Abstract

Using data on 50 million home sales from the last U.S. housing cycle, we document that much of the variation in volume came from the rise and fall in speculation. Cities with larger speculative booms have larger price booms, sharper increases in unsold listings as the market turns, and more severe busts. 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.
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 et al., 2018) and contributed to a capital overhang that slowed the economic recovery from the subsequent recession (Rognlie et al., 2017). During the bust, millions of households lost their homes in foreclosure, and falling house prices led many others to cut consumption (Mayer et al., 2009; Mian et al., 2013, 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. 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 from CoreLogic on 50 million home sales between 1995 and 2014. We measure speculative buying and selling across 115 metropolitan statistical areas (MSAs), which represent 48% of the U.S. housing stock. We pursue two complementary approaches to identify speculative activity. First, following Bayer et al. (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 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

---

1Harrison 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.”
the role of speculative buyers for inventory dynamics across MSAs.

The data reveal a strong relation between the differential entry of speculative buyers and
the size of the cycle. While overall volume increases substantially during the boom of 2000–
2005, both short-term and non-occupant volume rise dramatically more. In an accounting
sense, growth in speculative volume explains 40% to 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 slows but remains positive, and unsold list-
ings 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. We find that speculative volume is larger
when house price growth over the past year is greater, which suggests that extrapolation—
the belief that prices continue to rise after recent gains—draws speculators into the housing
market. Consistent with our interpretation of the data, a National Association of Real-
tors 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.

The second part of the paper builds a quantitative model to match these novel facts
about the housing market. Our approach adapts core insights from Cutler et al. (1990),
De Long et al. (1990), and Hong and Stein (1999) to study the housing market. As in these
papers, extrapolation causes a predictable boom and bust in 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 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. The model is able to match key facts from our empirical work, including the lead–lag relation between prices and volume, the excess growth of short-term and non-occupant volume during the boom, and a growth in listings during the quiet coming disproportionately from short-holding-period sales. 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 evaluate the effect of speculation on the housing cycle. When we shut down speculation by imposing rational expectations, almost all of the salient aspects of the housing cycle disappear completely or become quantitatively insignificant. We find similar patterns when we remove short-term and non-occupant buyers from the model. Therefore, speculators amplify the effects of non-rational expectations on prices and quantities over the housing cycle.

The model also allows us to measure the relative importance of short-term versus non-occupant buyers. Removing either group from the model attenuates the housing cycle, but there is substantial overlap between these two groups, a fact we confirm in the data. When we eliminate non-occupants while keeping constant the share of short-term buyers, the housing cycle remains strong, which suggests that short horizons are the key amplifying force in the model, as opposed to non-occupancy. Motivated by this result, we study transaction taxes on non-occupant buyers as well as on all buyers, as governments have used such taxes in attempts to curb speculation (Chi et al., 2021). Taxing all buyers attenuates the housing cycle, but even a large 5% tax on just non-occupants has only a small effect on the price boom, price bust, and volume boom.

Previous empirical work has examined short-term buyers (Adelino et al., 2016; Bayer
et al., 2020, 2016) and non-occupant buyers (Haughwout et al., 2011; Bhutta, 2015; Gao et al., 2019; Chinco and Mayer, 2015) during this cycle. We add to this literature in three 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, relying on mortgage records likely undercounts speculation. Second, the number of MSAs in our sample—115—is considerably larger than in some other work, allowing us to establish cross-MSA relations between speculation and other aspects of the 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.

Prior theoretical work has explained aspects of the comovement of prices and volume in the housing market and asset markets more generally. As we discuss in Section 6, existing models struggle to generate simultaneously three key patterns from our empirical work: the existence of the quiet, the disproportionate growth in short-term volume during the boom and quiet, and the excess growth in non-occupant purchases during the boom. Our model, which features heterogeneous holding periods, non-Walrasian market clearing, and extrapolation, matches these facts quantitatively. In addition, our model illustrates a mechanism for how speculation amplifies the housing cycle and explains why short-term purchases have a larger effect than those from non-occupants.

1 Data

This section describes the data we use to establish the core motivating facts for our model and how we identify speculative buyers in that data. Further information regarding the data is in Appendix A.

1.1 Data Sources and Sample Selection

We use data on individual housing transactions from CoreLogic, a private vendor that collects and standardizes publicly available tax assessments and deeds records from 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 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 A describes how we 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 pass the following filters: (a) the transaction is categorized by CoreLogic as 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 selling 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. Appendix A specifies the 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. Appendix A 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 and MSA-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 res-
idence, 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.

1.2 Identifying Speculators

We identify speculators in our transaction-level data using two complementary approaches, each of which has been used in prior work. In the first approach, we categorize transactions based on 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 et al. (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.

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 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.

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 et al., 2011; Gao et al., 2019; Mian and Sufi, 2019). While based on similar ideas, such approaches run the risk of omitting a substantial fraction of speculative activity.

2 Dynamics of Prices, Volume, and Inventory

This section presents three stylized facts from the last U.S. housing cycle that guide our empirical and theoretical 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 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 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 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. 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. We normalize transaction volume in each MSA-month by dividing by the MSA’s total housing stock from the 2000 Census. We then run regressions of the form:

\[ p_{i,t} = \beta_{\tau} v_{i,t-\tau} + \eta_{i,t}, \]

where \( p \) is log price, \( v \) is volume, \( i \) indexes MSAs, and time is measured in months. To 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.\(^2\)

The coefficient \( \beta_{\tau} \) provides an estimate of how within-MSA–calendar-month movements in volume at a \( \tau \)-month lag correlate with contemporaneous within-MSA movements in prices. We run these regressions separately for up to 4 years of lags (\( \tau = 48 \)) and one year of leads (\( \tau = -12 \)). Figure 2 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.

\(^2\)For other work regressing house prices on lagged transaction volume, see Leung et al. (2002), Clayton et al. (2010), and Head et al. (2014).
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.\textsuperscript{3} During the period when the relation between volume and prices reverses, aggregate inventories rise dramatically to nearly double their level from earlier in 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 earlier levels. In Daytona Beach, inventories rise to 450\% of their pre-quiet levels.\textsuperscript{4}

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.

3 Speculators During the Cycle

3.1 Quantities and Prices in the Boom

Figure 4 presents a simple illustration of the quantitative importance of speculation during the 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 deeds data. Each series is normalized relative to its average value in the year 2000 and seasonally adjusted by removing calendar-month fixed effects. For reference, we also report the raw counts of each type of transaction in the years 2000, 2005, and 2010. 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 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

\textsuperscript{3}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.

\textsuperscript{4}Figures IA1, IA2, and IA3 repeat the analyses in Figures 1–3 for MSAs outside the sand states, revealing the patterns we document are not exclusive to these states.
of the overall increase in volume. 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% 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% 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% 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 cross-MSA differences in speculative volume growth explain much of the differences in the total volume growth. For each MSA, we plot the change in either short-holding-period volume (Panel C) or non-occupant volume (Panel D)

5Part of the late-boom increase in short-term sales may have been mechanically related to the overall increase in volume of all types during the early part of the boom. Appendix B.1 uses conditional selling hazards by buyer cohort to quantify the contribution of an overall increase in total volume to the share of late-boom volume coming from short-term sales. Approximately 90% of the rise in short-term volume appears due to the changing composition of buyers between 2000 and 2005, rather than mechanical forces.

6One concern with our short-term speculation measure is that it is 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. Appendix B.2 presents instrumental variable regressions that predict short-term volume using pre-cycle demographics. The change in realized short-term volume is quantitatively important for overall volume growth and the size of the price cycle, even when using only the portion of short-term volume growth predicted by ex-ante buyer characteristics.
divided by initial total volume on the y-axis against the percent change in total volume on the x-axis. The slope 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%, or 36% excluding new construction. 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 are a major determinant of changes in total volume during the boom.

Table 2 shows how speculative volume relates to the size of the price and quantity cycles in the cross-section of MSAs (Table 1 shows summary statistics). We estimate the correlation between growth in each speculative measure and various housing market outcomes by separately regressing these outcomes on each measure of speculation. To aid interpretation, we scale the change in outcomes for all quantity measures relative to total volume in 2003.

The first two columns of Panel A show that 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. 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% in the boom and quiet. Thus, the relation between speculative volume and prices is economically large in the cross-section of MSAs.

3.2 Quantities and Prices in the Quiet and Bust

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 3–
4, 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 quiet of 76.9 percentage points relative to 2003 total volume. 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.\textsuperscript{10}

Consistent with the aggregate evidence in Figure 3, which shows a modest increase in listings during the boom, we find a small, statistically insignificant relation across MSAs between speculative booms and the change in listings during the boom (Panel B, columns 1–2).\textsuperscript{11} Given the strong relation between the short-term and total volume booms, this fact suggests that the increase in demand during the boom was sufficient to absorb the rising flow of listings from short-term buyers. As demand slowed, the continuing flow of listings from recent buyers saturated the market, leading to accumulating inventories in the quiet.

Figure 6 demonstrates this idea with listings data linked to transaction data at the property level. The link to past transactions allows us to see whether recent purchases disproportionately contribute to the surge of listings in the quiet. We plot monthly aggregate series for total listings and short-holding-period listings, defined as a listing where the prior 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, thus measuring the flow of short-holding-period listings without double counting unsold listings. Each series is normalized relative to its average value in 2003 and seasonally adjusted by removing calendar-month effects.

The increase in listings during the quiet comes largely from recent purchases. While

\textsuperscript{10} Table 2 reports the change in the inventory of unsold listings. Table IA7 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 is 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. Tables IA8 and IA9 repeat the analysis in Tables 2 and IA7, while including an indicator for whether the MSA is in a sand state. The results are similar, though somewhat weaker for for the non-occupant volume boom.

\textsuperscript{11} Table 1 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%).
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. 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 to 590 thousand, thus accounting for 55% of the rise in total listings. 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.\footnote{This evidence complements Genesove and Mayer (1997, 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.}

Table 2, Panel C and columns 3–4 of Panel A, show how 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) that are 13.5 and 13.9 percentage points larger. The analogous results for prices imply 7.4 and 4.5 percentage point larger 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 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 in the bust. While not a focus of our theoretical analysis, this outcome has received much attention in prior research given its importance for policy and macroeconomic outcomes. We find that the short-term speculative boom coincides with a larger number of foreclosures in 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) in the bust, equal to 370 thousand more foreclosures. During this time, there were 2.68 million foreclosures across the 115 MSAs 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.\textsuperscript{13}

4 The Role of Extrapolation among Speculators

In this section, we document that house price growth strongly predicts subsequent speculative purchases and beliefs in the housing market. This evidence, which appears in Figure 7, motivates including extrapolative beliefs in the model in Section 7.

Our first measures of speculative purchases use our deeds dataset. For each MSA and year from 2000 to 2011, we count total non-occupant purchases and divide by the equivalent count from 1999 as a normalization. We do the same for short-term purchases, defined here as those for which we observe another sale on the same property in the next three years. Panels A and B present binned scatter plots of normalized speculative purchases against house price growth in the past year. Both non-occupant and short-term purchases are much higher in the years and MSAs that witness higher house price appreciation in the last year.\textsuperscript{14}

The second measure uses responses from the NAR’s Investment and Vacation Home Buyers Survey. For each year of the survey, we calculate 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 home by the time of the survey. This measure of speculation captures the intention of buyers at the time of purchase. Thus, it complements our transaction-based metric that relies on realizations of short horizons after the fact. Panel C plots the short-term buyer share from the NAR against annual house price growth at the national level. A gain of 10\% in house prices over the past year is associated with an 8.2 percentage point larger short-term buyer share.

Our final measures use responses from the 2014–2017 waves of the Federal Reserve Bank

\textsuperscript{13}Related work documents a disproportionate share of investors among delinquencies and foreclosures. See, e.g., Haughwout et al. (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.

\textsuperscript{14}In Appendix B.4, we estimate higher-frequency panel VAR specifications of speculative volume and lagged house price appreciation, in the style of Chinco and Mayer (2015). The positive relation between prices and speculative purchases continues to hold.
of New York’s Survey of Consumer Expectations.\textsuperscript{15} This survey asks respondents’ views on housing as an investment as well as their probability of buying a non-primary home in the next three years. Thus, the survey directly queries non-occupant housing demand, complementing the measure of non-occupant purchases in our deeds data. Panels D and E present binned scatter plots of the survey measures against appreciation in the Zillow house price index over the past five years in the respondent’s ZIP code. The share of respondents saying that housing is a very good investment rises with local house price appreciation; the opposite is true for those calling housing a bad or very bad investment. The reported probability of buying a non-primary home also rises with lagged house price growth.\textsuperscript{16}

Prior work has also demonstrated the importance of extrapolative beliefs for understanding the housing market. Through surveys, Case et al. (2012) and Armona et al. (2019) document that expected future house price growth rises with realized past house price growth. On the theory front, Guren (2014) and Glaeser and Nathanson (2017) show that extrapolative expectations help explain predictable momentum and reversals in house price growth, and Glaeser et al. (2008) use extrapolative expectations to understand the variation in housing booms across MSAs. We build on this prior work, as well as the results in Figure 7, by incorporating these beliefs into our model.

5 Characterizing Speculative Buyers

Our results thus far indicate that speculative buyers are a major driver of changes in transaction volume and inventories during the boom and quiet, and that more speculative entry during the boom is associated with larger price cycles. These are the key empirical facts that we will target and seek to explain in our model. This section uses our microdata and data from other sources to shed additional light on the nature of these speculative purchases.\textsuperscript{17}

\textsuperscript{15}The data come from the replication files of Armona et al. (2019). We thank Andreas Fuster for sharing this evidence with us.

\textsuperscript{16}Figure IA4 shows this relation is stronger for respondents with high liquid savings (at least $175,000), which suggests that non-occupant housing demand is not only due to a home equity effect. We explore this idea further in Section 5.2.

\textsuperscript{17}Because this section focuses on the boom, we include foreclosures, resulting in slightly different aggregates than in Figure 4.
5.1 Overlap Between Short-Term and Non-Occupant Buyers

Our analysis has treated short-term and non-occupant buyers separately. Yet there may be substantial overlap between these groups. Here, we investigate this possibility.

We first look at data from the NAR’s Investor and Vacation Home Buyers Survey, available nationally from 2008 to 2015. These data report expected holding times separately for investor and non-investor buyers. As Figure 8 shows, about 20% of investor buyers report expected holding periods of under 3 years, larger than the corresponding share among non-investor buyers. Therefore, these data provide direct evidence of overlap between non-occupant and short-term buyers.

