The Role of Equity in the Housing Market: Empirical Evidence from 2007-2011

Issi Romem*

April 19, 2013

Abstract

Using a novel combination of administrative and proprietary data from 2007 to 2011 on King County, WA (metro Seattle), I estimate the effect of owner-occupants' home equity on their probability of sale and, indirectly, mobility. I exploit plausibly exogenous variation that follows only from changes in local housing price indices, and I account for confounding economic conditions that vary by time and location. The estimates indicate that sales decline dramatically over the combined loan-to-value ratio range from approximately 70% to 100%, well before homeowners reach negative equity levels.

Keywords:

Household mobility, negative equity, mortgage lock-in.

*Department of Economics, University of California, Berkeley. Email: `iromem@econ.berkeley.edu`. I thank my advisors, David Card, Bob Helsley and Enrico Moretti for their advice. Mark Borgschulte, Carola Conces, Barry Eichengreen, Fernando Ferreira, Willa Friedman, Francois Gerard, Joe Gyourko, Patrick Kline, Victor Lavy, Atif Mian, Roy Mill, Carl Nadler, Valentina Paredes-Haz, David Romer, Jesse Rothstein, Itay Saporta-Eksten, Michel Serafinelli, Edson Severnini, Ity Shurtz, Yannay Spitzer and seminar participants at UC Berkeley, Fannie Mae, the Federal Housing Finance Agency, Freddie Mac and at the 2012 EconCon conference at Princeton University provided many valuable comments and insights. I am grateful to Stan Humphries and Randy Newton for their help with data from Zillow.com and from the Washington State Voter Registry, and I am particularly grateful to Katie Dobbyn for her help with CoreLogic data. The CoreLogic Academic Research Council (CLARC) generously provided data for this research, although the views expressed here are my own, as are any remaining errors.

1 Introduction

As of mid 2012, fallen home prices left roughly 30% of mortgaged homeowners in the US with negative equity in their homes, while owner-occupant mobility was 30% below its level in 2005.[1] In theory, these observations are linked by the fact that falling home prices erode equity, preventing some homeowners from moving if selling their homes no longer enables them to pay off a mortgage or make a down payment on a new home. Yet, the empirical relationship between equity and mobility and its causal nature remain elusive: Whereas Ferreira et al. (2010, 2012) and earlier literature preceding the recent bust find that negative equity reduces mobility, recent studies such as Schulhofer-Wohl (2011) and Coulson and Grieco (2013) report the opposite.

Both theory and policy hang in the balance. In a seminal contribution to theory, Stein (1995) argues that the liquidity constraint imposed by diminished equity explains a fundamental feature of housing markets, whereby prices and trading volumes are positively correlated.[2] This feature, and Stein's explanation for it, form a basic tenet of our understanding of housing markets and their cycles. More recently, the possibility that diminished equity is adversely affecting the labor market by dampening mobility has drawn a great deal of attention and scrutiny, and the notion that resolving the housing crisis is inextricably linked with the relaxation of equity-driven constraints to mobility looms large.[3,4]

Resolving the debate concerning the empirical relationship between equity and mobility is essentially a matter of rigorous identification. Using a novel combination of recent administrative and proprietary data on homes in King County, Washington (metro Seattle), I estimate the effect of owner-occupant equity on the probability of sale and, indirectly, on mobility. I use changes in local housing price indices to isolate the component of equity that is most plausibly exogenous. Equity is jointly determined by owners' financial deci-

[1] Zillow.com estimates 30.9% of US homeowners with mortgages had negative equity in the second quarter of 2012. Aggregate data from the CPS and the ACS indicate the annual rate of owner-occupant mobility fell by 33.1% and 29.4%, respectively, from 2005 to 2011.

[2] Positive price-volume correlation in housing markets is documented for the US in Stein (1995), Berkovec and Goodman (1996) and Ortalo-Magne and Rady (2006), the UK in Andrew and Meen (2003), Ortalo-Magne and Rady (2004) and Benito (2006), Sweden in Hort (2000) and Hong Kong in Leung et al. (2002).

[3] The role of labor mobility in adjusting to economic shocks has long been recognized (Blanchard and Katz, 1992). Recent studies proposing this theme and questioning it include Sterk (2010), Aaronson and Davis (2011), Donovan and Schnure (2011), Estevao and Tsounta (2011), Karahan and Rhee (2011), Kothari et al. (2012), Modestino and Dennett (2012), Nenov (2012), Sahin et al. (2012) and Valletta (2013).

[4] If higher equity returns repeat buyers to the housing market, and if the historic correlation between price and volume also reflects causality in the opposite direction, *from the volume of traders to prices* (e.g. because of more frequent multiple bidder situations), then there is scope for policy to launch home prices on an upward spiral.

sions, which are fraught with endogeneity concerns, and by changes in home value. Setting the down payment amount, for example, is a financial decision that influences equity and is likely to be endogenous. Larger down payments offer savings and a bargaining advantage, but they require liquidity and so they are likely to reflect a more prosperous home buyer and lower mobility.[5] Changes in home value, on the other hand, consist of aggregate housing price changes and idiosyncratic property-level changes. As shown in Case et al. (1997) and confirmed in this study, properties that sell more frequently tend to experience idiosyncratic changes in home value that are more accentuated, reflecting that some homes are (unobservably) "better" than others and therefore exhibit less "churn" and more stable demand over time.[6] This suggests that properties which lost more value in the recent bust and whose owners therefore have less equity are also likelier to sell at any moment, implying that idiosyncratic changes in home value, too, are endogenous. Aggregate changes in housing prices, on the other hand, are driven to a large extent by factors far removed from the individual homeowner and often even from the local housing market. To harness these changes I construct an instrument for equity based purely on housing price indices that I refer to as *predicted* equity, but it too presents identification challenges.

A key identification challenge stems from economic conditions that vary by time and location, and which influence both home prices and sales. Consider for example a recession or an economically distressed area in which job security is low. Insecurity can deter committment to long term housing debt, reducing both the price and the quantity of homes sold. But even though fewer home sales coincide with lower prices and diminished equity, it would be wrong to attribute reduced sales to diminished equity. Time- and location-varying economic conditions can shift home prices and sales in a confounding manner through other channels as well, such as household formation and separation rates (Farnham et al. (2011)), the leverage cycle (Geanakoplos (2010)) and prevailing expectations with respect to future housing price trajectories.[7]

I address this challenge by conditioning estimates on the time and location of observation, using quarter by zip code area fixed effects. Doing so corresponds to a thought experiment

[5] Additional borrowing against a home once it has appreciated is another financial decision that influences equity and is likely to be endogenous. Such borrowing is only worthwhile provided the owner intends to stay in the home beyond a certain time horizon, suggesting that when such borrowing takes place, lower mobility could be a cause of low equity rather than a consequence.

[6] This finding has known implications in terms of sample selection bias for repeat sales housing price indices.

[7] Note that any factor channeling contemporary local economic conditions into housing demand contributes to the positive price-volume correlation in housing markets.

in which I compare the outcomes of similar homes observed at the same time and in the same place - and which are therefore subject to the same economic conditions - but which have experienced different changes in local housing prices only because they were bought at different times.[8] Consider for example the homes at 311 and 324 NW 48th street in Seattle's popular Fremont neighborhood. They are within 150 feet of each other, have the same number of bedrooms and bathrooms, nearly identical square footage, and they were built in 1940 and 1949, respectively.[9] When I compare them at later dates, say in late 2011, their probability of sale is subject to the same economic conditions - namely, those prevailing in Seattle's Fremont neighborhood in late 2011. Commonly held housing price expectations for the neighborhood at that time, for example, or the income, job security or family situation of typical Fremont homebuyers and sellers around that time all affect the sale probabilities of these two homes similarly. The crucial difference between the homes is in the timing of their last purchase. 311 NW 48th street was purchased in late 2009, but the home on 324 NW 48th street was purchased in mid 2006, when prices in the neighborhood were substantially higher and were still rising. By late 2011 the zip code area housing price index was 7% below its late 2009 level and 15% below its level in mid 2006, implying that the owner of 324 NW 48th street was likely to have lost twice as much equity as his neighbor due to price changes.[10]

However this comparison raises the concern that homes, owners and loans may differ systematically in ways that affect mobility depending on the time of purchase. The typical home buyer profile, for example, may vary over time reflecting changing lending standards (e.g. "subprime" lending), and homeowners buying when housing prices are expected to fall may be selected upon having a low propensity to sell in the short term. I account for this possibility by controlling for a flexible function of the time of purchase. The corresponding thought experiment then becomes comparing the sales of similar homes observed at the same time and in the same place, which differ in their experienced housing price changes only because they were bought at different times, while accounting for the average sale rate of homes bought in any given time period as they are realized in *all* periods of observation.

My data combine county assessor records with properietary mortgage data from Core-

[8]The similarity of homes refers to conditioning on observables, which include a flexible function of homeowners' tenure duration and numerous owner, home and loan characteristics.

[9]All details given here are publicly available from the King County assessor's office. For privacy reasons I do not specify owner or loan characteristics in this illustrative example.

[10]Note that experienced changes in the local housing price *index* are orthogonal by construction to any difference between owners, homes and loans observed in the same quarter and zip code area, except through the time of purchase.

Logic for the years 2007 to 2011. The merged dataset contains sale and equity histories, and rich home and loan characteristics, which I augment with owner characteristics from voter registration records and from Home Mortgage Disclosure Act (HMDA) files. The final dataset spans roughly 107,000 owner-occupied properties with appropriate attributes.[11] I construct predicted equity - the instrument for equity based purely on housing price indices - using zip code area housing price indices from Zillow.com.

I find that the probability of sale falls substantially at lower equity levels. Reduced form estimates using predicted equity indicate that the probability of arm's length sale falls over the range of combined loan-to-value (CLTV) ratios from 70% to 100%, and estimates using predicted equity as an instrument for actual equity place the magnitude of this effect at roughly 2.1 percentage points.[12,13] Given that the average unconditional rate of arm's length sales for mortgaged homes observed with CLTV ratios up to 70% is only 3.07%, this estimate implies a dramatic reduction in mobility. In contrast, naive OLS estimates using actual equity indicate an implausibly non-monotonic effect, and in particular an interval of high equity in which greater equity appears to reduce mobility, raising the concern that these estimates capture systematic unobservable differences between homeowners.

Whereas the public debate tends to dwell on the share of homeowners with negative equity, the estimates reveal that mobility is reduced well before homeowners reach negative equity levels (over 100% CLTV). Thus, the share of homeowners with negative equity substantially understates the share whose mobility is impaired by insufficient equity.[14] This finding also indicates the range of equity - 70% to 100% CLTV - in which policies manipulating homeowner equity are most likely to effectively shift sales and mobility.

In directly related earlier work, Henley (1998) observes 3,500 British households from 1992 to 1994 and finds that higher equity mildly lowers mobility, but that negative equity reduces

[11] I limit properties to those whose owners have tenure duration of 1 to 10 years. The total number of quarterly observations is just under 1.5 million.

[12] Unless stated otherwise, "sales" refer arm's length sales throughout this paper. An arm's length sale is one in which the involved parties are independent of each other and enter the agreement freely. Sales between relatives or between a firm and its subsidiary, for example, are not conducted at arm's length. In the absence of a sharper definition classifying sales as being at arm's length is ultimately a matter of discretion. I adopt a conservative approach, as detailed in appendix section **??**.

[13] Adhering to the norm of using lending industry terminology, I measure equity in terms of CLTV ratios. An x% CLTV ratio is identical to $(100 - x)$% equity, so 70% CLTV means 30% equity and CLTV ratios above 100% imply negative equity. The word "combined" in CLTV refers to the summing of *all* loans secured against a home, and is meaningful given that from 2005 to 2010 over 25% of housing units in the US - and 40% of properties in my sample - served as collateral for two or more loans (ACS estimate).

[14] The same Zillow.com report citing that 30.9% of mortgaged US homeowners had negative equity in their homes in the second quarter of 2012 also indicates that approximately half of mortgaged US homeowners have CLTV ratios above 80%.

it sharply. Chan (2001) observes 5,800 adjustable rate mortgages in New York's tri-state area from 1989 to 1994 and also finds that shifting from 70% to 100% CLTV reduces mobility sharply, by roughly 80%. Observing 2,400 US households from 1979 to 1996, Engelhardt (2003) argues that loss aversion - not insufficient equity - is responsible for reducing mobility when home prices fall. More recently Ferreira et al. (2010) observe 20,000 US households over the years 1985 to 2007 and estimate that negative equity reduces mobility by roughly one third.[15] I make three contributions with respect to this literature, of which the first two involve identification. First, earlier work shares the identifying assumption that conditional on controls, equity levels are uncorrelated with omitted variables that influence mobility. Given that actual equity is likely to be endogenous, whereas predicted equity captures only a plausibly exogenous component of it, my version of the identifying assumption is likelier to hold and yield estimates whose causal interpretation is valid. Second, by conducting the analysis within quarter by zip code area cells I control for time- and location-varying economic conditions that pose a key challenge to identification. I am able to do so by constructing data that are an order of magnitude larger than those used in earlier studies. The third contribution of this paper is in bringing the literature up to date with estimates from the current housing crisis.

This paper also relates to the influential study of the Boston condo market in the early 1990's by Genesove and Mayer (2001) and to a more recent study of the San Francisco Bay Area housing market by Anenberg (2011), who observe and attribute higher asking prices and lower sale probabilities primarily to owners' loss aversion. A crucial distinction between our papers is that Genesove and Mayer (2001) and Anenberg (2011) observe homeowners conditional on having listed their homes for sale, whereas I observe them unconditionally (in this respect), and examine an outcome that incorporates the decision to list a home for sale. This distinction implies that there is no contradiction in our findings. Rather, the contrast between our findings suggests it is possible that even though loss aversion plays a more decisive role once the sale process is underway, equity plays an important role in owners' earlier decision to list their homes for sale.[16] Like Genesove and Mayer (2001), I contribute to the literature studying the causal relationships underlying the positive price-volume correlation in housing markets, and in particular causality running from prices to

[15] Ferreira et al. (2011) update this estimate with data from 2009.

[16] The decision to list a home for sale cannot be analyzed separately with either of our data sets, because the data in Genesove and Mayer (2001) and Anenberg (2011) consist of sale listings, so are conditioned on the decision to sell, and the sale outcome in my data reflects the decision to sell only in conjunction with ultimate success of the sale process.

trade volumes. Other papers disentangling the web of causal relationships underlying this correlation include Leung et al. (2002), Clayton et al. (2010) and Anenberg (2012).

The paper is organized as follows: section 2 briefly formalizes the liquidity constraint underlying the role of equity, section 3 describes the empirical strategy and section 4 describes the data. Section 5 reports empirical estimates and section 6 concludes.

2 The home buyer's liquidity constraint

How might falling home prices and subsequent diminished equity affect homeowner mobility? An example adapted from Stein (1995) illustrates the idea.

Example: consider a family whose home was initially worth \$100,000, has an outstanding mortgage of \$85,000 and no other assets, and suppose the family wants to move. If housing prices have fallen so that the home is worth less than \$85,000, the family's equity becomes negative and it cannot even afford to pay off the mortgage. Under the circumstances, the family cannot move at all in the traditional way of selling one home and buying another.[17]

Notice, however, that contrary to the common focus on "underwater" homeowners the family's mobility might be affected even without having negative equity. At a value of \$90,000 the family can pay off the mortgage and keep \$5,000 of the proceeds. Supposing a minimum down payment requirement of 10% this amount is sufficient or a \$50,000 home, but not for one that is equivalent to the old home (that would require a \$9,000 down payment). Whether the family chooses to make such a move depends on the trade off between the want - or need - to move and the reduction in housing service consumption implied by moving into a less valuable home.

Of course, there is also a potential upside to the family's leverage. Provided it has the necessary future income, an increase in value to \$110,000 allows the family to pay off the mortgage and keep \$25,000 of the sale proceeds - enough for a \$250,000 home. Upgrades such as this are referred to as trade-up buying.

Formalizing the liquidity constraint: Suppose a *first-time home buyer* with savings S has no other assets at his disposal. If he is unable to take out a mortgage, he can afford a home H, where H is the measure of housing services obtained from the home and P is

[17] Foreclosure, leaving behind a vacancy or engaging in two-way rental may remain viable options depending on the family's income flow, though these options preclude owner-occupancy.

their unit price, provided that his savings satisfy $S \geq PH$. A mortgage allows the buyer to borrow funds with which to buy a home in exchange for a guarantee to repay the funds in the future and to make a down payment of at least γPH immediately, with $\gamma \in [0, 1]$.[18] With the option of a mortgage in hand and provided sufficient future income, the buyer can afford any home H such that his savings exceed the minimum required down payment,

$$S \geq \gamma PH. \tag{1}$$

Re-arranging, this condition implies that the housing services provided by his new home, which I refer to as its "size", is limited by the leveraged buying power of his savings:[19]

$$H \leq \frac{S}{\gamma P}. \tag{2}$$

This is the first-time buyer's liquidity constraint. Clearly, the buyer can afford less housing if its price increases.

Now suppose the buyer is *a repeat buyer* who already owns a home, H_0, and has outstanding mortgage debt $M \geq 0$ on his old home. Contemporary mortgages almost always include a due-on-sale clause, obligating the borrower to repay the outstanding balance of the mortgage in the event of sale.[20] Thus, provided sufficient future income, the buyer can afford any home H such that

$$PH_0 + S \geq M + \gamma PH, \tag{3}$$

meaning that he can only afford a home if his savings and sale proceeds are enough to pay off his old mortgage and make the minimum down payment on that home. Expressing this constraint as

$$H \leq \gamma^{-1}(H_0 + \frac{S - M}{P}). \tag{4}$$

reveals that the "size" of the new home is limited by the leveraged buying power of the sum

[18] Of course, mortgages also involve interest, amortization and the use of the home as collateral, not to mention reams of paperwork.

