Mortgage Default with Positive Equity*

David Low†

Consumer Financial Protection Bureau

January 4, 2018

Most recent version here

Abstract

Frictionless models of mortgage default predict that defaulters are underwater. Available evidence suggests otherwise, but there are well-known concerns with this evidence. This paper provides the first formal estimates of the home equity of defaulters, using rich microdata and a robust Bayesian estimation procedure. It finds that 27-47% of foreclosed homeowners had positive equity from 2011-2013, which implies roughly 81-87% had positive equity in more "normal" times. Motivated by this evidence, the paper then develops a quantitative lifecycle model of mortgage default with search frictions. In the model, homeowners who miss a mortgage payment may make it up the next period. As a result, abovewater homeowners sometimes choose to miss a payment rather than sell after an income shock. If their income recovers, they make up their payment and keep their home. If it does not, they may sell their home in a frictional market. Otherwise, they lose their home to foreclosure. The estimated model matches key untargeted moments, including the foreclosure rate and the proportion of defaulters with positive equity. In the model, a policy called "lender recourse" – which allows lenders to seize the assets of underwater defaulters – is generally ineffective at reducing default rates, consistent with the evidence.

*This paper is a revised version of the first chapter of my Ph.D. dissertation. I thank Raquel Fernández, Christopher Flinn, Gianluca Violante, Kevin Thom, and Mervyn King for their comments, support, and guidance. I thank Andreas Fuster, Benedict Guttman-Kenney, and Andrew Haughwout for data on the distribution of household leverage over time. All errors are my own. The views expressed are those of the author and do not necessarily reflect those of the Consumer Financial Protection Bureau.

†Email address: david.low@cfpb.gov

1 Introduction

Mortgage debt is by far the largest class of consumer debt in the United States, and mortgage default has profound implications for consumer welfare and the broader economy. The causes of mortgage default are not fully understood, but the home equity of defaulters provides critical clues. In the frictionless models that dominate the literature, homeowners with positive equity will always sell their homes rather than default, and so all defaulters are underwater. Available evidence suggests otherwise, but there are many well-known concerns with this evidence. It is therefore unclear how many defaulters have equity, whether a structural model with search frictions can match the number of defaulters with equity, or what the policy implications of such a model would be. This paper aims to fill this gap.

This paper provides the first formal estimates of the home equity of defaulters, using a robust Bayesian estimation procedure developed by Korteweg and Sorensen (2016). I estimate that 27% - 47% of foreclosed homeowners had positive equity between 2011-2013, which implies that in more "normal" times roughly 81% to 87% did. Motivated by this evidence, I develop a lifecycle model of consumption, housing, and mortgage decisions with rational agents and search frictions. 83% of foreclosures in the estimated model have positive equity, demonstrating that abovewater default rates in the data are consistent with economic theory. Model predictions also match the empirical evidence that a policy intended to discourage default called "lender recourse" is generally ineffective because it targets underwater homeowners.

Estimating home equity is challenging because the value of a home is typically observed with significant error. There are three commonly-used kinds of data on home values, all with strengths but also weaknesses. First, there is transaction-level data. Transactions provide relatively accurate measurements of home values, but they are endogenous and raise selection issues. They are also rare, leaving large gaps of time in these data where property values are unknown. Between transactions, the change in a home's value can be approximated by the change in a local house price index, but this neglects idiosyncratic shocks. Second, there is survey data on home values. Surveys are typically more frequent than transactions and are subject to fewer selection issues. However, home values reported in surveys are typically biased and inaccurate (e.g. Kiel and Zabel (1999)). A third approach is to use the observable characteristics of a property to estimate its value. These "hedonic" models of property values are also generally inaccurate, but they are attractive for valuing properties in foreclosure since such properties may depreciate more quickly than others.

The basic empirical strategy of this paper is to combine the strengths of all three data sources to formally estimate the home equity of foreclosed homeowners using an extension of the Bayesian Gibbs sampling procedure developed by Korteweg and Sorensen (2016). This extension explicitly controls for changes in a property's observable characteristics, including several measures of property depreciation. It filters out the unobserved time-varying portion of a property's value by exploiting the information embedded in transaction prices and homeowner-reported values, while accounting for the bias and noise in these signals. The procedure yields a full posterior distribution for the Loan-to-Value ratio (LTV) of every property in the data. This distribution characterizes not only the estimated LTV of a given property, but also the uncertainty in that estimate arising from measurement error, changes in property value that occur between measurements, and uncertainty in the parameter estimates of the empirical model.

I estimate that between 2011 and 2013, 47% percent of foreclosed homeowners had positive equity. For 27% percent of these foreclosures, negative equity is outside the one-sided 95% credible interval. However, homeowner equity was considerably lower during the foreclosure crisis than before or since (Fuster et al. (2016)). The estimated foreclosure hazard rates from 2011-2013 suggest that from 1998-2001 roughly 87% of foreclosed homeowners had positive equity. This number falls to 81% if the 95th percentiles of the posterior LTV distributions, rather than the means, are used as the LTV estimates.

Motivated by this evidence, the paper then develops a model of mortgage default with rational agents and search frictions. In the model, households make housing, consumption, and mortgage decisions over the lifecycle and are subject to income shocks. Renters allocate expenditures between consumption, rent, and liquid assets. Nonhomeowners who wish to buy a home must search for one in a frictional matching market. They then decide whether to buy the property, and if so what size mortgage to get. Mortgages are priced endogenously by competitive, risk-neutral banks, but are subject to exogenous LTV and Payment-to-Income ("PTI") limits (Corbae and Quintin (2015), Greenwald (2017)). Homeowners who are current on their mortgage choose levels of consumption and savings, and also make a discrete choice. They can sell their home in a frictionless market.[1] If they do not sell, they can refinance their mortgage, they can pay it on schedule, or they can skip their mortgage payment and become delinquent. Delinquent homeowners lose their home to foreclosure at the beginning of the next period only if they do not make up their missed payment and they do not sell their home in a frictional market.[2]

The model's focus on understanding the equity of defaulters is unique, but it is broadly similar to many quantitative models of default (e.g. Jeske et al. (2013), Campbell and Cocco (2014), Chatterjee and Eyingungor (2014), Corbae and Quintin (2015), and Laufer (forthcoming)). As in other models, a homeowner may default on a mortgage with positive equity, if her equity is

[1]Unlike delinquent homeowners, current homeowners are not subject to search frictions. I make this assumption for two reasons. First, current homeowners have more time to find a buyer, since they have the entire period to sell but delinquent homeowners must find a buyer before foreclosure occurs. Second, this assumption makes all foreclosures in the model fully endogenous in the sense that every foreclosed homeowner could have sold their home but chose instead to stay in the home and become delinquent on the mortgage.

