Internet Appendices for "Speculative Dynamics of Prices and Volume"

ANTHONY A. DEFUSCO
CHARLES G. NATHANSON
ERIC ZWICK

A Data

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.[1] 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.

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

3. 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.

4. 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.

5. 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

[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.

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.

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

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

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

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:

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

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

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 $2000m1 \leq t' < t \leq 2005m12$, we construct the counterfactual selling hazard as

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

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 $1995m1 \leq t' < 2005m12$,

$$e^c_{t',t'} = e_{t',t'}$$
$$v^c_{t',t'} = v_{t',t'}.$$

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

$$e^c_{t',t} = e^c_{t',t-1} - v^c_{t',t-1}$$
$$v^c_{t',t} = \pi^c_{t',t} e^c_{t',t}.$$

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 $v^c_{t',t}$ 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 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.[4] 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

[4]Gao et al. (2020) 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.

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 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.

C Additional analysis of speculation

In this appendix, we provide details about the calculations using microdata in Section 4.

C.1 Overlap between short-term and non-occupant buyers

The statistics in the text 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). Therefore, non-occupants account for an excess share of the growth in short-term buyers.

In a related approach, we measure the share of 2000–2005 non-occupant purchases that later become short-term sales. These calculations afford a direct comparison to the 2000–2005 increase in non-occupant volume that we analyze in Section 3. However, they are not completely comparable to the ones above, because they look until 2008 to see if a purchase becomes a short-term sale. 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). These numbers imply there was not a shift in the composition of non-occupant buyers during the boom toward short-term behavior. However, it is difficult to measure short investment horizons of buyers at the end of the boom because many listings from 2006–2008 did not sell quickly. Another interpretation of these results is that there was secular growth in long-term non-occupants alongside the entry of short-term speculators during the boom.

C.2 Credit utilization

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 $\in [60\%, 85\%)$), and a high-leverage group (purchase LTV $> 85\%$). Of the short-term sellers in 2000–2005, 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: 22% were low-LTV buyers, 30% were medium-LTV buyers, and 48% were high-LTV buyers. 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.[5] As in our analysis of cash transactions among speculators, these

[5] Of the short-term sellers in 2000–2005 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.

statistics reveal that short-term volume is associated with lower use of leverage in the cross-section relative to the general population.[6] At the same time, the proportional growth in short-term buying is stronger among medium- and high-LTV sellers, making a larger relative contribution to the overall growth in short-term volume.

C.3 Buyer scale and experience

Scale. We mark transactions as developer purchases when the buyer's name is not parsed as a person by CoreLogic and contains strings reflecting developer names. We identify developer names using CoreLogic's internal new construction flag, as described in Online Appendix A. Both this analysis and the analysis of inexperienced investors below exclude transactions with missing buyer 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).

Experience. To flag non-developers as experienced or inexperienced, we count the total number of transactions for each unique buyer name in an MSA. 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). Thus, inexperienced buyers constitute 2.42 million of 3.95 million total short-term sales, or 61%. Between 2000 and 2005, the number of inexperienced short-term-buyer sales increases from 310 thousand to 560 thousand, or 66% of the growth in short-term sales (excluding developers). The quantitative relevance of inexperienced buyers for volume is consistent with the evidence in Bayer et al. (2020).

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" (Bayer et al., 2020; Choi et al., 2014).

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-

[6] In Table IA9 of the online appendix, we extend Table 3 to look at average purchase LTVs for short-term and non-occupant buyers. Both speculative buyer types have lower average LTVs, which is exclusively driven by their higher all-cash shares.

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.[7] 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.[8] 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.

D Relation of model to literature

As mentioned in the Introduction, existing theoretical papers explain the comovement of prices and volume. However, 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).[9] These models cannot

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

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

[9]Two exceptions are Hedlund (2016) and Ngai and Sheedy (2020), who respectively focus on credit

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.[10] 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, listings modestly rise during the boom in the aggregate. The sharp rise in listings during the quiet, far above their 2000 level, is perhaps the most salient aspect of Figure 3.

constraints and within-market moves. As we explain in Online Appendices C.2 and C.3, short-term volume increases significantly among low-LTV sellers, and most short-term sellers do not relocate within the same MSA. Therefore, these two papers do not explain a substantial share of the disproportionate rise in short-term volume during the boom.

[10] 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).

E Proofs

E.1 Lemma 1

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

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

$$\sigma_l^2 = (2\rho^2)^{-1}\left(-(1-\rho^2)\sigma^2_{\epsilon^d} - \sigma^2_{\epsilon^g} + \sqrt{((1-\rho^2)\sigma^2_{\epsilon^d} + \sigma^2_{\epsilon^g})^2 + 4\rho^2\sigma^2_{\epsilon^d}\sigma^2_{\epsilon^g}}\right).$$

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

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

Because $\hat{g}_t = (1-\rho)\mu_g + \rho g_t^*$, we have proved the formula in the lemma for $\hat{g}_t$. We showed above that $\hat{\sigma}_g^2 = \rho^2\sigma_l^2 + \sigma^2_{\epsilon^g}$. We have $d_t = d_{t-1} + g_t + \epsilon_t^d = (d_{t-1} - \tilde{d}_{t-1}) + \tilde{d}_{t-1} + (1-\rho)\mu_g + \rho g_{t-1} + \epsilon_t^g + \epsilon_t^d = (d_{t-1} - \tilde{d}_{t-1}) + \tilde{d}_{t-1} + \hat{g}_t + \rho\zeta_t^g + \epsilon_t^g + \epsilon_t^d$, which immediately gives $\hat{d}_t = \tilde{d}_{t-1} + \hat{g}_t$, with $\hat{\sigma}_d^2 = \rho^2\sigma_l^2 + \sigma^2_{\epsilon^g} + \sigma^2_{\epsilon^d}$.

The bound we assume for r (see Section 5.1) is

$$r > e^{\mu_g + \frac{(1-\alpha\rho)^2\hat{\sigma}_d^2}{2(1-\rho)^2}} - 1. \tag{E1}$$

E.2 Lemma 2

By Lemma 1, $\Delta\tilde{d}_t = \hat{g}_t + (\tilde{d}_t - \hat{d}_t)$. Furthermore, $\hat{g}_{t+1} = \mu_g + (\alpha\rho)(\hat{g}_t - \mu_g) + (1-\alpha)\rho(\Delta\tilde{d}_t - \mu_g) = (1-\rho)\mu_g + \rho\hat{g}_t + (1-\alpha)\rho(\tilde{d}_t - \hat{d}_t)$. Finally, $\hat{d}_{t+1} = \tilde{d}_t + \hat{g}_{t+1} = \hat{d}_t + (1-\rho)\mu_g + \rho\hat{g}_t + (1 + (1-\alpha)\rho)(\tilde{d}_t - \hat{d}_t)$. From the point of view of agents, $\tilde{d}_t = d_t$. Therefore,

$$\hat{d}_{t+1} = \hat{d}_t + (1-\rho)\mu_g + \rho\hat{g}_t + (1 + (1-\alpha)\rho)\zeta_t \tag{E2}$$

$$\hat{g}_{t+1} = (1-\rho)\mu_g + \rho\hat{g}_t + (1-\alpha)\rho\zeta_t, \tag{E3}$$

where $\zeta_t \equiv d_t - \hat{d}_t$.

