The Home Sales Volatility Puzzle: An Empirical Exploration

Joshua Abel

February 12, 2019

Abstract

The recent housing cycle in the United States saw a large swing not only in home prices but in the number of home sales as well. This uses a comprehensive dataset on US home sales to investigate two popular explanations for the cyclicality of selling activity: "house lock," which conjectures that falling prices cause down-payment constraints to bind and prevent current homeowners from selling their homes; and nominal loss aversion, which proposes that cognitive frictions prevent homeowners from selling when doing so would not garner a price as high as the one they originally paid for the house. I find that while there is evidence that both of these mechanisms are active at the household level, they explain a fairly small portion of the decline in sales from boom to bust: likely no more than 10%.

1 Introduction

The surge and collapse in home prices in the recent housing cycle coincided with a dramatic rise and fall in the number of homes sold.[1] This dramatic swing—along with the financial panic and global recession it ignited—has renewed interest for researchers and housing market participants into why the volume of homes sold is so volatile and, in particular, why it seems to be so strongly associated with home prices. In this paper, I use a comprehensive dataset on US home sales to empirically evaluate the contributions of down-payment constraints and nominal loss aversion to these gyrations in home sales. I find that while household behavior is consistent with these mechanisms, their ability to explain the movements in the aggregate time series is extremely limited. Succinctly put, the movements in home sales are driven by homeowners who are not up against down-payments constraints and who are not facing nominal losses.

Why study this phenomenon? The surge and collapse in home sales over the past 25 years coincided with a dramatic rise and fall in home prices and housing investment. These developments precipitated the global financial crisis in the mid-late 2000s and led to a devastating recession, the ramifications of which we still feel. In a direct sense, the decline in prices crippled financial institutions and the implosion of the construction sector wiped out builders and left millions of them and their employees without work—yet the direct relevance of sales volume is less clear. But the fact that turnover of the existing housing stock moved nearly in lockstep with these other series, as shown in Figure 1, suggests it is a fruitful dimension to explore in order to understand the frenzy in the housing market in the boom and the trepidation that pervaded it in the bust. A complete story of this important housing cycle must account for the swings in home sales, not just movements in home prices and the size of the housing stock.

Understanding movements in home sales provides an interesting challenge for economists. Our models, even the simplest supply-and-demand diagrams taught to students studying the field for the first time, make stark and compelling predictions about prices and quantities. Yet in many foundational models, trading volume is indeterminate, as buying 1 share of an asset is equivalent to buying 2 and selling 1. In these models, trading volume is indeterminate because it is irrelevant: agents do not care how much paper they shuffled, they only care about their final positions. The housing market is an interesting setting to think about sales volume because this irrelevance argument does not hold: a household that sells a home and buys another will end up with the same number of houses as before, but changing their place of residence likely has a great impact on their consumption of housing services and/or local amenities. Far from shuffling paper, most households probably consider the decision of whether to move one of great importance and consideration in their lives.

Two popular explanations in the previous literature for why falling house prices coincide with declining housing market turnover are house lock and nominal loss aversion. House lock attributes

[1] See Figure 1.

the decline in sales during times of declining home prices to the fact that existing homeowners lose their home equity and so face frictions when trying to sell. In a similar way, if households are averse to realizing nominal losses, then a price decline may decrease the volume of home sales because homeowners do not want to realize a loss on their investments. I will confirm that these descriptions of household decision-making are borne out in the data, but I will also show that they are able to explain only a very small portion of the aggregate decline in home sales from boom to bust.

House lock was first proposed as a determinant of home sales by Stein (1995). Genesove and Mayer (1997) then used data on listings of condos in the Boston area to show that home sellers with high LTVs set higher list prices, wait longer to sell their homes, and receive higher prices. Anenberg (2011), in a more recent study of the San Francisco area, also finds that high LTVs are associated with higher sale prices. The evidence on the effect on sales has been more mixed. Ferreira et al. (2010) and Ferreira et al. (2012) use the American Housing Survey (AHS) to argue that mobility is reduced by negative equity, though Schulhofer-Wohl (2012) casts doubt on these findings. Engelhardt (2003) has trouble finding an effect on mobility in the National Longitudinal Survey of Youth (NLSY). More recent work focusing on home sales and using approaches similar to what I will use below do find an effect in Florida (Andersson and Mayock (2014)) and the United Kingdom (Bracke and Tenreyro (2018)).

The seminal paper on nominal loss aversion in the housing market was Genesove and Mayer (2001), which found that home sellers facing nominal losses set higher list prices, though the evidence on whether they received higher sale prices or waited longer to sell their homes was mixed. Since then, Anenberg (2011) confirmed in his San Francisco sample that nominal losses do seem to lead to higher sale prices, and Engelhardt (2003) and Bracke and Tenreyro (2018) have argued that nominal losses reduce measures of turnover (mobility in the NLSY in the former, sales in the UK in the latter). I contribute to this literature by using a dataset that comprises a large share of the universe of US home sales (and corresponding mortgage information) during a period of tremendous interest in the housing market. The sheer size of the dataset allows me to estimate models that are quite a bit more flexible than what is standard in the literature. Most importantly, I am able to use my estimates to make credible statements about the impact of these mechanisms on the aggregate time series of home sales.

Specifically, I empirically evaluate the quantitative power of house lock and nominal loss aversion. While a household's hazard of sale does depend on its loan-to-value ratio (LTV), as the theory predicts, the housing bust generated only a relatively modest movement in the LTV distribution. Sales fell not because of this movement but because of a downward shift in the hazard function across the LTV support. In total, house lock can account for only about 2% of the decline in sales. A similar approach suggests that nominal loss aversion cannot account for more than 8%. I go on to show that while these factors may cause homeowners to leave their homes on the market somewhat longer once they are listed, as has been emphasized in the previous literature, the speed of sale is a quantitatively trivial determinant of sales volume. So, while it is quite common to see these mechanisms used to explain the decline in sales, the vast majority of the decline is accounted for by homeowners who do not appear to have been facing equity constraints and who had not suffered nominal losses on their housing investments.

The paper proceeds as follows. Section 2 provides a model of house lock and nominal loss aversion and their impact on households' incentives to sell. In Section 3, I describe my main data source. Section 4 uses that data to evaluate the models of house lock and nominal loss aversion at both the household and aggregate levels. I conclude in Section 5.

2 Model of House Lock and Nominal Loss Aversion

This section presents a simple model of housing transactions that highlights the role of mortgage debt and nominal losses in determining transactions volume. The model is an extended version of the one found in Stein (1995). Households have some degree of mismatch to their current homes, generating an incentive to move. Importantly, the cost of moving must be financed out of the equity from the household's initial home, meaning that households with high LTVs will be forced to down-size if they choose to move. As a result, high LTVs can deter home sales. In addition, I allow for the possibility of default, so that borrowers with *very high* LTVs may in fact have heightened incentive to move, defaulting on their cumbersome mortgage and allowing a lender to sell their home. Finally, I will assume that households have some reference price against which they evaluate the sale price of their initial home. We can interpret the reference price as the price at which they *bought* the home initially, allowing for nominal loss aversion to play a role in determining sales volume.

Now, consider a household that owns a house of size 1. The household paid price p_0 for it and currently has a mortgage balance of L. Assume that the household lives for one more period, and the key decision it will have to make is whether to move and, if it does so, how large its new home will be. The timing of the model is as follows:

1. The household learns its mismatch to its current home, θ, and its income level, Y, which it will receive at the end of the period;
2. The household decides whether to move and how much housing, h, to purchase if it does;
3. If it moves, the household sells the current house at price p, pays off the mortgage, and then buys the new home at price $p \cdot h$;
4. The household receives income, consumes (net of any remaining debt), and lives in its final house.

Utility is given by:

$$U = f + v(h) + I_{MOVE} \cdot \left(\theta + \alpha_G (p - p_0)^+ - \alpha_L (p - p_0)^-\right). \tag{1}$$

The household gets utility from consuming the numeraire, food (f), and living in a larger house (h), where $v(\cdot)$ is a concave function. I assume the above quasilinear form for utility to abstract from wealth effects. θ is a reduced-form measure of the household's mismatch to its current home. Therefore, if the household moves ($I_{MOVE} = 1$), it will gain θ. The final terms parameterize the idea of nominal loss aversion, where it is typically assumed that $\alpha_L > \alpha_G \geq 0$. This captures the possibility that the initial purchase price, which rational models predict will not affect agents' decisions, will affect the household's valuation of a transaction. $(p - p_0)^+$ is the positive piece of the difference between the current price and the purchase price (0 if negative), while $(p - p_0)^-$ is the negative piece (0 if positive).

In a model with no equity constraints and no default option, the single constraint on the household would be:

$$Y + p - L = f + ph. \tag{2}$$

This says that net worth (income plus the value of its house, less debt) is equal to spending. However, as in Stein (1995), I will add the key friction that a household must pay for the move with money raised from selling the initial home.[2] This adds an additional constraint:

$$h \leq \frac{p - L}{p}. \tag{3}$$

This generates the possibility of "house lock": a sufficiently indebted household (L high) may be constrained by its current level of home equity and be unable to purchase its desired house.

