Mortgage Choices and Housing Speculation*

Gadi Barlevy
Federal Reserve Bank of Chicago
gbarlevy@frbchi.org

Jonas D.M. Fisher
Federal Reserve Bank of Chicago
jfisher@frbchi.org

November 19, 2010

Abstract

We describe a rational expectations model in which speculative bubbles in house prices can emerge. Within this model, when there is a bubble, both speculators and their lenders use interest-only mortgages (IOs) rather than traditional mortgages. Absent a bubble, there is no tendency for IOs to be used. These insights are used to assess the extent to which house prices in US cities were driven by speculative bubbles over the period 2000-2008. We find that IOs were used sparingly in cities where elastic housing supply precludes speculation from arising. In cities with inelastic supply, where speculation is possible, there was heavy use of IOs, but only in cities that had boom-bust cycles. Peak IO usage predicts rapid appreciations that cannot be explained by standard correlates and this variable is more robustly correlated with rapid appreciations than other mortgage characteristics, including sub-prime, securitization and leverage. Where IOs were popular, their use does not appear to have been a response to houses becoming more expensive. Indeed, their use anticipated future appreciation. Finally, consistent with the reason why lenders prefer IOs, these mortgages are more likely to be repaid earlier or foreclose. Combined with our model, this evidence suggests that speculative bubbles were an important factor driving prices in cities with boom-bust cycles.

JEL Classification Numbers: E0, O4, R0
Keywords: House prices, interest-only mortgages, subprime mortgages, securitization, speculation

*We thank Sumit Agarwal, Gene Amromin, Marco Bassetto, Huberto Ennis, Simon Gilchrist, Lars Hansen, Ned Prescott, Amir Sufi, for helpful comments. We also thank Ryan Peters and Shani Schechter for superb research assistance. The views expressed herein are those of the authors and do not necessarily represent those of the Federal Reserve Bank of Chicago or the Federal Reserve System.

1 Introduction

The financial crisis of 2007 has refocused attention on the housing market and its apparent vulnerability to boom-bust cycles in which house prices appreciate dramatically over a relatively short time period and then collapse. As evident from the U.S. experience, such cycles have the potential to severely disrupt the functioning of the financial sector given its exposure to house price risk, which in turn can affect real economic activity. Consequently, economists and policymakers have sought to understand when and why boom-bust cycles can arise in the housing market. Are such price movements driven by fundamentals, or do they reflect speculation in which prices increasingly drift away from the expected value of the services the underlying assets can offer? Are there any indicators that can predict where such boom-bust episodes might occur if policymakers wish to intervene before they develop?

This paper examines whether data from the mortgage market can help to address these questions. Our focus on the mortgage market is motivated by theoretical work that suggests credit markets can play a key role in allowing for speculative bubbles, e.g. Allen and Gorton (1993) and Allen and Gale (2000). These papers show that if traders finance their asset purchases with borrowed funds, they are willing to pay more for a risky asset than its expected value. This is because they can default on their creditors should their gamble fail. Lenders would naturally be reluctant to finance such speculative activity that comes at their expense. But, if lenders are unable to distinguish speculators from safe, profitable borrowers, they may end up financing such speculative purchases after all.

If credit markets indeed play a role in allowing for speculative bubbles, then, if at least some of the boom-bust cycles in the housing market reflect speculation, credit market data might be relevant for predicting the occurrence of such episodes. For example, if borrowers temporarily bid up prices above their true value because they can default should their speculative purchases fail, boom-bust cycles might be more likely to emerge if and when borrowers are able to leverage themselves to a greater extent and thus default against a larger share of the assets they purchase. Indeed, previous work by Lamont and Stein (1999) has already argued that house prices tend to be more volatile in cities where a larger proportion of mortgages are highly leveraged.[1]

[1] More precisely, Lamont and Stein (1999) show that in cities with a large share of mortgages with a loan-to-value ratio of over 80%, house prices respond more to income shocks than in cities with a small share of such mortgages. Their work was not motivated by interest in speculation, but by work in Stein (1995) on down-payment constraints. In Stein's model, house prices reflect fundamentals. However, down-payment constraints impede the efficient allocation of houses and make the fundamentals more volatile, similarly to Kiyotaki and Moore (1997), which implies more volatile house prices. This hypothesis is distinct from, but not mutually exclusive of, the model that motivates our analysis.

While previous work has been concerned with leverage, here we consider other mortgage market characteristics that are motivated by the work of Barlevy (2009). That paper argues that if lenders cannot avoid lending to speculators, they would have an incentive to offer particular types of contracts to influence the behavior of these speculators. Here, we build on this insight by focusing specifically on mortgage contracts, and argue that when lenders know that some of those they lend to are buying overvalued assets to speculate, it will be possible to make both lenders and speculators better off using contracts with back-loaded payments, i.e. contracts where the initial payments stipulated in the contract are low while later payments are onerously high. Lenders prefer these contracts because they preclude the borrowers from gambling at their expense for too long, given that speculators will be forced to sell the asset once payments rise (or else refinance with another borrower if possible). At the same time, borrowers prefer these contracts because they can defer building up equity in what they know is a risky asset, leaving them with the option to default on all of the principal they borrowed should the prices collapse early. Thus, these contracts effectively get the borrower to commit to settling his debt earlier than he would under a traditional mortgage contract. We further show that the fact that back-loaded contracts can make both parties better off is intimately related to the fact that the asset is the target of speculation; if it were not, then absent any other frictions, it would no longer be possible for back-loaded contracts to make both the lender and the borrower better off.

These results lead us to look at whether markets with boom-bust cycles also relied on back-loaded contracts.[2] We find that the use of back-loaded payments, specifically interest-only mortgages (IOs), was highly concentrated in cities that experienced boom-bust cycles.[3] In particular, we find that IOs were used only sparingly in areas with few restrictions on the supply of housing, i.e. cities where the supply response would prevent speculation from ever arising. But in cities where geographical and regulatory restrictions could have inhibited supply, so that speculation would have been possible under our theoretical setup, such contracts were used quite prevalently, but only in cities that experienced boom-bust cycles.

To convey the spirit of our findings, consider two cities: Phoenix, Arizona and Laredo, Texas. Laredo is a low income border city in a state with little regulation and vast open spaces on which new homes can be built. As such, we would expect that if house prices

[2]For a different perspective on boom-bust cycles see Burnside, Eichenbaum, and Rebelo (2010).

[3]Contemporaneous work by Amromin, Huang, and Sialm (2010) also finds that the contracts we focus on were associated with high price appreciation in the boom phase, although they emphasize the complexity of these contracts rather than their back-loaded nature. The fact that house price appreciation was concentrated in areas using back-loaded contracts was also noted in congressional testimony by Sandra Thompson of the FDIC back in September of 2006 during a hearing regarding nontraditional mortgage products, although this testimony points to a statistical pattern without studying it rigorously.

in Laredo were ever to rise above their fundamental value, the supply of housing would quickly rise in response and drive prices back down. This would limit the growth of housing prices to the growth of fundamentals. By contrast, although Phoenix also has plenty of open space, it also ranks fairly high on the Wharton Residential Land Use Index complied by Gyourko, Saiz, and Summers (2008). These restrictions could have prevented home builders from responding quickly if house prices exceeded fundamentals, so that house prices could have grown faster than fundamentals. Figure 1 shows the Federal Housing Finance Agency (FHFA) house price index for these two cities, deflated by the Consumer Price Index. In Laredo, real house prices grew at a steady rate of roughly 2.5% per year between 2000 and 2008. In Phoenix, house price growth was indeed much higher, on average 9.5% per year between 2001 and 2006 and 36% in 2005 alone. House prices then declined sharply, reverting to their 2001 levels by 2010. The fact that cities with geographical and regulatory restrictions on housing supply have more volatile housing prices has been pointed out before; see, for example, Krugman (2005) and Glaeser, Gyourko, and Saiz (2008). But Panel A of Figure 1 also shows that home buyers in the two cities relied on different types of mortgage contracts to finance their purchases: IOs grew to over 40% of all mortgages for purchase in Phoenix as home prices climbed, but accounted for at most 2% of mortgages for purchase in any given quarter in Laredo.

The association between extensive use of IOs and rapid house price appreciation evident in these two cities remains when we look at a cross-section of over 200 cities, and is robust to controlling for various city-level characteristics. In particular, the peak share of IOs among first lien mortgages for purchase turns out to be a better indicator of rapid house price appreciation than other variables that have been shown to be useful in predicting unusually high house price appreciation, e.g. restrictions on housing supply and whether the city previously experienced boom-bust cycles in housing prices. The peak share of IOs also does better than other characteristics of mortgages we consider, including the share of mortgages with high leverage ratios and the share of mortgages privately securitized within a year of origination. Even more noteworthy, we find no relationship between house price appreciation at the city level and the peak share of subprime mortgages. This result, which may seem surprising at first, reflects the fact that subprime mortgages were more common in low income cities while boom-bust cycles were more common in middle income and wealthy cities. Panel B of Figure 1 is consistent with this pattern. Laredo, a lower-income city, had one of the largest shares of subprime mortgages during this period, peaking at over 27%. By contrast, the share of subprime mortgages in Phoenix peaked at 11% of all home purchases. Thus, subprime borrowers do not appear to have played an important role in accounting for the boom-bust pattern in housing prices observed at the metropolitan level. This finding does not deny a role for subprime lending in explaining within-city price patterns as shown by Mian and Sufi (2009), the rise in home ownership over the period, as argued by Chambers, Garriga, and Schlagenhauf (2009), or the subsequent default wave, as argued by Corbae and Quintin (2009).

While our theory suggests IOs are heavily used because of the presence of speculative behavior, their use could just reflect rapid house price appreciation if price appreciation forces borrowers to turn to these mortgages for reasons of affordability. We offer evidence against this interpretation. First, we find that in cities where IOs were popular, their use took off before house prices. In other words, the use of these contracts anticipated the growth in housing prices rather than the other way around. This pattern can be seen in Panel A of Figure 1: The use of IOs in Phoenix began in early 2004, while house prices only took off in late 2004 and early 2005. We also find that the peak share of IOs continues to predict house price appreciation even after controlling for the level of the median house price at the peak and a measure of housing affordability. Finally, although other affordable mortgage products such as longer-term and hybrid mortgages were more common in cities with high price appreciation, the share of IOs turns out to be a much better predictor of which cities experience rapid house price appreciation.

Our findings offer several new insights for understanding boom-bust cycles in the housing market. First, they provide evidence that the boom-bust cycles in the housing market could have reflected a speculative bubble given that the contracts which are strictly preferred in this situation are more heavily used in cities that exhibited these cycles. In particular, the evidence we provide does not involve comparing the level of house prices to some measure of the true fundamental value of housing, but behavioral patterns that we show should be observed when assets are overvalued due to speculation. Although our evidence does not definitively prove that boom-bust cycles were driven by a speculative bubble, it does point to a pattern that any theory of boom-bust cycles must explain: in cities where such cycles occurred, home buyers took out back-loaded contracts in anticipation of the appreciation of house prices and not in response to them.

On the face of it, it might seem puzzling that lenders in cities with rapid price appreciation would be willing to expose themselves to more housing risk by letting borrowers avoid building equity in the homes they purchase. Although IOs did charge a premium, the premium was typically small, and lenders would collect less on these mortgages early on than on conventional mortgages with a slightly lower rate. Yet our theory can explain this pattern – it argues that lenders agreed to these contracts because they force borrowers to repay more quickly.

Lastly, since our results suggest that back-loaded contracts were used *before* house prices appreciated, policymakers might be able to use the type of contracts used to finance home purchases to anticipate boom-bust cycles. While IOs appeared to have been the relevant contract for forecasting rapid house price appreciation during the recent episode, more generally our model suggests a preference for back-loaded contracts, and in future episodes the contracts used to achieve back-loading may differ in details from those of this episode.

The paper is organized as follows. In the next section, we discuss the theoretical environment that motivates us to look at back-loaded mortgages as a predictor of boom-bust cycles. In Section 3, we describe the data we use. Section 4 documents the cross-sectional relationship between house price appreciation and mortgage characteristics. Section 5 shows that in cities where IO contracts were common, their use anticipated rather than followed house price appreciation. Section 6 examines whether borrowers with back-loaded contracts do indeed repay their debt more quickly, which in our model is the reason lenders prefer these contracts. Section 7 concludes.

2 Theory

In this section, we develop a model of speculation in the housing market due to risk-shifting following Allen and Gorton (1993), Allen and Gale (2000), and Barlevy (2009). In particular, we show that a combination of constrained housing supply, uncertain demand, and leverage allow for speculative bubbles in equilibrium. We then look at the implications of our model for which mortgages will be used to finance house purchases. We start our discussion by describing the economic environment. Next, we show that some agents in our model buy houses in the hope of selling them later for a higher price, pushing house prices above fundamentals. Finally, we discuss mortgage choice when house prices exceed fundamentals.

2.1 Environment

A key feature of our model is that we allow people to differ in how much they value home ownership: Some derive high service flows from owning a home as opposed to renting, e.g. because they can customize the house to suit their personal tastes, while others derive no additional value from owning over renting. The former are natural home owners, and have an incentive to hold on to their house even if house prices fall. The latter are ordinarily indifferent between owning and renting, but in some circumstances will opt to borrow and buy houses for speculative purposes. Since lenders cannot observe preferences, they can end up financing the latter type as they seek to lend to those who truly value home ownership.

Formally, suppose agents are infinitely-lived, risk-neutral, and discount at rate β. Low valuation types derive a per-period flow $(\beta^{-1}-1)\,d$ from occupying a house regardless of whether they rent or own. The value to buying a house for these types is thus

$$\sum_{t=1}^{\infty}\beta^{t}\left(\beta^{-1}-1\right)d=d \tag{1}$$

High valuation types receive the same flow utility of $(\beta^{-1}-1)\,d$ if they occupy a rented house. However, if they own the house they occupy and then customize it, they can receive a higher flow $(\beta^{-1}-1)\,D$ where $D>d$. The value to owning a house for these types is

$$\sum_{t=1}^{\infty}\beta^{t}\left(\beta^{-1}-1\right)D=D \tag{2}$$

Individuals can live in one house, and high types receive a higher flow only by occupying a house. We further assume that once high types customize a house, they will find it impossible to derive the same high flow elsewhere. This ensures that even if house prices fall, high types who borrowed to buy their house will not be tempted to default and move to an identical house elsewhere at a lower price, an implication that will be important below.

We analyze the equilibrium for a single city in isolation. The initial stock of houses in this city at date 0 is normalized to 1. The mass of agents who could own a house at date 0 is assumed to be strictly larger than the stock of housing. This allows us to avoid indeterminacies that can occur when there are as many houses as potential buyers. Potential buyers who opt not to own a house in this city can instead rent in the same city or move to another city. The initial distribution of houses in our city across potential owners at date 0 turns out to be irrelevant for our analysis, so we need not specify it. As an aside, we can allow for additional individuals who are either unable or uninterested in owning property in the city but would agree to rent for $(\beta^{-1}-1)\,d$ per period. That is, we need not allow all those who rent in the city to also have the option to own a home in that city.

As a preliminary step, consider equilibrium house prices when the number of potential buyers and houses are both constant over time. Since there are more potential buyers in total than houses, the price of houses cannot fall below d without resulting in excess demand. This gives us a lower bound on house prices. Now, suppose there are more houses than high type buyers. If house prices were above d, then homeowners who are either landlords or low type occupants could sell the house for more than they would get by holding on to the house indefinitely. For them to agree to own a house, they must expect to sell it for an even higher price in the future. However, the equilibrium price of housing must be such that no agent violates his transversality conditions. This requires that asymptotically, house prices grow less than the discount rate. Hence, if an agent were to buy a house intending to resell it, there would be a finite date at which it will be optimal to sell it. But if the price at this date exceeded d, there would be an excess supply of houses at this date. Hence, the price of houses could not have exceeded d. With more houses than high type buyers, then, the price of housing must equal d. Similarly, if there were more high types than houses, the equilibrium price of housing would have to be D: If the price were lower there would be excess demand for houses, while if the price were higher than either a transversality condition or market clearing at some future date would be violated.

To allow for nontrivial price dynamics, from now on we assume there are more houses than high types that at date 0. That is, some mass $\phi_0 \in (0, 1)$ of houses is initially either owned and occupied by low types or used as rental property. We then introduce the notion that some unanticipated shock creates more potential home buyers in the environment above. For example, a wave of immigrants might unexpectedly show up in the city, or as is more relevant for the period we study, financial innovation might make it possible for agents previously shut out of credit markets to now borrow and buy a home in this city.[4] This new group can only participate in the housing market after date 0. We assume they own no resources initially, in line with the interpretation that these are agents who previously had difficulty accessing credit. Thus, to buy a home these agents must first borrow an amount equal to the price of the house. The only loan arrangements we allow are non-recourse mortgages. That is, agents who borrow must repay their loan in a sequence of payments $\{m_\tau\}_{\tau=1}^{T}$ that yield a rate of return r to the lender on any outstanding principle, where τ indexes time since the loan was taken out and T represents the term of the mortgage. If an agent fails to make any payment m_τ, he is found in default and ownership of the house transfers to the lender. Non-recourse means that once a lender takes possession of a house, he cannot go after the borrower's other income sources.[5] To ensure borrowers could potentially pay back their loan, we assume they receive an income flow $\{\omega_\tau\}_{\tau=1}^{\infty}$ after taking out the loan that is large enough to structure a sequence of mortgage payments $m_\tau \leq \omega_\tau$ under which the borrower could pay off his debt in finite time. At the same time, we will need income not to be too high so that the borrower's debt obligation can only be discharged at a slow rate under any feasible contracts, in a sense we make precise below.

