exercise, we use a finite r_m and measure the extent to which price growth expectations depend on recent price growth.

---

[24] In general, movers may be indifferent between different list prices, or they may prefer to set an infinite list price when the right side of (7) is unbounded. We rule out these possibilities by verifying that a unique price in a fine mesh maximizes the right side of (7), and that the value function at this price exceeds the limiting value as $P \to \infty$.

[25] In particular, price growth expected over the next period is an affine function of an exponential weighted average of past growth. In our context, that affine function is

$$E_t \Delta \log P_{t+1} = \mu + \frac{\rho^2(1 - \alpha)}{1 + \rho(1 - \alpha)} \sum_{j=0}^{\infty} \rho^j (\Delta \log P_{t-j} - \mu).$$

7.4 Buyer Composition

Potential buyers decide whether to buy in light of their beliefs and flow utility. The value to potential buyer i at time t of owning a house is

$$V_{i,t}^b = \sum_{j=1}^{\infty} \lambda_i (1 - \lambda_i)^{j-1} \left(\sum_{k=1}^{j} \frac{e^{d_i}}{(1+r)^k} + \frac{E_{i,t} V^m(\hat{d}_{t+j}, \hat{g}_{t+j})}{(1+r)^j} \right), \tag{8}$$

where $E_{i,t}$ denotes the potential buyer's expectation conditional on her information set. Potential buyer i imputes $\hat{d}_t$ using the equation

$$\hat{d}_t = \log P_{i,t} - \log p(\hat{g}_t),$$

which holds due to Lemma 2. Because she observes the history of price growth and sales shares, she directly calculates $\hat{g}_t$ using the formula in Lemma 1. By (5) and (6), future values of these variables depend on the innovations $\tilde{d}_{t+j} - \hat{d}_{t+j}$ for $j \geq 0$. From the standpoint of the potential buyer, these innovations are distributed independently as $\mathcal{N}(0, \tilde{\sigma}^2)$ for $j > 1$. For $j = 0$, however, her information about her own flow utility is informative, and her posterior on this innovation is

$$d_t - \hat{d}_t \sim \mathcal{N} \left(\frac{\tilde{\sigma}^2 (d_i - \mu_{n_i} - \hat{d}_t)}{\tilde{\sigma}^2 + \sigma_a^2}, \frac{\tilde{\sigma}^2 \sigma_a^2}{\tilde{\sigma}^2 + \sigma_a^2} \right). \tag{9}$$

A purchase occurs when $V_{i,t}^b \geq P_{i,t}$. Lemma 3 uses (8) to simplify this decision rule.

Lemma 3. *Potential buyer i purchases a house at t if and only if*

$$e^{d_i} \geq \kappa_{n_i}^{\lambda_i}(\hat{g}_t) P_{i,t}.$$

The cutoff rule that potential buyers use to determine whether to purchase resembles the belief that movers have under the cap rate error except for the functions $\kappa_n^\lambda(\cdot)$, which are no longer constant and instead depend on the potential buyers' expected horizon λ_i, occupancy type n_i, and demand growth expectations $\hat{g}_t$.

While it is difficult to fully characterize the properties of the $\kappa_n^\lambda(\cdot)$ functions analytically, in the quantitative exercise below we document three properties of these functions that are helpful for understanding how the composition of buyers varies over the housing cycle. First,

each $\kappa_n^\lambda(\cdot)$ decreases in $\hat{g}_t$, with steeper slopes for larger values of λ. Intuitively, when $\hat{g}_t$ is high potential buyers expect larger capital gains in the future and will therefore be willing to purchase at higher prices today. Moreover, potential buyers with larger λ expect to sell sooner, so their demand is more sensitive to expected capital gains. In the limiting case of an infinite horizon investor $(\lambda \to 0)$, equation (8) makes clear that the buying decision does not depend on $\hat{g}_t$; in this case, κ_n^λ limits to a constant value of r. Second, $\kappa_0^\lambda(\cdot)$ is nearly identical to $\kappa_1^\lambda(\cdot)$ for each λ. The cutoffs depend very little on occupancy type because σ_a is much larger than $\tilde{\sigma}$. Finally, $\kappa_n^\lambda(\cdot)$ is typically larger for greater values of λ, reflecting higher cutoffs for short-term buyers. Because listings do not sell immediately, there is an endogenous illiquidity cost to becoming a mover. Short-term buyers expect to pay this cost sooner, so they are less inclined to purchase a house ex ante.

Together with Lemma 2, Lemma 3 gives the following equation for the realized share of listings that sell:

$$\pi_t = 1 - \sum_{n,\lambda} \beta_n^\lambda \Phi \left(\log p(\hat{g}_t) + \log \kappa_n^\lambda(\hat{g}_t) + \hat{d}_t - d_t - \mu_n \right), \tag{10}$$

where Φ is the CDF of $\mathcal{N}(0, \sigma_a^2)$. Holding $\hat{d}_t$ and d_t constant, π_t increases in $\hat{g}_t$ when each $\kappa_n^\lambda(\cdot)$ decreases and when r_m is large, so that $p(\cdot)$ is nearly constant. In this case, $\hat{g}_t$ tends to raise $\hat{d}_t$ due to (4), leading agents to overestimate time-t demand when the expected growth rate is high that period. This error raises $\hat{g}_{t+1}$ via (5), leading to positive feedback over time.

The share of sales going to each type of buyer is

$$b_{n,t}^\lambda = \pi_t^{-1} \beta_n^\lambda \left(1 - \Phi \left(\log p(\hat{g}_t) + \log \kappa_n^\lambda(\hat{g}_t) + \hat{d}_t - d_t - \mu_n \right) \right). \tag{11}$$

When $\mu_1 > 0$—so that non-occupants benefit less from housing on average—log non-occupant demand is more sensitive than log occupant demand to the demand shifter, d_t, and the belief about its growth rate, $\hat{g}_t$. This result holds because the normal distribution has the monotone likelihood property, so that $\Phi'/(1 - \Phi)$ is an increasing function. Because $\kappa_n^\lambda(\cdot)$ quantitatively does not depend on n, the argument of $\Phi(\cdot)$ is always larger for non-occupants than occupants of the same λ type when $\mu_1 > 0$.

In the quantitative exercise, the log of short-term buyer demand is more sensitive than long-term buyer demand to $\hat{g}_t$ for two reasons. First, $\kappa_n^\lambda(\cdot)$ decreases more sharply for larger

values of λ. Second, κ_n^λ is greater for larger values of λ, meaning that an equal decrease in $\log \kappa_n^\lambda(\hat{g}_t)$ boosts demand more for short-term buyers than long-term buyers due to the monotone likelihood property. For a similar reason, the demand shifter, d_t, increases short-term buying more strongly than long-term buying.

7.5   Quantities

The following accounting identities characterize the evolution of inventories, I_t, new listings, L_t, and volume, V_t, given sales probabilities, π_t, and the composition of buyers, $b_{n,t}^\lambda$:

$$I_t = (1 - \pi_{t-1})I_{t-1} + L_t$$

$$V_t = \pi_t I_t$$

$$L_t = \sum_\lambda \lambda S_{t-1}^\lambda$$

$$S_t^\lambda = (1 - \lambda)S_{t-1}^\lambda + (b_{0,t}^\lambda + b_{1,t}^\lambda)V_t,$$

where S_t^λ measures end-of-period stayers of type λ. Volume to buyers of occupancy type n equals $\sum_\lambda b_{n,t}^\lambda V_t$. To track realized short-term sales, the quantity we observe in the data, we define I_t^k to be the inventory of listings at t of homes purchased at time $t-k$. This quantity satisfies the recursion

$$I_t^k = (1 - \pi_{t-1})I_{t-1}^{k-1} + \sum_\lambda \lambda(1 - \lambda)^{k-1}(b_{0,t-k}^\lambda + b_{1,t-k}^\lambda)V_{t-k}$$

for $k > 0$, with initial condition $I_t^0 = 0$. The sales volume of houses purchased within the last j periods equals $V_t^j = \sum_{k=1}^j \pi_t I_t^k$. In the data we track new short-term listings; here, new listings of homes purchased within the last j periods equals

$$L_t^j = \sum_{k=1}^j \sum_\lambda \lambda(1 - \lambda)^{k-1}(b_{0,t-k}^\lambda + b_{1,t-k}^\lambda)V_{t-k}.$$

As these equations make clear, the current composition of buyers affects the composition of stayers, thereby altering future listings and volume. Volume rises when there are more listings or when the selling probability is higher.

8 Model Results

8.1 Simulation Methodology

Solving the model requires calculating the functions $p(\hat{g}_t)$ and $\kappa_n^\lambda(\hat{g}_t)$. To do so, we discretize $\hat{g}$ using the Rouwenhorst (1995) method and then calculate the function values at these discrete points. To evaluate the functions outside these points, we use cubic splines between mesh points and linear splines beyond the boundaries.

Each simulation of our model corresponds to 148 sequential realizations of ϵ_t^d and ϵ_t^g. The first 100 periods burn in the simulation, leaving 48 analysis periods. Each period represents a quarter, so our analysis spans 12 years. We draw a control sample of 1,000 independent simulations to analyze the baseline properties of the model. To analyze the impulse response to a shock, we draw a treatment sample of 1,000 additional simulations that are identical to the control except in periods 101–104 during which the growth rate shocks ϵ_t^g are two standard deviations higher.[26] Impulse responses are average differences between the treatment and control outcomes.

We set $r = 0.012$ and $\rho = 0.880$, corresponding to annual values of 5% and 0.51 in Guren (2018) and Glaeser and Nathanson (2017), respectively. We select values of the remaining parameters so that moments from our simulation match the empirical counterparts in Table 5. The composition of buyers and the volatility of demand growth determine β_n^λ and σ, respectively, and the selling hazard disciplines r_m, as more patient movers take longer to sell by setting higher prices. We target three features of the national U.S. housing cycle: the ratio of price boom to bust, the volume boom relative to the price boom, and the degree to which the non-occupant volume boom exceeds the occupant boom. Intuitively, these moments determine γ, σ_a, and μ_0 through quantifying extrapolation, the elasticity of demand, and the excess sensitivity of non-occupants.

8.2 Parameter Estimates and Buyer Cutoff Rules

Table 6 reports parameter values that match the moments in Panels B and C of Table 5. Non-occupant flow utility is 0.9% less than occupant flow utility on average, corresponding to

---

[26]We shock ϵ^g instead of ϵ^d so that in the rational benchmark, prices never overshoot. A sequence of 4 shocks matches the experiment in Barberis, Greenwood, Jin and Shleifer (2018). We choose 2 standard deviations to explore a large but plausible increase in demand.

less than a standard deviation in each group's flow utility distribution. The mover discount rate is 14%. To map this number into a flow cost of moving, we calculate how much higher the mover value function would be if the mover discount rate were r for a single period. The average difference is 3.7% of the list price, in line with the typical costs of selling a house (Han and Strange, 2015) and smaller than the estimate in Guren (2018) of 2.1% per month.

Panel B reports the magnitude of extrapolative expectations implied by our parameter estimates. Following Armona, Fuster and Zafar (2016), we focus on the regression coefficients on last year's price growth of expected annualized price growth over the next 1 and 2–5 years. We calculate these coefficients by regressing movers' expectations in period 105 of the control simulations against price growth in the previous 4 periods. The resulting values of 0.127 and 0.042 are somewhat smaller than corresponding values of 0.226 and 0.047 that Armona, Fuster and Zafar (2016) find through a survey (see their Table 5). Therefore, to match the key housing cycle moments in panel C of Table 5, our model requires a smaller amount of extrapolation than these authors found.

In Figure 8 we plot the potential buyer cutoff functions $\kappa_n^\lambda(\hat{g}_t)$ given our chosen parameters over a wide range of expected demand growth rates. These functions determine the relative sensitivity of buyer demand across buyers with different expected holding periods and occupancy types. Three features of these functions stand out: (1) each $\kappa_n^\lambda(\cdot)$ decreases, with steeper slopes for larger values of λ, (2) $\kappa_0^\lambda(\cdot)$ and $\kappa_1^\lambda(\cdot)$ are nearly identical for each λ, and (3) $\kappa_n^\lambda(\cdot)$ are generally larger for greater values of λ. These results imply that the sensitivity of buyer demand to the expected growth rate will be larger among buyers with shorter expected holding periods and that short-term buyers are more likely to be marginal. Because holding periods and occupancy status are correlated according to our estimates in Table 6, the similarity in cutoff rules between occupants and non-occupants implies that non-occupants are also more likely to be marginal entrants into the market when expected capital gains are high.

8.3   Impulse Responses

Figure 9 plots impulse responses. As with the national U.S. cycle in Figures 1 and 3, the cycle in the model progresses through a boom, quiet, and bust (Panels A and B). We use grey shading to mark the transition points between these phases, defined as the peaks of volume and prices. The quiet lasts 8 quarters, close to the duration in Figure 1 and the

correlation-maximizing lag in Figure 2.

In the boom, demand rises because its true level, d_t is higher and because the expected growth rate, $\hat{g}_t$, rises in response to price growth. Both channels differentially stimulate buying from potential buyers with higher λ (Panel C) and non-occupants (Panel D). The overall increase in housing demand pushes up the share of listings that sell, π_t (Panel E). Short-term buyers re-list their houses quickly, increasing the flow of listings during the boom (Panel F). Prices and volume increase as a result. Tempering the volume boom is the decline in inventory (Panel B), which occurs as the stock of unsold listings diminishes.

The qualitative behavior of volume, inventories, and sale probabilities during the boom is similar in search and matching models, such as Guren (2014). The key difference is the increasing *flow* of listings coming differentially from short-term buyers (Panel F). This flow limits the decline in inventories to 1.5 log points, amplifying and sustaining the rise in volume. Relative to the price boom, this decline in inventories is an order of magnitude smaller than in Guren (2014). Furthermore, the differential flow of short-term listings leads to the short-term volume boom in Panel C, which matches Figure 4. The disproportionate increase in demand from non-occupants, together with the overall rise in volume, produces the strong non-occupant volume boom in Panel D that also matches Figure 4.

In the quiet, demand begins to fall because the price level has risen so high. Due to the cap rate error, agents misattribute demand growth during the boom entirely to d_t, even though much of it comes from $\hat{g}_t$, the expected capital gains channel. As a result, agents over-estimate the demand level, and $\hat{d}_t - d_t$ becomes increasingly positive. As (10) shows, sales probabilities then fall (Panel E). Movers increase their list prices throughout the quiet because they continue to revise upward $\hat{d}_t$, their estimate of the level of demand, for two reasons. First, because of past price growth, the expected growth rate, $\hat{g}_t$, remains high, which mechanically causes upward revisions to the expected level of demand. Second, the sale probability, π_t, remains high even though it is falling, and these high realizations constitute positive surprises about demand that cause movers to increase their beliefs. Eventually, π_t falls below its pre-shock average, ending these upward revisions and the concomitant increase in list prices.

One of the distinguishing features of the quiet in both the model and the data is the sharp rise in unsold inventories. At their peak, unsold listings are 1.4 log points above their pre-shock level. The two causes of the glut of inventories are the fall in selling probabilities

37

(Panel E) and the elevated flow of short-term listings continuing throughout the quiet (Panel F), which matches the data in Figure 6. This second cause is novel to our model and may explain why inventories rise above their pre-shock level here whereas they fail to do so in models lacking this channel, such as Guren (2014).[27]

The bust begins as movers cut list prices. Agents revise down their expectations of the growth rate, which further depresses demand and sale probabilities. Because the cap rate error leads movers to ignore this channel, movers do not cut prices enough to restore demand, and the bust continues over several periods. Volume falls below its pre-shock level, as in Figure 1. The decline in $\hat{g}_t$ leads to a smaller share of short-term buyers, depressing the flow of new listings (Panel F), which allows inventories to recover (Panel B).

The model generates a second boom in prices, volume, and listings in the last 5 years of the simulation. This second boom occurs because prices overshoot on the way down, as is common in models with extrapolative expectations (Hong and Stein, 1999; Glaeser and Nathanson, 2017). Underpricing occurs when agents think that demand is *lower* than its true value, so that $\hat{d}_t - d_t$ becomes negative. As (10) shows, sale probabilities then rise, increasing volume. This increase in demand disproportionately affects short-term buyers, so short-term volume and listings rise during the second boom.

8.4   Counterfactuals

Many features of the impulse responses discussed above closely match the patterns observed in the data. However, the fact that our model matches these patterns does not directly speak to the role that speculators play in generating those patterns. To quantify the contribution of speculators to the housing cycle, we rerun the simulation under counterfactuals with only long-term buyers or occupants. Doing so allows us to make causal statements within the model's framework that are not feasible in our empirical analysis.

To study the role of short-term buyers, we re-run the simulations setting $\beta^\lambda = 0$ for all values of λ except for $\lambda = 0.03$, which is close to the reciprocal of the average horizon among potential buyers in the baseline. By assigning all potential buyers the same (low) value of λ this counterfactual removes both short-term buyers and the heterogeneity in holding

---

[27]Our model understates the rise in listings during the quiet because of our simplifying assumption that each mover matches to a potential buyer regardless of the number of contemporaneous movers. With a more realistic matching function, such as the one in Guren (2014), our model might hit the peak of listings (relative to price growth) that appears in Figure 3.

periods that generates variation in the composition of buyers. We run two versions of this counterfactual: one in which we keep the share of non-occupants among potential buyers with $\lambda = 0.03$ equal to its baseline, and one in which we change this ratio to the non-occupant share in the whole baseline population. The second version controls for the non-occupant share as we alter the λ distribution.

We perform a similar pair of counterfactual exercises to measure the effect of removing non-occupant buyers. The first counterfactual sets the non-occupant shares, β_0^λ, to zero, and then scales up the occupant shares, β_1^λ, so that they sum to one. This method skews the λ distribution toward long-term buyers because occupants have longer horizons than non-occupants (Table 6). Therefore, we explore a second counterfactual in which we maintain the original λ distribution while eliminating non-occupants. We continue to set each β_0^λ to zero, but now we update β_1^λ to the baseline λ share among all potential buyers.

Table 7 reports key outcomes from the impulse responses under the baseline and each of these four counterfactuals. In the counterfactuals with only long-term buyers, the price bust nearly disappears, the volume boom is half its baseline size, and sale probabilities rise less. Inventories fall more during the boom and attain a smaller level at the end of the quiet.[28] Therefore, eliminating short-term buyers prevents the model from matching the key aggregate facts (Figures 1 and 3). Interestingly, short-term volume still rises more than total volume, even though the composition of buyers remains the same throughout the boom (by construction). This pattern is a manifestation of the mechanical increase in short-term volume that we quantify in Section 5.2.

We obtain similar results in the first counterfactual with only occupants: the price bust, volume boom, rise in sale probabilities, and end-of-quiet listings become significantly smaller. However, when we adjust the λ distribution in the last counterfactual, eliminating non-occupants fails to attenuate the cycle. In fact, the cycle outcomes grow in this scenario. Evidently, non-occupants amplify the housing cycle, but only because many of them have short horizons. Long-term non-occupants fail to amplify the cycle and may even dampen it.

These results speak to the finding in Tables 2 and 3 that a short-holding-period boom more robustly predicts price booms and busts than does a non-occupant volume boom. Our

---

[28]These counterfactuals do a better job matching inventory levels during the bust, which reach 1.6 log points, a higher level than the baseline. In the baseline, new listings fall sharply during the bust because short-term buyers exit the market (Panel F of Figure 9). Thus, the baseline does a better job matching listing behavior in the boom and quiet than in the bust.

findings are also consistent with Gao, Sockin and Xiong (2019), who find that non-occupants amplify the housing bust, as that paper does not look separately at long-term versus short-term non-occupants. Chinco and Mayer (2015) find a stronger effect of out-of-town than local non-occupants buyers on subsequent house price growth. This finding is consistent with our results if the out-of-town buyers have shorter horizons than local ones. Finally, our results echo Nathanson and Zwick (2018), who theoretically predict larger house price booms in cities with a greater share of non-occupant buyers when those buyers disagree about future house prices and when the housing stock is fixed. Static disagreement in that model functions similarly to how, in this model, variation in horizons interacts with extrapolative expectations to generate heterogeneous expected returns.