Write $V^m(\hat{d}_t, \hat{g}_t) = e^{\hat{d}_t}v^m(\hat{d}_t, \hat{g}_t)$ and $P = e^{\hat{d}_t}p$. Then $\tilde{\pi}(P, d_t) = 1 - F(\log p + \log\overline{\kappa} - \zeta_t)$, which we denote $\tilde{\pi}(p, \zeta_t)$ by abuse of notation. Substituting these expressions into (6) and using (E2) yield

$$v^m(\hat{d}_t, \hat{g}_t) = \sup_p E\left(\tilde{\pi}(p, \zeta_t)p + \frac{(1-\tilde{\pi}(p,\zeta_t))e^{(1-\rho)\mu_g + \rho\hat{g}_t + (1+\rho-\alpha\rho)\zeta_t} v^m(\hat{d}_{t+1}, \hat{g}_{t+1})}{1+r_m}\right), \quad \text{(E4)}$$

with the expectation over $\zeta_t \sim \mathcal{N}(0, \hat{\sigma}_d^2)$ and $\hat{g}_{t+1}$ given by (E3). Because $\hat{d}_t$ and $\hat{d}_{t+1}$ appear only in the first argument of v^m, this function does not depend on $\hat{d}_t$, so

$$V^m(\hat{d}_t, \hat{g}_t) = e^{\hat{d}_t} v^m(\hat{g}_t). \quad \text{(E5)}$$

It follows that the argmax of (E4) does not depend on $\hat{d}_t$. We denote it $p(\hat{g}_t)$.

E.3 Lemma 3

When $r_m \to \infty$, $p(\cdot)$ becomes constant, as is clear from (E4). In this case, the formula for $\hat{d}_{t+1}$ at the beginning of the proof of Lemma 2 implies that $\Delta \log P_{t+1} = (1-\rho)\mu_g + \rho\hat{g}_t + (1+(1-\alpha)\rho)(\tilde{d}_t - \hat{d}_t)$. Solving for $\tilde{d}_t - \hat{d}_t$ and substituting it into the formula for $\hat{g}_{t+1}$ there yields $\hat{g}_{t+1} = (1+(1-\alpha)\rho)^{-1}((1-\rho)\mu_g + \rho\hat{g}_t + (1-\alpha)\rho\Delta\log P_{t+1})$. Iterating this formula backwards (and then subtracting 1 from the time subscripts) gives

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

Conditional on market data before t, agents at t believe that $E(\tilde{d}_t - \hat{d}_t) = 0$. Therefore, $E\Delta\log P_{t+1} = (1-\rho)\mu_g + \rho\hat{g}_t$. Substituting in the expression just derived for $\hat{g}_t$ gives the first equation in the lemma.

To derive the second equation, we let $\overline{p}$ denote the constant value of $p(\cdot)$ that holds in the limit as $r_m \to \infty$. From (5), $\tilde{d}_t = \hat{d}_t + \log(\overline{\kappa}\,\overline{p}) - F^{-1}(1-\pi_t)$. Therefore, the equation above for $\Delta\log P_{t+1}$ implies that $\Delta\log P_{t+1} = E\Delta\log P_{t+1} + (1+(1-\alpha)\rho)\left(\log(\overline{\kappa}\,\overline{p}) - F^{-1}(1-\pi_t)\right)$, as claimed.

E.4 Lemma 4

A potential buyer at t observes the history of price changes, $P_{t'}/P_{t'-1}$, but not past price levels. Therefore, her information set is different than the one in the statement of Lemma 1. Nonetheless, she still computes $\hat{g}_t$ using the formula in Lemma 1, as that formula depends only on past price changes and not past price levels. However, the formula for $\hat{d}_t$ does not work because it requires knowledge of P_{t-1}. Therefore, she imputes $\hat{d}_t$ using her knowledge of equilibrium and the list price she observes. In particular, given Lemma 2, $P = e^{\hat{d}_t}p(\hat{g}_t)$, which implies that $\hat{d}_t = \log(P/p(\hat{g}_t))$. The potential buyer's decision rule is therefore

$$V^b\left(\log\left(\frac{P}{p(\hat{g}_t)}\right), \hat{g}_t; \lambda, \delta, n\right) \geq P. \quad \text{(E6)}$$

The proof proceeds by showing that this inequality is equivalent to the one in Lemma 4 through suitable choice of $\kappa_{n,j}(\hat{g}_t)$.

Write $V^s(\hat{d}_t, \hat{g}_t; \lambda, \delta) = (r+\lambda)^{-1}e^\delta + e^{\hat{d}_t}v^s(\hat{d}_t, \hat{g}_t; \lambda, \delta)$. Substituting this equation, (E2), and (E5) into (8) yields

$$v^s(\hat{d}_t, \hat{g}_t; \lambda, \delta) = (1+r)^{-1}E\left(e^{(1-\rho)\mu_g+\rho\hat{g}_t+(1+\rho-\alpha\rho)\zeta_t}(\lambda v^m(\hat{g}_{t+1}) + (1-\lambda)v^s(\hat{d}_{t+1}, \hat{g}_{t+1}; \lambda, \delta))\right), \tag{E7}$$

with the expectation over $\zeta_t \sim \mathcal{N}(0, \hat{\sigma}_d^2)$ and $\hat{g}_{t+1}$ given by (E3). Because $\hat{d}_t$, $\hat{d}_{t+1}$, and δ appear only in the arguments of v^s, that function does not depend on $\hat{d}_t$ and δ, allowing us to write $V^s(\hat{d}_t, \hat{g}_t; \lambda, \delta) = (r+\lambda)^{-1}e^\delta + e^{\hat{d}_t}v^s(\hat{g}_t; \lambda)$. Substituting this equation, (E2), and (E5) into (7) yields

$$V^b(\hat{d}_t, \hat{g}_t; \lambda, \delta, n) = \frac{e^\delta}{r+\lambda} + \frac{e^{\hat{d}_t}}{1+r}E\left(e^{(1-\rho)\mu_g+\rho\hat{g}_t+(1+\rho-\alpha\rho)\zeta_t}\left(\lambda v^m(\hat{g}_{t+1}) + (1-\lambda)v^s(\hat{g}_{t+1}; \lambda)\right)\right),$$