As a second approach, we use our CoreLogic data to measure the overlap in the 2000–2005 increase in volume between the two speculative groups. This approach focuses on the marginal speculators who entered during the boom. We find that 27% of 2000–2005 short-term volume came from non-occupant buyers, while 41% of the increase in short-term volume over this time came from non-occupants. Therefore, non-occupants account for an excess share of the growth in short-term buyers. Conversely, 26% of 2000–2005 non-occupant purchases were resold within 3 years, but short-term sales account for only 23% of the growth in non-occupant purchases over this time. That is, there was not a shift in the composition of non-occupant buyers toward short-term behavior during the boom. These facts suggest continued growth in long-term non-occupants alongside the entry of short-term speculators.18

While Table 2 shows that the short-term volume boom appears as a stronger correlate relative to the non-occupant boom for some housing cycle outcomes, separating these two types of speculation empirically is difficult. The correlation of these two booms across MSAs is 0.82. Given the significant overlap between the short-holding-period and non-occupant measures of speculation, we caution that direct comparisons between the magnitudes reported in Table 2 may be misleading. We will instead rely on our model to tease apart the relative influence of these two types of speculation.

---

18These statistics focus on the non-occupant sample of 102 MSAs. Of 2000–2005 short-term volume, 790 thousand out of 2.93 million (27%) were non-occupant buyers (excluding developers). Short-term-non-occupant-buyer transactions increase over 2000–2005 from 90 thousand to 230 thousand, 41% of the overall growth in short-term transactions (370 thousand to 710 thousand, excluding developers). Of 2000–2005 non-occupant volume, 930 thousand out of 3.60 million (26%) become short-term sellers (excluding developers). Non-occupant purchases that become short-term sales increase over 2000–2005 from 110 thousand to 210 thousand, 23% of the overall growth in non-occupant transactions (440 thousand to 880 thousand, excluding developer buyers).
5.2 The Role of Credit and Professional Investors

Here we consider additional factors that have received attention in the literature on the housing cycle. First, we ask what role credit plays in enabling speculative volume. Table 3 presents 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 do not use a mortgage. 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.\(^{19}\) Thus, while credit may have enabled speculation, there is also a disproportionately large group of speculators who do not use credit at all. The behavior of these buyers goes unobserved in any analysis of speculative activity based on mortgage data alone.

To further investigate the role of credit, we decompose the increase in short-term selling into groups of 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 ∈ \([60\%, 85\%]\)), and a high-leverage group (purchase LTV > 85%). Of the 2000–2005 short-term sellers, 31% were low-LTV buyers, 33% were medium-LTV buyers, and 36% were high-LTV buyers. In contrast, for the long-term sellers for whom 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, low-LTV, medium-LTV, and high-LTV short-term-buyer transactions account for 15%, 58%, and 27% of the growth in short-term transactions, respectively.\(^{20}\)

As in our analysis of cash transactions among speculators, these statistics reveal that short-term volume is associated with lower use of leverage in the cross-section relative to the general population.\(^{21}\) At the same time, the proportional growth in short-term buying

---

\(^{19}\)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.

\(^{20}\)Of the 2000–2005 short-term sellers with non-missing LTV, 1.24 million were low-LTV buyers, 1.33 million were medium-LTV buyers, and 1.46 million were high-LTV buyers. Between 2000 and 2005, the number of low-LTV, medium-LTV, and high-LTV short-term-buyer transactions increases from 210 to 270 thousand, from 140 to 380 thousand, and from 190 to 300 thousand, respectively.

\(^{21}\)Table IA10 extends Table 3 to look at average purchase LTVs for short-term and non-occupant buyers.
is stronger among medium- and high-LTV sellers, making a larger relative contribution to the overall growth in short-term volume. This result might explain the positive relation between short-term volume growth and foreclosures in Table 2, if many of the marginal short-term buyers who used higher LTV loans defaulted during the bust. This evidence is also consistent with high credit growth among speculative buyers during the boom, as documented by Haughwout et al. (2011), Bhutta (2015), and Mian and Sufi (2019). While speculative buyers were not all 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.

Next, we ask what share of short-term volume is due to professional investors like real estate developers. 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. In our sample, these transactions account for 6% of total volume and 9% of the growth in volume between 2000 and 2005. Of the 3.95 million short-term sales in 2000–2005, the initial purchases for 580 thousand (15%) were from developer buyers. From 2000 to 2005, the number of short-term-buyer sales increases from 530 thousand to 930 thousand while the number of short-term-developer-buyer sales increases from 100 thousand to 130 thousand, or 8% of the growth in short-term volume. Though developers were active in the housing market, they did not contribute disproportionately to short-term volume growth in the boom. A possible reason is that developers were more likely to engage in speculation in the raw land market (Nathanson and Zwick, 2018).

Finally, we ask what share of short-term buyers were experienced investors holding multiple properties versus inexperienced speculators in just one or two homes. 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 sales, 2.42 million of 3.36 million (72%) were inexperienced buyers at the time of purchase (excluding developers). Between 2000 and 2005, the number of inexperienced shares.

Both speculative buyer types have lower average LTVs, which is exclusively driven by their higher all-cash shares.

22 We identify developer names using CoreLogic’s internal new construction flag, as Appendix A describes. Both this analysis and the analysis of inexperienced investors exclude transactions with missing buyer names.
short-term-buyer sales increases from 310 thousand to 560 thousand, or 66% of the growth in short-term sales (excluding developers).

Consistent with the evidence in Bayer et al. (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 contemporaneous patterns in prices. Bayer et al. (2020) and Bayer et al. (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. Both of these patterns are consistent with the notion of extrapolation-induced entry of short-term buyers we consider in our model.

Taken together, the results in this section 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 Appendix B.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.

6 Summary of Findings and Theoretical Motivation

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 due to overconfidence (Daniel et al., 1998, 2001; Scheinkman and Xiong, 2003). The second exploits features specific to the housing market, such as credit constraints (Stein, 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 et al., 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 et al. (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 et al. (2014) and Anenberg and Bayer (2020). These models cannot explain the result that homeowners who bought later in the boom were more likely to resell than homeowners who bought earlier. Overconfidence models, such as Daniel et al. (1998, 2001), generate speculative trading that accompanies booms and busts in asset prices. In these models, an initial increase in asset prices boosts the confidence of optimistic investors, leading them to push prices up further. However, these models are not designed to fit the rise in short-term volume that occurs during booms, because the same overconfident investors buy the asset in the early as well as the late stages of a boom. Other disagreement and extrapolation–psychology papers can generate a disproportionate short-term volume boom, as long as rising prices generate more disagreement or stronger psychological urges to both buy and sell.

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 et al., 2014; Burnside et al., 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 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.\(^{24}\) Barberis et al. (2018) and Liao and Peng (2018) 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 et al. (2016). In contrast, Guren (2014) matches all three variables. However, in that model, listings sharply decline during the boom (more than one-for-one with respect to prices), and they never rise above their pre-shock level in the impulse response. Empirically, we find that listings modestly rise during the boom in the aggregate and in most MSAs (Section 3.2). The sharp rise in listings during the quiet, far above their 2000 level, is perhaps the most salient aspect of Figure 3.

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 relations between speculative volume and price and quantity outcomes. 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. While our empirical findings suggest that speculation amplifies the housing cycle, we are not identifying a causal effect. Our model permits stronger causal statements and allows us to study the speculative mechanism in more detail.

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: 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

\(^{24}\)An exogenous increase in overconfidence raises volume in Daniel et al. (2001) and Scheinkman and Xiong (2003); it raises conditional return volatility in Daniel et al. (2001) while raising the price level in Scheinkman and Xiong (2003). Disagreement accounts for some of the average prices and volume in the housing market (Bailey et al., 2016) and can generate dispersion in beliefs about house price growth over the period we are studying (Piazzesi and Schneider, 2009; Burnside et al., 2016). By definition, disagreement is less suited to explain the high average level of these beliefs (Case et al., 2012; Foote et al., 2012; Cheng et al., 2014).
whether to purchase at the listed price or exit the housing market permanently.\textsuperscript{25} If the potential buyer chooses to purchase then the mover receives the list price and exits the market. A purchasing potential buyer $i$ becomes a stayer and receives flow utility $e^{d_i}$ at the beginning of each future period until she randomly becomes a mover, which happens with Poisson hazard $\lambda_i$.

At $t$, potential buyer flow utility satisfies

$$d_i = d_t + a_i,$$  \hspace{1cm} (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 $\mu_1$ 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:

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

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

Potential buyers vary in $\lambda_i$, $a_i$, and $n_i$. The mover hazard, $\lambda_i$, follows a discrete distribution $\beta^\lambda$. The share of each occupancy type is $\beta^\lambda_n$; $\beta^\lambda_{n,\lambda}$ is the share of each $(n, \lambda)$ pair. To match the data on expected holding times (Figure 8), we allow non-zero correlation between $\lambda_i$ and $n_i$. We denote the CDF of $a_i$ across potential buyers by $F$, a mixture of two normals.\textsuperscript{26}

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

\textsuperscript{25}In other models, some movers fail to match to a potential buyer due to search frictions (Head et al., 2014; Guren, 2018). We abstract from this possibility.

\textsuperscript{26}Potential buyer types in our model bear some similarities to the taxonomies in Frankel and Froot (1986), Cutler et al. (1990) and De Long et al. (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.
\( r_m \geq r \), 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 \( e^{dt} \) 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 \( \pi_t \). 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, \( \lambda_i \), and the list price to which she 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 \( \kappa \) 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.\(^{27}\) 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 (conditions for this outcome are below), which we denote \( P_t \). In this case, the cap rate error leads agents to infer the demand shifter from market data as follows. Equations (2) and (3)\(^{27}\) As Glaeser and Nathanson (2017) explain, Eyster and Rabin (2010, 2014) propose a similar social learning rule and motivate departing from rationality when modeling social learning.
together imply that a purchase occurs if and only if:

\[ a_i \geq \log P_t + \log \kappa - d_t. \]

That is, according to the cap rate error, a potential buyer completes a sale when her idiosyncratic flow utility is at least as large as a threshold, which varies only according to the current level of prices and the demand shifter. The share of listings that sell, \( \pi_t \), therefore equals:

\[ \pi_t = 1 - F (\log P_t + \log \kappa - d_t). \]

Using this equation, agents impute the contemporaneous level of the demand shifter as:

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

where the tilde denotes an inference true under the cap rate error (but not necessarily in reality). Holding the price of housing constant, a larger \( \pi_t \) indicates to agents that a greater share of listings sold, and therefore that the concurrent demand shifter must be higher.

Because movers observe \( \pi_t \) with a one-period lag, they infer the demand shifter with this lag as well. In particular, 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 \( t' < t \).

Agents view changes to inferred demand, \( \Delta \tilde{d}_{t-j} \) for \( j \geq 1 \), as a sequence of noisy observations of the growth rate, because \( \Delta d_{t-j} = g_{t-j} + \epsilon_{t-j}^g \). We assume that agents are Bayesian updaters, which implies that they employ a standard Kalman filter to form a posterior expectation of the current growth rate, \( g_t \), as a function of this sequence of observations. The following lemma characterizes this posterior, as well as the implied posterior on \( d_t \).

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

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

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

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

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

$$\hat{d}_{t+1} = \hat{d}_t + \hat{g}_{t+1} + (\tilde{d}_t - \hat{d}_t),$$ \hspace{1cm} (6)

which are useful for defining value functions below. Agents at $t+1$ revise up their expectations to the extent that $\tilde{d}_t$, the level of demand that they infer for the previous period, exceeds $\hat{d}_t$, the level that they expected. According to (4), $\tilde{d}_t$ is larger when $P_t$ and $\pi_t$ are higher. Therefore, the expected growth rate rises with past price growth, as in Glaeser and Nathanson (2017), and also with growth in the speed at which listings are selling.

### 7.3 Prices

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

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

The mover value function satisfies the recursion

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

where the expectation is over $d_t \sim N(\hat{d}_t, \tilde{\sigma}^2)$. If the random potential buyer who matches to the mover buys, the mover receives $P$ and exits the market. The first term, $\tilde{\pi}(P,d_t)P$, reflects the perceived probability of this event times the payoff. If such a sale does not occur, the mover continues to the next period, the value of which is $V^m(\hat{d}_{t+1}, \hat{g}_{t+1})$. The probability of continuing is $1 - \tilde{\pi}(P,d_t)$, and movers discount the next period at the mover discount rate $r_m$. The expectation operator in (7) reflects uncertainty over the current demand shifter, $d_t$, which pins down the perceived demand curve $\tilde{\pi}(P,d_t)$ as well as the future state variables $\hat{d}_{t+1}$ and $\hat{g}_{t+1}$ via (5) and (6) (because movers believe that $\tilde{d}_t = d_t$).

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. 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 C.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 et al. (2015, 2018) and Liao and Peng (2018) assume (see Appendix C.3). In our quantitative exercise, we use a finite $r_m$ and measure the extent to which price growth expectations depend on recent price growth.

### 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

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

(8)

where $E_{i,t}$ denotes the potential buyer’s expectation conditional on her information set. The outer sum runs over possible numbers of periods, $j$, from the time of purchase until becoming a mover. The probability of a spell of length $j$ depends on the potential buyer’s mover hazard, $\lambda_j$. For such a spell, the potential buyer receives flow utility $e^{d_i}$ for the next $j$ periods (corresponding to the inner sum) and the value of being a mover in $j$ periods, which depends on the state variables at that time, $\hat{d}_{t+j}$ and $\hat{g}_{t+j}$.

---

28 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 $P \to \infty$.

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

25
To forecast these state variables, potential buyer $i$ first 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. She uses the price of the specific house to which she matches, $P_{i,t}$, because the only public information on prices is the history of price changes. She calculates $\hat{g}_t$ using (4) and Lemma 1, which together express $\hat{g}_t$ in terms of the history of price growth and sales shares. By (5) and (6), future values of $\hat{d}$ and $\hat{g}$ are linear combinations of $\hat{d}_t$, $\hat{g}_t$, and the innovations $\tilde{d}_{t+j} - \hat{d}_{t+j}$ for $j \geq 0$. From her perspective, these innovations are distributed independently as

$$\tilde{d}_{t+j} - \hat{d}_{t+j} \sim \begin{cases} 
N\left(\frac{\sigma^2(d_i - \mu_n - \hat{d}_t)}{\sigma^2 + \sigma_n^2}, \frac{\sigma^2 \sigma_n^2}{\sigma^2 + \sigma_n^2}\right) & \text{if } j = 0 \\
N(0, \tilde{\sigma}^2) & \text{if } j > 0.
\end{cases}$$

(9)

Her flow utility, $d_i$, gives her non-public information about $d_t$, so her posterior at time $t$ differs from the public posterior $N(0, \tilde{\sigma}^2)$. At $t + 1$, agents believe that they observe $d_t$ via $\hat{d}_t$. Therefore, $d_t$ cannot improve the Kalman filter from time $t + 1$ onward, and the potential buyer’s posterior coincides with the public one for $j > 0$.