[19] The term *leveraged* buying power refers to the fact that every dollar of savings permits buying $1/\gamma$ dollars of housing with the help of a mortgage.

[20] This clause become prevalent in the late 1970's in response to rising interest rates. These prompted lenders to phase out the previously ubiquitous *assumable* mortgage, which allowed buyers to retain previous owners' lower rates.

of the sale proceeds and the buyer's net financial assets. This is the repeat-buyer's liquidity constraint. Whether rising home prices relax or tighten the constraint now depends on the sign of the buyers' net financial assets, $S - M$. Crucially, if his mortgage debt outweighs his savings, i.e. $S - M < 0$, then *rising home prices relax the liquidity constraint.*[21,22] This situation is typical of most repeat buyers, who tend to be within years - not decades - of purchasing their old home.

Note that if the mortgage debt exceeds the sum of sale proceeds and savings, i.e. $M > PH_0 + S$, then the buyer cannot afford to pay off the old mortgage, so he cannot afford any positive amount of housing and is unable to move in the traditional manner. Ignoring the role of savings, this situation is referred to as being "underwater" and amounts to having negative equity.

Suggestive stylized facts: An increase in house prices reduces the amount of housing that first time home buyers can afford, but it increases the amount that repeat buyers whose mortgage debt exceeds their savings can afford.[23] If enough repeat buyers have mortgage debt in excess of their savings, we should expect to see the share of first-time buyers decrease and the share of repeat buyers increase when housing prices rise, and the opposite when they fall. Figure 1 shows that the share of repeat buyers in the US increased until 2006 and then decreased, in tandem with housing prices.

Figure 2 goes further and shows that *within* the set of repeat buyers, the share reporting savings and the share reporting home equity as sources of funding increased and decreased, respectively, as housing prices fell from 2006 to 2010, suggesting that buyers whose wealth was in the form of financial assets rather than home equity were increasingly the ones capable of moving.[24]

[21] An implicit assumption is that all homes' prices change in proportion to their "size", H. Exceptions to the above rule may occur if this assumption does not hold. In this case, (4) becomes $\frac{H_1}{P_1} \leq \frac{H_0+S-M}{P_0}$, where P_0 and P_1 are the prices of the old and new homes, H_0 and H_1, respectively. A straightforward example would be a differential increase in P_0 and P_1 such that the liquidity constraint tightens.

[22] Note that non-mortgage debt - omitted from the above formulation above for simplicity - is really no different than mortgage debt in this context, and can be included in M.

[23] Repeat buyers whose savings exceed their mortgage debt are affected similarly to first time buyers.

[24] Savings are here defined broadly to include stocks and other financial assets. Equity refers primarily to the proceeds of selling a home, but also, e.g., home equity loans on additional homes owned.

Figure 1: The share of repeat home buyers. Whereas first-time home buyers can afford more housing when home prices are lower, repeat buyers whose mortgage exceeds their savings can afford more (less) housing when home prices are higher (lower). The housing price shown is the real average sale price of all US home sales (in 2012 dollars).

Source: National Association of Realtors and US Census.

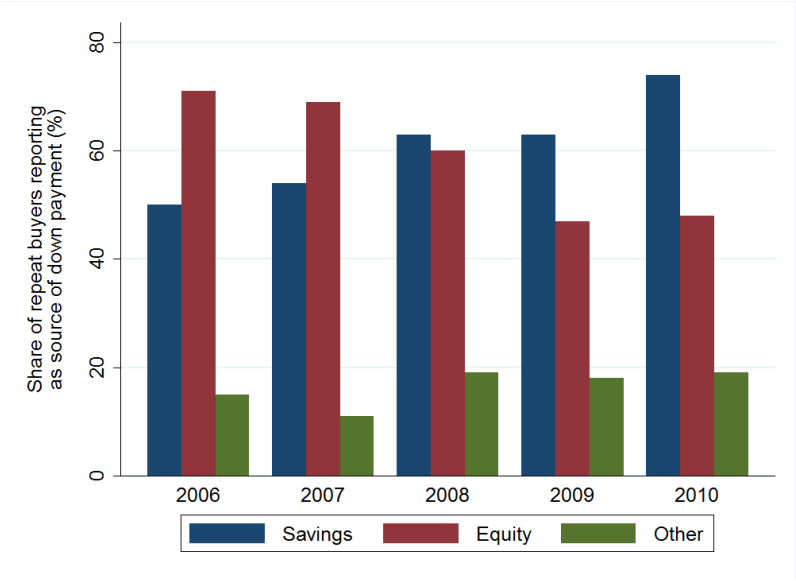


Figure 2: Sources of down payment funding reported by repeat buyers. As housing prices declined from their peak, a decreasing (increasing) share of funding came from housing equity (savings).

Source: National Association of Realtors.

3 Challenges to identification and solutions

3.1 Earlier approaches

To gauge the effect of owner-occupants' equity on mobility, earlier studies estimate the following generic form,

$$\Pr(move_{it}|duration_{it}) = f\left(CLTV_{it}, \mathbf{X}_{it}, \epsilon_{it}\right), \tag{5}$$

where $move_{it}$ is an indicator for some measure of mobility from property i at time t, $duration_{it}$ refers to the owner's tenure duration at the time, $\mathbf{X}_{it}$ is a vector of covariates and ϵ_{it} is a remainder term. The measure of mobility $move_{it}$, the model reflected by $f(\cdot)$ and the form in which CLTV enters it take on different forms in each study, as summarized in Table 1. The covariates observed in each study differ depending on the data used and are too numerous to list here, but they broadly tend to consist of owners' demographic and socio-economic characteristics, observable financial traits and environmental variables such as contemporary local unemployment rates.[25]

Table 1: Earlier studies of the effect of equity on mobility

Study	Mobility measure	Model	Functional form of CLTV
Ferreira et al. (2010, 2011)	Sale	Probit, LPM	Equity $\gtrless 0$
Engelhardt (2003)	Direct	LPM, PH	Initial CLTV in 80-90%,...,>95%
Chan (2001)	Mortgage prepayment	PH	Cltv in 40-50%,...,>95%
Henley (1998)	Direct	MPH	Piecewise linear in equity $\gtrless 0$

Notes: LPM refers to a linear probability model, PH to a proportional hazard model, and MPH to a mixed proportional hazard model that accounts for unobserved heterogeneity.

The concern in estimating (5) is that $CLTV_{it}$ might be correlated with the remainder term, ϵ_{it}. Such correlation may be present for several reasons.

3.2 Challenges to identification

The challenge posed by household-level determinants of equity: A property's CLTV ratio is fully determined by four factors: the initial down payment, payments made

[25] Some studies include year and MSA fixed effects, but not the cross of the two or more local or frequent fixed effects.

towards principal (as per the amortization schedule), adjustments of the combined loan amount and changes in the value of the property.[26] Each of these factors provides scope for correlation.

A larger down payment requires more liquidity, indicating greater wealth that may correlate with characteristics that independently influence homeowners' likelihood of moving, like their stage in the life cycle or the location of relevant employment opportunities if, say, their wealth is derived from a highly specialized occupation. The concern remains even conditional on observable demographic and socio-economic characteristics: a home buyer who is more prosperous in the narrow financial sense is likelier to also be prosperous in the broad sense of being satisfied and secure in his or her present life circumstances. This type of broad prosperity is practically unobservable and is likely to correspond with lower mobility and greater willingness to make a larger down payment (and in so doing to reap financial and bidding advantages). Similarly, the size of amortized payments towards principal increases with shorter mortgage terms and larger payments may indicate greater wealth, suggesting a re-iteration of the previous line of reasoning. Also along these lines, delinquent payments may indicate a lack of wealth.

Adjustments of the combined loan amount may also be problematic. Refinancing, prepayment and additional borrowing against a home all involve substantial transaction costs. Thus, even if the circumstances otherwise justify an adjustment - for instance if refinancing is appealing because interest rates have declined - an adjustment is only worthwhile for a homeowner who plans to stay put for a long enough time. This means that observing a homeowner make or not make adjustments provides a signal that their propensity to move, at least in the short run, is respectively lower or higher than it would be otherwise, for reasons that are independent of the subsequent level of equity. If adjustments to mortgages disproportionately involve additional borrowing which updates CLTV ratios upwards, then the effect of higher CLTV ratios on the probability of sale is biased downwards.

Changes in property value, too, may be endogenous, especially inasmuch as individual property values are considered. When an owner-occupant renovates, for example, the owner's propensity to move is likely to fall while the property value rises. Changes in *aggregate* local housing prices such as those measured by local housing price indices are not subject to concerns like this one, but even variation in equity that owes only to changes in aggregate local housing prices presents identification challenges.

[26] Adjustments of the combined loan amount include, for example, pre-payment, refinancing with cash-in or cash-out, or borrowing through any of the multitude of additional types of loans secured against the home.

The key challenge posed by confounding economic conditions: A key challenge to identification is that local housing prices and home sales are co-determined by factors that essentially reflect the business cycle. Because rising home prices raise equity levels, these factors make higher equity coincide with higher rates of home sale, regardless of any causal effect of equity on sales. In other words the business cycle, and differences in its manifestation across locations, confound estimates of the effect of equity on home sales.

To be more concrete, consider the following examples. Job security tends to vary over time owing to the business cycle, and may be more or less pronounced in different locations depending on the local composition of industry and on the typical demographic and socio-economic characteristics of residents, e.g. a blue-collar neighborhood versus a white-collar one. If job security facilitates commitment to long-term housing debt then it increases housing demand (and supply, inasmuch as home buyers are also sellers), influencing both the price and quantity of homes sold irrespective of equity. Voluntary job turnover, which has been shown by Akerlof et al. (1988) to be pro-cyclical, can have the same confounding effect if with some probability every job switch involves relocation or a housing upgrade. So can household formation and separation rates, which vary systematically along the business cycle, too.[27] The leverage cycle (Geanakoplos, 2010) can also render equity endogenous. When housing prices rise (fall) lenders may become less (more) wary of the risk of default and are likely to provide credit at higher (lower) leverage levels, thereby influencing the price and quantity of home sales in a way that is correlated with equity, but not caused by it.[28] Finally, housing price expectations are likely to vary given the past - and especially the recent - trajectory of housing prices, feeding into the demand for housing, and hence into housing prices and sale volumes.[29]

The challenge posed by the cohort of purchase: Although they are more subtle, exploiting variation in equity derived only from changes in aggregate local housing prices entails additional identification challenges as well. Conditional on the time and place of observation, changes in local housing prices since a home was purchased could reflect systematic differences in unobservables between homeowners who bought at different times, and who have chosen whether or not to sell every period since. An important example concerns *housing*

[27]Note that in addition to divorce, household separation includes young adults' decision to leave their parents' homes, too. With respect to divorce, Farnham et al. (2011) show that the way in which it is affected by the business cycle is anything but clear, a-priori.

[28]Central banks' management of interest rates in response to the business cycle can plays a similar role, too.

[29]Case et al. (2012) provide valuable empirical information on housing price expectations.

price expectations. When housing prices are expected to fall, potential buyers with intentions of selling in the short-run are less likely to buy. Thus, the pool of buyers when housing price expectations are dire is likely to be characterized, on average, by a lower propensity to sell than the pool of buyers in more optimistic times. To the extent that homeowners' current predicted equity levels reflect their cohort of purchase, such average tendencies of cohorts may confound the effect of equity on home sales.

The challenge posed by dynamic selection: A closely related identification challenge is dynamic selection, which is related to homeowners' repeated sale decision each period from the time of purchase to the time of observation. A stylized example can help clarify matters. Suppose that the pool of homeowners consists of two types, starter-home folk who have a high propensity to move, and dream-home folk who have a low one. Every period, the share of starter-home folk in each cohort dwindles, as starter-home folk sell and move (possibly re-entering the analysis in a different cohort). Inasmuch as this process is homogeneous across cohorts, it can be accounted for by including a polynomial of tenure duration as a control. However, if the decision to sell is influenced by other factors, such as home equity, then the "weeding out" process of starter-home folk occurs at a faster pace in cohorts experiencing higher equity, and is not adequately accounted for simply by controlling for a polynomial of tenure duration.

The challenge posed by loss aversion: A remaining concern stems from the fact that falling home values reduce equity mechanically, so experiencing a net loss in individual property value often coincides with low equity. Genesove and Mayer (2001) show that sellers experiencing a loss are less likely to succeed in selling their homes, suggesting that falling home values reduce sale rates by invoking loss aversion, rather than through diminished equity. Fortunately this concern has testable implications. If it is only loss aversion that reduces home sales when prices fall, then a sample consisting only of homeowners experiencing *gains* should not exhibit any correlation between equity and home sales.

3.3 Solutions

To gauge the effect of owner-occupants' equity on mobility, I adapt the generic form in (5) and estimate linear probability models of the following type:

$$sale_{it} = \sum_j \beta_j \mathbf{1}\{CLTV_{it} \in \mathcal{B}_j\} + \mathbf{X}_{it}\delta + \theta_{lt} + \epsilon_{it}, \tag{6}$$

where $sale_{it}$ is an indicator for the arm's length sale of property i in quarter t, the function $\mathbf{1}\{CLTV_{it} \in \mathcal{B}_j\}$ is an indicator that the owner's CLTV ratio falls in the interval ("bin") $\mathcal{B}_j \in \mathbb{R}$ belonging to some set $\{\mathcal{B}_j | j = 1, 2, 3, \ldots\} \in \mathbb{R}^J$, the vector $\mathbf{X}_{it}$ consists of property-level covariates, θ_{lt} is a saturated set of quarter by zip code area fixed effects where $l \equiv l(i)$ is property i's zip code area, and ϵ_{it} is a remainder term. The covariate vector $\mathbf{X}_{it}$ includes the observed owner, home and loan characteristics listed in Table 2.

In what follows, I explain how this framework addresses the various identification challenges one by one.

Addressing the challenge posed by household-level determinants of equity: Recall that a property's CLTV ratio is fully determined by four factors: the initial down payment, payments made towards principal, adjustments of th combined loan amount and changes in the value of the property, which may be specific to an individual home or aggregate in nature. To isolate variation in equity that comes only from aggregate changes in property values, I construct an instrument for actual CLTV using only changes in local housing prices indices, which I refer to as *predicted CLTV*, as follows.

$$\tilde{CLTV}_{it} \equiv \frac{0.8 \cdot hpi_{lc}}{hpi_{lt}}, \tag{7}$$

where hpi_{lt} is the value of the local housing price index for zip code area $l \equiv l(i)$ in the quarter of observation, t, and hpi_{lc} is the value of the same local housing price index in the quarter of property i's previous purchase, or cohort, $c \equiv c(i, t)$. The intuition behind predicted CLTV is that it mimics the CLTV ratio that would result for property i at time t if its value identically tracked that of the local housing price index, and if it had been purchased at time c with a down payment of 20%. Figure 3 is a scatterplot comparing predicted and actual CLTV ratios. Predicted and actual CLTV ratios are positively correlated, but they are far from identical. Whereas actual CLTV is contaminated by endogenous, household-level determinants of equity such as homeowners' financial decisions or renovation decisions,

Table 2: Owner, home and loan characteristics included as controls

Owner	Home	Loan
Tenure duration^3	Year built^3	Initial CLTV ratio at purchase^3
Cohort of purchase^3	Square footage^3	Mortgage term
Age^3	Bedroom count^2	Adjustable rate mortgage (ARM)
Average race and ethnicity of	Bathroom count^2	Going interest rate
local buyers in cohort	Real purchase price^3	
Local median real income^3		
Local average household size		
Local % college graduates		
Local % high-school graduates		

Notes: Superscripts $(\cdot)^2$ and $(\cdot)^3$ indicate the inclusion of second or third order polynomials, respectively. Tenure duration is measured in quarters. Cohort refers to the quarter in which the current owner purchased the home. The age variable measures the age of the oldest actively registered voter living in a property. The race and ethnicity variables are the mean shares of successful mortgage applicants in a property's census tract and its owner's year of purchase who are asian, black, hispanic and white. Local median real income, average household size and percent of college and high school graduates are observed at the census block-group level from the 2000 Census. Real purchase price is obtained by deflating the actual purchase price with respect to the Zillow.com housing price index for King County, WA, with the aim of obtaining an indication of value that is comparable over time. The mortgage term is captured by indicators for 10, 15, 30 and 40 year mortgages. The going interest rate refers to the average Freddie Mac rate for the appropriate term fixed rate mortgage, 30 days prior to origination (obtained by Corelogic).

predicted CLTV is free of these influences.