Below, we show that a speculative bubble can emerge in this economy, but only if there is some uncertainty that borrowers can gamble on and shift their losses on to their creditors

[4]For evidence on the expansion of credit, see Mian and Sufi (2009).

[5]For a discussion on recourse mortgages in the U.S., see Ghent and Kudlyak (2009). They argue that even in states that allow for recourse, lenders often find it unprofitable to go after other sources of income.

should the gamble fail. In what follows, we model this uncertainty as stemming from incomplete information about the number of new potential buyers that will ultimately arrive. As we shall see below, this uncertainty translates in the model into uncertainty about future house prices. We further assume this uncertainty is resolved only gradually. Formally, suppose that new potential buyers arrive sequentially in cohorts of mass n per period until some date $\mathcal{T}$ that is distributed geometrically, i.e. $\Pr(\mathcal{T} = t) = (1-q)^{t-1} q$ for $t = 1, 2, ...$ Agents do not know when the flow of new arrivals will stop until date $\mathcal{T}+1$, the first date in which no new buyers arrive. Beforehand, agents assign probability q that new buyers will cease to arrive next period. We assume a fraction ϕ of arrivals are low types, where ϕ is not too large to ensure lenders are willing to lend even at interest rates close to the risk-free rate. Lenders know the fraction of low types ϕ but not the preferences of any one borrower.

To simplify the analysis, we further assume that between dates 1 and $\mathcal{T}$, only those who arrive within a period can buy houses that period. That is, buyers must buy a house immediately or move to another city; they cannot time their purchases. For reasons we discuss below, this restriction may bind for high types who wish to delay buying. Letting low types delay their purchases complicates the analysis, but does not change our results.

Finally, since one way to accommodate new buyers is to build new homes, we need to specify the cost of building additional houses. We restrict these costs be less than or equal to d. Absent constraints on supply, then, new homes can be added at a cost that is no more than the valuation of low types. For simplicity, suppose costs are constant and equal to d.

2.2 Equilibrium House Price

The effect of new potential buyers on the housing market naturally depends on housing supply conditions. Absent any restrictions on supply, the equilibrium house price will simply equal building costs. This is because positive profits from construction would lead to infinite housing supply. Given our assumption on costs, the price is therefore equal to d. At this price, newly arriving high types will want to borrow to buy a home as long as the interest rate on the mortgage was below some cutoff that exceeds $\beta^{-1} - 1$. Low types will be indifferent about buying a home at an interest rate of $\beta^{-1} - 1$ and would avoid buying a house at a higher interest rate. Since loans are fully collateralized given house prices never decline, the equilibrium interest rate will equal the risk-free rate $\beta^{-1} - 1$. Hence, the equilibrium in this case is identical to the one in the benchmark case with more houses than high type buyers.

The more interesting case is when there are constraints on new construction. To capture this, we can interpret n as the mass of new arrivals per period net of new construction that period, and ϕ as the ratio of the mass of low types among new arrivals to n. Since the total number of arrivals is uncertain, agents cannot be sure whether the number of arrivals through date $\mathcal{T}$ will be large enough so that there will be more high types than houses or vice versa. More precisely, since the mass of homes not initially owned by high types is ϕ_0, and each period $(1-\phi)\,n$ units of the housing that belong to these original owners at date 0 would have to be reallocated to newly arriving high types to ensure all high type potential buyers can occupy a house, buyers would need to keep arriving for at least $\frac{\phi_0}{n(1-\phi)}$ periods before the number of high types surpasses the available number of houses. Let t^* denote the smallest integer strictly greater than $\frac{\phi_0}{n(1-\phi)}$. By date t^*, any lingering uncertainty about the housing market will be resolved. If buyers stop arriving before t^*, i.e. if $\mathcal{T} < t^*$, then there will be fewer high types than houses, and the house price will equal d from date $\mathcal{T}$ on. But if buyers keep arriving through t^*, i.e. if $\mathcal{T} \geq t^*$, then there will be more high types than houses, and the equilibrium price at date t^* will equal D.[6]

When the size of arriving cohorts is very small or very large, there will be no uncertainty as to whether there will be more high types or houses. In particular, if $n = 0$, then $t^* = \infty$ so that $\Pr(\mathcal{T} < t^*) = 1$. Just as in the benchmark case, house prices will equal d forever. This will also be the case if some new potential buyers did arrive, but agents remains certain that the stock of houses always surpassed the number of high type buyers. In the opposite direction, when $n > \frac{\phi_0}{1-\phi}$, the mass of high types that arrive in the first period is enough to buy out any houses not yet occupied by high types. In this case, $t^* = 1$ so that $\Pr(\mathcal{T} \geq t^*) = 1$. The price of housing would then jump to D. More generally, if agents are certain that at some point there will be more high types than houses, house prices will jump immediately to ensure agents cannot profit from buying houses now and selling when high type buyers show up, even if high types trickle in gradually. Although house prices rise in response to the unanticipated arrival of new buyers, there is no sense in which this should be viewed as a bubble. Rather, house prices rise because scarce land becomes more valuable the more high types arrive, since the same land now yields more housing services.

For intermediate cohort sizes, i.e. $0 < n < \frac{\phi_0}{1-\phi}$, the price of housing that will prevail at date t^* remains uncertain as long as new potential buyers keep arriving. We now argue that this uncertainty can lead to a speculative bubble in housing, i.e. a situation in which agents bid up the price of housing above its inherent worth because they expect that they might sell the house for an even higher price in the future. To see this, we must first define the

[6]Our setup assumes that if $\mathcal{T} \geq t^*$, the number of high types exceeds the stock of houses indefinitely. But this is not essential. We could have instead assumed that building constraints are temporary and new construction would eventually drive house prices down to d. Any temporary shortage would still lead to prices at t^* that are higher than d, and our results only require that the price at t^* be uncertain.

inherent worth of a house, i.e. its fundamental value. We define this as the value to society of building another house in the current period. Since we assumed there were more potential buyers than houses at date 0, an additional house can always be used to provide at least $(\beta^{-1}-1)\,d$ in housing services per period, and the higher value $(\beta^{-1}-1)\,D$ if the number of high types exceeds the stock of houses. Since the stock of available housing exceeds the number of high types before t^*, the value of housing services up to date t^* is $(\beta^{-1}-1)\,d$. Beyond this date, the value of services will be permanently either high or low, depending on whether enough new high types arrived to exhaust the stock of housing. Formally, if we let v_{t^*} denote the present discounted value of services beyond date t^*, then v_{t^*} is equal to d if $\mathcal{T} < t^*$ and D if $\mathcal{T} \geq t^*$. The fundamental value of a house as of date t is then just

$$f_t = \sum_{s=t+1}^{t^*} \beta^{s-t} \left(\beta^{-1}-1\right) d + \beta^{t^*-t} E_t \left[v_{t^*}\right] \tag{3}$$

Since we know from the benchmark case earlier that the price of housing at date t^* will equal v_{t^*} since all uncertainty at t^* is resolved, we can rewrite f_t as

$$f_t = \sum_{s=t+1}^{t^*} \beta^{s-t} \left(\beta^{-1}-1\right) d + \beta^{t^*-t} E_t \left[p_{t^*}\right] \tag{4}$$

That is, our definition corresponds to the value of holding the house until date t^* and selling it at that date. This notion can be reconciled with the alternative definition for fundamentals which holds that the fundamental value of an asset reflects the value of holding the asset indefinitely. In particular, when agents value the asset differently, as in our environment, Allen, Morris, and Postlewaite (1993) proposed to define the fundamental value as the price at which "every agent holding or buying the asset would willingly do so even if he were to be forced to maintain his holding of the asset forever" (p209). In our setting, every agent holding or buying a house at date t^* would willingly do so at price p_{t^*} even if forced to hold it forever. Thus, the equilibrium price at date t^* is the fundamental value by their definition. Our notion of the fundamental value is then consistent with a recursive version of the above definition, i.e. the fundamental value prior to t^* is defined as the price at which all agents either holding or buying the asset would willingly do so even if they were forced to maintain the asset until t^* and received the fundamental value of the asset at this date.

We shall now argue that equilibrium house prices can exceed f_t. Consider first whether $p_t = f_t$ can be an equilibrium. If it were an equilibrium and the interest rate r were close enough to $\beta^{-1}-1$, which will be true when ϕ isn't too large, then both high and low types would prefer to buy houses before date t^* when they arrive: High types value a house at $D > f_t$ and must buy or move on, while low types can assure themselves positive expected profits by buying a house, waiting one period, then selling if new buyers arrive and defaulting otherwise. In fact, prior to date t^*, low types should hold on to a house for as long as their outstanding debt obligation at the end of the period is d: They are always better off waiting one more period and then defaulting if no new traders arrive. Hence, if satisfying equilibrium demand required low types who owe at least d to sell their houses, prices would have to rise above f_t to induce the latter to sell.

Since the mass of housing that is not already occupied by high types at date 0 is ϕ_0, and since it is easy to check that lenders would refuse to lend against more than one house, then for the first $\frac{\phi_0}{n}$ periods there must be some original owners who have yet to sell their houses. Let t^{**} as the smallest integer strictly greater than $\frac{\phi_0}{n}$, so t^{**} is the earliest period in which demand from new buyers might have to be met by agents who purchased their houses after date 0. Recall that it takes $\frac{\phi_0}{(1-\phi)n}$ periods for the number of high types to exceed the stock of housing, so $t^{**} \leq t^*$. If either $t^{**} = t^*$ or mortgage contracts are structured so that the outstanding principal after t^{**} is less than d, there will be no need to induce any low types who owe at least d to sell their homes. In this case, $p_t = f_t$ will indeed be an equilibrium. But if $t^{**} < t^*$ and all agents who arrived between date 0 and t^{**} still owe at least d to their respective lenders, the price will have to exceed fundamentals at date t^{**} for the housing market to clear. Otherwise, all new buyers would wish to buy houses but all existing owners would prefer to hold on to their houses. By backwards induction, if $p_t > f_t$ at $t = t^{**}$, then $p_t > f_t$ for $t < t^{**}$. Otherwise, the original owners would wait to sell at t^{**} than sell at date t, and the market for houses would fail to clear. Thus, the equilibrium price is given by

$$p_t = f_t + b_t \tag{5}$$

where $b_t > 0$ as long as buyers keep arriving, at least until date t^*. That is, we have a bubble that bursts with constant probability q until date t^*, at which point it bursts with certainty. If there is a bubble, then prior to period t^{**}, the equilibrium price path will resemble the stochastically bursting bubble posited in Blanchard and Watson (1982), i.e.

$$b_t = \begin{cases} (1+g)\, b_{t-1} & \text{with probability } 1-q \\ 0 & \text{with probability } q \end{cases} \tag{6}$$

where $g > 0$. This is because up to date t^{**}, the original owners must be indifferent between selling at date t and holding the asset into period $t+1$, entitling them to rents for one more period, and then selling the asset. This indifference implies

$$p_t = \beta \left[\left(\beta^{-1} - 1\right) d + (1-q)\, p_{t+1} + qd \right]. \tag{7}$$

Substituting in $p_t = f_t + b_t$ and using the fact that $f_t = \beta \left[\left(\beta^{-1} - 1\right) d + qd + (1-q)\, f_{t+1} \right]$ yields $b_{t+1} = \left(\beta\,(1-q)\right)^{-1} b_t$. Intuitively, owners who wait to sell are risking losing the chance to sell an asset while it overvalued. As such, they need to be compensated for waiting, and this compensation accrues as additional capital gains from owning the asset. Beyond date t^{**}, the bubble will grow at a rate less than $(\beta(1-q))^{-1}$. This is because agents who value default do not need prices to rise as much if the bubble doesn't burst to be willing to hold the asset. Since some agent who values default must be just indifferent about selling the asset, an agent who does not value default – either because he is unleveraged or owes less than the lowest the house could be worth – would value the asset at less than the price. Thus, beyond date t^{**}, agents with no debt against the asset would strictly prefer to sell it.

In short, a combination of constraints on supply, uncertain demand, and mortgage contracts that repay slowly can lead to speculative bubbles in which houses trade above their intrinsic worth. These bubbles arise when low types who are otherwise indifferent between owning and renting shift to buying, not because they enjoy housing services but because they can profit if house prices go up and default if they do not. Lenders are willing to fund these speculators only because they cannot distinguish them from high types who value their homes and would not default even if prices fall. While we argue house prices are overvalued in some well-defined sense, the model also suggests this notion is subtle and may be hard to detect in practice. First, our model implies speculation occurs when fundamentals appreciate. Thus, if and when a bubble occurs, there will also be fundamental reasons for prices to grow, and one must distinguish between this growth and the part due to speculation. Second, even when house prices are overvalued, high type agents will value houses at above the traded price, which seems to contradict the notion that houses are overvalued. Yet these agents would view houses as too expensive, and if sufficiently patient would delay buying a house if they could until the bubble bursts and the price is expected to fall. Hence, as noted above, forcing high types to buy immediately may be a necessary constraint.

2.3 Mortgage Contract Choice

Finally, we turn to the question of how speculative bubbles affect mortgage choices. Recall that a mortgage corresponds to a set of payments $\{m_\tau\}_{\tau=1}^{T}$ where τ denotes time since the loan originated. A traditional fixed-rate mortgage involves a constant payment

$$m_\tau = \frac{r(1+r)^T}{(1+r)^T - 1} L \quad \text{for } \tau = 1, ..., T \tag{8}$$

We allow lenders to offer both the traditional mortgage above and an IO mortgage. The latter requires the borrower to only pay back interest for the first T_0 periods of the contract, and then repay as under a traditional mortgage with term $T - T_0$. That is,

$$\widehat{m}_\tau = \begin{cases} \widehat{r}L & \text{if } \tau = 1, .., T_0 \\ \\ \dfrac{\widehat{r}\,(1+\widehat{r})^{T-T_0}}{(1+\widehat{r})^{T-T_0}-1}L & \text{if } \tau = T_0+1, .., T \end{cases} \tag{9}$$

We focus on IO mortgages because Barlevy (2009) suggests lenders prefer to offer speculators loans with rising payments. Although various mortgages backload payments, the IO mortgage seems to have been most popular for the period we consider.

In line with our previous discussion, we assume the income stream $\{\omega_\tau\}_{\tau=1}^{\infty}$ allows borrowers to cover all payments under the equilibrium fixed-rate contract in (8) if they choose, i.e. $\omega_\tau > m_\tau$ for $\tau = 1, ..., T$. We now add the assumption that this same income stream falls short of the higher payment required under the interest only contract at the same rate r, i.e. $\omega_t < \widehat{m}_\tau$ for $\tau > T_0$ when $\widehat{r}$ is set equal to r, the equilibrium rate on the fixed-rate mortgage. In other words, we assume that the disposable income available to agents only barely covers the flow of mortgage payments for the fixed-rate mortgage. We offer two justifications for this. First, we argued above that mortgages in which debt is discharged slowly encourage speculation. This implies lenders would have an incentive to limit their loans to short duration mortgages to avoid taking on speculators. The fact that they don't suggests borrowers must already be pretty close to their obligated payments so that such contracts are infeasible.[7] Second, several observers of the mortgage market over the period we study argue that high payments on backloaded mortgages were often onerous for borrowers.[8]

Although borrowers are assumed unable to cover the higher required payment under the IO contract out of their income, borrowers could in principle draw on previous savings to cover the shortfall or refinance to a new mortgage with lower payments. We rule out both possibilities, and assume borrowers must default at date T_0+1. Arguably, neither restriction is consequential. First, with regard to savings, as long as $r > \beta^{-1} - 1$, the value of payments discounted at the risk-free rate will be higher under the IO contract with rate r than under a fixed rate contract with the same rate. Thus, as long as borrowers cannot afford the higher debt obligation of the IO contract, they would have to default at some point. Assuming this happens immediately rather than eventually is simply convenient. As for refinancing, lenders in our environment prefer backloaded contracts because they encourage speculators to sell earlier and repay their loan. If backloaded contracts induce borrowers to refinance

[7]A caveat to this argument is that even if borrowers cannot afford to take on shorter maturity fixed-rate loans, they might still afford backloaded mortgages it heir income grows sufficiently over time. We are implicitly assuming income growth for new arrivals is low enough to rule out this possibility.

[8]See, for example, Congressional testimony by Thompson (2006).

with another lender, this would achieve the same goal and thus only strengthen the appeal of such contracts. The difficulty with modelling refinancing is that lenders would naturally try to make inference about borrowers from the terms of their previous mortgage. Ensuring speculators can refinance in equilibrium even after choosing a backloaded contract requires a more complicated model.

We can now be more precise about our requirement that mortgage contracts retire debt only slowly. Let L_τ denote the amount the agent owes τ periods after taking out the loan. L_τ can be constructed recursively as follows. The initial loan amount L_0 is equal to the price of the house, p_0, which in a slight abuse of notation refers to the price when the mortgage is taken out rather than calendar date 0. Outstanding debt evolves as follows:

$$\begin{aligned} L_0 &= L = p_0 \\ L_{\tau+1} &= (1+r)\,L_\tau - m_\tau \end{aligned} \tag{10}$$

A necessary condition for a bubble is that at date t^{**}, agents value the option of holding on to the asset into date $t^{**}+1$ and defaulting if the value of the house collapses. Default at date $t^{**}+1$ is valuable if

$$L_{t^{**}+1} > \left(\beta^{-1}-1\right) d + d = d/\beta \tag{11}$$

where $L_{t^{**}+1}$ is constructed using the payments m_τ for the fixed-rate mortgage in (8). More generally, we could require this condition hold when we use ω_τ in lieu of m_τ. This would ensure that any feasible contract could leave enough debt at $t^{**}+1$ to make default valuable. In what follows, we assume $L_{t^*+1} > d/\beta$, implying borrowers always prefer to default if prices fall. This simplifies the exposition but is not essential.