[2]Search frictions are modeled for delinquent mortgagors because of they limited time they have to sell before foreclosure occurs. In this market, delinquent mortgagors are allowed to decrease the price of their home to increase the probability of sale.

not sufficient to cover the transaction costs associated with selling a home. Such a homeowner, while technically abovewater, is said to be "effectively" underwater. More importantly, this paper follows a small but growing literature in macroeconomics (Hedlund (2016a), Hedlund (2016b), Head et al. (2016), and Garriga and Hedlund (2017)) by including search frictions in models of mortgage default.[3] As in these papers, distressed homeowners may default even with positive effective equity if they are unable to find a buyer for their home. This paper builds on this literature by modeling foreclosure as a process that takes more than one period (Herkenhoff and Ohanian (2015)). This allows homeowners to respond to the housing market frictions by selling their home earlier in the foreclosure process.

The estimated model matches key untargeted moments. 83% of defaulters in the model have positive equity, which compares well to the estimates of 81% to 87% discussed above. The aggregate foreclosure rate in the model is .45% , which is quite close to the "long-run" foreclosure rate Jeske et al. (2013) target of .5%. Thus this paper shows that abovewater mortgage default rates seen in the data are compatible with rational agents in a model with search frictions.

The model generates valuable insights on lender recourse. This policy, which is intended to discourage default, allows lenders to seize the non-housing assets of defaulters to cover the difference between the outstanding mortgage balance and the value of the home.[4] This difference is positive for underwater defaulters, so under recourse they may lose assets if they default. Several structural models find that recourse is effective at discouraging default, e.g. Quintin (2012), Campbell and Cocco (2014), Hatchondo et al. (2014), Li et al. (2014), and Corbae and Quintin (2015). However, empirical evidence generally shows that recourse has no effect on default rates (e.g. Clauretie (1987), Ghent and Kudlyak (2011), and Li and Oswald (2014)).

In line with this evidence, in the model recourse is ineffective at discouraging default. The reason is simple: abovewater defaulters are not subject to recourse, because their home is worth more than the outstanding mortgage balance. Therefore recourse has no effect on abovewater homeowners, who make up the majority of defaulters. Recourse does discourage underwater homeowners from defaulting in the model, but there are few of them. This is consistent with the empirical evidence in Clauretie (1987), Ghent and Kudlyak (2011), and Li and Oswald (2014) that recourse does not lower default rates, while still being consistent with the evidence in Ghent and Kudlyak (2011) that recourse does lower the default rates of underwater homeowners. It is also consistent with Dobbie and Goldsmith-Pinkham (2015), who show that recourse lowers the default rates of underwater homeowners but not of abovewater homeowners.[5]

[3] Search frictions are a fundamental feature of housing markets, and modeling them helps theory match the data in a number of ways. For a comprehensive review of the literature, see Han and Strange (2015).

[4] Ghent and Kudlyak (2011) note "the fair market value restriction is likely present because the lender is often the only bidder at the foreclosure sale... In the absence of such a restriction, the lender could profit from a foreclosure by placing an artificially low bid."

[5] Dobbie and Goldsmith-Pinkham (2015) (using data from the foreclosure crisis) find that recourse does lower

Section 2 provides a brief overiew of existing evidence on the equity of defaulters. Section 3 discusses the data and empirical methodology used to estimate the equity of foreclosed homeowners. Section 4 presents the results of this estimation. Section 5 develops the theoretical model, which is estimated in Section 6. Section 7 presents results from the model.

2 Existing Evidence

There is a considerable amount of existing evidence on the equity of defaulters, but it all has significant issues. The evidence can be divided into three types.

First, most existing evidence on the relationship between equity and default comes from estimated default rates as a function of equity (e.g. Deng et al. (2000), Foote et al. (2008), Gerardi et al. (2009), Elul et al. (2010), and Gerardi et al. (2013)). These estimates uniformly show that abovewater homeowners are less likely to default, but by themselves they do not provide direct evidence on how many defaulters are abovewater. To give these estimates a quantitative interpretation, Figure 1 combines them with data on the equity distribution of homeowners from Fuster et al. (2016) to approximate the proportion of foreclosures with equity. For comparison, Figure 1 performs the same calculation using data on default rates as a function of equity and the equity distribution from CoreLogic.[6]

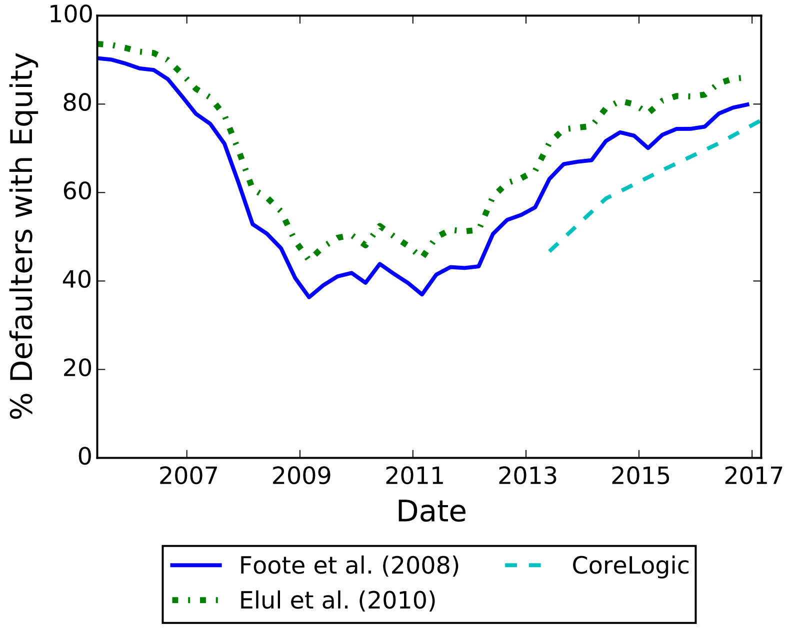
Figure 1 suggests that 80% or more of defaulters had positive equity before 2006, a number that dropped to 40% or less during the foreclosure crisis before climbing again. To understand why, Figure 2 displays estimates of the default hazard rate of underwater homeowners relative to abovewater homeowners, and of the number of abovewater homeowners relative to underwater homeowners over time. Figure 2 shows that according to typical estimates underwater homeowners are roughly 5 to 10 times more likely to default. It also shows abovewater homeowners outnumbered underwater homeowners by more than 50 to 1 before the foreclosure crisis, by 5 to 1 or less during the crisis, and by roughly 15 or 30 to 1 in 2017.