Addressing the key challenge posed by confounding economic conditions: To address the key identification challenge I condition estimates on the time and location of observation. The quarter by zip code area fixed effects, θ_{lt} in (6), effectively transform the analysis into one that is within time by location cells.[30] The thought experiment that

[30]Conditioning on this set of fixed effects requires sufficient observations in each time by location cell. With insufficient observations, one might condition only additively on time fixed effects and location fixed effects, or even just on a set time fixed effects. While doing so is far better than ignoring the issue, it fails to sweep away all of the endogenous variation. The reason is that location fixed effects account only for time-invariant differences between locations, and time fixed effects account only for location-invariant differences over time (i.e. a common time trend). Realistically, however, variation is neither time-invariant across locations, nor location-invariant over time. Consider for example the housing markets in California's hard-hit Central Valley and in far less impacted San Francisco: controlling for a common time trend and for

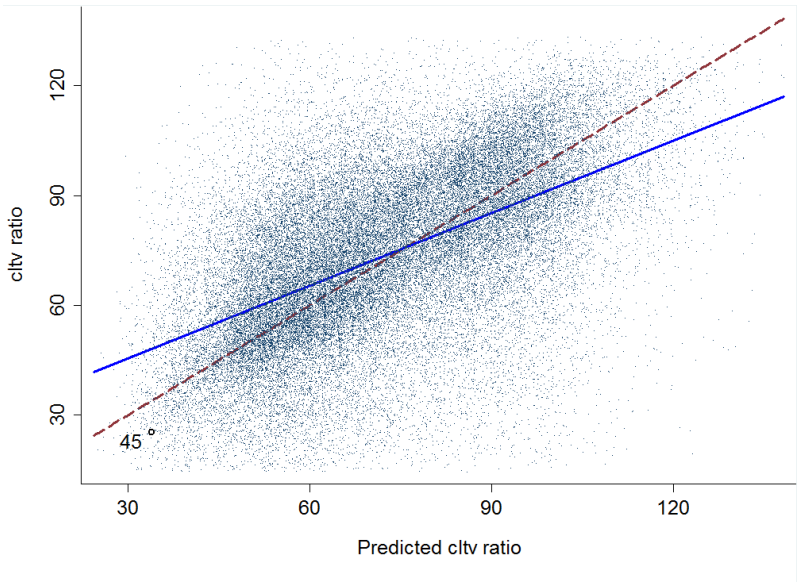


Figure 3: Predicted vs. actual CLTV ratios. By construction, predicted CLTV reflects only the component of actual CLTV driven by changes in local housing price indices. Actual CLTV also reflects changes in individual property values, as well as homeowners' financial decisions such as their initial down payment and changes in their combined loan amount (e.g. due to further borrowing against equity), and amortized payments towards principal. The linear fit shown is $actual\ CLTV = 39.17(0.92) + 0.45(0.01) \times predicted\ CLTV$, and has an R^2 of 0.30 (standard errors clustered by zip code area).

parallels this strategy is comparing the sale of homes observed at the same time and in the same place, but whose owners nevertheless differ in their equity levels. Note that while actual CLTV may differ within time by location cells because of any of the household-level determinants of equity, predicted CLTV can only vary within these cells due to homes' time of purchase.

Addressing the challenge posed by the cohort of purchase: To account for the possibility that homeowners' current CLTV levels reflect omitted variables correlated with their cohort of purchase, I include in the vector $\mathbf{X}_{it}$ a third-order polynomial of the cohort of purchase. This control accounts for selection into purchase cohorts due to housing price expectations at the time of purchase, as well as any other unobservable owner, home or loan characteristics that vary systematically by cohort of purchase.

time-invariant differences between these two locations can hardly account for the differential impact of the housing crisis on both prices and home sale volumes in the two locations.

Addressing the challenge posed by dynamic selection: The extent to which dynamic selection is a concern is not a-priori clear. I address dynamic selection in an appendix section by gauging its extent and direction.

Addressing the challenge posed by loss aversion: I address the possible role of loss aversion by estimating the effect of equity on home sales in a sample limited only to homeowners who have experienced a net gain in their home value. While effects observed in the full sample may reflect the influence of loss aversion, any effect observed in this "gain-only" sample cannot be driven by loss aversion.[31]

3.4 Limitations

The empirical strategy detailed above is not without limitations, as follows.

Endogenous savings: It is possible that homeowners' savings constitute an endogenous response to their level of equity. If so then savings, which affect homeowners' ability to sell and move, may vary systematically with equity and confound estimates. Because I cannot observe savings, I cannot control for this possibility directly. However, assuming that homeowners' saving behavior is compensating in nature, meaning that they save more when their equity is low and vice versa, then ignoring homeowners' endogenous savings would attenuate any observed effect of equity to zero. Thus, any non-zero estimates can be considered to be a lower bound (in absolute value) of the true effect of equity on sales.

Wealth effects: The strategy does not distinguish between different mechanisms through which equity affects sales. One such mechanism is the effect of changes in equity on homeowners' liquidity constraints, as detailed in section 2. Another involves wealth effects that accompanies changes in equity. Analyzing such wealth effects is not trivial (see e.g. Grossman and Laroque (1990)), and is beyond the scope of this paper. Nevertheless, the ultimate finding of this study whereby the effect of equity on home sales occurs within the confined equity range of 70%-100% CLTV, suggests that liquidity constraints rather than wealth effects are the more important mechanism.

[31] Note that estimates obtained from a sample limited only to homeowners who have experienced a net *loss* may still be subject to the influence of loss aversion. Suppose for example that loss averse homeowners set their asking price equal to the amount they paid for their home. In this case, homeowners with greater net losses set prices farther above their homes' market value which reduces their probability of sale more sharply, and this reduction in sales correlates with lower (or more negative) equity, simply because the falling home values that generate the loss also reduce equity.

General equilibrium effects of equity: The empirical strategy detailed in above estimates only a "partial equilibrium" effect of equity on sales, in the sense that the effect of an individual homeowner's equity is estimated on his or her sale probability, *all else equal*. The effect of changes in the equity levels of a broader group of homeowners, e.g. higher equity across the board raising all homeowners' likelihood of mobility, constitute part of the contemporary local economic conditions soaked up by the time $\times$ location fixed effects. Estimating these broader, "general equilibrium" effects of equity requires a different framework than the one presented here, and is beyond the scope of this paper as well.

4 Data

This study uses data from King County, Washington, which is the core county of the Seattle metropolitan area. Section 4.1 describes the data sources and construction, and provides summary statistics. Section 4.3 provides information on the evolution of housing prices in King County and on the Seattle metro area's relative standing in terms of diminished equity.

4.1 Data sources and construction

This study combines data from several sources: the core data are from administrative records kept by the King County assessor and from a matched proprietary dataset on housing loans provided by CoreLogic. In addition to these two main sources, further information is obtained from the Home Mortgage Disclosure Act's (HMDA) loan application registry, from Washington State's Voter Registration Database (VRDB), from the US Census. Local housing price indices at the zip code area level are obtained from Zillow.com.

The King County assessor maintains numerous files. The key file used in this study is the real property sales file, which documents the date, dollar amount and identities of the parties involved in every real estate transaction in the county. I shape the file into a panel, such that properties are observed repeatedly over time at a quarterly frequency, and each quarter every property's sale or lack thereof is indicated. Many sales - almost 42% - are attributed reasons such as placement in a trust, foreclosure, divorce settlement or gift transfer. I focus on the remaining 58% of sales for which a sale reason is not provided, and consider them to have been conducted at arm's length.[32] Additional county assessor files provide information on property characteristics such as their address, year of construction, square footage and

[32] A summary list of sale reasons and their frequency is given in appendix Table 3.

bedroom and bathroom count.

Table 3: Frequency of Sale by Reason

	Frequency	Percent
None, i.e. arm's length sale	147,868	58.18
Other	23,888	9.40
Placement in trust	14,992	5.90
Foreclosure	14,442	5.68
Establishment of community property	12,109	4.76
Property settlement	10,871	4.28
Estate settlement	10,019	3.94
Partial interest (love, affiliation, gift)	7,216	2.84
Divorce settlement	7,084	2.79
Tenancy partition	3,024	1.19
Correction (refiling)	1,904	0.75
Will-related transfers	434	0.17
Trade	103	0.04
Other settlement	70	0.03
Quit claim deed	62	0.02
Assumption	55	0.02
Easement	17	0.01
Total	254,158	100.00

Notes: the reported sale frequencies are for all observed residential properties over the period 2007-2010, regardless of tenure duration, price at previous sale or owner-occupancy status.

The CoreLogic data provide information on housing loans in the county, and they allow me to observe properties' estimated CLTV ratios. The data consist of a sequence of quarterly extracts from CoreLogic's database, each of which is a snapshot of the open liens on properties in King County on the last day of the quarter. Approximately 75% of observations in the county assessor data are successfully matched with the corresponding CoreLogic records.[33] Every quarterly observation in the data is associated with up to four loans, and among

[33]The unmatched observations in the county assessor data reflect either properties that are not associated with a loan, or properties with address discrepancies that - despite some effort - preclude successful matching. Properties without associated loans are likely to comprise the bulk of unmatched observations.

the characteristics observed for each loan are its dollar amount, origination date, term and interest rate type (fixed or adjustable).[34] The CoreLogic data also contain estimates of the value of every property each quarter. The estimates are generated by an automatic valuation model (AVM) whose details are proprietary, but which is essentially a sophisticated hedonic regression model whose predictive capability has been honed over time.[35] With quarterly estimates of each property's value in hand, the information available on the associated loans allows me to estimate each property's quarterly CLTV ratio, as described in appendix section A.1.

In order to construct the predicted CLTV variable, each observation is assigned two levels of the Zillow housing price index for the appropriate zip code area: One recorded the quarter preceding observation - reflecting the level of the index roughly when decision to sell or not was made - and another recorded earlier, when the quarter the property was bought by its current owner. The Zillow housing price index is described in more detail in appendix section A.2.

The Home Mortgage Discolsure Act (HMDA) data allow me to observe homeowner characteristics recorded at the time of loan application, including loan applicants' reported income, race, ethnicity, joint application status and gender. Unfortunately loan-level HMDA data only report location at the Census tract level, so they cannot be perfectly matched with other data sources and I was only able to obtain a 55% success rate.[36] The analysis in this study is therefore conducted using aggregate HMDA variables (census tract averages) which can be assigned to all observations, and the loan-level HMDA data are used only for the purpose of reporting summary statistics.

Finally, the State Voter Registration Database (VRDB) provides information on property occupants - as opposed to owners - which I use to infer owner-occupancy. The process of inferring owner-occupancy is described in appendix A.3 and is asymmetric in the sense that properties inferred to be owner-occupied are in fact such, whereas properties *not* inferred to

[34] Only a fraction of a percent of properties are observed to have more than two loans secured against them at any one time, let alone more than four.

[35] In fact, the estimates are generated by a cascade of several automated valuation models, where the term "cascade" refers to an algorithm that pre-determines the contingent sequence in which the different models - each with its particular context-dependent pros and cons - is invoked. CoreLogic marketing materials report that automatic valuation has made significant headway over the last decade and is increasingly accepted as a substitute for human appraisal. CoreLogic is a large player in this market and it is a reasonable assumption that its AVM cascade defines the state-of-the-art.

[36] A key variable that permits matching HMDA data with other sources is the mortgage lender identity. Unfortunately, the CoreLogic data do not include this variable. See, e.g. Stroebel (2012) and Bayer et al. (2011) for examples of previous merges with loan-level HMDA data.

be owner-occupied may nevertheless still be owner-occupied. The VRDB data also provide occupants' date of birth and hence their age.

4.2 Sample selection

I restrict the sample in several ways: First, the data are limited to properties inferred to be owner-occupied, as this is the population of interest. Second, the data are limited to properties whose owners' tenure duration is between 1 and 10 years. Properties with owner tenure duration below 1 year are omitted to avoid capturing the effect of individuals "flipping" properties for a profit. Properties with owner tenure duration above 10 years are omitted because Zillow.com housing price indices only go back only to mid 1996, preventing the construction of the predicted CLTV variable for observations from 2007 with longer tenure duration. Third, the data are limited to properties whose owners do not concurrently own more than 2 properties in order to avoid capturing real estate investors whose sale decisions are likely to reflect an entire portfolio of properties. Fourth, one percent of properties with estimated CLTV values in the top and bottom half percentiles are omitted in order to avoid extreme values, which are likely to be erroneous. Finally, only observations not missing any variables necessary for the reported regressions are kept.

In total the data used consist of 1,451,696 complete observations on 107,440 properties and 116,551 property by ownership spells.[37] Table 4 reports summary statistics.

4.3 Housing prices in King County

Figure 4 reports the trajectory of the Zillow housing price index (HPI) for King County, and contrasts it with the national and California HPIs. Housing prices in the county peaked only at the end of the second quarter of 2007, a year later than housing prices did nationally. From then on they dropped continuously through the end of the sample period. As can be seen, all three HPIs follow roughly a similar trajectory, but the California HPI is more accentuated - exhibiting a steeper rise and a steeper fall - than the King County and national HPIs.

Figure 5 displays local zip code area housing price indices within the county. There is a clear common trend underlying these local HPIs, in line with the national and Cali-

[37] Of these, loan-level HMDA data is available for 963,035 quarterly observations on 70,027 properties and 72,112 property by ownership spells.

Table 4: Summary statistics

	mean	s.d.
	(1)	(2)
Owner characteristics		
Tenure duration (years)	4.83	2.34
Age (years)	45.0	11.6
Real income* ('000s 2012)	139.7	99.5
Local share college grad. (%)	49.7	17.6
Asian* (%)	6.6	
Black* (%)	1.6	
Hispanic* (%)	1.7	
Home characteristics		
Year built	1,966.9	29.2
Square footage	2,062	849
Normalized purchase price ('000s)	382	217
Loan characteristics		
Cltv ratio at purchase (%)	84.3	17.3
Mortgage term (years)	29.0	4.2
Adjustable rate mortgage (%)	26.0	
Going interest rate at purchase (%)	5.72	0.91
N observations	1,451,696	
N properties	107,440	

Notes: For detailed description of the variables see Table 2 and the accompanying notes. Variables marked with an asterisk are observed only for sub-sample successfully merged with loan-level HMDA data.

fornia HPIs. This is reassuring, because it suggests that the forces driving these housing price indices even at this local level are far removed, and are even external to the Seattle metropolitan area. Nevertheless, there is also a substantial amount of heterogeneity across zip code areas within the county.

Table 5 reports the share of mortgaged homeowners who had negative equity in their homes in the second quarter of 2012 for the 30 largest US metro areas. At 37.8%, Seattle is above the national average of 30.9%. Although it is not as hard hit as the non-coastal southwest and Florida MSAs, it has a higher share of negative equity homeowners than most of the other coastal cities.

Table 5: Percent of mortgaged owner-occupied homes with negative equity

Metro areas:					United States 30.9%
1	Pittsburgh	15.6%	16	Columbus	33.4%
2	Boston	19.6%	17	San Diego	33.9%
3	San Jose	20.3%	18	Charlotte	36.4%
4	New York	20.7%	19	**Seattle**	**37.8%**
5	Philadelphia	25.4%	20	Minneapolis-St. Paul	38.7%
6	Denver	27.1%	21	Chicago	39.2%
7	San Francisco	28.5%	22	Miami-Ft. Lauderdale	43.7%
8	Los Angeles	28.6%	23	Tampa	46.6%
9	Dallas-Ft. Worth	28.9%	24	Detroit	48.3%
10	St. Louis	30.2%	25	Sacramento	49.3%
11	Cincinnati	30.3%	26	Riverside	51.2%
12	Baltimore	30.8%	27	Phoenix	51.6%
13	Washington DC	31.3%	28	Orlando	51.9%
14	Cleveland	32.9%	29	Atlanta	54.4%
15	Portland, Oregon	33.2%	30	Las Vegas	68.5%

Notes: metro Seattle, whose core is King County, is not extreme in terms of its share of properties with negative equity, but it is above the national average and is harder hit than most of the coastal US cities.

Source: Zillow.com, 2012 Q2.

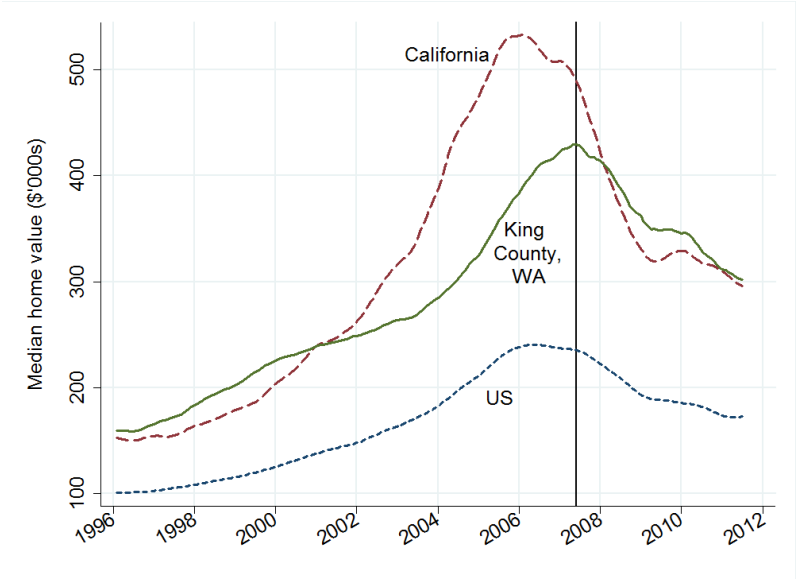


Figure 4: The King County, WA, housing price index vs. those of California and US.

Source: Zillow.com.

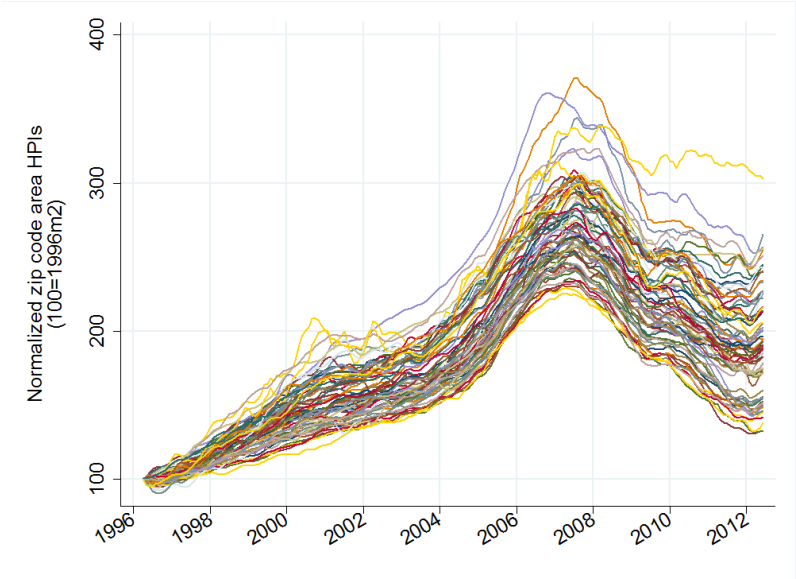


Figure 5: 72 zip code area housing price indices within King County. The indices share a clear common trend that aligns with the national HPI, suggesting that they are in fact driven by factors far removed from the zip code areas and from the Seattle metro area in general. Nevertheless, they also exhibit substantial variation within the county.