with the expectation over $\zeta_t \sim \mathcal{N}\left(\frac{\hat{\sigma}_d^2(\delta-\hat{d}_t-\mu_n)}{\sigma_a^2+\hat{\sigma}_d^2}, \frac{\sigma_a^2\hat{\sigma}_d^2}{\sigma_a^2+\hat{\sigma}_d^2}\right)$ and $\hat{g}_{t+1}$ given by (E3). Let $\Psi(\zeta_t, \hat{g}_t; \lambda)$ denote the argument inside the expectation. We can simplify the buying decision, (E6), to

$$\frac{e^\delta}{P} \geq (r+\lambda)\left(1 - \frac{E\Psi(\zeta_t, \hat{g}_t; \lambda)}{(1+r)p(\hat{g}_t)}\right), \tag{E8}$$

with the expectation over $\zeta_t \sim \mathcal{N}\left(\frac{\hat{\sigma}_d^2(\log(e^\delta/P)+\log p(\hat{g}_t)-\mu_n)}{\sigma_a^2+\hat{\sigma}_d^2}, \frac{\sigma_a^2\hat{\sigma}_d^2}{\sigma_a^2+\hat{\sigma}_d^2}\right)$.

To proceed, we use the following lemma about $v^m(\hat{g}_t)$ and $v^s(\hat{g}_t; \lambda)$:

Lemma IA1. *For all $\lambda > 0$, $v^m(\hat{g}_t)$ and $v^s(\hat{g}_t, \lambda)$ are continuous and weakly increasing functions of $\hat{g}_t$.*

Proof. Appendix E.5. □

From IA1, it follows immediately that $\Psi(\zeta_t, \hat{g}_t; \lambda)$ is a continuous and weakly increasing function of ζ_t for any g_t and $\lambda > 0$, which implies that the right side of (E8) continuously and weakly decreases in e^δ/P. The left side continuous and strictly increases in e^δ/P. Therefore, for n, j, and $\hat{g}_t$ such that the right side does not limit to a positive number as $\delta \to \infty$ for $\lambda = \lambda_j$, then (E8) holds for all δ, meaning that Lemma 4 holds with $\kappa_{n,j}(\hat{g}_t) = 0$. If the right side limits to a positive number as $\delta \to -\infty$ when $\lambda = \lambda_j$, then by the Intermediate Value Theorem, there exists a unique $\kappa_{n,j}(\hat{g}_t)$ such that the inequality holds if and only if $e^\delta/P \geq \kappa_{n,j}(\hat{g}_t)$, which proves Lemma 4.

E.5 Value Function Monotonicity

This section establishes that the functions $v^m(\cdot)$ and $v^s(\cdot; \lambda)$, which we define in the proofs of Lemmas 2 and 4, 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 IA2. *Let $f : (0,\infty) \times \mathbb{R} \to \mathbb{R}$ be continuous. Suppose there exists functions $g_0 : \mathbb{R} \to \mathbb{R}$ and $g_\infty : \mathbb{R} \to \mathbb{R}$ such that $\lim_{x\to 0} f(x,y) = g_0(y)$ and $\lim_{x\to\infty} g_\infty(y)$ uniformly. Define $\tilde{f} : [0,\infty)^* \times \mathbb{R} \to \mathbb{R}$ by $\tilde{f}(x,y) = f(x,y)$ for $x \in (0,\infty)$ and $\tilde{f}(x,y) = g_x(y)$ for $x \in \{0,\infty\}$. Then $\tilde{f}$ is continuous.*

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

We next establish the existence of a continuous solution $v^m(\cdot)$ to (E4). Let $\mathcal{C}$ be the space of bounded continuous functions from $\mathbb{R}$ to itself. Let $a > 0$ be a constant. For $v \in \mathcal{C}$, we define the operator T by $(Tv)(\hat{g}) = \sup_p f(p,\hat{g})$, where

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

where $\phi(\cdot)$ is the probability density function of $\mathcal{N}(0,\hat{\sigma}_d^2)$. If v is a fixed point of T, then $v^m(\hat{g}) = (a + e^{\frac{\rho\hat{g}}{1-\rho}})v(\hat{g})$ solves (E4). We find a fixed point by demonstrating that $T : \mathcal{C} \to \mathcal{C}$ and then showing that for a sufficiently small value of a, T satisfies the Blackwell conditions and is hence a contraction mapping.

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

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

so Tv is bounded.

Demonstrating continuity is much more complicated. We first apply Lemma 12.14 of Stokey 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.[11] Given our specification of Q, this integral reduces to $\int_{-\infty}^{\infty} h(\hat{g},\zeta')\phi(\zeta')d\zeta'$, which is trivially continuous in ζ. To demonstrate continuity in $\hat{g}$, we closely follow the proof of their Lemma 9.5. Choose a sequence $\hat{g}_n$ converging to $\hat{g}$. Then $|\int_{-\infty}^{\infty} h(\hat{g}_n,\zeta')\phi(\zeta')d\zeta' - \int_{-\infty}^{\infty} h(\hat{g},\zeta')\phi(\zeta')d\zeta'| \leq \int_{-\infty}^{\infty} |h(\hat{g}_n,\zeta') - h(\hat{g},\zeta')|\phi(\zeta')d\zeta'$. Each function $\zeta' \mapsto |h(\hat{g}_n,\zeta') - h(\hat{g},\zeta')|$ converges pointwise to the zero function (by the continuity of h), so by the Lebesgue Dominated Convergence Theorem (their Theorem 7.10), this integral limits to zero. Therefore, $\hat{g} \mapsto \int_{-\infty}^{\infty} h(\hat{g},\zeta')\phi(\zeta')d\zeta'$ is continuous in $\hat{g}$, and Q has the Feller property. As a result, $f(\cdot,\cdot)$ is continuous on $(0,\infty)\times\mathbb{R}$.

The next step is to invoke our Lemma IA2. To do so, we must show uniform converge of $f(p,\hat{g})$ for $p \to 0$ and $p \to \infty$. In the first limit, $f(p,\hat{g}) \to 0$, and this convergence is uniform because terms with $\hat{g}$ multiplying the terms with p are uniformly bounded in $\hat{g}$. In the second limit, the convergence is to the integral in which $\tilde{\pi} = 0$, and the convergence is uniform for the same reason. Hence, Lemma IA2 applies, and the induced $\tilde{f}$ is continuous.