However, there are serious concerns with the evidence in Figure 1. Property values and mortgage balances are measured with error, which leads to attenuation bias in estimates of the correlation between equity and foreclosure hazard rates. Elul et al. (2010) and CoreLogic measure default before foreclosure occurs, so homeowners with equity in these data may have avoided foreclosure by selling their home. The samples used by Foote et al. (2008) and Elul et al. (2010) to calculate default rates may not be compatible, across space or time, with the nationwide equity estimates

default rates overall. A natural explanation for the discrepancy between their findings and those of Clauretie (1987) and Ghent and Kudlyak (2011) (who use data from before the foreclosure crisis) and Li and Oswald (2014) (who use data on mortgages originated in Nevada after September 2009) is that the collapse in house prices from 2006-2009 drove many more homeowners underwater than is typical. See Section 2.

[6] CoreLogic is a data and analytics company. Data is from CoreLogic (2013), CoreLogic (2014), CoreLogic (2015), CoreLogic (2016), and CoreLogic (2017).

Figure 1: Interpretation of Default Rate Estimates



Notes: Default rates from Foote et al. (2008) and Elul et al. (2010) are combined with data on the number of homeowners in LTV bins of width 5 provided by Fuster et al. (2016). Default rates from CoreLogic are combined with data on the proportion of homeowners in LTV bins from CoreLogic. The estimated number of foreclosures in each bin is the number of homeowners in that bin times the foreclosure hazard rate of that bin.

from Fuster et al. (2016).[7] Finally, Fuster et al. (2016) note for several reasons that they may underestimate the proportion of homeowners with negative equity during the mortgage crisis.

A second and more direct kind of evidence comes from estimates of the number of defaulters with positive equity. Sources using older data report high numbers, e.g. 90.3% in Ambrose and Capone (1998), 90.8% in Deng et al. (2000), and 99.7% in Pennington-Cross (2003).[8] Estimates from the foreclosure crisis are lower, e.g. 31% in Haughwout and Okah (2009) and 35% in Laufer (forthcoming).[9] In line with intuition provided by Figure 2, estimates appear to be going up again. For example, Realtytrac estimates that the percentage of foreclosures with positive equity increased from 24% to 49.7% from 2013 to 2015.[10]

There are also serious concerns with these estimates. Except for RealtyTrac, none come from

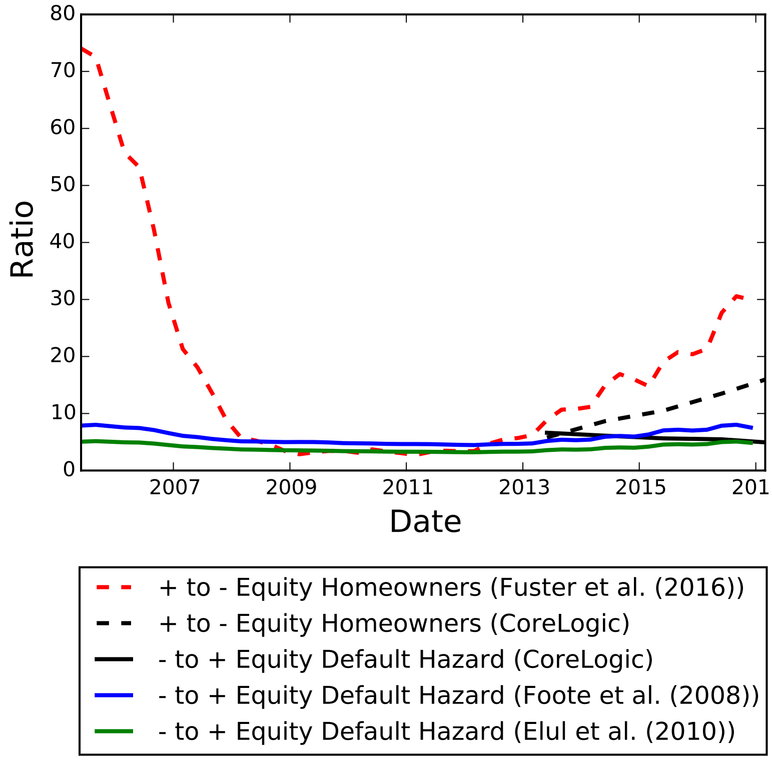
[7]For example, Foote et al. (2008) use data from Massachusetts, which is a recourse state. The relative forecelosure hazard rate of underwater homeowners is likely higher in non-recourse states.

[8]The samples differ between these papers. Ambrose and Capone (1998) study defaulted FHA mortgages from 1988 through 1994. Deng et al. (2000) use Freddie Mac data from 1976-1992. Pennington-Cross (2003) studies foreclosed homeowners in Fannie Mae and Freddie Mac from 1995-1999

[9]Haughwout and Okah (2009) study subprime and Alt-A securitized mortgages in December 2008 that are in foreclosure or have been foreclosed on. Laufer (forthcoming) studies homeowners in Los Angeles County who had purchased their homes between 2000 and 2003 and defaulted by 2009.

[10]See `https://www.washingtonpost.com/news/wonk/wp/2014/04/18/more-homeowners-no-longer-need-to-be-in-foreclosure-and-they-may-not-even-know-it/?utm_term=.53aee2008d01` and `https://www.realtytrac.com/news/home-prices-and-sales/2015-year-end-home-equity-and-underwater-report/`.

Figure 2: Ratio of Base Rates and Default Hazard Rates



Notes: Data on the number of homeowners in LTV bins of width 5 are provided by Fuster et al. (2016). Foreclosure hazard rates of each bin are from Foote et al. (2008) and Elul et al. (2010). The estimated number of foreclosures in each bin is the number of homeowners in that bin times the foreclosure hazard rate of that bin. The estimated foreclosure hazard rate of homeowners with positive (negative) equity is the estimated number of foreclosures with positive (negative) equity divided by the estimated number of homeowners with positive (negative) equity.