To gauge the role of various model ingredients for producing the results, Internet Appendix IV simulates a rational version of our model and a version with Walrasian market clearing. The rational model dampens or eliminates most cycle dynamics, except for the disproportionate non-occupant volume boom. In the Walrasian model, prices and volume continue to go through a boom and bust cycle, but volume peaks after prices, thus eliminating the quiet. Therefore, departing from rationality seems necessary to produce the price and quantity dynamics in the data, and relaxing Walrasian market clearing appears necessary to generate the quiet.

9   Final Remarks

Our paper raises additional lines of inquiry within the housing market. We have argued, theoretically and empirically, that short-term investors play a crucial role in the housing cycle. Do the expansions in credit that typically accompany housing booms appeal disproportionately to short-term investors? Barlevy and Fisher (2011) document a strong correlation across U.S. metropolitan areas between the size of the 2000s house price boom and the take-up of interest-only mortgages. These mortgages back-load payments by deferring principal repayment for some amount of time and thus might appeal especially to buyers who expect to resell quickly. The targeting of credit expansions to short-term buyers might explain the amplification effects of credit availability on real estate booms documented by Favara and Imbs (2015), Di Maggio and Kermani (2017), and Rajan and Ramcharan (2015). Mian and Sufi (2019) explore this channel in contemporaneous work.

A second line of inquiry within housing concerns tax policy. The capital gains tax discourages housing speculation by lowering expected after-tax capital gains. However, it discourages productive residential investment as well. Is this tax optimal, and if not, what type of tax policy would be better? Many economists have analyzed or proposed transaction taxes (Tobin, 1978; Stiglitz, 1989; Summers and Summers, 1989; Dávila, 2015). It is unclear whether these taxes would particularly discourage short-term investors, given that the incidence of this tax might fall more on buyers than sellers.

A third research question involves new construction, which is absent from our model. In a static model, Nathanson and Zwick (2018) predict that undeveloped land amplifies house price booms by facilitating speculation by developers. Developers have short investment horizons because the time from land purchase to home sale ranges from a few months to a few years. Moreover, because developers do not receive flow utility, their payoffs resemble those of the non-occupants in our model. Therefore, adding construction to the model in this paper might further clarify the role of land markets and new construction in housing cycles.

Although this paper focuses on the housing market, many of the patterns we study appear in other asset markets as well. Several famous bubbles involve large movements in transaction volume (Cochrane, 2011). The lead–lag relation between prices and volume holds, albeit at different frequencies, in four other boom-bust episodes shown in Figure 10: the 1995–2005 market in technology stocks, the 1985–1995 Japanese stock market, the experimental bubbles studied by Smith, Suchanek and Williams (1988), and the 1985–1995 bubble in Postwar art. Short-horizon trading was prevalent during the technology boom (Cochrane, 2002; Ofek and Richardson, 2003). Even outside of bubbles, stock market volume increases following high returns and predicts negative subsequent returns (Lee and Swaminathan, 2000; Jones, 2002; Statman, Thorley and Vorkink, 2006; Griffin, Nardari and Stulz, 2007).

Cutler, Poterba and Summers (1991) document price dynamics such as momentum and mean reversion in many asset classes. They conclude the generality of these patterns suggests that inherent features of the speculative process likely explain them. Can our model of speculation explain the joint dynamics of prices and volume outside the housing market? In the Walrasian variant of our model in Internet Appendix IV, we do not generate a consistent lead–lag relation, but we find that some movers do not sell when the expected growth rate is very large. We conjecture that it may be possible to generate price and volume outcomes

41

that resemble the quiet but in a market design that more closely resembles the Walrasian benchmark. We hope that future work will investigate the striking similarity of volume dynamics in other markets.

References

Adelino, Manuel, Antoinette Schoar, and Felipe Severino. 2016. "Loan Originations and Defaults in the Mortgage Crisis: The Role of the Middle Class." *Review of Financial Studies*, 29(7): 1635–1670.

Aliprantis, Charalambos D., and Kim C. Border. 2006. *Infinite Dimensional Analysis: A Hitchhiker's Guide.* Springer.

Anenberg, Elliot, and Patrick Bayer. 2013. "Endogenous Sources of Volatility in Housing Markets: The Joint Buyer-Seller Problem." National Bureau of Economic Research.

Anenberg, Elliot, and Patrick Bayer. 2020. "Endogenous Sources of Volatility in Housing Markets: The Joint Buyer-Seller Problem." Forthcoming, *International Economic Review*.

Armona, Luis, Andreas Fuster, and Basit Zafar. 2016. "Home Price Expectations and Behavior: Evidence from a Randomized Information Experiment." Working Paper.

Bailey, Michael, Ruiqing Cao, Theresa Kuchler, and Johannes Stroebel. 2016. "Social Networks and Housing Markets." Working Paper.

Barberis, Nicholas, Robin Greenwood, Lawrence Jin, and Andrei Shleifer. 2015. "X-CAPM: An Extrapolative Capital Asset Pricing Model." *Journal of Financial Economics*, 115(1): 1–24.

Barberis, Nicholas, Robin Greenwood, Lawrence Jin, and Andrei Shleifer. 2018. "Extrapolation and bubbles." *Journal of Financial Economics*, 129(2): 203–227.

Barlevy, Gadi, and Jonas D.M. Fisher. 2011. "Mortgage Choices and Housing Speculation." Working Paper, Federal Reserve Bank of Chicago.

Bayer, Patrick, Christopher Geissler, Kyle Mangum, and James W. Roberts. 2020. "Speculators and Middlemen: The Strategy and Performance of Investors in the Housing Market." *Review of Financial Studies*.

Bayer, Patrick, Kyle Mangum, and James W. Roberts. 2016. "Speculative Fever: Investor Contagion in the Housing Bubble." NBER Working Paper 22065.

Bhutta, Neil. 2015. "The Ins and Outs of Mortgage Debt during the Housing Boom and Bust." *Journal of Monetary Economics*, 76: 284–298.

Burnside, Craig, Martin Eichenbaum, and Sergio Rebelo. 2016. "Understanding Booms and Busts in Housing Markets." *Journal of Political Economy*, 124(4): 1088–1147.

Case, Karl E. 2008. "The Central Role of Home Prices in the Current Financial Crisis: How Will the Market Clear?" *Brookings Papers on Economic Activity*, 2: 161–193.

Case, Karl E., Robert J. Shiller, and Anne K. Thompson. 2012. "What Have They Been Thinking? Homebuyer Behavior in Hot and Cold Markets." *Brookings Papers on Economic Activity*, Fall: 265–298.

Charles, Kerwin Kofi, Erik Hurst, and Matthew J. Notowidigdo. 2018. "Housing Booms and Busts, Labor Maket Opportunities, and College Attendance." *American Economic Review*, 108(10): 2947–94.

Cheng, Ing-Haw, Sahil Raina, and Wei Xiong. 2014. "Wall Street and the Housing Bubble." *American Economic Review*, 104(9): 2797–2829.

Chinco, Alex, and Christopher Mayer. 2015. "Misinformed Speculators and Mispricing in the Housing Market." *Review of Financial Studies*, 29(2): 486–522.

Choi, Hyun-Soo, Harrison G. Hong, and José Scheinkman. 2014. "Speculating on Home Improvements." *Journal of Financial Economics*, 111(3): 609–624.

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

Cochrane, John H. 2002. "Stocks as Money: Convenience Yield and the Tech-Stock Bubble." Working Paper.

Cochrane, John H. 2011. "Discount Rates." *Journal of Finance*, 66: 1047–1108.

Cutler, David M., James M. Poterba, and Lawrence H. Summers. 1990. "Speculative Dynamics and the Role of Feedback Traders." *American Economic Review, Papers and Proceedings*, 80(2): 63–68.

Cutler, David M., James M. Poterba, and Lawrence H. Summers. 1991. "Speculative Dynamics." *Review of Economic Studies*, 58(3): 529–546.

Dávila, Eduardo. 2015. "Optimal Financial Transactions Taxes." Working Paper, NYU Stern.

De Long, J. Bradford, Andrei Shleifer, Lawrence H. Summers, and Robert J. Waldmann. 1990. "Positive Feedback Investment Strategies and Destabilizing Rational Speculation." *Journal of Finance*, 45(2): 379–395.

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

Di Maggio, Marco, and Amir Kermani. 2017. "Credit-Induced Boom and Bust." *The Review of Financial Studies*, 30(11): 3711–3758.

Edelstein, Robert, and Wenlan Qian. 2014. "Short-Term Buyers and Housing Market Dynamics." *Journal of Real Estate Finance and Economics*, 49: 654–689.

Favara, Giovanni, and Jean Imbs. 2015. "Credit Supply and the Price of Housing." *American Economic Review*, 105(3): 958–992.

Foote, Christopher L., Kristopher S. Gerardi, and Paul S. Willen. 2012. "Why Did So Many People Make So Many *Ex Post* Bad Decisions? The Causes of the Foreclosure Crisis." *Federal Reserve Bank of Boston Public Policy Discussion Papers*, 12(2).

Frankel, Jeffrey A., and Kenneth A. Froot. 1986. "Understanding the U.S. Dollar in the Eighties: The Expectations of Chartists and Fundamentalists." *Economic Record*, 24–38.

Gao, Zhenyu, Michael Sockin, and Wei Xiong. 2019. "Economic Consequences of Housing Speculation." NBER Working Paper 26457.

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

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

Glaeser, Edward L. 2013. "A Nation of Gamblers: Real Estate Speculation and American History." *American Economic Review*, 103(3): 1–42.

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

Griffin, John M., Federico Nardari, and René M. Stulz. 2007. "Do Investors Trade More When Stocks Have Performed Well? Evidence from 46 Countries." *Review of Financial Studies*, 20(3): 905–951.

Griffin, John M, Samuel Kruger, and Gonzalo Maturana. 2020. "What drove the 2003–2006 house price boom and subsequent collapse? Disentangling competing explanations." *Journal of Financial Economics*.

Guren, Adam. 2014. "The Causes and Consequences of House Price Momentum." Working paper, Harvard University.

Guren, Adam. 2018. "House Price Momentum and Strategic Complementarity." Forthcoming, *Journal of Political Economy*.

Guren, Adam M, and Timothy J McQuade. 2020. "How Do Foreclosures Exacerbate Housing Downturns?" *Review of Economic Studies*, 87(3): 1331–1364.

Han, Lu, and William C. Strange. 2015. "The Microstructure of Housing Markets: Search, Bargaining, and Brokerage." In *Handbook of Regional and Urban Economics*. Vol. 5B, , ed. Gilles Duranton, J. Vernon Henderson and William C. Strange, Chapter 13, 813–886. Elsevier.

Harrison, J. Michael, and David M. Kreps. 1978. "Speculative Investor Behavior in a Stock Market with Heterogeneous Expectations." *Quarterly Journal of Economics*, 92(2): 323–336.

Haughwout, Andrew, Donghoon Lee, Joseph Tracy, and Wilbert van der Klaauw. 2011. "Real Estate Investors, the Leverage Cycle, and the Housing Market Crisis." Federal Reserve Bank of New York Staff Report no. 514.

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.

Hedlund, Aaron. 2016. "The Cyclical Dynamics of Illiquid Housing, Debt, and Foreclosures." *Quantitative Economics*, 7: 289–328.

Hong, Harrison, and Jeremy C. Stein. 1999. "A Unified Theory of Underreaction, Momentum Trading, and Overreaction in Asset Markets." *Journal of Finance*, 54(6): 2143–2184.

Jones, Charles M. 2002. "A Century of Stock Market Liquidity and Trading Costs." Working Paper.

Kelley, John L. 1955. *General Topology.* Springer-Verlag. pg. 150.

Lee, Charles, and Bhaskaran Swaminathan. 2000. "Price Momentum and Trading Volume." *Journal of Finance*, 55(5): 2017–2069.

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

Liao, Jingchi, and Cameron Peng. 2018. "Price and Volume Dynamics in Bubbles." Working Paper.

Mayer, Christopher. 2011. "Housing Bubbles: A Survey." *Annual Review of Economics*, 3: 559–577.

Mayer, Christopher, Karen Pence, and Shane M. Sherlund. 2009. "The Rise in Mortgage Defaults." *Journal of Economic Perspectives*, 23(1): 27–50.

Mian, Atif, Amir Sufi, and Francesco Trebbi. 2015. "Foreclosures, House Prices, and the Real Economy." *Journal of Finance*, 70(6): 2587–2634.

Mian, Atif, and Amir Sufi. 2019. "Credit supply and housing speculation." NBER Working Paper 24823.

Mian, Atif, Kamalesh Rao, and Amir Sufi. 2013. "Household Balance Sheets, Consumption, and the Economic Slump." *Quarterly Journal of Economics*, 128(4): 1687–1726.

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

Ngai, L. Rachel, and Kevin D. Sheedy. 2016. "The Decision to Move House and Aggregate Housing-Market Dynamics." Working Paper.

Ofek, Eli, and Matthew Richardson. 2003. "Dotcom Mania: The Rise and Fall of Internet Stock Prices." *Journal of Finance*, 58(3): 1113–1137.

Ortalo-Magné, François, and Sven Rady. 2006. "Housing Market Dynamics: On the Contribution of Income Shocks and Credit Constraints." *Review of Economic Studies*, 73(2): 459–485.

Penasse, Julien, and Luc Renneboog. 2016. "Bubbles and Trading Frenzies: Evidence from the Art Market." Working Paper.

Piazzesi, Monika, and Martin Schneider. 2009. "Momentum Traders in the Housing Market: Survey Evidence and a Search Model." *American Economic Review*, 99(2): 406–411.

Piskorski, Tomasz, and Amit Seru. 2018. "Debt Relief and Slow Recovery: A Decade after Lehman." National Bureau of Economic Research Working Paper 25403.

Rajan, Raghuram, and Rodney Ramcharan. 2015. "The Anatomy of a Credit Crisis: The Boom and Bust in Farm Land Prices in The United States in the 1920s." *American Economic Review*, 105(4): 1439–1477.

Rognlie, Matthew, Andrei Shleifer, and Alp Simsek. 2017. "Investment Hangover and the Great Recession." *American Economic Journal: Macroeconomics*, 10(2): 113–53.

Rouwenhorst, K. Geert. 1995. "Asset Pricing Implications of Equilibrium Business Cycle Models." In *Frontiers of Business Cycle Research.* , ed. Thomas F. Cooley, 294–330. Princeton, NJ:Princeton University Press.

Scheinkman, Jose A., and Wei Xiong. 2003. "Overconfidence and Speculative Bubbles." *Journal of Political Economy*, 111(6): 1183–1220.

Smith, Vernon L., Gerry L. Suchanek, and Arlington W. Williams. 1988. "Bubbles, Crashes, and Endogenous Expectations in Experimental Spot Asset Markets." *Econometrica*, 56(5): 1119–1151.

Statman, Meir, Steven Thorley, and Keith Vorkink. 2006. "Investor Overconfidence and Trading Volume." *Review of Financial Studies*, 19(4): 1531–1565.

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

Stiglitz, Joseph E. 1989. "Using Tax Policy To Curb Speculative Short-Term Trading." *Journal of Financial Services Research*, 3: 101–115.

Stokey, Nancy L., Robert E. Lucas, Jr., and Edward C. Prescott. 1989. *Recursive Methods in Economic Dynamics.* Harvard University Press.

Summers, Lawrence H., and Victoria P. Summers. 1989. "When Financial Markets Work Too Well: A Cautious Case For a Securities Transaction Tax." *Journal of Financial Services Research*, 3: 261–286.

Tobin, James. 1978. "A Proposal for International Monetary Reform." *Eastern Economic Journal*, 4(3): 153–159.

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

FIGURE 1
The Dynamics of Prices and Volume
Panel A. National
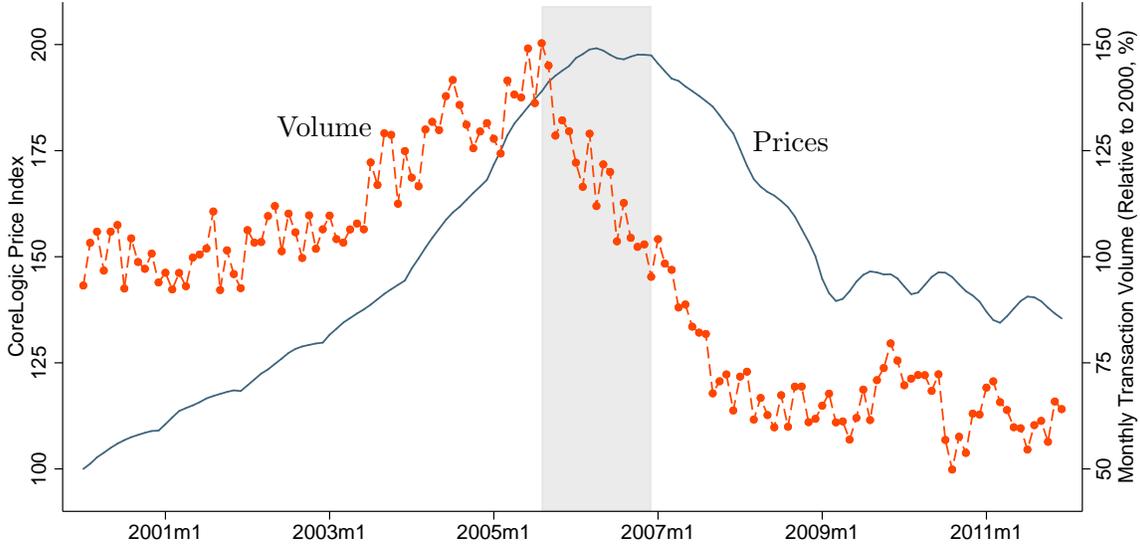
Panel B. Phoenix, AZ
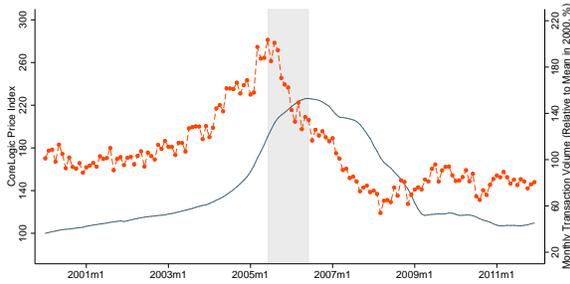
Panel C. Las Vegas, NV
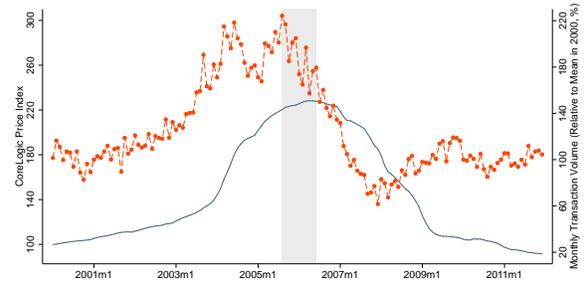
Panel D. Orlando, FL
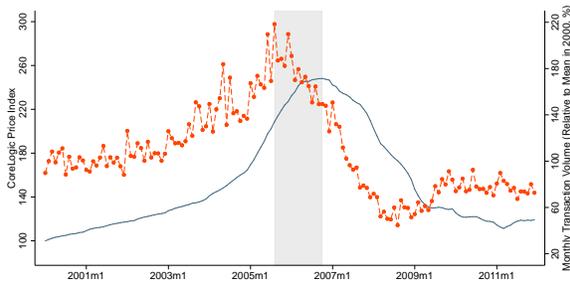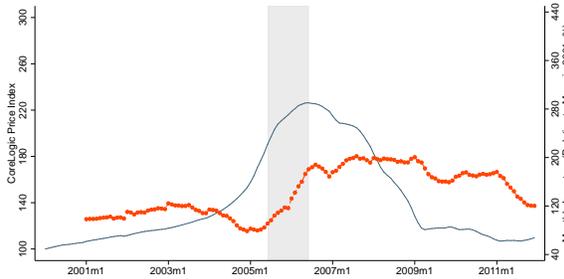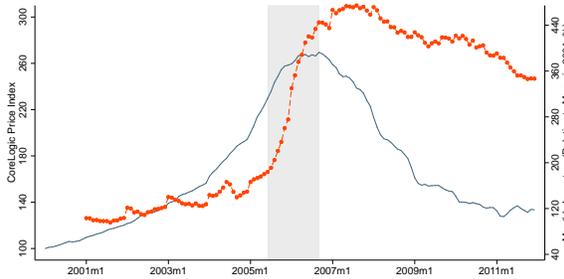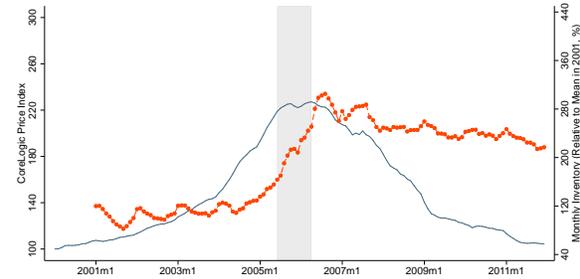
Panel E. Bakersfield, CA

Notes: This figure displays the dynamic relation between prices and volume in the U.S. housing market between 2000 and 2011. Panel A plots monthly prices and sales volume at the aggregate level. Panels B through E plot analogous series for a set of cities that represent regions with the largest boom–bust cycles during this time: Phoenix, AZ; Las Vegas, NV; Orlando, FL; and Bakersfield, CA. Monthly price index information comes from CoreLogic and monthly sales volume is based on aggregated transaction data from CoreLogic for 115 MSAs representing 48% of the U.S. housing stock. We apply a calendar-month seasonal adjustment for volume. Shaded regions denote the *quiet*, defined as the period between the peak of volume and the last peak of prices before their pronounced decline.