To determine which of the two mortgages a speculator prefers, let V_τ denote the expected value to a low type who still owns a house τ periods after buying it and before knowing whether new buyers will arrive at date τ. The speculator can either sell the house and pay $(1+r)\,L_{\tau-1}$; pay m_τ and retain ownership of the house; or default. The payoffs to the three options are $(\beta^{-1}-1)\,d + p_\tau - (1+r)\,L_\tau$, $(\beta^{-1}-1)\,d + \beta V_{\tau+1} - m_\tau$, and 0, respectively. Let τ^* denote the number of periods between when the contract originated and t^*. At $\tau = \tau^*$, the optimal strategy is to sell the asset if new buyers arrive and default if they don't. Hence,

$$V_{\tau^*} = (1-q)\left[\left(\beta^{-1}-1\right) d + D - (1+r)\,L_{\tau^*-1}\right] \tag{12}$$

For $\tau < \tau^*$, V_τ is defined recursively as

$$V_\tau = (1-q)\max\left[\left(\beta^{-1}-1\right) d + p_\tau - (1+r)\,L_{\tau-1}, \left(\beta^{-1}-1\right) d + \beta V_{\tau+1} - m_\tau, 0\right] \tag{13}$$

We can use these equations to compute the value of speculating under each contract. However, recall that under the interest only contract, the speculator would have to either sell or default at date $T_0 + 1$. If $T_0 + 1 < \tau^*$, we must replace the boundary condition (12) with

$$V_{T_0+1} = (1-q) \max \left[\left(\beta^{-1} - 1 \right) d + p_{T_0+1} - (1+\widehat{r}) L_{T_0}, 0 \right] \tag{14}$$

and then use (13) to recursively compute values earlier. Computing the values back to date $\tau = 1$ reveals which contract borrowers prefer when they take out the loan.

As for lenders, let Π_τ denote the expected revenue to the lender τ periods into the loan, before knowing if buyers will arrive at date τ. To allow for foreclosure costs, we assume the lender receives a fraction $\theta \in [0,1]$ of the value of the house after default. Hence,

$$\Pi_{\tau^*} = q\theta \frac{d}{\beta} + (1-q)(1+r) L_{\tau^*-1}$$

For $\tau < \tau^*$, the value Π_τ is given by

$$\Pi_\tau = q\theta \frac{d}{\beta} + (1-q) \pi_\tau$$

where expected profits π_τ if new buyers arrive at date τ depend on what the borrower does. If he sells the asset, $\pi_\tau = (1+r) L_{\tau-1}$. If he remains current on his payments, $\pi_\tau = m_\tau + \beta \Pi_{\tau+1}$. If he defaults, $\pi_\tau = \theta \left[(\beta^{-1} - 1) d + p_\tau \right]$. This assumes the lender sells the house after the borrower defaults. But given our result above that low types who have no liens against the house will always be willing to sell the asset in equilibrium, this will indeed be optimal. Expected profits from funding a speculator are thus $\beta \Pi_1 - L$.

In what follows, we take the price path p_τ as given and examine which mortgage lenders and borrowers prefer. We first consider the case of a bubble, i.e. $p_\tau = f_\tau + b_t$. While in the model the bubble grows at a rate that declines over time, we instead assume the bubble grows at a constant rate until τ^*.[9] That is, we posit a stochastically bursting bubble as in (6), where b_0 and g are parameters. However, we impose that the price path we feed in satisfies the restrictions on the equilibrium price path if there is a bubble: p_τ must be between d and D for all τ, the growth rate g cannot exceed $(\beta(1-q))^{-1} - 1$, and at some point before date τ^* unleveraged agents must strictly prefer to sell the asset at price p_τ.

The model does not yield a simple characterization for when borrowers and lenders prefer certain contracts. But the model is trivial to solve numerically. Solving it for various parameters reveals that when we parameterize p_τ so that unleveraged agents strictly prefer

[9] A price path that grows at a constant rate rather than a rate that declines over time can be sustained if we let the size of the cohorts that arrive decline with time.

to sell the asset, we can usually find some IO contract that both parties prefer to the fixed-rate contract. The intuition for this is as follows. Given our parameterization, the joint interests of borrower and lender are maximized if they sell the asset early, since collectively they have no debt obligations when they purchase the asset, and above we argued that in equilibrium agents with no liens against an asset would at some point strictly prefer to sell the asset. However, the fixed-rate contract encourages the borrower to hold on to the asset because of the option to default on the lender. Hence, both borrower and lender can be made better off if they could agree to sell the asset earlier than under the fixed-rate mortgage. A properly designed IO contract can capture these gains: It forces the borrower to sell the asset by date T_0+1, but at the same time compensates him for this by allowing him to avoid building equity in the asset so that he can default on a larger amount should the house price collapse. The IO contract thus redistributes the gains from selling the asset early to make both parties better off.

As an illustration, consider the following parameterization. Set the mortgage term $T = 30$ and $T_0 = 5$, in line with the modal IO period on mortgages during the period we study. We set $\beta = 0.97$, implying a discount rate of 3% per year. We set the real interest rate $r = 0.04$ to exceed the discount rate. We normalize $d = 1$ and set $D = 30$. The higher this ratio, the more appreciation will be possible in equilibrium. We set q to 0.2, implying the average duration of a bubble is 5 years. For now, we abstract from foreclosure costs and consider the case where $\theta = 1$. To allow for a bubble, we set $b_0 = 0.1$, and let the bubble term grow at a rate $g = 0.05$. The implied growth rate in the price of the asset p_t ranges between 10 and 15% in the first five years. This in on par with the average annual house price appreciation in the top cities in our sample. Finally, we set t^* to 15, i.e. any uncertainty about housing prices would be resolved after 15 years. For these parameter values, both borrower and lender prefer the IO mortgage over the fixed-rate mortgage, i.e. V_t and Π_t at $t = 1$ are higher under the IO mortgage.

The fact that the parties specifically prefer the IO option with $T_0 = 5$ is sensitive to the parameters we choose. For example, if we increased q to 0.5, the lender will no longer agree to the IO contract with $T_0 = 5$, since the odds that the bubble bursts within the first five years are now higher. However, we should still be able to find a shorter IO term that both parties prefer. Indeed, when $q = .5$, both parties will prefer a contract with an IO period of $T_0 = 4$ to the traditional mortgage product. Varying other parameters may similarly lead the borrower of the lender to no longer prefer the IO contract with $T_0 = 5$, but both will still prefer some shorter or longer contract.

If both borrower and lender prefer some IO contract to a fixed-rate contract with the same rate, it seems reasonable that in equilibrium they will use the IO contract. We confirm this in Appendix A, where we show that in equilibrium lenders offer both contracts, and high types choose the fixed-rate mortgage and low types choose the IO mortgage. Furthermore, IO loans carry a higher interest rate in equilibrium. For example, for our numerical example above, given $r = 0.04$ on the IO contract, the equilibrium interest rate on traditional mortgages will be 0.035, or 50 basis points lower. This is on par with the empirical penalty for the IO option.[10] Of course, such perfect sorting is unrealistic; in practice non-speculators may also prefer backloaded mortgages for reasons not captured by our model, such as liquidity constraints. Indeed, if mortgages were perfectly separating, backloaded mortgages would be revealed to be unprofitable, whereas in practice these mortgages were bought and sold at positive prices.[11] Thus, in reality agents must have believed some IO mortgages were profitable. But since our main focus is the implication that speculation encourages the use of IO mortgages rather than what mortgages are offered to those who cross-subsidize losses from speculators, this is not an issue for our analysis.

Next, we turn to the case where prices are equal to fundamentals, i.e. $p_\tau = f_\tau$. Although this means there is no bubble, it need not mean that low types do not engage in speculation. In the case where supply is unconstrained, so $p_\tau = f_\tau = d$ at all dates, there is indeed no way to profit from buying the asset. But when housing supply is constrained and $p_\tau = f_\tau$ where f_t rises over time as in (4), low types would find it profitable buy a house and then sell it if new buyers arrive so that house prices rise and default if no buyers arrive. However, regardless of whether low types speculate, we next show that absent other frictions, moving to an IO contract can no longer make both parties better off.

Proposition: Suppose $p_\tau = f_\tau$ for all dates τ and $\theta = 1$. Then $V_\tau + \Pi_\tau = (\beta^{-1} - 1)\, d + E_{\tau-1}[f_\tau]$ for any contract $\{m_\tau\}_{\tau=1}^{T}$. Hence, if a mortgage contract makes one party better off relative to some benchmark contract, it must make the other party worse off.

Thus, in the absence of a bubble, it will no longer be possible to make both borrowers and lenders better off by using an IO mortgage. This doesn't tell us what type of mortgage will be observed in equilibrium. If $p_\tau = f_\tau = d$, all loans are fully collateralized, so the equilibrium interest rate will equal $\beta^{-1} - 1$. In that case, which contract will be used in equilibrium is inherently indeterminate: When the interest rate r is the same as the risk-free rate, borrowers can save enough before T_0 to meet the higher payments from date $T_0 + 1$ on,

[10] For example, Lacour-Little and Yang (2008) cite a spread of 25 basis points from lender pricing sheets. Jack Guttentag also looks at wholesale prices on mortgages in 2006 and reports a larger spread ranging between 37.5 to 100 basis points. See http://www.mtgprofessor.com/A%20-%20Interest%20Only/how_much_more_does_interest-only_cost.htm.

[11] That said, since many of the participants in the market for mortgage securities were themselves leveraged, the same risk-shifting issues raised here may have inflated the price of risky securities.

and thus are indifferent between the two contracts. This will not be true when $p_\tau = f_\tau > d$, since the interest rate exceeds $\beta^{-1} - 1$. In that case, if the IO contract forces borrowers to default or sell the asset at date $T_0 + 1$, then in equilibrium they will receive the fixed-rate mortgage. A corresponding IO contract will either not be offered or, if offered, will be rejected by speculators.

To summarize, our analysis suggests that if house price appreciation was due to speculative bubbles, we should also observe more IO contracts used in those cities where prices took off and then crashed. In cities where there is no rapid appreciation or price crashes, the model is more ambiguous as to what contracts would be used. Since the model predicts backloaded contracts are used in anticipation of a bubble, we can observe these contracts in bubble cities even before asset prices start to rise. Finally, since our model suggests lenders agree to extend backloaded mortgages only because they encourage borrowers to repay their debt more quickly, we should observe that IO mortgages are more likely to be paid off early. The remainder of the paper confirms these predictions empirically.

3 Data

We now describe the data we use to explore the implications of our model. More details are in the appendix. We describe house prices first, then the mortgages data we use, and finally other variables that *a priori* should affect house prices.

3.1 House Prices

For house prices, we use the Federal Housing Finance Agency (FHFA) house price index, previously known as the OFHEO house price index. The FHFA house price index is compiled quarterly from home prices in mortgages purchased or securitized by Fannie Mae and Freddy Mac, agencies that are subject to FHFA oversight. The FHFA house price index has several advantages. First, it is a repeat-sales index based on the change in price for the same home over time. This makes it less vulnerable to changes in the composition of which houses get sold over time than indices based on the median transaction price. Second, it tracks a large number of cities over a long time period. However, the FHFA house price index also has some well-known shortcomings. For example, since it only includes mortgages purchased or securitized by Fannie Mae and Freddy Mac, it excludes homes that were financed with non-conforming mortgages such as jumbo or subprime loans. Such mortgages are used in computing the rival Case-Shiller index. However, the Case-Shiller index is only publicly available for 20 cities. We did confirm that for these 20 cities, the rate of price appreciation during the boom phase was similar under both indices. This is comforting, since our analysis mostly involves this phase. When we recompute our measure of house price appreciation for each city using the Case-Shiller index, the statistics were similar in magnitude to those using the FHFA index, and the correlation between the two measures is 0.98.[12]

Recall that in our model, cities with speculative price dynamics exhibit faster real price growth until the price collapses. To get at this notion empirically, we first deflated the FHFA index by the consumer price index, and then identified the peak real price between 2000q1 and 2008q4 for each city. We then computed the maximum 4-quarter log real price growth between 2000q1 and the city-specific peak date. Thus, for each city we measure price appreciation during the boom phase using the fastest rate at which real house prices before they reach their peak. However, we also experimented with the average price appreciation between 2000q1 and the peak. As we discuss below, the share of IOs is highly correlated with this measure as well, but this measure also seems more correlated with other characteristics of the mortgage market. Our reason for preferring the maximum 4-quarter growth rate is that it emphasizes especially rapid house price growth concentrated over a short time period. That is, given two cities with the same average growth rate up to the peak, this measure ranks a city in which house prices grow slowly at first but then surge above a city that grows at a stable rate. As a result, the maximal 4-quarter price growth seems to better identify those cities that are most singled out for the boom-bust cycle they experienced. For example, the two cities with the highest maximal 4-quarter price growth are Las Vegas and Phoenix, respectively, yet they rank only 53rd and 57th among all cities in terms of their average price appreciation from 2000 to their peak.

3.2 Mortgages

For data on mortgages, we turned to the Lender Processing Services (LPS) Applied Analytics dataset, previously known as the McDash dataset. The data consists of information on mortgages collected from the servicers that process payments and monitor the status of mortgages on behalf of lenders. The LPS dataset includes data from 9 out of the 10 top

[12] We also explored two other issues concerning the FHFA home price index. First, the index relies on both market transactions and appraised home values from refinancings. Since the FHFA also reports a purchase-only house price index for 25 cities based solely on transaction prices, we compared our price growth measures for those cities where both indices are reported and found they were almost perfectly correlated. Second, the FHFA index uses a simple average of house price appreciation rather than weighting by house value. We therefore looked at the Conventional Mortgage House Price Index, which is essentially a value-weighted version of the FHFA index available for the same large set of cities. Again, we found that our price growth measures were nearly identical to those based on the FHFA index.

mortgage servicers, and has fairly broad coverage of all mortgages, some 60% of the mortgage market by value.[13] However, the dataset is not meant to be a representative sample of mortgages. Indeed, it tends to underrepresent mortgages held by banks on their portfolios, since smaller and mid-size banks often service their own loans and do not report their data to LPS. The dataset also appears to undersample subprime mortgages, which again tend not to be serviced by those outfits that report to LPS. However, since our analysis uses variation across cities, this selection would only compromise our analysis if this selection varies systematically across cities. While we do not have data on the universe of mortgages to address how selection varies across cities, we did compare the foreclosure rates on mortgages in LPS to annual foreclosure rates across a large set of cities published by RealtyTrac in 2007-2009. The latter are based on public records such as default notices and court filings. The foreclosure rates we compute were consistent with those complied by RealtyTrac, and we found no evidence of systematic bias across cities. Since the model suggests speculators are both likely to take out back-loaded mortgages and default if the price collapses, the absence of systematic bias in foreclosure rates across cities is reassuring.

Given that our model suggests speculation would tend to favor back-loaded contracts, we need a measure of how pervasive such contracts are in each city. For this, we first need to take a stand on which mortgages should count as back-loaded. There are several mortgage products that involve unambiguously back-loaded payments. One such mortgage is the graduated payment mortgage, first introduced in the 1970s. As suggested by its name, this mortgage offered payments that gradually increased over the duration of the loan, often during the first five years. However, these mortgages were rarely used in the period we look at. Another contract with back-loaded payments is the IO we discussed in the previous section, in which the borrower only pays interest for some specified period and only then repays both principal and interest. In contrast to the graduated payment mortgage, this product was used quite extensively, at least in certain cities. Finally, another popular but less widely used product is the option-ARM mortgage, in the borrower has the option each month to pay both principal and interest, pay only the interest portion, or pay less than the required interest and add to his total principal, at least up to some maximum amount set forth in the contract. Table 1 reports some characteristics for IO and option-ARM mortgages from the LPS dataset, as well as non-back-loaded fixed-rate and adjustable rate mortgages for comparison.

Although both IO and option-ARM are essentially back-loaded contracts, we chose to focus on IOs in our empirical work. We do this for two reasons. First, the LPS dataset

[13]This estimate of the coverage is reported in **?**.

only began to identify mortgages as IO or option-ARM mortgages from 2005 on. Mortgages that both originated and terminated before January 2005 were not classified. However, we can directly identify which of these mortgages were IO using the scheduled payment, since the scheduled payment would exactly equal the interest rate times the loan amount for IOs. Unfortunately, there is no analogous way to identify option-ARMs, since the scheduled payment can reflect any of the options available to the borrower. Thus, we trust the time series on IOs more than we do the series on option-ARMs.[14] Second, as suggested by Table 1, the two types of mortgages appear to have served somewhat different purposes. In particular, option-ARMs were associated with an extremely high rate of prepayment penalties. This is inconsistent with the notion that back-loaded contracts were designed to induce early repayment of the loan.[15] By contrast, the fraction of IOs with prepayment penalties is only a little higher than the fraction of all mortgages with such penalties. Thus, IOs seem more similar to the type of back-loaded contract that would be mutually preferred according to the model. However, as a robustness exercise we also considered using only option-ARM mortgages and both types of mortgages combined. The results were qualitatively similar.