It is noteworthy that the potential buyer’s expectation of $\hat{d}_{t+j}$ and $\hat{g}_{t+j}$ is linear in her own flow utility, $d_i$, as well as $\hat{d}_t$ and $\hat{g}_t$. By (4) and Lemma 1, the current state variables themselves are linear in the history of log prices, $\log P_{t'}$ for $t' < t$, and a simple transformation of past sale probabilities, $F^{-1}(1 - \pi_{t'})$ for $t' < t$. This linearity is noteworthy because there is no linear or log-linear relationship between the demand shifter, $d_i$, and prices or sale shares. Linearity results from the cap rate error, which leads agents to impute demand in a simple fashion from $P_t$ and $\pi_t$, as well as the normal distributions of $\epsilon^d_t$, $\epsilon^g_t$, and $a_i$, which imply linear updating of posteriors from priors.

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

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

**Proof.** Appendix C.4.

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 $\kappa_n(\cdot)$, which are no longer constant and instead depend on the potential buyers’ expected horizon $\lambda_i$, occupancy type $n_i$, and demand growth expectations $\hat{g}_t$.

While it is difficult to fully characterize the properties of the $\kappa_n(\cdot)$ 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 $\kappa_n(\cdot)$ decreases in $\hat{g}_t$, with steeper slopes for larger values of $\lambda$. 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 $\lambda$ 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, $\kappa_n$ limits to a constant value of $r$. Second, $\kappa^\lambda_n(\cdot)$ is nearly identical to $\kappa^1_n(\cdot)$ for each $\lambda$. The cutoffs depend very little on occupancy type because one’s flow utility $d_i$ is not very informative about market demand $d_t$ (as $\sigma_a$ is much larger than $\tilde{\sigma}$). Finally, $\kappa^\lambda_n(\cdot)$ is typically larger for greater values of $\lambda$, 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 implies that a purchase occurs when

$$a_i \geq \log p(\hat{g}_t) + \log \kappa_n^\lambda_i(\hat{g}_t) + \hat{d}_t - d_t.$$ 

A potential buyer is marginal when this expression is an equality. The idiosyncratic flow utility $a_i$ of the marginal buyer may differ across $\lambda_i$ and $n_i$, and this dependence itself may depend on the expected growth rate, $\hat{g}_t$. Furthermore, the marginal $a_i$ is lower for all potential buyer types when $\hat{d}_t - d_t$ is lower, that is, when agents underestimate the current state of demand. In this case, movers post list prices that are low relative to actual demand, and more potential buyers decide to buy. Because $\hat{d}_t - d_t$ is unknown at time $t$, it is impossible to identify the quantiles of the marginal buyers in the $F(\cdot)$ distribution until $t + 1$, when $\pi_t$ becomes known.

Due to this inequality, sales to buyers of type $\lambda$ and $n$ equal $L_t \beta_n^\lambda(1 - \Phi(\log p(\hat{g}_t) + \log \kappa_n^\lambda(\hat{g}_t) + \hat{d}_t - d_t - \mu_n))$, where $L_t$ is total listings and $\Phi$ is the CDF of $\mathcal{N}(0, \sigma_a^2)$. Summing
this expression over $\lambda$ and $n$ and then dividing by $L_t$ gives the share of listings that sell:

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

To calculate the share of sales to each buyer type ($b^\lambda_{n,t}$), we divide sales to each type by $\pi_t L_t$ (total sales):

$$b^\lambda_{n,t} = \pi_t^{-1} \beta^\lambda_n \left( 1 - \Phi \left( \log p(\hat{g}_t) + \log \kappa^\lambda_n(\hat{g}_t) + \hat{d}_t - d_t - \mu_n \right) \right).$$ (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 $\Phi'/\Phi$ is an increasing function. Because $\kappa^\lambda_n(\cdot)$ quantitatively does not depend on $n$, the argument of $\Phi(\cdot)$ is always larger for non-occupants than occupants of the same $\lambda$ 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, $\kappa^\lambda_n(\cdot)$ decreases more sharply for larger values of $\lambda$. Second, $\kappa^\lambda_n$ is greater for larger values of $\lambda$, meaning that an equal decrease in log $\kappa^\lambda_n(\hat{g}_t)$ 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.

To clarify the mechanism that generates positive feedback within the model, we substitute (10), the equation for the share of listings that sell, into (4), the equation describing how agents infer demand using the cap rate error. We also substitute the equation that Lemma 2 gives for prices into (4). After using the definition of $F(\cdot)$ and rearranging terms, we obtain the following equation relating the inferred level of demand $\tilde{d}_t$ to other state variables:

$$\sum_{\lambda, n} \beta^\lambda_n \Phi \left( \log p(\hat{g}_t) + \log \kappa + \hat{d}_t - \tilde{d}_t - \mu_n \right) = \sum_{\lambda, n} \beta^\lambda_n \Phi \left( \log p(\hat{g}_t) + \log \kappa^\lambda_n(\hat{g}_t) + \hat{d}_t - d_t - \mu_n \right).$$

This equation shows how $\hat{g}_t$ drives misestimation of demand. When $\hat{g}_t$ is large, $\kappa$ exceeds $\kappa^\lambda_n(\hat{g}_t)$ and agents overestimate the level of demand at $t$, i.e., $\tilde{d}_t > d_t$. Similarly, when $\hat{g}_t$ is small, $\kappa^\lambda_n(\hat{g}_t)$ exceeds $\kappa$ and agents underestimate demand, i.e., $\tilde{d}_t < d_t$. These errors create positive feedback because $\tilde{g}_{t+1}$ is increasing in $\tilde{d}_t$. Therefore, a large $\hat{g}_t$ predictably causes
agents to be surprised about the level of demand at \( t \), which itself causes an upward revision of \( \hat{g}_{t+1} \). Separately, a large \( \hat{g}_t \) also raises \( \hat{d}_{t+1} \). Both of these effects contribute to raise house prices at \( t + 1 \).

Two aspects of this feedback mechanism are particularly useful for understanding our later results. First, it is strongest when the agent’s cutoff rule is more sensitive to expected growth, that is, when \( \kappa(\cdot) \) falls sharply with \( \hat{g}_t \). This relation is strongest for short-term (high \( \lambda \)) potential buyers, so our model generates the largest house price cycles when the highest-\( \lambda \) type is marginal. Second, if \( \hat{g}_t \) is sufficiently positive and true demand is much lower than expected demand \( (d_t \ll \hat{d}_t) \), then agents overestimate demand \( (\check{d}_t > d_t) \) at the same time that they are revising their expectations downward \( (\check{d}_t < \hat{d}_t) \). This situation describes the quiet. Movers expect continued demand growth and post high prices, but few of these listings sell because they overestimate demand. Price growth nevertheless slows down because agents are revising their expectations downward.

### 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, \( \pi_t \), and the composition of buyers, \( b_{\lambda n, t} \):

\[
I_t = (1 - \pi_{t-1}) I_{t-1} + L_t,
\]

\[
V_t = \pi_t I_t,
\]

\[
L_t = \sum_{\lambda} \lambda S_{\lambda t-1}^\lambda,
\]

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

where \( S_{\lambda t}^\lambda \) measures end-of-period stayers of type \( \lambda \). Volume to buyers of occupancy type \( n \) equals \( \sum_{\lambda} b_{n, t}^\lambda V_t \). To track realized short-term sales, as observed in the data, 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 \( p(\hat{\gamma}_t) \) and \( \kappa_\lambda^\gamma(\hat{\gamma}_t) \). To do so, we discretize \( \hat{\gamma} \) 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 \( \epsilon_{d,t} \) and \( \epsilon_{g,t} \). 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 model’s baseline properties. To analyze the impulse response to a shock, we draw a treatment sample of 1,000 additional simulations identical to the control except in periods 101–104 during which the growth rate shocks \( \epsilon_{g,t} \) are two standard deviations higher.\(^{30}\) Impulse responses are average differences between 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 4. The composition of buyers and the volatility of demand growth determine \( \beta_n^\lambda \) and \( \sigma \), 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

\(^{30}\)We shock \( \epsilon^g \) instead of \( \epsilon^d \) so that in the rational benchmark, prices never overshoot. A sequence of 4 shocks matches the experiment in Barberis et al. (2018). We choose 2 standard deviations to explore a large but plausible increase in demand.
the non-occupant volume boom exceeds the occupant boom. Intuitively, these moments determine $\gamma$, $\sigma_a$, and $\mu_1$ through quantifying extrapolation, the elasticity of demand, and the excess sensitivity of non-occupants.

### 8.2 Parameter Estimates and Buyer Cutoff Rules

Table 5 reports parameter values that match the moments in Panels B and C of Table 4. Non-occupant flow utility is 0.9% less than occupant flow utility on average, corresponding to 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 et al. (2019), we focus on the 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 prior 4 periods. The values of 0.127 and 0.042 are somewhat smaller than corresponding values of 0.226 and 0.047 that Armona et al. (2019) find through a survey (see their Table 5). Therefore, to match the key housing cycle moments in Panel C of Table 4, our model requires a smaller amount of extrapolation than these authors found.

Figure 9 plots the potential buyer cutoff functions $\kappa_n^\lambda(\hat{g}_t)$ given our chosen parameters for 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 stand out: (1) each $\kappa_n^\lambda(\cdot)$ decreases, with steeper slopes for larger values of $\lambda$, (2) $\kappa_0^\lambda(\cdot)$ and $\kappa_1^\lambda(\cdot)$ are nearly identical for each $\lambda$, and (3) $\kappa_n^\lambda(\cdot)$ are generally larger for greater values of $\lambda$. These results imply that the sensitivity of buyer demand to the expected growth rate is 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 5, the similarity in cutoff rules between occupants and non-occupants implies that non-occupants are also more likely to be marginal entrants when expected capital gains are high.
8.3 Impulse Responses

Figure 10 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).\footnote{The price boom in our model is smaller than the national boom in Figure 1. Potentially, the shocks that generated the national boom are stronger than the one year of two-standard-deviation shocks we feed into our model. Another possibility is that our assumed value of 0.023 for the annual volatility of demand growth (see Table 4) is too low. Finally, new construction and credit, which our model omits, may have amplified the national boom (Nathanson and Zwick, 2018; Favilukis et al., 2017). To ease comparison with the national cycle, we analyze outcomes in our model relative to the price boom it generates.} 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 $\lambda$ (Panel C) and non-occupants (Panel D). The overall increase in housing demand pushes up the share of listings that sell, $\pi_t$ (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 $d_t$, though much of it comes from $\hat{g}_t$, the expected capital gains channel. Thus, agents over-estimate the demand level, and $\hat{d}_t - d_t$ becomes increasingly positive. As (10) shows, sale 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, \( \pi_t \), remains high even though it is falling, and these high realizations constitute positive surprises about demand that cause movers to increase their beliefs. Eventually, \( \pi_t \) 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% above their pre-shock level. The two causes of the excess inventories are the fall in selling probabilities (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).\(^{32}\)

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

\(^{32}\)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 also hit the peak of listings (relative to price growth) that appears in Figure 3.
to the role that speculation plays in generating those patterns. To quantify the contribution of speculation to the housing cycle, we rerun the simulation under three counterfactuals, each of which shuts down a different aspect of our baseline model.

8.4.1 Rational expectations

In the fully rational counterfactual, movers are no longer subject to the cap rate error and therefore have rational instead of extrapolative expectations about future price growth. Because they correctly understand the problem that potential buyers are solving, movers recognize that the true demand curve is:

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

Using this demand curve, movers at $t$ correctly infer the history of the demand shifter, $d_{t'}$ for $t' < t$, from past price levels and sale shares. The mover value function under rational expectations is:

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

where the expectation is over $d_t \sim \mathcal{N}(\hat{d}_t, \tilde{\sigma}^2)$. By an argument analogous to the proofs of Lemmas 2 and 3, the optimal price takes the form $e^{d_t} p(\hat{g}_t)$, and potential buyer $i$ buys when $e^{d_i} \geq \kappa_n^\lambda(\hat{g}_t) P_{i,t}$, although $p(\cdot)$ and $\kappa_n^\lambda(\cdot)$ may differ from the corresponding functions in the baseline model.

In Panels A through D of Figure 11, we solve for these functions and then compute impulse responses using the same parameters and sequence of shocks in the baseline model.33 When expectations are rational, 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. The short- and long-horizon volume booms are nearly identical in size. In contrast, non-occupant volume continues to rise much more than occupant volume.34 Therefore, even when potential buyers have rational expectations, non-occupants react more strongly to the demand shock underlying the impulse response, but this reaction does not

---

33 Figure IA5 reports impulse responses of sale probabilities and new listings for all counterfactuals.

34 As discussed in Section 7, this result reflects the fact that non-occupant demand is more elastic to the level of the demand shifter, $d_t$, because $\mu_1 > 0$ and the normal distribution has the monotone likelihood property.
generate any positive feedback.

In summary, the price bust and the rise in listings above their initial value—two salient features of the data in Figure 3—depend on departing from rational expectations. These features appear in the baseline model but not the rational version. Quantitatively, a large volume boom, and one that is disproportionately short-term, likewise depend on departing from rationality. An excess non-occupant volume boom does not.

8.4.2 Walrasian market clearing

In the Walrasian version of our model, a mechanism selects a price $P_t$ each period so that the number of potential buyers willing to buy at that price equals the number of movers willing to sell. Appendix D solves this model and describes technical changes to the model setup and parameters that aid comparison to the baseline.

We find that the equilibrium price is $P_t = e^{dt} p(\hat{g}_t)$, where $p(\cdot)$ is a function. The true level of demand, $d_t$, directly affects prices instead of agents’ estimation of this level, $\hat{d}_t$, as in the baseline model. Here, demand from buyers directly pins down the price; in the baseline model, movers choose the price and demand pins down the share of listings that sell. As a result, prices incorporate changes to demand more quickly with Walrasian market clearing. In the Walrasian model, agents believe that the equilibrium house price is $P_t = e^{dt} \hat{p}$, where $\hat{p}$ is a constant. Therefore, when $\hat{g}_t$ is high, equilibrium prices exceed what agents expect, which leads them to think mistakenly that $d_t$ is high. This force in turn pushes up $\hat{g}_{t+1}$, which increases $P_{t+1}$. This positive feedback mechanism is similar to the one in the baseline model.

Results appear in Panels E through H of Figure 11. Prices and volume both go through a large 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 nine 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. Listings rise in the Walrasian model, but listings and volume coincide due to Walrasian market clearing, so these two variables never diverge as in the baseline model. Finally, short-term and non-occupant volume continue to rise in a large and disproportionate fashion in the Walrasian model.

In summary, most of the features of the baseline impulse response do not require departing
from Walrasian market clearing, as they continue to appear in the Walrasian extension. These features include large price and volume cycles, high levels of listings while prices fall, and disproportionate volume booms from short-term sales and non-occupant purchases. However, the existence of the quiet—a period right after the boom in which volume falls while prices and listings rise—does require departing from Walrasian market clearing.