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

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

$$(1+r_m)^{-1}e^{(1-\rho)\mu_g}\sup_x \frac{ae^{\rho x+\frac{(1+\rho-\alpha\rho)^2\hat{\sigma}_d^2}{2}} + e^{\rho\mu_g+\frac{\rho x}{1-\rho}+\frac{(1-\alpha\rho)^2\hat{\sigma}_d^2}{2(1-\rho)^2}}}{a+e^{\frac{\rho x}{1-\rho}}} < 1.$$

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

$$(1+r_m)^{-1}e^{\mu_g+\frac{(1-\alpha\rho)^2\hat{\sigma}_d^2}{2(1-\rho)^2}} < 1.$$

This inequality holds because $r_m \geq r$ and we assume that (E1) holds. 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 $\mathcal{C}$, as desired. Call this function v^*. As mentioned above, $v^m(\hat{g}) = v^*(\hat{g})(a+e^{\frac{\rho\hat{g}}{1-\rho}})$ then solves (E4); this function clearly inherits the continuity of v^*.

Finally, we show that v^m is weakly increasing. Let $\mathcal{C}' \subset \mathcal{C}$ be the set of v such that $v(\hat{g})(a+e^{\frac{\rho\hat{g}}{1-\rho}})$ weakly increases. We claim that $\mathcal{C}'$ is closed. Let $\{v_n\}$ be in $\mathcal{C}'$ converging in $\mathcal{C}$ to v. For any $\hat{g}_0 < \hat{g}_1$, $v_n(\hat{g}_1)(a+e^{\frac{\rho\hat{g}_1}{1-\rho}}) - v_n(\hat{g}_0)(a+e^{\frac{\rho\hat{g}_0}{1-\rho}}) \geq 0$. Because v_n converges pointwise to v, we must have $v(\hat{g}_1)(a+e^{\frac{\rho\hat{g}_1}{1-\rho}}) - v(\hat{g}_0)(a+e^{\frac{\rho\hat{g}_0}{1-\rho}}) \geq 0$ as well. Therefore, Corollary 1 to Theorem 3.2 of Stokey et al. (1989) shows that $v^m \in \mathcal{C}'$ as long as $T : \mathcal{C}' \to \mathcal{C}'$, which is immediate from (E4).

The task remaining for this appendix is to show that each $v^s(\cdot;\lambda)$ weakly and continuously

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

increases. The argument proceeds as with $v^m(\cdot)$, but we use (E7), and we can skip the steps involving a supremum. Define the map T on $\mathcal{C}$ by

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

where $g' = (1-\rho)\mu_g + \rho\hat{g} + \rho(1-\alpha)\zeta$, and $a > 0$ is a constant to be specified later. If v is a fixed point of T, then $v^s(\hat{g};\lambda) = (a + e^{\frac{\rho\hat{g}}{1-\rho}})v(\hat{g})$ solves (E7). Clearly, Tv is bounded. To prove continuity, we again apply Lemma 12.14 of Stokey et al. (1989), this time with $X = Z = \mathbb{R}$, our $\hat{g}$ corresponding to their y, and our ζ corresponding to their z. In order to apply their lemma, we have to absorb the ζ terms into the Q transition function so that their f is bounded. Using the identity $e^{-z^2/(2\sigma^2)+bz} = e^{\sigma^2 b^2/2} e^{-(z-\sigma^2 b)^2/(2\sigma^2)}$, we have

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

and

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

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

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

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

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

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

which always holds because $\lambda \in [0,1]$ and we assume that (E1) holds. Therefore, T satisfies the Blackwell conditions and is a contraction mapping. As a result, it has a unique fixed point in $\mathcal{C}$. Call it v^{**}. Then $v^s(\hat{g};\lambda) = (a + e^{\frac{\rho\hat{g}}{1-\rho}})v^{**}(\hat{g})$ solves (E7).

Finally, we show that $v^s(\cdot;\lambda)$ weakly and continuously increases. Continuity follows from the continuity of v^{**}. As argued above, weak monotonicity holds as long as $T : \mathcal{C}' \to \mathcal{C}'$, where this set is defined as above. That T maps $\mathcal{C}'$ into itself is immediate from (E7) and the fact that v^m weakly increases. QED

F Details on Counterfactuals

F.1 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, (3) implies:

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

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

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

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

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

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

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

Mover optimality therefore requires that

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

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

We now solve for the optimal potential buyer decision, which determines the true pricing function. For $j \geq 1$, 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, $V^m = \tilde{p}e^{\tilde{d}_t}$. (In fact, even in the exact equilibrium, the mover value coincides with the price because movers are indifferent between selling and not.) The remainder of the derivation follows the proof of Lemma 4 closely, so we omit it. That is, there exist functions $\kappa_j(\hat{g}_t)$ such that a potential buyer purchases a house if and only if $e^{\delta} \geq \kappa_j(\hat{g}_t)P_t$. The functions no longer depend on n because the private flow utility δ 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

$$I_t = NI_t \left(1 - \sum_{j=1}^{J}(\beta_{0,j} + \beta_{1,j})F(\log \kappa_j(\hat{g}_t) + \log P_t - d_t)\right),$$

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 $\kappa_j(\cdot)$ functions at these values, interpolating/extrapolating in between and beyond the mesh.

To maintain comparability with the main model, we decrease γ to 0.042 so that the price overshoot is the same in the Walrasian model as in the main model, and we update $\overline{\kappa}$ 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 γ leads to more stable price paths as in the baseline model. Other parameters remain the same.

F.2 Comparing Short-term and Non-Occupant Buyers

To study the role of short-term buyers, we re-run the simulations setting $\beta_{n,j} = 0$ for all values of j except that for which $\lambda_j = 0.03$. Unlike the counterfactual in Section 6.5.3, we keep a positive mass of non-occupant potential 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 λ distribution.

We perform a similar pair of counterfactual exercises to measure the effect of removing non-occupant buyers. The first counterfactual sets the non-occupant shares, $\beta_{0,j}$, to zero, and then scales up the occupant shares, $\beta_{1,j}$, so that they sum to one. This method skews the λ distribution toward long-term potential buyers because occupants have longer horizons than non-occupants. Therefore, we explore a second counterfactual in which we maintain the original λ distribution while eliminating non-occupants. We continue to set each $\beta_{0,j}$ to zero, but now we update $\beta_{1,j}$ to the baseline share of all potential buyers for whom $\lambda = \lambda_j$.

Table IA10 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.[12] Therefore, eliminating short-term buyers prevents the model from matching key aggregate facts (Figures 1 and 3).[13]

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

One concern is that the occupant premium, μ_1, is about 7 times smaller than the standard deviation of flow utility, σ_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 IA10 under the larger values of $\mu_1 = 0.033$ and $\mu_1 = 0.066$, corresponding to 50% and 100% of the baseline value of σ_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 λ 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. (2020), 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.