To measure the extent to which IOs were used in each city, we first computed for each city the fraction of all first-lien mortgages for purchase (as opposed to refinancing) that involved an IO feature. We also considered the share of IOs weighted by loan size, but this ratio turned out to be very similar to the simple share. This is because even though the average loan size for IOs in Table 1 is considerably larger than the average loan size for all mortgages, this is driven by the fact that IOs were more common in more expensive cities. Within cities, the average loan size for IO was more closely related to traditional mortgages. To remain consistent with our approach of measuring house price appreciation using the maximal 4-quarter growth rate, we summarize the use of IOs in each city using the maximum share of these mortgages over the sample period. We constructed similar statistics for other mortgage characteristics, specifically the share of 30-year hybrid mortgages (2/28 and 3/27 mortgages with a short fixed-rate component of 2-3 years followed by an adjustable rate),

[14] In private correspondence, Paul Willen pointed out to us an additional data issue stemming from the fact that option-ARM mortgages before 2003 were largely held in portfolio because lenders were attracted to the fact that these mortgages were readjusted monthly, thus providing a better hedge against interest-rate movements than conventional ARM mortgages that adjusted more slowly and whose changes were capped. Since loans held in portfolio are underrepresented in the LPS, the data is likely to missrepresent the time series pattern for these mortgages. Willen's points are mirrored in press releases from Golden West Financial Corp, one of the leading issuers of option-ARM mortgages prior to 2003, on their website, www.goldenwestworld.com. See, in particular, the note "History of the Option ARM" posted on the site.

[15] Prepayment penalties come in two varieties; hard penalties, which penalize any early repayment, and soft penalties, which waive the penalty if the house is sold. Anecdotal evidence suggests that penalties on option-ARMs were increasingly shifted towards the soft variety, i.e. lenders were increasingly allowing borrowers to sell the asset without penalty. But soft prepayment penalties are still puzzling for our model, since one would expect lenders to want speculators to refinance.

the share of mortgages with a term of 30 years or more, the share of subprime mortgages, the share of mortgages that were privately securitized one year of origination, and the share of mortgages from non-occupant investors. In each case, we use the maximum share of such mortgages in each city over the sample period. An additional mortgage characteristic of interest we already mentioned is leverage. However, given the growing custom of using second "piggyback" loans to cover down payments, measuring the true extent of leverage for our sample period requires knowing the combined loan-to-value (CLTV) of *all* loans against a given property. Unfortunately, LPS does not match first and second liens taken against the same property, and even if it did, the second lien would not be reported if it was serviced by a servicer who did not report to LPS. In lieu of this, we turned to the LoanPerformance ABS database on non-prime privately securitized mortgages. This data is reported by trustees of privately securitized mortgage pools rather than by servicers. Since CLTV data is highly relevant for investors, such data is reported for all mortgages, even when the second lien is not directly in the dataset. Thus, we have some data on leverage for certain types of mortgages, but for a different set of mortgages than for our other mortgage variables. Following Lamont and Stein (1999), we consider the share of mortgages with a CLTV exceeding 80%. Since the subprime mortgage market collapsed soon after the crisis began, the time series does not span the same period as in the LPS dataset. We therefore chose to use the average share of mortgages with CLTV above 80% instead of the maximum share of such mortgages. Table 2 reports descriptive statistics for how the various mortgage variables we use vary across cities in our sample.

3.3 Other Data

Lastly, we compiled various data on factors that should presumably affect house prices and have been used in previous work that has attempted to account for variation in house price appreciation, e.g. Case and Shiller (2003), Himmelberg, Mayer, and Sinai (2005), and Glaeser et al. (2008). Our list of variables includes real per capita income, unemployment, and population, which ought to affect demand for housing; the share of land in each city that is undevelopable because of bodies of water or land terrain that is too steep, which we take from Saiz (2010), and the Wharton Residential Land Use Index that is reported in Gyourko et al. (2008), variables which ought to affect supply of housing; and property tax rates, which affect both supply and demand.

4 Cross-Sectional Evidence

This section describes our evidence on the relationship between the use of back-loaded contracts, specifically IOs, and rapid house price appreciation in the cross-section of cities. Recall that our approach presumes speculative behavior could have emerged in the early 2000s, but only in cities where housing supply was somehow constrained, and we want to explore whether participants in the mortgage market respond differently in cities where speculation could emerge and those where it cannot.

4.1 House Prices and Housing Supply

To distinguish between cities that are and are not vulnerable to speculation, we ranked each city by the share of undevelopable land and by the value of their land use index. We view cities those which rank in the bottom half of both measures as unlikely to experience speculation, and those which rank in the top half of both measures as cities where speculation could have occurred in principle. Figure 2 plots house price appreciation against the maximal share of IOs for both groups. Not surprisingly, cities with few restrictions on supply exhibit low rates of house price appreciation. This was essentially already demonstrated in Glaeser et al. (2008). However, Figure 2 also shows that these cities tended to forgo IOs. For cities where there are some restrictions on supply, there is wide variation in both house price appreciation and the use of IOs. That is, not all cities that were prone to speculation by our classification exhibited a boom-bust cycle. However, house price appreciation and the use of IO contracts does appear to be strongly correlated for these cities. This pattern anticipates one of our findings below, namely that the share of IOs is a better predictor of house price appreciation than data on supply constraints. This because of absence of geographic and regulatory constraints tells us that these cities will exhibit lower rates of house price appreciation, but the presence of constraints does not guarantee that a city will exhibit high rates of house price appreciation. By contrast, knowing that a city tended to rely on IOs quite heavily is a good predictor of whether that city experienced high price appreciation. Consequently, data on the share of these mortgages renders information on supply constraints redundant in predicting house price appreciation.

4.2 Baseline Estimates

Figure 2 is meant to be illustrative. A more rigorous analysis should draw on data for all cities, including those that do not rank in either the top of bottom half of both supply measures, as well as control for other observable city characteristics. This analysis is reported in Table 3. The first column shows that the positive relationship between house price appreciation and IO contracts survives when we expand our sample to all cities for which we have complete data for all our variables. The coefficient on the share of IOs is statistically significant at the 1% level. To help interpret the coefficient 0.416, note that the difference in the share of IOs between the city with smallest such share in our data and the city with the largest such share is equal to $0.609 - 0.017 = 0.592$. Multiplying this by 0.416 implies that the largest log price appreciation we observe in the data should be 0.246 higher than the smallest log price appreciation we observe, which implies that the maximum 4-quarter growth rate in the city with the largest share of back-loaded mortgages will exceed the growth rate in the city with smallest share by $\exp(0.246) - 1 = 27.9\%$. This is comparable to the difference in peak growth rates between Phoenix (36%) and Laredo (7.8%) in Figure 1.

Of course, some of the variation in house price appreciation across cities might be due to differences in other factors that help determine the value of housing services in different cities. The second column in Table 3 includes only these factors, both in levels and in annualized changes from the beginning of our sample and the peak date in each city. The change in population growth, unemployment, and property tax rates all enter significantly with the expected signs, as do the two variables affecting the supply of housing. Interestingly, the R^2 for these variables combined is not much larger than for the share of IOs. In the third column of Table 3, we use these variables as controls when looking at the relationship between house price appreciation and the share of IOs. The coefficient on the share of IOs is smaller than in the first column, although we cannot reject that the two coefficients are equal at the 5% level. More importantly, the coefficient remains tightly estimated and significantly different from zero. Hence, the share of IOs is significantly related to the residual variation in house price appreciation across cities that cannot be explained by the set of typical covariates used to predict house price appreciation. As anticipated earlier, accounting for the share of IOs renders the two variables that capture constraints on housing supply statistically insignificant; given information on which cities rely on IOs, these variables provide little additional information to help predict which cities experience house price booms. In the last column, we add state fixed effects so that our identification relies on variation from cities in the same state. The coefficient falls to 0.225, but remains highly significant.

4.3 The Affordability Hypothesis

Table 3 confirms that home buyers in cities with unusually high price appreciation tended to rely more on IOs. However, this correlation may arise for reasons that have nothing to do with speculation. For example, cities with faster house price appreciation may have faster income growth, and individuals who expect their income to grow may prefer back-loaded mortgages for liquidity reasons even when assets are priced at their fundamental value. The fact that the correlation survives even when we control for growth in per capita income across cities should discount this particular explanation.[16] But an even more mundane explanation for the correlation is that as houses become expensive, borrowers may be forced to resort to mortgage products that remain affordable. IOs have the advantage that they offer very low payments during the IO period, and so they offer at least temporary affordability. By this view, the use of IOs simply mirrors the rapid appreciation of housing prices for whatever reason and need not reflect a strategic response to the presence of speculation. One way to check this is to include the level of housing prices as a control variable to see if it drives out the correlation between house price appreciation and the use of IOs. More precisely, for each city we first took the median price of single family homes reported by the National Association of Realtors at a common date, 2000q1. We then used the rate of real price appreciation in the FHFA to compute the implied price level in each city at its peak.

The first two columns in Table 4 show the effect of adding the log of the price level at the peak.[17] The other control variables in the regression, which are not reported, correspond to the set of explanatory variables in Table 3. When we include the log peak price by itself, this variable has a positive and statistically significant effect, confirming that places with high appreciation ended up as relatively expensive cities compared with price levels in other cities. However, when we add the share of IOs, the coefficient on log peak price becomes statistically insignificant, while the coefficient on the share of IOs remains highly significant, lower than in Table 3 but the difference is not statistically significant. We also considered the ratio of the peak price to per-capita income in the year of the peak as an alternative affordability measure. These are reported in the third and fourth columns of Table 4. The results are the same. Finally, we also examined whether cities that experienced previously high rates of house appreciation might have been more willing to embrace IOs because they were more attuned to issues of affordability. In particular, we focused on the period between 1985 and 1989 when once again there were simultaneous boom-bust cycle in real housing

[16] A related hypothesis is that income is more volatile in cities with faster house price appreciation, and back-loaded contracts allow borrowers to avoid default if they suffer a negative income shock early on. We tried to control for this in two ways. First, we added the variance of annual growth in real log per-capita income between 1969 and 2000 as an additional variable. The coefficient on this variable was positive but not statistically significant, and including it had virtually no effect on the coefficient on IOs or its standard error. However, this variable only captures the volatility of aggregate income. We therefore also tried to control for employment shares by industry for 8 categories (agriculture and mining, construction, manufacturing, transportation and utilities, trade, finance and real-estate, services, and government) since income volatility varies by industry. Even after adding these variables, the share of IOs remained significant at the 1% level.

[17] Using the peak price instead of log peak price yielded similar results.

prices in many U.S. cities.[18] Again, we find that on its own, higher price appreciation in that previous episode is significantly correlated with price appreciation in the more recent period. But this correlation turns insignificant (and with the wrong sign) once we control for the share of IOs, which remains highly statistically significant. While it is true that cities with rapid house price appreciation from 2000 on tended to be places that were ultimately more expensive and which experienced rapid house price appreciation previously, these features are not as good at predicting which particular cities experience rapid price appreciation from 2000-2007 as information on whether borrowers in that city tended to use IOs. For example, some expensive cities like New York and Newark had only modest house price appreciation from 2000 even as relatively inexpensive cities like Boise, Idaho had unusually high house price appreciation. But the latter city tellingly had a higher share of IOs. Essentially, there are too many examples of relatively inexpensive cities with both high price appreciation and shares on IOs and relatively expensive cities with little price appreciation or IOs that our findings can be attributed to affordability.

4.4 The Role of Other Mortgage Characteristics

Table 4 shows that the share of IOs is related to housing prices only through contemporaneous house price appreciation rather than the level of house prices or previous price appreciation. This suggests that there is a connection between the rate of house price appreciation and mortgage choice. But since our discussion so far has focused only on IOs, it is not obvious that back-loadedness is indeed the most prominent mortgage characteristic associated with more rapid house price appreciation. To address this question, we consider several alternative mortgage characteristics that were already summarized in Table 2. The first two mortgage characteristics we consider, the share of hybrid mortgages and the share of mortgages with a term of at least 30 years, are meant to capture alternative affordability products. The next three correspond to various theories that have been proposed for the boom-bust cycle in the housing market: the share of subprime borrowers, the share of mortgages that were privately securitized soon after origination, and the share of highly leveraged mortgages (with a CLTV of at least 80%). Lastly, we also consider the share of mortgages taken out by investors, i.e. individuals who do not plan to occupy the property they are borrowing against. Recent work by Robinson and Todd (2010) suggests that the share of investors in the mortgage pool played a significant role in explaining the rise of foreclosures when house prices began to

[18] We also examined house price appreciation between 1985 and 1999 to capture whether cities had recently become expensive. This rate did worse in predicting price appreciation after 2000, and also had no affect the share of IOs.

decline, and possibly the rate of house price appreciation for some states in the boom phase. It is also noteworthy from Table 1 that investors were over-represented among back-loaded mortgages, so we want to explore whether our back-loaded mortgage measure is really a proxy for the presence of more investors in cities with rapidly rising house prices.

The results for including these mortgage variables are reported in Table 5. We examine the effect of adding each of these alternatives by itself, as well as combining all in a single regression. As evident from the first row of the table, including any one of these variables by itself has no significant impact on the coefficient on the share of IOs. However, when we combine all of these measures, our measured coefficient falls, although it remains statistically significant at the 1% level. Moreover, since the standard error on this variable doubles when we combine all of these alternative measures, we cannot reject that the coefficient is the same as in our benchmark specification. The two affordability measures, the share of hybrids and the share of long term mortgages, are not significant at the 5%, although both are highly significant and positive when we omit the share of IOs. The share of subprime mortgages is statistically significant when it is the only variable we use in addition to the share of IOs. However, the sign of the coefficient is the opposite of what we might expect: cities where there were more subprime borrowers had lower house price appreciation. Moreover, the significance is not robust to including other mortgage characteristics, and the share comes in insignificant when we only use the share of subprime mortgages and drop the share of IOs. As noted above, the negative correlation in column (4) most likely reflects the prevalence of subprime mortgages in lower income cities, while house price appreciation was mostly concentrated in medium and high income cities. Thus, the expansion of the subprime market does not seem to be relevant for explaining why some cities experienced large boom-bust cycles in housing prices while others did not. However, this does not mean that the expansion of the subprime market was unimportant. It may have played an important role in the dramatic rise in foreclosures when house prices declined, and these may be as important if not a more important concern for policymakers as the boom-bust cycle in housing prices.

Turning to the share of mortgages that were privately securitized shortly after origination, we again find it is not statistically significant for explaining house price appreciation once we control for the share of IOs. However, on its own this variable is significant and positively to house price appreciation. The average share of mortgages with a CLTV of over 80% is statistically significant, but again comes with a surprising sign: cities in which where borrowers were more leveraged tended to experience less house price appreciation during the boom phase. One explanation for this is that in cities where there was house price speculation, lenders tried to protect themselves by insisting that borrowers put up larger equity stakes. However, this correlation is not robust, and when we combine all mortgage characteristics together, this variable is no longer statistically significant. Finally, the share of mortgages taken out by investors is significantly correlated with price appreciation, and this correlation is robust to adding other variables. Interestingly, the share of investor mortgages is only weakly correlated with the share of IOs. Thus, the share of investors appears to be a distinct factor that accounts for rapid house price appreciation than the one captured by the share of IOs.

As noted earlier, we also considered the average house price appreciation between 2000q1 and the peak in each city. Most of the results were qualitatively similar for this measure: The share of IOs is statistically significant at the 1% level in predicting this measure of house price appreciation, and including the share of IOs as an explanatory variable renders insignificant the share of undevelopeable area in a city, the Wharton regulatory index, the level of house prices, and the rate of house price appreciation between 1985 and 1989. However, adding other mortgage characteristics produced somewhat different results for this price measure than are reported in Table 5. In particular, the share of mortgages that were privately securitized is statistically significant at the 5% level both when we add it by itself as well as when we include all of our mortgage variables. Moreover, when we include all of our alternative mortgage characteristics together in the same regression, the share of IOs only enters significant at the 5% level rather than at the 1% level, and its coefficient fall by half. Thus, we continue to find a strong correlation between the share of IOs and house price appreciation for this measure, but there appears to be some other factor, possibly associated with private securitization, that also varies with average house price appreciation. This factor also appears to be correlated with the share of IOs, since including all of the factors appears to mitigate the importance of IOs even if it doesn't eliminate it.

4.5 Interest-Only Mortgages During Price Declines

So far, our results show that the use of IOs appears to be strongly associated with rapid house price appreciation before house prices peak. However, the model would suggest that the use of these mortgages would also be associated with rapid declines in house prices if and when house prices collapse. To investigate this feature, we constructed measures of house price decline following the peak in an analogous way to the way we constructed measures of price appreciation before the peak. In particular, for each city we measure the decline as the largest 4-quarter decline between the peak and the end of our sample, i.e. 2008q4. In 31 cities, the highest price recorded occurred in the last quarter in the sample, 2008q4, so there were no period after the peak to analyze. In another 12 cities, the peak occurred sometime in 2008, so the period of decline was not long enough to compute a 4-quarter growth rate. Among the remaining cities in which the peak price occurred before 2008, the largest 4-quarter price declined occurred in the same period in 85% of the cities, between 2007q3 and 2008q3, even though these cities peaked at different dates, some as early as 2003. Nearly all registered their sharpest decline starting in 2007. In what follows, we adopt of the convention of using the negative of the price change for losses, so bigger decline corresponds to a larger number.

Given that many of the cities with rapid house price appreciation also experienced significant subsequent price declines, it is not surprising that we find a similarly strong correlation between the share of IOs and the decline in house prices following the peak to the one we found for house price appreciation. This correlation remains highly significant after controlling for the level of home prices at the peak and the rate of house price appreciation between 1985 and 1989, as in Table 4. As with house price appreciation, the share of IOs renders variables like geographical constraints and previous house price appreciation redundant for forecasting large price declines. However, as was the case when we used our alternative measure of average price appreciation before the peak, there appears to be an additional factor that can predict which cities have large house price declines and which appears to be correlated with the share of IOs. We show this in Table 6, which reproduces Table 5 using the maximum 4-quarter decline following the peak, and where the values of population growth, unemployment, per capita income, and property tax rates are taken for the period following the peak. We now find that almost all of the mortgage characteristics we consider are statistically significant at the 5% level, and some are significant at the 1% level; only the share of mortgages with terms of 30 years or more fails to enter significantly. When we include all of these explanatory variables, only the share of hybrid mortgages and the share of investors are statistically significant at the 5% level. The share of IOs remains significant when we include these additional features, but only at the 5% level rather than the 1% level. Moreover, the coefficient on the share of IOs is less than half the size as when we only control for this mortgage characteristic.