Source: Zillow.com.

5 Empirical estimates

This section is organized as follows. Section 5.1 observes the data unconditionally, and section 5.2 reports naive and reduced form estimates using actual and predicted equity, respectively. Section 5.3 reports IV estimates. Section 5.4 provides evidence on the sensitivity of sale rates to price innovations and section 5.5 addresses loss aversion.

5.1 Unconditional observation

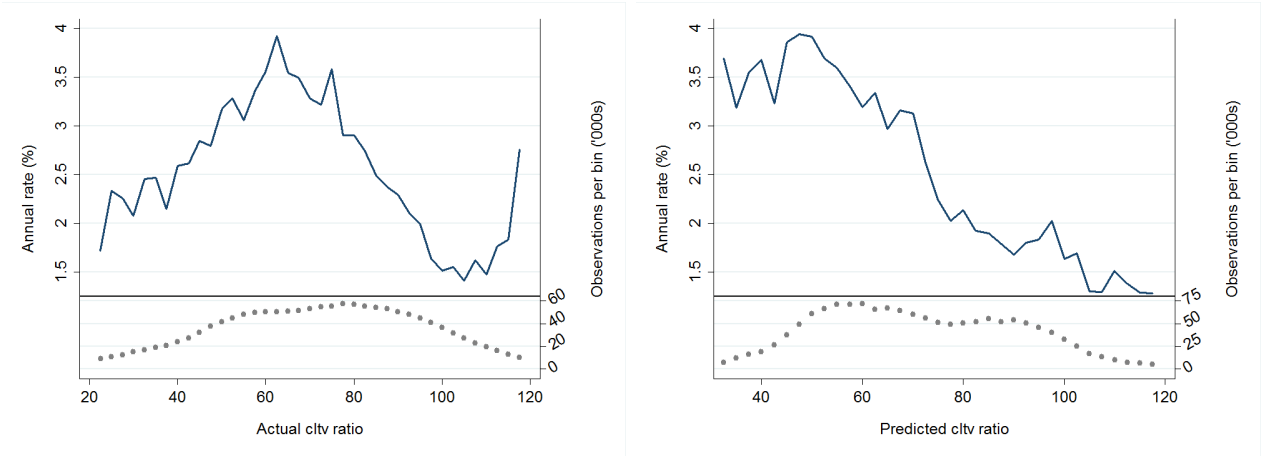
Unconditional sale rates: Figure 6a shows the unconditional annual rate of arm's length home sales by actual CLTV ratio. Observations on the left side are properties with low CLTV ratios, meaning their owners have high equity in them, whereas observations on the right are properties with high CLTV ratios and low homeowner equity. Properties with CLTV ratios above 100% have negative equity. The Figure indicates that the sale rate is increasing for low CLTV ratios, peaks at approximately 65% CLTV, and then declines until roughly 105%, after which it rises again sharply.

The downward slope from roughly 65% to 105% CLTV is consistent with the notion that diminished equity reduces the probability of sale and, indirectly, mobility. In contrast with the focus on negative equity in previous research and in the media, the raw data suggest that the effect of equity on mobility begins well before homeowners reach negative equity levels. It is intuitive to attribute the sharp upward slope at the right end of the Figure to distressed sales at highly negative levels of equity.[38] This upward slope does not have a counterpart in the previous literature, perhaps because distressed sales (and very negative equity) were not as prevalent in the real estate downturn of the late 1980's and early 1990's. The upward slope below 65% CLTV probably reflects systematic differences between homeowners with different CLTV ratios.[39]

In contrast, Figure 6b shows the unconditional rate with respect to predicted CLTV. The same downward slope emerges between 65% and 105% predicted CLTV, but on either side of this interval are two regions that appear flat in comparison to the previous Figure. The

[38] The sale indicator which I use does not include sales labeled by the county assessor as foreclosures, but it does include short sales. Although some sales are marked as short sales in assessor comments, these markings do not appear to be systematic and there is no reason to believe that they capture all short sales.

[39] A possible explanation for this slope that can be ruled out is that of dynamic selection, whereby homeowners with higher propensity to sell - e.g. owners of so-called starter homes - are more likely to have sold by the time they would otherwise have obtained low CLTV ratios. However, as shown in appendix section B this upward slope is roughly similar in raw data compiled for homeowners with tenure duration capped at 10, 5 and even just 3 years, suggesting that dynamic selection is not the culprit.



(a) Actual CLTV

(b) Predicted CLTV

Figure 6: Unconditional annual arm's length sale rates by actual and predicted CLTV ratios. The downward slope from roughly 65% to 105% CLTV in both panels is consistent with the notion that diminished equity reduces the probability of sale and, indirectly, mobility. The upward slope on the right hand side of panel *a* likely reflects short sales, and does not have a counterpart in earlier studies. In contrast, the upward slope on the left hand side of panel *a* raises the concern that actual CLTV is endogenous.

Notes: Bin means are reported for 2.5% CLTV ratio bins. The number of observations per bin refers to property-by-quarter observations.

upward sloping region on the left hand side of the actual CLTV (6a) Figure is a primary cause for concern that estimates using actual CLTV are endogenous. Its elimination in the predicted CLTV Figure (6b) is reassuring, and suggests that estimates using predicted CLTV can be more safely interpreted to reflect a causal effect of equity. The second upward slope, at right end of the actual CLTV Figure (6a), does not raise the same degree of concern with respect to endogeneity, but that it too disappears in the predicted CLTV Figure (6b) suggests that it may reflect systematic differences between homeowners with different CLTV ratios as well.

The distribution of actual and predicted CLTV ratios: the lower panels of Figures 6a and 6b show the distribution of observations over the ranges of actual and predicted CLTV ratios. Casual inspection of these panels is informative. Recall that actual CLTV reflects homeowners' financial decisions, whereas predicted CLTV does not. Homeowners who made down payments greater (lesser) than 80% have actual CLTV ratios below (above)

their predicted CLTV. Comparing the left tail of the Figures, which is thicker in the Figure 6a, suggests that homeowners with low actual CLTV probably made down payments greater than the ubiquitous 20%. In addition, homeowners with low predicted CLTV ratios are generally able to engage in additional borrowing against their homes, and in so doing raise their actual CLTV ratio above their predicted one. Such borrowing shifts density to the right in Figure 6a relative to Figure 6b, and is apparent owing to the roughly bi-modal distribution in both Figures: the modes are similarly located in both Figures, but the right hand one is denser for actual CLTV than it is for predicted CLTV. Such borrowing is also apparent in the right tail, which is thicker in the Figure 6a.

Balancing: Table 6 sheds light on observable differences between properties with different actual CLTV ratios. The first three columns report a set of covariate means for sub-samples with low, medium and high CLTV ratios, defined as [30,60), [60,90) and [90,120) percent actual CLTV, respectively. Homeowners with higher actual CLTV ratios tend to be younger, poorer and less educated, and to live in smaller, less expensive homes.[40] Consistent with the being less affluent, they are also more likely to have adjusTable rate mortgages (featuring lower monthly payments, all else equal). Homeowners with higher actual CLTV ratios belong to this category because they make smaller down payments and pay off their mortgages more slowly (over longer terms), and also because they tend to have shorter tenure durations, which over the observed period mostly implies having experienced more negative housing prices changes.[41] All in all, homeowners with different actual CLTV ratios are quite dissimilar.[42]

The last three columns of Table 6 report the same covariate means for sub-samples with low, medium and high *predicted* CLTV ratios. With respect to most characteristics, the stark differences between homeowners with different actual CLTV ratios shrink dramatically or even disappear when predicted CLTV is considered instead. That predicted CLTV, which derives only from changes in local housing prices, is substantially more balanced than actual CLTV, suggests that estimates using predicted CLTV can more confidently be regarded to reflect a causal effect of equity.

[40] And also in slightly newer homes, perhaps suggesting that they tend to be farther from the city center.

[41] Some additional notes with respect to actual CLTV in Table 6: (1) while minorities are only a small share of the population in metro Seattle, it is noteworthy that higher actual CLTV ratios are more prevalent among black and Hispanic homeowners, while the opposite is true of Asian homeowners. (2) The higher going interest rates associated with higher CLTV ratios should not be interpreted as evidence that such homeowners receive "worse" loans, because these are going rates determined by the timing of purchase - not the rate actually associated with an observed property's mortgage.

[42] To what degree this holds for the data used in previous studies is uncertain, although Table 2 of Chan (2001) suggests that her sample was qualitatively similar.

Table 6: Average observable covariates by CLTV ratio

	Actual CLTV ratio in			Predicted CLTV ratio in		
	[30,60)	[60,90)	[90,120)	[30,60)	[60,90)	[90,120)
	(1)	(2)	(3)	(4)	(5)	(6)
			Owner characteristics			
Tenure duration (years)	5.80	4.59	4.03	6.33	4.48	3.45
Age (years)	47.6	44.4	42.3	46.3	44.6	43.8
Real income* ('000s)	147	140	130	141	139	141
Local share college grad. (%)	53.9	50.1	44.4	50.5	50.3	47.7
Asian* (%)	6.9	6.6	6.1	5.9	6.9	7.4
Black* (%)	0.9	1.6	2.4	1.5	1.6	1.6
Hispanic* (%)	1.0	1.8	2.5	1.4	1.8	2.0
			Home characteristics			
Year built	1966.0	1967.1	1967.6	1963.0	1967.7	1970.5
Square footage	2,219	2,068	1,864	2,033	2,084	2,061
Normalized purchase price ('000s)	435	381	318	390	386	365
			Loan characteristics			
Cltv ratio at purchase (%)	75.7	86.1	93.1	84.1	84.5	83.8
Mortgage term (years)	27.8	29.6	30.1	28.3	29.2	29.5
Adjustable rate mortgage (%)	18.6	28.4	31.4	23.9	28.5	23.2
Going interest rate (%)	5.57	5.68	5.96	5.72	5.69	5.81
N observations	377,331	642,620	360,250	440,528	693,648	307,301
N properties	30,093	49,832	29,823	37,015	51,447	26,812

Notes: Whereas owners, homes and loans differ systematically across *actual* equity levels, they differ far less starkly across *predicted* equity levels, along most dimensions. For detailed description of the variables see Table 2 and the accompanying notes. Variables marked with an asterisk are observed only for sub-sample successfully merged with loan-level HMDA data.

However, predicted CLTV does correlate with homes' year of construction and with homeowners' tenure duration and age. These correlations are artifacts of the prolonged monotonic rise in housing prices that preceded the current housing crisis, and that generates a positive correlation between low predicted CLTV and any variable that positively increases with tenure duration. As mentioned in section 3, controlling for a polynomial of tenure duration helps alleviate concerns of endogeneity that derive from these correlations, and related concerns with respect to dynamic selection are addressed in appendix section B and ultimately relaxed.[43]

5.2 Naive and reduced form estimates

I estimate two sequences of linear probability models that are specific forms of (6). The first sequence is of the form

$$sale_{it} = \Sigma_j \beta_j \mathbf{1}\{CLTV_{it} \in [j, j+h)\} + \theta_{lt} + \mathbf{X}_{it}\delta + \epsilon_{it}, \tag{8}$$

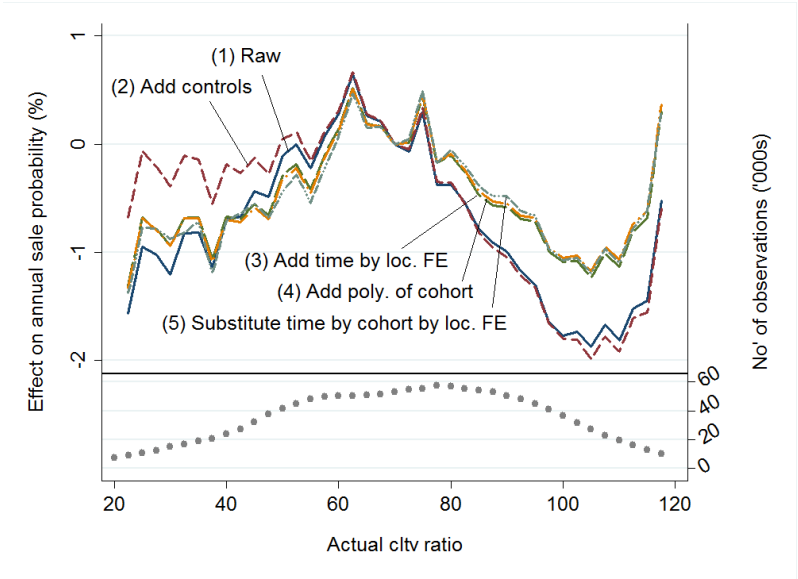
where each CLTV ratio bin, j, has bandwidth h set to 2.5%. At each step in the sequence I alter the set of covariates $\mathbf{X}_{it}$, except for one key step in which I add the time by location fixed effects, θ_{lt}, instead. Owing to the way the CLTV bins are set up, this specification is convenient for visual presentation. Naive estimates with respect to actual CLTV are reported in Figure 7a, and reduced form estimates with respect to predicted CLTV are reported in Figure 7b. In parallel, I estimate another sequence of the form

$$sale_{it} = \Sigma_j \beta_j \mathbf{1}\{CLTV_{it} > j\} + \theta_{lt} + \mathbf{X}_{it}\delta + \epsilon_{it}, \tag{9}$$

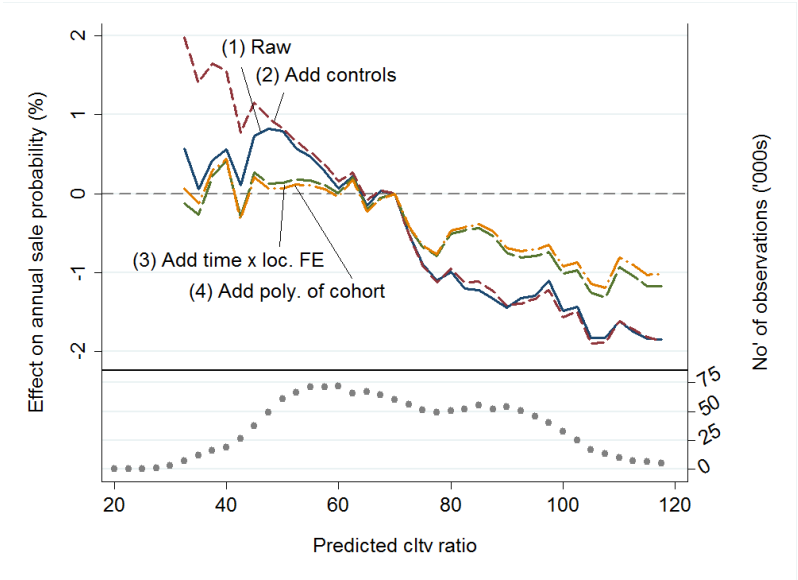
in which each of the summed $\mathbf{1}\{CLTV_{it} > j\}$ components is an indicator that the CLTV ratio exceeds a cutoff level, j, set at 10% intervals. This setup implies that the β_j coefficients capture the marginal effect of shifting from the $[j-10, j)$ CLTV bin to the current $[j, j+10)$ CLTV bin, and is convenient for numerical presentation. Naive estimates with respect to actual CLTV are reported in Table 7, and reduced form estimates with respect to predicted CLTV are reported in Table 8.[44]

[43] Mid-range homeowners are likelier to have an adjustable rate mortgage than both other groups, reflecting that adjustable rate mortgage originations peaked sharply between 2003 and 2006, and that the general price trajectory since then puts buyers from those years mostly in the 60% to 90% predicted CLTV bin.

[44] Standard errors in estimates of (9) are clustered at the zip code area level, which serves multiple purposes. First, such clustering allows for arbitrary correlation of errors within observations of the same property (within and between ownership spells), accounting for serial autocorrelation. Second, such clustering allows



(a)



(b)

Figure 7: Conditional effects of actual and predicted CLTV ratios on annual sale rates. Estimates with respect to actual and predicted CLTV ratios both indicate a decrease in sale probabilities from roughly 70% to 100% CLTV. However, actual CLTV estimates exhibit implausible non-monotonicity. While the upward slope at the right end of the actual CLTV panel can intuitively be attributed to distressed sales, the upward slope on the left hand side is probably the result of systematic differences between homeowners with different actual CLTV ratios. That these upward slopes do not appear in the predicted CLTV estimates is reassuring with respect to the validity of these estimates' causal interpretation.

Notes: Reported estimates are coefficients of fixed effects for 2.5% CLTV bin indicators from (8), whose levels have been normalized such that the effect is zero at 70% CLTV. The steps of the analysis, 1 to 4, correspond to columns 1 to 4 of Table 7. The number of observations per bin refers to property-by-quarter observations.