49

FIGURE 2

The Lead–Lag Relationship between Prices and Volume

Notes: This figure shows that the correlation between prices and lagged volume is robust across MSAs and maximized at a positive lag of 24 months. We regress the demeaned log of prices on seasonally adjusted lagged volume divided by the 2000 housing stock following equation (1) for each lag from -12 months to 48 months and plot the implied correlation and its 95% confidence interval calculated using standard errors that are clustered by month. The implied correlation equals $\beta_\tau \, \text{std}(v_{i,t-\tau})/\text{std}(p_{i,t})$, where $v_{i,t-\tau}$ and $p_{i,t}$ are the demeaned regressors.

FIGURE 3
The Dynamics of Prices and Inventories
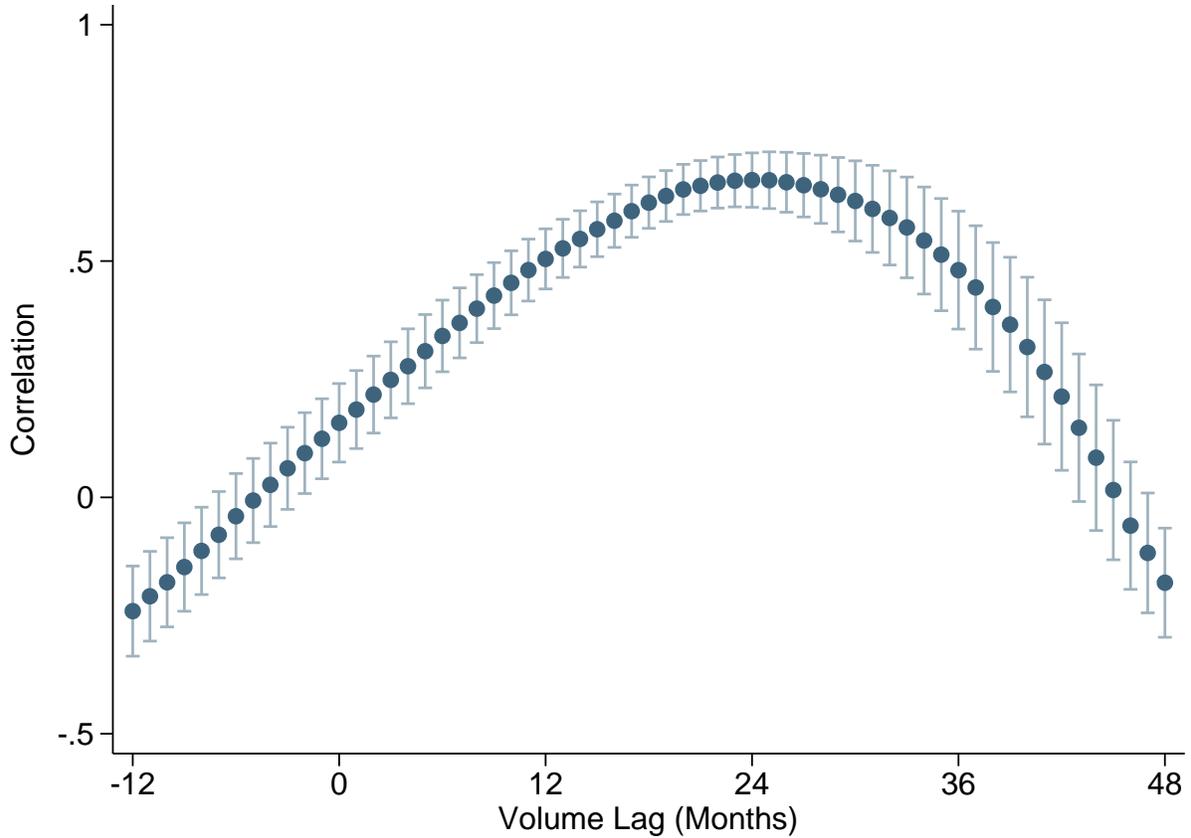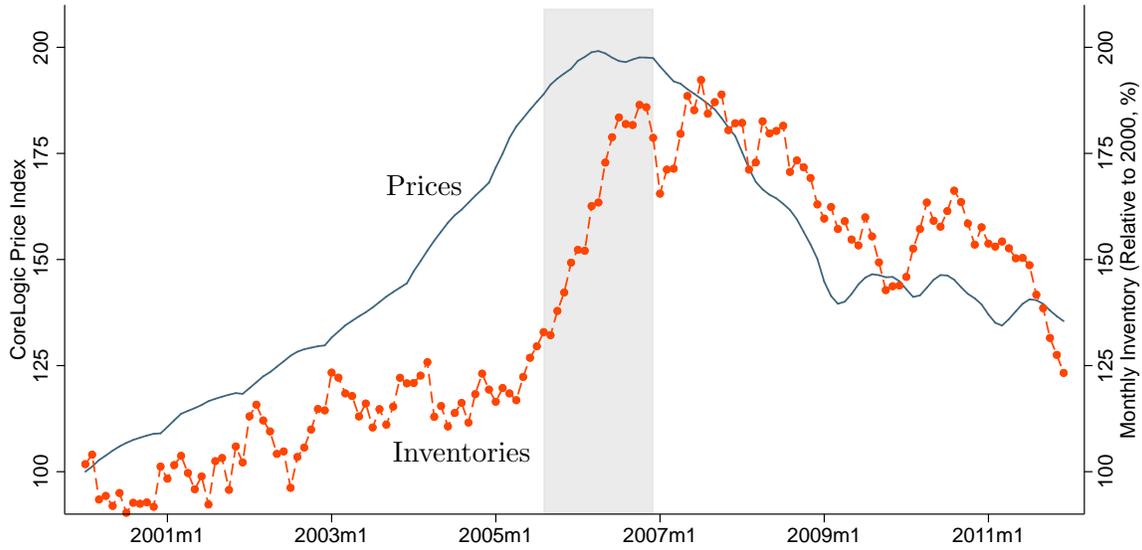Panel A. National

Notes: This figure displays the dynamic relation between prices and inventory in the U.S. housing market between 2000 and 2011. Panel A plots monthly prices and the inventory of listings at the aggregate level. Panels B through E plot analogous series for a set of cities that represent regions with the largest boom–bust cycles during this time: Phoenix, AZ; Reno, NV; Daytona Beach, FL; and Bakersfield, CA. Aggregate inventory information comes from the National Association of Realtors, which are available starting in 2000. Our MSA-level inventory data are available for these cities starting in 2001. Monthly price index information comes from CoreLogic and monthly inventory by MSA is based on aggregated data from CoreLogic for 57 of the 115 MSAs in our main sample for which listings data are available. We apply a calendar-month seasonal adjustment for inventories. Shaded regions denote the *quiet*, defined as the period between the peak of volume and the last peak of prices before their pronounced decline.

51

FIGURE 4

Normalized Aggregate Volume by Transaction Type

Volume (000s)	2000	2005	2010
Short$_{S1}$	510	940	150
Existing$_{S1}$	2,130	2,880	930
Total$_{S1}$	2,730	3,820	1,150
Non-Occupant$_{S2}$	510	1,030	290
Total$_{S2}$	2,310	3,290	990

- - ■ - - Short-Holding-Period Volume   ● Total Volume (Excluding New Construction)

— ◆ — Non-Occupant Volume   ▲ Total Volume (Including New Construction)
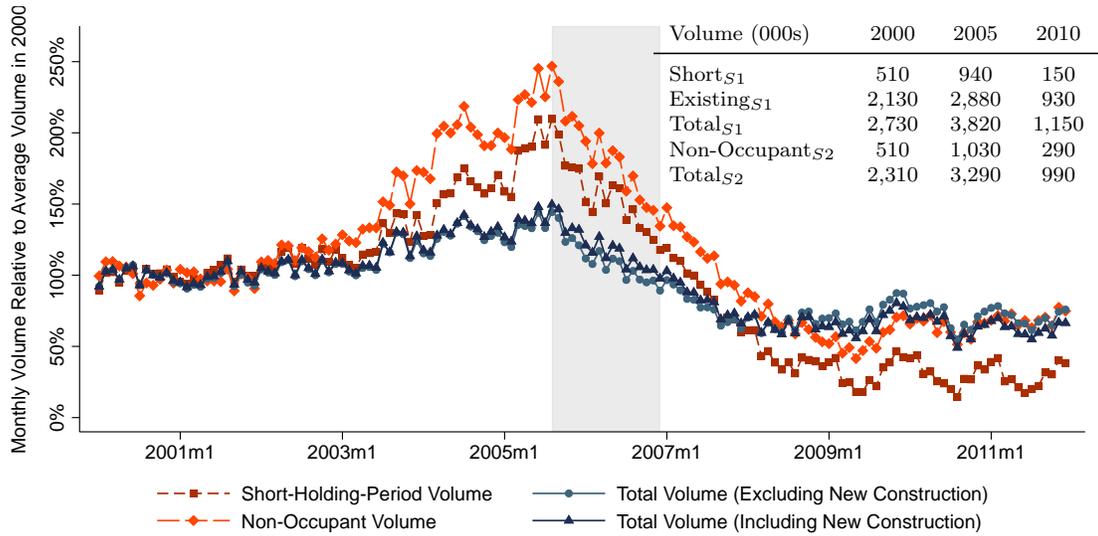
Notes: This figure plots monthly aggregate time series for total transaction volume (navy triangles), total volume excluding new construction (blue circles), short-holding-period volume (red squares), and non-occupant volume (orange diamonds) between 2000 and 2011. All series exclude foreclosures. The non-occupant volume series only includes observations from the 102 MSAs for which we can consistently identify these transactions; the other series include observations for all 115 MSAs. Each series is separately normalized relative to its average value in the year 2000 and seasonally adjusted by removing calendar-month fixed effects. For reference, the raw counts of each type of transaction in the years 2000, 2005, and 2010 are reported in the upper right corner of the figure. In the table, $S1$ refers to the short-holding-period sample of 115 MSAs and $S2$ refers to the non-occupant sample of 102 MSAs.

FIGURE 5

Short-Holding-Period, Non-Occupant, and Total Volume Growth Across MSAs

Panel A. Total Volume Versus Volume by Holding Period

Holding Period < 3 Years   Holding Period ≥ 3 Years

Panel B. Total Volume Versus Volume by Occupancy Status

Non-Occupant Buyers   Occupant Buyers

Panel C. Role of Short Volume in Total Volume Growth

slope = 0.36
slope = 0.30

Including New Construction   Excluding New Construction

Panel D. Role of Non-Occupant Volume in Total Volume Growth

slope = 0.54
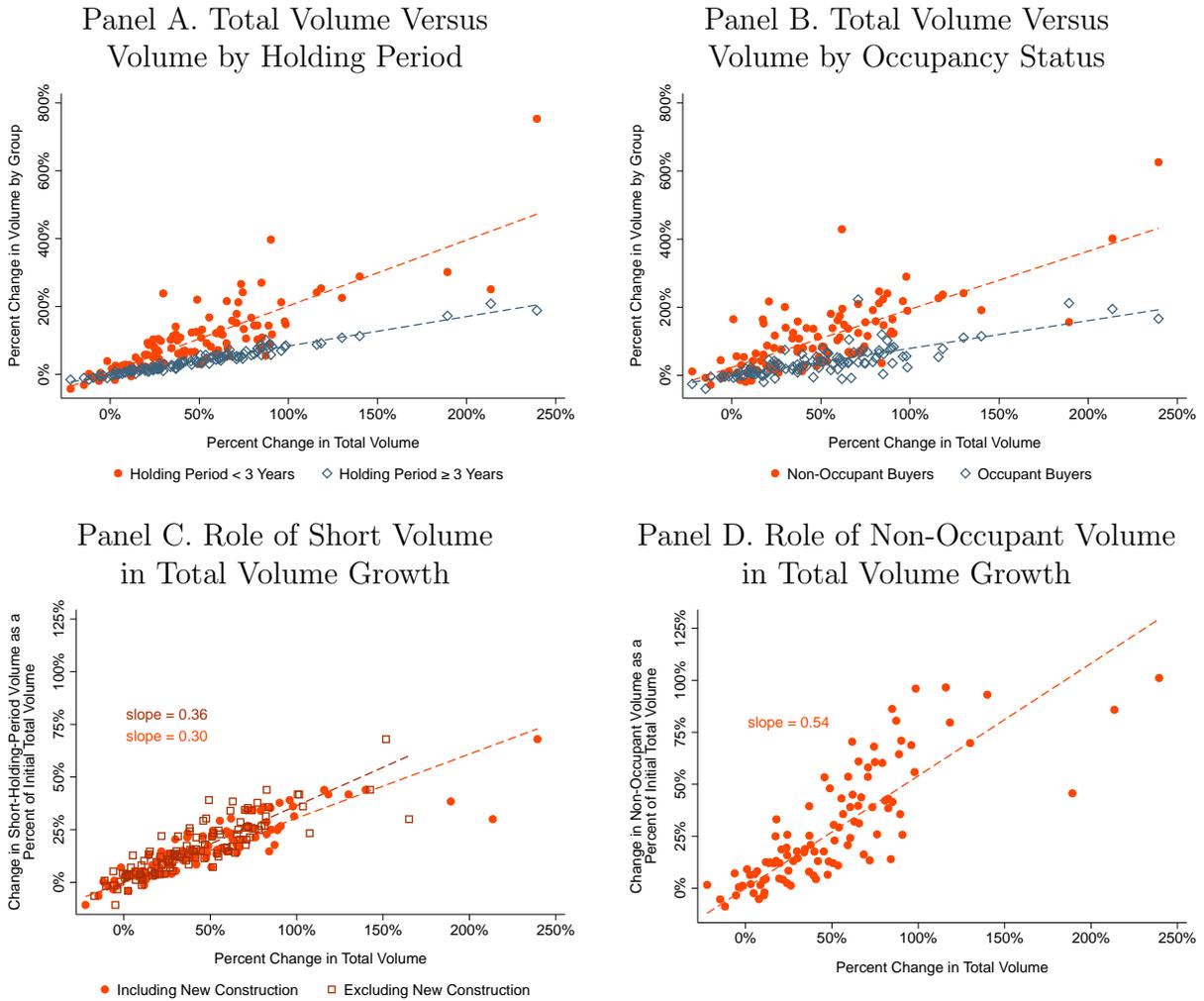
Notes: This figure illustrates the quantitative importance of short-holding-period and non-occupant volume in accounting for the increase in total volume across MSAs between 2000 and 2005. The top two panels present MSA-level scatter plots of the percent change in total volume from 2000 to 2005 versus the percent change in volume for short and long holding periods (Panel A) and the percent change in volume for occupant and non-occupant buyers (Panel B). The bottom two panels show that the growth in short-holding-period and non-occupant volume were quantitatively important components of the growth in total volume across MSAs. For each MSA, we plot the change in short-holding-period volume (Panel C) and non-occupant volume (Panel D) divided by initial total volume on the y-axis against the percent change in total volume on the x-axis. Because short-holding-period volume is based on the holding period of the seller and therefore cannot, by definition, include sales of newly constructed homes, Panel C also includes a version of the scatter plot that excludes new construction from total volume for reference.

FIGURE 6

The Flow of Listings for Short-Holding-Period Buyers

Listings (000s)	2003	2007	2010
Short	280	590	170
Total	1,170	1,730	1,380

Notes: This figure illustrates the time variation in propensities to list among recent buyers versus all buyers between 2000 and 2011 in the U.S. We link listings micro data to transaction data at the property level to identify short-holding-period listings. We plot monthly aggregate time series for total listings (blue circles) and short-holding-period listings (red squares), defined as a listing where the previous sale occurred within the past three years. The series include observations for the 57 MSAs in our listings sample. Each series is separately normalized relative to its average value in the year 2003 and seasonally adjusted by removing calendar-month fixed effects. For reference, the raw counts of each type of listing in the years 2003, 2007, and 2010 are also reported in the upper right corner of the figure.

FIGURE 7
Expected Holding Times of Homebuyers, 2008–2015

Panel A. Response Heterogeneity by Property Type
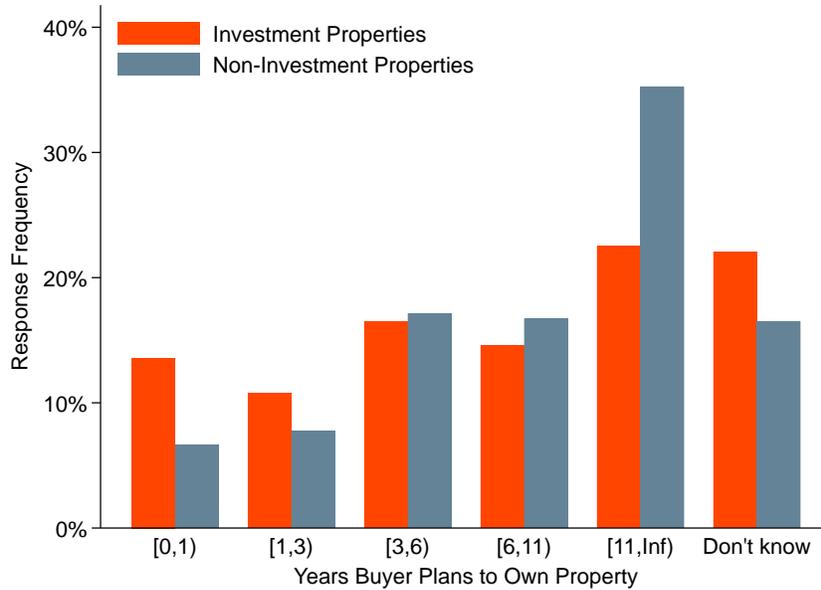
Panel B. Short-Term Buyers and Recent House Price Growth

Notes: This figure presents evidence on heterogeneity in expected holding times among recent homebuyers and the correlation between expected holding times and recent price changes from the National Association of Realtors' Investment and Vacation Home Buyers Survey. Panel A plots the response frequency averaged equally over each survey year from 2008 to 2015. In Panel B, "annual house price growth" equals the average across that year's four quarters of the log change in the all-transactions FHFA U.S. house price index from four quarters ago, and "short-term buyer share" equals the share of respondents other than those reporting "don't know" who report an expected horizon of less than three years. We reclassify buyers who have already sold their properties by the time of the survey as having an expected holding time in [0,1). We use the FHFA index here because, unlike the CoreLogic indices used elsewhere in the paper, the FHFA house price index covers the period 2015–2016.