Whatever the additional factor that appears to be associated with rapid declines in prices, then, it seems correlated with most of the other mortgage characteristics we consider, including variables like the share of subprime mortgages which appeared unrelated to house price appreciation before the peak. Since previous work has argued that the growth of the subprime market contributed to the increase in foreclosures, one reasonable candidate for the missing factor is the foreclosure rate once house prices began to decline. In our model, the foreclosure rate would contain no additional information beyond the share of

IOs, since the only home buyers who default in our model are those who speculate. But in practice, there does appear to be significant variation in foreclosure rates that cannot be explained by differences in the propensity to use IOs. As an example, consider Washington DC and Stockton, CA. The two cities have similar shares of IOs at their peak, 47% and 49%, respectively. They also experiences similar rates of house price appreciation, where the maximum 4-quarter growth in house prices was 22.5% and 28.2%, respectively. But the two differ dramatically in terms of foreclosures. Stockton had a foreclosure rate of 9.5% in 2008 according to RealtyTrac, a firm which records foreclosure rates from public records and court notices, the highest in the country. In Washington DC, the foreclosure rate was less than a third, at 3.0%. Consistent with this, house prices fell much more in Stockton than in Washington DC when housing prices generally started to decline: The largest 4-quarter price decline after the peak are quite different in the two cities: 17.3% in Washington DC, but 42.3% in Stockton. Although differences in foreclosure rates are outside the model, there are various reasons for why they might differ. For example, if defaults tend to be more concentrated in certain areas in some cities, they are more likely to drive down the value of neighboring properties facing abandoned buildings. Campbell, Giglio, and Pathak (2009) provide evidence of such spillover effects.

To explore whether the additional mortgage characteristics we consider are significant because they help to predict foreclosure rates, we considered the maximum share of outstanding mortgages each quarter that entered into foreclosure for each city. We then regressed this measure on the maximum share of IOs. The residual from this regression represents the rate of foreclosure that could not be predicted based on the use of IOs. When we add this variable in addition to the list of characteristics we control for in Table 6, the share of hybrid mortgages is no longer significant, and the coefficients on the remaining mortgage characteristics are pushed towards zero. The one mortgage characteristic that does enter significantly is the share of nonprime mortgages with a CLTV of over 80%, but it enters with the opposite sign than when we control for the variable by itself. Moreover, when we tried an F-test all six mortgage characteristics, we could not reject the hypothesis that they were all zero. The share of IOs is now significant at the 1% level. Thus, the factors that seem to best predict which cities will experience large price declines following their peak are whether these cities relied on IOs, even though these were largely taken out before house prices peaked, and whether these cities experienced an unexpectedly large number of foreclosures[19]

[19] As a robustness check, we also looked at whether the maximum foreclosure rate was useful for predicting price appreciation before the peak. Adding this variable did not have much impact on the coefficient on the share of IOs, and while the foreclosure rate was statistically significant at the 5% level when it was the only mortgage variable other than the share of IOs, it was not statistically significant when we included it together with the other six mortgage characteristics we considered.

5 Time Series Evidence

In the previous section we demonstrated that use of IOs is a powerful indicator of whether a city experienced rapid house price appreciation in the period 2000q1-2008q4. Our hypothesis is that this reflects the contract choices of house buyers and mortgage lenders amidst a speculative bubble. However, it could be that back-loaded mortgages are mechanically associated with house price appreciation, because as houses become expensive, borrowers choose such mortgages for affordability reasons. We have already provided some evidence from the cross-section that rejects this interpretation. In this section, we offer additional evidence based on the time series information in our panel of cities.

Our strategy for doing this involves testing for Granger-causality. If use of IOs is driven by past house price appreciation then this appreciation should Granger-cause increased use of IOs. Alternatively, if anticipated house price appreciation leads agents to choose IOs then increased use of these mortgages should Granger-cause house price appreciation. In this latter case we would not interpret Granger-causality as meaning there literally is a causal connection between current use of IOs and future house price appreciation. Rather, consistent with our theory, we would interpret it as meaning simply that increased use of IOs *anticipates* future growth in house prices. This pattern is clearly evident for the case of Phoenix described in the introduction. Is it true more generally?

5.1 Sample of Cities to be Considered

Based on our theory we do not expect use of IOs to Granger-cause house price appreciation in *all* cities. We only expect this in cities where there was an added incentive to use IOs due to a speculative bubble. In these cities we expect to see that these mortgages attained a relatively high share of the market. Therefore in this section we focus on a subset of cities in which the peak share of IO mortages was relatively high, at least 40 percent. This leaves us with a sample of 29 cities. Using 40 percent as the cut-off is clearly arbitrary. Nevertheless, it has the advantage that the dynamics of house prices and mortgages in these cities appears relatively homogeneous. For example, all cities in this sample experience both relatively large price appreciations and peak usage of IOs in the middle quarters of the sample period. Such homogeneity justifies assuming that the coefficients in the statistical models we use to test for Granger-causality are common across these cities thereby increasing the power of these tests. Our results are robust to lowering the IO cutoff to 30%. In addition to restricting our sample of cities by peak IO usage, we focus attention on only the period of house price appreciation leading up to 2006q4. This quarter is the peak of the national

FHFA real price.[20] We focus on the period of house price appreciation instead of the entire boom-bust period since our theory has little to say about use of IOs after prices collapse. In summary, our dynamic analysis is based on a balanced panel of the 29 cities with the largest usage of IOs running from 2000q1 to 2006q4.

5.2 Dynamic Correlations

Some simple correlations demonstrate that the dynamics of IO use and house price appreciation evident for Phoenix is widespread in our sub-sample of 29 cities. Figure 3 displays dynamic correlations between changes in the share of IOs at date $t + j$, Δio_{t+j}, with log changes in real house prices at date t, Δp_t, for $j = -4, -3, \ldots, 4$. The key features of Figure 3 are that the correlations for $j \leq 2$ are positive, the correlations between changes in the IO share and future house price appreciation are larger than the corresponding correlations with past appreciation, and the peak correlation is at $j = -2$. In the language of time series analysis, this pattern of correlations indicates that increased use of IOs *leads* house price appreciation in cities which end up with relatively high shares of IOs. This leading relationship is consistent with IO use rising in anticipation of future house price appreciation.

5.3 Granger-Causality

To test for Granger-causality we estimate simple statistical models of house price appreciation and changes in use of IOs including from one to four lags of these variables. For simplicity we focus on models where the number of included lags is the same for each variable. Table 7 and Table 8 report coefficient estimates, specification tests, and tests of Granger-causality based on estimating the models in two ways, motivated by different underlying assumptions regarding the degree of homogeneity among the cities in our sample. Assuming that housing market dynamics are well-approximated by models with identical constant and slope coefficients, then Ordinary Least Squares (OLS) is appropriate. OLS estimates are reported in Table 7. To accommodate city-specific constants, that is "fixed effects," with homogeneous slope coefficients, we use the System-GMM estimator developed by Arellano and Bond (1991), Arellano and Bover (1995), Blundell and Bond (1998) and Holtz-Eakin, Newey, and Rosen (1988). Our GMM estimates are reported in Table 8.[21]

[20] We have redone our analysis using the first three quarters of 2006 as the end-date of the sample and obtain similar results.

[21] These estimates are computed using the `xtabond2` program for `STATA`, described in Roodman (2009). Our GMM estimates are cluster-robust and include the small sample correction developed in Windmeijer (2005).

5.3.1 OLS Estimates

Consider the OLS estimates in Table 7. Maintaining the homogeneity hypothesis, the results based on OLS are valid only if the estimated residuals are serially uncorrelated. Otherwise the estimates are inconsistent. The tables report p-values for Arellano and Bond (1991) tests of the null hypothesis that the residuals exhibit no serial correlation of order one through four, $AR(j), j = 1, 2, 3, 4$. Using conventional significance levels, these tests indicate that it is not possible to reject serial correlation for at least value of j when predicting house price appreciation or changes in the share of IOs, even with four lags of each variable. We have increased the number of lags to six and serial correlation is still present. The presence of serial correlation is a strong indication that fixed effects are present in our data and that our estimates and standard errors based on OLS are inconsistent.

Since the tests of serial correlation do not have perfect power it is still worth examining the tests for Granger-causality with OLS. Table 7 indicate we can reject the null hypothesis that house price appreciation does not Granger-cause future use of IOs with three lags at the 3 percent level of significance and with two lags at the 4 percent level. (The F-statistics associated with this hypothesis are displayed in the row labelled "F-stat" with the associated p-values below.) In both the two and three lag cases we cannot reject the hypothesis that the coefficients on house price appreciation sum to negative numbers. (The sum of coefficients with associated standard errors are reported in the rows labelled with summations.) Furthermore, the only individual coefficients that are significant are negative. So, while there is some evidence of Granger-causality running from prices to IOs, it is of the wrong sign. Table 7 also indicates the null hypothesis that changes in use of IOs do not Granger-cause house prices is strongly rejected in all cases. While a couple of the IO coefficients are negative, the magnitudes are small, and we cannot reject the hypothesis that the sum of IO coefficients is positive in all four models. Combined these results indicate that increased use of IOs is probably not driven by past house price appreciation and suggest that use of IOs anticipates future house price appreciation, and

5.3.2 GMM Estimates

In addition to rendering the estimates inconsistent, serial correlation in the OLS residuals suggests fixed effects are present. With fixed effects OLS potentially suffers from dynamic panel bias.[22] Our GMM estimates address this phenomenon. System-GMM involves esti-

[22] Whether or not fixed effects are included in the regression, if the data indeed have fixed effects, estimation by OLS mechanically induces correlation between the regression error and any included lagged dependent variable.

mating the coefficients of interest using a system of two equations. One equation involves differencing the original equation to remove the fixed effects and then using lagged variables as instruments. Instrumental variables are necessary in this case because differencing induces correlation between lagged dependent variables and the differenced error term. The second equation involves using lagged differences of the variables as instruments in the original equation. The validity of the instruments in this case is based on the assumption that the differenced variables are orthogonal to the fixed effects. Our GMM estimates in Table 8 are based on using lags three and four of both variables as instruments.[23] The validity of our instruments with GMM depends on the lack of serial correlation in the estimation errors for the differenced equation of order three and higher.[24] Table 8 indicates four lags are necessary for this condition to be satisfied when forecasting house price appreciation. Table 8 indicates two or more lags satisfy the serial correlation criterion when forecasting changes in the use of IOs.[25]

For the three and four lag models of IOs estimated by GMM we fail to reject the hypothesis that prices do not Granger-cause mortgages at greater than the 29 percent significance level. In the two lag case there is modest evidence that the null of no Granger-causality is rejected, at the 7 percent level. Notice, however, that in the two lag case the only coefficent that is significant is negative. Moreover, the sum of coefficients is negative in all three cases and significantly so at the 10 percent level with two lags. So, at best, house price appreciation predicts lower future use of IOs. With the four lag model of house price appreciation estimated using GMM the hypothesis that mortgages do not Granger-cause prices is easily rejected. Furthermore, the two coefficients on lagged IOs in the four lag model that are statistically significant are positive and the estimated sum of all these coefficients is positive, large and statistically significant. Since lags three and four of IOs are not significant we have explored estimating this model with just two lags on mortgages. We obtain similar

[23] All lags at least as large as three are valid instruments. Including all valid lags can lead to a very large number of instruments, which is problematic in small samples. For example, the p-value of the Hansen-Sargan test of the overidentifying restrictions is biased toward unity. Our implementation of System-GMM follows the convention of including orthogonality conditions that are valid at each date. That is, the expectations are evaluated over cities at each date. Consequently, even with just two lags as instruments the instrument count is quite large. While we fail to reject the J test of the overidentifying restrictions with a p-value of unity in all the models estimated by GMM, the actual values of the test statistic are relatively small. This indicates that the non-rejection of the overidentifying restrictions is not driven by noise in the data. Evaluating expectations over cities and dates dramatically reduces the number of instruments. There does not appear to be agreement in the empirical literature on the best way to proceed. However, our findings are robust to following this approach. Our findings are also robust to using just one lag as an instrument.

[24] Differencing induces first order serial correlation and the fact that the original equations already involve first differences suggests up to second order serial correlation is possible.

[25] We have explored using the lag selection criterion for dynamic panel data models introduced by Andrews and Lu (2001). For both house price appreciation and change in use of IOs this leads to the one lag model being chosen.

results. So, the evidence points toward our hypothesis that increased use of IOs occurs in anticipation of, and not because of previous, house price appreciation.

It is common in dynamic panel data analysis to include dummies for each time period in the estimation. Such a practise is typically justified by the concern that predictability at the individual level is not a direct manifestation of individual decisions, but is driven by an aggregate relationship. In our sample of 29 cities the run-up in house prices occurred in all cities around the same time, so including time dummies renders the other regressors insignificant. It is conceivable that innovation in credit provision is a leading indicator of the business cycle and house prices are a lagging indicator. If it just so happened that IOs were a product of such innovation then the relationship we observe at the city level could just reflect business cycle dynamics. We are skeptical of this hypothesis because it does not account for why use of IOs was concentrated in those cities that had the most house price appreciation. Nevertheless, we have considered including lags of log GDP growth and the Federal Funds interest rate in our statistical models and it does not change our findings.

6 Mortgage Pre-payment and Foreclosure

Our theory implies that lenders prefer IOs over traditional mortgages because they encourage borrowers who may be speculators to re-pay their mortgages sooner rather than later. An obvious check on our theory, then, is to assess whether the IOs in our sample are pre-paid at a faster rate than other kinds of mortgages. However confirmation of this feature of our model it does not prove our theory. For example, it could be that households who intend to move relatively soon choose IO mortgages. Still, if we were to find that IOs are pre-paid at a slower rate than other mortgages it would be evidence against our theory. Our theory also predicts that when prices collapse holders of IO mortgages who are speculators will default. This suggests looking at foreclosure in our data as well.

We estimate pre-payment and foreclosure rates using the Kaplan-Meier estimator, which is commonly used to calculate unemployment durations. For simplicity we focus on pre-payment, but the same approach applies to foreclosures. Let $S(t)$ be the probability that payments on a particular kind of mortgage originated at a fixed date continue at least until date t. For a sample from this population of size N, let the observed times until pre-payment be t_i, $i = 1, 2, \ldots, N$. Corresponding to each t_i is the number of mortgages that are at risk of being pre-paid at date $t_i - 1$, n_i, and the number of pre-payments d_i. The Kaplan-Meier estimator is the nonparametric maximum likelihood estimate of $S(t)$

$$\hat{S}(t) = \prod_{t_i \leq t} \frac{n_i - d_i}{n_i}$$

Our estimator of the pre-payment rate is $1 - \hat{S}(t)$. Note that we include refinances in our classification of pre-paid since lenders are indifferent between pre-payment due to selling the house and refinancing. An important advantage of the Kaplan-Meier estimator is that the method takes into account right-censoring, which occurs if a mortgage leaves our sample before pre-payment is observed. This can happen when the mortgage is sold to a servicer outside the sample or the mortgage is defaulted on, i.e. foreclosed. We do not report standard errors for our estimates since they are very small relative to the point estimates, due to the large size of the sample on which they are based.

Figure 4 displays estimated pre-payment and foreclosure rates for two types of mortgage: IO and non-back-loaded. By "non-back-loaded" we mean all mortgages that are neither IOs or Option-ARMs. The estimates are based on mortgages originated in 2005q1 for our sample of 29 cities with maximum share of IOs exceeding or equal to .4. This has the advantage of including a relatively large sample of IOs. Our estimates are similar for other quarters of origination in 2005 and 2004. They are less similar for data in 2003 where the sample is small.

Figure 4 indicates that pre-payment and foreclosure rates are higher for IOs than non-back-loaded mortgages. After 8 quarters 40 percent of IOs have been pre-paid while less than 30 percent of non-back-loaded mortgages are pre-paid. The difference in pre-payment rates stays roughly equal through 20 quarters. The foreclosure rates are indistinguishable through 8 quarters after origination and then start to diverge during 2007. By 2009q2 the difference in foreclosure rates is more than 10 percentage points. Using the log rank test of equality we confirm that the differences between the survival functions, $\hat{S}(t)$, are statistically significant at very high levels of significance. We conclude that our data is consistent with the theory along these dimensions.

7 Conclusion

In this paper we have argued that both speculators and those who lend to them prefer mortgages with back-loaded payments, in particular interest-only mortgages, over traditional mortgages when there is speculation in the housing market and lenders cannot determine which borrowers are speculators. This insight motivated our analysis of house prices and mortgages for a sample of US cities over the period 2000-2008. Our main findings are that IO usage is a robust predictor of which cities experienced high price appreciation; subprimes, securitization, leverage do not seem important for predicting price appreciation; that IOs do not seem to be used because of high house prices, but that IOs anticipate future price appreciation; following periods of high IO origination, IOs pre-paid and foreclosed at higher rates than mortgages without back-loaded payments. We think our findings represent compelling evidence that speculation was an important reason for the rapid acceleration in house prices in some cities over the period 2003 to 2006.

Some may interpret our findings to mean that back-loaded mortgages such as IO loans were essential for sustaining speculation, and so policymakers concerned about speculation should forbid these contracts. Our analysis does not allow us to say what would happen counterfactually if such a rule were in place. However, we note a few things. First, back-loaded contracts do serve a useful purpose in some cases, such as helping those expecting future income growth (young people starting their careers). Second, back-loaded contracts may be the "canary in the coal mine" in terms of anticipating future price increases, as our empirical evidence suggests. Finally, Barlevy (2009) suggests forbidding back-loaded contracts may not prevent speculation, and since back-loaded contracts encourage agents to sell, eliminating them may end up causing asset prices to be even more overvalued.