Table 7: The effect of actual CLTV ratio on annual sale probability

Actual cltv ratio	(1)	(2)	(3)	(4)	(5)
Over 30%	0.164	-0.050	0.103	0.112	0.105
	(0.149)	(0.148)	(0.145)	(0.146)	(0.152)
Over 40%	0.440***	0.078	0.242	0.223	0.319*
	(0.157)	(0.155)	(0.157)	(0.158)	(0.164)
Over 50%	0.495***	0.239*	0.422***	0.428***	0.304**
	(0.143)	(0.141)	(0.140)	(0.141)	(0.141)
Over 60%	0.408***	0.335***	0.518***	0.540***	0.601***
	(0.114)	(0.120)	(0.118)	(0.118)	(0.120)
Over 70%	-0.384***	-0.386***	-0.163	-0.147	-0.099
	(0.118)	(0.116)	(0.115)	(0.115)	(0.117)
Over 80%	-0.609***	-0.648***	-0.402***	-0.379***	-0.345***
	(0.109)	(0.112)	(0.109)	(0.110)	(0.113)
Over 90%	-0.612***	-0.631***	-0.373***	-0.364***	-0.371***
	(0.109)	(0.112)	(0.110)	(0.109)	(0.112)
Over 100%	-0.506***	-0.553***	-0.358***	-0.338***	-0.381***
	(0.106)	(0.107)	(0.111)	(0.111)	(0.115)
Over 110%	0.619***	0.618***	0.744***	0.760***	0.766***
	(0.143)	(0.143)	(0.143)	(0.145)	(0.150)
Owner/home/loan ctrls		+	+	+	+
Time by Loc. FE			+	+	
Poly. of Cohort				+	
Time by Cohort by Loc. FE					+
N observations			1,451,278		
N properties			107,412		
Avg. outcome level			2.72%		

Notes: this Table corresponds to Figure 7a. The dependent variable is 100× annualized arm's length sale indicator and reported coefficients are marginal effects, measured in percentage points, so shifting from 39% to 41% or to 51% actual CLTV raises the annual probability of sale estimated in column 1 by 0.440 or by 0.935 ($= 0.440 + 0.495$) percentage points, respectively. Sample includes only owner-occupied homes, whose occupants have tenure duration of 1 to 10 years and do not concurrently own more than two properties. For the list and description of included home, owner and loan characteristics see Table 2 and the accompanying notes. Standard errors clustered in 72 zip code areas. One, two and three asterisks reflect statistical significance at the 10, 5 and 1 percent levels, respectively.

Table 8: The effect of predicted CLTV ratio on annual sale probability

Predicted cltv ratio	(1)	(2)	(3)	(4)
Over 30%	-0.934	-0.919	0.347	-0.071
	(0.898)	(0.890)	(1.147)	(1.173)
Over 40%	0.337	-0.517**	0.250	-0.087
	(0.206)	(0.228)	(0.253)	(0.366)
Over 50%	-0.096	-0.468***	0.075	0.082
	(0.136)	(0.137)	(0.146)	(0.176)
Over 60%	-0.479***	-0.474***	-0.123	-0.020
	(0.107)	(0.109)	(0.120)	(0.160)
Over 70%	-0.633***	-0.686***	-0.405***	-0.302***
	(0.100)	(0.100)	(0.104)	(0.106)
Over 80%	-0.599***	-0.492***	0.005	0.083
	(0.096)	(0.098)	(0.111)	(0.115)
Over 90%	-0.112	-0.245***	-0.272***	-0.198*
	(0.078)	(0.082)	(0.104)	(0.110)
Over 100%	-0.288***	-0.316***	-0.310***	-0.244***
	(0.106)	(0.098)	(0.089)	(0.094)
Over 110%	-0.192	-0.069	0.127	0.213
	(0.165)	(0.162)	(0.175)	(0.171)
Owner/home/loan ctrls		+	+	+
Time by Loc. FE			+	+
Poly. of Cohort				+
N observations		1,451,278		
N properties		107,412		
Avg. outcome level		2.72%		

Notes: this Table corresponds to Figure 7b. The dependent variable is 100× annualized arm's length sale indicator and reported coefficients are marginal effects, measured in percentage points, so shifting from 59% to 61% or to 71% predicted CLTV raises the annual probability of sale estimated in column 1 by -0.479 or by -1.112 ($= -0.479 - 0.633$) percentage points, respectively. Sample includes only owner-occupied homes, whose occupants have tenure duration of 1 to 10 years and do not concurrently own more than two properties. For the list and description of included home, owner and loan characteristics see Table 2 and the accompanying notes. Standard errors clustered in 72 zip code areas. One, two and three asterisks reflect statistical significance at the 10, 5 and 1 percent levels, respectively.

Step 1: In the first step of the sequence I adapt the raw data into a form that is easily comparable to estimates from the subsequent steps of the sequence. The solid lines labeled 1 in Figures 7a and 7b are identical to the unconditional ones in Figures 6a and 6b, respectively, other than they have been shifted downwards so that the effect associated with a 70% CLTV ratio is normalized to zero.[45] The first columns of Table 7 and 8 report the parallel estimates numerically (from (9)), where $\mathbf{X}_{it}$ includes only a constant. The estimates record the same patterns as the Figures.

Step 2: In the second step of the sequence I condition the estimates on a broad set of observed owner, home and loan characteristics. In particular, these controls include a cubic of properties' initial CLTV ratios, reflecting the initial down payment, and a cubic of their owners' tenure duration, which controls for a so-called "seasoning effect" (Caplin et al., 1997) whereby owners propensity to sell evolves over time. The cubic of tenure duration also controls for the average degree of dynamic selection over time.

The dashed lines labeled 2 in Figures 7a and 7b report the effect associated with CLTV conditional on these controls. From roughly 50% CLTV up these lines track the average effect associated with CLTV in the raw data in both Figures, suggesting that non-random selection into CLTV bins above 50% CLTV is not of great concern inasmuch as the controls provide a good indication with respect to confounding unobservables. Below 50% CLTV, however, the controls influence the effect associated with CLTV, suggesting that non-random selection into CLTV bins is a concern. Unreported estimates conditioning on subsets of the controls indicate that the shift in estimates from step 1 to step 2 is driven primarily by homeowners' initial down payments.

Step 3: In this step I condition estimates on a set of quarter by zip code area fixed effects. The estimates from this step appear as long-dashed lines labeled 3 in Figures 7a and 7b. Conditioning on time by location cells causes a counter-clockwise rotation of the estimates.

for arbitrary correlation of errors within observations made in the same time period and the same place, and therefore experiencing the same contemporary local economic conditions. Third, such clustering allows for arbitrary correlation of errors within observations sharing the same cohort of purchase, and therefore experiencing common cohort-related shocks, such as a common set of housing price expectations. Finally, such clustering accounts for spatial and combined spatial and temporal correlation of errors among nearby properties, as long as they are within the same zip code area.

[45] This normalization anticipates that once a set of fixed effects is included in the regression, the estimates of β_j in (8) are no longer informative with respect to the absolute level of home sales by CLTV bin, but only with respect to the relative level across different CLTV bins. 70% CLTV proves to be a convenient point for normalization given the results that follow.

Low CLTV ratios - actual and predicted - tend to be observed at times and places in which the housing market was "hot", with both prices and transaction volumes high. Accounting for this tendency reduces the effect associated with low CLTV ratios, generating a downward shift on the left hand side of the Figures - from the lines labeled 2 to the lines labeled 3. Conversely, higher CLTV ratios tend to be observed at times and places in which the housing market was "cold", with both prices and transaction volumes low. Accounting for this raises the effect associated with high CLTV ratios, generating an equally substantial - but opposite - upward shift on the right hand side of the Figure.

Conditioning estimates on the time by location cell of observation reduces the estimated magnitude of the effect between 70% and 100% CLTV by half, from roughly 2 percentage points slightly less than 1. Given that the average annual sale rate at 70% actual (predicted) CLTV in the observed sample is only 3.3% (3.1%) per year, this difference is quite substantial. The third column of Tables 7 and 8 report numerical estimates corresponding to this step of the analysis. The marginal effect of each CLTV bin is estimated to be higher than in the previous column, consistent with a counter-clockwise rotation.

Step 4: In this step I account for the properties' cohort of purchase by including a cubic of the time of purchase in $\mathbf{X}_{it}$. The importance of this step is in accounting for systematic differences in the pool of buyers at different times. The perforated lines labeled 4 in Figures 7a and 7b report estimates of this step. These lines hardly differ from those of the previous step, implying that conditional on all previous controls, accounting for factors correlated with time of purchase barely influences the estimates.

A closely related issue is dynamic selection, whereby the composition of the pool of homeowners in a cohort systematically changes over time as homeowners sell their homes and leave the cohort selectively. Appendix section B provides evidence that dynamic selection is unlikely to be driving the observed results.

Step 5: In a final step of the sequence, I replace the set of time by location fixed effects, θ_{lt}, with a fully saturated three way set of time by cohort by location fixed effects, ψ_{lct}. This step can only be performed with respect to actual CLTV, because it eliminates all variation in predicted CLTV by construction. The double-perforated line labeled 5 in Figure 7a and column 5 of Table 7 report the estimates, which are almost indistinguishable from those of the previous two steps.

The purpose of this step is to shed light on the source of identifying variation driving the results with respect to actual CLTV. This result implies that, in contrast to the results

obtained using predicted CLTV, those obtained using actual CLTV are *not* driven by changes in aggregate local housing prices. Rather, they are driven primarily by homeowners' financial decisions (as well as by property-level home values within time by cohort by location cells). Given that variation derived from homeowners' financial decisions is likely to be endogenous, the fact that the results using actual CLTV persists through this step casts doubt on their causal nature and highlights the advantage of predicted CLTV as in reflecting more plausibly exogenous variation.

5.3 Instrumental variable estimates

This section reports instrumental variable estimates of the effect of equity on home sales, using predicted equity as an instrument for actual CLTV. In order to allow for a non-linear effect of equity I estimate the following specification that includes a cubic of actual CLTV, and I use a cubic of predicted CLTV as the set of instruments.[46]

$$sale_{it} = \beta_1 CLTV_{it} + \beta_2 CLTV_{it}^2 + \beta_3 CLTV_{it}^3 + \theta_{lt} + \mathbf{X}_{it}\delta + \epsilon_{it}. \tag{10}$$

The set of controls includes the full set of owner, home and loan characteristics, a set of quarter by zip code area fixed effects and a cubic of the cohort of purchase. Estimates are reported in Table 9. As the results using a polynomial of CLTV may differ from earlier results using CLTV bin indicators, I report naive OLS estimates of specification (10) in column 1 and reduced form estimates in column 2. Column 3 reports the IV estimates.

The naive OLS estimates in column 1 are similar to those in Table 7. The effect associated with higher actual CLTV ratios is positive at first, peaking aroung 65% CLTV, and then falls until a CLTV ratio of approximately 100%. The sharp upward slope observed for homeowners with negative equity at the right end of Figure 7a does show up in the estimate, but not with sufficient precision to be statistically significant at standard levels. Similarly, the reduced form estimates in column 2 are similar to those obtained in Table 8, in that the predicted CLTV range below 70% CLTV is essentially flat, followed by a steep, statistically significant drop from 70% to 100% predicted CLTV, and then a flat region above 100% predicted CLTV.

The IV estimates are useful for obtaining the magnitude of the effect of predicted CLTV on sale probabilities in terms of actual CLTV.[47] The marginal effect of shifting from an actual

[46] A cubic is the lowest-order polynomial capable of capturing the rough shape of the curve in Figure 7a. Results using higher-order polynomials do not differ substantially.

[47] The IV estimates are larger in (absolute) magnitude than the reduced form estimates because every

Table 9: IV estimates of the effect of CLTV ratio on home sale rates

	OLS	Reduced form	IV
Effect of shifting cltv	(1)	(2)	(3)
From 30% to 40%	0.324***	0.180	1.069
	(0.060)	(0.503)	(1.464)
From 40% to 50%	0.512***	-0.014	0.315
	(0.048)	(0.290)	(0.524)
From 50% to 60%	0.402***	-0.147	-0.248
	(0.045)	(0.189)	(0.460)
From 60% to 70%	0.125***	-0.225	-0.618
	(0.035)	(0.142)	(0.491)
From 70% to 80%	-0.189***	-0.253**	-0.792**
	(0.031)	(0.105)	(0.376)
From 80% to 90%	-0.409***	-0.237***	-0.768***
	(0.036)	(0.074)	(0.268)
From 90% to 100%	-0.405***	-0.183***	-0.545**
	(0.038)	(0.061)	(0.239)
From 100% to 110%	-0.047	-0.097	-0.118
	(0.049)	(0.065)	(0.190)
From 110% to 120%	0.797***	0.016	0.513
	(0.111)	(0.127)	(0.705)
Owner/home/loan ctrls	+	+	+
Time by Loc. FE	+	+	+
Poly. of Cohort	+	+	+
N observations		1,451,696	
N properties		107,440	
Avg. outcome level (percent)		2.72%	

Notes: The dependent variable is 100× annualized arm's length sale indicator. The explanatory variables of interest form a quartic of actual CLTV, and are instrumented by a quartic of predicted CLTV. Reported coefficients are specific marginal effects, measured in percentage points. Angrist-Pischke F-statistics (p-values) for the CLTV, $CLTV^2$, $CLTV^3$ and $CLTV^4$ first stage regressions are 59.00 (0), 12.31 (< 0.001), 42.21 (0) and 61.16 (0), respectively, and the partial R^2 for each of the first-stage regressions is 0.026, 0.036, 0.044 and 0.048, respectively. Sample includes only owner-occupied homes, whose occupants have tenure duration of 1 to 10 years and do not concurrently own more than two properties. For the list and description of included home, owner and loan characteristics see Table 2 and the accompanying notes. Standard errors clustered in 72 zip code areas. One, two and three asterisks reflect statistical significance at the 10, 5 and 1 percent levels, respectively.

CLTV ratio just below 70% to just over 100% is approximately a 2.1 percentage point drop. Because the annual rate of arm's length sales observed in the sample below 70% CLTV is roughly 3.07%, a 2.1% drop is quite dramatic and means that a shift to 100% CLTV almost completely eliminates arm's length home sales.[48] As in the naive OLS estimate, the upward slope for homeowners with negative equity does show up, but it can not be significantly distinguished from zero.

5.3.1 The OLS bias

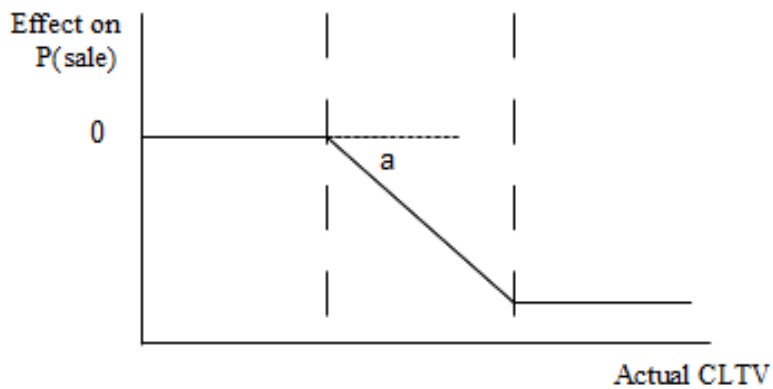
Focusing on the coefficients in the 70% to 100% CLTV range, note that the IV estimates are greater in (absolute) magnitude than the naive OLS estimates, indicating that the latter are biased towards zero. Figures 8a-8c illustrate the situation in a stylized manner. Figure 8a depicts the true effect of equity on the probability of home sale, reflecting the reduced form and IV results with respect to the range and magnitude of the effect, respectively. The true effect is confined to the 70% and 100% CLTV range, in which its slope is steep and negative. If the sum of confounding effects is *positive* and has a moderate slope, as depicted in Figure 8b, then the super-position of the true and confounding effects yields a non-monotonic pattern resembling the naive OLS estimates. In this non-monotonic pattern shown in Figure 8c sales increase in CLTV below 70%, then decrease between 70% and 100% CLTV - but with a more moderate slope than the true effect - and then increase again above 100%. Thus, a positive sum of confounding effects can explain the bias of the naive OLS estimates towards zero in the 70% to 100% CLTV range. But why might the sum of confounding effects associated with equity be positive?

There could be any number of confounding effects that generate a positive slope. An example mentioned earlier is that greater household prosperity may induce both larger down payments and lower propensities to move, thereby biasing OLS regressions of sales on CLTV upward (similarly, greater prosperity may induce shorter mortgage terms, to the same effect). Another example mentioned earlier involves renovations: As long as renovations which increase the value of a home and decrease CLTV tend to coincide with intentions to stay in the home for an extended period, they give rise to a positive slope.[49] A separate source of

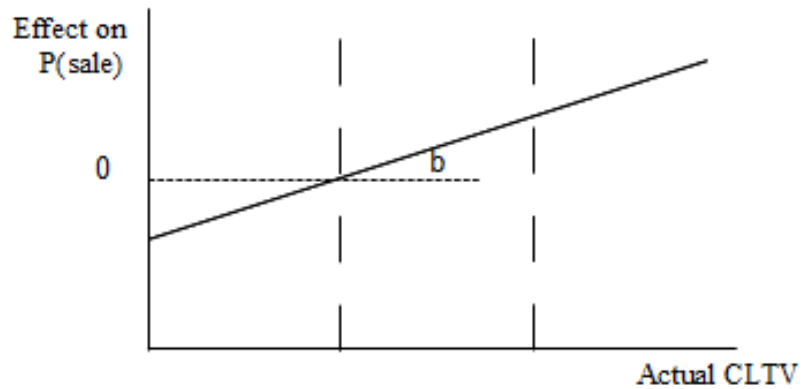
unit of predicted CLTV shifts actual CLTV only by a fraction of a unit, as indicated (unconditionally and linearly) in Figure 3. For a textbook example see Cameron and Trivedi (2005) section 4.8.3 (page 98).

[48] At its highest, in 2007, the annual rate of arm's length sales observed in the sample below 70% CLTV is roughly 4.5%, and at its lowest, in 2011, it is roughly 2.0%. The observed 2.1% drop is substantial compared to either of these figures.