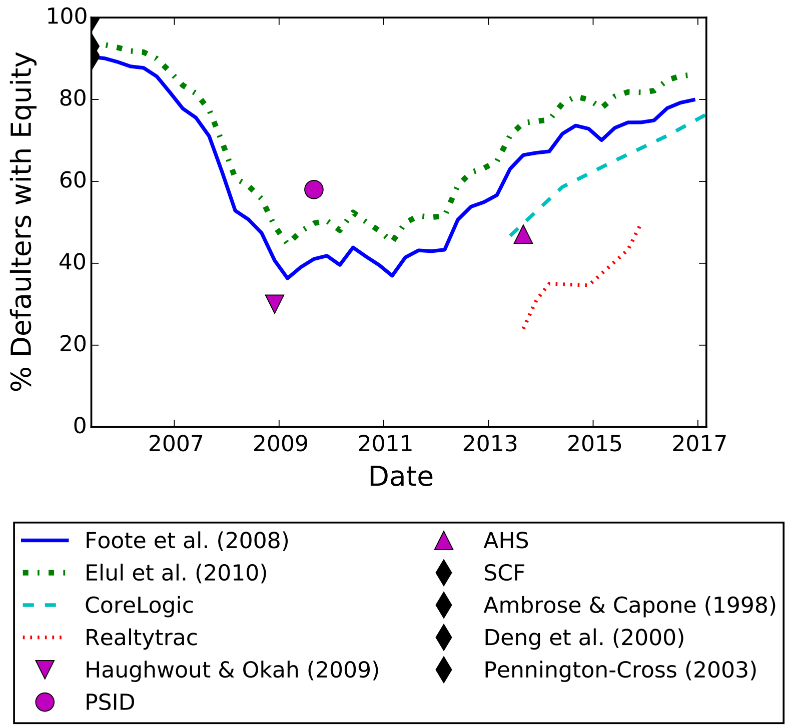
a sample likely to be representative of all U.S. mortgagors. Many use only the first lien against a property to estimate mortgage balances outstanding, and so they understimate the combined LTV of homeowners with other liens. Most use transaction prices updated with a regional HPI to estimate the value of the property, which neglects idiosyncratic shocks (Korteweg and Sorensen (2016) and may therefore overstate the equity of defaulters. Again, the samples typically include homeowners who have not yet lost their home to foreclosure, and therefore may still sell before foreclosure occurs.

Finally, there is publicly-available survey evidence on the equity of defaulters. This evidence is largely consistent with estimates from other sources. For example, 93% of defaulters in the 1998 and 2001 Survey of Consumers Finances report having positive equity, while 58% of homeowners in the Panel Survey of Income Dynamics (PSID) who lost their home to foreclosure between 2008 and 2010 reported positive equity in 2008.[11] In the 2013 American Housing Survey (AHS), 47% of homeowners who report a "somewhat" or "very" high probability of losing their home to foreclosure in the next two months also report having positive equity.

These sources, like the others, raise serious concerns. Homeowners generally report the value of their homes with bias and noise (Kiel and Zabel (1999)). Also, out of necessity, equity in these datasets is measured before foreclosure occurs, so homeowners with positive equity may still sell their homes to avoid foreclosure, or lose equity before foreclosure occurs.

To summarize, Figure 3 plots estimates from existing sources on the percent of defaulters with equity over time.

Figure 3: Existing Evidence on Defaulters with Equity



Notes: Default rates from Foote et al. (2008) and Elul et al. (2010) are combined with data on the number of homeowners in LTV bins of width 5 provided by Fuster et al. (2016). Default rates from CoreLogic are combined with data on the proportion of homeowners in LTV bins from CoreLogic. The estimated number of foreclosures in each bin is the number of homeowners in that bin times the foreclosure hazard rate of that bin. Numbers from the SCF, PSID, and AHS are calculated by the author. Numbers from other sources are provided directly by those sources.

To the best of my knowledge, this is a comprehensive review of existing evidence on the percent of defaulters with equity. As emphasized, all of this evidence has significant issues. Given the considerable implications of the equity of defaulters for mortgage default, the lack of higher-quality

[11] Details on these calculations are available from the author.

evidence is a significant gap in the literature. This gap is especially concerning because the evidence that does exist is not consistent with the frictionless models that dominate the literature, which predict that few or no defaulters have equity. The next two sections provide the first formal estimates of the equity of defaulters. These estimates are not subject to the concerns raised above.

3 Empirical Methodology

3.1 Data

The primary data source for estimation is the American Housing Survey (AHS), a publicly-available biannual panel dataset of properties produced by the Census Bureau. The AHS has unparalleled detail on time-varying idiosyncratic property depreciation. This is critical since properties in foreclosure many depreciate more quickly than other properties. The AHS also has useful controls for geographic shocks, although these are not as extensive. For example, the finest observed geographic unit in the public version of the AHS is the Census division. The estimation procedure outlined in Subsection 3.2 directly controls for changes in observable property characteristics.

The AHS also includes questions on the transaction prices and homeowner-estimated values of properties. The estimation procedure outlined in Subsection 3.2 optimally filters these signals to account for unobserved shocks affecting property values, including unobserved geographic shocks. These signals are also used to characterize the uncertainty in property values arising from unobserved shocks.

Data is from the national AHS samples from 1997 - 2013.[12] The full dataset has 108,368 properties. I first drop properties that are ever reported to be anything besides a house or apartment, leaving 85,600 properties. I then drop properties that are not regular, owner-occupied interviews at least half the time. This is a significant restriction, leaving 43,653 properties. Next I drop properties for which we have no reported value or transaction price, leaving 41,922 properties. Finally, I drop properties with any of three significant issues in value: (1) a reported value ever in the bottom percentile, (2) a reported value topcoded by the Census Bureau, or (3) a highest reported value *or* transaction price that is more than 10 times the lowest reported value *or* transaction price. This final restriction leaves 33,874 properties in the full sample. The full sample is used to estimate the parameters of the model.

The AHS unfortunately does not directly identify properties that went through foreclosure. However, the 2013 wave of the AHS does include questions that identify two subsamples of interest. The first are properties that were occupied in 2011, but vacant during the 2013 AHS because of

[12] The AHS drew a new sample for 2015, and so I do not use data beyond 2013.

foreclosure.[13] Relative to an ideal sample of all properties foreclosed between 2011-2013, this sample includes properties vacated but not yet foreclosed in 2013, and does not include properties vacated but re-occupied by the time of the 2013 AHS. The former likely oversamples from underwater homeowners, since they should be less likely to stay in the property to fight the foreclosure. The latter likely consists disproportionately of abovewater foreclosures, because banks are less likely to offload foreclosed properties in depressed markets. Therefore, this sample is likely to understate the equity of foreclosed homeowners.