Appendix A: Proofs of Theoretical Results

Equilibrium Contracting with Bubbles: In this section, we show that if given the equilibrium interest rate on a traditional mortgage, there exists an IO mortgage that low type borrowers and lenders both prefer, then (1) both traditional and IO mortgages will be offered in equilibrium, with low-types choosing the IO contract and high types choosing the traditional mortgage; (2) IO contracts will carry a higher interest charge in equilibrium.

First, we argue that low-types must receive IO contracts in equilibrium. For suppose not, i.e. they receive a traditional mortgage contract with interest rate r^*. We first argue that r^* must exceed $\beta^{-1} - 1$ to ensure non-negative profits. If r^* were equal to $\beta^{-1} - 1$, high types could choose this contract as well. Thus, in equilibrium, high types must find the contract they choose at least as profitable as choosing the traditional mortgage with rate $\beta^{-1} - 1$. Since a traditional mortgage at rate $\beta^{-1} - 1$ allows the high types to keep their homes indefinitely and requires them to pay the opportunity cost of funds, lenders will not be able to earn positive profits from high types. Since expected profits from low types are negative given they will choose to default, lenders will not be able to earn positive profits. Hence, $r^* > \beta^{-1} - 1$.

Now, consider a lender who offers the same set of contracts in equilibrium, but also offers an IO contract with interest rate $\widehat{r}^*$. High types will not choose this contract, since given $\widehat{r}^* > \beta^{-1} - 1$, the present discounted value they would have to pay under the traditional contract is lower. Moreover, they might have to default on the house when payment rise. Hence, these types will stick with whatever contract they were originally choosing in equilibrium. However, both borrowers and lenders are better off under the IO contract. Since lenders earn zero profits in equilibrium, this implies the lender who offers the IO contract at rate $\widehat{r}^*$ will earn a strictly positive profit. Hence, the original contracting arrangement could not have been an equilibrium.

Next, we argue that high types will receive the traditional mortgage in equilibrium. For suppose in equilibrium all borrowers took the IO mortgage with interest rate r^*. Consider a lender who offers a traditional mortgage at rate r^*. Such a lender would not attract low types. This is because by assumption, low types prefer the IO contract at whatever rate is charged on traditional mortgages in equilibrium. That rate must be higher than r^*, or else high types would have chosen it. But since low types will find IO contracts even more attractive at lower rates, they must also prefer the IO mortgage at rate r^*. Thus, a lender offering a traditional mortgage with rate r^* will not attract low types. But he will attract high types. Since lending to high types at an interest rate that exceeds the discount rate yields positive profits, such a lender will earn a strictly positive profit. But then the original contracting arrangement could not have been an equilibrium.

Hence, both contracts will be offered in equilibrium. We now argue that in equilibrium, low types must be indifferent between the types of mortgages contracts in equilibrium. For suppose not, i.e. low-types strictly prefer the IO contract. Consider a lender who offers only the traditional mortgage contract offered in equilibrium, but lowering the interest rate by ε. Given low-types strictly prefer the IO contract, there exists an ε such that they would still prefer the IO contract. High types will prefer this contract. But there exists an ε small enough that the lender offering this contract and attracting only the safe borrowers will earn a strict profit.

Since low types prefer the IO contract with rate r^* to the traditional mortgage contract with rate r^*, the only way to ensure they are indifferent between the two contracts is to charge a lower rate on the traditional mortgage in order to make it more attractive. Hence, the equilibrium rate on the traditional mortgage contract will be lower than on the IO contract. To solve for this rate, we need to find the interest rate such that V_1 at this rate is equal to V_1 under the IO contract with rate r^*. For the same reason as above, the equilibrium interest on both mortgages must exceed $\beta^{-1} - 1$, since otherwise expected profits to the lender will be negative.

Proof of Proposition: Substituting in the fact that $p_\tau = f_\tau$, it follows from the definition of $V_\tau + \Pi_\tau$ that if either the borrower sells the asset or defaults, this sum is equal to $(\beta^{-1} - 1) d + E_{\tau-1}[f_\tau]$. If the borrower keeps making payments, we have

$$V_\tau + \Pi_\tau = \left(\beta^{-1} - 1\right) d + \beta \left(V_{\tau+1} + \Pi_{\tau+1}\right)$$

For this case, we need to separately consider the two cases, $f_\tau = d$ for all dates and f_τ given by (4). In the first case, note that at $\tau = T + 1$, all debt would have been retired. Hence, $V_{T+1} = (\beta^{-1} - 1) d + d$ and $\Pi_{T+1} = 0$. It follows that

$$\begin{aligned} V_{T+1} + \Pi_{T+1} &= \left(\beta^{-1} - 1\right) d + d \\ &= \left(\beta^{-1} - 1\right) d + E_{T-1}[f_T] \end{aligned}$$

Next, suppose $V_s + \Pi_s = (\beta^{-1} - 1) d + E_{s-1}[f_s]$ for $s = \tau + 1, ..., T + 1$. Then at date τ, we have

$$\begin{aligned} V_\tau + \Pi_\tau &= \left(\beta^{-1} - 1\right) d + \beta \left(\left(\beta^{-1} - 1\right) d + E_\tau[f_{\tau+1}]\right) \\ &= \left(\beta^{-1} - 1\right) d + \beta \left(\left(\beta^{-1} - 1\right) d + d\right) \\ &= \left(\beta^{-1} - 1\right) d + E_{\tau-1}[f_\tau] \end{aligned}$$

This establishes the claim for $f_\tau = d$.

Next, let $f_\tau = \sum_{s=\tau+1}^{\tau^*} \beta^{s-t} (\beta^{-1} - 1) d + \beta^{\tau^* - \tau} E_\tau [p_{\tau^*}]$, where p_{τ^*} is equal to D with probability $1 - q$ and d with probability q. At date τ^*, since all uncertainty is resolved, meeting payments will be weakly dominated by selling the asset and strictly dominated if $r > \beta^{-1} - 1$. Regardless of whether the borrower defaults or not, we have

$$\begin{aligned} V_{\tau^*} + \Pi_{\tau^*} &= \left(\beta^{-1} - 1\right) d + E_{\tau^*-1}[p_{\tau^*}] \\ &= \left(\beta^{-1} - 1\right) d + E_{\tau^*-1}[f_{\tau^*}] \end{aligned}$$

Finally, suppose $V_s + \Pi_s = (\beta^{-1} - 1) d + E_{s-1}[f_s]$ for $s = \tau + 1, ..., T + 1$. Then at date τ, we have

$$V_\tau + \Pi_\tau = \left(\beta^{-1} - 1\right) d + \beta E_{\tau-1} \left(\left(\beta^{-1} - 1\right) d + E_\tau[f_{\tau+1}]\right)$$

However, from the definition of f_τ, we have

$$
\begin{aligned}
f_\tau &= \sum_{s=\tau+1}^{\tau^*} \beta^{s-\tau}\left(\beta^{-1}-1\right)d + \beta^{\tau^*-\tau}E_\tau\left[p_{\tau^*}\right] \\
&= \beta\left(\beta^{-1}-1\right)d + \beta\left[\sum_{s=\tau+2}^{\tau^*} \beta^{s-\tau-1}\left(\beta^{-1}-1\right)d + \beta^{\tau^*-\tau-1}E_\tau\left[p_{\tau^*}\right]\right] \\
&= \beta E_{\tau-1}\left[\left(\beta^{-1}-1\right)d + \sum_{s=\tau+2}^{\tau^*} \beta^{s-\tau-1}\left(\beta^{-1}-1\right)d + \beta^{t^*-\tau-1}E_{\tau+1}\left[p_{\tau^*}\right]\right] \\
&= \beta E_{\tau-1}\left[\left(\beta^{-1}-1\right)d + E_\tau\left[f_{\tau+1}\right]\right]
\end{aligned}
$$

This allows us to rewrite the sum $V_\tau + \Pi_\tau$ as

$$V_\tau + \Pi_\tau = \left(\beta^{-1}-1\right)d + E_{\tau-1}\left[f_\tau\right]$$

which establishes the claim. ■

Appendix B: Data

In this appendix, we describe our data sources and how we constructed the variables used in our empirical work.

B1: House Price Data

Our primary data source for housing prices is the Federal Housing Finance Agency (FHFA) house price index, formerly the OFHEO House Price Index, for metropolitan staitical areas (MSAs). MSAs are defined by the Office of Management and Budget (OMB). If specified criteria are met and an MSA contains a single core population greater than 2.5 million, the MSA is divided into Metropolitan Divisions.[26] For these MSAs, FHFA reports data for each Division, rather than the MSA as a whole. We follow this convention thruought, using Metropolitan Divisions instead of MSA where applicable.

To arrive at a measure of changes in real house prices over time, we divide the FHFA index by the consumer price index for urban consumers for all items, as reported by the Bureau of Labor Statistics.

Because of data limitations for mortgage data before 2000, we begin our dataset in the first quarter of 2000, even though the FHFA index is available earlier. However, we do use

[26] The following MSAs have been divided into Metropolitan Divisions: Boston-Cambridge-Quincy, MA-NH; Chicago-Naperville-Joliet, IL-IN-WI; Dallas-Fort Worth- Arlington, TX; Detroit-Warren-Livonia, MI; Los Angeles-Long Beach-Santa Ana, CA; Miami- Fort Lauderdale-Miami Beach, FL; New York-Northern New Jersey-Long Island, NY-NJ-PA; Philadelphia-Camden-Wilmington, PA-NJ-DE-MD; San Francisco-Oakland-Fremont, CA; Seattle-Tacoma-Bellevue, WA; and Washington-Arlington-Alexandria, DC-VA-MD-WV

earlier data to construct a measure of real house price appreciation between 1985q1 and 1989q4. Given limitations on some of the explanatory variables we use, such as real income per capita, we only consider data through the end of 2008. To simplify the notation, we index quarters by t and refer to the initial date, 2000q1, as $t = 1$ and the terminal date, 2008q4, as $t = T$.

For each city i, we identify the quarter where the real price of housing reaches its peak. That is, if p_{it} denotes the real house price index in city i at date t, then

$$t_i^* = \arg\max_t \{p_{it}\}$$

To simplify notation, we will henceforth omit the reference to city i. As our measure of real house price appreciation before the peak, we use the highest 4-quarter growth in real house prices up date t^*. That is, the rate of house price appreciation in the boom phase is given by

$$\max \left\{ \ln \left(\frac{p_t}{p_{t-4}} \right) \right\}_{t=5}^{t^*}$$

for all cities where $t^* \geq 5$.

In addition to the maximum rate of appreciation, we also calculate average annualized real house price appreciation prior to peak as

$$\frac{4}{t^* - 1} \ln \frac{p_{t^*}}{p_1}$$

for each city.

To study the post-peak period, we calculate the largest 4-quarter decline after the peak as

$$\max \left\{ \ln \left(\frac{p_{t-4}}{p_t} \right) \right\}_{t=t^*}^{T}$$

where $T \geq t^* + 4$.

Since the FHFA is an index, we need to turn to alternative sources for data on price levels. We chose to use the median sales price of single family homes in each city as reported by the National Association of Realtors in 2000q1. Denote this price as P_1^{NAR} to reflect that we only use this price from date $t = 1$. To compute the level of the peak price in each city in a way that would not be vulnerable to changes in the composition of housing that are brought to the market, we use the appreciation rate in the FHFA index. That is, the peak price in each city is given by

$$P_{t^*} = P_1^{NAR} \times \frac{p_{t^*}}{p_1}$$

When we use this price as a control in Table 2, we use $\ln P_{t^*}$.

B2: Mortgage Data

Our primary data source for mortgage information is the Lender Processing Services (LPS) mortgage performance data, formerly McDash, reported by loan servicers for the mortgages they service. The LPS data is reported monthly. We refer to these files herein as the dynamic files, as they are used to track mortgages through time. From these dynamic files, we construct a static file that contains information for each loan we ever observe. This is generally the file we use to construct our mortgage data. Any exceptions will be noted.

One issue we have to grapple with in this dataset is that different servicers begin to report to LPS at different dates. When a servicer first begins to report, they report on all of their outstanding loans. For loans that were originated before the servicer began to report, the relevant data is backfilled, i.e. data for quarters before the servicer began to report are entered not in real time but retrospectively. This pattern of reporting gives rise to survivorship bias, since the only loans that are reported retrospectively are those which survive until the reporting period. One approach that some researchers have taken to address this, e.g. Foote, Gerardi, Goette, and Willen (2010), is to only include mortgages that are reported within a short time of when they were originated, e.g. a few months or one year. However, restricting attention to data that is reported in real time can give rise to composition bias if the servicers that report to LPS later tend to specialize in certain types of loans. In particular, throwing out backfilled data appears to undercount the share of back-loaded mortgages earlier on in our sample, since these mortgages appear to have been disproportionately originated by services who report to LPS at later dates. Since we only use data starting in 2000q1, we opted to use backfilled data on mortgages to avoid composition bias. This restriction has little bearing on our cross-sectional analysis, which mostly relies on origination in the peak years of 2005 and 2006 when most services were already reporting to McDash. The restriction does matter for the time-series analysis. However, since we find that IO mortages were less likely to survive, the survivorship bias in our data biases against our findings. Looking only at surviving mortgages will tend to underrepresent the fraction of mortgages originating in previous years that would have involved a back-loaded payment pattern.

The primary mortgage variable of concern to us is the (quarterly) share of IO mortgages. To identify IO mortgages, we use the IO flag (IO_FLG) which is derived from the payment frequency type variable (PMT_FREQ_TYPE). Unfortunately, LPS only started collecting data on whether loans are IO in 2005. Therefore, the IO flag is only available for loans that were still being serviced in 2005 or later. To identify IO loans that were both issued and discharged prior to 2005, we use the payment ammount. IO loans should have payment ammounts approximately equal to the principal times the monthly interest rate. We use a close band around the exact value to compensate for variable month lengths, rounding errors and other small deviations. Testing this method on the post-2005 data proved highly effective at identifying IO loans.[27] Specifically, in our test sample we correctly identify 98.5% of IO mortgages while falsely identifying about 1.5% of non-IO loans as IO. In our cross-sectional analyses, we use the maximum value of the IO share over the period. Maximums are constructed analogously for each of the mortgage variables described below.

Building our other mortgage variable is simpler, since we do not have the non-reporting is-

[27] For an unknown reason, one of the servicers reports payments for IO mortgages as twice the principal times monthly interest. Therefore, we also use a tight band arround this value to identify some additional IO loans

sues that we had with IOs. From the LPS data, we construct the share of "hybrid" mortgages (2/28 and 3/27 loans, i.e. loans that have a fixed rate for the first two or three years and then adjust) from the ARM initial rate adjustment period (variable FIRST_RATE_NMON). We use the amortization term at origination (TERM_NMON) to identify the share of "long term" mortgages, those that have a term of at least 30 years. We identify subprime (Grade 'B' or 'C') mortgages by the grade from the mortgage type (MORT_TYPE). This classification is given by the servicers themselves and is the same one used by Foote et al. (2010). Mortgages with pre-payment penalties are identified by a pre-payment penalty flag (PP_PEN_FLG) as reported by the servicer.

In calculating the share of privately securitized mortgages, we are concerned with whether or not a mortgage was packaged and sold off in the first twelve months after origination. In order to see this, we need to turn to the dynamic LPS files. There are two cases that need to be considered. If a mortgage has lasted at least twelve months before being discharged, in which case we use the investor status (INVESTOR_TYPE) at twelve months after origination. If the mortgage was discharged within the first eleven months after origination, we use the investor status at the last date at which we observe the mortgage in the dynamic files. As above, in our cross-sectional analysis we use the maximum value taken over the period.

We identify the foreclosure share in a quarter as the ratio of mortgages that are actively reporting in a quarter who report being in foreclosure for the first time over the number of mortgages actively reporting. As above, we use the highest value over our period of interest in the cross sectional regressions.

B2.1: LP Mortgage Data

Since first mortgages in the LPS data set are not matched to their second leins, we are unable to calculate a combined loan-to-value (CLTV) ratio from that data. Therefore, in order to calculate the share of martgages with a CLTV above eighty percent we look to a different data set, the LoanPerformance subprime data set, from First American CoreLogic, which covers a different universe of mortgages[28], but which matches first lien mortgages to second lein (and above) mortgages, allowing us to calculate CLTVs.

B3: Other Data

Finally, we compile data on cities from various sources that we can use to control for differences across cities that may directly account for differences in prices. We gather many of these variables from the Haver Analytics Database. In these cases, we will specify where Haver Analytics gets their data. These variables are included in the cross-sectional regressions as both levels and growth rates, as explained below.

Not every data source uses the same geographical breakdown as the FHFA. In these cases we use translation tables from MABLE/Geocorr2K, the Geographic Correspondence Engine

[28] The LoanPerformance subprime data set consists of the underlying collateral of pools of private-label subprime securitizations

with Census 2000 Geography from the Missouri Census Data Center.

B3.1: Population

Population for each city is from the Census Bureau's Current Population Reports, P-60 , via Haver, and available annually. For the level in the cross-sectional analysis for price increases, we use the log of the mean level of population from t to t^*. For decreases we use the log of the mean level of population from t^* to T. For the growth rate durring the rise in prices we use the average annual growth over the period, and for the rate of decline durring the rise in prices we use the average annual decline over the period, calculated, respectively, as

$$\ln(\frac{pop_{t^*}}{pop_1})\frac{4}{t^*-1} \quad \text{and} \quad \ln(\frac{pop_T}{pop_{t^*}})\frac{4}{T-t^*}$$

where pop_t is the population at date t.