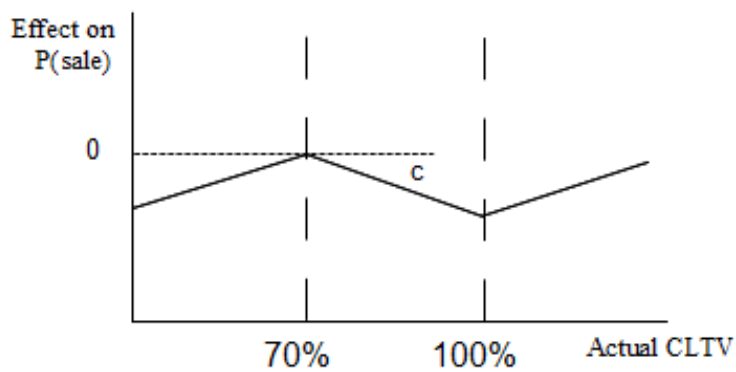
[49] Renovations may show up in the data if they are of a type that is recorded by the county assessor's office, and if they the recorded variables are reflected in CoreLogic's AVM.



(a) The causal effect of equity. This is the effect reflected by the IV estimates.



(b) The sum of confounding effects associated with equity.



(c) The super-position of 8a and 8b given $|a| > |b|$. This is the confounded "effect" reflected by the naive OLS estimates.

Figure 8: Stylized illustration of the relationship between naive OLS and IV estimates. Figure 8a depicts the true effect of equity on the probability of home sale, reflecting the reduced form and IV results with respect to the range and magnitude of the effect, respectively. Figure 8b portrays the sum of confounding effects as having a moderate positive slope. The super-position of Figures 8a and 8b yields a non-monotonic pattern resembling the naive OLS estimates, shown in Figure 8c.

bias could be measurement error in the (actual) CLTV variable, which may arise for multiple reasons. For instance, as I do not observe payment histories directly, I assume that payments towards principal are paid on time and set according to the standard (fixed payment) amortization formula. Neither of these assumptions is likely to be true in general, attenuating the naive OLS estimates to zero.[50]

The "better homes" hypothesis: Perhaps the most important confounding effect stems from the distinction between aggregate and individual changes in home values (recall that both types of changes are reflected in actual CLTV, but only aggregate changes are reflected in predicted CLTV). Suppose certain homes within a quarter $\times$ zipcode cell are, on average, more appealing to incumbent and potential residents - e.g. because of their layout or precise location - in a way that is unobservable to the econometrician. The incumbent residents of such "better homes" will be less likely to sell them, so compared to observably similar homes they will have a lower sale rate, or lesser "churn". At the same time, the demand for "better homes" will be more robust, in the sense that when the demand for housing falls, these will be the last homes in which buyers loose interest. Thus, when housing prices fall "better homes" will lose less of their value - thereby maintaining lower CLTV ratios - than the average observably similar home. Combining these two observations implies that "better homes" will tend to have both lower sale rates (less "churn") and lower CLTV ratios in periods when housing prices are falling, generating the upward slope shown in Figure 8b. A simulation and empirical evidence reported in appendix section C indicate that the "better homes" hypothesis is likely to explain at least part of the OLS bias.

5.4 Sensitivity of sales to housing price innovations

In this section I ask whether the sale probability of properties with low or negative equity is more sensitive to the latest changes in local housing prices. If the observed reduction in the home sale rates from 70% to 100% CLTV is caused by insufficient equity then a drop in housing prices should reduce the sale probability *around this range* by further eroding equity, and an increase in prices should do the opposite. There is no clear reason why housing price innovations should affect sale probabilities *differentially* by CLTV ratio, unless insufficient equity is an issue.

To estimate the sensitivity of sale probabilities with respect to housing price innovations

[50] Attenuation to zero amounts to an upward bias in the 70% to 100% CLTV range, but attenuation alone is not sufficient to produce the slope shown in Figure 8b.

I estimate a sequence of linear probability models of the following form

$$\begin{aligned} sale_{it} =& \alpha(P_{i,t} - P_{i,t-4}) + \Sigma_j \beta_j \mathbf{1}\{CLTV_{i,t-4} > j\} \\ &+ \Sigma_j \gamma_j [(P_{i,t} - P_{i,t-4})\mathbf{1}\{CLTV_{i,t-4} > j\}] + \theta_{lt} + \mathbf{X}_{it}\delta + \epsilon_{it}. \end{aligned} \tag{11}$$

This specification is similar to that in (9), but it contains CLTV ratios that are lagged by one year (expressed as four quarters). In addition, it includes the change in the value of property i over the last year, $(P_{i,t} - P_{i,t-4})$, as a covariate, and a set of interaction terms between this price innovation and the one-year lagged CLTV ratios. The γs are the coefficients of interest.

Table 10 reports estimates of (11) for specifications that parallel steps 1 to 5 of the previous analysis. All four specifications reveal a similar picture: the sensitivity of sale probabilities with respect to price innovations first become more positive at lagged CLTV ratios of roughly 70%, and then become more negative above 100%. This finding means that in the window ranging from roughly 70% CLTV to 100%, price innovations have a more positive effect on the likelihood of sale than they do at outside this range, consistent with a causal effect of equity.

Note that the negative interaction effect above 100% CLTV is even greater in magnitude than the sum of the positive effects at 60% and 70% CLTV. This result is consistent with the possibility that most sales observed to occur with negative equity are short sales, and it reveals that the likelihood of such sales is negatively correlated with price innovations, i.e. short sales are more common when prices are falling than when they are rising.

5.5 Loss Aversion

Genesove and Mayer (2001) show that sellers who have lost in home value tend to set higher asking prices and are therefore less likely to succeed in selling their homes. The fact that falling values mechanically reduce equity suggests that low equity may non-causally correlate with depressed home sales simply because of loss aversion. However, loss aversion should not affect the sale probability of sellers who experience both falling home values and net gain in equity. Thus, to ensure that my results are not driven by this alternative explanation, I re-estimate the models above on the sub-sample of owners who experience a net gain in equity.[51]

[51] Net loss and net gain refer to a home's current value minus its purchase price being negative or positive, respectively. A homeowner who first experiences a gain in home value and then experiences a greater loss is

Table 10: Sensitivity of annual sale rates to price innovations by lagged actual CLTV

1 yr lagged CLTV ratio × year-on-year Δ value (×10,000)	(1)	(2)	(3)	(4)	(5)
Over 30%	0.019	0.019	-0.005	-0.005	0.002
	(0.037)	(0.037)	(0.036)	(0.036)	(0.037)
Over 40%	-0.062**	-0.065**	-0.077**	-0.081**	-0.088***
	(0.031)	(0.031)	(0.032)	(0.032)	(0.031)
Over 50%	0.044	0.035	0.017	0.013	0.007
	(0.033)	(0.033)	(0.033)	(0.034)	(0.036)
Over 60%	0.099***	0.103***	0.080**	0.079**	0.070**
	(0.033)	(0.033)	(0.032)	(0.032)	(0.032)
Over 70%	0.046	0.062*	0.064**	0.064**	0.083**
	(0.032)	(0.032)	(0.032)	(0.032)	(0.032)
Over 80%	-0.008	0.009	0.019	0.019	0.022
	(0.035)	(0.035)	(0.035)	(0.035)	(0.036)
Over 90%	-0.111**	-0.088*	-0.058	-0.058	-0.026
	(0.048)	(0.048)	(0.049)	(0.049)	(0.052)
Over 100%	-0.268***	-0.282***	-0.204**	-0.204**	-0.226**
	(0.098)	(0.098)	(0.097)	(0.097)	(0.101)
Over 110%	0.132	0.088	0.119	0.125	0.100
	(0.138)	(0.139)	(0.139)	(0.139)	(0.146)
1 yr lagged CLTV ratio bins	+	+	+	+	+
YoY change in value	+	+	+	+	+
Owner/home/loan ctrls		+	+	+	+
Time by Loc. FE			+	+	
Poly. of Cohort				+	
Time by Cohort by Loc. FE					+
N observations			1,451,278		
N properties			107,412		
Avg. outcome level			2.72%		
Median home value ('000s)			422		

Notes: The dependent variable is 100× annualized arm's length sale indicator. Reported coefficients are marginal effects, measured in percentage points. For example: the coefficient for "over 60%" in column 1 indicates that if actual CLTV is greater than 60%, a $10,000 year-on-year increase in value raises the annual prob. of sale by 0.099 percentage points. Sample includes only owner-occupied homes, whose occupants have tenure duration of 1 to 10 years and do not concurrently own more than two properties. For the list and description of included home, owner and loan characteristics see Table 2 and the accompanying notes. Standard errors clustered in 72 zip code areas. One, two and three asterisks reflect statistical significance at the 10, 5 and 1 percent levels, respectively.

Figures 9a and 9b summarize the analysis within the gain-only sample, and their resemblance to the corresponding Figures for the full sample (7a and 7b) is striking. Corresponding estimates are reported in Table 11 for specifications that include controls for the full set of owner, home and loan characteristics as well as quarter by zip code area fixed effects and a cubic of the time of purchase (for completeness, Table 12 reports the parallel estimates for the Loss only sample). Column 1 reports naive OLS estimates using actual CLTV which reflect the non-monotonic pattern in Figure 9a.[52] Column 2 reports reduced form estimates using predicted CLTV, omitting the ranges below 30% and above 90% predicted CLTV which have few observations. A predicted CLTV ratio in excess of 70% reduces the probability of sale by 0.4 percentage points in this sample, a slightly larger effect than in the full sample.[53] Columns 3, 4 and 5 of the Table report separate naive and reduced form estimates, and of course IV estimates, using a third order polynomial of CLTV. These results, too, are qualitatively similar to the corresponding full sample results (Table 9), although they are substantially less precise.[54]

The results for the gain-only sample allow me to rule out that the full sample results are driven purely by loss aversion, especially considering that roughly 60% of the full sample also belongs to the gain-only sample. However, I cannot rule out that the effect of equity on mobility estimated in the full sample is *at least partially* driven by loss aversion. An observed correlation between low equity and reduced sale rates within a similarly defined *loss*-only sample can still reflect loss aversion, as can differences between gain-only and loss-only samples.[55]

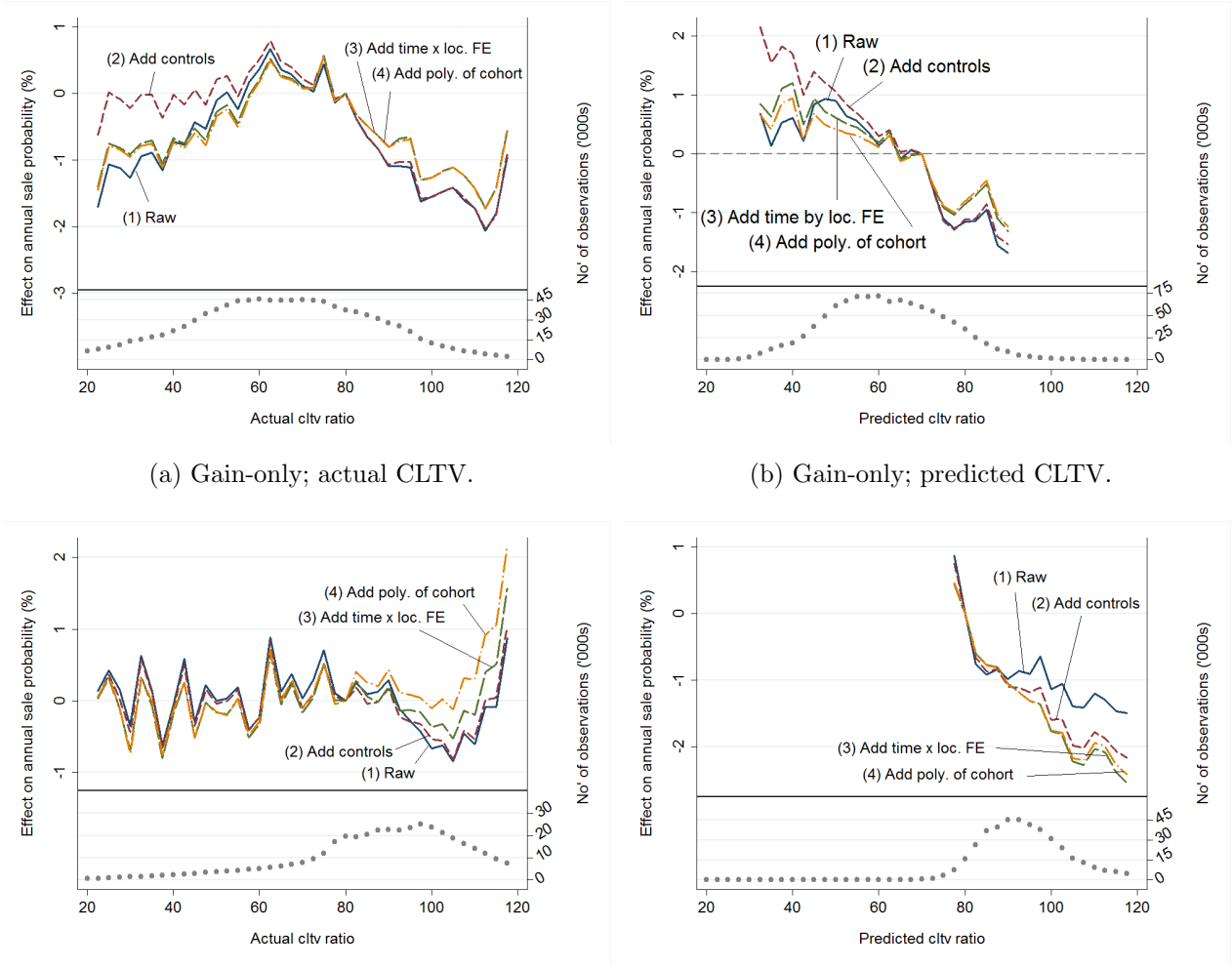
considered to experience only *net* loss and not *net* gain, even though he has experienced both gain and loss during his tenure.

[52]The upward slope at high negative equity levels that emerges in the full sample does not emerge as sharply in Figure 9a, and does not show up at all in Table 11. Estimates obtained from an opposite loss-only sample, as well as the results in Table 10, indicate that the increase in sales (short sales) at high negative equity levels is driven by homeowners experiencing net loss.

[53]The continued decline above 90% predicted CLTV that is observed in the full sample cannot be observed in the gain-only sample, as very few homeowners experiencing gains have such high predicted CLTV ratios (many have actual CLTV ratios above 90% due to their financial decisions, but not due to changes in price).

[54]The intuition behind this loss of precision is that restricting the sample only to homeowners with net gain substantially reduces variation in predicted CLTV, but does not affect the amount of variation in other determinants of actual equity, such as down payment amounts. Consequently predicted CLTV is not as good a predictor of actual CLTV in the gain-only sample as it is in the full sample, implying a weaker first stage and less precise IV estimates. This problem is even more severe in the loss-only sample, in which IV results are too imprecise to be informative.

[55]See footnote 31.



(a) Gain-only; actual CLTV.

(b) Gain-only; predicted CLTV.

(c) Loss-only; actual CLTV.

(d) Loss-only; predicted CLTV.

Figure 9: Gain-only and loss-only sample results. The gain-only results in panels *a* and *b* are remarkably similar to those in the full sample (Figures 7a and 7b), and they are not subject to the influence of loss aversion. Panel *d* indicates that in the loss-only sample sale rates increase (decrease) with predicted equity (CLTV), which is consistent with equity affecting sales through the liquidity constraint mechanism, but not through a wealth effect. The contrast between the loss-only result in panel *c* - in which sale rates do not decrease with actual CLTV (when fully conditioned in line (4)) - and those in panel *d*, stems from the fact that predicted and actual CLTV ratios are only weakly correlated within the loss-only sample (confining the sample only to properties with net loss substantially reduces variation in predicted equity without reducing variation in other determinants of actual equity, such as down payment amounts, with the result that predicted equity explains a much smaller share of variation in actual equity within this sample).

Table 11: Estimates for gain-only sample

Cltv ratio	cltv bins actual	cltv bins predicted	Effect of shifting cltv	3rd-order polynomial of cltv OLS	Reduced form	IV
	(1)	(2)		(3)	(4)	(5)
Over 40%	0.171	-0.124	From 30% to 40%	0.442***	0.264	2.630*
	(0.167)	(0.367)		(0.085)	(0.579)	(1.355)
Over 50%	0.441***	0.008	From 40% to 50%	0.297***	-0.030	0.806
	(0.152)	(0.181)		(0.056)	(0.373)	(0.772)
Over 60%	0.563***	0.004	From 50% to 60%	0.150***	-0.230	-0.486
	(0.127)	(0.160)		(0.050)	(0.237)	(0.529)
Over 70%	-0.131	-0.412***	From 60% to 70%	0.003	-0.335*	-1.247**
	(0.128)	(0.117)		(0.049)	(0.173)	(0.525)
Over 80%	-0.493***	0.000	From 70% to 80%	-0.145***	-0.346**	-1.476***
	(0.148)	(0.127)		(0.045)	(0.156)	(0.549)
Over 90%	-0.501***		From 80% to 90%	-0.293***	-0.263*	-1.174*
	(0.130)			(0.038)	(0.150)	(0.637)
Over 100%	-0.359**		From 90% to 100%	-0.443***	-0.085	-0.340
	(0.176)			(0.052)	(0.164)	(0.985)
Over 110%	-0.021		From 100% to 110%	-0.593***	0.188	1.025
	(0.248)			(0.096)	(0.246)	(1.643)
Owner, home and loan ctrls	+	+	Owner, home and loan ctrls	+	+	+
Time by Loc. FE	+	+	Time by Loc. FE	+	+	+
Poly. of Cohort	+	+	Poly. of Cohort	+	+	+
N observations	1,028,770		N observations	1,028,770		
N properties	90,530		N properties	90,530		
Avg. outcome level	3.03%		Avg. outcome level	3.03%		

Notes: the dependent variable is 100× annualized arm's length sale indicator. This Table reports results for the gain-only sample, which includes all properties in the full sample whose estimated value at the time of observation exceeds the purchase price. The results correspond to full-sample results as follows: columns 1 and 2 correspond to the fourth columns of Tables 7 and 8, respectively, and columns 3, 4 and 5 correspond to the columns of Table 9 - see notes in the corresponding full-sample Tables. With respect to column 5, Angrist-Pischke F-statistics (p-values) for the CLTV, $CLTV^2$ and $CLTV^3$ first stage regressions are 335.28 (0), 208.77 (0) and 132.59 (0), respectively, and the partial R^2 for each of the first-stage regressions is 0.0017, 0.0034 and 0.0045 respectively. These low partial R^2 values make the estimates substantially less precise than the full sample ones - see footnote 54 for the intuition. For the list and description of included home, owner and loan characteristics see Table 2 and the accompanying notes. Standard errors clustered in 72 zip code areas. One, two and three asterisks reflect statistical significance at the 10, 5 and 1 percent levels, respectively.