To allow for mortgage default, assume the household has the choice to live in a home of size 0 and not pay back its debt. Therefore, the constraint 2 is replaced by 2':

$$Y + p - L \cdot (1 - I_{Default}) = f + ph, \tag{2'}$$

along with the requirement that defaulting households live in a home of size 0:

$$h = 0 \text{ if } I_{Default} = 1. \tag{4}$$

To solve the model, begin by defining h^* to be the level of housing a household would choose if

[2]For simplicity of notation, I will assume that the new home is purchased in cash - with no mortgage. One could allow the households to buy a new home while only putting down some fraction of its value, but this has no effect on the results.

unconstrained by their home equity, conditional on moving.[3] It is characterized by equality of the marginal rate of substitution between housing and food and the ratio of their prices:

$$v'(h^*) = p. \tag{5}$$

The moving household's choice can then be written as:

$$h = \max\left\{\min\left\{h^*, \frac{p-L}{p}\right\}, 0\right\}. \tag{6}$$

The inner minimum function incorporates the requirement that current home equity be sufficient to pay for the new house, while the outer maximum function incorporates the default option, which says that a household can walk away from its current home and cannot be forced to pay off its debt. In that case, it will set $h = 0$ and $f = Y$. As a result, utility from moving can be expressed as follows:

$$U_{\text{M}} = \begin{cases} Y - L - p(h^* - 1) + v(h^*) + \theta + \alpha_G(p - p_0)^+ - \alpha_L(p - p_0)^- & \text{if } L \le p(1 - h^*) \\ Y + v(\frac{p-L}{p}) + \theta + \alpha_G(p - p_0)^+ - \alpha_L(p - p_0)^- & \text{if } p(1 - h^*) < L \le p \\ Y + v(0) + \theta + \alpha_G(p - p_0)^+ - \alpha_L(p - p_0)^- & \text{if } L > p. \end{cases} \tag{7}$$

Utility from not moving is simply:

$$U_{\text{N}} = Y - L + v(1). \tag{8}$$

Define $\Delta(L, p, Y, \theta, p_0)$ to be the difference between utility when moving and utility when not moving. First, consider the effect of L on the household's incentive to move:

$$\frac{\partial \Delta}{\partial L} \begin{cases} = 0 \text{ if } LTV < (1 - h^*) \\ < 0 \text{ if } LTV \in \big[(1 - h^*), 1\big] \\ > 0 \text{ if } LTV > 1 \end{cases}, \tag{9}$$

where $LTV \equiv L/p$. Intuitively, when LTV is low and the equity constraint does not bind, tightening it with a little more debt does not affect the household's incentive to move: every additional dollar of debt is a dollar that is not consumed, regardless of whether the household moves or not.[4] With an intermediate level of debt, when the down-payment constraint binds, the

[3]This concept is independent of loss aversion because, unlike mortgage debt, the reference point only affects the decision of whether to move (as we will see), not the size of the home conditional on moving.

[4]This is the one place where the assumption of quasilinear utility is substantive. With wealth effects, the additional

household is being forced to under-consume housing. The more that constraint tightens (i.e. LTV increases), the further away from the optimum h^* a moving household will be, so the incentive to move decreases. Effectively, these households are being forced to down-size if they move, and so only those with relatively high mismatch realizations θ will find it worthwhile to do so. Finally, when the household's debt is so high that it is underwater ($LTV > 1$), it cannot afford to buy a house larger than $h = 0$ if it moves. Therefore, a marginal increase in L does not affect house size. Instead, every additional dollar of debt is a dollar that must be paid if the household stays in their current home but that the household can walk away from by defaulting and allowing the lender to sell the foreclosed home. As a result, sales become more likely when debt increases in this region—though they are distressed sales by lenders.

If the distribution of θ is similar across LTV levels, this simple model predicts the following:

1. The probability of a home selling is flat at low levels of LTV;
2. For some $\tilde{LTV} < 1$, the probability of a home selling is a decreasing function of LTV when $LTV \in [\tilde{LTV}, 1]$;
3. The probability of a home selling is an increasing function of LTV when $LTV > 1$.

Turning to nominal loss aversion, the effect of p_0 is quite straightforward. Note that unlike L, p_0 does not affect the household's decision of home choice, conditional on moving. It does, however, affect whether the household decides to move in the first place. In particular, if $p < p_0$, moving requires the household to realize a nominal loss. As this has a negative effect on utility, a higher θ is required to induce the household to transact, and so the probability of sale falls, as the household will require a higher mismatch (θ) to be induced to move. Therefore, we expect the probability of sale to be an increasing function of the difference between p and p_0. Note a subtlety here, which is that this may only apply to non-distressed sales, where the homeowner is actually receiving the proceeds of the sale. When the household defaults, it does not receive any payment for the home and may in fact have moved well before the sale occurs. As a result, I will focus on non-distressed sales in the empirical tests of this portion of the model.

3 Deeds Records

My primary dataset comprises property-level public records on home sales and mortgages. This data is provided by CoreLogic, which scraped electronic records from county registers of deeds from across the country. The data covers the years 2000-14, and its geographic coverage is strong and improves over the sample period. In 2000, about 50% of U.S. counties are included, covering over 85% of the U.S. population. By 2004, counties with around 90% of the population are covered, and by the end of the sample, counties including over 99% of the population are included.

dollar of consumption would be valued differently depending on whether the household moves, since the mixture of f and h depends on that decision. As a result, marginal increases in debt could affect the household's incentive to move.

When a home is sold, the address of the property being transferred is observed, as well as the names of (up to two) buyers and sellers involved in the transaction, the price and date of the transaction, and the terms of any mortgage taken out by the buyer. In addition to transactions, the data records mortgage refinances, when mortgage terms change but there is no transfer of ownership.

3.1 Primary Estimation Sample

My main empirical results come from a property-by-month panel constructed from a 1% random sample of the dataset described above. Specifically, I randomly select 1% of properties in the CoreLogic dataset. Next, I drop sales that are classified by CoreLogic as nominal (or "non-arm's-length") transactions, which is about 1/3 of all sale records. I then generate an observation for each property for each month between January 2000 and December 2014. From this panel of properties, I generate "ownership spells," which begin with the sale of a property and end with its subsequent sale. Foreclosures require some further care. When a lender seizes a house from a homeowner, a record is generated in the deeds data, but I do not count this as a sale. Rather, the ownership spell ends when the lender (or a subsequent lender to whom the property has been transferred) makes an arm's-length sale of the property.

One final sample selection criterion is important to discuss. In order to compute a homeowner's LTV, I need to know the value of the house, which requires knowing the previous sale price. This means that, while the dataset contains nearly the universe of transactions, the panel used in estimation will not contain the universe of properties. For instance, if a property sells for the first time in, say, June 2004, it cannot be included in the sample until July 2004 because its LTV in the months January 2000-June 2004 is unknown. This means the sample size will increase over time. As a result, I will not be discussing the *count* of sales but rather the *hazard* of a sale occurring. Furthermore, I begin the analysis in 2004 rather than 2000, as this allows the full sample to have a more representative set of properties,[5] both because of the issue detailed in this paragraph as well as the improved coverage of the dataset over time as discussed earlier.

Figure 2 shows that the remaining sample, while not fully representative due to the necessary sample restriction discussed above,[6] still exhibits a collapse in sales during the housing bust—and slow recovery thereafter—that is extremely similar to the aggregate data provided by the National Association of Realtors. This offers reassurance that subsequent analysis of this sample is likely to apply to the broader population of properties.

[5]For instance, in the early years of the sample, the *LTV* distribution is very tightly packed around 80% and 100%, since the sample is comprised of people who have only very recently purchased their homes and who therefore have not experienced much movement in local home prices (or amortized much of their mortgage debt). By waiting a handful of years, I allow the sample to look more like the population: as discussed below, the *LTV* distribution in my sample in later years is quite similar to the distribution calculated from a separate dataset that does not have this same sample selection issue.

[6]It is also not fully representative because, as discussed, some counties are omitted. Furthermore, there are instances when a sale occurs but a price is not listed. LTV cannot be computed in these cases, so they are also not included in the estimation.

All told, this sample has 26,498,330 property-month observations spanning 265,832 properties and 392,074 ownership spells. There are 127,334 sales in this sample,[7] generating a monthly hazard of 48bp. Figure 3 shows that, during the crash, there was a very large increase in the share of these sales that were "distressed." In 2004, over 95% percent of sales were non-distressed—sales by homeowners who were able to pay back the mortgage in full. In the height of the crisis, nearly 50% of sales were either short sales[8] or sales out of foreclosure. While this number declined throughout the recovery, it was still close to 20% in 2014, well above the historical norm.

3.1.1 Computing LTV

A homeowner's LTV is a key variable of interest in Section 4, so I now describe in detail how this measure is calculated.

The numerator is the mortgage debt being collateralized by the house. As discussed above, the deeds records provide information about mortgages taken out when a home is sold, as well as subsequent refinances. This means that at the time of purchase, the loan balance is directly observed (inclusive of "piggyback" second liens). In subsequent months when the property is not observed in the data, I amortize the loan balance according to the interest rate and mortgage term listed,[9] assuming that the payments are being made.

Some ambiguity arises when a household refinances. While the balance and terms of the new loan are listed, there is not explicit information about what has happened with the previous loan. Specifically, it is not known whether the new loan is a replacement of the prior loan or an additional loan on top of it. In the results that follow, I assume that if the new loan has a balance that is less than 25% of the outstanding balance on the old loan, then it is a cashout refinance and I add the two balances together. If it is more than 25% of the old loan's outstanding balance, I assume it is a replacement, so that the new level of total mortgage debt is just the balance of the new loan.[10]

The denominator of the LTV is the market price of the home. This is of course only observed when the home sells. To get an estimate of the LTV in months between sales, I update the sale price using local[11] home price appreciation, as measured by CoreLogic's repeat sale home price index, which is standard practice in this literature. Specifically, if home i in geography g sells in months t_0 and $t_1 > t_0$, I compute the market price in month $t \in [t_0, t_1]$ as:

[7] The number of sales is nearly equal to the difference between the number of ownership spells and the number of properties because each sale begins a new ownership spell. There are 1,092 "extra" sales because this is precisely the number that occurred in December 2014, the last month of the sample. As a result, these sales do not begin new ownership spells.

[8] Short sales refer to cases where the proceeds from selling the home are insufficient to pay off the mortgage, but the lender agrees to allow the borrower to sell the house and not pay back the entire loan, rather than go through the foreclosure process.

[9] If the term is missing, I assume it is a 30-year mortgage. If the interest rate is missing, I assume it is the interest rate reported by Fannie Mae as the average contract rate on a conforming loan for that month.

[10] I have run the analysis using cutoffs of 20% and 0% and while the shape of the LTV distribution is somewhat sensitive to this, the results later in the paper are hardly affected at all.