Still, for robustness, this section also estimates the LTVs of another subsample of interest: homeowners who in 2013 reported the probability of moving because of foreclosure within the next two months to be “somewhat” or “very” high.[14] Homeowners with equity in this state may still be able to sell their home to avoid foreclosure. They may also fight harder against foreclosure and thus remain in this state for more time. Therefore, this sample seems likely to have more equity than a sample of foreclosed homeowners.

Data on house price indices is from Black Knight.

3.2 Methodology

The estimation procedure I use is an extension of the one developed by Korteweg and Sorensen (2016). This extension exploits past and future data on transaction prices, homeowner-reported values, and hedonic characteristics of a property to estimate its current value. It accounts for unobserved shocks to the value of a property, the bias and noise in homeowner-reported values, and the noise in reported transaction prices that may result from measurement error or search frictions.

I assume the log market value V of property i at time t consists of an observed component and an unobserved component, with:

$$V_i(t) = X^P \beta^P + U_i(t) \tag{1}$$

X^P are observable characteristics of the property, and β^P is the vector of their hedonic prices. $U_i(t)$ consists of the unobservable characteristics of the property that affect its value. $U_i(t)$ follows a Markov process, with

$$U_i(t) = U_i(t-1) + \epsilon_i^U(t) \tag{2}$$

The market value of a property is never directly observed. However, the AHS provides two noisy measurements of the market value of a property: homeowner-reported values of the property and reported transaction prices. These noisy measurements are filtered to estimate a property’s market value.

[13] After restricting the data as outlined above, this sample contains 46 properties.
[14] After restricting the data as outlined above, this sample contains 44 properties.

I assume observed (log) transaction prices are a function of the market value of a property as follows:

$$P_i(t) = V_i(t) + \epsilon_i^{SF}(t) + \epsilon_i^{ME}(t)$$

where the idiosyncratic, temporary, mean zero shocks to P consist of shocks due to search frictions, ϵ^{SF}, and shocks due to measurement error, ϵ^{ME}. The variances of these shocks are not separately identified, so the equation I estimate is:

$$P_i(t) = V_i(t) + \epsilon_i^{P}(t) \tag{3}$$

where the idiosyncratic, temporary shocks to P, ϵ^P, have mean zero and variance σ_P^2.

The market value of a property is also measured with error when survey respondents report the value of their home. As is well-known, there is considerable bias and noise in these self-reported values (e.g. Kiel and Zabel (1999).) I therefore assume homeowners report the log value of their home as follows:

$$R_i(t) = V_i(t) + X^R\beta^R + \epsilon_i^{R}(t) \tag{4}$$

where $R_i(t)$ is the self-reported home value, X^R are observable characteristics of the homeowner or property that bias the report by β^R, and ϵ_i^R is normally distributed with mean zero and variance σ_R^2.

Equations 1, 2, 3, and 4 together define the model, which I estimate at a quarterly frequency. Theoretically, the parameters of the model could be estimated by maximum likelihood. However, maximum likelihood is numerically intractable because the state variable U needs to be estimated for every quarter *for every property* in the dataset.[15] Instead, following Korteweg and Sorensen (2016), I implement a Bayesian Markov Chain Monte Carlo (MCMC) method called Gibbs sampling.

Because Gibbs sampling is described in detail in Korteweg and Sorensen (2016) and many other places, I provide only an overview of the specific procedure used here. The parameters that need to be estimated are the β and $var(\epsilon)$ terms in Equations 1, 2, 3 and 4 and the unobserved component of value $U_i(t)$ for every property in every time period.

The first step of the Gibbs sampler recovers the posterior distributions for U. To understand how, note that Equation 2 defines a dynamic linear state space model for U. U is never directly observed, but transactions provide noisy observations of U through Equations 1 and 3, while homeowner-reported values provide noisy observations of U through Equations 1 and 4. Conditional

[15] There are over 30,000 properties in the data. Each of these properties is tracked for 65 quarters, yielding over 2 million parameters to be estimated.

on values of the β and $var(\epsilon)$ terms in Equations 2, 3 and 4, a Kalman filter could generate the posterior for U at time t, conditional on information before time t. However, a Kalman filter alone would be inefficient, since information after time t should also be included in estimates of U at time t. Therefore, I implement the Forward-Filtering Backwards-Sampling (FFBS) algorithm of Carter and Kohn (1994) and Fruhwirth-Schnatter (1994). To provide some rough intuition, the FFBS algorithm may be thought of as running a Kalman filter forward and then backwards in time; it is described in more detail in the appendix.

The second step of the Gibbs sampler estimates the β and $var(\epsilon)$ terms in Equations 1, 2, 3 and 4. Conditional on draws of $U_i(t)$ provided by the FFBS algorithm, slightly modified versions of Equations 1, 2, 3, and 4 provide regressions that can be estimated by standard Bayesian Ordinary Least Squares. The regressions are described in the appendix. They provide posteriors for the β and $var(\epsilon)$ terms, which are used to provide draws for the next round of the FFBS algorithm.

Every round of the Gibbs sampler performs these two steps in turn to draw values for every parameter. These draws converge quickly to the posterior distributions of the parameters. I iterate the Gibbs sampler 2000 times, discarding the first 1500 iterations as a burn-in period and using the last 500 iterations to approximate the posterior distributions of the parameters.

4 Empirical Results

This section first presents estimates of the parameters of Equations 1, 2, 3 and 4, which are obtained by the second stage of the Gibbs sampler. Then it presents estimated LTVs, which are obtained in the first stage of the Gibbs sampler and are the main object of interest.

Because all estimates are obtained from the Gibbs sampler, the uncertainty in every parameter accounts for uncertainty in every other parameter. For example, the uncertainty in the value of a property comes not just from the measurement error in observations of that property, but also from uncertainty in the model parameters used to filter those measurements.

A major advantage of the AHS is its rich set of observables that can be included as controls in Equations 1, 2 and 4.[16] I introduce these variables in groups, yielding four specifications of the model. All specifications include a constant term in Equation 4, but not in Equation 1. A constant term in Equation 1 is not separately identified from the mean value of U.