FIGURE 8

Buying Cutoffs for Different Expected Growth Rates
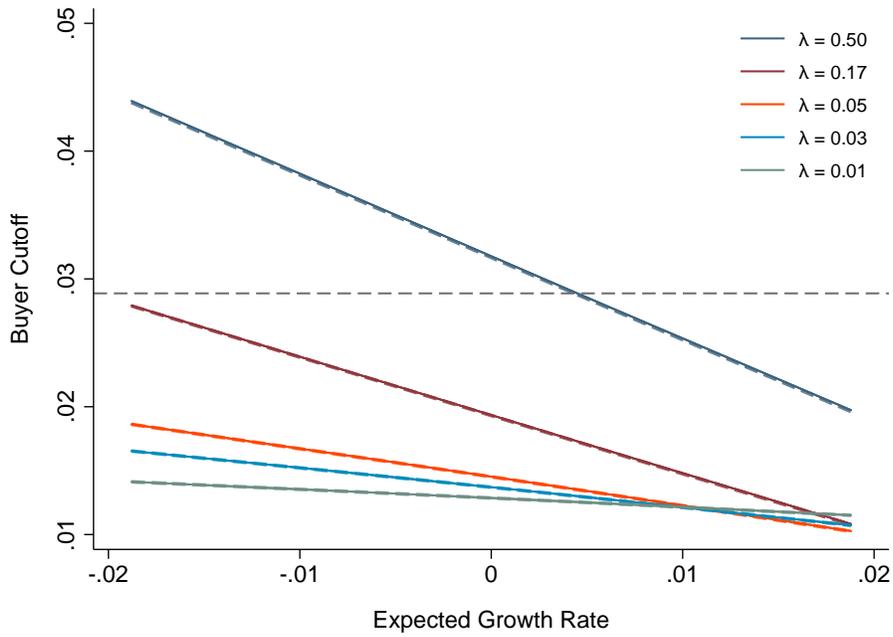
Notes: The buying cutoff, $\kappa_n^\lambda(\hat{g}_t)$, determines how large a potential buyer's flow utility must be relative to the price of a house for her to decide to buy. It depends on the potential buyer's quarterly moving hazard, λ, her occupancy type, n, and the current expected quarterly growth rate of the demand process, $\hat{g}_t$. We plot values of these functions for the λ values in our calibration, which appear in the legend. Solid lines correspond to occupants ($n = 1$); dashed lines correspond to non-occupants ($n = 0$). The horizontal grey dashed line gives κ, which agents mistakenly believe is the time-invariant buying cutoff for other potential buyers.

FIGURE 9
Impulse Responses
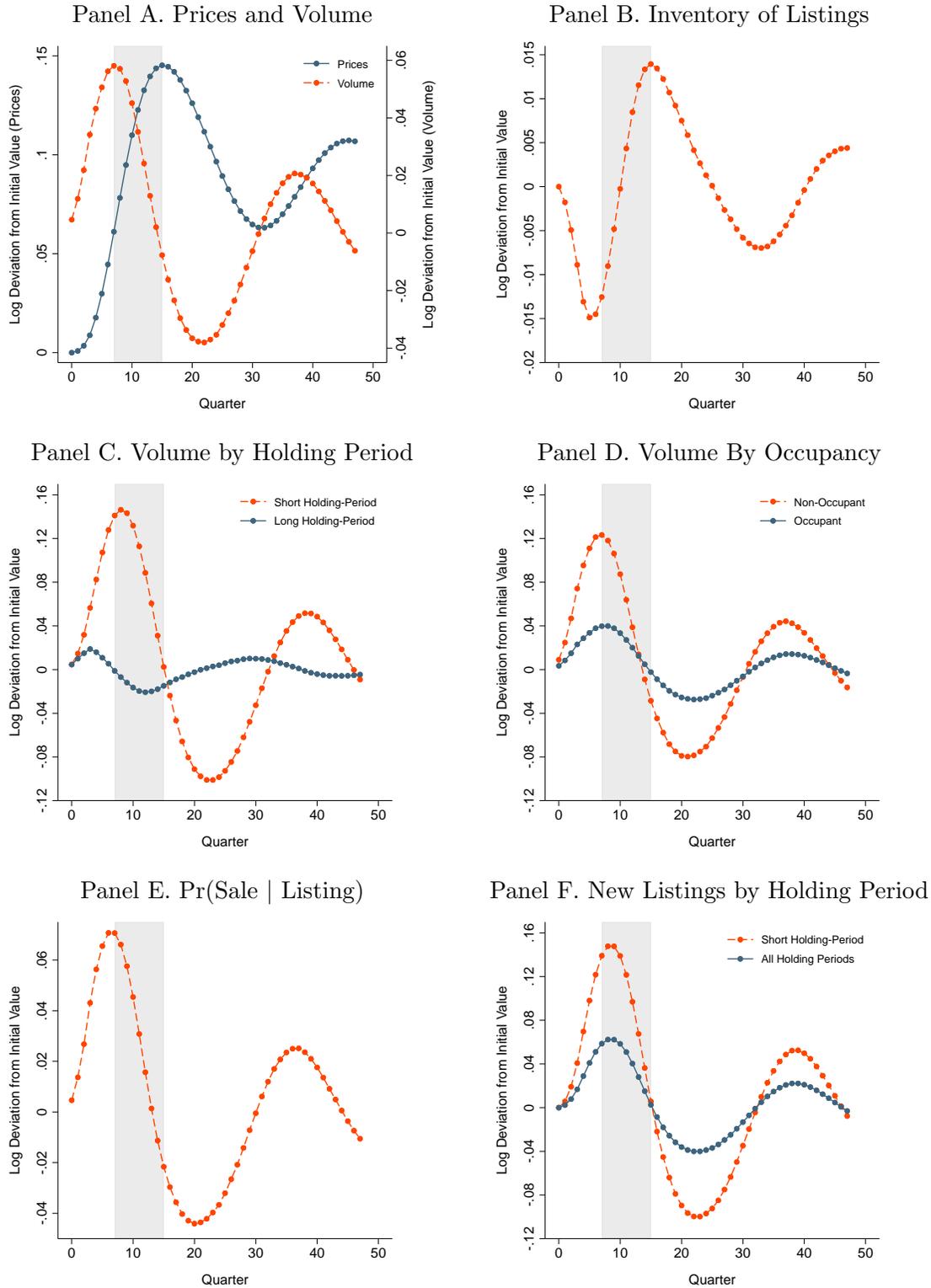
Panel A. Prices and Volume

Panel B. Inventory of Listings

Panel C. Volume by Holding Period

Panel D. Volume By Occupancy

Panel E. Pr(Sale | Listing)

Panel F. New Listings by Holding Period

Notes: Impulse responses are average differences between log outcomes in control simulations and treatment simulations, in which a 2-standard-deviation shock to ϵ_t^g (the demand growth innovation) occurs in quarters 0 through 3. The shaded grey area denotes the beginning and end of the quiet. A short holding period is defined as less than or equal to 12 quarters and a long holding period is defined as greater than 12 quarters.

57

FIGURE 10

The Joint Dynamics of Prices and Volume

Panel A. U.S. Equities, Tech (1995–2005)
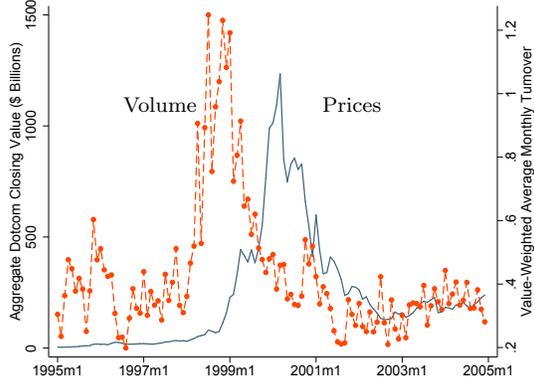
Panel B. Japan Equities (1985–1995)
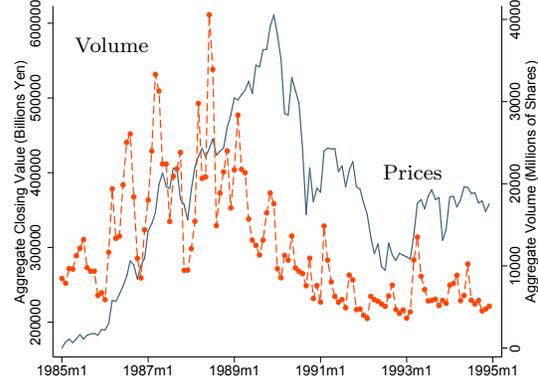
Panel C. Experimental Markets, SSW (1988)
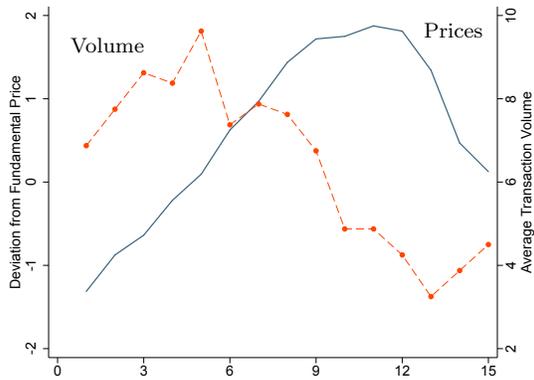
Panel D. Art Market (1985–1995)
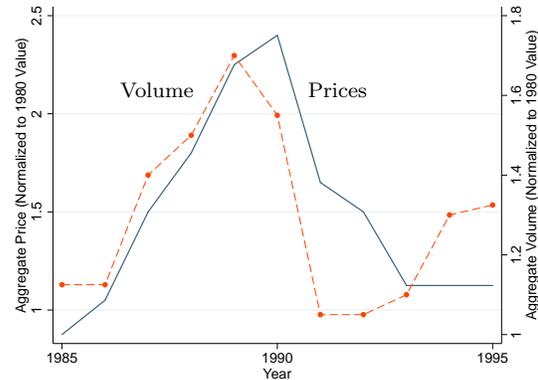
Notes: These figures display the dynamic relation between prices and transaction volume for four distinct bubble episodes: the 1995–2005 market in technology stocks (Panel A), the 1985–1995 Japanese stock market (Panel B), the bubbles in experimental asset markets (Panel C), and the 1985–1995 bubble in the Postwar art market (Panel D). Panel A data come from CRSP and cover the Dotcom sample in Ofek and Richardson (2003). For prices, we plot aggregate Dotcom market capitalization. For volume, we plot average monthly turnover (shares traded/shares outstanding), weighted by market cap. Panel B data come from the Tokyo stock exchange online archive and cover all first- and second-tier (i.e., large and micro-cap) stocks. For volume, we plot total shares traded per month (shares-outstanding data are not available). For prices, we plot aggregate market capitalization. Panel C data were manually entered from the published Smith, Suchanek and Williams (1988) manuscript and cover all eight experiments that include a price boom and bust (IDs are 16, 17, 18, 26, 124xxf, 39xsf, 41f, 36xx). For prices, we plot average deviations from fundamental value. For volume, we plot the average number of trades. Panel D data come from Figure 1 of the working paper version of Penasse and Renneboog (2016) and cover aggregate art prices and transaction volume from auction houses for paintings and works on paper for more than 10,000 artists.

TABLE 1
All-Cash Buyer Shares

	Transaction-Level	MSA-Level			
	All Months	All Months	Boom	Quiet	Bust
Short Buyers	0.29	0.38	0.29	0.28	0.52
		(0.21)	(0.16)	(0.17)	(0.20)
Non-Occupant Buyers	0.38	0.41	0.36	0.32	0.50
		(0.18)	(0.15)	(0.14)	(0.18)
All Buyers	0.20	0.25	0.22	0.20	0.30
		(0.16)	(0.15)	(0.14)	(0.16)

Notes: This table presents statistics on the share of buyers of various types who purchased their homes without the use of a mortgage ("all-cash buyers"). In column 1, the all-cash buyer share is measured at the transaction level and includes all transactions recorded between January 2000 and December of 2011 from the CoreLogic deeds records described in Section 2.1. The first row includes only transactions by homebuyers who are observed to have sold the home within three years of purchase. The second row includes only non-occupant buyers. The third row includes all buyers. In columns 2–5, all-cash buyer shares are first calculated at the MSA-by-month level and then averaged across MSA-months within a given time period. The standard deviation of these MSA-month means is reported in parentheses for reference. Column 2 includes all MSA-months between January 2000 and December 2011. Column 3 includes only MSA-months between January 2000 and August 2005. Column 4 includes only MSA-months between August 2005 and December 2006. Column 5 includes only MSA-months between December 2006 and December 2011. All statistics are calculated in the full sample of 115 MSAs with the exception of those for non-occupants, which are calculated in the sample of 102 MSAs with valid non-occupancy data.

TABLE 2
Speculative Booms and Housing Market Outcomes

Panel A. MSA-Level Prices

	Price Boom			Price Bust		
Short-Volume Boom	1.930***		3.104***	-0.571***		-0.898***
	(0.297)		(0.564)	(0.083)		(0.158)
Non-Occupant Volume Boom		0.570***	-0.714**		-0.166***	0.206***
		(0.173)	(0.279)		(0.049)	(0.078)
Number of Observations	115	102	102	115	102	102
R-squared	0.272	0.098	0.309	0.293	0.103	0.323

Panel B. MSA-Level Inventories

	Δ Listings Boom			Δ Listings Quiet		
Short-Volume Boom	-1.133		-1.564	5.961***		7.022***
	(1.027)		(1.956)	(1.353)		(2.593)
Non-Occupant Volume Boom		-0.070	0.545		2.645***	-0.119
		(0.505)	(0.922)		(0.718)	(1.222)
Number of Observations	57	48	48	57	48	48
R-squared	0.022	0.000	0.014	0.261	0.228	0.336

Panel C. MSA-Level Volume Quiet and Bust

	Δ Volume Quiet + Bust			Foreclosures Bust		
Short-Volume Boom	-1.047***		-0.663***	0.895**		2.863***
	(0.096)		(0.177)	(0.398)		(0.746)
Non-Occupant Volume Boom		-0.512***	-0.237***		-0.060	-1.245***
		(0.051)	(0.087)		(0.215)	(0.368)
Number of Observations	115	102	102	115	102	102
R-squared	0.515	0.505	0.567	0.043	0.001	0.130

Notes: This table reports estimates of the relation between speculative volume and housing cycle measures at the MSA level. Short-Volume Boom has a mean of 16.0% and a standard deviation of 12.9%. Non-Occupant Volume Boom has a mean of 29.3% and a standard deviation of 27.1%. Δ Volume Quiet + Bust is defined as the change in total volume from 2005 through 2011. Δ Listings Boom is defined as the change in total listings from 2003 through 2005. Δ Listings Quiet is defined as the change in total listings from 2005 through 2007. Foreclosures Bust is defined as total foreclosures from 2007 through 2011. Price Boom is defined as the change in prices from 2000 through 2006. Price Bust is defined as the change in prices from 2006 through 2011. To aid interpretation of these relations, we scale the change in outcomes for all quantity measures relative to total volume in 2003 and multiply by 100. Internet Appendix Table IA.II presents summary statistics for each sample. Significance levels 10%, 5%, and 1% are denoted by *, **, and ***, respectively.

TABLE 3

House Price Appreciation and Speculative Buyer Shares (Monthly Panel VAR)

	House Price Appreciation Rate		
Lagged Price Appreciation	0.375***	0.387***	0.372***
	(0.026)	(0.027)	(0.026)
Lagged Short-Buyer Share	0.021***		0.023***
	(0.005)		(0.005)
Lagged Non-Occupant Share		0.009	0.006
		(0.008)	(0.006)

	Short-Buyer Share		
Lagged Price Appreciation	0.163***		0.162***
	(0.048)		(0.048)
Lagged Short-Buyer Share	0.780***		0.781***
	(0.024)		(0.023)
Lagged Non-Occupant Share			0.001
			0.017

	Non-Occupant Share		
Lagged Price Appreciation		0.124***	0.172***
		(0.044)	(0.045)
Lagged Short-Buyer Share			-0.071***
			(0.016)
Lagged Non-Occupant Share		0.892***	0.900***
		(0.025)	(0.021)

Notes: This table presents estimates from MSA-by-month panel vector autoregressions (pVARs) describing the relation between house price growth and the share of purchases made by non-occupant buyers and "short buyers," defined as buyers who will sell within three years of purchase. The left-hand-side variables are house price appreciation from $t-1$ to t, the short-buyer share of total volume in t, and the non-occupant share of total volume in t. The right-hand-side variables are lagged versions of these variables. The sample includes 8,568 observations for 102 MSAs for which we can consistently identify non-occupant buyers. House price appreciation has a mean of 0.84% and a standard deviation of 1.32%. Short-buyer share has a mean of 21.0% and a standard deviation of 5.5%. Non-occupant share has a mean of 32.8% and a standard deviation of 18.9%. Column (1) includes only house price appreciation and the short-buyer share. Column (2) includes only house price appreciation and the non-occupant share. Column (3) includes both speculative volume measures. The sample period includes the boom and quiet, which runs from January 2000 through December 2006. Regressions include MSA and month fixed effects and thus report the average autoregressive relations within MSAs over time. We seasonally adjust house prices by removing MSA-by-calendar-month fixed effects before computing house price growth. Standard errors are clustered at the MSA level.

TABLE 4
The Speculative Share of Total Volume in the Boom

Panel A. Accounting Regressions

	OLS: Volume Boom				IV: Volume Boom	
Short-Volume Boom	2.30		2.67		2.28	2.84
	(0.11)		(0.12)		(0.18)	(0.46)
Non-Occupant Volume Boom		1.25		1.74		
		(0.08)		(0.11)		
MSA-level	X	X			X	
ZIP-level (MSA Effects)			X	X		X
Number of Observations	115	102	6826	5662	102	6826
R-squared	0.78	0.69	0.67	0.66	0.79	0.60

Panel B. Variance Decomposition

Variance of LHS	0.083	0.093	0.268	0.289
Variance of RHS	0.012	0.041	0.020	0.057
Contribution to Boom (%)	33	55	20	34

Notes: This table presents regressions at the MSA and ZIP levels of the percentage change in total volume from 2000 to 2005 on the change in short-holding-period volume or the change in non-occupant volume from 2000 to 2005 relative to total volume in 2000. Panel A presents OLS regressions and IV regressions, where the short-volume boom is instrumented with demographic data from the 2000 Census 5% microdata. The instruments are the share of recent buyers under 35 and the share of recent buyers aged 65 or older. The ZIP-level regression is estimated with MSA fixed effects and with standard errors clustered at the MSA level. Census microdata was not available for 13 MSAs in our sample, hence the lower sample count in column 5. See Internet Appendix Table IA.IV for first-stage regressions. The F-statistics in the MSA-level and ZIP-level (Kleibergen-Paap Wald F-statistic reflecting MSA-level clustering) regressions are 40 and 8, respectively. Panel B presents the inputs needed to interpret the Panel A regressions in terms of a variance decomposition that matches the plots in Figure 5.