[11] For properties in ZIP codes covered by CoreLogic's index, I use ZIP code-level appreciation. If that is not possible, I use county-level appreciation, followed by MSA and finally state.

$$\hat{P}_{i,t,g} = P_{i,t_0} \frac{HPI_{g,t}}{HPI_{g,t_0}}. \tag{10}$$

Figure 4 shows the LTV distribution in this sample for selected years (excluding the roughly 30% of properties that have no mortgage at any given time). I discuss the evolution of the LTV distribution in much greater detail below, but for now I note that it compares very favorably with Figure 5, which shows the LTV distribution computed in a separate dataset.[12] Specifically, Figure 5 comes from a 0.1% random sample of the Credit Risk Insights Servicing McDash (CRISM) dataset, which is a monthly mortgage loan performance dataset (from Lender Processing Services) with borrowers' credit records (from Equifax) merged in. While the price has to be imputed in a similar way (using local price indexes), the mortgage balance is directly observed each month, precluding any need to impute it based on the loan's terms. Furthermore, the credit bureau data lists the borrower's outstanding debt on home equity lines of credit (HELOC) and closed-end second mortgage balances.[13] That the LTV distributions look so similar when computed in these two very different datasets is important validation for the construction of this key measure.[14]

3.1.2 Computing Nominal Gains/Losses

Another important measure for my analysis is nominal gains/losses that a homeowner faces if she sells her home. I will focus only on nominal gains/losses coming from local home price appreciation. If home i in geography g sold in month t_0 and sells again in month t_1, then for $t \in [t_0, t_1]$:

$$NomGain_{i,t} = \frac{HPI_{g,t}}{HPI_{g,t_0}} - 1. \tag{11}$$

So a homeowner faces nominal losses if and only if local home prices have declined since the time of purchase. This assumes that when the household initially bought the home, the price did not include any discount or premium. Some papers studying nominal loss aversion (e.g. Genesove and Mayer (2001), Anenberg (2011)) relax this assumption and use home characteristics to estimate a "fair" market price using a hedonic model. However, given the time period I am studying, the

[12] This alternative dataset, CRISM, is not available until the middle of 2005, which is why I do not show a comparison for 2004.

[13] Note that CRISM is not perfect for this exercise either. The reason is that the second mortgage balances are given at the borrower level, so for borrowers with multiple properties, I could be including second mortgage debt in the LTV when it really is collateralized by a different property. However, this problem is quite distinct from the issues with the deeds data discussed in the text, so the corroboration of the two datasets is still quite reassuring.

[14] Two differences between the datasets are apparent. First, because CRISM is based on mortgage data, it does not contain the 30% of properties that do not have mortgages. Therefore, the densities are a bit higher in general in the CRISM data. Secondly, the deeds data shows much tighter bunching around notable cutoffs like 80%. This is because the measure of value being used in the deeds data is the price, whereas the available measure in CRISM is the appraisal. Since many borrowers take out loans for exactly 80% of the purchase price, there is sharp bunching in the deeds records. The bunching is mitigated somewhat in CRISM to the extent that appraisals and sale prices differ.

preponderance of variation in nominal gains/losses will come from market-level movements in prices, so I choose to focus on market-level variation, as in Engelhardt (2003) and Bracke and Tenreyro (2018).

Figure 6 shows the distribution of this variable in my sample for selected years. I will discuss this in greater detail below, but the reader can immediately note that, as expected, this distribution deteriorates badly as the price decline unfolds.

3.2 100% Seattle Sample

In addition to the results from the 1% national sample, I provide estimates coming from the universe of transactions in the Seattle-Tacoma-Bellevue MSA. Results from this sample mostly serve as robustness checks for the main results—showing that what's true across MSAs is also true within a particular MSA—and so will be largely confined to Appendix A.1. However, some of the analysis below will leverage Multiple Listing Service (MLS) data on home listings, which is linked to the deeds data at the property level. The coverage and quality of that data is heterogeneous across MSAs, and the data quality in Seattle is very high, so it is a sensible place to focus on. Furthermore, the size of the price boom and bust in Seattle was roughly equal to the national experience, further bolstering its appropriateness for closer inspection.

The Seattle sample has 47,518,278 property-month observations from 446,286 properties and 662,170 ownership spells. There were 217,698 sales. Figure A-1 shows the evolution of the sale hazard in the Seattle sample, which is quite similar to the national series, while Figures A-2 and A-3 show the distributions in selected years of LTV and nominal gains/losses, respectively. Comparing these with the national distributions in Figures 4 and 6, we see some of the geographic heterogeneity of the housing cycle. Because prices continued to grow in Seattle for an extra year (compared to national measures of prices), the 2009 distributions of both LTV and nominal gains/losses are fatter in both tails than in 2004, as households who bought early in the sample period got to experience an additional year of appreciation rather than depreciation.

4 The Role of House Lock and Nominal Loss Aversion

I now use the above data to empirically evaluate the contributions of house lock and nominal loss aversion to the decline in home sales from boom to bust. Following the model presented in Section 2, we expect the probability of a home selling to have a non-monotonic relationship with LTV, as a result of the competing effects of down-payment constraints and the default option, and to have a negative relationship with nominal losses. I will show that the data is consistent with these predictions, and then I will turn to evaluating the quantitative effects of the mechanisms.

4.1 House Lock

My workhorse econometric framework is a linear probability model.[15] I estimate the equation:

$$Sale_{i,t} = \mathbf{LTV}_{i,t} \cdot \gamma_1 + \mathbf{X}_{i,t} \cdot \gamma_2 + u_{i,t}, \tag{12}$$

where $Sale_{i,t}$ is an indicator function for whether property i was sold in month t; $\mathbf{LTV}_{i,t}$ is a set of K indicator functions for whether property i has an LTV in month t in one of K regions of the LTV support; $\mathbf{X}_{i,t}$ is a set of controls that varies by specification, as described below; and $u_{i,t}$ is a residual. In particular, I partition the LTV support into 24 bins,[16] with the partition being particularly fine in the range $LTV \in [0.7, 1.1]$, as this region is both well-populated throughout the sample period and likely to be the region where the down-payment and mortgage default incentives described above are quite active. This specification allows the effect of LTV to vary arbitrarily across these bins, so I can estimate a very flexible relationship. In the reported results that follow, the omitted category will be those observations with $LTV_{i,t} \in (0, 0.2]$, so all effects are relative to being in that group.

I will show results from 4 specifications, which differ in what is included in $\mathbf{X}_{i,t}$. All specifications include flexible controls for months since purchase ("duration"), calendar month (to soak out any effects of seasonality), and nominal gains/losses. The "Baseline" specification controls for only these. The specification "+ Covariates" further controls for county-level economic conditions (county unemployment rate 6 months prior, change in the county unemployment rate from 12 to 6 months prior, the gross number of jobs created in the county in that year [expressed as a percentage of employment]), ZIP code-level measures of financial health (average FICO score on purchase mortgages, average household income, and share of adults with at least a college—all measured soon before the start of the sample period), and some property level controls (initial down-payment and percentile rank of property's value within MSA). The specification "+ MSA-,year-FEs" adds in indicators for each MSA and year in the sample, while the final specification, "+ MSA-by-year FEs" adds in indicators for each MSA in each year. Standard errors are clustered at the MSA-year level.

Figure 7 shows the results of this estimation on the 1% national sample. The probability of sale is quite stable for values of LTV between 0.2 and 0.8, but it falls monotonically and dramatically between 0.8 and 1. Depending slightly on the specification, the monthly probability of a sale is about 15bp lower when $LTV = 1$ than when $LTV = 0.8$, which is about 30% of the sample average (48bp). This confirms the primary prediction of the house lock model, which is that high LTVs

[15] I have also run the analysis using a Cox proportional hazard model, and results are hardly changed. I present the results using the linear model to allow for a more exhaustive set of controls while maintaining computational tractability.

[16] The 24 bins are as follows: [0], (0,0.2], (0.2,0.4], (0.4,0.6], (0.6,0.7], (0.7,0.75], (0.75,0.8], (0.8,0.82], (0.82,0.84], (0.84,0.86], (0.86,0.88], (0.88,0.9], (0.9,0.92], (0.92,0.94], (0.94,0.96], (0.96,0.98], (0.98,1], (1,1.05], (1.05,1.1], (1.1,1.2], (1.2,1.3], (1.3,1.4], (1.4,1.6], (1.6,∞).

(less than 1) deter homeowners from selling their properties.

The results also confirm the model's secondary prediction, which is that the likelihood of sale is an *increasing* function of LTV when $LTV > 1$, as homeowners become increasingly likely to default. The estimates suggest that while this effect begins to take hold immediately at $LTV = 1$, it does not begin to outweigh the house lock effect until the LTV is quite high, around 1.4. Homes with the highest LTVs (in the range of 1.6 and above) are in fact the most likely to be sold.

In addition to confirming the household predictions of the house lock model and laying the groundwork for evaluating its quantitative implications for aggregate movements in home sales (see below), these results are consistent with a growing body of work concluding that borrowers are not inclined to default on a mortgage in the absence of a sharp cash flow shock until their LTVs are quite high. The estimates suggest that this "strategic" incentive[17] does not outweigh the other consequences of default (including being forced to move and perhaps down-size, having diminished access to credit markets, and moral qualms associated with not paying back debts as in Guiso et al. (2013)) until the mortgage debt reaches about 150% of the home's value, a number similar to that estimated by Bhutta et al. (2017).

The fact that the sale hazard depends on LTV allows for the possibility that changes in the distribution of LTV over time may be responsible for changes in aggregate sales volume. Figure 4 shows the LTV distribution's rightward shift during the housing bust, as home prices fell. Whereas very few borrowers were underwater on their mortgages in 2004—near the end of a long period of increasing home prices—by 2009 around 20% found themselves in this position. I will now assess the magnitude of the effect of this shift on overall selling activity. In particular, I will look at the decline in the sale hazard from the period 2004-6 ("boom") to the period 2007-11 ("bust").

To start, I will treat the hazard function[18] in Figure 7 as constant across time. Let s_k be the hazard coefficient of a home in bin k, and let $w_{k,t}$ be the percent of homes in period t that are in bin k. We can then compute the change in the sale hazard generated by a shift in the LTV distribution as:

$$\Delta_{t,t-1} = \sum_{k=1}^{K} s_k \cdot (w_{k,t} - w_{k,t-1}). \tag{13}$$

To the extent that the shift in the distribution moved mass from bins with high hazard rates (e.g. $LTV \in (0, 0.2]$) to bins with low hazard rates (e.g. $LTV \in (0.98, 1]$), the average hazard

[17] The distinction between "strategic" and "liquidity-based" default is a bit contrived. As Campbell and Cocco (2015) show, if a homeowner can save a sufficient amount on monthly payments by defaulting (say, if local rents are less than the mortgage payments), then there will be some threshold LTV above which she will default. Liquidity shocks can move that threshold around, so ultimately every default decision is based on both liquidity and LTV. Nonetheless, despite this imprecision, the term "strategic default" is useful shorthand for describing the role that LTV has in determining households' incentives to default.