B3.2: Real Per Capita Income

MSA-level per capita personal income is from the Bureau of Economic Analysis (BEA), via Haver, and available annually. This series is deflated by the chained price index for personal consumption expenditures. For the level in the cross-sectional analysis, we use the log of the mean level of real per capita personal income from t to t^*. For decreases we use the log of the mean level of real per capita personal income from t^* to T. For the growth rate durring the rise in prices we use the average annual growth over the period, and for the decline we use the average annual decline over the period, calculated as population above.

B3.3: Unemployment

The MSA-level unemployment rate is from the Bureau of Labor Statistics (BLS), via Haver, and available monthly. We aggregate to a quarterly frequency using a simple average of the months in a quarter. For the level, we use the mean rate (in percentage terms) from t to t^*. For declines we use the mean rate from t^* to T. For the growth rates, we use the average annual change over the period (rise or fall), ie.

$$\ln(\frac{U_{t^*}}{U_1})\frac{4}{t^*-1} \quad \text{and} \quad \ln(\frac{U_T}{U_{t^*}})\frac{4}{T-t^*}$$

where U_t is the unemployment rate at time t.

B3.4: Property taxes

Our level version of the property tax measures are median tax rate (taxes paid / house value) in 2000 for the increases and median tax rate at t^* for the decreases. For the change in tax rates over the period, we use the average annual change in tax rates between $t = 1$ and t^* for the increase and the average annual change between t^* and T for the decrease, calculaved as

$$(tax_{t^*} - tax_1) * \frac{4}{t^* - 1} \quad \text{and} \quad (tax_T - tax_{t^*}) * \frac{4}{T - t^*}$$

respectively, where tax_t is the tax rate at time t.

Data was extracted from the American Community Survey (ACS) from the US Census Bureau using an extract request from IPUMS USA [http://usa.ipums.org]. The variables of interest from the ACS are annual property taxes (PROPTX99) and House value (VALUEH). PROPTX99 is a categorical variable, so there are data ranges. We use mid-level values, so for example where PROPTX99 is equal to 66, we code property tax as 7500, since the value of 66 in PROPTX99 indicates a range of \$7,001-\$8,000. Anything above \$10,000 is simply coded as \$10,000+ (proptx99==69). This doesn't affect our analysis since we use median property taxes.

Since the ACS has its own definition of metro areas, we need to use the IPUMS metro area-to-MSA/PMSA translation table and then use a MSA/PMSA-to-CBSA table from GEOCORR2K. We also weight households by household weight (HHWT).

B3.5 Static Variables

We use two static variables in the cross-sectional analyses. The first is the Wharton residential land use Regulation Index (WRI) from Gyourko, Saiz, and Summers (2008), a summary measure of the stringency of the local regulatory environment in each community. The second static variable is undevelopeable area (undevarea) from Saiz (2010). Undevarea is meant to be a precise measure of exogenously undevelopable land in cities estimated from satelite data. Both of these variables are meant to capture the relative supply elasticity of each city.[29]

[29] Thanks to Albert Saiz for providing the necessary data to construct both of the series.

References

Allen, F. and D. Gale (2000). Bubbles and crises. *The Economic Journal 110*(460), 236–255.

Allen, F. and G. Gorton (1993). Churning bubbles. *Review of Economic Dynamics 60*(4), 813–836.

Allen, F., S. Morris, and A. Postlewaite (1993). Finite bubbles with short sale constraints and asymmetric information. *Journal of Economic Theory 61*, 206–229.

Amromin, G., Huang, and Sialm (2010). Mortgage complexity and house price dynamics. Chicago Fed manuscript.

Andrews, D. W. and B. Lu (2001). Consistent model and moment selection procedures for GMM estimation with application to dynamic panel data models. *Journal of Econometrics 101*, 123–164.

Arellano, M. and S. Bond (1991). Some tests of specifiation for panel data: Monte carlo evidence andan application to employment equations. *Review of Economic Studies 58*, 277–297.

Arellano, M. and O. Bover (1995). Another look at the instrumental variables estimation of error-components models. *Journal of Econometrics 68*, 29–51.

Barlevy, G. (2009). A leveraged based model of speculative bubbles. Chicago Fed manuscript (revised version of Working Paper 2008-01).

Blanchard, O. and M. Watson (1982). Bubbles, rational expectations, and financial markets. In P. Wachtel (Ed.), *Crisis in the Economic and Financial Structure: Bubbles, Bursts, and Shocks*. Lexington Books.

Blundell, R. and S. Bond (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics 87*, 11–143.

Burnside, C., M. Eichenbaum, and S. Rebelo (2010). Understanding booms and busts in housing markets. Northwestern University manuscript.

Campbell, J., S. Giglio, and P. Pathak (2009). Forced sales and house prices. NBER Working Paper 14866.

Case, K. E. and R. J. Shiller (2003). Is there a bubble in the housing market? *Brookings Papers on Economic Activity 2003*(2), 299–342.

Chambers, M., C. Garriga, and D. E. Schlagenhauf (2009). Accounting for changes in the homeownership rate. *International Economic Review 50*(3), 677–726.

Corbae, D. and E. Quintin (2009). Mortgage innovation and the foreclosure boom. University of Texas at Austin manuscript.

Foote, C., K. Gerardi, L. Goette, and P. Willen (2010). Reducing foreclosures: No easy answers. In D. Acemoglu, K. Rogoff, and M. Woodford (Eds.), *NBER Macroeconomics Annual 2010*, Volume 204, pp. 89–138.

Ghent, A. C. and M. Kudlyak (2009). Recourse and residential mortgage default: Theory and evidence from US states. Federal Reserve Bank of Richmond Working a[er No. 09-10,.

Glaeser, E. L., J. Gyourko, and A. Saiz (2008). Housing supply and housing bubbles. *Journal of Urban Economics 64*(2), 198–217.

Gyourko, J., A. Saiz, and A. Summers (2008). A new measure of the local regulatory environment for housing markets. *Urban Studies 45*(3).

Himmelberg, C., C. Mayer, and T. Sinai (2005). Assessing high house prices: Bubbles, fundamentals and misperceptions. *The Journal of Economic Perspectives 19*(4), 67–92.

Holtz-Eakin, D., W. Newey, and H. Rosen (1988). Estimating vector autoregressions with panel data. *Econometrica 56*, 1371–1395.

Kiyotaki, N. and J. Moore (1997). Credit cycles. *Journal of Political Economy 105*(2).

Krugman, P. (2005, August 8). That hissing sound. *New York Times*.

Lacour-Little, M. and J. Yang (2008). Pay me now or pay me later: Alternative mortgage products and housing consumption. California State University at Fullerton manuscript.

Lamont, O. and J. C. Stein (1999). Leverage and house-price dynamics in us cities. *RAND Journal of Economics 30*(3), 498–514.

Mian, A. and A. Sufi (2009). The consequences of mortgage credit expansion: evidence from the US mortgage default crisis. *Quarterly Journal of Economics 124*(4), 1449–1496.

Robinson, B. and R. Todd (2010). The role of non-owner-occupied homes in the current housing and foreclosure. Federal Reserve Bank of Richmond No. 10-11.

Roodman, D. (2009). How to do xtabond2: An introduction to difference and system gmm in stata. Center for Global Development Working Paper No. 103.

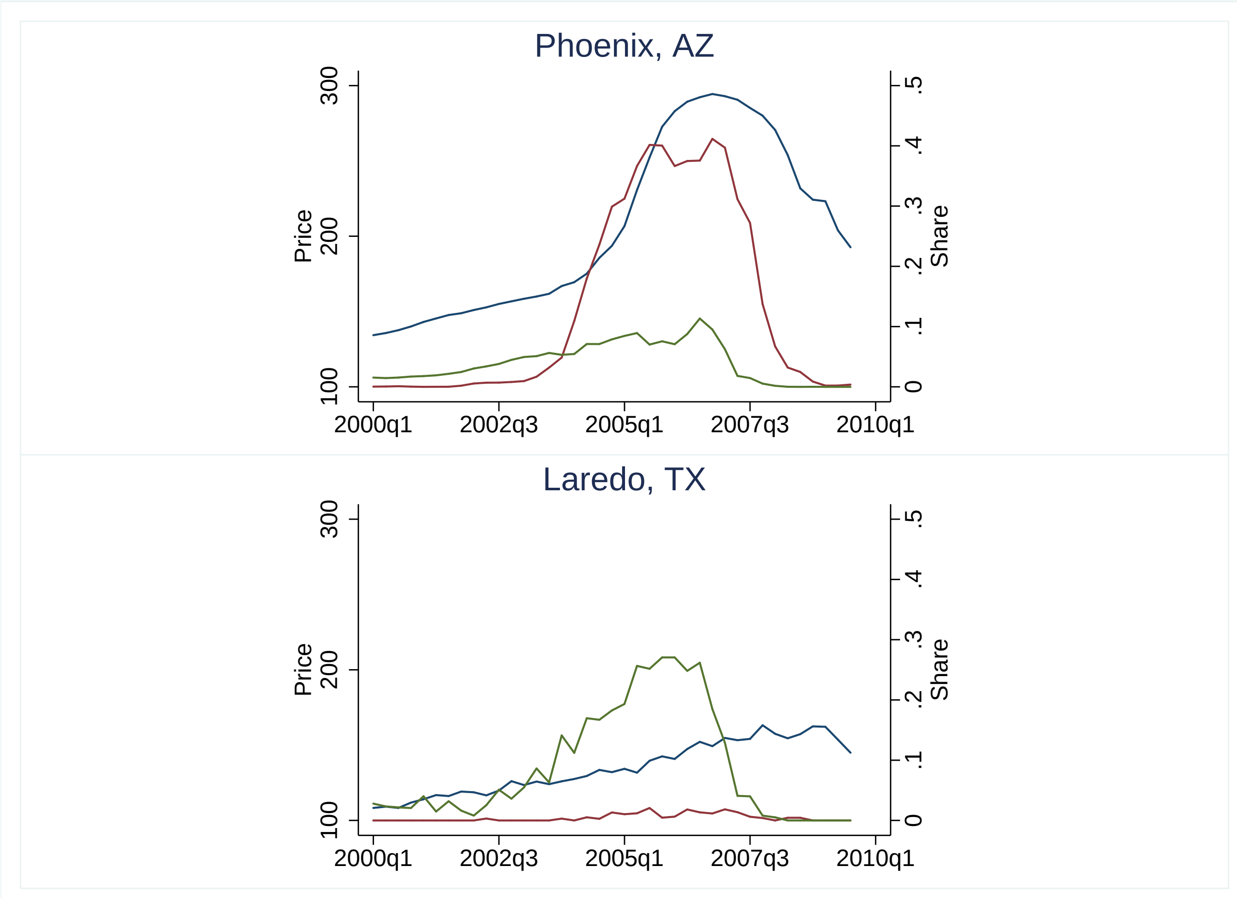
Saiz, A. (2010). The geographic determinants of housing supply. *Quarterly Journal of Economics 125*(3), 198–217.

Stein, J. C. (1995). Prices and trading volume in the housing market: A model with down-payment effects. *Quarterly Journal of Economics 110*(2).

Thompson, S. L. (2006). Nontraditional mortgage products. Statement before the Senate Subcommittee on Economic Policy, Subcommittee on Housing and Transportation and Committee on Banking, Housing and Urban Affairs.

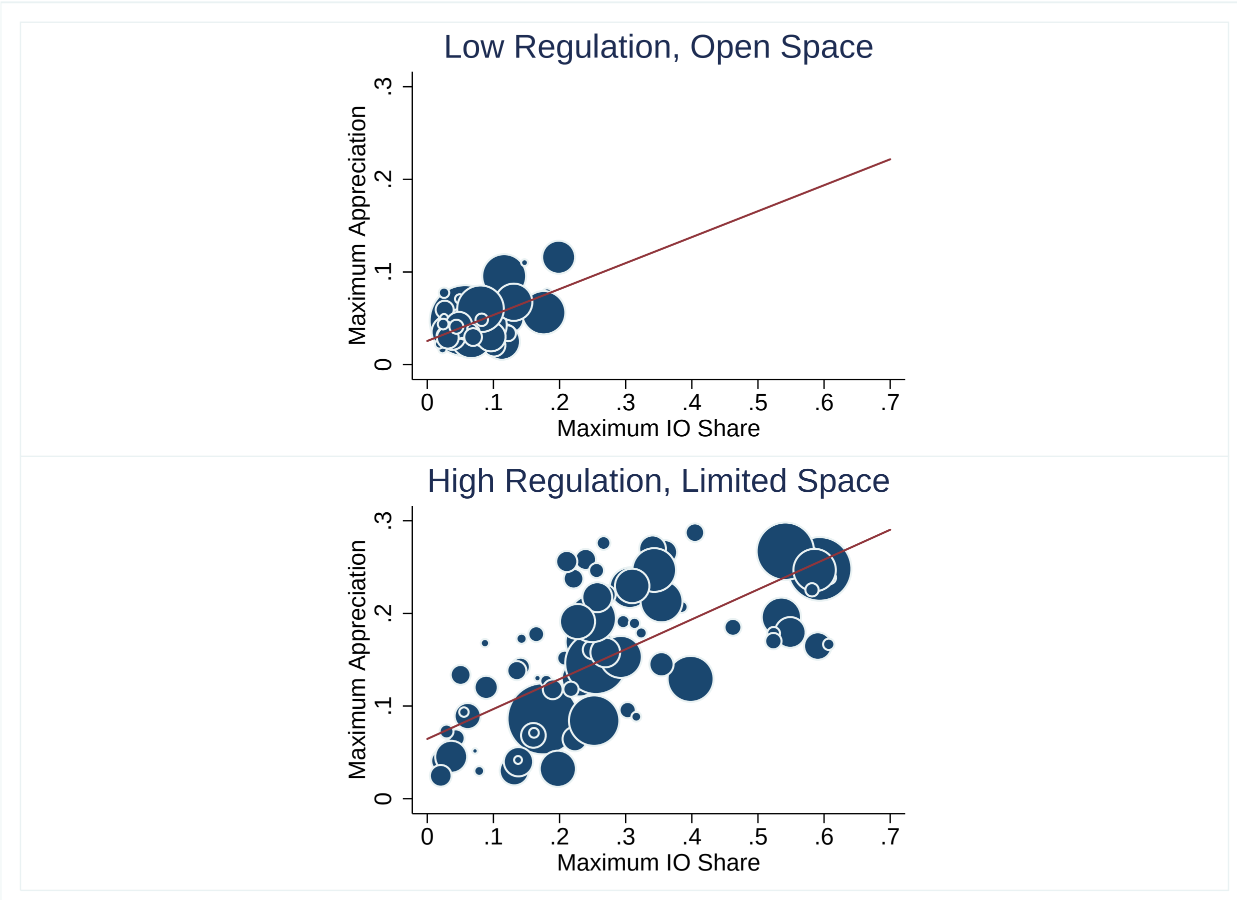
Windmeijer, F. (2005). A finite sample correction for the variance of linear efficient two-step GMM estimators. *Journal of Econometrics 126*, 25–51.

Figure 1: House Prices and Mortgage Use in Two Cities



Note: Blue lines – Real Price, Red lines – IO Share, Green lines – Sub-prime Share.

Figure 2: Maximum 4 Quarter Appreciation versus Maximum IO Share



Note: Circles indicate relative size of city, Red lines correspond to OLS regressions.