8.4.3 Absence of speculative buyers

The last counterfactual shuts down speculation by adjusting the distribution of potential buyer types while leaving the framework of the model unchanged. In particular, we set $\beta^\lambda_n = 0$ for all $\lambda$ and $n$ except for $\lambda = 0.03$ and $n = 1$, meaning that all potential buyers are occupants with a horizon of about 8 years, which is close to the average horizon among potential buyers in the baseline. By assigning all potential buyers the same (low) value of $\lambda$, this counterfactual removes both short-term buyers and the heterogeneity in holding periods that generates variation in the composition of buyers. We update $\kappa$ so that the demand error is still zero and keep other parameters unchanged.

Panels I through L display the results. Prices and volume still go through a cycle, but the volume boom is 3 times smaller, and the price overshoot almost disappears. Listings fall 7%, much more than the baseline decline of 1.5%. There is a quiet during which listings rise, but they reach a smaller value of 0.4% (versus the baseline of 1.4%) at the end of this period. Short-term volume rises more than long-term volume, but by a factor of only 1.3 relative to the baseline of 7.8. Finally, non-occupant volume equals zero by assumption.

In summary, removing speculation dramatically reduces the magnitude of the cycle relative to the baseline model. Speculative buyers in the baseline also limit the decline in listings during the boom relative to the no-speculation counterfactual. These results suggest that the 2000s housing cycle would have been far less severe in the absence of speculation. There would have been essentially no price bust, the volume cycle would have been much smaller, and inventories would have fallen substantially as demand grew.

---

35 The disproportionate rise in short-term volume in this setting, where all potential buyers have the same $\lambda$, is due to a mechanical channel that we discuss and estimate in Appendix B.1.
8.5 Comparing Short-Term and Non-Occupant Buyers

The counterfactuals up to this point have explored the effect of shutting down speculation entirely, either through rational expectations or the distribution of potential buyers. In this section, we explore the distinct roles of short-term and non-occupant potential buyers in amplifying the cycle. 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 $\lambda$ except for $\lambda = 0.03$. Unlike the previous counterfactual, we keep a positive mass of non-occupant buyers, and we do so in two ways. In the first, the share of non-occupants among potential buyers with $\lambda = 0.03$ equals its baseline. In the second, 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 $\lambda$ 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, $\beta_0^\lambda$, to zero, and then scales up the occupant shares, $\beta_1^\lambda$, so that they sum to one. This method skews the $\lambda$ distribution toward long-term buyers because occupants have longer horizons than non-occupants (Table 5). Therefore, we explore a second counterfactual in which we maintain the original $\lambda$ distribution while eliminating non-occupants. We continue to set each $\beta_0^\lambda$ to zero, but now we update $\beta_1^\lambda$ to the baseline $\lambda$ share among all potential buyers.

Table 6 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 boom falls to 8.7% from its baseline of 14.5%, meaning that short-term buyers amplify the price boom by 67%. Furthermore, in the counterfactuals, 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. Therefore, eliminating short-term buyers prevents the model from matching key aggregate facts (Figures 1 and 3).

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.

---

36 These counterfactuals do a better job matching inventory levels during the bust, which reach 1.6% above the initial level, a higher peak than the baseline. In the baseline, new listings fall sharply during the bust because short-term buyers exit the market (Panel F of Figure 10). Thus, the baseline does a better job matching listing behavior in the boom and quiet than in the bust.

37 The occupant adjustment does not affect the cycle because agents in the model correctly understand the distribution of flow utility, meaning that changing the flow utility distribution does not destabilize prices.
However, when we adjust the $\lambda$ 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.

One concern is that the occupant premium, $\mu_1$, is about 7 times smaller than the standard deviation of flow utility, $\sigma_a$. Therefore, non-occupants may play a small role in amplifying the cycle solely because of parameter values in which non-occupants closely resemble occupants. To investigate this possibility, Table IA11 regenerates the first, third, and fifth columns of Table 6 under the larger values of $\mu_1 = 0.033$ and $\mu_1 = 0.066$, corresponding to 50% and 100% of the baseline value of $\sigma_a$. We continue to find significant attenuation of the cycle with all long-term buyers if we adjust for the occupant distribution, but not with all occupant buyers if we adjust for the $\lambda$ distribution.

These results speak to the finding in Table 2 that a short-volume boom more robustly predicts price booms and busts than does a non-occupant boom. Our findings are consistent with Gao et al. (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-occupant buyers on subsequent price growth. This finding is consistent with our results if 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 prices and 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.

8.6 Transaction Taxes

Several governments have introduced taxes on real estate transactions in an attempt to curb speculative activity (see Chi et al. (2021) for a list). In this section, we use our model to study an ad valorem tax that buyers must pay at the time of purchase. To study taxes that target non-occupants, we allow the tax rate to depend on the buyer’s occupancy type $n$, so that buyer $i$ pays a tax $\tau_n P_{i,t}$ when purchasing a home. Analyzing capital gains taxes would
complicate our model significantly, so we leave that to future work.\footnote{With a capital gains tax, movers who bought their house at different past prices face different optimality problems when selecting a list price. In general, movers will post different list prices at the same time, meaning that we must keep track of the time-varying distribution of list prices. In the current model (and the extensions with buyer taxes), all movers post the same list price.}

Holding prices constant, the share of potential buyers who complete a purchase is lower in the presence of this new tax. As a result, $\kappa$ must go up, as we select this constant so that the average value of $d_t - \hat{d}_t$ equals zero. Intuitively, because the cap rate error on average leads to the correct estimation of demand, the threshold $\kappa$ must rise to reflect the decrease in housing demand from the new tax. We denote this new value $\kappa(\tau)$, where $\tau = (\tau_0, \tau_1)$.

By analyzing the mover value function, it is straightforward to show that the new optimal price is $P_t = e^{\hat{d}_t} p(\hat{g}_t) \kappa(0)/\kappa(\tau)$, where $p(\cdot)$ is the same function that is in Lemma 2. That is, prices scale down by a constant amount that reflects the reduced demand due to the tax.

The reduction in housing demand operates through the cutoff functions, $\kappa^\lambda(\cdot)$. Due to the proportional nature of the tax, Lemma 3 continues to hold, but now these cutoff functions depend on the tax. We denote them $\kappa^\lambda_n(\hat{g}_t, \tau)$. Potential buyer $i$ decides to buy if

$$a_i \geq \log p(\hat{g}_t) + \log \left(\frac{\kappa(0)\kappa^\lambda_n(\hat{g}_t, \tau)}{\kappa(\tau)}\right) + \hat{d}_t - d_t.$$ 

We explore a tax that binds equally on all buyers, so that $\tau_0 = \tau_1$, and a tax that affects only non-occupant buyers, so that $\tau_1 = 0$. We consider taxes of 0.5%, 1%, and 5%, which span the tax rates in many large cities (Chi et al., 2021).

Table 7 reports the same outcomes in Table 6 for each tax regime. A 5% tax on all buyers significantly attenuates the price cycle, reducing the bust from 8.2% to 1.1%. It also reduces the volume boom, but this reduction is smaller than the corresponding one for prices. Smaller taxes of 0.5% and 1% also reduce the cycle amplitude, but these effects are much smaller.

The last three columns of Table 7 report results for the tax on non-occupant buyers. This tax is a weaker instrument for attenuating the house price cycle: the 5% tax reduces the price bust only to 5.8%, and the lower taxes have a smaller effect. The 5% tax nearly eliminates the non-occupant volume boom, reducing it to 0.1% from 12.3%. Therefore, targeting the tax to non-occupants limits its efficacy in reducing the house price cycle, as even a tax that nearly eliminates the non-occupant volume boom still leaves much of the house price cycle.

To understand the mechanism behind these results, Figure IA6 plots the adjusted buying
cutoffs, \( \kappa(0) \kappa^\lambda_n(\hat{g}, \tau)/\kappa(\tau) \), for both 5% tax scenarios. Comparing this figure to Figure 9 shows how each tax changes housing demand. The 5% tax on all buyers raises the cutoffs for the \( \lambda = 0.5 \) group by about half a standard deviation (\( \sigma_a \)), which makes the \( \lambda = 0.17 \) group more marginal than before. Therefore, the tax effectively skews the composition of buyers towards those with longer horizons. The tax on non-occupants similarly raises the cutoffs, but only for non-occupants. As a result, both the \( \lambda = 0.5 \) occupants and the \( \lambda = 0.17 \) non-occupants are marginal. Therefore, many of the buyers with the shortest horizons are still active in the market, explaining why this tax has a weaker effect.

9 Conclusion

Our paper raises additional lines of inquiry within the housing market. We have argued with theory and evidence that speculators in general and short-term speculators in particular 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. While we analyze a fixed transactions tax in this paper, in the spirit of Tobin (1978), Stiglitz (1989), Summers and Summers (1989), and Dávila (2015), natural alternatives such as a short-term capital gains tax might discourage housing speculation by lowering expected after-tax capital gains. However, such taxes discourage productive residential investment as well. Is this tax optimal, and if not, what type of tax policy would be better? It is also unclear empirically whether transaction and capital gains 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. Adding construction to the model in this paper might further clarify the role of land markets and new construction in housing cycles.

References


FIGURE 1
The Dynamics of Prices and Volume
Panel A. National

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.
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 $\beta_\tau \frac{\text{std}(v_{i,t-\tau})}{\text{std}(p_{i,t})}$, where $v_{i,t-\tau}$ and $p_{i,t}$ are the demeaned regressors.
FIGURE 3
The Dynamics of Prices and Inventories
Panel A. National

Panel B. Phoenix, AZ

Panel C. Reno, NV

Panel D. Daytona Beach, FL

Panel E. Bakersfield, CA

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.
FIGURE 4
Normalized Aggregate Volume by Transaction Type

<table>
<thead>
<tr>
<th>Volume (000s)</th>
<th>2000</th>
<th>2005</th>
<th>2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short(_{S1})</td>
<td>510</td>
<td>940</td>
<td>150</td>
</tr>
<tr>
<td>Existing(_{S1})</td>
<td>2,130</td>
<td>2,880</td>
<td>930</td>
</tr>
<tr>
<td>Total(_{S1})</td>
<td>2,730</td>
<td>3,820</td>
<td>1,150</td>
</tr>
<tr>
<td>Non-Occupant(_{S2})</td>
<td>510</td>
<td>1,030</td>
<td>290</td>
</tr>
<tr>
<td>Total(_{S2})</td>
<td>2,310</td>
<td>3,290</td>
<td>990</td>
</tr>
</tbody>
</table>

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, \(S_1\) refers to the short-holding-period sample of 115 MSAs and \(S_2\) 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

Panel B. Total Volume Versus Volume by Occupancy Status

Panel C. Role of Short Volume in Total Volume Growth

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

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

<table>
<thead>
<tr>
<th>Listings (000s)</th>
<th>2003</th>
<th>2007</th>
<th>2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short</td>
<td>280</td>
<td>590</td>
<td>170</td>
</tr>
<tr>
<td>Total</td>
<td>1,170</td>
<td>1,730</td>
<td>1,380</td>
</tr>
</tbody>
</table>

**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
Speculative Homebuying and Recent House Price Appreciation

Panel A. Short-Holding-Period Buyers

Panel B. Non-Occupant Volume

Panel C. Expected Short-Term Buyers

Panel D. View of Housing as Investment

Panel E. P(Buying Non-Primary Home)

Notes: Panels A and B use CoreLogic data to show the relation between short-holding-period volume and non-occupant volume at the MSA level, respectively, and the past year’s house price appreciation. Volume measures are scaled relative to their level in 1999. Short-holding-period volume in Panel A is forward-looking, i.e., it is based on whether the buyer sells within three years. Panel C uses data from the NAR Investment and Vacation Home Buyers Survey; “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 use the FHFA index here because it covers the period 2015–2016. Panels D and E use data from the Federal Reserve Survey of Consumer Expectations and Armona et al. (2019) to study the relation between recent house price growth and the probability of buying a non-primary home. In these data, local house price appreciation is computed at the ZIP-level from Zillow.
FIGURE 8
Expected Holding Times of Homebuyers, 2008–2015

Notes: This figure presents evidence on heterogeneity in expected holding times among recent homebuyers from the NAR Investment and Vacation Home Buyers Survey. We plot the response frequency averaged equally over each survey year from 2008 to 2015. We reclassify buyers who have already sold their properties by the time of the survey as having an expected holding time in [0,1).
FIGURE 9
Buying Cutoffs for Different Expected Growth Rates

Notes: The buying cutoff, $\kappa_{\lambda}^n(\hat{g}_t)$, 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, $\lambda$, 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 $\lambda$ 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 $\kappa$, which agents mistakenly believe is the time-invariant buying cutoff for other potential buyers.

"FIGURE 9"
"Buying Cutoffs for Different Expected Growth Rates"

Notes: The buying cutoff, $\kappa_{\lambda}^n(\hat{g}_t)$, 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, $\lambda$, 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 $\lambda$ 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 $\kappa$, which agents mistakenly believe is the time-invariant buying cutoff for other potential buyers.
Notes: Impulse responses are average differences between log outcomes in control simulations and treatment simulations, in which a 2-standard-deviation shock to $\epsilon_g^t$ (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.
FIGURE 11
Impulse Responses in Counterfactuals

Panel A. Prices and Volume, Rational
Panel B. Inventory of Listings, Rational
Panel C. Volume by Holding Period, Rational
Panel D. Volume by Occupancy, Rational

Panel E. Prices and Volume, Walrasian
Panel F. Inventory of Listings, Walrasian
Panel G. Volume by Holding Period, Walrasian
Panel H. Volume by Occupancy, Walrasian

Panel I. Prices and Volume, No Speculation
Panel J. Inventory of Listings, No Speculation
Panel K. Volume by Holding Period, No Speculation
Panel L. Volume by Occupancy, No Speculation