[18] Specifically, the hazard function corresponding to the fourth specification, which includes MSA-by-year fixed effects.

rate will decline. This exercise is depicted visually in Figure 8, where I show the hazard function plotted against the difference in LTV distributions between the two periods.

The headline result is that $\Delta_{Bust,Boom}$ = -0.5bp. This is compared to the actual decline in the hazard rate of 32.5bp, meaning this channel is able to account for about 1.5% of the decline in sales. Figure A-4 shows that the story is very similar if one focuses exclusively on a single MSA, as the house lock effect can account for 3.3% of the decline in sales in the Seattle area.

Why is the effect so small? One might be tempted to attribute it to the offsetting impact of two competing effects: yes, sales declined due to the lock-in effect of the down-payment constraint as borrowers shifted into the high-LTV region, but this was offset by the increase in sales coming from defaulting households who had been pushed into the *very*-high-LTV region. While this is true, it is only a small part of the story. If we shut down the default channel by assuming the hazard function is constant for $LTV > 1$, where that constant is the depressed level for the group with $LTV \in (0.98, 1]$, we would still only get a decline of 2.0bp, or just 6.2% of the actual decline.

To get a better sense of why the house lock story is quantitatively insufficient to explain the drop in sales, consider what would happen if all homeowners had been in the reference bin ($LTV \in (0, 0.2]$) in the boom and had all moved to the $LTV \in (0.98, 1]$ bin in the bust, so that the sale hazard was as depressed as possible. This would generate a decline in the sale hazard of 20bp, about 60% of the true decline. In other words, it would have required a massive shift in the LTV distribution to generate movement in the overall sale hazard on a comparable scale to what happened to the true hazard rate. While the LTV distribution did of course deteriorate badly, it was not nearly on that scale.

Instead, there was a downward shift in the hazard function. To see this, I re-estimate the model allowing the effect of LTV to vary over time:

$$Sale_{i,t} = (\mathbf{Period}_{i,t} \times \mathbf{LTV}_{i,t}) \cdot \gamma_1 + \mathbf{X}_{i,t} \cdot \gamma_2 + u_{i,t}, \tag{14}$$

where **Period** is a set of indicators for whether the observation is occurring in the "boom" (2004-6), "bust" (2007-11), or "recovery" (2012-4). Figure 9 shows the results of this estimation, where each coefficient is interpreted as the basis point difference in the sale hazard relative to borrowers with $LTV \in (0, 0.2]$ *in the boom*.

The results are stark: the hazard function for homes with LTVs between 0 and 0.8 fell by 30-35bp. As this is where the large majority of homeowners are positioned in the distribution (about 80% in the boom, 70% in the bust), it is this decline—in a region where house lock and mortgage default are irrelevant—that is mostly responsible for the 32.5bp decline in sales. Most of the remainder comes from downward shifts in the hazard function in other parts of the support, with only a very small part (again, 1.5%) coming from shifts in the distribution along the hazard function. Figure A-5 shows that this conclusion remains true when focusing exclusively on Seattle.

Therefore, despite its success in explaining an element of household behavior, the house lock story is incapable of explaining the aggregate decline in home sales—in an accounting sense, this decline was driven by homeowners who do not face constraints from their mortgages and yet became far less likely to sell in the bust than they were in the boom. An explanation for this development must come from elsewhere.

4.2 Nominal Loss Aversion

Nominal loss aversion provides another mechanism through which falling home prices can lead to decreased sales activity. If households get disutility from realizing a loss on an investment—or more generally, if utility is an increasing function of realized gains on investments, independent of any impact on lifetime wealth or consumption—then homeowners will become less likely to sell after home prices fall because realized gains will be lower. As this mechanism only applies to homeowners and not their lenders if and when they foreclose on a home, I will focus in this subsection on non-distressed sales. Therefore, I will censor an ownership spell when a lender forecloses on the home or when it is sold via short sale. Figure 10 shows the non-distressed sale hazard in this censored sample. As expected, the decline in this measure of sales is more dramatic than the overall decline in sales, as the increase in distressed sales is not present to buffer the collapse.

I use the same empirical framework to estimate the impact of nominal loss aversion as I did in the previous subsection for house lock. I partition the support of nominal gains/losses into 8 bins: ($-\infty$,-20%], (-20%,-10%], (-10%,-5%], (-5%,0%], (0%,10%], (10%,20%], (20%,40%], (40%,∞). In the reporting of results that follows, the omitted group will be those in the lowest bin with nominal gains in the range ($-\infty$,-20%].

Figure 11 shows the results of this estimation. First note the sensitivity of the estimates to the set of controls. When no controls are included for time period, the effects appear extremely large. This is essentially a restatement of this paper's motivating fact, which is that periods of booming prices have high levels of home sales. When MSA-by-year indicators are included,[19] the effect is weakened substantially, so that homeowners experiencing nominal gains in excess of 20% are about 12bp more likely to sell than homeowners experiencing nominal losses of over 20%, compared to an average non-distressed sale probability of 38bp in the sample.

To assess the aggregate impact, I perform the same sort of calculation as in the previous subsection on house lock:

$$\tilde{\Delta}_{t,t-1} = \sum_{k=1}^{\tilde{K}} \tilde{s}_k \cdot (\tilde{w}_{k,t} - \tilde{w}_{k,t-1}), \tag{15}$$

where $\tilde{s}_k$ is the sale hazard coefficient for bin k and $\tilde{w}_{k,t}$ is the share of nominal gain/loss

[19] Further stringency in the set of controls seems to impact the results very little. For instance, when including ZIP code-by-month FEs rather than MSA-by-year FEs, the coefficients are hardly changed.

distribution in period t that is in bin k. Figure 12 shows the components of this calculation: the estimated hazard coefficients and the change in the nominal gain/loss distribution between the boom and bust.

Performing this calculation, the change in the nominal gain/loss distribution can account for a 3.5bp decline in the non-distressed sale hazard between boom and bust, or about 7.9% of the actual decline. While this is hardly trivial, it suggests that, as in the case of house lock investigated earlier, other explanations are required to explain the dramatic decline in home sales experienced during the crash.

The reason nominal loss aversion fails to account for a large share of the decline in aggregate sales is somewhat different than why house lock is not capable of doing so. Whereas, as discussed previously, only about 10% of the LTV distribution shifted from being below 0.8 to above 0.8, about 50% of the nominal gain distribution shifted from being positive to negative between boom and bust. However, the effects on household behavior are not strong enough to convert that large change in the distribution into a steep decline in sales. Even if the entire distribution where shifted from the bin with highest gains into the bin with highest losses, this would generate only about 30% of the observed decline in non-distressed sales. Given the more modest shift, the true number appears to be quite a bit smaller than that.

4.3 Time on Market and Sales Volume

I conclude this section by discussing how the time a listed home spends on the market relates to this broader discussion of sales volume. This is relevant because some prominent empirical work has focused on this part of the selling process. Seminal work in Genesove and Mayer (1997) and Genesove and Mayer (2001) used listings data to demonstrate that house lock and nominal loss aversion affect the amount of time a house spends on the market after being listed for sale. Table A-1 uses the Seattle (non-REO[20]) listings data to verify the results of those studies. In particular, I find that listed homes becomes less likely to be sold as their LTVs increase past 0.8, consistent with the house lock story. Homes that have experienced more appreciation since they were last purchased are also more likely to sell, with this effect being stronger in the region where homeowners have experienced losses than the region with gains. This is consistent with the literature's results with regards to nominal loss aversion.

While this suffices to qualitatively demonstrate that homeowners are affected by these channels when selling their homes, it tells us very little about their quantitative impact—even at the household level. The reason is that the speed at which homes turn over is not very dependent on how quickly listed homes sell. Rather, what is critical is the frequency with which they are listed for sale in the first place, a point emphasized by Ngai and Sheedy (2017).

To see why, let n be the hazard that an unlisted home is put on the market and s be the hazard

[20]This stands for "real estate owned," and refers to properties owned by lenders.

of a home selling, conditional on being listed. In steady state, then, the unconditional hazard of sale is given by:

$$V = \frac{sn}{s+n}. \tag{16}$$

This allows us to decompose variation in the unconditional sale hazard into a part coming from variation in the probability of being listed and a part coming from the probability of a listed home selling:[21]

$$dV \approx \frac{n^2}{(n+s)^2}ds + \frac{s^2}{(n+s)^2}dn. \tag{17}$$

The contribution of each hazard is scaled by the square of the other hazard. Intuitively, if n is small and s is large (as is the case), the set of listed homes will be a fairly low share of the housing stock. As a result, changes in the rate at which they sell will not have a large impact on overall sales. An increase in n, however, will get a lot more homes out on the market, and since they sell quickly, this well have large effect on overall sales. In the listings data available for Seattle, $n \approx \frac{1}{100}$ and $s \approx \frac{1}{4}$, so the variation in V comes almost entirely from movement in n, not s.[22] Figure 13 shows this explicitly—almost all of the variation in V comes from movements in n. As a result, while studying listings data can be informative about how housing markets function, ultimately time-on-market is not a quantitatively important determinant of the frequency with which homes sell. The results I presented above, then, show that the mechanisms uncovered by Genesove and Mayer (1997) and Genesove and Mayer (2001) extend to the more important decision of whether to list the home in the first place. In this sense, then, these results extend and bolster that older literature. Nonetheless, while the household results are compelling, they simply cannot account for much of the decline in sales, as demonstrated.

5 Concluding Remarks

This paper has demonstrated that house lock and nominal loss aversion were not quantitatively meaningful drivers of the surge and collapse of home sales during the recent housing cycle. Importantly, these results are actually entirely consistent with a large literature demonstrating the

[21] Technically, this variation has be to between steady states, but the high rate at which listed homes sell makes the transition period relatively brief.