Table 12: Estimates for loss-only sample

Cltv ratio	cltv bins: actual	cltv bins: predicted	Effect of shifting cltv	3rd-order polynomial of cltv: OLS	Reduced form	IV
	(1)	(2)		(3)	(4)	(5)
Over 40%	0.291		From 30% to 40%	0.647***	-13.067***	55.943
	(0.378)			(0.192)	(4.895)	(43.366)
Over 50%	-0.031		From 40% to 50%	0.307**	-9.954***	32.480
	(0.324)			(0.128)	(3.566)	(27.303)
Over 60%	0.467		From 50% to 60%	0.076	-7.267***	13.861
	(0.288)			(0.084)	(2.449)	(14.429)
Over 70%	-0.000		From 60% to 70%	-0.047	-5.007***	0.085
	(0.241)			(0.061)	(1.545)	(5.187)
Over 80%	0.130	-2.141***	From 70% to 80%	-0.061	-3.174***	-8.846**
	(0.187)	(0.558)		(0.053)	(0.856)	(4.114)
Over 90%	-0.030	-0.166	From 80% to 90%	0.033	-1.767***	-12.935*
	(0.165)	(0.147)		(0.050)	(0.391)	(7.140)
Over 100%	-0.139	-0.412***	From 90% to 100%	0.235***	-0.788***	-12.179
	(0.142)	(0.147)		(0.055)	(0.178)	(7.691)
Over 110%	1.362***	0.116	From 100% to 110%	0.546***	-0.235*	-6.581
	(0.160)	(0.197)		(0.079)	(0.136)	(5.374)
Owner, home and loan ctrls	+	+	Owner, home and loan ctrls	+	+	+
Time by Loc. FE	+	+	Time by Loc. FE	+	+	+
Poly. of Cohort	+	+	Poly. of Cohort	+	+	+
N observations	422,926		N observations	422,926		
N properties	46,865		N properties	46,865		
Avg. outcome level	1.95%		Avg. outcome level	1.95%		

Notes: the dependent variable is 100× annualized arm's length sale indicator. This Table reports results for the gain-only sample, which includes all properties in the full sample whose estimated value at the time of observation exceeds the purchase price. The results correspond to full-sample results as follows: columns 1 and 2 correspond to the fourth columns of Tables 7 and 8, respectively, and columns 3, 4 and 5 correspond to the columns of Table 9 - see notes in the corresponding full-sample Tables. With respect to column 5, Angrist-Pischke F-statistics (p-values) for the CLTV, $CLTV^2$ and $CLTV^3$ first stage regressions are 8.00 (0.006), 11.21 (0.001) and 24.08 (0), respectively, and the partial R^2 for each of the first-stage regressions is 0.0248, 0.0318 and 0.0346 respectively. The low F-statistics imply that the estimates are not informative, but I report them for the sake of completeness - see footnote 54 for the intuition. For the list and description of included home, owner and loan characteristics see Table 2 and the accompanying notes. Standard errors clustered in 72 zip code areas. One, two and three asterisks reflect statistical significance at the 10, 5 and 1 percent levels, respectively.

6 Conclusion

In this study, I estimate the effect of owner-occupant equity on the probability of home sale and, indirectly, on mobility. Most importantly, I establish the causal nature of this empirical relationship. I do so primarily by exploiting plausibly exogenous variation in equity that stems only from changes in local housing price indices, and by comparing properties observed in the same time and location, thereby controlling for a wide variety of confounding time- and location-varying economic conditions. In addition, I bring the literature on the equity-mobility nexus up to date, using a novel combination of proprietary and administrative data to provide estimates from the current housing crisis.

I find that sale probabilities decline well before homeowners reach negative equity levels, suggesting that the share of homeowners who have negative equity substantially understates the share of homeowners whose mobility is impaired by insufficient equity. More importantly, my findings indicate that policy attempting to influence the housing market or the economy at large by manipulating homeowners' equity - e.g. by some means of reducing mortgage principal - is likely to be most effective in inducing mobility if it targets homeowners whose equity ranges from 70% to 100% CLTV. Admittedly, targeting this group of homeowners rather than those who are deeper "underwater" may not be an attractive policy, but recognizing the range in which equity shifts the probability of sale and mobility can nevertheless help shape more effective and informed policy. Increasing mortgaged homeowners' equity across the board by generating mild inflation, for example, is likely to induce mobility among homeowners in the 70% to 100% CLTV range, whereas principal reductions for deeply "underwater" homeowners may fail to induce any mobility whatsoever.

Abstracting from policy and the present context, the theoretical link put forth in Stein (1995) whereby owner-occupant mobility hinges upon sufficient equity and its relation to the positive price-volume correlation in housing markets form a basic tenet of our understanding of housing markets and their cycles. This study underpins our understanding by providing evidence that the corresponding *empirical* relationship between equity and mobility is causal. By shedding light on a causal effect of price on trading volume, this study contributes to the literature disentangling the multiple causal relationships driving the positive price-volume correlation in housing markets. Establishing both theoretical and causal empirical relationships running in the opposite direction, from volume to prices, could provide a foundation for a more complete, dynamic understanding of housing markets and their cycles, and may be a fruitful avenue for future research.

References

Aaronson, D. and J. Davis (2011). How much has house lock affected labor mobility and the unemployment rate? *Chicago Fed Letter* (Sep).

Akerlof, G. A., A. K. Rose, J. L. Yellen, L. Ball, and R. E. Hall (1988). Job Switching and Job Satisfaction in the U.S. Labor Market. *Brookings Papers on Economic Activity 1988*(2), 495–594.

Andrew, M. and G. Meen (2003). House Price Appreciation, Transactions and Structural Change in the British Housing Market: A Macroeconomic Perspective. *Real Estate Economics 31*(1), 99–116.

Anenberg, E. (2011). Loss aversion, equity constraints and seller behavior in the real estate market. *Regional Science and Urban Economics 41*(1), 67 – 76.

Anenberg, E. (2012). Information Frictions and Housing Market Dynamics. *Finance and Economics Discussion Series 2012-48. Board of Governors of the Federal Reserve System (U.S.).*

Bayer, P., R. McMillan, A. Murphy, and C. Timmins (2011, July). A Dynamic Model of Demand for Houses and Neighborhoods. Working Paper 17250, National Bureau of Economic Research.

Benito, A. (2006). The down-payment constraint and UK housing market: Does the theory fit the facts? *Journal of Housing Economics 15*(1), 1 – 20.

Berkovec, J. A. and J. L. Goodman (1996). Turnover as a Measure of Demand for Existing Homes. *Real Estate Economics 24*(4), 421–440.

Blanchard, O. J. and L. Katz (1992). Regional Evolutions. *Brookings Papers on Economic Activity 1992*(1), pp. 1–75.

Cameron, C. and P. Trivedi (2005). *Microeconometrics: Methods and Applications.* Cambridge University Press.

Caplin, A., C. Freeman, and J. Tracy (1997, November). Collateral Damage: Refinancing Constraints and Regional Recessions. *Journal of Money, Credit and Banking 29*(4), 496–516.

Case, B., H. O. Pollakowski, and S. M. Wachter (1997). Frequency of Transaction and House Price Modeling. *The Journal of Real Estate Finance and Economics 14*(1-2), 173–87.

Case, K. E., R. J. Shiller, and A. Thompson (2012, September). What Have They Been Thinking? Home Buyer Behavior in Hot and Cold Markets. Working Paper 18400, National Bureau of Economic Research.

Chan, S. (May 2001). Spatial Lock-in: Do Falling House Prices Constrain Residential Mobility? *Journal of Urban Economics 49*, 567–586.

Clayton, J., N. Miller, and L. Peng (2010). Price-volume Correlation in the Housing Market: Causality and Co-movements. *The Journal of Real Estate Finance and Economics 40*, 14–40.

Coulson, N. E. and P. L. Grieco (2013). Mobility and mortgages: Evidence from the PSID. *Regional Science and Urban Economics 43*(1), 1 – 7.

Donovan, C. and C. Schnure (2011). Locked in the House: Do Underwater Mortgages Reduce Labor Market Mobility? Unpublished manuscript.

Engelhardt, G. V. (2003). Nominal loss aversion, housing equity constraints, and household mobility: evidence from the United States. *Journal of Urban Economics 53*(1), 171 – 195.

Estevao, M. and E. Tsounta (2011). Has the Great Recession raised US structural unemployment? *IMF Working Papers*, 1–46.

Farnham, M., L. Schmidt, and P. Sevak (2011, May). House Prices and Marital Stability. *American Economic Review 101*(3), 615–19.

Ferreira, F., J. Gyourko, and J. Tracy (2010). Housing busts and household mobility. *Journal of Urban Economics 68*(1), 34 – 45.

Ferreira, F., J. Gyourko, and J. Tracy (2011, September). Housing Busts and Household Mobility: An Update. Working Paper 17405, National Bureau of Economic Research.

Ferreira, F., J. Gyourko, and J. Tracy (2012). Housing busts and household mobility: an update. *Economic Policy Review* (Nov), 1–15.

Geanakoplos, J. (2010). The Leverage Cycle. *NBER Macroeconomics Annual 24*(1), pp. 1–66.

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

Grossman, S. and G. Laroque (1990). Asset Pricing and Optimal Portfolio Choice in the Presence of Illiquid Durable Consumption Goods. *Econometrica 58*, 25–51.

Henley, A. (1998). Residential Mobility, Housing Equity and the Labour Market. *The Economic Journal 108*(447), pp. 414–427.

Hort, K. (2000). Prices and turnover in the market for owner-occupied homes. *Regional Science and Urban Economics 30*(1), 99 – 119.

Karahan, F. and S. Rhee (2011). Geographical Reallocation and Unemployment during the Great Recession: The Role of the Housing Bust.

Kothari, S., I. Saporta-Eksten, and E. Yu (2012). The (Un)importance of Mobility in the Great Recession. Stanford University, unpublished manuscript.

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

Modestino, A. S. and J. Dennett (2012). Are American homeowners locked into their houses? The impact of housing market conditions on state-to-state migration. *Regional Science and Urban Economics*.

Nenov, P. (2012). Labor Market and Regional Reallocation Effects of Housing Busts. Unpublished manuscript.

Ortalo-Magne, F. and S. Rady (2004). Housing transactions and macroeconomic fluctuations: a case study of England and Wales. *Journal of Housing Economics 13*(4), 287 – 303.

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

Sahin, A., J. Song, G. Topa, and G. L. Violante (2012, August). Mismatch Unemployment. Working Paper 18265, National Bureau of Economic Research.

Schulhofer-Wohl, S. (2011, January). Negative Equity Does Not Reduce Homeowners' Mobility. Working Paper 16701, National Bureau of Economic Research.

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

Sterk, V. (2010, October). Home Equity, Mobility, and Macroeconomic Fluctuations. DNB Working Papers 265, Netherlands Central Bank, Research Department.

Stroebel, J. (2012). The impact of asymmetric information about collateral values in mortgage lending. Unpublished manuscript.

Valletta, R. G. (2013). House Lock and Structural Unemployment. *Labour Economics*.

A Data Appendix

A.1 Measuring equity

By convention, home equity is most often measured inversely, in terms of a property's combined loan-to-value ratio (CLTV), and this is the measure of equity used in (5). The CLTV ratio of property i at time t is the sum of principal owed at that time on all loans secured by the property, divided by the property's contemporary value.[56] Formally,

$$CLTV_{it} \equiv \frac{\sum_{j \in J(i,t)} loan_{jt}}{value_{it}}, \tag{12}$$

where $J(i,t)$ is the set of loans for which property i serves as collateral and $loan_{jt}$ is the amount of principal owed on loan j.[57] The CLTV ratio is inversely related to the property owner's share of equity, which is simply

$$\%\ equity_{it} \equiv \frac{value_{it} - \sum_{j \in J(i,t)} loan_{jt}}{value_{it}} \equiv 1 - CLTV_{it}. \tag{13}$$

[56] Each loan secured against a property has its own loan-to-value (ltv) ratio. The word "combined" simply reflects that in instances in which there is more than one loan, the ratio involves their sum.

[57] The primary loan secured against a property is typically a mortgage obtained when the property is purchased, but a sizeable share of mortgaged properties - 40% in my sample - are observed serving as collateral for further borrowing, either since the time of purchase ("piggy back" loans) or from later on. Additional borrowing comes in many shapes and forms, including secondary mortgages, refinancing (often involving cash-in or cash-out), home equity loans or lines of credit, and various other methods means of borrowing against one's home. A property's combined loan-to-value (CLTV) ratio is simple the sum of loan-to-value (ltv) ratios for any individual loans secured against the property.

The CLTV ratio used in the paper is an estimate. First, because property values can never be observed directly, $value_{it}$ is taken to be CoreLogic's corresponding Automated Valuation Model (AVM) estimate, described in the main text.[58] Second, only the initial balance of the loans secured against a property is observed, not their outstanding balance once payments have been made towards principal, as CoreLogic does not observe these payments. I estimate payments towards principal in a very crude way, assuming standard (fixed payment) amortization over the observed term of the loan and given the going interest rate associated with the loan by CoreLogic.[59]

A.2 The Zillow housing price index

The Zillow housing price index is hedonic in the sense that it relies on estimating home values based on their observable characteristics, rather than relying on repeat sales to account for all fixed observable and unobservable home attributes. However, this index is not simply a normalized sequence of time fixed effects estimated in a hedonic regression. In order to minimize the selection bias inherent in hedonic regressions because the set of properties selling in a given period is non-random, Zillow hedonically estimates the value of each and every home in every period. The Zillow housing index value for a set of homes in a certain period is then taken to be the median home value for that set in that period.[60]

A.3 Inferring owner-occupancy status

I infer that a property was owner-occupied in a given quarter if at least one contemporary occupant shared a last name with a contemporary owner. Thus, I broadly interpret owner-occupancy to include relatives of owners (as well as occupants who by coincidence share a last name with an owner). I observe the occupants of a property by matching each property-by-quarter observation with contemporary voter registry records by address.[61] Registering to vote is thus used as a proxy for occupancy. Occupants who do not register to vote, or

[58] Even at times when a property is sold, one could deliberate whether the sale price reflects the property's value in a way that generalizes beyond the buyer and seller's agreed upon price.

[59] For a loan with principal P, an N period term, an interest rate of r per period and a fixed per period payment c, the balance after n periods is $(1+r)^n P \cdot [1 - \frac{1-(1+r)^{-n}}{1-(1+r)^{-N}}]$.

[60] Additional information on Zillow's indexing methodology is available online at the time of writing, at: http://www.zillowblog.com/research/2012/01/21/zillow-home-value-index-methodology.

[61] Roughly 82% of voter registry records are successfully matched with properties; the unmatched voter registry records primarily reflect condos with idiosyncratically recorded numbering, and apartments for which the county assessor record encompasses the entire multi-unit property.

prefer to use another permanent address, e.g. of a parent's home, remain unobserved. Within property-by-ownership spells, I take a property's owner-occupancy status to be that of the last period of the ownership spell. This way, occupants who take their time about updating their voter registry after moving are still captured as occupants, as long as they update their voters registry either before the last period of their ownership spell or before the last period observed (2011Q4). This method of inferring owner-occupancy is asymmetric, in the sense that properties *not* inferred to be owner-occupied may in fact still be owner-occupied, e.g. if their owner-occupants failed to update their voter registration records.

B Dynamic selection

Dynamic selection refers to the process whereby the composition of the pool of homeowners in a cohort systematically changes over time as homeowners sell their homes and exit the cohort selectively. A simple way of thinking about the problem is to suppose that homeowners belong to one of two types: owners of starter homes who have a high propensity to move and owners of permanent homes who have a low propensity to move. As a purchase cohort's tenure duration grows longer its relative share of starter-home owners dwindles. This "weeding out" process is of no concern inasmuch as it is identical across cohorts, because then it gets soaked up by the flexible tenure duration control included in the regressions. However if the "weeding out" process is differential across cohorts, i.e. if starter-home owners are "weeded out" at a quicker pace in some cohorts than in others because of conditions that facilitate or hinder mobility, then the flexible tenure duration control is inadequate and captures only the average effect of dynamic selection across all cohorts.[62]

Dynamic selection is a potential concern if increasing home values raise equity and in so doing facilitate mobility, because then cohorts that experience greater increases in home values during their tenure are likely to experience a faster pace of dynamic selection. Homeowners remaining in such cohorts long enough to have low CLTV ratios are likely to have a relatively low propensity to move as well, biasing estimates of the effect of equity (CLTV) on mobility downward (upward).