Notes: Impulse responses are average differences between log outcomes in control simulations and treatment simulations, in which a 2-standard-deviation shock to $\epsilon^*_g$ (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>
<thead>
<tr>
<th>Panel A. Short-Volume Sample</th>
<th>Variable</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Short-Volume Boom</td>
<td>15.97</td>
<td>12.93</td>
<td>115</td>
</tr>
<tr>
<td></td>
<td>Price Boom</td>
<td>97.06</td>
<td>47.88</td>
<td>115</td>
</tr>
<tr>
<td></td>
<td>Price Bust</td>
<td>-27.9</td>
<td>13.64</td>
<td>115</td>
</tr>
<tr>
<td></td>
<td>Δ Volume Quiet + Bust</td>
<td>-62.96</td>
<td>18.87</td>
<td>115</td>
</tr>
<tr>
<td></td>
<td>Foreclosures Bust</td>
<td>82.84</td>
<td>55.96</td>
<td>115</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B. Non-Occupant Volume Sample</th>
<th>Variable</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Non-Occupant Volume Boom</td>
<td>29.29</td>
<td>27.05</td>
<td>102</td>
</tr>
<tr>
<td></td>
<td>Short-Volume Boom</td>
<td>16.88</td>
<td>13.36</td>
<td>102</td>
</tr>
<tr>
<td></td>
<td>Price Boom</td>
<td>100.57</td>
<td>49.27</td>
<td>102</td>
</tr>
<tr>
<td></td>
<td>Price Bust</td>
<td>-28.99</td>
<td>13.97</td>
<td>102</td>
</tr>
<tr>
<td></td>
<td>Δ Volume Quiet + Bust</td>
<td>-63.32</td>
<td>19.47</td>
<td>102</td>
</tr>
<tr>
<td></td>
<td>Foreclosures Bust</td>
<td>86.57</td>
<td>58.08</td>
<td>102</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel C. Short-Volume Sample with Listings</th>
<th>Variable</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Short-Volume Boom</td>
<td>14.64</td>
<td>12.33</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td>Δ Listings Boom</td>
<td>91.67</td>
<td>94.93</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td>Δ Listings Quiet</td>
<td>178.39</td>
<td>143.86</td>
<td>57</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel D. Non-Occupant Volume Sample with Listings</th>
<th>Variable</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Non-Occupant Volume Boom</td>
<td>27.81</td>
<td>27.32</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>Short-Volume Boom</td>
<td>15.84</td>
<td>12.88</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>Δ Listings Boom</td>
<td>82.11</td>
<td>93.67</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>Δ Listings Quiet</td>
<td>171.74</td>
<td>151.29</td>
<td>48</td>
</tr>
</tbody>
</table>

Notes: This table reports summary statistics for MSA-level variables in different samples of MSAs. 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 2
Speculative Booms and Housing Market Outcomes

### Panel A. MSA-Level Prices

<table>
<thead>
<tr>
<th></th>
<th>Price Boom</th>
<th>Price Bust</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short-Volume Boom</td>
<td>1.930***</td>
<td>-0.571***</td>
</tr>
<tr>
<td></td>
<td>(0.297)</td>
<td>(0.083)</td>
</tr>
<tr>
<td>Non-Occupant Volume Boom</td>
<td>0.570***</td>
<td>-0.166***</td>
</tr>
<tr>
<td></td>
<td>(0.173)</td>
<td>(0.049)</td>
</tr>
<tr>
<td>Number of Observations</td>
<td>115</td>
<td>102</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.272</td>
<td>0.098</td>
</tr>
</tbody>
</table>

### Panel B. MSA-Level Inventories

<table>
<thead>
<tr>
<th></th>
<th>Δ Listings Boom</th>
<th>Δ Listings Quiet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short-Volume Boom</td>
<td>-1.133</td>
<td>5.961***</td>
</tr>
<tr>
<td></td>
<td>(1.027)</td>
<td>(1.353)</td>
</tr>
<tr>
<td>Non-Occupant Volume Boom</td>
<td>-0.070</td>
<td>2.645***</td>
</tr>
<tr>
<td></td>
<td>(0.505)</td>
<td>(0.718)</td>
</tr>
<tr>
<td>Number of Observations</td>
<td>57</td>
<td>48</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.022</td>
<td>0.000</td>
</tr>
</tbody>
</table>

### Panel C. MSA-Level Volume Quiet and Bust

<table>
<thead>
<tr>
<th></th>
<th>Δ Volume Quiet + Bust</th>
<th>Foreclosures Bust</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short-Volume Boom</td>
<td>-1.047***</td>
<td>0.895**</td>
</tr>
<tr>
<td></td>
<td>(0.096)</td>
<td>(0.398)</td>
</tr>
<tr>
<td>Non-Occupant Volume Boom</td>
<td>-0.512***</td>
<td>-0.060</td>
</tr>
<tr>
<td></td>
<td>(0.051)</td>
<td>(0.215)</td>
</tr>
<tr>
<td>Number of Observations</td>
<td>115</td>
<td>115</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.515</td>
<td>0.043</td>
</tr>
</tbody>
</table>

**Notes:** This table reports estimates of the relation between speculative volume and housing cycle measures at the MSA level. 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. Table 1 presents summary statistics for each sample. Significance levels 10%, 5%, and 1% are denoted by *, **, and *** respectively.
### TABLE 3

All-Cash Buyer Shares

<table>
<thead>
<tr>
<th></th>
<th>Transaction-Level</th>
<th>MSA-Level</th>
<th>MSA-Level</th>
<th>MSA-Level</th>
<th>MSA-Level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All Months</td>
<td>All Months</td>
<td>Boom</td>
<td>Quiet</td>
<td>Bust</td>
</tr>
<tr>
<td>Short Buyers</td>
<td>0.29</td>
<td>0.38</td>
<td>0.29</td>
<td>0.28</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td>(0.21)</td>
<td>(0.16)</td>
<td>(0.17)</td>
<td>(0.20)</td>
<td></td>
</tr>
<tr>
<td>Non-Occupant Buyers</td>
<td>0.38</td>
<td>0.41</td>
<td>0.36</td>
<td>0.32</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td>(0.18)</td>
<td>(0.15)</td>
<td>(0.14)</td>
<td>(0.18)</td>
<td></td>
</tr>
<tr>
<td>All Buyers</td>
<td>0.20</td>
<td>0.25</td>
<td>0.22</td>
<td>0.20</td>
<td>0.30</td>
</tr>
<tr>
<td></td>
<td>(0.16)</td>
<td>(0.15)</td>
<td>(0.14)</td>
<td>(0.16)</td>
<td></td>
</tr>
</tbody>
</table>

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 1.1. 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 4
Inputs into model calibration

<table>
<thead>
<tr>
<th>Parameter or target</th>
<th>Value</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: Assumed parameters</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( r ) (non-mover discount rate)</td>
<td>0.012</td>
<td>Guren (2018)</td>
</tr>
<tr>
<td>Potential ( \lambda ) values</td>
<td>{0.50, 0.17, 0.05, 0.03, 0.01}</td>
<td>Figure 8</td>
</tr>
<tr>
<td>( \rho ) (demand growth persistence)</td>
<td>0.880</td>
<td>GN (2017)</td>
</tr>
<tr>
<td><strong>Panel B: Steady-state targets</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Occupant buyer shares</td>
<td>(0.06, 0.07, 0.16, 0.16, 0.34)</td>
<td>Figure 8</td>
</tr>
<tr>
<td>Non-occupant buyer shares</td>
<td>(0.04, 0.03, 0.04, 0.04, 0.06)</td>
<td>Figure 8</td>
</tr>
<tr>
<td>Annual volatility of demand growth</td>
<td>0.023</td>
<td>GN (2017)</td>
</tr>
<tr>
<td>Quarterly selling hazard</td>
<td>0.75</td>
<td>Guren (2018)</td>
</tr>
<tr>
<td>Mean demand error</td>
<td>0</td>
<td>Model</td>
</tr>
<tr>
<td>Mean demand growth</td>
<td>0</td>
<td>Model</td>
</tr>
<tr>
<td><strong>Panel C: Cycle targets</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Price overshoot</td>
<td>2.3</td>
<td>Figure 1</td>
</tr>
<tr>
<td>Volume boom/price boom</td>
<td>0.4</td>
<td>Figure 1</td>
</tr>
<tr>
<td>Non-occupant boom/occupant boom</td>
<td>3.1</td>
<td>Figure 4</td>
</tr>
</tbody>
</table>

**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 5
Outputs from model calibration

<table>
<thead>
<tr>
<th>Parameter or outcome</th>
<th>Interpretation</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: Derived parameters</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_a$</td>
<td>Flow utility dispersion</td>
<td>0.066</td>
</tr>
<tr>
<td>$\mu_1$</td>
<td>Occupant premium</td>
<td>0.009</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>$g$ variance share</td>
<td>0.070</td>
</tr>
<tr>
<td>$\kappa$</td>
<td>Cap rate error</td>
<td>0.029</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>Demand volatility</td>
<td>0.011</td>
</tr>
<tr>
<td>$\mu$</td>
<td>Mean demand growth</td>
<td>$-0.000$</td>
</tr>
<tr>
<td>$r_m$</td>
<td>Mover discount rate</td>
<td>0.141</td>
</tr>
<tr>
<td>$\beta_0^\lambda$</td>
<td>Non-occupant shares</td>
<td>(0.143, 0.022, 0.030, 0.030, 0.045)</td>
</tr>
<tr>
<td>$\beta_1^\lambda$</td>
<td>Occupant shares</td>
<td>(0.185, 0.052, 0.119, 0.119, 0.254)</td>
</tr>
<tr>
<td><strong>Panel B: Steady-state outcomes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-year extrapolation</td>
<td>–</td>
<td>0.127</td>
</tr>
<tr>
<td>2-5-year extrapolation</td>
<td>–</td>
<td>0.042</td>
</tr>
</tbody>
</table>

*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 4. 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>
<thead>
<tr>
<th>Outcome</th>
<th>All long-term buyers</th>
<th>All occupants</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Baseline</td>
<td>No occupant adjustment</td>
</tr>
<tr>
<td>Price boom</td>
<td>14.5</td>
<td>8.7</td>
</tr>
<tr>
<td>Price bust</td>
<td>−8.2</td>
<td>−0.4</td>
</tr>
<tr>
<td>Volume boom</td>
<td>5.8</td>
<td>2.9</td>
</tr>
<tr>
<td>Listings, end of boom</td>
<td>−1.3</td>
<td>−3.1</td>
</tr>
<tr>
<td>Listings, end of quiet</td>
<td>1.4</td>
<td>0.4</td>
</tr>
<tr>
<td>Short volume boom</td>
<td>14.1</td>
<td>3.4</td>
</tr>
<tr>
<td>Non-occupant volume boom</td>
<td>12.3</td>
<td>3.6</td>
</tr>
<tr>
<td>Sale probability boom</td>
<td>7.1</td>
<td>6.0</td>
</tr>
</tbody>
</table>

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 counterfactuals involve different values of the underlying distribution of potential buyers, $\beta_n$, that the text describes. We alter $\kappa$ in each counterfactual to maintain a zero demand error while keeping other parameters the same. The baseline values correspond to Figure 10.
## TABLE 7
Outcomes for different tax regimes

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Baseline</th>
<th>Tax on all buyers</th>
<th></th>
<th>Tax on non-occupant buyers</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0.5%</td>
<td>1%</td>
<td>5%</td>
<td></td>
</tr>
<tr>
<td>Price boom</td>
<td>14.5</td>
<td>13.1</td>
<td>12.2</td>
<td>9.7</td>
<td>13.5</td>
</tr>
<tr>
<td>Price bust</td>
<td>-8.2</td>
<td>-6.4</td>
<td>-5.1</td>
<td>-1.1</td>
<td>-7.0</td>
</tr>
<tr>
<td>Volume boom</td>
<td>5.8</td>
<td>5.5</td>
<td>5.2</td>
<td>4.1</td>
<td>5.3</td>
</tr>
<tr>
<td>Listings, end of boom</td>
<td>-1.3</td>
<td>-1.2</td>
<td>-1.1</td>
<td>-1.0</td>
<td>-0.8</td>
</tr>
<tr>
<td>Listings, end of quiet</td>
<td>1.4</td>
<td>1.3</td>
<td>1.2</td>
<td>0.6</td>
<td>1.2</td>
</tr>
<tr>
<td>Short volume boom</td>
<td>14.1</td>
<td>13.7</td>
<td>13.2</td>
<td>10.3</td>
<td>13.5</td>
</tr>
<tr>
<td>Non-occupant volume boom</td>
<td>12.3</td>
<td>11.7</td>
<td>11.1</td>
<td>8.9</td>
<td>7.5</td>
</tr>
<tr>
<td>Sale probability boom</td>
<td>7.1</td>
<td>6.7</td>
<td>6.3</td>
<td>5.1</td>
<td>6.2</td>
</tr>
</tbody>
</table>

**Notes:** We report 100 times changes in log outcomes between treatment and control simulations. See notes to Table 6 for outcome definitions. The tax is relative to the purchase price, payable at time of sale. We alter $\kappa$ in each column to maintain a zero demand error while keeping other parameters the same. The baseline values correspond to Figure 10.
Internet Appendix

A 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.\footnote{We rely on CoreLogic’s internal transaction-type categorization to determine whether a transaction occurred at arms length.} 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 IA1 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.
(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.\(^2\) We define a vacant land sale to be any transaction where the sale occurs a year or more before the property was built.

Table IA2 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.\(^3\) The 102 MSAs that remain after these two filters are marked with an “x” in columns 3 and 7 of Table IA1.

---

\(^2\)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.

\(^3\)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 IA1.

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.

B Robustness and Alternative Explanations

B.1 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:

\[
\pi_{t', t} = \alpha_{\text{buy}}^{y(t')} + \alpha_{\text{sell}}^{y(t)} + \alpha_{\text{duration}}^{t-t'} + \epsilon_{t', t},
\]

where \( y(\cdot) \) gives the year of the month. The first set of fixed effects, \( \alpha_{\text{buy}}^{y(t')} \), captures the
average propensity of buyer cohorts from year \( y(t') \) to sell in any future year. The second
set of fixed effects, \( \alpha_{\text{sell}}^{y(t)} \), captures the average propensity of all owners to sell in year \( y(t) \).
The third set of fixed effects, \( \alpha_{\text{duration}}^{t-t'} \), measures time-invariant selling hazard profiles as a
function of time elapsed since purchase \( t - t' \). We interpret year-to-year movements in \( \alpha_{\text{buy}}^{y(t')} \)
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.

Table IA3 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 \( 2000 \leq t' < t \leq 2005 \), we construct the counterfactual selling
hazard as

\[
\pi_{c, t'} = \pi_{t', t} - \left( \alpha_{\text{buy}}^{y(t')} - \hat{\alpha}_{\text{buy}}^{2000} \right),
\]
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 1995 \( m_1 \leq t' < 2005 \),

\[
\begin{align*}
    e_{t',t}^c &= e_{t',t} \\
    v_{t',t}^c &= v_{t',t}.
\end{align*}
\]

We then iteratively update the counterfactuals over \( t \) running from \( t'+1 \) to 2005:

\[
\begin{align*}
    e_{t',t}^c &= e_{t',t}^c - v_{t',t}^c \\
    v_{t',t}^c &= \pi_{t',t}^c e_{t',t}^c.
\end{align*}
\]

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 IA3 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.

### B.2 Endogenous Holding Periods

The empirical evidence presented in Section 3 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. This section addresses this issue further using an instrumental variables strategy.

Our 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 potential homebuyers. 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.

The strength of this instrument is that it is predetermined relative to the realized holding
periods for sellers in the boom and may therefore help purge our estimates of mechanical bias arising from endogenous changes in holding periods over the course of ownership spells. We stress this instrument does not remove the 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 between total and short-term volume.

Table IA4 presents the results. Column 1 presents first stage regressions of the short-term volume boom on the old and young shares. The F-statistic of 39.95 indicates the IV regressions are well powered. Column 2 shows that an 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. 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. We rescale the coefficient using a variance decomposition, which indicates that 33 percent of the variation in total volume growth across MSAs can be explained by changes in short-term volume, thus matching the short-term volume result from Figure 5.