[22] This is not to say that that time-on-market does not matter for anything. First of all, as the search process is costly—a pillar of the search-and-matching literature—periods of extended time-on-market can be harmful to home sellers. Second of all, while it is true that home sellers waiting a bit longer to sell their homes does not impact sales volume, that extra time could be quite important for them to garner higher bids and thus sell at higher prices. Finally, in terms of aggregate variables, inventory—the stock of listed homes—does depend quite a bit on both n and s. Defining L to be the share of homes that are listed, an analogous analysis to the one above yields $dL \approx -\frac{n}{(s+n)^2}ds + \frac{s}{(s+n)^2}dn$. Now each hazard's contribution is no longer scaled by the *square* of the other, which allows s to be an important determinant of inventory. Intuitively, if listed homes sell very quickly, not many homes will be listed at any given point in time. So while overall sales are determined almost entirely by the flow of homes onto the market, the stock of homes on the market does depend on both flows.

relevance of these factors in household decision-making: in a dataset representing a large fraction of U.S. home sales, I very flexibly estimate the dependence of home-selling behavior on LTV and nominal losses and find evidence for these effects consistent with less-flexibly estimated empirical results as well as theoretical predictions. However, since my dataset represents the near-universe of home sales, I am able to perform simple shift-share analyses to gauge the strengths of these effects for aggregate sales volume, and they are very small. In an accounting sense, the decline in sales from boom to bust was driven almost entirely by households for whom the house lock and nominal loss mechanisms are not active.

As a secondary point, the paper emphasizes that sales volume depends almost exclusively on how frequently homes are listed for sale, rather than how quickly listed homes sell. This was discussed in the context of the seminal empirical papers Genesove and Mayer (1997) and Genesove and Mayer (2001), which empirically evaluated the role of the mechanisms in the second flow (from listed to sold). While the results in those papers are quite interesting, by focusing on the flow from being listed to being sold, they do not shed much light on overall sales volume. The same point, though, casts immediate doubt on any theoretical explanation that operates through time on market. Many models have been written[23] that use search frictions to generate longer selling times in downturns, thereby generating a qualitatively correct correlation between prices and volume. However, these explanations cannot be all that compelling since they operate on the wrong flow. Quantitatively meaningful explanations must explain why homeowners list their homes at much lower rates in downturns than in booms, not why it takes longer to sell in the former than the latter.

This of course begs the question of what the correct explanation is, since this paper has argued that many off-the-shelf explanations are in fact quite weak relative to the observed fluctuations in sales. A series of papers[24] has emphasized the participation of "speculators," agents who buy homes not because they want to live in them but rather because they think they can predict and exploit short-term price changes. These types of agents are likely an important part of the story. In contrast, the following chapter of this dissertation lays out an explanation that focuses on household behavior (i.e. agents who actually live in the homes that they own). I argue that an aggregate constraint, the irreversibility of housing construction, creates aggregate dynamics in a model of the housing market that in many ways match the evidence from the recent housing cycle, especially the rise and fall in sales volume. I turn to that next.

[23]Prominent examples include Wheaton (1990), Krainer (2001), Novy-Marx (2009), Head et al. (2014), Guren (2018).

[24]See Haughwout et al. (2011), Bayer et al. (2017b), Bayer et al. (2017a), and DeFusco et al. (2018).

Bibliography

Andersson, F. and T. Mayock (2014): "How Does Home Equity Affect Mobility?" *Journal of Urban Economics*, 84, 23–39.

Anenberg, E. (2011): "Loss Aversion, Equity Constraints and Seller Behavior in the Real Estate Market," *Regional Science and Urban Economics*, 41, 67–76.

Bayer, P., C. Geissler, K. Mangum, and J. Roberts (2017a): "Speculators and Middlement: The Strategy and Performance of Invesetors in the Housing Market," Nber working paper 16784.

Bayer, P., K. Mangum, and J. Roberts (2017b): "Speculative Fever: Investor Contagion in the Housing Boom," Nber working paper 22065.

Bhutta, N., J. Dokko, and H. Shan (2017): "Consumer Ruthlessness and Mortgage Default during the 2007 to 2009 Housing Bust," *The Journal of Finance*, 72, 2433–2466.

Bracke, P. and S. Tenreyro (2018): "History Dependence in the Housing Market," *Bank of England Staff Working Paper (No. 630)*.

Campbell, J. and J. Cocco (2015): "A Model of Mortgage Default," *Journal of Finance*, 70, 1495–1554.

DeFusco, A., C. Nathanson, and E. Zwick (2018): "Speculative Dynamics of Prices and Volume," Nber working paper 23449.

Engelhardt, G. (2003): "Nominal Loss Aversion, Housing Equity Constraints, and Household Mobility: Evidence from the United States," *Journal of Urban Economics*, 53.

Ferreira, F., J. Gyourko, and J. Tracy (2010): "Housing Busts and Household Mobility," *Journal of Urban Economics*, 68, 34–45.

——— (2012): "Housing Busts and Household Mobility: An Update," *FRBNY Economic Policy Review*.

Genesove, D. and C. Mayer (1997): "Equity and Time to Sale in the Real Estate Market," *American Economic Review*, 87, 255–269.

——— (2001): "Loss Aversion and Seller Behavior: Evidence from the Housing Market," *The Quarterly Journal of Economics*, 116, 1233–1260.

Guiso, L., P. Sapienza, and L. Zingales (2013): "The Determinants of Attitudes toward Strategic Defaults on Mortgages," *Journal of Finance*, 68, 1473–1515.

Guren, A. (2018): "House Price Momentum and Strategic Complementarity," *Journal of Political Economy*, 126, 1172–1218.

Haughwout, A., D. Lee, J. Tracy, and W. Van der Klaauw (2011): "Real Estate Investors, the Leverage Cycle, and the Housing Market Crisis," Staff report 514, Federal Reserve Bank of New York.

Head, A., H. Lloyd-Ellis, and H. Sun (2014): "Search, Liquidity, and the Dynamics of House Prices and Construction," *American Economic Review*, 104, 1172–1210.

Krainer, J. (2001): "A Theory of Liquidity in Residential Real Estate Markets," *Journal of Urban Economics*, 49, 32–53.

Ngai, L. R. and K. Sheedy (2017): "The Decision to Move House and Aggregate Housing-Market Dynamics," Working Paper.

Novy-Marx, R. (2009): "Hot and Cold Markets," *Real Estate Economics*, 37, 1–22.

Schulhofer-Wohl, S. (2012): "Negative Equity Does Not Reduce Homeowners' Mobility," *Federal Reserve Bank of Minneapolis Quarterly Review*, 35.

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

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

Figures

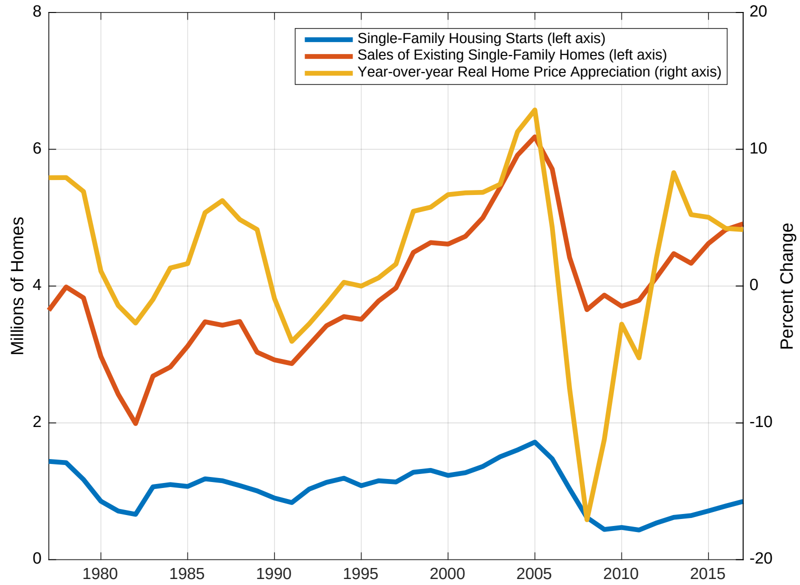

Figure 1: Housing starts for single-family homes (from Census Bureau) and sales of existing single-family homes (from National Assocation of Realtors) are plotted against the lefthand axis, and year-over-year growth of CoreLogic's national home price index is shown on the righthand axis.

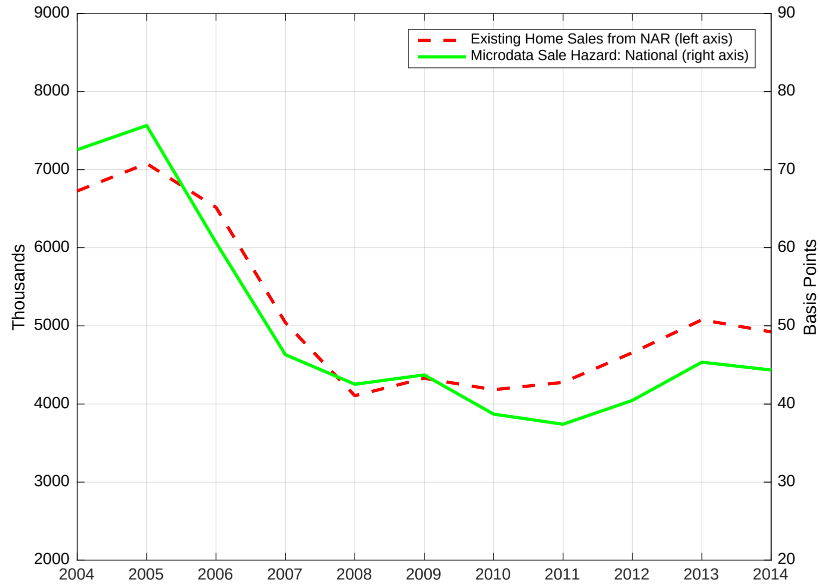

Figure 2: Comparison of National Association of Realtor's count of existing sales with the sale hazard from the final sample of microdata.