[12]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 11). Thus, the baseline does a better job matching listing behavior in the boom and quiet than in the bust.

[13]The occupant adjustment does not affect the cycle because agents in the model correctly understand the distribution of housing utility, meaning that changing the housing utility distribution does not destabilize prices.

References

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

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

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

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

Bayer, Patrick, Geissler, Christopher, Mangum, Kyle and Roberts, James W. (2020). 'Speculators and Middlemen: The Strategy and Performance of Investors in the Housing Market', *Review of Financial Studies* 33(11), 5212–5247.

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

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

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

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

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

Daniel, Kent, Hirshleifer, David and Subrahmanyam, Avanidhar. (1998). 'Investor Psychology and Security Market Under- and Overreactions', *Journal of Finance* 53(6), 1839–1885.

Daniel, Kent, Hirshleifer, David and Subrahmanyam, Avanidhar. (2001). 'Overconfidence, Arbitrage, and Equilibrium Asset Pricing', *Journal of Finance* 56(3), 921–965.

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

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

Gao, Zhenyu, Sockin, Michael and Xiong, Wei. (2020). 'Economic Consequences of Housing Speculation', *Review of Financial Studies* 33(11), 5248–5287.

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

Head, Allen, Lloyd-Ellis, Huw and Sun, Hongfei. (2014). 'Search, Liquidity, and the Dynamics of House Prices and Construction', *American Economic Review* 104(4), 1172–1210.

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

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

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

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

Ngai, L. Rachel and Sheedy, Kevin D. (2020). 'The Decision to Move House and Aggregate Housing-Market Dynamics', *Journal of the European Economic Association* 18(5), 2487–2531.

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

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

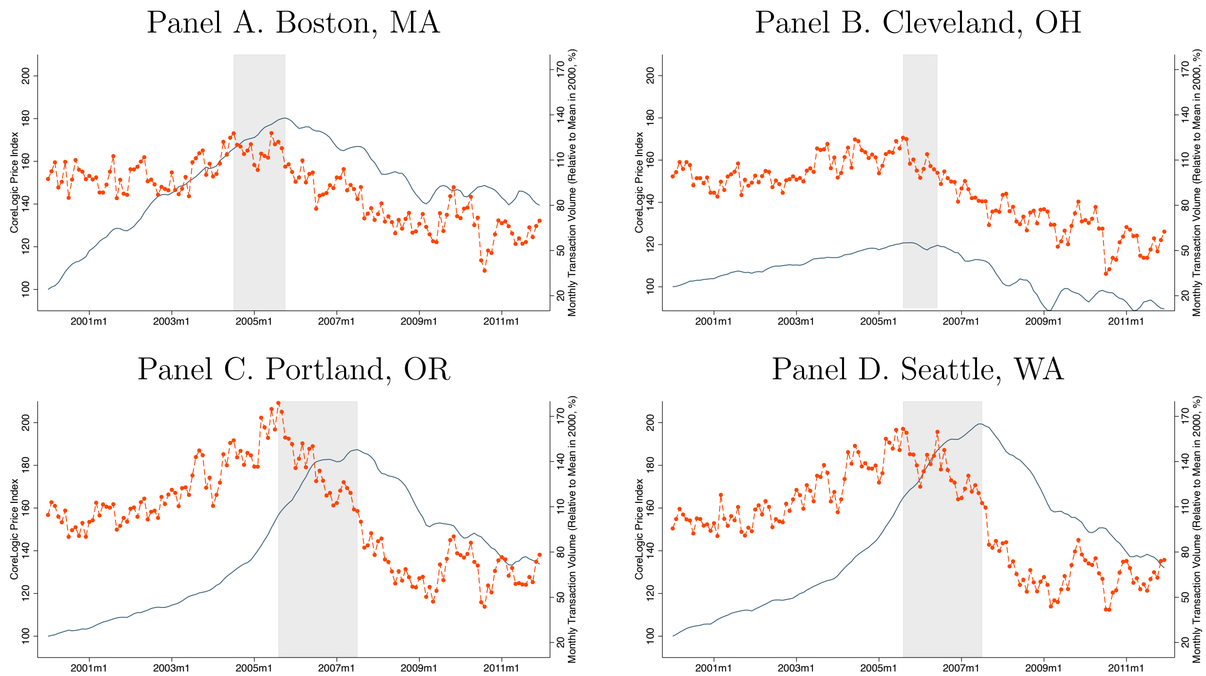
Scheinkman, Jose A. and Xiong, Wei. (2003). 'Overconfidence and Speculative Bubbles', *Journal of Political Economy* 111(6), 1183–1220.

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

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

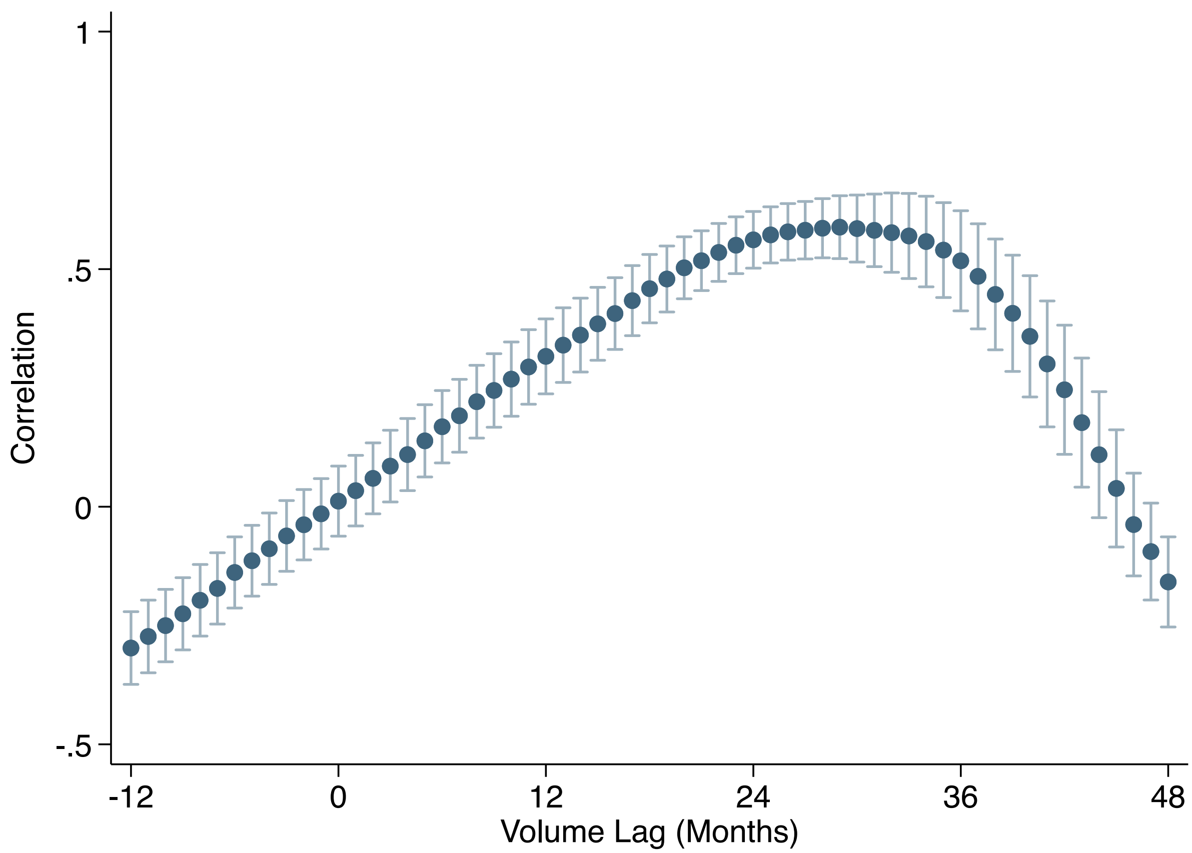
Wheaton, William C. (1990). 'Vacancy, Search, and Prices in a Housing Market Matching Model', *Journal of Political Economy* 98(6), 1270–1292.