Specfication 1 includes only the Census-division HPI in Equation 1. This is the finest HPI level available in the public version of the AHS.

Specification 2 introduces promising controls into Equation 4 describing self-reported home values. These controls are (1) a dummy variable for whether the respondent has graduated from

[16] Note that, due to the time-varying property-specific $U_i(t)$ in Equation 1, all coefficients of Equation 1 are identified by within-property changes in observed value.

high school, (2) a dummy variable for whether the respondent has owned a home before, and (3) the quarters elapsed since the purchase of the property (capped at 12). The goal of this specification is to control for observables that may bias self-reported home values in ways that systematically affect foreclosed properties. For example, foreclosed homeowners may be more biased than other homeowners when reporting the value of their homes. This would bias estimates of the LTVs of foreclosures downwards.

Specification 3 addresses the concern that foreclosed properties may experience worse local price shocks than other properties. The Census division divides the U.S. into only seven regions, so the HPI measure is quite coarse. To the extent that more local HPI movements are reflected in observed transaction prices and homeowner-reported values, they are still incorporated into LTV estimates.[17] However, foreclosures were clustered in areas with the greatest HPI declines, so this is a potentially serious issue. To check, Specification 3 includes three additional control variables. The first is "Fair Market Rent", which is a valuable proxy for a local HPI because it is available at a finer geographical level than Census division.[18] Second is the quality of the neighborhood, as reported by the homeowner, on a scale from 1 to 10. The third control is the presence of abandoned or vandalized buildings within half a block of the property, which is mainly intended to control for nearby foreclosures.[19]

Specification 4 addresses the concern that foreclosed properties may experience worse idiosyncratic depreciation shocks than other properties. It controls for a number of variables likely to proxy for depreciation. These are (1) whether a window in the property is broken or boarded up, (2) whether there are cracks or holes wider than the edge of a dime in the inside walls or ceilings, (3) whether there are holes, cracks, or crumbling in the property foundation, (4) whether there are holes in the roof, (5) whether there were water leaks from inside the property in the past year, and (6) whether there were water leaks from outside the property in the past year. These variables control for property depreciation to the extent that they are causal as well as to the extent they are correlated with otherwise unobserved depreciation.

Estimates of the parameters of Equations 1 and 2 are shown in Table 1. Estimates of the parameters of Equation 4 are shown in Table 2.

There are few surprises in Table 1. Almost all control variables have the expected sign and are highly statistically significant. Homeowner-reported neighborhood quality is not significant, which

[17]Because house prices declined significantly between 2006 and 2009 before stabilizing, and foreclosure LTVs are measured between 2011 and 2013, there is considerable time for these signals to be incorporated into the estimates.

[18]In the public use file, 'Fair Market Rent" is available for unique values of Census division crossed with urban crossed with temperature. "Fair Market Rent" is calculated by HUD; for more details, see https://www.huduser.gov/portal/datasets/fmr.html

[19]Unfortunately this variable is only observed for half the sample in 2013 . It can still be included in the regressions under the assumption that it does not change for properties for which it is not observed. Alternatively, one can think of the variable not as indicating a nearby abandoned or vandalized building, but as indicating the observation of a nearby abandoned or vandalized building.

may indicate that it is a measure of the homeowner's valuation of the neighborhood more than the market's. Surprisingly, holes in the roof has an estimated effect near zero that is fairly precise. This may be because the control for outside leaks picks up most of the effect for holes in the roof. It may also be because roofs are expensive to fix, and so a homeowner may allow roof problems to develop while otherwise maintaining the property well. In contrast, the estimated effect of a broken window is surprisingly large. This is likely because windows are relatively cheap to replace, so if a broken window is not repaired it indicates the property is generally not being maintained. It is also interesting to note that controlling for idiosyncratic property depreciation decreases the standard errors of the effects of the controls for local neighborhood shocks. This may be because depreciating properties tend to cluster. Estimates of σ_V^2 are substantial, indicating significant search frictions or measurement error (or both) in the AHS.

Table 1: Estimated Parameters of Equations 1 and 2.

	A	B	C	D
HPI	0.6677***	0.662***	0.6585***	0.6558***
	(0.0105)	(0.0097)	(0.0232)	(0.011)
Abandoned Buildings			-0.1317	-0.1679***
			(0.1505)	(0.0248)
Fair Market Rent			0.348***	0.3215***
			(0.087)	(0.0175)
Neighborhood Quality			-0.0044	-0.0024
			(0.0065)	(0.0021)
Broken Window				-0.1533***
				(0.0198)
Cracks in Walls				-0.0693***
				(0.0154)
Crumbling Foundation				-0.1534***
				(0.0202)
Holes in Roof				0.0141
				(0.027)
Inside Leaks				-0.0261**
				(0.011)
Outside Leaks				-0.0612***
				(0.0102)
σ_U^2	0.0037***	0.0037***	0.004***	0.0037***
	(0.0)	(0.0)	(0.0007)	(0.0002)
σ_V^2	0.0275***	0.0273***	0.0269***	0.0271***
	(0.0005)	(0.0005)	(0.0013)	(0.0006)

Notes: Standard errors are in parentheses. *, **, and *** indicate statistical significance at the 10, 5, and 1 percent level, respectively.

Table 2 indicates that homeowners overestimate the value of their home on average by roughly 5-7%, a value well in line with others in the literature. High school graduation appears to mitigate this bias, but previous homeownership does not. Bias also goes down with tenure, which may be evidence that homeowners learn the value of their home over time (Kiel and Zabel (1999), Davis and Quintin (2017)).

Table 2: Estimated Parameters of Equations 4.

	A	B	C	D
Constant	0.0602***	0.074***	0.0789***	0.0789***
	(0.0015)	(0.0033)	(0.004)	(0.0032)
First Time Homeowner		-0.0002	0.0	-0.0003
		(0.0029)	(0.0022)	(0.0026)
High School Grad		-0.0127***	-0.0175***	-0.0145***
		(0.0027)	(0.004)	(0.0026)
Quarters Since Purchase		-0.0007***	-0.0019***	-0.0019***
		(0.0002)	(0.0005)	(0.0002)
σ_R^2	0.0195***	0.0194***	0.0194***	0.0215***
	(0.0002)	(0.0002)	(0.0014)	(0.0006)

Notes: Standard errors are in parentheses. *, **, and *** indicate statistical significance at the 10, 5, and 1 percent level, respectively.