In Table IA4, column 3, the short-term volume coefficient does not change 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. Columns 4 through 7 show that the relations between the price boom and bust and the short-term volume boom strengthen in the IV specifications. This result suggests a modest negative feedback between price growth and holding period, perhaps reflecting a disposition effect force in which price growth induces buyers to sell earlier than they otherwise would. Overall, the IV results present strong evidence that the change in realized short-term volume is quantitatively important for determining overall volume growth and the size of the price cycle, even when using only the portion of short-term volume growth predicted by ex-ante buyer characteristics.

B.3 Repeat Buyers

The patterns we document are consistent with speculative motives leading short-term buyers to enter and exit the 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 in the boom (Stein, 1995; Ortalo-Magné and Rady, 2006), or repeated buying and selling of homes within the same market by experienced “flippers” (Choi et al., 2014; Bayer et al., 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. 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. These 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 engaging in repeat purchases do not account for most of the short-term volume and its growth, even during the cycle’s peak.

B.4 High Frequency Analysis of Price Growth and Speculative Volume (pVAR)

To further investigate the link between house price changes and speculative entry, we examine higher frequency data. Speculative buyers may both cause and respond to house price changes. Because of the potential for this type of feedback mechanism, we do not attempt to directly identify the “causal” effect of speculators on house prices. 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 prices. In particular, we estimate a series of panel vector auto-regressions (pVARs) that relate house price growth to 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 IA5 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

4These 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.

5The 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.

6Gao et al. (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.
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 non-occupant share of purchases. Unlike short-buyer entry, non-occupant entry does not appear to be predictive for 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 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.

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.

\section{Omitted Proofs of Mathematical Statements}

\subsection{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 $\sigma^2_t$ its variance. We solve for these outcomes using standard Kalman filtering. Denote $\sigma^2_\gamma = (1 - \gamma)\sigma^2$ and $\sigma^2_g = \gamma(1 - \rho^2)\sigma^2$.

We have $g_{t-1} = g_t^* + \zeta_t^g$, where $\zeta_t^g \sim \mathcal{N}(0, \sigma^2_t)$. Therefore, $g_t = (1 - \rho)\mu + \rho g_{t-1} + \epsilon_t^g = \cdots$
Because $\hat{r}$, we can express $\hat{r}$ + $\epsilon_t^d$. Therefore, the new posterior variance satisfies $\sigma_t^2 = \sigma_t^2(\rho^2\sigma_t^2 + \sigma_y^2)(\sigma_t^2 + \rho^2\sigma_t^2 + \sigma_y^2)^{-1}$. Solving yields

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

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

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

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

### C.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 $\bar{\pi}(P, d_t) = 1 - F(\log p + \log \kappa - \zeta_t)$, where we denote $\bar{\pi}(p, \zeta_t)$ by abuse of notation. Substituting these expressions into (7) and using (5) and (6) yield

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

with the expectation over $\zeta_t \sim \mathcal{N}(0, \sigma^2)$. 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 $p(\hat{g}_t)$.

### C.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

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

Therefore, $\hat{g}_t$ proxies for expected future price growth, with $\rho$ 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

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

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 \( (1 - \rho)\mu + \rho \hat{d}_t + (1 + \rho(1 - \alpha))(\hat{d}_t - \hat{d}_t) \), and from (4), we have \( \hat{d}_t - \hat{d}_t = \log \kappa P - F^{-1}(1 - \pi_t) \).

Substituting the equation just derived for \( \hat{g}_t \) yields

\[
\log P_{t+1} - \log P_t = \\
\mu + (1 + \rho(1 - \alpha)) (\log \kappa P - F^{-1}(1 - \pi_t)) + \frac{\rho^2(1 - \alpha)}{1 + \rho(1 - \alpha)} \sum_{j=0}^{\infty} \rho^j (\Delta \log P_{t-j} - \mu).
\]

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 \( \mu \), 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 \( \pi_t \), as his rule-of-thumb sellers do not adjust list prices in response to market data other than past prices.

### C.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 \( \lambda \) 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 \( V_\lambda^s(\hat{d}_t, \hat{g}_t) = e^{\hat{d}_t} V_\lambda^s(\hat{g}_t) \). 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

\[
v_\lambda^s(\hat{d}_t, \hat{g}_t) = (1 + r)^{-1} E_t \left( e^{(1-\rho)\mu + \rho \hat{d}_t + (1+\rho(1-\alpha))\zeta_t} \left( (1 - \lambda)v_\lambda^s(\hat{d}_{t+1}, \hat{g}_{t+1}) + \lambda V^m(\hat{g}_{t+1}) \right) \right), \tag{13}
\]

with the expectation over \( \zeta_t \sim N(0, \sigma^2) \). Because \( \hat{d}_t \) enters only \( v^s_\lambda \), 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

\[
V_{i,t}^b = \frac{e^{d_t}}{r + \lambda_i} + \frac{P_{i,t}}{(1 + r)p(\hat{g}_t)} E_{i,t} \left( e^{(1 - \rho)s + \rho \delta + (1 + \rho - \alpha)\zeta} \left( (1 - \lambda_i)v_{\lambda_i}(\hat{g}_t + 1) + \lambda_i v^m(\hat{g}_t + 1) \right) \right),
\]

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

\[
e^{d_t} \geq P_{i,t}(r + \lambda_i) \left( 1 - \frac{\int_{-\infty}^{\infty} \Psi_{\lambda_i} \left( \hat{g}_t, \zeta + \frac{\delta^2(\mu_{1i} - \hat{g}_t)}{\sigma^2 + \sigma^2_a} \right) \phi(\zeta) d\zeta}{(1 + r)p(\hat{g}_t)} \right),
\]

where \( \phi \) is a mean-zero normal pdf with variance \( \sigma^2(\sigma^2_\alpha + \sigma^2_\alpha)^{-1} \). Write \( e^{d_t} = \kappa_i P_{i,t} \). Then the equation becomes

\[
\kappa_i \geq (r + \lambda_i) \left( 1 - \frac{\int_{-\infty}^{\infty} \Psi_{\lambda_i} \left( \hat{g}_t, \zeta + \frac{\delta^2(\mu_{1i} - \hat{g}_t)}{\sigma^2 + \sigma^2_a} \right) \phi(\zeta) d\zeta}{(1 + r)p(\hat{g}_t)} \right).
\]

In Appendix C.5, we prove that \( v^m(\cdot) \) and \( v^*_{s}(\cdot) \) are continuous functions that weakly increase. As a result, the right side of the above inequality continuously and weakly decreases in \( \kappa_i \). The left side continuously and strictly increases in \( \kappa_i \). 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 \( \kappa_{ni}^\lambda(\hat{g}_t) = 0 \). If the right side limits to a positive number as \( \kappa_i \to 0 \), then by the Intermediate Value Theorem, there exists a unique \( \kappa_{ni}^\lambda(\hat{g}_t) > 0 \) such that the inequality holds if and only if \( \kappa_i \geq \kappa_{ni}^\lambda(\hat{g}_t) \), which proves the Lemma.

### C.5 Value Function Monotonicity

This section establishes that the functions \( v^m(\cdot) \) and \( v^*_{s}(\cdot) \), which we define in the proofs of Lemmas 2 and 3, weakly and continuously increase. We follow Stokey et al. (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 1A.1

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 \( \tilde{f}(x, y) = g_\infty(y) \) 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 \( (x, y) \in \tilde{f}^{-1}(Z) \), 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.

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

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

where \( \phi(\cdot) \) is the probability density function of \( \mathcal{N}(0, \tilde{\sigma}^2) \). If \( v \) is a fixed point of \( T \), then \( v^m(\hat{g}) = (a + e^{\frac{\rho p}{1-\rho}}) v(\hat{g}) \) solves (12). We find a fixed point by demonstrating that \( T : C \to C \) 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 \( Tv \in C \). We have the bound

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

so \( Tv \) is bounded.

Demonstrating continuity is much more complicated. We first apply Lemma 12.14 of Stokey et al. (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. 7 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 \( \zeta \). 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

\[
\left| \int_{-\infty}^{\infty} h(\hat{g}_n, \zeta') \phi(\zeta') d\zeta' - \int_{-\infty}^{\infty} h(\hat{g}, \zeta') \phi(\zeta') d\zeta' \right| \leq \int_{-\infty}^{\infty} |h(\hat{g}_n, \zeta') - h(\hat{g}, \zeta')| |\phi(\zeta') d\zeta' |.
\]

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

7Their 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 \( \zeta \). This boundedness holds because \( v \) is bounded, because \( \lim_{p \to \infty} \hat{p}(\zeta, p) p = 0 \), and because \( \lim_{\zeta \to \infty} (1 - \tilde{\pi}(p, \zeta)) e^{\zeta c} = 0 \) for any \( c > 0 \).
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 IA1. 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 \( \pi = 0 \), and the convergence is uniform for the same reason. Hence, Lemma IA1 applies, and the induced \( \hat{f} \) 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 \( \sup_{p \in (0, \infty)} f(p, \hat{g}) = \sup_{p \in [0, \infty)^*} \hat{f}(p, \hat{g}) \) and because \( [0, \infty)^* \) is compact. Therefore, \( Tv \in C \).

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

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

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

\[
(1 + r_m)^{-1} e^{\frac{(1-\alpha)^2}{2}} < 1.
\]

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

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

\[
= \sigma^2 \left(1 + \rho^2(1 - 2\gamma) + \sqrt{(1 - \rho^2)^2 + 4(1 - \gamma)\gamma\rho^2(1 - \rho^2)}\right).
\]

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

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

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 et al. (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 \( C \), as desired. Call this function \( v^* \). As mentioned above, \( v^m(\hat{g}) = v^*(\hat{g})(a + e^{\frac{\phi d}{1-\rho}}) \) then solves (12); this function clearly inherits the continuity of \( v^* \).

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

\[
(Tv)(\hat{g}) = (1+r)^{-1} \int_{-\infty}^{\infty} \left( \frac{ae^{(1-\rho)\mu+\rho\hat{g}+(1+\rho-\alpha)\zeta}}{a+e^{\rho \hat{g}}} + \frac{e^{\mu+\rho \hat{g}+(1+\rho-\alpha)\zeta}}{a+e^{\rho \hat{g}}}ight) ((1-\lambda)v(g') + \lambda v^s(g')) \phi(\zeta) d\zeta,
\]

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 \( v^s(\hat{g}) = (a+e^{\rho \hat{g}})v(\hat{g}) \) solves (13). Clearly, \( Tv \) is bounded. To prove continuity, we again apply Lemma 12.14 of Stokey et al. (1989), this time with Theorem 3.2 of Stokey et al. (1989) shows that \( \hat{g} \) corresponding to their \( \zeta \), and our \( \zeta \) corresponding to their \( z \). In order to apply their lemma, we have to absorb the \( \zeta \) terms into the \( Q \) transition function so that their \( f \) is bounded. Using the identity \( e^{-z^2/(2\sigma^2)+bz} = e^{\sigma^2b^2/2}e^{-(z-\sigma b)^2/(2\sigma^2)} \), we have

\[
e^{(1+\rho-\alpha)\zeta} \phi(\zeta) = e^{\sigma^2(1+\rho-\alpha)\phi(\zeta - \sigma)(1 + \rho - \alpha)}
\]

and

\[
e^{(1+\rho-\alpha)\zeta} \phi(\zeta) = e^{\sigma^2(1-\alpha)\phi(\zeta - \sigma)(1 - \rho)}.
\]

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

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

\[
1 - \lambda \sup_{\hat{g}} \frac{ae^{(1-\rho)\mu+\rho\hat{g}+(1+\rho-\alpha)\zeta}}{a+e^{\rho \hat{g}}} + e^{\mu+\rho \hat{g}+(1+\rho-\alpha)\zeta} < 1.
\]

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

\[
(1-\lambda)e^{\mu+(1+\rho-\alpha)\zeta} < 1+r,
\]

which always holds, because \( \alpha \in (0,1), \sigma \leq \sigma \) (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 \( C \). Call it \( v^{**} \). Then \( v^*_s(\hat{g}) = (a+e^{\rho \hat{g}})v^{**}(\hat{g}) \) solves (13).
Finally, we show that $v^*$ weakly and continuously increases. Continuity follows from the continuity of $v^{**}$. As argued above, weak monotonicity holds as long as $T : C' \rightarrow C'$, where this set is defined as above. That $T$ maps $C'$ into itself is immediate from (13) and the fact that $v^m$ weakly increases. QED

### D Walrasian Extension

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

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

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

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

In equilibrium, movers must weakly prefer selling at this price versus waiting to sell next period. Therefore, we must have $e^{dt} \geq (1 + r_m)^{-1} E_t e^{dt+1}$, 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 $\tilde{d}_{t-j} = \log(\tilde{P}_{t-j})$ 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

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

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

$$E_t e^{dt+1} = e^{dt} e^{(1-\rho)\mu + \rho \hat{g}_t^m} e^{(\rho^2 \sigma_t^2 + \sigma^2)/2}.$$

Mover optimality therefore requires that

$$\hat{g}_t^m \leq \rho^{-1} (\log(1 + r_m) - (1 - \rho) \mu - (\rho^2 \sigma_t^2 + \sigma^2)/2).$$

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 \( j \geq 1 \), 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, \( V^m_t = \tilde{p}\tilde{d}_t \). (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 \( e^{d_t} \geq \kappa^\lambda(\hat{g}_t)P_t \). 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.