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



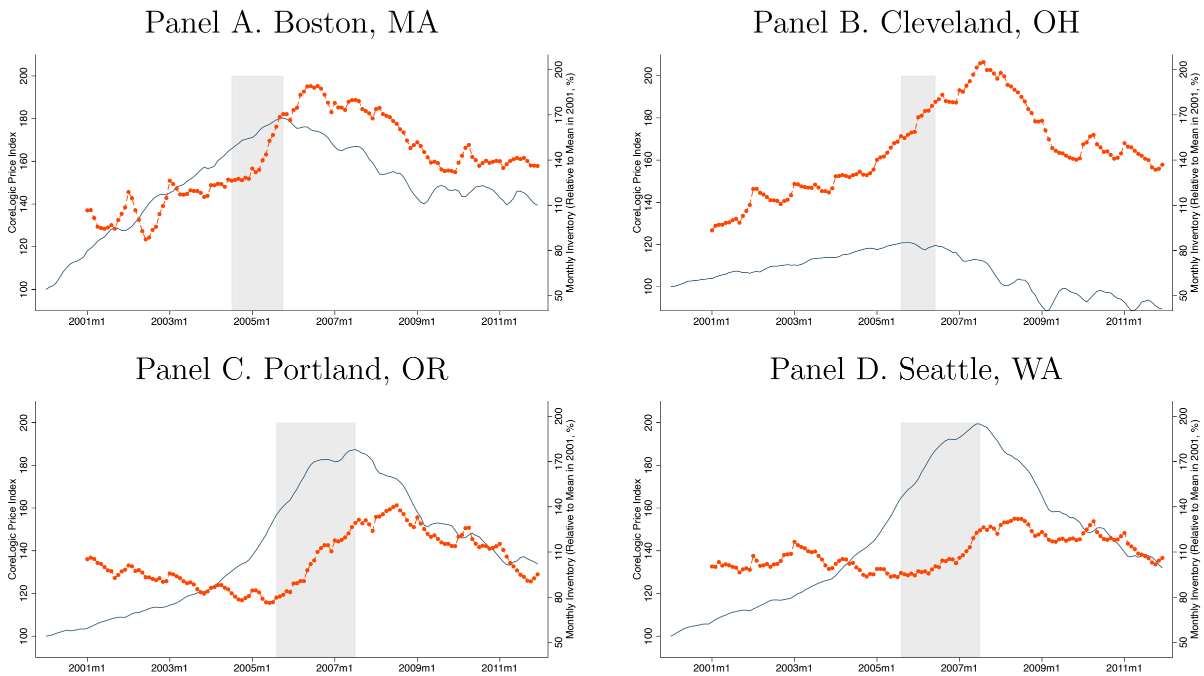
Notes: This figure displays the dynamic relation between prices (solid blue) and volume (dotted orange) in the U.S. housing market between 2000 and 2011. In Figure 1, we focus on cities that represent the largest boom–bust cycles. Here, we focus on the largest cities outside of the sand states for which 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.

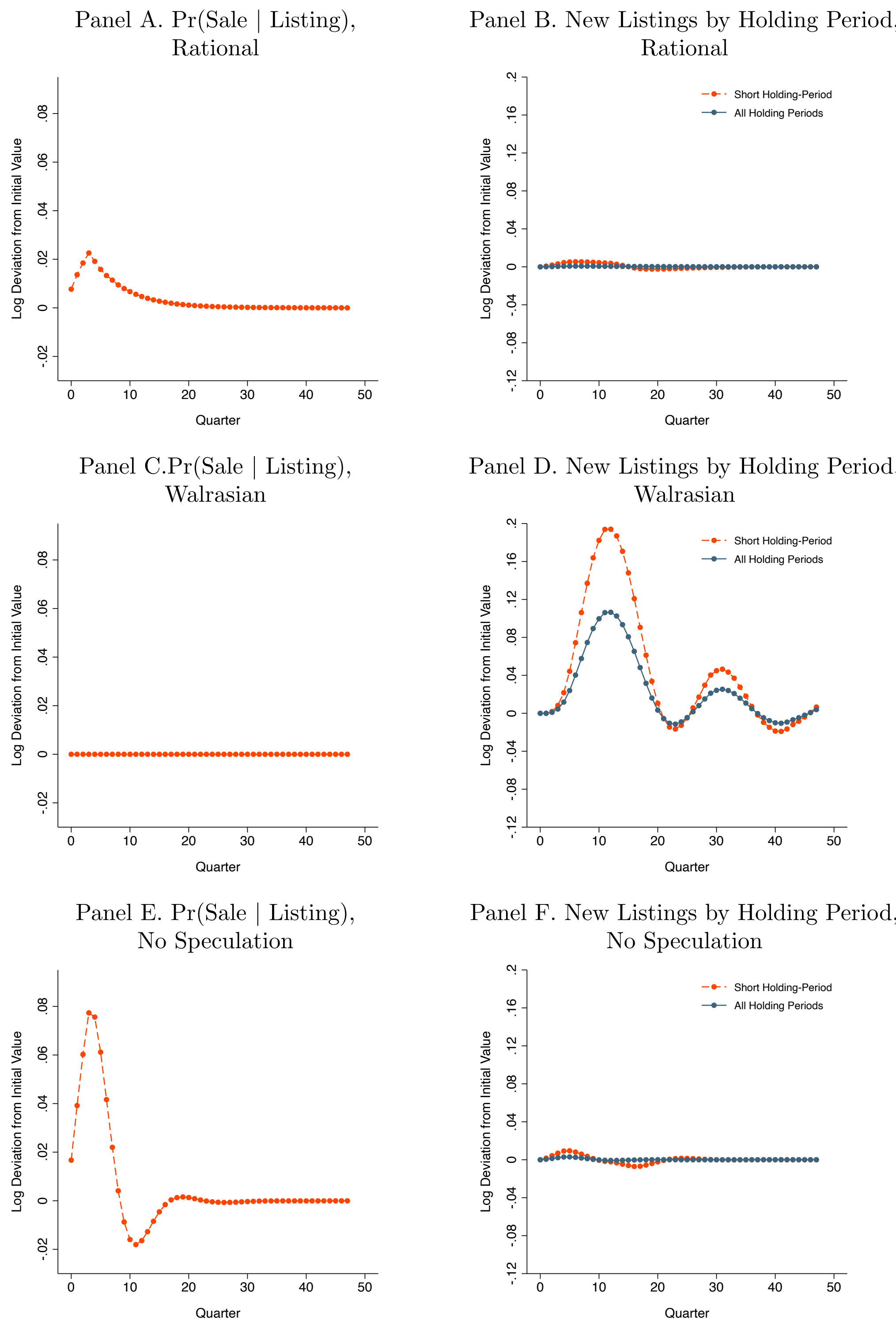
FIGURE IA3
The Dynamics of Prices and Inventories (Non-Sand-State Cities)