TABLE 5

Inputs into model calibration

Parameter or target	Value	Source
Panel A: Assumed parameters		
r (non-mover discount rate)	0.012	Guren (2018)
Potential λ values	$\{0.50, 0.17, 0.05, 0.03, 0.01\}$	Figure 7
ρ (demand growth persistence)	0.880	GN (2017)
Panel B: Steady-state targets		
Occupant buyer shares	$(0.06, 0.07, 0.16, 0.16, 0.34)$	Figure 7
Non-occupant buyer shares	$(0.04, 0.03, 0.04, 0.04, 0.06)$	Figure 7
Annual volatility of demand growth	0.023	GN (2017)
Quarterly selling hazard	0.75	Guren (2018)
Mean demand error	0	Model
Mean demand growth	0	Model
Panel C: Cycle targets		
Price overshoot	2.3	Figure 1
Volume boom/price boom	0.4	Figure 1
Non-occupant boom/occupant boom	3.1	Figure 4

Notes: This table reports parameters that we assume in the calibration, as well as targets we use to determine the remaining parameters. In the model, we target the mean buyer shares, quarterly selling hazard, and demand error across all analysis periods in control simulations. We theoretically derive the annual volatility of demand growth as well as the mean demand growth as functions of parameters. Price overshoot is the ratio of log price growth from the beginning to peak to log price growth from the beginning to the trough after the peak. Volume boom/price boom is the ratio of log existing volume growth from the beginning to the peak of volume (2000 to 2005, using numbers from Figure 4) to aforementioned log price growth. Non-occupant boom/occupant boom is the ratio of each category of log volume growth from 2000 to 2005 in the sample of MSAs we use for non-occupant analysis. In the model, we use quarterly minimums and maximums instead of aggregating at the year level. We match all targets to within rounding. GN (2017) denotes Glaeser and Nathanson (2017).

TABLE 6

Outputs from model calibration

Parameter or outcome	Interpretation	Value
	Panel A: Derived parameters	
σ_a	Flow utility dispersion	0.066
μ_1	Occupant premium	0.009
γ	g variance share	0.070
κ	Cap rate error	0.029
σ	Demand volatility	0.011
μ	Mean demand growth	-0.000
r_m	Mover discount rate	0.141
β_0^λ	Non-occupant shares	$(0.143, 0.022, 0.030, 0.030, 0.045)$
β_1^λ	Occupant shares	$(0.185, 0.052, 0.119, 0.119, 0.254)$
	Panel B: Steady-state outcomes	
1-year extrapolation	–	0.127
2–5-year extrapolation	–	0.042

Notes: See text for definitions of parameters in Panel A. We find these values by searching for parameters such that moments from the model match targets in Table 5. Panel B reports regression coefficients of annualized price growth in the next year and between 2 and 5 years from now on last year's price growth. We run these regressions across control simulations at the beginning of the analysis period.

TABLE 7
Model counterfactuals

Outcome	Baseline	All long-term buyers		All occupants	
		No occupant adjustment	Occupant adjustment	No short-term adjustment	Short-term adjustment
Price boom	14.5	8.7	8.7	9.4	14.6
Price bust	−8.2	−0.4	−0.4	−0.6	−8.3
Volume boom	5.8	2.9	2.9	2.1	5.8
Listings, end of boom	−1.3	−3.1	−3.1	−0.2	−1.3
Listings, end of quiet	1.4	0.4	0.4	0.0	1.4
Short volume boom	14.1	3.4	3.4	6.4	14.1
Non-occupant volume boom	12.3	3.6	3.6	–	–
Sale probability boom	7.1	6.0	6.0	2.3	7.1

Notes: We report 100 times changes in log outcomes between treatment and control simulations. We define the end of the quiet as the first local maximum in the impulse response of log prices, and we measure the following outcomes at that time: price boom and listings end of quiet. We define the end of the boom as the first local maximum in the impulse response of log volume before the end of the quiet, and we measure the following outcomes at that time: volume boom, listings end of boom, short volume boom, non-occupant volume boom, and sale probability boom. The price bust is the change from the end of the quiet to the first local minimum of the impulse response of log prices after the end of the quiet. A two-sided minimum does not occur in the 48 analysis periods in the fourth column, so we extend the analysis 60 additional periods to find such a minimum. The counterfactals involve different values of the underlying distribution of potential buyers, β_n^λ, that the text describes. We alter κ in each counterfactual to maintain a zero demand error while keeping other parameters the same. The baseline values correspond to Figure 9.

A   Omitted Proofs of Mathematical Statements

A.1   Proof of Lemma 1

Movers at t believe that they observe $d_{t-j} = \tilde{d}_{t-j}$ for all $j > 0$. Let g_t^* denote the mean of the posterior on g_{t-1} from this information, and σ_l^2 its variance. We solve for these outcomes using standard Kalman filtering. Denote $\sigma_d^2 = (1-\gamma)\sigma^2$ and $\sigma_g^2 = \gamma(1-\rho^2)\sigma^2$.

We have $g_{t-1} = g_t^* + \zeta_t^g$, where $\zeta_t^g \sim \mathcal{N}(0, \sigma_l^2)$. Therefore, $g_t = (1-\rho)\mu + \rho g_{t-1} + \epsilon_t^g = (1-\rho)\mu + \rho g_t^* + \rho\zeta_t^g + \epsilon_t^g$. The prior on g_t at $t+1$ is thus $\mathcal{N}((1-\rho)\mu + \rho g_t^*, \rho^2\sigma_l^2 + \sigma_g^2)$. The information is $\Delta\tilde{d}_t$, which according to movers equals $g_t + \epsilon_t^d$. Therefore, the new posterior variance satisfies $\sigma_l^2 = \sigma_d^2(\rho^2\sigma_l^2 + \sigma_g^2)(\sigma_d^2 + \rho^2\sigma_l^2 + \sigma_g^2)^{-1}$. Solving yields

$$\sigma_l^2 = (2\rho^2)^{-1}\left(-(1-\rho^2)\sigma_d^2 - \sigma_g^2 + \sqrt{((1-\rho^2)\sigma_d^2 + \sigma_g^2)^2 + 4\rho^2\sigma_d^2\sigma_g^2}\right).$$

The new posterior mean satisfies $g_{t+1}^* = (1-\alpha)\Delta\tilde{d}_t + \alpha((1-\rho)\mu + \rho g_t^*)$, where $\alpha = \sigma_d^2/(\sigma_d^2 + \rho^2\sigma_l^2 + \sigma_g^2)$. Iterating (and then subtracting one from the time subscripts everywhere) gives

$$g_t^* = \mu + (1-\alpha)\sum_{j=1}^{\infty}(\alpha\rho)^{j-1}\left(\Delta\tilde{d}_{t-j} - \mu\right).$$

Because $\hat{g}_t = (1-\rho)\mu + \rho g_t^*$, we have proved the Lemma formula. We have $d_t = d_{t-1} + g_t + \epsilon_t^d = (d_{t-1} - \tilde{d}_{t-1}) + \tilde{d}_{t-1} + (1-\rho)\mu + \rho g_{t-1} + \epsilon_t^g + \epsilon_t^d = (d_{t-1} - \tilde{d}_{t-1}) + \tilde{d}_{t-1} + \hat{g}_t + \rho\zeta_t^g + \epsilon_t^g + \epsilon_t^d$, which immediately gives $\hat{d}_t = \tilde{d}_{t-1} + \hat{g}_t$, with $\tilde{\sigma}^2 = \rho^2\sigma_l^2 + \sigma_g^2 + \sigma_d^2$.

A.2   Proof of Lemma 2

Write $V^m(\hat{d}_t, \hat{g}_t) = e^{\hat{d}_t}v^m(\hat{d}_t, \hat{g}_t)$ and $P = e^{\hat{d}_t}p$. Denote $\zeta_t = d_t - \hat{d}_t$. Then $\tilde{\pi}(P, d_t) = 1 - F(\log p + \log\kappa - \zeta_t)$, which we denote $\tilde{\pi}(p, \zeta_t)$ by abuse of notation. Substituting these expressions into (7) and using (5) and (6) yield

$$v^m(\hat{d}_t, \hat{g}_t) = \sup_p E\left(\tilde{\pi}(p, \zeta_t)p + \frac{(1-\tilde{\pi}(p, \zeta_t))e^{(1-\rho)\mu + \rho\hat{g}_t + (1+\rho-\alpha\rho)\zeta_t}v^m(\hat{d}_{t+1}, \hat{g}_{t+1})}{1 + r_m}\right), \quad (12)$$

with the expectation over $\zeta_t \sim \mathcal{N}(0, \tilde{\sigma}^2)$. Because $\hat{d}_t$ appears only through the first argument of v^m, this function does not depend on $\hat{d}_t$. It follows that the argmax also does not depend on $\hat{d}_t$. We denote it $p(\hat{g}_t)$.

A.3   Limit of Infinite Mover Impatience

When $r_m \to \infty$, $p(\cdot)$ becomes constant, as is clear from the equation for v^m. In that case, $E\log(P_{t+j}/P_t) = E(\hat{d}_{t+j} - \hat{d}_t)$. From the point of view of movers at t, we can iteratively

apply (5) and (6) to obtain

$$E_t \log \left(\frac{P_{t+j}}{P_t} \right) = j\mu + \frac{\rho(1-\rho)^j}{1-\rho} \left(\hat{g}_t - \mu \right).$$

Therefore, $\hat{g}_t$ proxies for expected future price growth, with ρ controlling the term structure of future expectations. We can also express $\hat{g}_t$ in terms of past price growth. In particular, making substitutions to (5) gives $\hat{g}_{t+1} = (1-\rho)\mu + \rho\hat{g}_t + \rho(1-\alpha)(\Delta \log P_{t+1} - \hat{g}_{t+1})$. Recursively expanding this equation and moving back time subscripts gives

$$\hat{g}_t - \mu = \frac{\rho(1-\alpha)}{1+\rho(1-\alpha)} \sum_{j=0}^{\infty} \rho^j \left(\Delta \log P_{t-j} - \mu \right).$$

Therefore, beliefs about future price growth endogenously extrapolate from past price growth, as in Glaeser and Nathanson (2017). In contrast to that paper, here we allow forward-looking movers through finite r_m, in which case prices become less extrapolative. We choose r_m to match moments in our quantitative exercise.

We can also derive price setting at $t+1$ as a function of market data. In particular, $\log P_{t+1} - \log P_t = \hat{d}_{t+1} - \hat{d}_t$. From (5) and (6), we know that this difference equals $(1-\rho)\mu + \rho\hat{d}_t + (1+\rho(1-\alpha))(\tilde{d}_t - \hat{d}_t)$, and from (4), we have $\tilde{d}_t - \hat{d}_t = \log \kappa p - F^{-1}(1-\pi_t)$. Substituting the equation just derived for $\hat{g}_t$ yields

$$\log P_{t+1} - \log P_t =$$

$$\mu + (1+\rho(1-\alpha)) \left(\log \kappa p - F^{-1}(1-\pi_t) \right) + \frac{\rho^2(1-\alpha)}{1+\rho(1-\alpha)} \sum_{j=0}^{\infty} \rho^j \left(\Delta \log P_{t-j} - \mu \right).$$

Therefore, movers set list prices as a markup over last period's price, where the markup is the sum of three terms: the mean growth rate μ, the information learned from sales probabilities at t, and a weighted sum of past price growth. This rule closely resembles the "backward-looking rule of thumb" that Guren (2018) assumes. The formula there, however, lacks a term corresponding to the one here with π_t, as his rule-of-thumb sellers do not adjust list prices in response to market data other than past prices.

A.4 Proof of Lemma 3

We define $V_\lambda^s(\hat{d}_t, \hat{g}_t) = \sum_{j=1}^{\infty} \lambda(1-\lambda)^{j-1}(1+r)^{-j} E_t V^m(\hat{d}_{t+1}, \hat{g}_{t+1})$, where the expectation is conditional on mover information at t. This expression gives the resale value of owning a house at t for a stayer of type λ conditional on public information. We write this value recursively as

$$V_\lambda^s(\hat{d}_t, \hat{g}_t) = (1+r)^{-1} E_t \left((1-\lambda) V_\lambda^s(\hat{d}_{t+1}, \hat{g}_{t+1}) + \lambda V^m(\hat{d}_{t+1}, \hat{g}_{t+1}) \right).$$

We write $V_\lambda^s(\hat{d}_t, \hat{g}_t) = e^{\hat{d}_t} v_\lambda^s(\hat{d}_t, \hat{g}_t)$. Plugging in the result from the proof of Lemma 2 that $V^m(\hat{d}_t, \hat{g}_t) = e^{\hat{d}_t} v^m(\hat{g}_t)$, we get

$$v_\lambda^s(\hat{d}_t, \hat{g}_t) = (1+r)^{-1} E\left(e^{(1-\rho)\mu + \rho\hat{g}_t + (1+\rho-\alpha\rho)\zeta_t}\left((1-\lambda)v_\lambda^s(\hat{d}_{t+1}, \hat{g}_{t+1}) + \lambda v^m(\hat{g}_{t+1})\right)\right), \quad (13)$$

with the expectation over $\zeta_t \sim \mathcal{N}(0, \tilde{\sigma}^2)$. Because $\hat{d}_t$ enters only v_λ^s, that function does not depend on $\hat{d}_t$, so we can write $V_\lambda^s(\hat{d}_t, \hat{g}_t) = e^{\hat{d}_t} v_\lambda^s(\hat{g}_t)$. Substituting this expression into (8), using the recursive formulation for the resale value, and using the potential buyer imputation of $\hat{d}_t$ give

$$V_{i,t}^b = \frac{e^{d_i}}{r + \lambda_i} + \frac{P_{i,t}}{(1+r)p(\hat{g}_t)} E_{i,t}\left(e^{(1-\rho)\mu + \rho\hat{g}_t + (1+\rho-\alpha\rho)\zeta_t}\left((1-\lambda_i)v_{\lambda_i}^s(\hat{g}_{t+1}) + \lambda_i v^m(\hat{g}_{t+1})\right)\right),$$

where the expectation is over ζ_t drawn from the normal in (9). Letting $\Psi^{\lambda_i}(\hat{g}_t, \zeta_t)$ denote the argument inside the expectation, we can then simplify the buying decision, $V_{i,t}^b \geq P_{i,t}$, as

$$e^{d_i} \geq P_{i,t}(r + \lambda_i)\left(1 - \frac{\int_{-\infty}^\infty \Psi^{\lambda_i}\left(\hat{g}_t, \zeta + \frac{\tilde{\sigma}^2(d_i - \mu_{n_i} - \hat{d}_t)}{\tilde{\sigma}^2 + \sigma_a^2}\right)\phi(\zeta)d\zeta}{(1+r)p(\hat{g}_t)}\right),$$

where ϕ is a mean-zero normal pdf with variance $\tilde{\sigma}^2 \sigma_a^2 (\tilde{\sigma}^2 + \sigma_a^2)^{-1}$. Write $e^{d_i} = \kappa_i P_{i,t}$. Then the equation becomes

$$\kappa_i \geq (r + \lambda_i)\left(1 - \frac{\int_{-\infty}^\infty \Psi^{\lambda_i}\left(\hat{g}_t, \zeta + \frac{\tilde{\sigma}^2(\log \kappa_i + \log p(\hat{g}_t) - \mu_{n_i} - \hat{d}_t)}{\tilde{\sigma}^2 + \sigma_a^2}\right)\phi(\zeta)d\zeta}{(1+r)p(\hat{g}_t)}\right).$$

In Internet Appendix III, we prove that $v^m(\cdot)$ and $v_s^\lambda(\cdot)$ are continuous functions that weakly increase. As a result, the right side of the above inequality continuously and weakly decreases in κ_i. The left side continuously and strictly increases in κ_i. Therefore, if the right side limits to a non-positive number as $\kappa_i \to 0$, then the inequality holds for all $\kappa_i > 0$, meaning we can set $\kappa_{n_i}^{\lambda_i}(\hat{g}_t) = 0$. If the right side limits to a positive number as $\kappa_i \to 0$, then by the Intermediate Value Theorem, there exists a unique $\kappa_{n_i}^{\lambda_i}(\hat{g}_t) > 0$ such that the inequality holds if and only if $\kappa_i \geq \kappa_{n_i}^{\lambda_i}(\hat{g}_t)$, which proves the Lemma.

For Online Publication

I Data Appendix

To conduct our empirical analysis we make use of a transaction-level data set containing detailed information on individual home sales taking place throughout the US between 1995 and 2014. The raw data was purchased from CoreLogic and is sourced from publicly available tax assessment and deeds records maintained by local county governments. In some analyses we supplement this transaction-level data with additional data on the listing behavior of individual homeowners. Our listings data is also provided by CoreLogic and is sourced from a consortium of local Multiple Listing Service (MLS) boards located throughout the country.

Selecting Geographies

To select our sample of transactions, we first focus on a set of counties that have consistent data coverage going back to 1995 and which, together, constitute a majority of the housing stock in their respective MSAs. In particular, to be included in our sample a county must have at least one "arms length" transaction with a non-negative price and non-missing date in each quarter from 1995q1 to 2014q4.[29] Starting with this subset of counties, we then further drop any MSA for which the counties in this list make up less than 75 percent of the total owner-occupied housing stock for the MSA as measured by the 2010 Census. This leaves us with a final set of 250 counties belonging to a total of 115 MSAs. These MSAs are listed below in Table IA.VI along with the percentage of the housing stock that is represented by the 250 counties for which we have good coverage. Throughout the paper, when we refer to counts of transactions in an MSA we are referring to the portion of the MSA that is accounted for by these counties.

Selecting Transactions

Within this set of MSAs, we start with the full sample of all arms length transactions of single family, condo, or duplex properties and impose the following set of filters to ensure that our final set of transactions provides an accurate measure of aggregate transaction volume over the course of the sample period:

1. Drop transactions that are not uniquely identified using CoreLogic's transaction ID.

2. Drop transactions with non-positive prices.

3. Drop transactions that are recorded by CoreLogic as nominal transfers between banks or other financial institutions as part of a foreclosure process.

4. Drop transactions that appear to be clear duplicates, identified as follows:

   (a) If a set of transactions has an identical buyer, seller, and transaction price but are recorded on different dates, keep only the earliest recorded transaction in the set.

---

[29]We rely on CoreLogic's internal transaction-type categorization to determine whether a transaction occurred at arms length.

(b) If the same property transacts multiple times on the same day at the same price keep only one transaction in the set.

5. If more than 10 transactions between the same buyer and seller at the same price are recorded on the same day, drop all such transactions. These transactions appear to be sales of large subdivided plots of vacant land where a separate transaction is recorded for each individual parcel but the recorded price represents the price of the entire subdivision.

6. Drop sales of vacant land parcels in MSAs where the CoreLogic data includes such sales.[30] We define a vacant land sale to be any transaction where the sale occurs a year or more before the property was built.

Table IA.VII shows the number of transactions that are dropped from our sample at each stage of this process as well as the final number of transactions included in our full analysis sample.

Identifying Occupant and Non-Occupant Buyers

We identify non-occupant buyers using differences between the mailing addresses listed by the buyer on the purchase deed and the actual physical address of the property itself. In most cases, these differences are identified using the house numbers from each address. In particular, if both the mailing address and the property address have a non-missing house number then we tag any instance in which these numbers are not equal as a non-occupant purchase and any instance in which they are equal as occupant purchases. In cases where the mailing address property number is missing we also tag buyers as non-occupants if both the mailing address and property address street names are non-missing and differ from one another. Typically, this will pick up cases where the mailing address provided by the buyer is a PO Box. In all other cases, we tag the transaction as having an unknown occupancy status.