To maintain comparability with the main model, we decrease \( \gamma \) to 0.042 so that the price overshoot is the same in the Walrasian model as in the main model, and we update \( \kappa \) so that the demand error is still zero on average. Under the baseline parameters, the price paths in the Walrasian model seem to be explosive. We believe that prices explode because they adjust more quickly with Walrasian market clearing. Choosing a lower \( \gamma \) leads to more stable price paths as in the baseline model. Other parameters remain the same.
References


FIGURE IA1
The Dynamics of Prices and Volume (Non-Sand-State Cities)

Panel A. Boston, MA

Panel B. Cleveland, OH

Panel C. Portland, OR

Panel D. Seattle, WA

Notes: This figure displays the dynamic relation between prices and volume in the U.S. housing market between 2000 and 2011. In Figure 1, we focus on cities that represent the largest boom–bust cycles. Here, we focus on the largest cities outside of the sand states for which we have both volume and listings data. Variables are defined as in Figure 1. Shaded regions denote the quiet, defined as the period between the peak of volume and the last peak of prices before their pronounced decline.
FIGURE IA2
The Lead–Lag Relationship between Prices and Volume (No Sand States)

Notes: This figure shows that the correlation between prices and lagged volume is robust across MSAs. The figure is constructed as in Figure 2 but excludes MSAs in Arizona, California, Florida, and Nevada.
FIGURE IA3
The Dynamics of Prices and Inventories (Non-Sand-State Cities)

Panel A. Boston, MA
Panel B. Cleveland, OH
Panel C. Portland, OR
Panel D. Seattle, WA

Notes: This figure displays the dynamic relation between prices and inventory in the U.S. housing market between 2000 and 2011. In Figure 3, we focus on cities that represent the largest boom-bust cycles. Here, we focus on the largest cities outside of the sand states for which we have both volume and listings data. Variables are defined as in Figure 3. Shaded regions denote the quiet, defined as the period between the peak of volume and the last peak of prices before their pronounced decline.
FIGURE IA4
Speculative Homebuying and House Price Appreciation, Savings Split

Notes: This figure uses data from Armona et al. (2019) to study the relation between recent house price growth and the probability of buying a non-primary home. The variables are defined as in Figure 7, except we divide the sample based on the household’s level of liquid savings. 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 IA5
Additional Impulse Responses in Counterfactuals

Panel A. \Pr(\text{Sale} \mid \text{Listing}), Rational
Panel B. New Listings by Holding Period, Rational

Panel C. \Pr(\text{Sale} \mid \text{Listing}), Walrasian
Panel D. New Listings by Holding Period, Walrasian

Panel E. \Pr(\text{Sale} \mid \text{Listing}), No Speculation
Panel F. New Listings by Holding Period, No Speculation

Notes: Impulse responses are average differences between log outcomes in control simulations and treatment simulations, in which a 2-standard-deviation shock to $\epsilon_t^g$ (the demand growth innovation) occurs in quarters 0 through 3. A short holding period is defined as less than or equal to 12 quarters.
FIGURE IA6
Adjusted Buying Cutoffs for Different Expected Growth Rates

Panel A. Tax on all buyers

\[ \kappa_{\lambda}(\hat{g}, \tau) \kappa(0)/\kappa(\tau), \]

where \( \tau \) is the tax rate. For both panels we use a tax rate of \( \tau = 0.05 \). Solid lines correspond to occupants (\( n = 1 \)); dashed lines correspond to non-occupants (\( n = 0 \)). The horizontal grey dashed line gives \( \kappa(0) \).

Notes: The adjusted buying cutoff for horizon type \( \lambda \) and occupancy type \( n \) is \( \kappa_{\lambda}(\hat{g}, \tau) \kappa(0)/\kappa(\tau) \), where \( \tau \) is the tax rate. For both panels we use a tax rate of \( \tau = 0.05 \). Solid lines correspond to occupants (\( n = 1 \)); dashed lines correspond to non-occupants (\( n = 0 \)). The horizontal grey dashed line gives \( \kappa(0) \).
## TABLE IA1
List of Metropolitan Statistical Areas Included in the Analysis Sample

<table>
<thead>
<tr>
<th>Metropolitan Statistical Area</th>
<th>Share of Housing Stock Represented</th>
<th>Included in Non-Occupant Analysis</th>
<th>Included in Listings Analysis</th>
<th>Metropolitan Statistical Area</th>
<th>Share of Housing Stock Represented</th>
<th>Included in Non-Occupant Analysis</th>
<th>Included in Listings Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Akron, OH</td>
<td>1.00</td>
<td>x</td>
<td>x</td>
<td>New York–Norwalk–Bridgeport, NY-NJ-CT</td>
<td>0.97</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Ann Arbor, MI</td>
<td>1.00</td>
<td>x</td>
<td>x</td>
<td>North Port–San Antonio–Bradenton, FL</td>
<td>1.00</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Atlantic City–Hammondton, NJ</td>
<td>0.80</td>
<td>x</td>
<td>x</td>
<td>Orlando–Kissimmee–Orlando–Dayton, FL</td>
<td>1.00</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Bakersfield, CA</td>
<td>1.00</td>
<td>x</td>
<td>x</td>
<td>Pensacola–Panama City, FL</td>
<td>1.00</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Baltimore–Columbia–Towson, MD</td>
<td>1.00</td>
<td>x</td>
<td>x</td>
<td>Phoenix–Scottsdale–Avondale, AZ</td>
<td>1.00</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Barnstable, MA</td>
<td>1.00</td>
<td>x</td>
<td>x</td>
<td>Providence–Warwick–Newport, RI-MA</td>
<td>0.76</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Bellingham, WA</td>
<td>1.00</td>
<td>x</td>
<td>x</td>
<td>Providence–Warwick–Newport, RI-MA</td>
<td>0.76</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Bellport-Staten–Norwalk–NY</td>
<td>0.90</td>
<td>x</td>
<td>x</td>
<td>Reno–Sparks–Humboldt, NV</td>
<td>1.00</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Birmingham–Anniston–Tuscaloosa, AL</td>
<td>1.00</td>
<td>x</td>
<td>x</td>
<td>Stockton–Tracy–Lodi, CA</td>
<td>0.94</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Buffalo–Cheektowaga–Niagara Falls, NY</td>
<td>0.90</td>
<td>x</td>
<td>x</td>
<td>Suisun–Vallejo–Rodeo, CA</td>
<td>1.00</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Buffalo–Cheektowaga–Niagara Falls, NY</td>
<td>0.90</td>
<td>x</td>
<td>x</td>
<td>Santa Cruz–Watsonville, CA</td>
<td>1.00</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Camden–Camden County, NJ</td>
<td>1.00</td>
<td>x</td>
<td>x</td>
<td>Santa Maria–Santa Barbara, CA</td>
<td>1.00</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Canton–Massillon, OH</td>
<td>0.92</td>
<td>x</td>
<td>x</td>
<td>Santa Monica–Los Angeles, CA</td>
<td>0.97</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Cape Coral–Fort Myers, FL</td>
<td>0.80</td>
<td>x</td>
<td>x</td>
<td>San Diego–Carlsbad, CA</td>
<td>1.00</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Chico, CA</td>
<td>1.00</td>
<td>x</td>
<td>x</td>
<td>San Francisco–Oakland–Hayward, CA</td>
<td>1.00</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Cleveland–Elyria–OH</td>
<td>0.78</td>
<td>x</td>
<td>x</td>
<td>San Jose–San Mateo–San Francisco, CA</td>
<td>1.00</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Colorado Springs, CO</td>
<td>0.95</td>
<td>x</td>
<td>x</td>
<td>Seattle–Bellevue–Redmond, WA</td>
<td>1.00</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Corpus–Cameron–Brownsville, TX</td>
<td>1.00</td>
<td>x</td>
<td>x</td>
<td>Sebastian–Vero Beach, FL</td>
<td>1.00</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Dayton, OH</td>
<td>0.86</td>
<td>x</td>
<td>x</td>
<td>Solano, CA</td>
<td>1.00</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Des Moines–Ames–Oskaloosa, IA</td>
<td>0.95</td>
<td>x</td>
<td>x</td>
<td>Sonoma–Rohnert–Sebastopol, CA</td>
<td>1.00</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>El Centro, CA</td>
<td>1.00</td>
<td>x</td>
<td>x</td>
<td>Sonoma–Rohnert–Sebastopol, CA</td>
<td>1.00</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>El Paso, TX</td>
<td>0.99</td>
<td>x</td>
<td>x</td>
<td>South Carolina–Myrtle Beach, SC</td>
<td>1.00</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Elyria, OH</td>
<td>1.00</td>
<td>x</td>
<td>x</td>
<td>South Florida–Miami–Fort Lauderdale, FL</td>
<td>1.00</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Erie, PA</td>
<td>1.00</td>
<td>x</td>
<td>x</td>
<td>Stockton–Lodi–Tracy, CA</td>
<td>0.94</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Eugene–OR</td>
<td>1.00</td>
<td>x</td>
<td>x</td>
<td>Surprise–Peoria–Phoenix–Yuma, AZ</td>
<td>1.00</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Gainesville–GA</td>
<td>0.95</td>
<td>x</td>
<td>x</td>
<td>Springfield–Springfield–Springfield, IL</td>
<td>0.94</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Hanford–Coeur d’Alene, CA</td>
<td>1.00</td>
<td>x</td>
<td>x</td>
<td>Springfield–Springfield–Springfield, IL</td>
<td>0.94</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Hartford–West Hartford–East Hartford, CT</td>
<td>1.00</td>
<td>x</td>
<td>x</td>
<td>Springfield–Springfield–Springfield, IL</td>
<td>0.94</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Homestead, FL</td>
<td>1.00</td>
<td>x</td>
<td>x</td>
<td>Springfield–Springfield–Springfield, IL</td>
<td>0.94</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Ithaca, NY</td>
<td>1.00</td>
<td>x</td>
<td>x</td>
<td>Steubenville–Princeton–Pomeroy, OH</td>
<td>0.92</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Jacksonville, FL</td>
<td>0.96</td>
<td>x</td>
<td>x</td>
<td>Tecumseh–Port Huron–Saginaw, MI</td>
<td>1.00</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Kansas–Olathe–Leawood, KS</td>
<td>1.00</td>
<td>x</td>
<td>x</td>
<td>Texarkana–Texarkana–Williamson, TX</td>
<td>1.00</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Kingston, NY</td>
<td>1.00</td>
<td>x</td>
<td>x</td>
<td>Toms River–Millville–North Wildwood, NJ</td>
<td>1.00</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>La Crosse–Onalaska–La Crosse, WI</td>
<td>1.00</td>
<td>x</td>
<td>x</td>
<td>Toledo–OH</td>
<td>1.00</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Lancaster, PA</td>
<td>1.00</td>
<td>x</td>
<td>x</td>
<td>Toledo–OH</td>
<td>1.00</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Las Vegas–Henderson–Paradise, NV</td>
<td>1.00</td>
<td>x</td>
<td>x</td>
<td>Tonopah–Pahrump–Pahrump, NV</td>
<td>1.00</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Los Angeles–Long Beach, CA</td>
<td>1.00</td>
<td>x</td>
<td>x</td>
<td>Towson–Baltimore–Baltimore–Arlington, MD-WV</td>
<td>0.95</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Madison, WI</td>
<td>1.00</td>
<td>x</td>
<td>x</td>
<td>Towson–Baltimore–Baltimore–Arlington, MD-WV</td>
<td>0.95</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Merced, CA</td>
<td>1.00</td>
<td>x</td>
<td>x</td>
<td>Towson–Baltimore–Baltimore–Arlington, MD-WV</td>
<td>0.95</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Miami–Fort Lauderdale–West Palm Beach, FL</td>
<td>1.00</td>
<td>x</td>
<td>x</td>
<td>Vero Beach–Fort Pierce, FL</td>
<td>1.00</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Modesto, CA</td>
<td>1.00</td>
<td>x</td>
<td>x</td>
<td>Vista–Escondido–San Marcos, CA</td>
<td>1.00</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Naples, FL</td>
<td>1.00</td>
<td>x</td>
<td>x</td>
<td>Virginia Beach–Norfolk–Hampton, VA</td>
<td>1.00</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>New Haven–Milford, CT</td>
<td>1.00</td>
<td>x</td>
<td>x</td>
<td>Virginia Beach–Norfolk–Hampton, VA</td>
<td>1.00</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>

**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 IA2  
Number of Transactions Dropped During Sample Selection

<table>
<thead>
<tr>
<th>Dropped:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-unique CoreLogic ID</td>
<td>50</td>
</tr>
<tr>
<td>Non-positive price</td>
<td>3,309,100</td>
</tr>
<tr>
<td>Nominal foreclosure transfer</td>
<td>531,786</td>
</tr>
<tr>
<td>Duplicate transaction</td>
<td>609,756</td>
</tr>
<tr>
<td>Subdivision sale</td>
<td>1,304,920</td>
</tr>
<tr>
<td>Vacant lot</td>
<td>831,774</td>
</tr>
</tbody>
</table>

| Final Number of Transactions | 51,080,640 |

Notes: This table shows the number of transactions dropped at each stage of our sample-selection procedure.
### TABLE IA3
Mechanical Short-Term Volume Estimates

<table>
<thead>
<tr>
<th>Year</th>
<th>( \hat{\alpha}<em>{buy} - \hat{\alpha}</em>{2000} )</th>
<th>Total Volume</th>
<th>Actual Short-Term Volume</th>
<th>Counterfactual Short-Term Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>0</td>
<td>2821596</td>
<td>512787</td>
<td>512787</td>
</tr>
<tr>
<td>2001</td>
<td>0.0003</td>
<td>2757954</td>
<td>499643</td>
<td>494741</td>
</tr>
<tr>
<td>2002</td>
<td>0.0008</td>
<td>2985550</td>
<td>556987</td>
<td>534342</td>
</tr>
<tr>
<td>2003</td>
<td>0.0014</td>
<td>3226968</td>
<td>614492</td>
<td>557701</td>
</tr>
<tr>
<td>2004</td>
<td>0.0023</td>
<td>3667997</td>
<td>772708</td>
<td>659111</td>
</tr>
<tr>
<td>2005</td>
<td>0.0027</td>
<td>3857236</td>
<td>909976</td>
<td>725847</td>
</tr>
<tr>
<td>2000–2005 growth</td>
<td>–</td>
<td>36.7%</td>
<td>77.5%</td>
<td>41.5%</td>
</tr>
</tbody>
</table>

**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 \( \alpha_{buy} \), a fixed effect for the propensity to sell a house having bought it in year \( y \), using the regression equation in Section B.1. In the counterfactual, we assume that \( \alpha_{buy} \) remains constant at its level in \( y = 2000 \) for \( y \in \{2001, 2002, 2003, 2004, 2005\} \).
<table>
<thead>
<tr>
<th></th>
<th>First Stage</th>
<th>Volume Boom</th>
<th>Price Boom</th>
<th>Price Bust</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>OLS</td>
<td>IV</td>
<td>OLS</td>
</tr>
<tr>
<td>Short-Volume Boom</td>
<td></td>
<td>2.28***</td>
<td>(0.12)</td>
<td>2.28***</td>
</tr>
<tr>
<td>Old Share</td>
<td></td>
<td>1.69***</td>
<td>(0.26)</td>
<td></td>
</tr>
<tr>
<td>Young Share</td>
<td></td>
<td>0.66**</td>
<td>(0.32)</td>
<td></td>
</tr>
<tr>
<td>Number of Observations</td>
<td></td>
<td>102</td>
<td>102</td>
<td>102</td>
</tr>
<tr>
<td>R-squared</td>
<td></td>
<td>0.45</td>
<td>0.79</td>
<td>0.79</td>
</tr>
<tr>
<td>F-Statistic</td>
<td></td>
<td>39.95</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: This table presents OLS and IV regressions at the MSA level of price and volume housing cycle measures on the change in short-holding-period volume from 2000 to 2005 relative to total volume in 2000. In the IV regressions, 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. Census microdata was not available for 13 MSAs in our sample, hence the lower sample count in this table. The first column presents the first-stage regression and F-statistic.
### TABLE IA5
House Price Appreciation and Speculative Buyer Shares (Monthly Panel VAR)