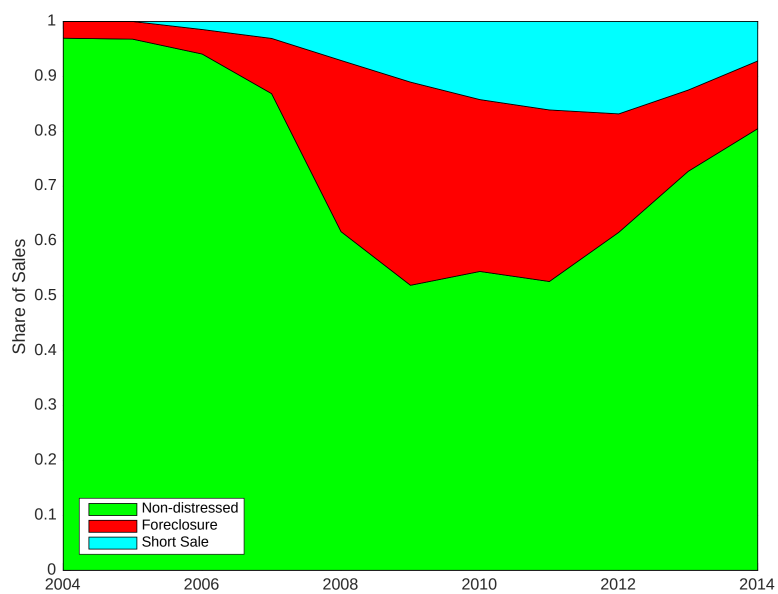

Figure 3: Sales in the deeds data are either non-distressed, short sales, or sales out of foreclosure. I show how the share of sales in each category evolves over time.

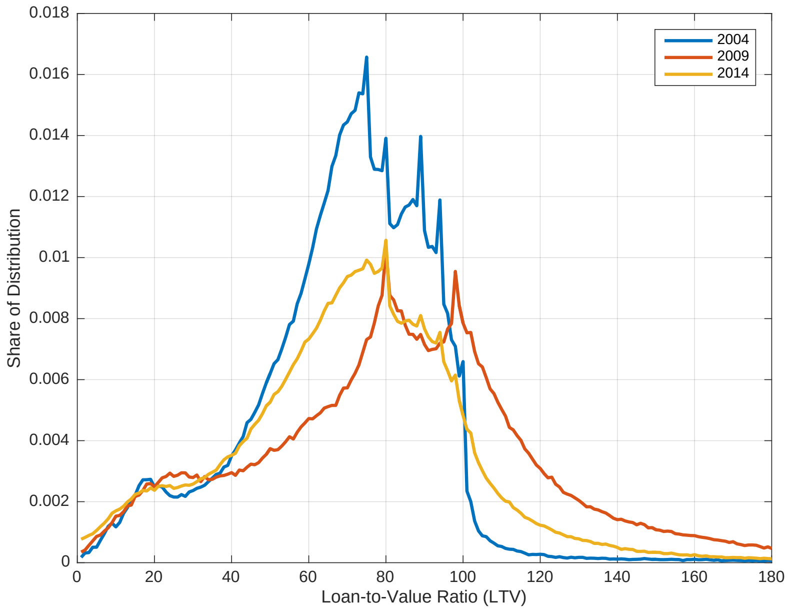

Figure 4: LTV distribution in estimation sample for selected years.

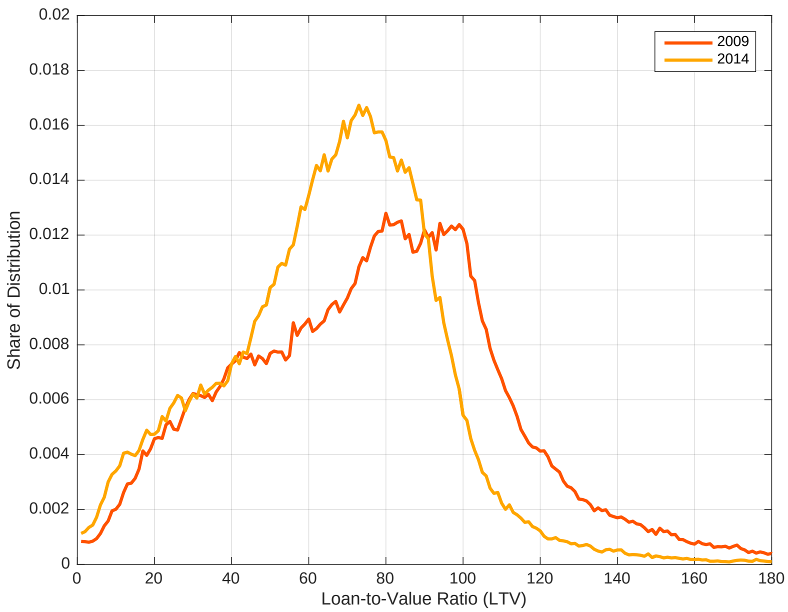

Figure 5: LTV distribution in CRISM dataset for selected years.

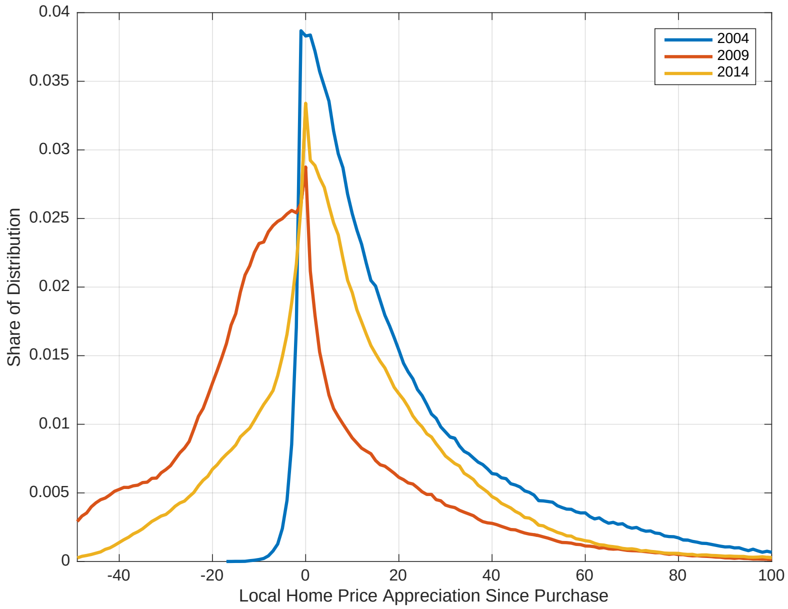

Figure 6: Distribution of local home price appreciation since purchase in estimation sample for selected years.

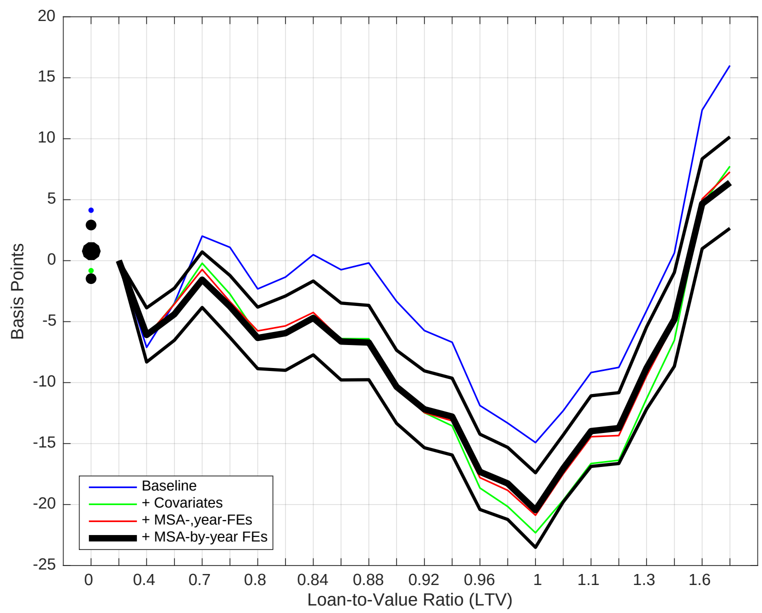

Figure 7: Relationship between sale probability and LTV, 1% national sample. Results come from a linear probability model. All specifications include months since purchase ("duration"), calendar month (to soak out any effects of seasonality), and a flexible step function for nominal gains/losses. The "Baseline" specification controls for only these. The specification "+ Covariates" further controls for county-level economic conditions (county unemployment rate 6 months prior, change in the county unemployment rate from 12 to 6 months prior, the gross number of jobs created in the county in that year [expressed as a percentage of employment]), ZIP code-level measures of financial health (average FICO score on purchase mortgages in 2003, average household income in 2004, and share of adults with at least a college degree in the 2000 Census), and some property level controls (initial down-payment and percentile rank within MSA of property's value). The specification "+ MSA-,year-FEs" adds in indicators for each MSA and year in the sample, while the final specification, "+ MSA-by-year FEs" adds in indicators for each MSA in each year. Standard errors are shown for the final specification, and they are clustered at the MSA-level level. There are 8 bins for nominal gains/losses: [0], (0,0.2], (0.2,0.4], (0.4,0.6], (0.6,0.7], (0.7,0.75], (0.75,0.8], (0.8,0.82], (0.82,0.84], (0.84,0.86], (0.86,0.88], (0.88,0.9], (0.9,0.92], (0.92,0.94], (0.94,0.96], (0.96,0.98], (0.98,1], (1,1.05], (1.05,1.1], (1.1,1.2], (1.2,1.3], (1.3,1.4], (1.4,1.6], (1.6,∞).

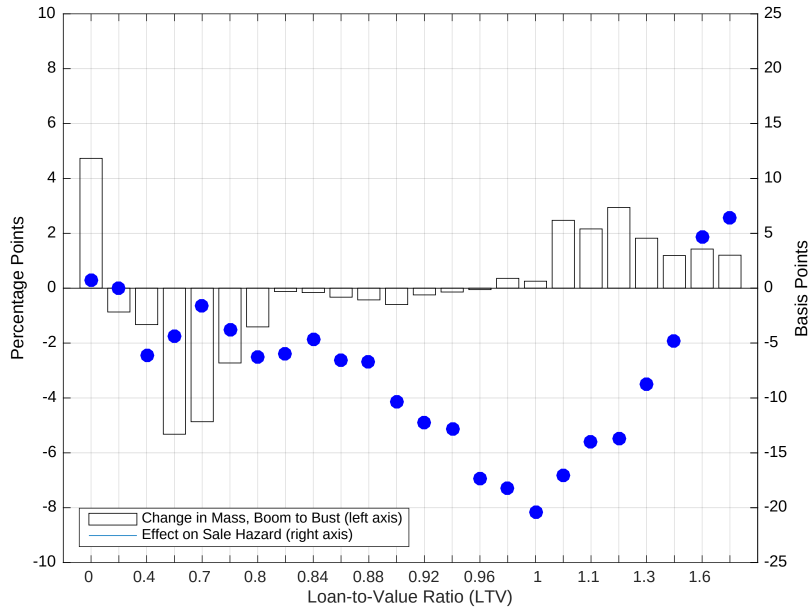

Figure 8: The relationship between the probability of sale and LTV is shown in the blue dots. The white bars show the change from 2004-6 to 2007-11 in the percent of borrowers in each bin.