Restricting the Sample for the Non-Occupant Analysis

Our analysis of non-occupant buyers focuses on the growth of the number of purchases by these individuals between 2000 and 2005. To be sure that this growth is not due to changes in the way mailing addresses are coded by the counties comprising the MSAs in our sample, for the non-occupant buyer analysis we keep only MSAs for which we are confident such changes do not occur between 2000 and 2005. In particular, we first drop any MSA in which the share of transactions in any one year between 2000 and 2005 with unknown occupancy status exceeds 0.5. Of the remaining MSAs, we then drop those for which the increase in the number of non-occupant purchases between any year and the next exceeds 150%, with the possible base years being those between 2000 and 2005.[31] The 102 MSAs that remain after these two filters are marked with an "x" in columns 3 and 7 of Table IA.VI.

---

[30]MSAs are flagged as including vacant land sales if more then 5 percent of the sales in the MSA occur more then two years before the year in which the property was built.

[31]This step drops only Chicago-Naperville-Elgin, IL-IN-WI.

Restricting the Sample for Listings Analysis

The geographic and time series coverage of the CoreLogic MLS data is not as comprehensive as the transaction-level data. As a result, our analysis of listings behavior is restricted to a subset of markets for which we can be relatively certain that the MLS data is representative of the majority of owner-occupied home sales in the area. We impose several filters to identify this subset of MSAs. First, starting with the full set of 115 MSAs contained in the transaction-level data, we drop any MSA for which there is not at least one new listing in every month and in every county subcomponent of the MSA between January 2000 and December 2014. Within the remaining set of MSAs we then drop any MSA for which the number of new listings between 2006 and 2008 is more than 2.5 times the number of new listings between 2003 and 2005. This filter eliminates MSAs that experience large jumps in coverage during the quiet. Finally, we also drop any MSA for which the number of sold listings (from the MLS data) is less than 25 percent of total sales volume (from the transaction data) over the period 2003-2012. This filter eliminates MSAs for which the listings data is likely to be unrepresentative of sales activity during our main sample period. This leaves a final sample of 57 MSAs for our listings analysis. These MSAs are marked with an "x" in columns 4 and 8 of Table IA.VI.

Identifying New Construction Sales

In several parts of our analysis we omit new construction sales from the calculation of total transaction volume. To identify sales of newly constructed homes, we start with the internal CoreLogic new construction flag and make several modifications to pick up transactions that may not be captured by this flag. CoreLogic identifies new construction sales primarily using the name of the seller on the transaction (e.g. "PULTE HOMES" or "ROCKPORT DEV CORP"), but it is unclear whether their list of home builders is updated dynamically or maintained consistently across local markets. To ensure consistency, we begin by pulling the complete list of all seller names that are ever identified with a new construction sale as defined by CoreLogic. Starting with this list of sellers, we tag any transaction for which the seller is in this list, the buyer is a human being, and the transaction is not coded as a foreclosure sale by CoreLogic as a new construction sale. We use the parsing of the buyer name field to distinguish between human and non-human buyers (e.g. LLCs or financial institutions). Human buyers have a fully parsed name that is separated into individual first and name fields whereas non-human buyer's names are contained entirely within the first name field.

This approach will identify all new construction sales provided that the seller name is recognized by CoreLogic as the name of a homebuilder. However, many new construction sales may be hard to identify simply using the name of the seller. We therefore augment this definition using information on the date of the transaction and the year that the property was built. In particular, if a property was not already assigned a new construction sale using the builder name, then we search for sales of that property that occur within one year of the year that the property was built and record the earliest of such transactions as a new construction sale.

Finally, for properties that are not assigned a new construction sale using either of the

two above methods, we also look to see if there were any construction loans recorded against the property in the deeds records. If so, we assign the earliest transaction to have occurred within three years of the earliest construction loan as a new construction sale. We use a three-year window to allow for a time lag between the origination of the construction loan and the actual date that the property was sold. Construction loans are identified using CoreLogic's internal deed and mortgage type codes.

II Mechanical Short-Term Volume

Table IA.V reports the buy-year fixed effects estimates for years 2000 to 2005 relative to 2000. The fixed effects are linear differences of a monthly selling hazard, so multiplying by 12 roughly gives the effect on the annualized selling probability. Therefore, buyers in 2005 have a 3.2 percentage point larger annual selling hazard than buyers in 2000 (12 times 0.0027 equals 0.0324).

We use these estimates to construct counterfactual growth of short-term volume from 2000 to 2005. For each $2000m1 \leq t' < t \leq 2005m12$, we construct the counterfactual selling hazard as

$$\pi_{t',t}^c = \pi_{t',t} - \left(\hat{\alpha}_{y(t')}^{buy} - \hat{\alpha}_{2000}^{buy} \right),$$

which subtracts away any increase due to the change in the composition of buyers from 2000 to the year of t'. We then compute the counterfactual of $v_{t',t}$, the volume of homes bought in t' and sold in t, using the following iterative procedure. Let $e_{t',t}$ count homes bought in t' that have not yet sold by t, and let c superscripts mark counterfactual values. We initialize counterfactuals with actuals: for each $1995m1 \leq t' < 2005m12$,

$$e_{t',t'}^c = e_{t',t'}$$
$$v_{t',t'}^c = v_{t',t'}.$$

We then iteratively update the counterfactuals over t running from $t' + 1$ to $2005m12$:

$$e_{t',t}^c = e_{t',t-1}^c - v_{t',t-1}^c$$
$$v_{t',t}^c = \pi_{t',t}^c e_{t',t}^c.$$

To compute short-term volume in year y, we sum $v_{t',t}$ across all subscripts for which $y(t) = y$ and $0 < t - t' < 36$; we sum $v_{t',t}^c$ across the same indices for counterfactual short-term volume.

The remaining columns of Table IA.V report the results. Between 2000 and 2005, total volume grows 36.7% and short-term volume grows 77.5% in the actual data. The disproportionate rise in short-term volume is the difference, 40.8%. Counterfactual short-term volume rises 41.5% between 2000 and 2005, giving a disproportionate rise of 4.8%. Therefore, $4.8\%/40.8\% = 11.8\%$ of the disproportionate rise in short-term volume remains in the counterfactual. We attribute the 88.2% that disappeared to the changing composition of buyers between 2000 and 2005.

III Value Function Monotonicity

This section establishes that the functions $v^m(\cdot)$ and $v^s_\lambda(\cdot)$, which we define in the proofs of Lemmas 2 and 3, weakly and continuously increase. We follow Stokey, Lucas and Prescott (1989). To apply their results, we need to work with a one-point (Alexandroff) compactification of a subset of the real numbers. For a topological set X, the Alexandroff compactification is the set $X^* = X \cup \{\infty\}$, whose open sets are those of X together with sets whose complements are closed, compact subsets of X; X^* is compact (Kelley, 1955).

Lemma 4. *Let $f : (0, \infty) \times \mathbb{R} \to \mathbb{R}$ be continuous. Suppose there exists functions $g_0 : \mathbb{R} \to \mathbb{R}$ and $g_\infty : \mathbb{R} \to \mathbb{R}$ such that $\lim_{x \to 0} f(x, y) = g_0(y)$ and $\lim_{x \to \infty} g_\infty(y)$ uniformly. Define $\tilde{f} : [0, \infty)^* \times \mathbb{R} \to \mathbb{R}$ by $\tilde{f}(x, y) = f(x, y)$ for $x \in (0, \infty)$ and $\tilde{f}(x, y) = g_x(y)$ for $x \in \{0, \infty\}$. Then $\tilde{f}$ is continuous.*

Proof. Let $Z \subset \mathbb{R}$ be open. We show that $\tilde{f}^{-1}(Z)$ is open by demonstrating that for each $(x, y) \subset \tilde{f}^{-1}(Z)$, there exists an open set U such that $(x, y) \in U \subset \tilde{f}^{-1}(Z)$. If $x \in (0, \infty)$, then set $U = f^{-1}(Z)$, which is open by the continuity of f. Consider the case $x = 0$. Because Z is open, there exists $\epsilon > 0$ such that all z with $|z - g_0(y)| < \epsilon$ are in Z. By uniform convergence, there exists $\delta > 0$ such that $|f(x', y') - g_0(y)| < \epsilon$ for all $x \in [0, \delta)$ and $y \in \mathbb{R}$. Therefore, $U = [0, \delta) \times \mathbb{R}$ suffices. Consider the case $x = \infty$. There likewise exists $\epsilon > 0$ such that all z with $|z - g_\infty(y)| < \epsilon$ are in Z. By uniform convergence, there exists $N > 0$ such that $|f(x', y') - g_\infty(y)| < \epsilon$ for all $x > N$ and $y \in \mathbb{R}$. Therefore, $U = (N, \infty) \times \mathbb{R}$ suffices. $\qquad\square$

We next establish the existence of a continuous solution $v^m(\cdot)$ to (12). Let $\mathcal{C}$ be the space of bounded continuous functions from $\mathbb{R}$ to itself. Let $a > 0$ be a constant. For $v \in \mathcal{C}$, we define the operator T by $(Tv)(\hat{g}) = \sup_p f(p, \hat{g})$, where

$$
f(p, \hat{g}) = \int_{-\infty}^{\infty} \left(\frac{\tilde{\pi}(p, \zeta)p}{a + e^{\frac{\rho\hat{g}}{1-\rho}}} + \frac{(1 - \tilde{\pi}(p, \zeta))e^{(1-\rho)\mu + \rho\hat{g} + (1+\rho-\alpha\rho)\zeta}}{1 + r_m} \times \right.
$$
$$
\left. \frac{\left(a + e^{\rho\mu + \frac{\rho^2 \hat{g}}{1-\rho} + \frac{\rho^2(1-\alpha)\zeta}{1-\rho}} \right) v((1-\rho)\mu + \rho\hat{g} + \rho(1-\alpha)\zeta)}{a + e^{\frac{\rho\hat{g}}{1-\rho}}} \right) \phi(\zeta)d\zeta,
$$

where $\phi(\cdot)$ is the probability density function of $\mathcal{N}(0, \tilde{\sigma}^2)$. If v is a fixed point of T, then $v^m(\hat{g}) = (a + e^{\frac{\rho\hat{g}}{1-\rho}})v(\hat{g})$ solves (12). We find a fixed point by demonstrating that $T : \mathcal{C} \to \mathcal{C}$ and then showing that for a sufficiently small value of a, T satisfies the Blackwell conditions and is hence a contraction mapping.

We first show that $Tv \in \mathcal{C}$. We have the bound

$$
||Tv|| \leq \sup_p \int_{-\infty}^{\infty} a^{-1}\tilde{\pi}(p, \zeta)p\phi(\zeta)d\zeta +
$$
$$
(1 + r_m)^{-1}e^{(1-\rho)\mu}||v|| \sup_x \frac{ae^{\rho x + \frac{(1+\rho-\alpha\rho)^2\tilde{\sigma}^2}{2}} + e^{\rho\mu + \frac{\rho x}{1-\rho} + \frac{(1-\alpha\rho)^2\tilde{\sigma}^2}{2(1-\rho)^2}}}{a + e^{\frac{\rho x}{1-\rho}}},
$$

so Tv is bounded.

Demonstrating continuity is much more complicated. We first apply Lemma 12.14 of Stokey, Lucas and Prescott (1989) to establish the continuity of $f(\cdot, \cdot)$.

In their terminology, $X = (0, \infty)$, $Z = \mathbb{R}^2$, their y corresponds to our p, their z corresponds to our $(\hat{g}, \zeta)$, and the transition function Q puts mass $\phi(\zeta')$ on $(\hat{g}, \zeta')$ and mass 0 on other elements of Z. To apply their lemma, we must show that Q has the Feller property, which means (see their page 375) that $\int h(z')Q(z, z')dz'$ is continuous in z as long as h is continuous and bounded.[32] Given our specification of Q, this integral reduces to $\int_{-\infty}^{\infty} h(\hat{g}, \zeta')\phi(\zeta')d\zeta'$, which is trivially continuous in ζ. To demonstrate continuity in $\hat{g}$, we closely follow the proof of their Lemma 9.5. Choose a sequence $\hat{g}_n$ converging to $\hat{g}$. Then $|\int_{-\infty}^{\infty} h(\hat{g}_n, \zeta')\phi(\zeta')d\zeta' - \int_{-\infty}^{\infty} h(\hat{g}, \zeta')\phi(\zeta')d\zeta'| \leq \int_{-\infty}^{\infty} |h(\hat{g}_n, \zeta') - h(\hat{g}, \zeta')|\phi(\zeta')d\zeta'$. Each function $\zeta' \mapsto |h(\hat{g}_n, \zeta') - h(\hat{g}, \zeta')|$ converges pointwise to the zero function (by the continuity of h), so by the Lebesgue Dominated Convergence Theorem (their Theorem 7.10), this integral limits to zero. Therefore, $\hat{g} \mapsto \int_{-\infty}^{\infty} h(\hat{g}, \zeta')\phi(\zeta')d\zeta'$ is continuous in $\hat{g}$, and Q has the Feller property. As a result, $f(\cdot, \cdot)$ is continuous on $(0, \infty) \times \mathbb{R}$.

The next step is to invoke our Lemma 4. To do so, we must show uniform converge of $f(p, \hat{g})$ for $p \to 0$ and $p \to \infty$. In the first limit, $f(p, \hat{g}) \to 0$, and this convergence is uniform because terms with $\hat{g}$ multiplying the terms with p are uniformly bounded in $\hat{g}$. In the second limit, the convergence is to the integral in which $\tilde{\pi} = 0$, and the convergence is uniform for the same reason. Hence, Lemma 4 applies, and the induced $\tilde{f}$ is continuous.

The final step is to show that $(Tv)(\hat{g})$ is continuous. This statement follows immediately from Berge's Maximum Theorem on general topological spaces (see, for instance, page 570 of Aliprantis and Border (2006)) because $\sup_{p \in (0, \infty)} f(p, \hat{g}) = \sup_{p \in [0, \infty)^*} \tilde{f}(p, \hat{g})$ and because $[0, \infty)^*$ is compact. Therefore, $Tv \in \mathcal{C}$.

We next verify the Blackwell conditions for T (Theorem 3.3 in Stokey, Lucas and Prescott (1989)). Monotonicity is trivial. Given the bound above, discounting holds as long as

$$(1 + r_m)^{-1} e^{(1-\rho)\mu} \sup_x \frac{a e^{\rho x + \frac{(1+\rho-\alpha\rho)^2 \tilde{\sigma}^2}{2}} + e^{\rho\mu + \frac{\rho x}{1-\rho} + \frac{(1-\alpha\rho)^2 \tilde{\sigma}^2}{2(1-\rho)^2}}}{a + e^{\frac{\rho x}{1-\rho}}} < 1.$$

We are free to choose any positive value of a. By considering the limit as $a \to 0$, we find that we can choose such an a to satisfy this inequality as long as

$$(1 + r_m)^{-1} e^{\mu + \frac{(1-\alpha\rho)^2 \tilde{\sigma}^2}{2(1-\rho)^2}} < 1.$$

Because $0 \leq \alpha \leq 1$, it is sufficient for $e^{\mu + \tilde{\sigma}^2/2} < 1 + r_m$. If we can show that $\tilde{\sigma} \leq \sigma$, then we are done because we assumed in Section 7 that $e^{\mu + \sigma^2/2} < 1 + r \leq 1 + r_m$. From the proof of

---

[32]Their lemma also requires that the term inside the integral defining $f(\cdot, \cdot)$, other than $\phi(\zeta)d\zeta$, is bounded in p, $\hat{g}$, and ζ. This boundedness holds because v is bounded, because $\lim_{p \to \infty} \tilde{p}(\zeta, p)p = 0$, and because $\lim_{\zeta \to \infty} (1 - \tilde{\pi}(p, \zeta))e^{c\zeta} = 0$ for any $c > 0$.

Lemma 1, we have

$$\tilde{\sigma}^2 = \frac{\sigma_g^2}{2} + \frac{(1+\rho^2)\sigma_d^2}{2} + \frac{\sqrt{((1-\rho^2)\sigma_d^2 + \sigma_g^2)^2 + 4\rho^2\sigma_d^2\sigma_g^2}}{2}$$
$$= \sigma^2 \left(\frac{1+\rho^2(1-2\gamma)}{2} + \frac{\sqrt{(1-\rho^2)^2 + 4(1-\gamma)\gamma\rho^2(1-\rho^2)}}{2} \right).$$

We want to show that the term inside the large parentheses is no greater than 1. By isolating the square root and then squaring, we reduce this inequality to

$$(1-\rho^2)^2 + 4(1-\gamma)\gamma\rho^2(1-\rho^2) \leq (1-\rho^2(1-2\gamma))^2,$$

which simplifies to $0 \leq \gamma(2-\rho^2)$, which is true because $0 \leq \gamma, \rho \leq 1$. Therefore, by Theorem 3.3 of Stokey, Lucas and Prescott (1989), T is a contraction mapping. By the Contraction Mapping Theorem (their Theorems 3.1 and 3.2), T has a unique fixed point in $\mathcal{C}$, as desired. Call this function v^*. As mentioned above, $v^m(\hat{g}) = v^*(\hat{g})(a + e^{\frac{\rho\hat{g}}{1-\rho}})$ then solves (12); this function clearly inherits the continuity of v^*.

Finally, we show that v^m is weakly increasing. Let $\mathcal{C}' \subset \mathcal{C}$ be the set of v such that $v(\hat{g})(a+e^{\frac{\rho\hat{g}}{1-\rho}})$ weakly increases. We claim that $\mathcal{C}'$ is closed. Let $\{v_n\}$ be in $\mathcal{C}'$ converging in $\mathcal{C}$ to v. For any $\hat{g}_0 < \hat{g}_1$, $v_n(\hat{g}_1)(a+e^{\frac{\rho\hat{g}_1}{1-\rho}}) - v_n(\hat{g}_0)(a+e^{\frac{\rho\hat{g}_0}{1-\rho}}) \geq 0$. Because v_n converges pointwise to v, we must have $v(\hat{g}_1)(a + e^{\frac{\rho\hat{g}_1}{1-\rho}}) - v(\hat{g}_0)(a + e^{\frac{\rho\hat{g}_0}{1-\rho}}) \geq 0$ as well. Therefore, Corollary 1 to Theorem 3.2 of Stokey, Lucas and Prescott (1989) shows that $v^m \in \mathcal{C}'$ as long as $T : \mathcal{C}' \to \mathcal{C}'$, which is immediate from (12).