Figure 3: Correlations between Δio_{t+j} and Δp_t

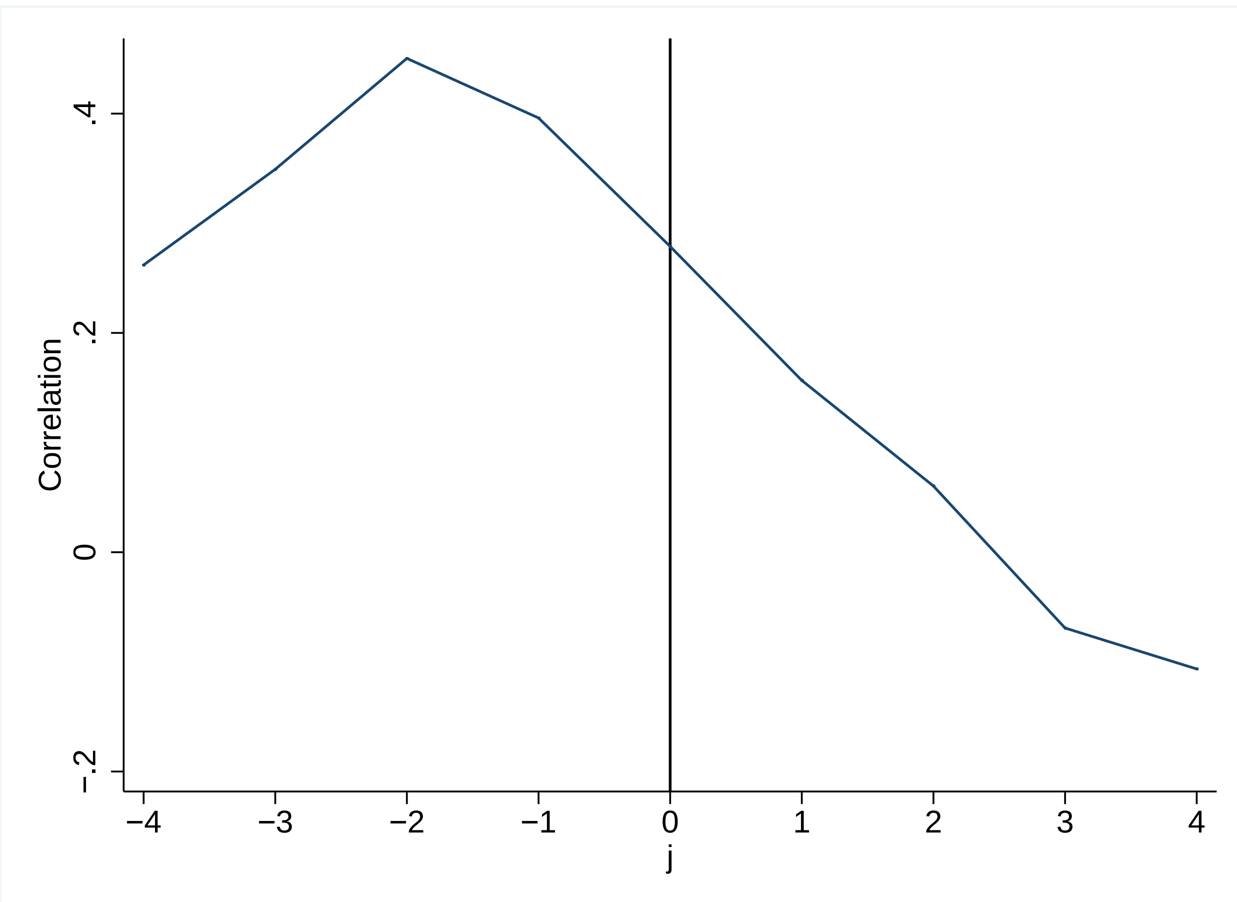
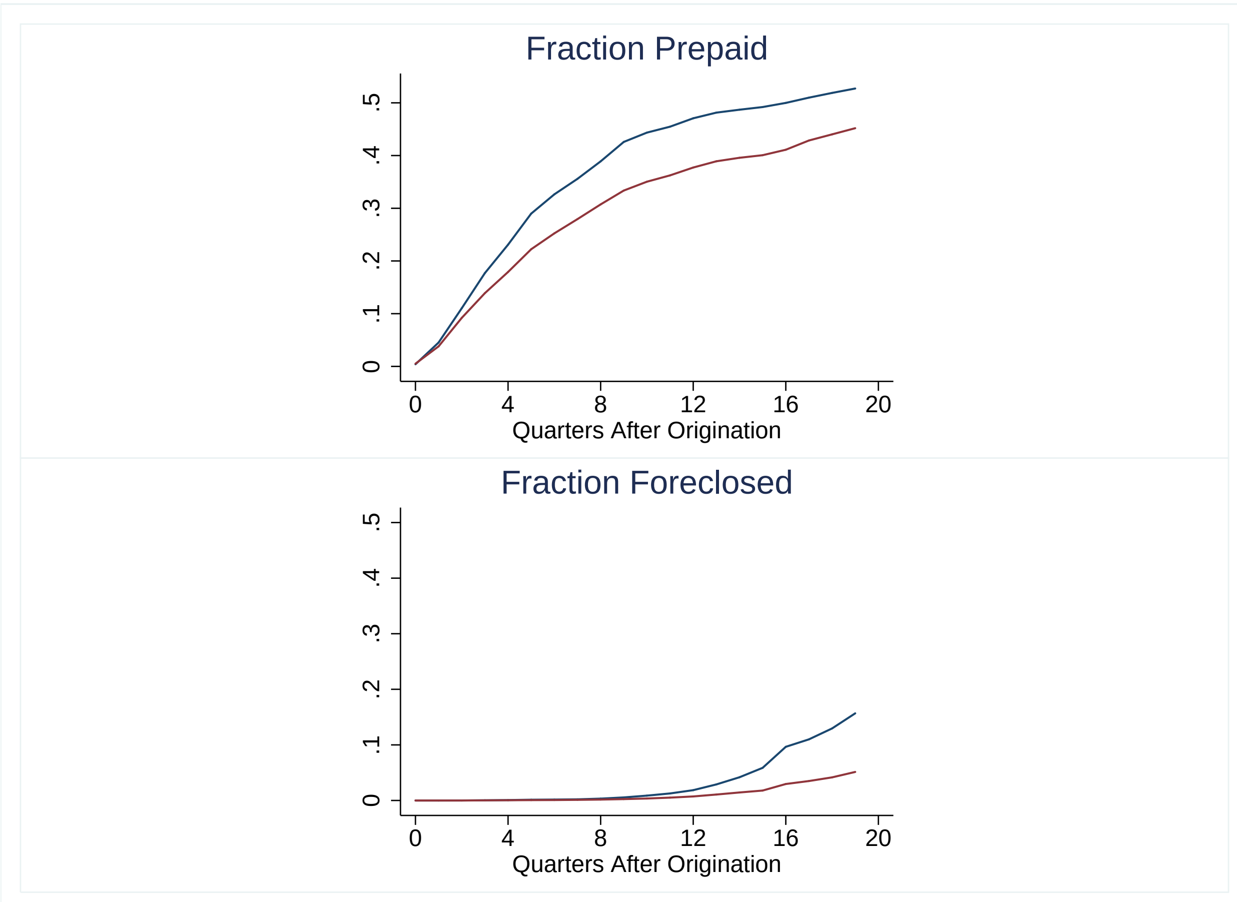


Figure 4: Pre-Payment and Foreclosure of Mortgages Originated in 2005q1



Note: Blue lines – IO Mortgages, Red lines – Non-back-loaded Mortgages.

Table 1: Mortgage Characteristics

Type	Year	Mean Amount	All	Owner	Investor	Long Term	Priv. Sec.	Sub-prime	Pre Pay Penalty	ARM
IO	2000	451.0	0.1	97.1	2.9	58.9	58.7	0.2	0.3	92.8
	2005	322.3	19.7	88.6	11.4	99.4	52.5	9.3	17.8	91.8
Option ARM	2000	192.4	1.1	97.3	2.7	98.6	N/A*	1.2	31.9	100.0
	2005	342.2	8.2	84.7	15.3	99.9	58.4	24.3	64.2	100.0
Fixed	2000	120.9	85.2	95.8	4.2	90.9	9.4	0.6	4.5	0.0
(Not IO)	2005	185.5	57.3	92.6	7.4	91.3	14.3	2.1	2.0	0.0
ARM	2000	217.9	10.6	98.1	1.9	94.1	17.6	7.6	14.5	100.0
(Not back-loaded)	2005	215.7	14.2	90.4	9.6	99.8	45.0	29.8	28.0	100.0
Other	2000	138.7	3.0	99.9	0.1	99.1	23.7	0.0	0.0	0.0
	2005	195.6	0.5	96.4	3.6	98.4	4.2	9.8	0.3	0.0
All	2000	132.6	100.0	96.1	3.9	91.5	10.7	1.3	5.7	11.7
	2005	225.6	100.0	90.9	9.1	94.9	29.8	9.3	13.9	40.5

Note: Entries are percent of indicated type of mortgage except for "Mean Amount" which is in units of thousands of current dollars.

Table 2: Summary Statistics for Price and Mortgage Variables

	Mean	St. Dev.	Min	25p	Median	75p	Max
Prices							
Max 4Q Growth	0.10	0.07	0.02	0.04	0.07	0.15	0.33
Mean Growth	0.05	0.04	0.01	0.02	0.03	0.07	0.14
Max 4Q Decline	0.10	0.10	0.00	0.05	0.07	0.11	0.65
Mean Decline	0.05	0.06	0.00	0.02	0.03	0.06	0.35
Mortgages							
Max IO Share	0.15	0.14	0.02	0.05	0.09	0.22	0.61
Max Hybrid Share	0.14	0.06	0.04	0.10	0.13	0.16	0.45
Max Long Term Share	0.97	0.02	0.90	0.96	0.97	0.98	1.00
Max Sub-prime Share	0.11	0.05	0.03	0.08	0.10	0.14	0.32
Max Securitiezed Share	0.28	0.10	0.12	0.21	0.25	0.32	0.64
Mean 80+ Share	0.68	0.08	0.38	0.63	0.69	0.73	0.85
Max Investor Share	0.11	0.04	0.04	0.08	0.10	0.14	0.38

Table 3: Baseline Models of Maximum 4 Quarter Price Appreciation in Boom Phase

	(1)	(2)	(3)	(4)
Max IO Share	.42		.35	.22
	(.03)***		(.03)***	(.04)***
Population Growth		2.76	1.35	.86
		(.64)***	(.36)***	(.27)***
Population		.008	-.002	.001
		(.006)	(.004)	(.002)
Per Capita Income Growth		.5	.87	.21
		(.44)	(.32)***	(.15)
Per Capita Income		-.04	-.1	-.06
		(.04)	(.03)***	(.01)***
Property Tax		-.04	.02	-.006
		(.01)***	(.01)*	(.01)
Property Tax Change		-.85	-.58	-.64
		(.2)***	(.1)***	(.1)***
Unemployment		.001	-.005	.001
		(.004)	(.003)	(.002)
Unemployment Growth		-.42	-.3	-.06
		(.11)***	(.08)***	(.08)
Regulation		.02	.004	-.004
		(.01)**	(.008)	(.005)
Undevelopable Land		.09	.02	.02
		(.03)***	(.02)	(.01)
State Fixed Effects	No	No	No	Yes
Observations	237	237	237	237
R^2	.65	.74	.87	.97

Note: OLS regressions of Maximum 4 Quarter Price Appreciation on indicated variables. Most variables are mean values from 2000q1 to quarter of peak real house price. Property taxes are for 2000 and the change between 2000 and the year of the peak price. The Regulation variable is the Wharton Regulation Index and the Undevelopable Land variable is from Saiz (2010). ***, **, and * denote significance at the 1, 5 and 10 percent levels respectively.

Table 4: Controlling for Affordability and Past Price Appreciation

	(1)	(2)	(3)	(4)	(5)	(6)
Max IO Share		.28		.29		.39
		(.07)***		(.07)***		(.04)***
Price at Peak	.13	.04				
	(.02)***	(.04)				
Price/Income at Peak			.13	.03		
			(.02)***	(.03)		
Appreciation 1985-1989					.32	-.19
					(.13)**	(.13)
Observations	105	105	105	105	171	171
R^2	.86	.9	.86	.9	.76	.88

Note: OLS regressions of Maximum 4 Quarter Price Appreciation on indicated variables plus the variables in Table 3 excluding Regulation, Undevelopable Land and State Fixed Effects. ***, **, and * denote significance at the 1, 5 and 10 percent levels respectively.

Table 5: Controlling for Additional Mortgage Characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Max IO Share	.35	.4	.33	.36	.35	.31	.34	.25
	(.03)***	(.04)***	(.04)***	(.03)***	(.05)***	(.04)***	(.02)***	(.07)***
Max Hybrid Share		-.14						.01
		(.09)						(.17)
Max Long Term Share			.43					.43
			(.24)*					(.24)*
Max Sub-prime Share				-.19				-.2
				(.09)**				(.21)
Max Securitized Share					.008			.1
					(.06)			(.06)
Mean 80+ Share						-.15		-.06
						(.07)**		(.07)
Max Investor Share							.62	.58
							(.13)***	(.12)***
Observations	237	237	237	237	237	237	237	237
R^2	.87	.88	.87	.88	.87	.88	.9	.91

Note: OLS regressions of Maximum 4 Quarter Price Appreciation on indicated variables plus the variables in Table 3 excluding Regulation, Undevelopable Land and State Fixed Effects. ***, **, and * denote significance at the 1, 5 and 10 percent levels respectively.

Table 6: Controlling for Additional Mortgage Characteristics With Price Declines

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Max IO Share	.59	.43	.6	.59	.39	.7	.58	.31	.35
	(.08)***	(.08)***	(.08)***	(.07)***	(.09)***	(.08)***	(.07)***	(.12)***	(.08)***
Max Hybrid Share		.51						.6	.26
		(.15)***						(.3)**	(.24)
Max Long Term Share			-.33					-.64	-.17
			(.49)					(.53)	(.36)
Max Sub-prime Share				.47				-.39	-.27
				(.15)***				(.3)	(.23)
Max Securitized Share					.32			.27	.12
					(.09)***			(.1)***	(.08)
Mean 80+ Share						.39		.27	-.13
						(.15)**		(.14)*	(.11)
Max Investor Share							.37	.56	-.04
							(.2)*	(.19)***	(.15)
Foreclosure									12.3
									(1.9)***
N	188	188	188	188	188	188	188	188	188
R^2	.84	.85	.84	.85	.85	.85	.84	.88	.93

Note: OLS regressions of Maximum 4 Quarter Price Declines on indicated variables plus the variables in Table 3 excluding Regulation, Undevelopable Land and State Fixed Effects. ***, **, and * denote significance at the 1, 5 and 10 percent levels respectively.

Table 7: Granger-Causality Based on OLS Regressions

	Δp_t				Δio_t			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Δp_{t-1}	.50	.37	.37	.42	.04	-.02	.05	.04
	(.04)***	(.05)***	(.05)***	(.04)***	(.06)	(.07)	(.07)	(.07)
Δp_{t-2}		.14	.10	.12		-.14	-.03	.008
		(.04)***	(.04)**	(.04)***		(.07)*	(.08)	(.08)
Δp_{t-3}			.15	.27			-.21	-.18
			(.04)***	(.05)***			(.08)***	(.08)**
Δp_{t-4}				-.31				-.04
				(.04)***				(.07)
Δio_{t-1}	.17	.13	.15	.13	.43	.32	.33	.31
	(.03)***	(.03)***	(.03)***	(.03)***	(.05)***	(.05)***	(.05)***	(.05)***
Δio_{t-2}		.12	.13	.12		.29	.32	.34
		(.03)***	(.03)***	(.03)***		(.05)***	(.05)***	(.05)***
Δio_{t-3}			-.07	-.07			-.10	-.07
			(.04)**	(.04)**			(.06)	(.07)
Δio_{t-4}				.02				-.11
				(.03)				(.06)*
$\sum \Delta p$	0.50	0.51	0.63	0.50	0.04	-0.16	-0.19	-0.17
	(.04)***	(.04)***	(.05)***	(.06)***	(.06)	(.07)**	(.09)**	(.10)*
$\sum \Delta io$	0.17	0.25	0.21	0.20	0.43	0.61	0.54	0.46
	(.03)***	(.03)***	(.03)***	(.04)***	(.05)***	(.05)***	(.06)***	(.06)***
AR(1)	0.00	0.58	0.00	0.01	0.00	0.36	0.02	0.07
AR(2)	0.20	0.60	0.00	0.00	0.00	0.01	0.65	0.00
AR(3)	0.00	0.00	0.00	0.09	0.40	0.62	0.87	0.04
AR(4)	0.00	0.00	0.00	0.71	0.60	0.09	0.42	0.14
F Statistic	36.2	29.6	23.0	15.4	0.47	3.21	3.12	1.42
P-value	0.00	0.00	0.00	0.00	0.49	0.04	0.03	0.23
Observations	754	725	696	667	754	725	696	667

Note: OLS regressions of log price growth or change in IO share on indicated variables. ***, **, and * denote significance at the 1, 5 and 10 percent levels respectively. "$\sum x$" denotes sum of coefficients associated with variable x. "$AR(j)$" indicates the p-value of the Arellano and Bond (1991) test for serial correlation in the residuals of order j. "F Statistic" is the test statistic for the null that the non-regressor lag coefficients are all zero with the p-value below.

Table 8: Granger-Causality Based on System-GMM

	Δp_t				Δio_t			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Δp_{t-1}	.45	.33	.35	.40	.02	-.003	.10	.16
	(.05)***	(.05)***	(.05)***	(.06)***	(.08)	(.11)	(.09)	(.11)
Δp_{t-2}		.12	.12	.09		-.22	-.10	-.17
		(.03)***	(.04)***	(.06)		(.10)**	(.11)	(.17)
Δp_{t-3}			.20	.26			-.24	-.37
			(.06)***	(.07)***			(.14)*	(.22)*
Δp_{t-4}				-.32				-.19
				(.09)***				(.20)
Δio_{t-1}	.22	.16	.21	.17	.60	.50	.52	.45
	(.02)***	(.04)***	(.03)***	(.04)***	(.05)***	(.08)***	(.09)***	(.10)***
Δio_{t-2}		.12	.12	.11		.22	.25	.24
		(.03)***	(.03)***	(.04)***		(.08)***	(.08)***	(.09)***
Δio_{t-3}			-.10	-.07			-.15	-.13
			(.05)**	(.06)			(.07)**	(.07)*
Δio_{t-4}				.05				-.02
				(.05)				(.08)
$\sum \Delta p$	0.45	0.45	0.67	0.43	0.02	-0.22	-0.24	-0.58
	(.05)***	(.06)***	(.08)***	(.17)**	(.08)	(.12)*	(.18)	(.44)
$\sum \Delta io$	0.22	0.28	0.23	0.26	0.60	0.72	0.62	0.55
	(.02)***	(.03)***	(.04)***	(.06)***	(.05)***	(.06)***	(.05)***	(.14)***
J-stat	28.9	28.7	28.2	28.4	29.0	28.9	28.7	26.6
dof	140	138	134	126	140	138	134	126
AR(1)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
AR(2)	0.99	0.01	0.37	0.68	0.001	0.14	0.39	0.54
AR(3)	0.00	0.00	0.01	0.37	0.05	0.43	0.85	0.77
AR(4)	0.00	0.00	0.00	0.52	0.78	0.95	0.89	0.79
F Statistic	83.8	62.6	35.4	16.8	0.07	2.87	1.29	0.89
P-value	0.00	0.00	0.00	0.00	0.80	0.07	0.30	0.48
Observations	754	725	696	667	754	725	696	667

Note: System-GMM estimates of log price growth or change in IO share on indicated variables. ***, **, and * denote significance at the 1, 5 and 10 percent levels respectively. See Table 7 for descriptions of "$\sum x$," "$AR(j)$" and "F-statistic". "J-stat" indicates Hansen-Sargan test statistic for the over-identifying restrictions, where "dof" is the degrees of freedom of the test. In all cases the p-value is 1.