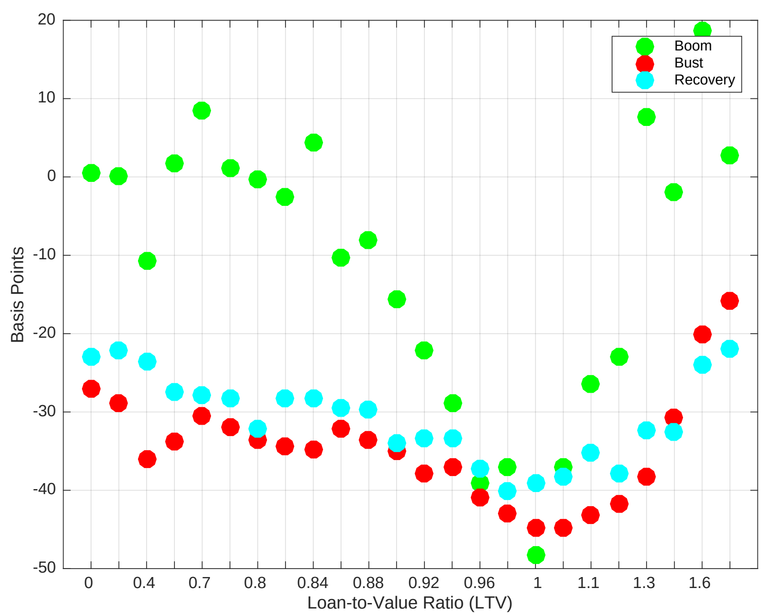

Figure 9: The time-varying relationship between the probability of sale and LTV. The "boom" (2004-6) is shown in green dots, the "bust" (2007-11) is shown in red dots, and the "recovery" (2012-4) is shown in blue dots. The estimates come from a regression with all covariates described in the caption of Figure 7 and MSA fixed effects).

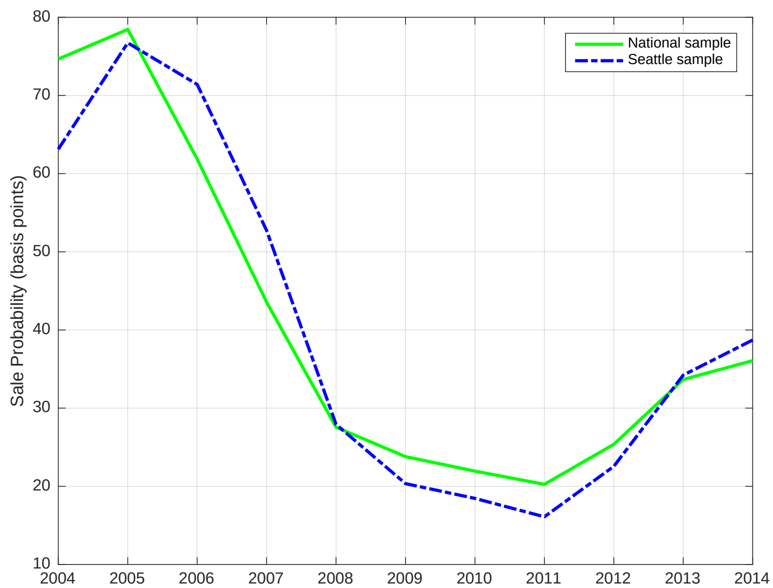

Figure 10: Hazard of a non-distressed sale in the deeds microdata, for both the 1% U.S. sample and the 100% Seattle-Tacoma-Bellevue sample.

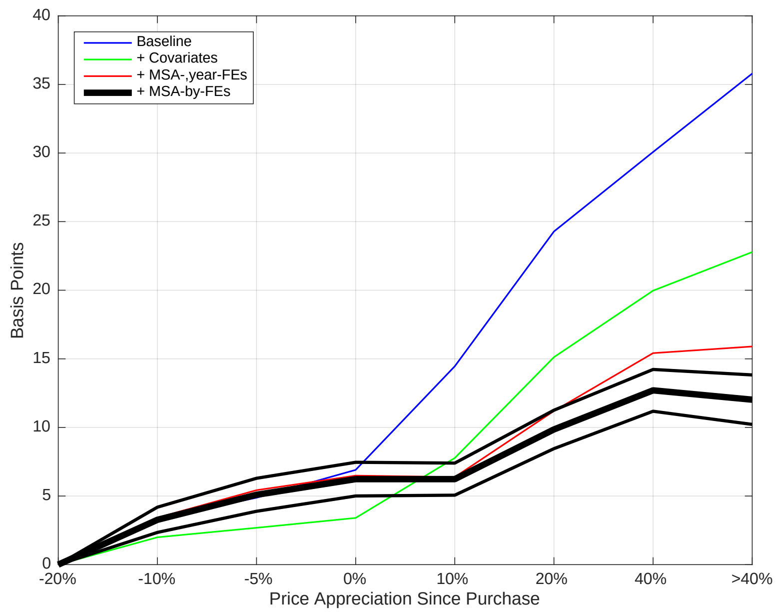

Figure 11: Relationship between non-distressed sale probability and nominal gains/losses, 1% national sample. Results come from a linear probability model. All specifications include months since purchase ("duration"), calendar month (to soak out any effects of seasonality), and a flexible step function for LTV. The "Baseline" specification controls for only these. The specification "+ Covariates" further controls for county-level economic conditions (county unemployment rate 6 months prior, change in the county unemployment rate from 12 to 6 months prior, the gross number of jobs created in the county in that year [expressed as a percentage of employment]), ZIP code-level measures of financial health (average FICO score on purchase mortgages in 2003, average household income in 2004, and share of adults with at least a college degree in the 2000 Census), and some property level controls (initial down-payment and percentile rank within MSA of property's value). The specification "+ MSA-,year-FEs" adds in indicators for each MSA and year in the sample, while the final specification, "+ MSA-by-year FEs" adds in indicators for each MSA in each year. Standard errors are shown for the final specification, and they are clustered at the MSA-level level. There are 8 bins for nominal gains/losses: ($-\infty$,-20%], (-20%,-10%], (-10%,-5%], (-5%,0%], (0%,10%], (10%,20%], (20%,40%], (40%,∞).

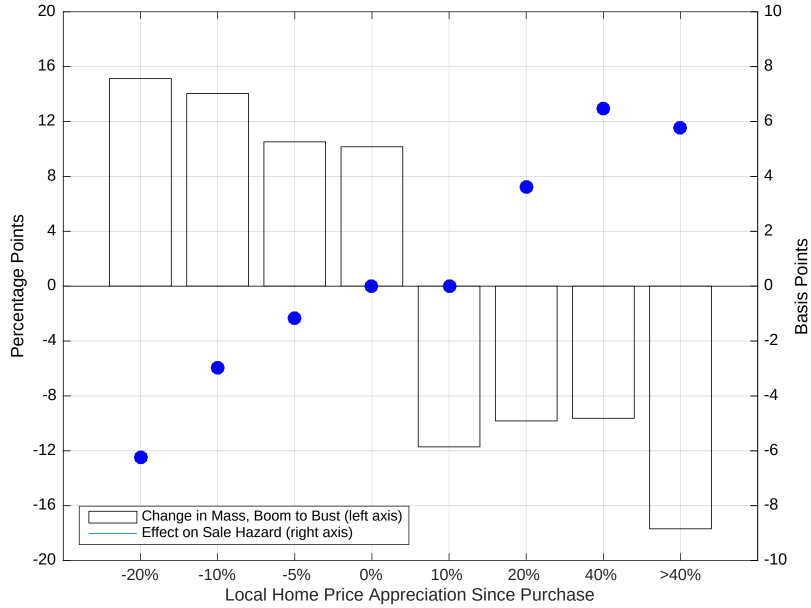

Figure 12: The relationship between the probability of sale and nominal gains/losses is shown in the blue dots. The white bars show the change from 2004-6 to 2007-11 in the percent of borrowers in each bin.

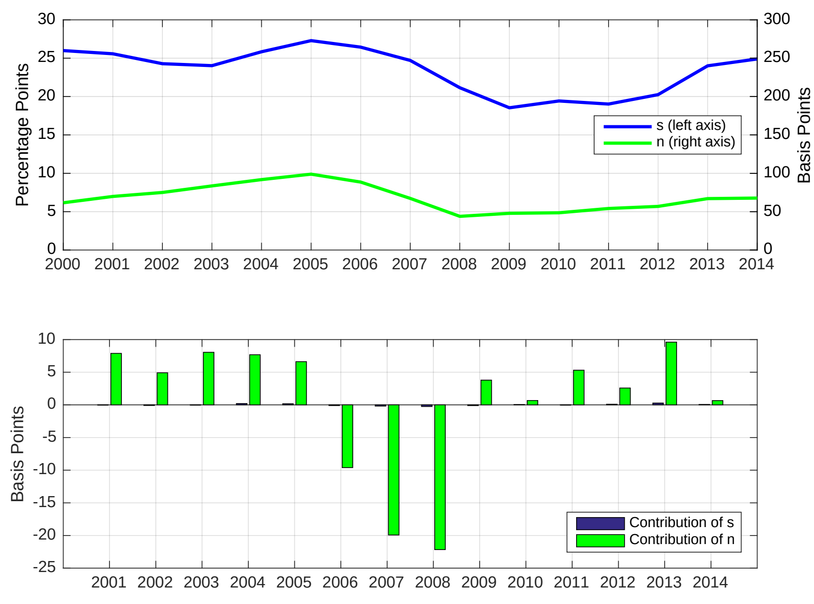

Figure 13: The top panel shows the variation in the probability in the Seattle sample that a listed home sells (s) and that an unlisted home is listed (n). (Note that the 2 series are plotted on different axes.) The bottom panel shows the share of the change in the overall sale hazard that can be attributed to changes in each s and n.