The task remaining for this appendix is to show that each $v^s(\cdot)$ weakly and continuously increases. The argument proceeds as with $v^m(\cdot)$, but we use (13), and we can skip the steps involving a supremum. Define the map T on $\mathcal{C}$ by

$$(Tv)(\hat{g}) = (1+r)^{-1} \int_{-\infty}^{\infty} \left(\frac{ae^{(1-\rho)\mu+\rho\hat{g}+(1+\rho-\alpha\rho)\zeta}}{a + e^{\frac{\rho\hat{g}}{1-\rho}}} + \right.$$
$$\left. \frac{e^{\mu+\frac{\rho\hat{g}}{1-\rho}+\frac{(1-\alpha\rho)\zeta}{1-\rho}}}{a + e^{\frac{\rho\hat{g}}{1-\rho}}} \right) ((1-\lambda)v(g') + \lambda v^*(g'))\phi(\zeta)d\zeta,$$

where $g' = (1-\rho)\mu + \rho\hat{g} + \rho(1-\alpha)\zeta$, and $a > 0$ is a constant to be specified later. If v is a fixed point of T, then $v_\lambda^s(\hat{g}) = (a + e^{\frac{\rho\hat{g}}{1-\rho}})v(\hat{g})$ solves (13). Clearly, Tv is bounded. To prove continuity, we again apply Lemma 12.14 of Stokey, Lucas and Prescott (1989), this time with $X = Z = \mathbb{R}$, our $\hat{g}$ corresponding to their y, and our ζ corresponding to their z. In order to apply their lemma, we have to absorb the ζ terms into the Q transition function so that their f is bounded. Using the identity $e^{-z^2/(2\sigma^2)+bz} = e^{\sigma^2b^2/2}e^{-(z-\sigma^2b)^2/(2\sigma^2)}$, we have

$$e^{(1+\rho-\alpha\rho)\zeta}\phi(\zeta) = e^{\tilde{\sigma}^2(1+\rho-\alpha\rho)^2/2}\phi(\zeta - \tilde{\sigma}^2(1+\rho-\alpha\rho))$$

and

$$e^{\frac{(1-\alpha\rho)\zeta}{1-\rho}}\phi(\zeta) = e^{\frac{\tilde{\sigma}^2(1-\alpha\rho)^2}{2(1-\rho)^2}}\phi\left(\zeta - \frac{\tilde{\sigma}^2(1-\alpha\rho)}{1-\rho}\right).$$

These functions serve as constants times a valid transition function (we showed above that the normal distribution with 0 mean and variance $\tilde{\sigma}^2$ has the Feller property), and the remainder of the integrand is bounded in both $\hat{g}$ and ζ. Thus, Lemma 12.14 applies and Tv is continuous. As a result, $T : \mathcal{C} \to \mathcal{C}$.

Next we verify the aforementioned Blackwell conditions for T. Monotonicity again is trivial. Discounting holds if

$$\frac{1-\lambda}{1+r}\sup_{\hat{g}}\frac{ae^{(1-\rho)\mu+\rho\hat{g}+\frac{(1+\rho-\alpha\rho)^2\tilde{\sigma}^2}{2}} + e^{\mu+\frac{\rho\hat{g}}{1-\rho}+\frac{(1-\alpha\rho)^2\tilde{\sigma}^2}{2(1-\rho)^2}}}{a + e^{\frac{\rho\hat{g}}{1-\rho}}} < 1.$$

Because we are free to pick any $a > 0$, the inequality holds for some such a if

$$(1-\lambda)e^{\mu+\frac{(1-\alpha\rho)^2\tilde{\sigma}^2}{2(1-\rho)^2}} < 1+r,$$

which always holds, because $\alpha \in (0,1)$, $\tilde{\sigma} \leq \sigma$ (see above), $e^{\mu+\sigma^2/2} < 1+r$ by assumption, and $\lambda \in (0,1)$. Therefore, T satisfies the Blackwell conditions and is a contraction mapping. As a result, it has a unique fixed point in $\mathcal{C}$. Call it v^{**}. Then $v_\lambda^s(\hat{g}) = (a + e^{\frac{\rho\hat{g}}{1-\rho}})v^{**}(\hat{g})$ solves (13).

Finally, we show that v_λ^s weakly and continuously increases. Continuity follows from the continuity of v^{**}. As argued above, weak monotonicity holds as long as $T : \mathcal{C}' \to \mathcal{C}'$, where this set is defined as above. That T maps $\mathcal{C}'$ into itself is immediate from (13) and the fact that v^m weakly increases. QED

IV   Model Variants

Rational

In the fully rational variant of our model, movers know the true cutoff functions $\kappa_n^\lambda(\hat{g}_t)$ that potential buyers use. These functions affect the mover value function, which in turn determines the $\kappa_n^\lambda(\cdot)$, so we iterate until finding a fixed point to solve the model, using the same parameters as the baseline.

We now spell out this procedure in more detail. Movers recognize that the true demand curve is

$$\pi(P, d_t, \hat{g}_t) = 1 - \sum_{n,\lambda}\beta_n^\lambda\Phi\left(\log P + \log\kappa_n^\lambda(\hat{g}_t) - d_t - \mu_n\right).$$

Their value function is then

$$V^m(\hat{d}_t, \hat{g}_t) = \sup_P E\left(\pi(P, d_t, \hat{g}_t)P + (1+r_m)^{-1}(1 - \pi(P, d_t, \hat{g}_t))V^m(\hat{d}_{t+1}, \hat{g}_{t+1})\right),$$

where the expectation is over $d_t \sim \mathcal{N}(\hat{d}_t, \tilde{\sigma}^2)$. By an argument analogous to the proof of

Lemma 2, the solution takes the form $e^{\hat{d}_t}v^m(\hat{g}_t)$ with argmax $e^{\hat{d}_t}p(\hat{g}_t)$, although $v^m(\cdot)$ and $p(\cdot)$ may differ from the corresponding functions in the baseline model. Because the mover value function takes this form, an argument analogous to the proof of Lemma 3 confirms the existence of functions $\kappa_n^\lambda(\cdot)$ such that potential buyers buy when $e^{d_i} \geq \kappa_n^\lambda(\hat{g}_t)P_{i,t}$. These functions depend on $v^m(\cdot)$, which depends on the $\kappa_n^\lambda(\cdot)$ functions. We iteratively solve for these functions using the same discretization as in the baseline model and then compute impulse responses using the same sequences of shocks.

The results appear in Figure IA.II. For ease of comparison with Figure 9, we use the same axis ranges for corresponding panels. Prices no longer overshoot, inventories never rise above their pre-shock value, and the volume boom lasts only four quarters and is only about one quarter of its size in the baseline model. Interestingly, non-occupant volume continues to rise much more than occupant volume. As Section 7 discusses, non-occupant demand is more elastic to the level of the demand shifter, d_t, because $\mu_0 < 0$ and due to a property of the normal distribution. Therefore, even when potential buyers have rational expectations, non-occupants react more strongly to the demand shock underlying the impulse response.

Walrasian

In the Walrasian version of our model, a mechanism selects a price each period so that the number of potential buyers willing to buy at that price equals the number of movers willing to sell. The main model assumes that each mover matches to a potential buyer with probability one, which implicitly assumes that the potential buyer population moves in proportion to the mover population. To maintain comparability with the main model, we make an analogous assumption in the Walrasian variant that the number of potential buyers at time t is NI_t, where $N > 1$ is a constant.

Here, we describe equilibrium in which all movers sell. In this case, the cap rate error implies the equation

$$I_t = NI_t\left(1 - F\left(\log\kappa + \log P_t - d_t\right)\right).$$

Solving for P_t yields what agents believe is the equilibrium pricing function:

$$\tilde{P}(d_t) = \kappa^{-1}e^{F^{-1}(1-N^{-1})}e^{d_t} = \tilde{p}e^{d_t}.$$

In equilibrium, movers must weakly prefer selling at this price versus waiting to sell next period. Therefore, we must have $e^{d_t} \geq (1 + r_m)^{-1}E_te^{d_{t+1}}$, where E_t denotes the mover expectation that we now specify. By observing the current and past prices, movers believe that they observe the history of demand as $\tilde{d}_{t-j} = \log(\tilde{p}^{-1}P_{t-j})$ for $j \geq 0$. By a Kalman filtering argument similar to the proof of Lemma 1, the mover posterior on g_t at t has mean

$$\hat{g}_t^m = \mu + (1-\alpha)\sum_{j=0}^{\infty}(\alpha\rho)^j\left(\Delta\tilde{d}_{t-j} - \mu\right) = \mu + (1-\alpha)\sum_{j=0}^{\infty}(\alpha\rho)^j\left(\Delta\log P_{t-j} - \mu\right)$$

and variance σ_l^2. We have $d_{t+1} = d_t + g_{t+1} + \epsilon_{t+1}^d = d_t + (1-\rho)\mu + \rho g_t + \epsilon_{t+1}^g + \epsilon_{t+1}^d = d_t + (1-\rho)\mu + \rho\hat{g}_t^m + \rho\zeta_t^g + \epsilon_{t+1}^g + \epsilon_{t+1}^d$. Therefore,

$$E_te^{d_{t+1}} = e^{d_t}e^{(1-\rho)\mu+\rho\hat{g}_t^m}e^{(\rho^2\sigma_l^2+\sigma^2)/2}.$$

Mover optimality therefore requires that

$$\hat{g}_t^m \le \rho^{-1} \left(\log(1 + r_m) - (1 - \rho)\mu - (\rho^2 \sigma_l^2 + \sigma^2)/2 \right).$$

This inequality cannot hold at all times because $\hat{g}_t^m$ is unbounded. Therefore, when the expected growth rate is sufficiently high, some movers will refrain from selling their homes at the Walrasian equilibrium price. However, we check that the inequality holds for all $\hat{g}_t^m$ in the discrete mesh and also for all realized values in the simulations. For our parameters, the right side equals 0.15, which is much larger than the maximal realized value of 0.03. Therefore, in our simulations, we assume the approximation that the equilibrium always features full sale by all movers at all times.

We now solve for the optimal potential buyer decision, which determines the true pricing function. For $j \ge 1$, potential buyers set $\Delta \tilde{d}_{t-j} = \Delta \log P_{t-j}$. They face the same filtering problem on g_t as potential buyers in the main model, so their posterior mean $\hat{g}_t$ follows the formula in Lemma 1. Because they sell immediately in the approximate equilibrium we consider, the mover value is just the price, $V_t^m = \tilde{p}e^{\tilde{d}_t}$. (In fact, even in the exact equilibrium, the mover value coincides with the price because movers are indifferent between selling and not.) The remainder of the derivation follows the proof of Lemma 3 closely, so we omit it. That is, there exist functions $\kappa^\lambda(\hat{g}_t)$ such that a potential buyer purchases a house if and only if $e^{d_i} \ge \kappa^{\lambda_i}(\hat{g}_t)P_t$. The functions no longer depend on n because the private flow utility d_i is uninformative about d_t, as potential buyers believe that they observe d_t perfectly via $\tilde{d}_t = \log(\tilde{p}^{-1}P_t)$. The actual equilibrium price must satisfy

$$I_t = NI_t \left(1 - \sum_\lambda \beta^\lambda F(\log \kappa(\hat{g}_t) + \log P_t - d_t) \right),$$

for which it is clear that a unique solution always exists of the form $P_t = p(\hat{g}_t)e^{d_t}$. We discretize the $\hat{g}_t$ space and solve for the pricing function $p(\cdot)$ and the $\kappa^\lambda(\cdot)$ functions at these values, interpolating/extrapolating in between and beyond the mesh.

We then simulate the model as in the main text. The price paths seem to be explosive under the baseline parameters. We believe that prices explode because they adjust more quickly with Walrasian market clearing. In any event, to maintain comparability with the main model, we decrease γ to 0.042 so that the price overshoot is the same in the Walrasian model as in the main model, and we update κ so that the demand error is still zero on average. Other parameters remain the same.

Results appear in Figure IA.III. Prices and volume both go through a boom and bust cycle in the Walrasian model, as in the main model. However, volume now peaks after prices so there is no longer a quiet. The price boom is faster, with prices reaching their peak 9 quarters after the shock instead of 15. Under Walrasian market clearing, prices react more quickly to new information, explaining the absence of the quiet and the shorter duration of the price boom. Short-term and non-occupant volume continue to rise disproportionately in the Walrasian model, so these aspects of the baseline model do not depend on our departure from Walrasian market clearing.

V   Additional Exhibits

FIGURE IA.I

Non-Primary Homebuying and House Price Appreciation

Panel A. View of Housing as Investment vs. P(Buying Non-Primary Home)
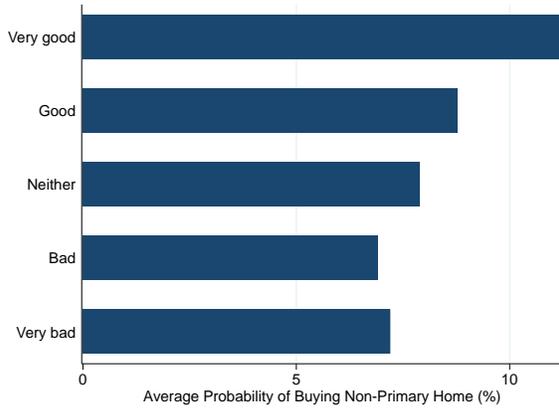
Panel B. View of Housing as Investment vs. Recent House Price Appreciation
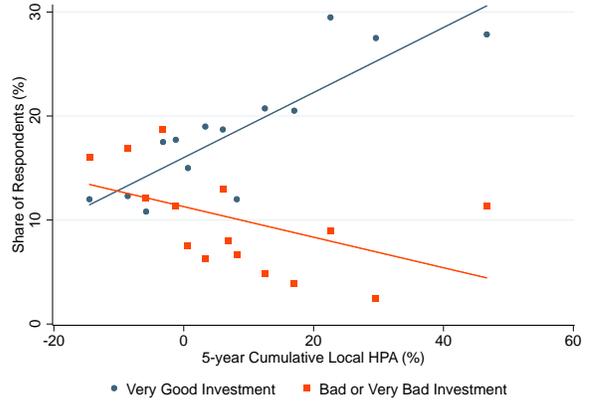
Panel C. P(Buying Non-Primary Home) vs. Recent House Price Appreciation
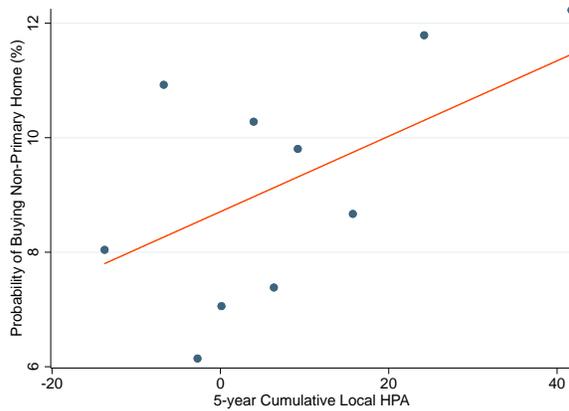
Panel D. P(Buying Non-Primary Home) vs. Recent House Price Appreciation, Savings
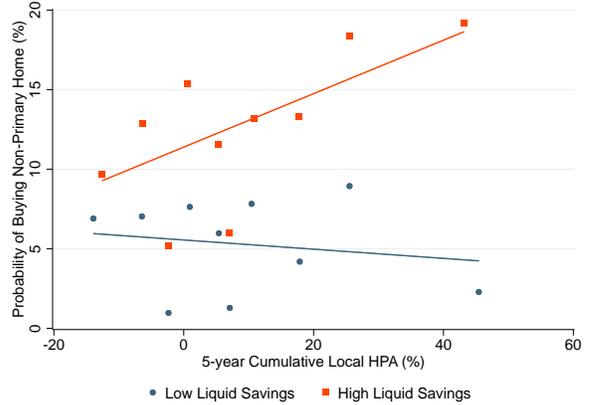
Notes: This figure uses data from the Federal Reserve Survey of Consumer Expectations and Armona, Fuster and Zafar (2016) to study the relation between recent house price growth and the probability of buying a non-primary home. In this data, local house price appreciation is computed at the ZIP-level from Zillow. High versus low liquid savings refer to those below the 25th and above the 75th percentiles, respectively, where the 25th percentile is $1,500 and the 75th percentile is $175,000.

FIGURE IA.II
Impulse Responses, Rational Model
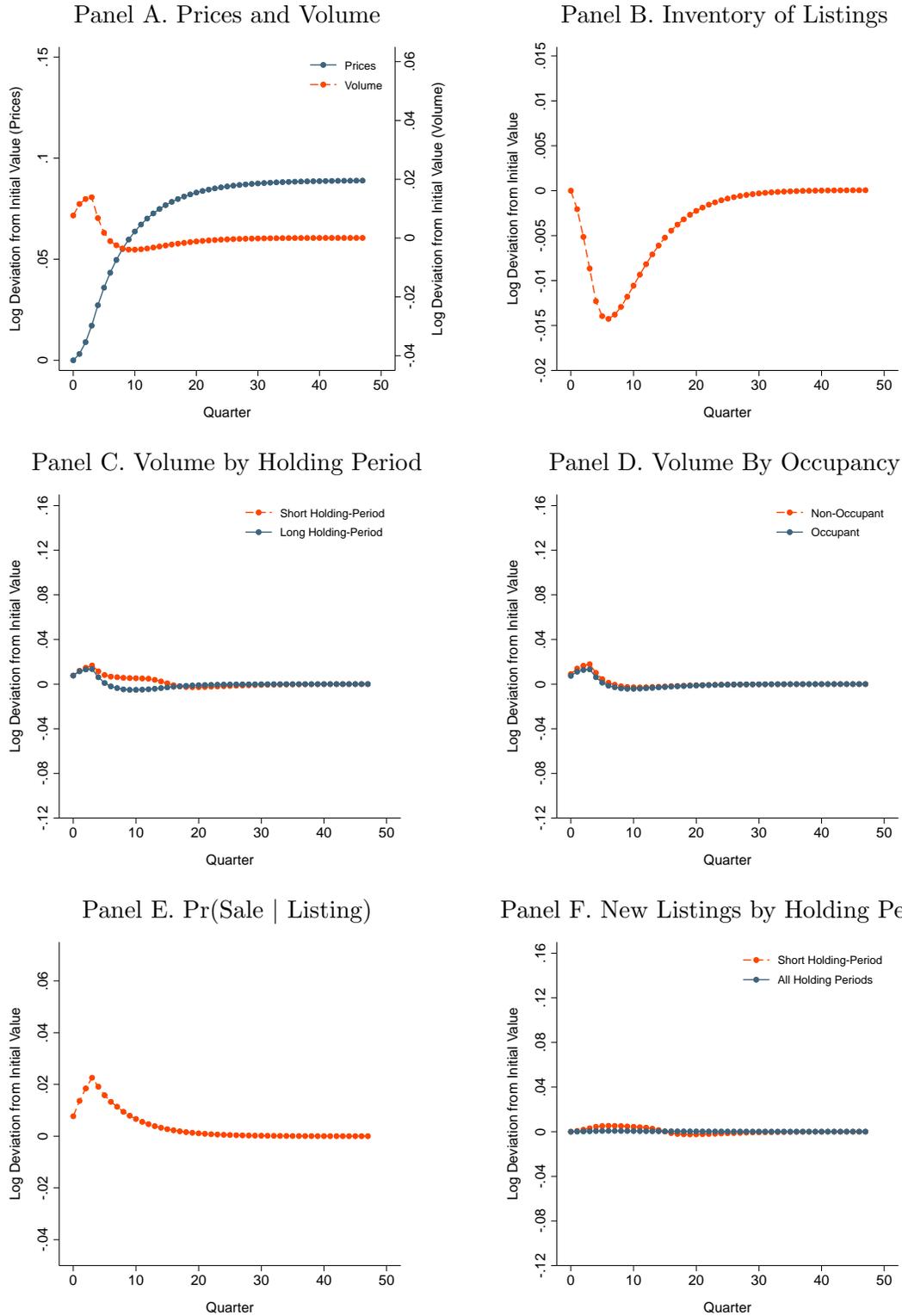
Panel A. Prices and Volume

Panel B. Inventory of Listings

Panel C. Volume by Holding Period

Panel D. Volume By Occupancy

Panel E. Pr(Sale | Listing)

Panel F. New Listings by Holding Period

Notes: Impulse responses are average differences between log outcomes in control simulations and treatment simulations, in which a 2-standard-deviation shock to ϵ_t^g (the demand growth innovation) occurs in quarters 0 through 3. A short holding period is defined as less than or equal to 12 quarters and a long holding period is defined as greater than 12 quarters.