Notes: This figure displays the dynamic relation between prices (solid blue) and inventory (dotted orange) 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

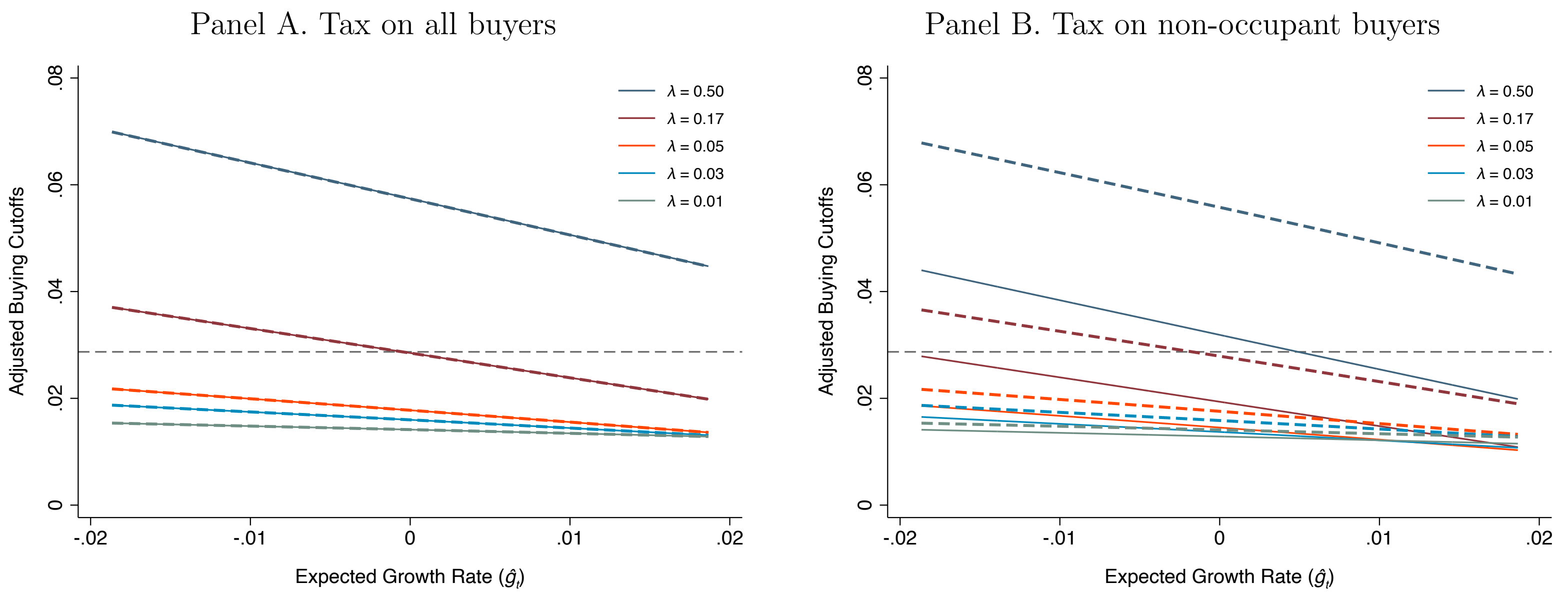
Additional Impulse Responses in Counterfactuals



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

FIGURE IA5

Adjusted Buying Cutoffs for Different Expected Growth Rates



Notes: The adjusted buying cutoff for occupancy type n and horizon type λ_j is $\overline{\kappa}\kappa^{\tau}_{n,j}(\hat{g})/\overline{\kappa}^{\tau}$, where $\tau = (\tau_0, \tau_1)$ is the vector of tax rates. In Panel A, we explore a 5% on all buyers, so that $\tau = (0.05, 0.05)$. In Panel B, we explore a tax that binds only on non-occupants, so that $\tau = (0.05, 0)$. Solid lines correspond to occupants ($n = 1$); dashed lines correspond to non-occupants ($n = 0$). The horizontal grey dashed line gives $\overline{\kappa}$.

TABLE IA1
List of Metropolitan Statistical Areas Included in the Analysis Sample

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

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

TABLE IA2
Number of Transactions Dropped During Sample Selection

Original number of Transactions	57,668,026
Dropped: Non-unique CoreLogic ID	50
Dropped: Non-positive price	3,309,100
Dropped: Duplicate transaction	618,129
Dropped: Subdivision sale	1,321,261
Dropped: Vacant lot	839,078
Final Number of Transactions	51,580,408

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

TABLE IA3

Mechanical Short-Term Volume Estimates

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

Notes: Total Volume gives annual transaction counts in our analysis sample. Actual Short-Term Volume are sales of properties for which the previous purchased occurred less than 36 months in the past. We estimate α_y^{buy}, a fixed effect for the propensity to sell a house having bought it in year y, using the regression equation in Online Appendix B.1. In the counterfactual, we assume that α_y^{buy} remains constant at its level in $y = 2000$ for $y \in \{2001, 2002, 2003, 2004, 2005\}$.