Appendix

A.1 Empirical Results for Seattle-Tacoma-Bellevue MSA

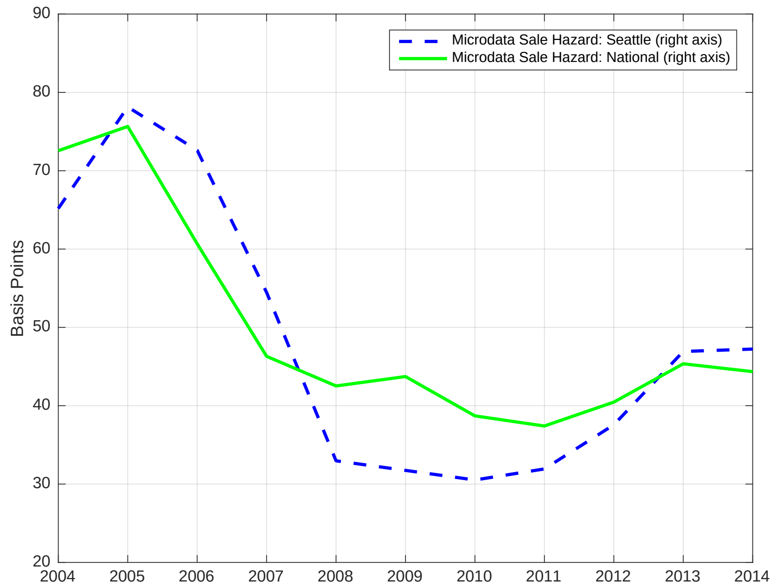

Figure A-1: Comparison of sale hazards in the Seattle and national samples.

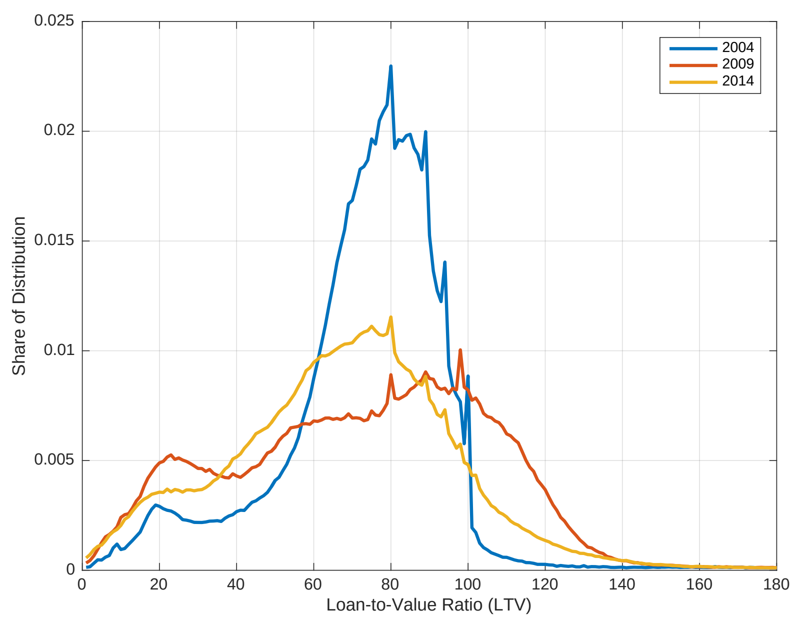

Figure A-2: LTV distribution in Seattle sample for selected years.

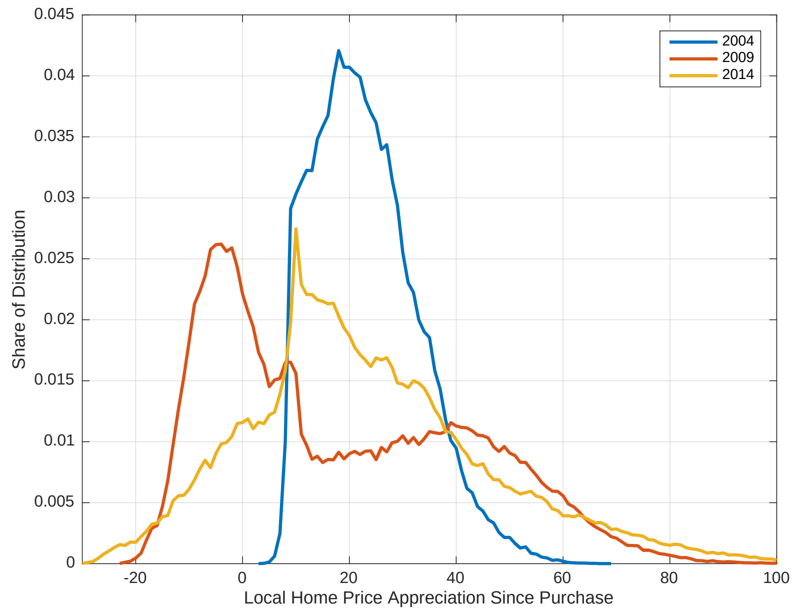

Figure A-3: LTV distribution in Seattle sample for selected years.

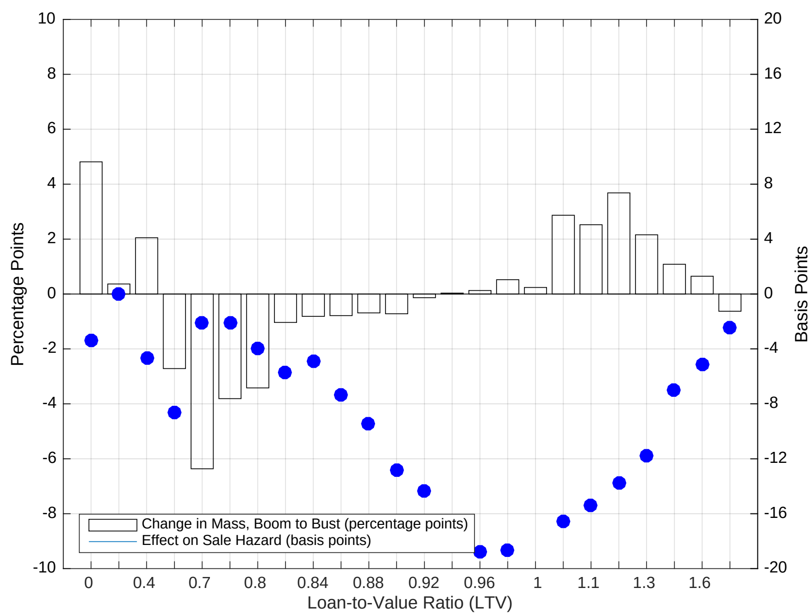

Figure A-4: The relationship between the probability of sale and LTV, estimated from the Seattle-Tacoma-Bellevue sample, is shown in the blue dots. The white bars show the change from 2004-6 to 2007-11 in the percent of borrowers in each bin in.

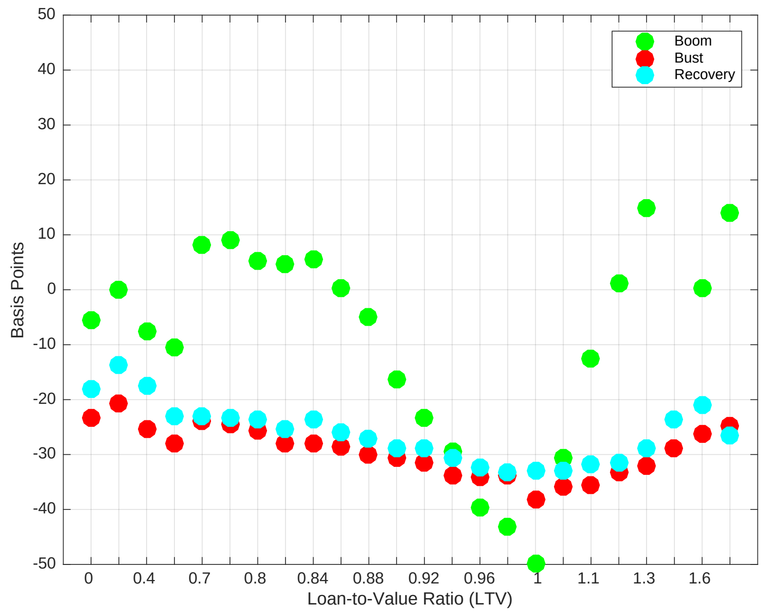

Figure A-5: The time-varying relationship between the probability of sale and LTV in the Seattle-Tacoma-Bellevue MSA. The "boom" (2004-6) is shown in green dots, the "bust" (2007-11) is shown in red dots, and the "recovery" (2012-4) is shown in blue dots. The estimates come from a regression with all covariates described in the caption of Figure 7 and MSA fixed effects).

	(1)	(2)	(3)
LTV	0.9	1.3	1.1
	(0.3)***	(0.4)***	(0.4)***
(LTV-0.8)$\cdot I_{\{LTV>0.8\}}$	-17.0	-14.6	-14.3
	(0.5)***	(0.5)***	(0.5)***
HPA	40.0	34.5	35.2
	(1.0)***	(1.0)***	(1.1)***
HPA$\cdot I_{\{HPA>0\}}$	-29.5	-21.8	-21.1
	(1.2)***	(1.3)***	(1.3)***
LTV + (LTV-0.8)$\cdot I_{\{LTV>0.8\}}$	-16.0	-13.3	-13.2
	(0.3)***	(0.3)***	(0.3)***
HPA + HPA$\cdot I_{\{HPA>0\}}$	10.5	12.7	14.1
	(0.4)***	(0.5)***	(0.5)***
Controls		✓	✓
Year FEs			✓

Table A-1: Regression results using MLS listings data in Seattle. Sample is non-REO homes listed for sale. "HPA" is home price appreciation since property was last purchased—my measure of nominal gains/losses. Due to reduced sample size (from only looking at listed properties), I use a parametric relationship (linear spline) to study the effects of LTV and nominal gains/losses on selling behavior. All specifications control for months since purchase, calendar month, and months since put on market (listed). Specification (2) further controls for variables listed in the footnote of Figure 7. The final columns adds indicators for each year. Note: $N = 405,702$.