FIGURE IA.III
Impulse Responses, Walrasian Model
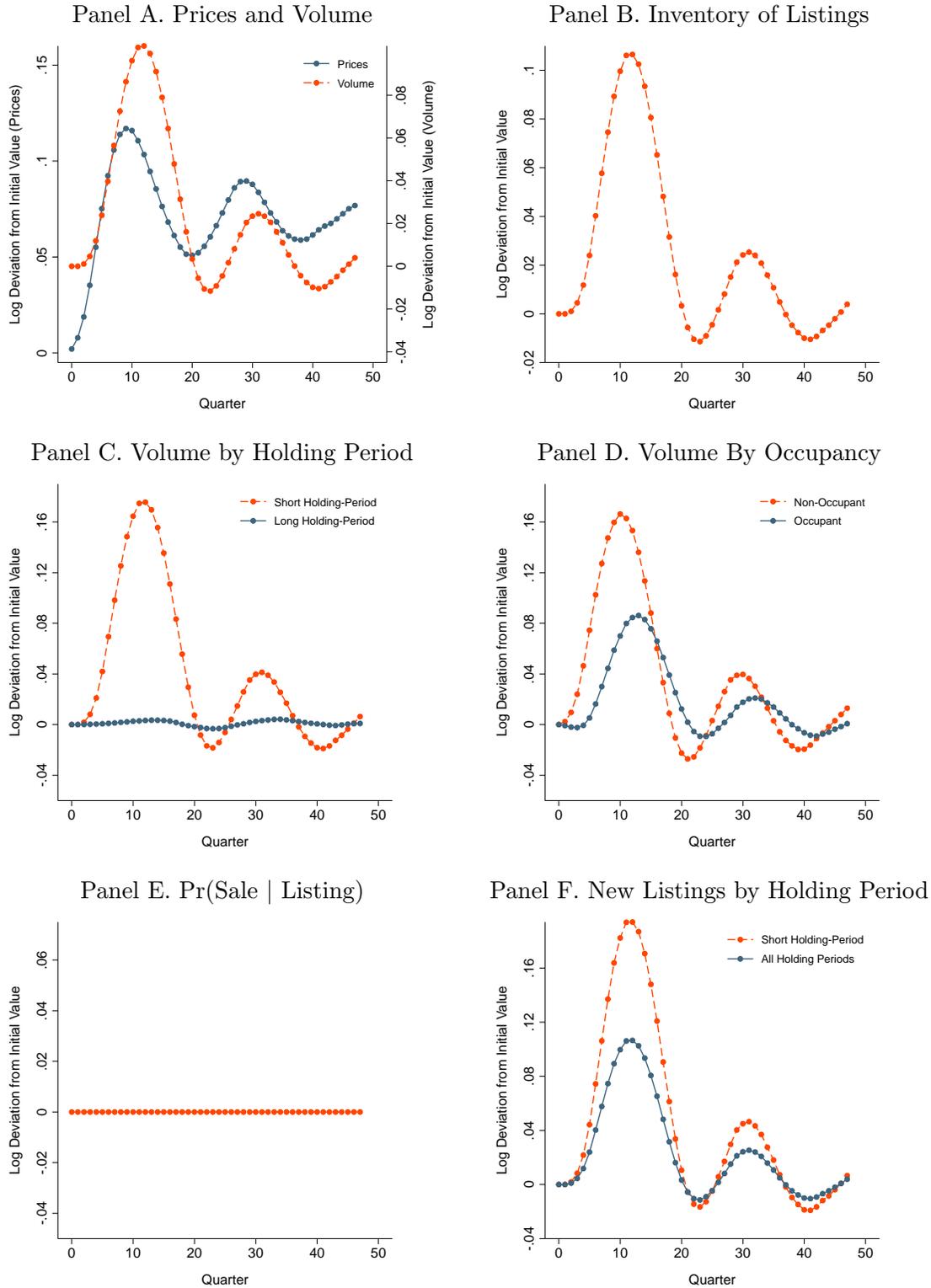
Panel A. Prices and Volume

Panel B. Inventory of Listings

Panel C. Volume by Holding Period

Panel D. Volume By Occupancy

Panel E. Pr(Sale | Listing)

Panel F. New Listings by Holding Period

Notes: Impulse responses are average differences between log outcomes in control simulations and treatment simulations, in which a 2-standard-deviation shock to ϵ_t^g (the demand growth innovation) occurs in quarters 0 through 3. A short holding period is defined as less than or equal to 12 quarters and a long holding period is defined as greater than 12 quarters.

TABLE IA.I
All-Cash Buyer Shares and Mean LTV by Buyer Type

	Transaction-Level	MSA-Level			
	All Months	All Months	Boom	Quiet	Bust
	All-Cash Buyer Share				
Short Buyers	0.29	0.38	0.29	0.28	0.52
		(0.21)	(0.16)	(0.17)	(0.20)
Non-Occupant Buyers	0.38	0.41	0.36	0.32	0.50
		(0.18)	(0.15)	(0.14)	(0.18)
All Buyers	0.20	0.25	0.22	0.20	0.30
		(0.16)	(0.15)	(0.14)	(0.16)
	Mean LTV				
Short Buyers	0.59	0.52	0.60	0.59	0.41
	(0.40)	(0.18)	(0.13)	(0.13)	(0.17)
Non-Occupant Buyers	0.50	0.48	0.52	0.54	0.41
	(0.41)	(0.14)	(0.12)	(0.11)	(0.15)
All Buyers	0.65	0.62	0.64	0.64	0.59
	(0.36)	(0.13)	(0.12)	(0.11)	(0.14)
	Mean LTV | LTV > 0				
Short Buyers	0.84	0.85	0.84	0.82	0.85
	(0.16)	(0.06)	(0.05)	(0.04)	(0.07)
Non-Occupant Buyers	0.81	0.82	0.82	0.80	0.82
	(0.17)	(0.06)	(0.06)	(0.05)	(0.06)
All Buyers	0.82	0.83	0.82	0.80	0.85
	(0.16)	(0.05)	(0.04)	(0.04)	(0.05)

Notes: This table presents statistics on LTV ratios and the share of buyers of various types who purchased their homes without the use of a mortgage. In column 1, statistics are measured at the transaction level and includes all transactions recorded between January 2000 and December 2011 from the CoreLogic deeds records described in Section 2.1. The first row of each panel includes only transactions by homebuyers who are observed to have sold the home within three years of purchase. The second row of each panel includes only non-occupant buyers. The third row of each panel includes all buyers. In columns 2–5, means are first calculated at the MSA-by-month level and then averaged across MSA-months within a given time period. The standard deviation of these MSA-month means is reported in parentheses. Column 2 includes all MSA-months between January 2000 and December 2011. Column 3 includes only MSA-months between January 2000 and August 2005. Column 4 includes only MSA-months between August 2005 and December 2006. Column 5 includes only MSA-months between December 2006 and December 2011. All statistics are calculated in the full sample of 115 MSAs with the exception of those for non-occupants, which are calculated in the sample of 102 MSAs with valid non-occupancy data.

TABLE IA.II
Speculators and Housing Market Outcomes (Summary Statistics)

Panel A. Short-Volume Sample

Variable	Mean	Standard Deviation	Observations
Short-Volume Boom	15.97	12.93	115
Price Boom	97.06	47.88	115
Price Bust	-27.9	13.64	115
Δ Volume Quiet + Bust	-62.96	18.87	115
Foreclosures Bust	82.84	55.96	115

Panel B. Non-Occupant Volume Sample

Variable	Mean	Standard Deviation	Observations
Non-Occupant Volume Boom	29.29	27.05	102
Short-Volume Boom	16.88	13.36	102
Price Boom	100.57	49.27	102
Price Bust	-28.99	13.97	102
Δ Volume Quiet + Bust	-63.32	19.47	102
Foreclosures Bust	86.57	58.08	102

Panel C. Short-Volume Sample with Listings

Variable	Mean	Standard Deviation	Observations
Short-Volume Boom	14.64	12.33	57
Δ Listings Boom	91.67	94.93	57
Δ Listings Quiet	178.39	143.86	57

Panel D. Non-Occupant Volume Sample with Listings

Variable	Mean	Standard Deviation	Observations
Non-Occupant Volume Boom	27.81	27.32	48
Short-Volume Boom	15.84	12.88	48
Δ Listings Boom	82.11	93.67	48
Δ Listings Quiet	171.74	151.29	48

Notes: This table reports summary statistics for MSA-level variables in different samples of MSAs in Table 2. Δ Volume Quiet + Bust is defined as the change in total volume from 2005 through 2011. Δ Listings Boom is defined as the change in total listings from 2003 through 2005. Δ Listings Quiet is defined as the change in total listings from 2005 through 2007. Foreclosures Bust is defined as total foreclosures from 2007 through 2011. Price Boom is defined as the change in prices from 2000 through 2006. Price Bust is defined as the change in prices from 2006 through 2011. To aid interpretation of these relations, we scale the change in outcomes for all quantity measures relative to total volume in 2003 and multiply by 100. Total volume in 2003 has mean 28,061 and standard deviation 43,708 in the Short Volume Sample and mean 25,167 and standard deviation 35,967 in the Short Volume Sample with Listings.

TABLE IA.III
Speculators and Housing Market Outcomes (Additional Listing Outcomes)

Panel A. Propensity to List

	Δ New Listings Boom			Δ New Listings Quiet		
Short-Volume Boom	0.270		0.540	0.649***		0.452
	(0.182)		(0.349)	(0.160)		(0.305)
Non-Occupant Volume Boom		0.115	-0.097		0.308***	0.130
		(0.092)	(0.165)		(0.080)	(0.144)
Number of Observations	57	48	48	57	48	48
R-squared	0.038	0.033	0.082	0.229	0.243	0.278

Panel B. Sale Probability

	Δ P(Sale) Boom			Δ P(Sale) Quiet		
Short-Volume Boom	0.142***		0.100	-0.163***		-0.273***
	(0.032)		(0.064)	(0.031)		(0.059)
Non-Occupant Volume Boom		0.058***	0.019		-0.047**	0.061**
		(0.017)	(0.030)		(0.018)	(0.028)
Number of Observations	57	48	48	57	48	48
R-squared	0.268	0.206	0.247	0.332	0.122	0.404

Notes: This table reports estimates of the relation between speculative volume and housing cycle measures at the MSA level. Short-Volume Boom has a mean of 16.0% and a standard deviation of 12.9%. Non-Occupant Volume Boom has a mean of 29.3% and a standard deviation of 27.1%. Δ New Listings Boom is defined as the change in the flow of listings from 2003 through 2005. Δ New Listings Quiet is defined as the change in the flow of listings from 2005 through 2007. These outcomes correspond to listing propensities among existing homeowners. Δ P(Sale) Boom is defined as the change in the probability of sale among the observed stock of listings from 2003 through 2005. Δ P(Sale) Quiet is defined as the change in the probability of sale among the observed stock of listings from 2005 through 2007. To aid interpretation of these relations, we scale the change in outcomes for all quantity measures relative to total volume in 2003. We do not scale the sale probability. Significance levels 10%, 5%, and 1% are denoted by *, **, and ***, respectively.

TABLE IA.IV

First-Stage Regressions of Demographics on Short Volume Boom

Old Share	1.69	0.31
	(0.26)	(0.11)
Young Share	0.66	0.51
	(0.32)	(0.15)
MSA-level	X	
ZIP-level (MSA Effects)		X
Number of Observations	102	6826
R-squared	0.45	0.32

Notes: This table presents first-stage regressions corresponding to the IV specification in Table 4. Demographic data come from the 2000 Census 5% microdata. The Young Share is the share of recent buyers under 35. The Old Share is the share of recent buyers aged 65 or older. The ZIP-level regression is estimated with MSA fixed effects and with standard errors clustered at the MSA level. The F-statistics in the MSA-level and ZIP-level (Kleibergen-Paap Wald F-statistic reflecting MSA-level clustering) regressions are 40 and 8, respectively.

TABLE IA.V

Mechanical Short-Term Volume Estimates

Year	$\hat{\alpha}_y^{buy} - \hat{\alpha}_{2000}^{buy}$	Total Volume	Actual Short-Term Volume	Counterfactual Short-Term Volume
2000	0	2821596	512787	512787
2001	0.0003	2757954	499643	494741
2002	0.0008	2985550	556987	534342
2003	0.0014	3226968	614429	557701
2004	0.0023	3667997	772708	659111
2005	0.0027	3857236	909976	725847
2000–2005 growth	–	36.7%	77.5%	41.5%

Notes: Total Volume gives annual transaction counts in our analysis sample. Actual Short-Term Volume are sales of properties for which the previous purchased occurred less than 36 months in the past. We estimate α_y^{buy}, a fixed effect for the propensity to sell a house having bought it in year y, using the regression equation in Section 5.2. In the counterfactual, we assume that α_y^{buy} remains constant at its level in $y = 2000$ for $y \in \{2001, 2002, 2003, 2004, 2005\}$.

List of Metropolitan Statistical Areas Included in the Analysis Sample

Metropolitan Statistical Area	Share of Housing Stock Represented	Included in Non-Occupant Analysis	Included in Listings Analysis
Akron, OH	1.00	x	x
Ann Arbor, MI	1.00	x	x
Atlanta-Sandy Springs-Roswell, GA	0.80		
Atlantic City-Hammonton, NJ	1.00	x	x
Bakersfield, CA	1.00	x	x
Baltimore-Columbia-Towson, MD	1.00	x	
Barnstable Town, MA	1.00		x
Bellingham, WA	1.00	x	x
Bend-Redmond, OR	1.00	x	
Boston-Cambridge-Newton, MA-NH	0.89		x
Boulder, CO	1.00	x	x
Bremerton-Silverdale, WA	1.00	x	x
Bridgeport-Stamford-Norwalk, CT	1.00		x
Buffalo-Cheektowaga-Niagara Falls, NY	0.80	x	x
California-Lexington Park, MD	1.00	x	
Canton-Massillon, OH	0.92	x	x
Cape Coral-Fort Myers, FL	1.00	x	x
Champaign-Urbana, IL	0.82	x	
Charleston-North Charleston, SC	0.79	x	
Chicago-Naperville-Elgin, IL-IN-WI	0.90		
Chico, CA	1.00	x	
Cincinnati, OH-KY-IN	0.78	x	x
Cleveland-Elyria, OH	1.00	x	x
Colorado Springs, CO	0.95	x	
Crestview-Fort Walton Beach-Destin, FL	1.00	x	
Dallas-Fort Worth-Arlington, TX	0.85	x	
Dayton, OH	0.86	x	x
Deltona-Daytona Beach-Ormond Beach, FL	1.00	x	x
Denver-Aurora-Lakewood, CO	0.95	x	
El Centro, CA	1.00	x	
El Paso, TX	0.99	x	x
Elmira, NY	1.00	x	
Erie, PA	1.00	x	
Eugene, OR	1.00	x	x
Flagstaff, AZ	1.00	x	
Fort Collins, CO	1.00	x	x
Fresno, CA	1.00	x	
Gainesville, FL	0.91	x	
Gainesville, GA	1.00		
Hanford-Corcoran, CA	1.00	x	
Hartford-West Hartford-East Hartford, CT	1.00		x
Homosassa Springs, FL	1.00	x	x
Ithaca, NY	1.00	x	x
Jacksonville, FL	0.98	x	
Kahului-Wailuku-Lahaina, HI	1.00	x	x
Kingston, NY	1.00	x	x
Lake Havasu City-Kingman, AZ	1.00	x	x
Lakeland-Winter Haven, FL	1.00	x	
Lancaster, PA	1.00	x	x
Las Vegas-Henderson-Paradise, NV	1.00	x	
Los Angeles-Long Beach-Anaheim, CA	1.00	x	x
Madera, CA	1.00	x	
Merced, CA	1.00	x	x
Miami-Fort Lauderdale-West Palm Beach, FL	1.00	x	
Modesto, CA	1.00	x	x
Napa, CA	1.00	x	
Naples-Immokalee-Marco Island, FL	1.00	x	x
New Haven-Milford, CT	1.00		x
New York-Newark-Jersey City, NY-NJ-PA	0.97	x	
North Port-Sarasota-Bradenton, FL	1.00	x	
Norwich-New London, CT	1.00		x
Ocala, FL	1.00	x	x
Ocean City, NJ	1.00	x	x
Olympia-Tumwater, WA	1.00	x	x
Orlando-Kissimmee-Sanford, FL	1.00	x	
Oxnard-Thousand Oaks-Ventura, CA	1.00	x	x
Palm Bay-Melbourne-Titusville, FL	1.00	x	
Pensacola-Ferry Pass-Brent, FL	1.00	x	
Philadelphia-Camden-Wilmington, PA-NJ-DE-MD	1.00	x	
Phoenix-Mesa-Scottsdale, AZ	1.00	x	x
Pittsfield, MA	1.00		
Portland-Vancouver-Hillsboro, OR-WA	0.97	x	x
Port St. Lucie, FL	1.00	x	x
Prescott, AZ	1.00	x	x
Providence-Warwick, RI-MA	0.78		x
Punta Gorda, FL	1.00	x	
Raleigh, NC	0.78	x	
Reading, PA	1.00	x	
Redding, CA	1.00	x	
Reno, NV	0.99	x	x
Riverside-San Bernardino-Ontario, CA	1.00	x	x
Rockford, IL	0.84	x	
Sacramento–Roseville–Arden-Arcade, CA	1.00	x	x
Salem, OR	0.79	x	
Salinas, CA	1.00	x	
San Diego-Carlsbad, CA	1.00	x	x
San Francisco-Oakland-Hayward, CA	1.00	x	x
San Jose-Sunnyvale-Santa Clara, CA	1.00	x	
Santa Cruz-Watsonville, CA	1.00	x	
San Luis Obispo-Paso Robles-Arroyo Grande, CA	1.00	x	x
Santa Maria-Santa Barbara, CA	1.00	x	
Santa Rosa, CA	1.00	x	
Seattle-Tacoma-Bellevue, WA	1.00	x	x
Sebastian-Vero Beach, FL	1.00	x	
Sebring, FL	1.00	x	
Sierra Vista-Douglas, AZ	1.00	x	
Spokane-Spokane Valley, WA	0.87	x	
Springfield, IL	0.93	x	
Springfield, MA	1.00		x
Springfield, OH	1.00	x	
Stockton-Lodi, CA	1.00	x	x
Tampa-St. Petersburg-Clearwater, FL	1.00	x	
The Villages, FL	1.00	x	
Toledo, OH	0.92	x	x
Trenton, NJ	1.00	x	
Tucson, AZ	1.00	x	x
Urban Honolulu, HI	1.00	x	x
Vallejo-Fairfield, CA	1.00	x	
Vineland-Bridgeton, NJ	1.00	x	x
Visalia-Porterville, CA	1.00	x	
Washington-Arlington-Alexandria, DC-VA-MD-WV	0.95	x	
Worcester, MA-CT	1.00		x
Youngstown-Warren-Boardman, OH-PA	0.80	x	x
Yuba City, CA	1.00	x	
Yuma, AZ	1.00	x	

Notes: This table lists the Metropolitan Statistical Areas that are included in the final analysis sample along with the share of the total 2010 owner-occupied housing stock for each MSA that is represented by the subset of counties for which CoreLogic has consistent data coverage back to 1995.

TABLE IA.VII

Number of Transactions Dropped During Sample Selection

Original Number of Transactions	57,668,026
Dropped: Non-unique CoreLogic ID	50
Dropped: Non-positive price	3,309,100
Dropped: Nominal foreclosure transfer	531,786
Dropped: Duplicate transaction	609,756
Dropped: Subdivision sale	1,304,920
Dropped: Vacant lot	831,774
Final Number of Transactions	51,080,640

Notes: This table shows the number of transactions dropped at each stage of our sample-selection procedure.