TABLE IA4
Instrumental Variables Estimation of the Role of Short-Term Volume

		Volume Boom		Price Boom		Price Bust	
	First Stage	OLS	IV	OLS	IV	OLS	IV
Short-Volume Boom		2.28***	2.28***	2.18***	2.78***	-0.77***	-1.05***
		(0.12)	(0.18)	(0.38)	(0.57)	(0.09)	(0.13)
Old Share	1.69***						
	(0.26)						
Young Share	0.66**						
	(0.32)						
Number of Observations	102	102	102	102	102	102	102
R-squared	0.45	0.79	0.79	0.25	0.23	0.45	0.39
F-Statistic	39.95						

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)

	House Price Appreciation Rate		
Lagged Price Appreciation	0.375***	0.387***	0.372***
	(0.026)	(0.027)	(0.026)
Lagged Short-Buyer Share	0.021***		0.023***
	(0.005)		(0.005)
Lagged Non-Occupant Share		0.009	0.006
		(0.008)	(0.006)
	Short-Buyer Share		
Lagged Price Appreciation	0.163***		0.162***
	(0.048)		(0.048)
Lagged Short-Buyer Share	0.780***		0.781***
	(0.024)		(0.023)
Lagged Non-Occupant Share			0.001
			0.017
	Non-Occupant Share		
Lagged Price Appreciation		0.124***	0.172***
		(0.044)	(0.045)
Lagged Short-Buyer Share			-0.071***
			(0.016)
Lagged Non-Occupant Share		0.892***	0.900***
		(0.025)	(0.021)

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

TABLE IA6

Speculators and Housing Market Outcomes (Extra Listing Outcomes)

Panel A. Propensity to List

	Δ New Listings Boom		Δ New Listings Quiet	
Short-Volume Boom	0.270		0.649***	
	(0.182)		(0.160)	
Non-Occupant Volume Boom		0.115		0.308***
		(0.092)		(0.080)
Number of Observations	57	48	57	48
R-squared	0.038	0.033	0.229	0.243

Panel B. Sale Probability

	Δ P(Sale) Boom		Δ P(Sale) Quiet	
Short-Volume Boom	0.142***		-0.163***	
	(0.032)		(0.031)	
Non-Occupant Volume Boom		0.058***		-0.047**
		(0.017)		(0.018)
Number of Observations	57	48	57	48
R-squared	0.268	0.206	0.332	0.122

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 IA7

Speculative Booms and Housing Market Outcomes (Sand State Control)

Panel A. MSA-Level Prices

	Price Boom		Price Bust	
Short-Volume Boom	1.022***		-0.237***	
	(0.272)		(0.061)	
Non-Occupant Volume Boom		0.228		-0.044
		(0.142)		(0.032)
Number of Observations	115	102	115	102
R-squared	0.514	0.453	0.696	0.662

Panel B. MSA-Level Inventories

	Δ Listings Boom		Δ Listings Quiet	
Short-Volume Boom	-1.581		4.276***	
	(1.163)		(1.461)	
Non-Occupant Volume Boom		-0.206		1.930***
		(0.525)		(0.642)
Number of Observations	57	48	57	48
R-squared	0.034	0.020	0.337	0.440

Panel C. MSA-Level Volume Quiet and Bust

	Δ Volume Quiet + Bust		Foreclosures Bust	
Short-Volume Boom	-1.145***		-0.233	
	(0.105)		(0.377)	
Non-Occupant Volume Boom		-0.516***		-0.451**
		(0.053)		(0.185)
Number of Observations	115	102	115	102
R-squared	0.533	0.505	0.317	0.333

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 IA8

Speculators and Housing Market Outcomes (Extra Listing Outcomes, Sand State Control)

Panel A. Propensity to List

	Δ New Listings Boom		Δ New Listings Quiet	
Short-Volume Boom	0.050		0.431**	
	(0.198)		(0.171)	
Non-Occupant Volume Boom		0.040		0.228***
		(0.087)		(0.072)
Number of Observations	57	48	57	48
R-squared	0.131	0.213	0.323	0.451

Panel B. Sale Probability

	Δ P(Sale) Boom		Δ P(Sale) Quiet	
Short-Volume Boom	0.146***		-0.086***	
	(0.036)		(0.028)	
Non-Occupant Volume Boom		0.058***		-0.021
		(0.018)		(0.013)
Number of Observations	57	48	57	48
R-squared	0.268	0.206	0.598	0.607

Notes: This table reports estimates of the relation between speculative volume and housing cycle measures at the MSA level. The table follows Table IA6 while adding a control for "Sand States," which is an indicator for MSAs in Arizona, California, Florida, and Nevada.

TABLE IA9
All-Cash Buyer Shares and Mean LTV by Buyer Type

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

Notes: This table presents statistics on LTV ratios and the share of buyers of various types who purchased their homes without the use of a mortgage. In column 1, statistics are measured at the transaction level and includes all transactions recorded between January 2000 and December 2011 from the CoreLogic deeds records described in Section 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 IA10
Model counterfactuals

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

Notes: We report 100 times changes in log outcomes between treatment and control simulations. See notes to Table 6 for outcome definitions. A two-sided minimum for prices 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 in order to measure the price bust. The counterfactals involve different values of the underlying distribution of potential buyers, $\beta_{n,j}$, that the text describes. We alter κ in each counterfactual to maintain a zero demand error while keeping other parameters the same. The baseline values correspond to Figure 11.

TABLE IA11

Robustness to larger occupant premium (μ_1)

Outcome	$\mu_1 = 0.033$			$\mu_1 = 0.066$		
	Baseline	All long-term buyers	All occupants	Baseline	All long-term buyers	All occupants
Price boom	14.0	8.7	14.6	12.8	8.6	14.6
Price bust	−7.6	−0.4	−8.3	−5.6	−0.3	−8.3
Volume boom	5.6	2.9	5.8	4.9	2.9	5.8
Listings, end of boom	−1.2	−3.1	−1.3	−1.0	−3.0	−1.3
Listings, end of quiet	1.3	0.4	1.4	0.9	0.3	1.4
Short volume boom	13.8	3.4	3.4	12.5	3.4	14.1
Non-occupant volume boom	11.3	5.6	–	7.5	9.1	–
Sale probability boom	6.8	6.0	7.1	6.0	5.9	7.1

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 μ_1, 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 $\beta_{n,j}$ distribution to put all weight on $\lambda = 0.03$ while keeping the occupancy distribution unchanged, corresponding to the third column of results in Table IA10. In the All occupants column, we further change the $\beta_{n,j}$ distribution to put all weight on $n = 1$ while keeping the λ distribution unchanged, corresponding to the fifth column of results in Table IA10.