With these estimates, we can now turn to the main objects of interest: the LTVs of foreclosed homeowners. Recall that LTV is defined as total mortgage debt outstanding divided by the value of the property, multiplied by 100.

Mortgage debt is the total principal outstanding on all mortgage loans and home equity lines of credit. Outstanding balance on some liens is sometimes missing. Because it is unlikely to be missing at random, and because of the small sample size, I impute outstanding balance when it is missing. I do so by assuming the respondent was current on the debt until five years before the survey, at which point delinquency started and late fees and interest began to accrue.[20]

$U_i(t)$ values, and hence LTVs, are random variables that are drawn in the second stage of the Gibbs sampler. The Gibbs sampler produces posteriors for the LTV of every property in the dataset. This distribution characterizes not only the estimated LTV of a given property, but also the uncertainty in that estimate arising from measurement error, changes in property value that occur between measurements, and uncertainty in the parameter estimates of the empirical model.

Figures 4 plots histograms of pooled LTV draws from all four model specifications for properties vacant in the 2013 AHS because of foreclosure. Figure 5 does the same for properties with a "somewhat" or "very" high reported probability of being foreclosed within the next two months. The percent of LTV estimates below 100 are reported directly by Table 3.

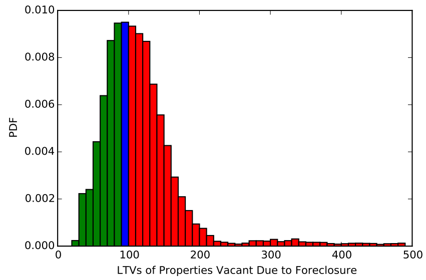
The results differ remarkably little between specifications. This could indicate that before foreclosure observables do not generally deteriorate for properties in the sample.[21] It could also indicate that homeowners accurately incorporate these observables into their estimates of their home's value. This latter interpretation is consistent with Kiel and Zabel (1999), who find that

[20]I assume the late fee on a missed payment is equal to 5% of its value, while the annualized net penalty interest rate on delinquent debt is 12%. Unfortunately I am not aware of any systematic data on these fees, but these values are reasonable. Other reasonable values change the results very little.

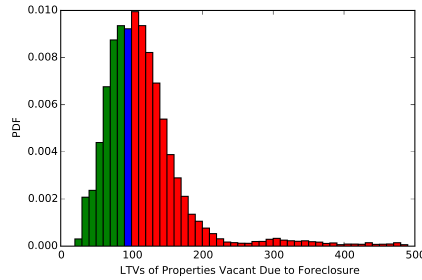
[21]For example, the depreciation often associated with foreclosed properties could occur after properties become vacant.

Figure 4: Distribution of LTV Draws For Properties Vacant Because of Foreclosure

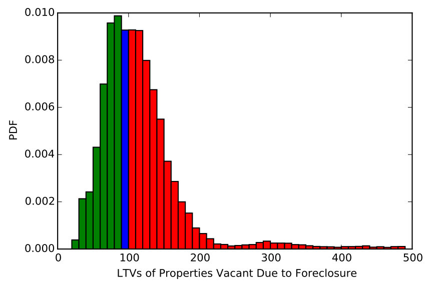
(a) Specification A



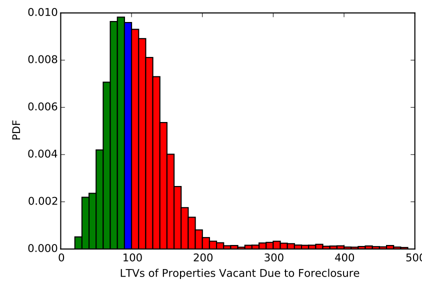
(b) Specification B



(c) Specification C



(d) Specification D



Notes: The Loan-to-Value ratio (LTV) is defined as 100 times the ratio of total mortgage debt outstanding to a draw of the property value from the Gibbs sampler. The sample consists of the 46 properties that were occupied in 2011 but vacant in 2013 because of foreclosure. Mortgage debt and house value are measured at the time the property became vacant. There are 500 draws of the property value, and hence LTV, for each property; all 500 draws for all 46 properties are pooled together. Red denotes draws of underwater LTVs above 100; blue denotes draws of abovewater but "effectively" underwater LTVs between 90 and 100; green denotes abovewater draws of LTVs below 90.

Table 3: % of Foreclosures Abovewater

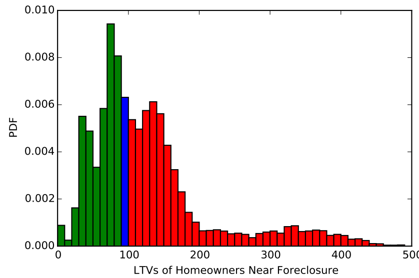
Specification	A	B	C	D
% Foreclosed with Estimated LTV<100	42.9	40.6	45.2	47.4
% Foreclosed with 95th LTV Percentile<100	22.6	27.1	27.1	27.0
% Near Foreclosure with Estimated LTV<100	51.1	47.1	51.1	51.1
% Near Foreclosure with 95th LTV Percentile<100	34.9	34.9	35.9	35.9

Notes: Table reports the percent of properties with mean and 95th percentile of Loan-to-Value ratio (LTV) posteriors below 100 for each of four model specifications. LTV is defined as 100 times the ratio of total mortgage debt outstanding to a draw of the property value from the Gibbs sampler. The sample of foreclosed properties consists of the 46 properties that were occupied in 2011 but were vacant in 2013 because of foreclosure. Mortgage debt and house value for this sample are measured at the time the property became vacant. The sample of properties near foreclosure consists of the 44 properties with data in the 2013 national AHS for which the reported probability of foreclosure within the next two months was "somewhat" or "very" high. Mortgage debt and house value for this sample are measured at the time of the survey.

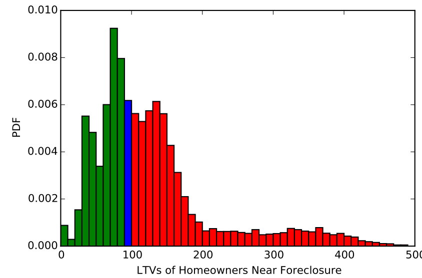
errors in homeowner-reported values are uncorrelated with observable characteristics of the owner or the property, except for the owner's tenure.