To gauge the extent of dynamic selection I compare the rates of home sale by CLTV ratio for three sets of homeowners. The first set is the full sample used in this study, which consists of properties whose owners have tenure duration of 1 to 10 years. The second and

[62]Dynamic selection is differential across cohorts if, for example, factors influencing mobility interact multiplicatively as they do in a mixed proportional hazard model.

third are subsets whose tenure duration is within 1 to 5 years and 1 to 3 years, respectively. If differences in sale rates across equity levels are driven by dynamic selection then sale rates should increase (decrease) with equity the most (least) for the 1 to 3 year group, and least (most) for the 1 to 10 year group.

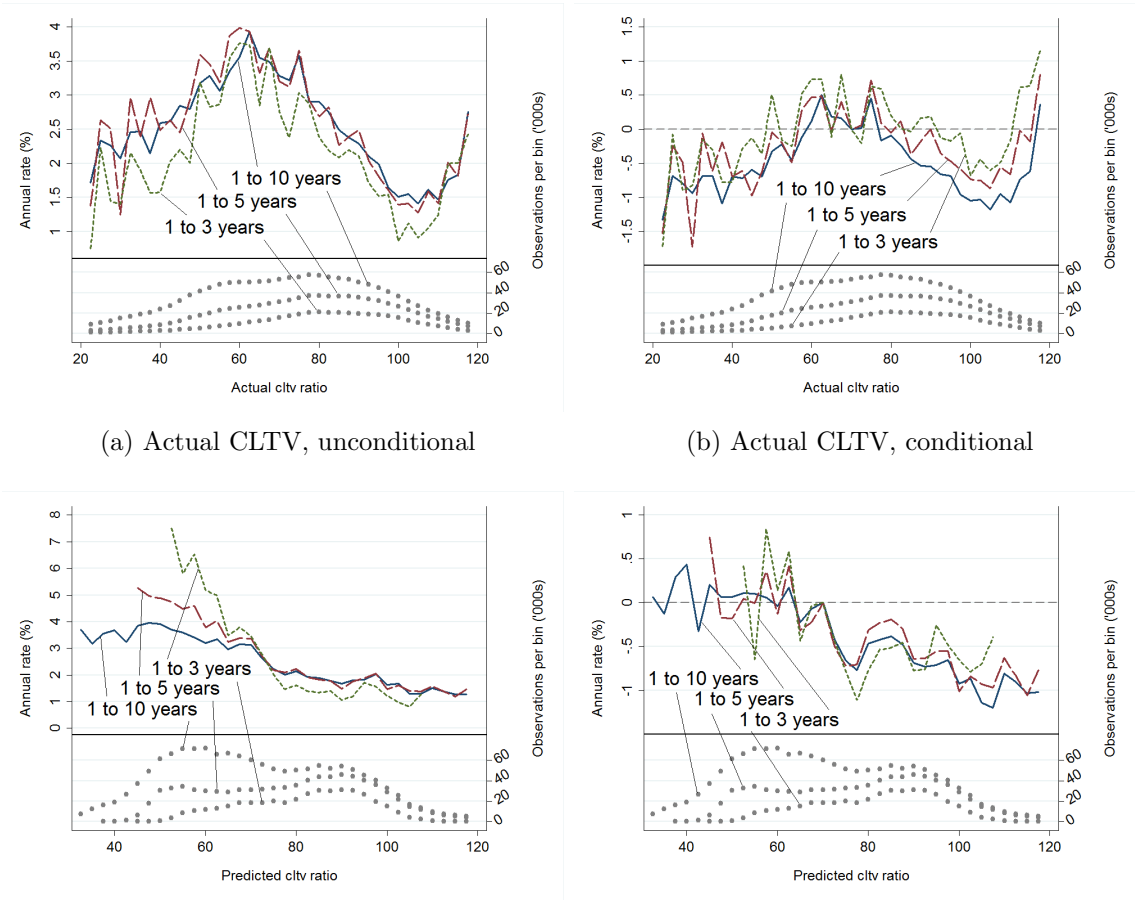
Figures B.1a shows the unconditional home sale rates for these three groups with respect to actual CLTV. The three groups do not differ substantially, suggesting that dynamic selection does not play an important role in shaping the relationship between actual equity and sale rates. Figure B.1b shows the corresponding home sale rates conditional on the full set of owner, home and loan characteristics - and in particular a cubic of tenure duration - as well as quarter by zip code area fixed effects and a cubic of purchase cohort, and does not suggest an important role for dynamic selection either. Figure B.1c, on the other hand, shows the unconditional home sale rates for these three groups by *predicted* CLTV. Below 80% predicted CLTV the sale rate increases the most with equity for the 1 to 3 year group, then for the 1 to 5 year group and least of all for the 1 to 10 year group, which suggests that dynamic selection is taking place. However, once the home sale rates are fully conditioned in Figure B.1d (as in Figure B.1b) the effect of equity on sale rates appears to be similar for the different tenure duration groups, suggesting that the controls adequately account for dynamic selection. Overall, these results alleviates concerns with respect to dynamic selection.[63]

C The "better homes" hypothesis

This appendix addresses the "better homes" hypothesis outlined in section 5.3. The hypothesis is that relative to observably similar homes, certain "better homes" tend to sell less frequently and to incur smaller losses in value when aggregate housing prices fall. If both traits coincide this generates a positive correlation between homes sales and CLTV ratios - which rise in proportion to loss in home value - causing naive OLS regressions of home sales on CLTV ratios to be biased upwards.

The observed empirical relationship between sale probability and actual CLTV is shown in figure 6a and is stylized in figure 8c. The argument put forth in section 5.3.1 is that this

[63]Moreover, note that the direction in which dynamic selection potentially biases results works *against* finding a positive effect of equity on home sales. Therefore if one remains concerned about dynamic selection despite the results shown here, the implication is that estimates of the effect of equity on mobility reported in this paper are in fact lower bounds.



(a) Actual CLTV, unconditional

(b) Actual CLTV, conditional

(c) Predicted CLTV, unconditional

(d) Predicted CLTV, conditional

Figure B.1: The effect of equity on sales by actual and predicted CLTV and by tenure duration group. Panels *a* and *b* (*c* and *d*) report unconditional sale rates and conditional effects on sale rates by actual (predicted) CLTV ratio for the full sample, which contains homeowners with tenure duration of 1 to 10 years, and for subsets with tenure duration of 1 to 5 years and of 1 to 3 years. Conditioning refers to controlling for the full set of owner, home and loan characteristics (including a cubic of tenure duration) as well as quarter by zip code area fixed effects and a cubic of purchase cohort - see Table 2 for a detailed account of the included controls. If dynamic selection is taking place then an increase (decrease) in equity (CLTV) should correspond to a greater increase in sale rates for groups with lower tenure duration. Dynamic selection does not appear to be taking place along the dimension of actual CLTV. Dynamic selection does appear to take place along the predicted CLTV dimension, but it is adequately accounted for by conditioning the estimates on the above set of controls.

pattern emerges from the superposition of a causal effect of CLTV on the probability of sale (from 70% to 100% CLTV) on one hand and confounding factors that bias the OLS estimate of the effect upward over the entire CLTV range on the other. These confounding factors include the "better homes" hypothesis. To illustrate how the observed pattern emerges I generate data and run a simulation in which:

- Aggregate housing prices fall.
- A higher CLTV ratio reduces the probability of sale over the 70% to 100% CLTV range for all homes.
- All homes are valued equally at the onset, but half of the homes are "better" and therefore, compared to the remaining "worse" homes, they sell less frequently and incur smaller losses in value when aggregate housing prices fall.
- The owner of "better" homes are assumed to have made somewhat larger initial down payments.[64]

Figure C.1 visualizes the results of the simulation.[65] The solid lines in the upper panel reflect the annual sale probability of "better" and "worse" homes, respectively, and the dashed line reflects weighted average given the density of "better" and "worse" homes at each level of CLTV (shown in the bottom panel). An increase in the CLTV ratio over the 70% to 100% range decreases the sale probability for both types of homes and as per the "better homes" hypothesis the sale rate is lower for "better" homes across the board.

However, the decreasing share of "better" homes at higher CLTV levels generates a composition effect whereby higher CLTV levels appear to raise the annual probability of

[64] As shown shortly, this assumption is supported by empirical evidence. This assumption is not strictly necessary, but it helps the simulation visually mimic the relationship between CLTV and annual sale probability observed in the raw data, as seen in Figure 6a (particularly at low levels of CLTV).

[65] In more detail: The data consist of 20 quarterly observations on 1000 simulated properties, half of which are "better" and half of which are "worse". All homes are valued equally when the simulation begins, but the value of "better" and "worse" homes evolves as 0.9 and 1.1 times the change in the housing pricing index (HPI), respectively, and the HPI is falls by 2% each quarter. Thus, "better" homes lose less value as aggregate housing prices fall. "Better" homes have an annual sale probability of 2% whereas "worse" homes have an annual sale probability of 4%, i.e. they sell more frequently. All homes experience a linear reduction in annual sale probability of 1.6% as they transition from 70% to 100% CLTV. In addition, the owners of "better" homes are assumed to have made down payment that is 5 percentage points larger on average. Specifically, down payments are assumed to be distributed lognormal (μ,0.25), with $\mu = 42.5\%$ for "better" homes and 37.5% for "worse" (this implies a mass of down payments centered at 40% with a left tail that extends to 0% and a longer right tail. The lognormal distribution captures the asymmetry of the distribution of down payments, and generates only a negligible density of down payments to the right of 100%.)

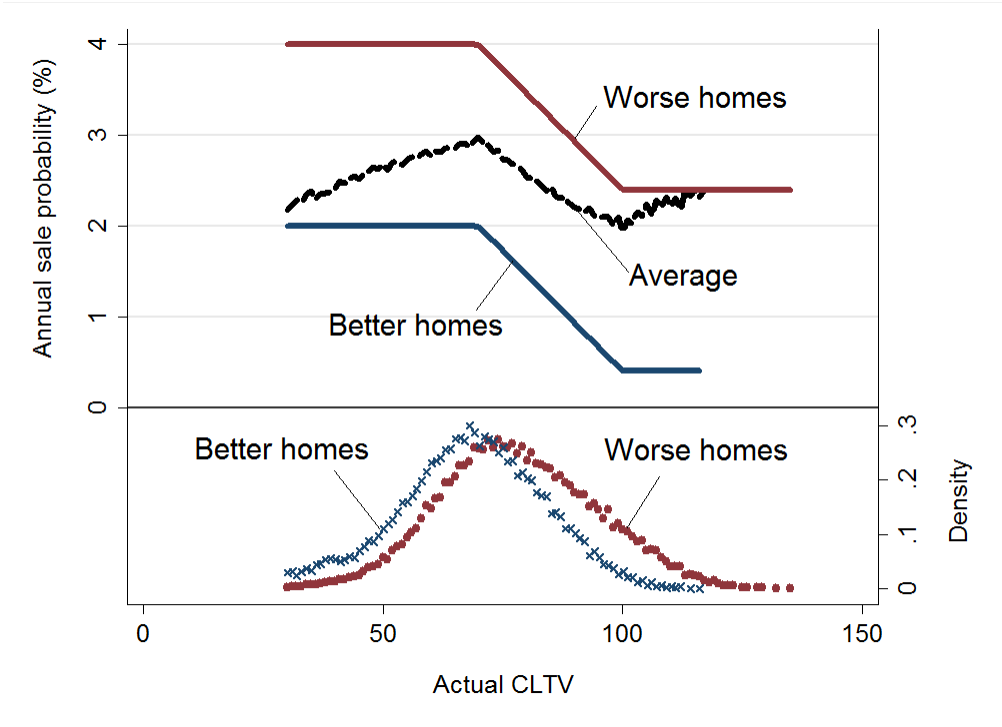


Figure C.1: Data generated under the assumptions of the "better homes" hypothesis mimic the relationship between the CLTV ratio and the annual probability of sale observed in the raw data, as in Figure 6a. The sale probability of all properties is reduced causally and similarly when CLTV increases from 70% to 100%, but "better" homes are less likely to sell across the board. Because "better" homes lose less value than "worse" homes when aggregate housing prices fall (and because their owners tend to make larger down payments), these homes tend to have lower CLTV ratios. The weighted average of homes' annual sale probabilities at each level of CLTV reflects the *composition* of homes at that level, reflected by an upward slope in the ranges below 70% and above 100% CLTV, and in a downward slope in the 70% to 100% CLTV range that is *moderate* than the underlying causal effect of CLTV. The moderate slope of the average annual sale probability with respect to CLTV in the 70% to 100% CLTV range qualitatively corresponds to the OLS estimates in column 1 of Table 9, whereas the steeper slopes of uniquely "better" or "worse" homes over this range correspond to the IV estimates in column 3 of Table 9.

sale. This composition effect does *not* reflect causality. Outside of the 70% to 100% CLTV range the composition effect is immediately evident in the positive slope of the dashed line, whereas within this range it merely flattens the negative slope of the dashed line compared to the causal effect captured by the solid lines. The dashed line essentially mimics the relationship between CLTV and the annual sale probability seen in the raw data, as in

Figure 6a. Further, the moderate slope of the dashed line qualitatively corresponds to the naive OLS estimates of the effect of CLTV on the probability of sale in the 70% to 100$ CLTV range, as they appear in column 1 of Table 9, whereas the steeper slopes of the solid lines correspond to the IV estimates over this range that reflect the *causal* effect of CLTV, as they appear in column 3 of Table 9.

But are the assumptions underpinning the "better homes" hypothesis and the simulation true to reality? To answer this question I test among observably similar homes whether those that tend to sell less frequently also tend to incur smaller losses in value when aggregate housing prices fall (and to have had larger down payments). I quantify properties' tendency to sell more or less frequently by observing their average annual sale rate in the 10 years preceding my sample, from Jan 1st 1997 through December 31st 2006 (omitting properties built after Jan 1st 1997). I then regress these properties' year-on-year changes in estimated value when housing prices fell during the recent housing crisis on their prior sale rate, while controlling for observables and conducting the estimate within sets of properties experiencing identical aggregate price changes. Specifically, I estimate the regression

$$\Delta P_{i,t,t-4} = \beta SR_i + \mathbf{X}_{it}\delta + \psi_{lct} + \epsilon_{it},$$

where $\Delta P_{i,t,t-4}$ is the percentage year-on-year change in the estimated value of property i between time (quarter) $t-4$ and t; the sale rate, SR_i, is the average annual number of sales recorded for property i, built no later than 1996, from Jan 1st 1997 through December 31st 2006; $\mathbf{X}_{it}$ is the full vector of control variables detailed in Table 2; ψ_{lct} is a saturated set of zipcode by quarter by time of purchase fixed effects and ϵ_{it} is an error term. The fixed effects ψ_{lct} ensure that the estimate is conducted within sets of properties with identical aggregate house price histories since their last purchase. All properties within such a cell experience - by construction - the same change in the aggregate local housing price index, so identifying variation in housing price changes stems from *changes in individual home values, conditional on aggregate changes in home values.* These fixed effects also eliminate any potential confounding effects generated by any other conditions that vary over time, location or cohort of purchase. The sample is limited to properties observed in quarter by zipcode cells experiencing a year-on-year decrease in the zipcode area housing price index. The sample is also limited to properties observed in 2008-2011, so that sales determining a property's sale rate do not coincide with the period of the latest year-on-year change in value. Finally, the sample is confined to properties built no later than 1996 in order to avoid a "mechanical" bias created by newly constructed properties. Such properties have a

Table C.1: The Correlation Between Properties' Prior Sale Frequency and the Sensitivity of Their Value to Decreases in the Local Housing Price Index

	Year-on-year change in home value (%)
Average annual sale rate (1997-2006)	-0.0060***
	(0.0009)
Owner, home and loan characteristics	+
Zip × qtr. of obs. × qtr. of purchase FE	+
N	757,639

Notes: Sample includes only owner-occupied homes, whose occupants have tenure duration of 1 to 10 years and do not concurrently own more than two properties. The sample is also limited to properties observed in 2008-2011, in zipcode by quarter cells experiencing year-on-year decreases in the zipcode area HPI, and that were built no later than 1996. For the list and description of included home, owner and loan characteristics see Table 2 and the accompanying notes. Standard errors clustered in 72 zip code areas. One, two and three asterisks reflect statistical significance at the 10, 5 and 1 percent levels, respectively.

"mechanically" higher than average sale rate in their first several years of existence because their history necessarily begins with a sale (e.g. a property built one month ago has a misleading average of 12 sales per year), and all else equal they are likely to have higher value than older properties.

The estimate reported in Table C.1 indicates that, on average, the value of homes which sold more frequently during 1997-2006 fell more sharply than observably similar homes which experienced the same aggregate changes in housing prices, validating the key assumption behind the "better homes" hypothesis. To assess whether the owners of "better" homes tend to make somewhat larger down payments I estimate a similar regression in which I replace $\Delta P_{i,t,t-4}$ with the initial CLTV ratio from the most recent purchase. The estimate reported in Table C.2 indicates that on average the owners of homes which sold more frequently during 1997-2006 had higher initial CLTV ratios, implying that they did in fact tend to make smaller down payments.

Table C.2: The Correlation Between Properties' Prior Sale Frequency and Down Payment

	Down Payment (% of purchase price)
Average annual sale rate (1997-2006)	3.136***
	(0.805)
Owner, home and loan characteristics	+
Zip × qtr. of obs. × qtr. of purchase FE	+
N	757,639

Notes: Sample includes only owner-occupied homes, whose occupants have tenure duration of 1 to 10 years and do not concurrently own more than two properties. The sample is also limited to properties observed in 2008-2011, in zipcode by quarter cells experiencing year-on-year decreases in the zipcode area HPI, and that were built no later than 1996. For the list and description of included home, owner and loan characteristics see Table 2 and the accompanying notes. Standard errors clustered in 72 zip code areas. One, two and three asterisks reflect statistical significance at the 10, 5 and 1 percent levels, respectively.