<table>
<thead>
<tr>
<th></th>
<th>House Price Appreciation Rate</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lagged Price Appreciation</td>
<td>0.375***</td>
<td>0.387***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.026)</td>
<td>(0.027)</td>
</tr>
<tr>
<td></td>
<td>Lagged Short-Buyer Share</td>
<td>0.021***</td>
<td>0.023***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.005)</td>
<td>(0.005)</td>
</tr>
<tr>
<td></td>
<td>Lagged Non-Occupant Share</td>
<td>0.009</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.008)</td>
<td>(0.006)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Short-Buyer Share</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lagged Price Appreciation</td>
<td>0.163***</td>
<td>0.162***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.048)</td>
<td>(0.048)</td>
</tr>
<tr>
<td></td>
<td>Lagged Short-Buyer Share</td>
<td>0.780***</td>
<td>0.781***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.024)</td>
<td>(0.023)</td>
</tr>
<tr>
<td></td>
<td>Lagged Non-Occupant Share</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.017</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Non-Occupant Share</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lagged Price Appreciation</td>
<td>0.124***</td>
<td>0.172***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.044)</td>
<td>(0.045)</td>
</tr>
<tr>
<td></td>
<td>Lagged Short-Buyer Share</td>
<td>-0.071***</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.016)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lagged Non-Occupant Share</td>
<td>0.892***</td>
<td>0.900***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.025)</td>
<td>(0.021)</td>
</tr>
</tbody>
</table>

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 IA6
Summary Statistics for Volume and Inventory Outcomes by MSA Characteristics

<table>
<thead>
<tr>
<th></th>
<th>Log Income</th>
<th>Log Population</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Tercile 1</td>
<td>Tercile 2</td>
</tr>
<tr>
<td>Δ Volume Boom</td>
<td>0.70</td>
<td>0.48</td>
</tr>
<tr>
<td></td>
<td>(0.49)</td>
<td>(0.30)</td>
</tr>
<tr>
<td>Δ Volume Quiet + Bust</td>
<td>-0.68</td>
<td>-0.65</td>
</tr>
<tr>
<td></td>
<td>(0.21)</td>
<td>(0.16)</td>
</tr>
<tr>
<td>Δ Listings Boom</td>
<td>1.04</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td>(0.99)</td>
<td>(0.97)</td>
</tr>
<tr>
<td>Δ Listings Quiet</td>
<td>1.77</td>
<td>1.90</td>
</tr>
<tr>
<td></td>
<td>(1.35)</td>
<td>(1.56)</td>
</tr>
<tr>
<td>Share Non-Hispanic White</td>
<td>0.52</td>
<td>0.48</td>
</tr>
<tr>
<td></td>
<td>(0.47)</td>
<td>(0.45)</td>
</tr>
<tr>
<td>Δ Volume Quiet + Bust</td>
<td>-0.62</td>
<td>-0.68</td>
</tr>
<tr>
<td></td>
<td>(0.17)</td>
<td>(0.20)</td>
</tr>
<tr>
<td>Δ Listings Boom</td>
<td>0.87</td>
<td>1.13</td>
</tr>
<tr>
<td></td>
<td>(0.66)</td>
<td>(1.00)</td>
</tr>
<tr>
<td>Δ Listings Quiet</td>
<td>1.99</td>
<td>1.72</td>
</tr>
<tr>
<td></td>
<td>(1.02)</td>
<td>(1.47)</td>
</tr>
<tr>
<td>Share College Educated</td>
<td>0.71</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>(0.53)</td>
<td>(0.28)</td>
</tr>
<tr>
<td>Δ Volume Quiet + Bust</td>
<td>-0.70</td>
<td>-0.58</td>
</tr>
<tr>
<td></td>
<td>(0.21)</td>
<td>(0.17)</td>
</tr>
<tr>
<td>Δ Listings Boom</td>
<td>1.06</td>
<td>0.74</td>
</tr>
<tr>
<td></td>
<td>(1.10)</td>
<td>(0.91)</td>
</tr>
<tr>
<td>Δ Listings Quiet</td>
<td>2.55</td>
<td>1.31</td>
</tr>
<tr>
<td></td>
<td>(2.05)</td>
<td>(0.83)</td>
</tr>
</tbody>
</table>

**Notes:** This table reports summary statistics for MSA-level variables in different samples of MSAs. For each MSA characteristic (e.g., Log Income), MSAs are sorted into three groups based on the terciles of that characteristic in the full sample of MSAs for which that characteristic is available. The mean of each variable and its standard deviation (in parentheses) within that group of MSAs is recorded in the table. The number of MSAs over which that mean and standard deviation are calculated is reported in brackets. This number can vary because the terciles are calculated in the full sample of MSAs but the variable being tabulated is only available for a subset of MSAs (e.g., the listings outcomes) or because the MSA characteristic is only available for a subset of MSAs. Housing supply elasticities are taken from Saiz (2010) and all other MSA characteristics are from the 2000 Census 5% microdata. Δ Volume Boom is defined as the change in total volume from 2000 through 2005. 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. 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.
### TABLE IA7
Speculators and Housing Market Outcomes (Extra Listing Outcomes)

#### Panel A. Propensity to List

<table>
<thead>
<tr>
<th></th>
<th>Δ New Listings Boom</th>
<th>Δ New Listings Quiet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short-Volume Boom</td>
<td>0.270 (0.182)</td>
<td>0.649*** (0.160)</td>
</tr>
<tr>
<td>Non-Occupant Volume Boom</td>
<td>0.115 (0.092)</td>
<td>0.308*** (0.080)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Number of Observations</th>
<th>R-squared</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>57</td>
<td>0.038</td>
</tr>
<tr>
<td></td>
<td>48</td>
<td>0.033</td>
</tr>
</tbody>
</table>

#### Panel B. Sale Probability

<table>
<thead>
<tr>
<th></th>
<th>Δ P(Sale) Boom</th>
<th>Δ P(Sale) Quiet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short-Volume Boom</td>
<td>0.142*** (0.032)</td>
<td>-0.163*** (0.031)</td>
</tr>
<tr>
<td>Non-Occupant Volume Boom</td>
<td>0.058*** (0.017)</td>
<td>-0.047** (0.018)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Number of Observations</th>
<th>R-squared</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>57</td>
<td>0.268</td>
</tr>
<tr>
<td></td>
<td>48</td>
<td>0.206</td>
</tr>
</tbody>
</table>

*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 IA8**  
Speculative Booms and Housing Market Outcomes (Sand State Control)

**Panel A. MSA-Level Prices**

<table>
<thead>
<tr>
<th></th>
<th>Price Boom</th>
<th>Price Bust</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Short-Volume Boom</strong></td>
<td>1.022*** (0.272)</td>
<td>-0.237*** (0.061)</td>
</tr>
<tr>
<td><strong>Non-Occupant Volume Boom</strong></td>
<td>0.228 (0.142)</td>
<td>-0.044 (0.032)</td>
</tr>
<tr>
<td><strong>Number of Observations</strong></td>
<td>115 102</td>
<td>115 102</td>
</tr>
<tr>
<td><strong>R-squared</strong></td>
<td>0.514 0.453</td>
<td>0.696 0.662</td>
</tr>
</tbody>
</table>

**Panel B. MSA-Level Inventories**

<table>
<thead>
<tr>
<th></th>
<th>Δ Listings Boom</th>
<th>Δ Listings Quiet</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Short-Volume Boom</strong></td>
<td>-1.581 (1.163)</td>
<td>4.276*** (1.461)</td>
</tr>
<tr>
<td><strong>Non-Occupant Volume Boom</strong></td>
<td>-0.206 (0.525)</td>
<td>1.930*** (0.642)</td>
</tr>
<tr>
<td><strong>Number of Observations</strong></td>
<td>57 48</td>
<td>57 48</td>
</tr>
<tr>
<td><strong>R-squared</strong></td>
<td>0.034 0.020</td>
<td>0.337 0.440</td>
</tr>
</tbody>
</table>

**Panel C. MSA-Level Volume Quiet and Bust**

<table>
<thead>
<tr>
<th></th>
<th>Δ Volume Quiet + Bust</th>
<th>Foreclosures Bust</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Short-Volume Boom</strong></td>
<td>-1.145*** (0.105)</td>
<td>-0.233 (0.377)</td>
</tr>
<tr>
<td><strong>Non-Occupant Volume Boom</strong></td>
<td>-0.516*** (0.053)</td>
<td>-0.451** (0.185)</td>
</tr>
<tr>
<td><strong>Number of Observations</strong></td>
<td>115 102</td>
<td>115 102</td>
</tr>
<tr>
<td><strong>R-squared</strong></td>
<td>0.533 0.505</td>
<td>0.317 0.333</td>
</tr>
</tbody>
</table>

**Notes:** This table reports estimates of the relation between speculative volume and housing cycle measures at the MSA level. The table follows Table 2 while adding a control for “Sand States,” which is an indicator for MSAs in Arizona, California, Florida, and Nevada.
### TABLE IA9
Speculators and Housing Market Outcomes (Extra Listing Outcomes, Sand State Control)

#### Panel A. Propensity to List

<table>
<thead>
<tr>
<th>Speculative Volume</th>
<th>Δ New Listings Boom</th>
<th>Δ New Listings Quiet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short-Volume Boom</td>
<td>0.050</td>
<td>0.431**</td>
</tr>
<tr>
<td></td>
<td>(0.198)</td>
<td>(0.171)</td>
</tr>
<tr>
<td>Non-Occupant Volume Boom</td>
<td>0.040</td>
<td>0.228***</td>
</tr>
<tr>
<td></td>
<td>(0.087)</td>
<td>(0.072)</td>
</tr>
<tr>
<td>Number of Observations</td>
<td>57</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td>48</td>
<td>48</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.131</td>
<td>0.323</td>
</tr>
<tr>
<td></td>
<td>0.213</td>
<td>0.451</td>
</tr>
</tbody>
</table>

#### Panel B. Sale Probability

<table>
<thead>
<tr>
<th>Speculative Volume</th>
<th>Δ P(Sale) Boom</th>
<th>Δ P(Sale) Quiet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short-Volume Boom</td>
<td>0.146***</td>
<td>-0.086***</td>
</tr>
<tr>
<td></td>
<td>(0.036)</td>
<td>(0.028)</td>
</tr>
<tr>
<td>Non-Occupant Volume Boom</td>
<td>0.058***</td>
<td>-0.021</td>
</tr>
<tr>
<td></td>
<td>(0.018)</td>
<td>(0.013)</td>
</tr>
<tr>
<td>Number of Observations</td>
<td>57</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td>48</td>
<td>48</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.268</td>
<td>0.598</td>
</tr>
<tr>
<td></td>
<td>0.206</td>
<td>0.607</td>
</tr>
</tbody>
</table>

Notes: This table reports estimates of the relation between speculative volume and housing cycle measures at the MSA level. The table follows Table IA7 while adding a control for “Sand States,” which is an indicator for MSAs in Arizona, California, Florida, and Nevada.
### TABLE IA10
All-Cash Buyer Shares and Mean LTV by Buyer Type

<table>
<thead>
<tr>
<th></th>
<th>Transaction-Level</th>
<th>MSA-Level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All Months</td>
<td>All Months</td>
</tr>
<tr>
<td></td>
<td></td>
<td>All-Cash Buyer Share</td>
</tr>
<tr>
<td>Short Buyers</td>
<td>0.29 (0.21)</td>
<td>0.38 (0.16)</td>
</tr>
<tr>
<td>Non-Occupant Buyers</td>
<td>0.38 (0.18)</td>
<td>0.41 (0.15)</td>
</tr>
<tr>
<td>All Buyers</td>
<td>0.20 (0.16)</td>
<td>0.25 (0.15)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Mean LTV</th>
<th>Mean LTV</th>
<th>Mean LTV</th>
<th>Mean LTV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short Buyers</td>
<td>0.59 (0.40)</td>
<td>0.52 (0.18)</td>
<td>0.60 (0.13)</td>
<td>0.59 (0.13)</td>
</tr>
<tr>
<td>Non-Occupant Buyers</td>
<td>0.50 (0.41)</td>
<td>0.48 (0.14)</td>
<td>0.52 (0.12)</td>
<td>0.54 (0.11)</td>
</tr>
<tr>
<td>All Buyers</td>
<td>0.65 (0.36)</td>
<td>0.62 (0.13)</td>
<td>0.64 (0.12)</td>
<td>0.64 (0.11)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Mean LTV</th>
<th>Mean LTV</th>
<th>Mean LTV</th>
<th>Mean LTV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short Buyers</td>
<td>0.84 (0.16)</td>
<td>0.85 (0.06)</td>
<td>0.84 (0.05)</td>
<td>0.82 (0.04)</td>
</tr>
<tr>
<td>Non-Occupant Buyers</td>
<td>0.81 (0.17)</td>
<td>0.82 (0.06)</td>
<td>0.82 (0.06)</td>
<td>0.80 (0.05)</td>
</tr>
<tr>
<td>All Buyers</td>
<td>0.82 (0.16)</td>
<td>0.83 (0.05)</td>
<td>0.82 (0.04)</td>
<td>0.80 (0.04)</td>
</tr>
</tbody>
</table>

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 1.1. 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>
<thead>
<tr>
<th>Outcome</th>
<th>μ₁ = 0.033</th>
<th>μ₁ = 0.066</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Baseline</td>
<td>All long-term buyers</td>
</tr>
<tr>
<td>Price boom</td>
<td>14.0</td>
<td>8.7</td>
</tr>
<tr>
<td>Price bust</td>
<td>−7.6</td>
<td>−0.4</td>
</tr>
<tr>
<td>Volume boom</td>
<td>5.6</td>
<td>2.9</td>
</tr>
<tr>
<td>Listings, end of boom</td>
<td>−1.2</td>
<td>−3.1</td>
</tr>
<tr>
<td>Listings, end of quiet</td>
<td>1.3</td>
<td>0.4</td>
</tr>
<tr>
<td>Short volume boom</td>
<td>13.8</td>
<td>3.4</td>
</tr>
<tr>
<td>Non-occupant volume boom</td>
<td>11.3</td>
<td>5.6</td>
</tr>
<tr>
<td>Sale probability boom</td>
<td>6.8</td>
<td>6.0</td>
</tr>
</tbody>
</table>

Notes: We report 100 times changes in log outcomes between treatment and control simulations. See notes to Table 6 for outcome definitions. For each value of μ₁, we re-choose the other parameters in Table 5 by matching the targets in Table 4 other than non-occupant boom/occupant boom. The Baseline column reports outcomes under each new set of parameters. In the All long-term buyers column, we further change the βₙ distribution to put all weight on λ = 0.03 while keeping the occupancy distribution unchanged, corresponding to the third column of results in Table 6. In the All occupants column, we further change the βₙ distribution to put all weight on n = 1 while keeping the λ distribution unchanged, corresponding to the fifth column of results in Table 6.