Figure 5: Distribution of LTV Draws For Properties Near Foreclosure

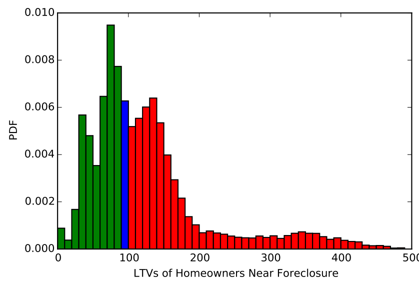
(a) Specification A



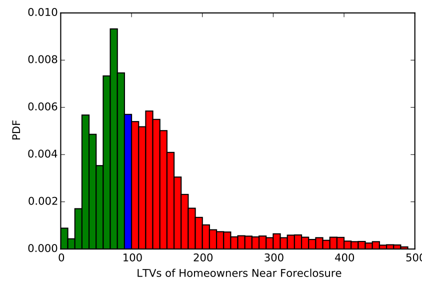
(b) Specification B



(c) Specification C



(d) Specification D



Notes: The Loan-to-Value ratio (LTV) is defined as 100 times the ratio of total mortgage debt outstanding to a draw of the property value from the Gibbs sampler. The sample consists of the 44 properties reported to have a "somewhat" or "very" high probability of foreclosure within the next two months. There are 500 draws of the property value, and hence LTV, for each property; all 500 draws for all 44 properties are pooled together. Red denotes draws of underwater LTVs above 100; blue denotes draws of abovewater but "effectively" underwater LTVs between 90 and 100; green denotes abovewater draws of LTVs below 90.

Results indicate that roughly 40-50% of foreclosures between 2011 and 2013 in the AHS had positive equity. These estimates should not be viewed as precise because of the small sample size of the AHS and the uncertainty in the LTV estimates. However, they provide strong support for the general conclusion from Section 2 that many defaulters have positive equity. Moreover, many more homeowners were underwater between 2011 and 2013 than is typical, and so the implications for more "normal" times are more precise. For example, under Specification D, 47% percent of foreclosed properties have a mean LTV below 100. Taking the 95th percentile of the posterior instead of its mean as the LTV estimate yields 27% percent of foreclosures with positive equity. However, combining the foreclosure hazard rates from Specification D with the LTV distribution of homeowners in the 1998 and 2001 SCF, the same change implies a drop in the estimated percent of foreclosures with positive equity from 87% to 81%.[22]

[22] The SCF has data on reported home values, not actual home values. One could correct for the bias and noise of reported home values using the parameter estimates in Table 2. However, doing so *decreases* the fraction of homeowners estimated to be underwater. This is because the noise in homeowner-reported values causes more abovewater homeowners to be incorrectly reported as underwater than vice-versa, since most homeowners are

Thus there is strong evidence that, outside the foreclosure crisis, a considerable majority of defaulters have positive equity. Section 5 develops a quantitative model with search frictions to see if it can rationalize these findings.

5 Model

This section builds a quantitative model of housing and mortgages over the lifecycle, which draws from the housing literature (Bajari et al. (2013), Li et al. (2016)) and mortgage literature (Jeske et al. (2013), Chatterjee and Eyingungor (2014), Corbae and Quintin (2015), Laufer (forthcoming)). I explicitly model the problems of both homeowners and renters, and model mortgages as long-term, refinancable and defaultable debt.

The model deviates from most models of foreclosure in two important ways. First, as in Hedlund (2016a), Hedlund (2016b), Head et al. (2016), and Garriga and Hedlund (2017), search frictions in the housing market mean that home sellers are not guaranteed to find a buyer. Second, foreclosure takes more than one period, following Herkenhoff and Ohanian (2015). This allows distressed homeowners in the model to respond to the search frictions by listing their home earlier in the foreclosure process. But if a homeowner does not make her mortgage payment, does not catch up on it later, and does not sell her home, she will lose her home to foreclosure, even if she has positive equity.

5.1 Environment

The baseline environment is a standard model of consumption, housing, and mortgage choice over the lifecycle, with similarities to Bajari et al. (2013) and Li et al. (2016). House purchases are characterized by a frictional search-and-matching process, along the lines of Genesove and Han (2012b) and Ngai and Sheedy (2017). The foreclosure process is the heart of the model, and is fairly unique to the literature, so it is discussed separately.

Baseline Time is discrete. Consumers receive an exogenous, stochastic income flow $\{y_t\}$. Agents face a chance of death every period, and die with certainty in period T. Consumers value consumption c, and discount the future at rate β.

Agents can purchase liquid, risk-free assets a, which earn a rate of return R. Agents can borrow against a fraction ξ of their income y, so the borrowing constraint is given by:

$$a^{'} \geq -\xi y \tag{5}$$

abovewater. This effect is stronger than the effect of the bias in homeowner-reported values. Therefore, to be conservative, I do not correct for the bias and noise in reported home values.

Agents care about the quality of their housing. Non-homeowners can choose to spend an arbitrary amount on rent r. The budget constraint for renters is:

$$R^{-1}a^{'} + c + r = a + y \tag{6}$$

Non-homeowners may also try to buy a house in a frictional search-and-matching market (Genesove and Han (2012b), Ngai and Sheedy (2017)). Hopeful buyers first choose the price p of a home to search for. After paying a financial search cost ι, they then draw the idiosyncratic quality q of the "best" house they find at price p from a distribution $\mathcal{Q}$. After observing q, potential buyers decide whether to buy the house and become a homeowner, or not to buy the house and remain a renter for the period. Buying a home incurs a proportional cost ϕ_b, so the total cost of buying a home is $(1+\phi_b)p$.

Home buyers without a default flag (discussed later) can finance their purchase with a mortgage. To keep the state space manageable, I assume that home buyers obtain fixed-rate mortgages that last until the terminal period. Agents' mortgages can therefore be summarized by the constant mortgage payment m.

Mortgage payments are offered to a mortgage lender in exchange for a loan of size L. L is chosen by risk-neutral mortgage lenders, who charge interest rate R_b, discount the future at rate β_B, observe the borrower's state and choice variables, and choose L to maximize their profits subject to perfect competition.[23] These conditions determine the function $L = L(a, p, q, m, y)$. Home buyers are also subject to exogenous LTV and "Payment-to-Income" ratio (PTI) limits (Corbae and Quintin (2015), Greenwald (2017)). House purchases are processed immediately, so home buyers instantly become a homeowner with house price p and mortgage payment m, and their assets a go down by their down payment, $(1+\phi_b)p - L(a, p, m, y)$.[24]

Houses can be sold, incurring a proportional cost ϕ_s that represents both broker fees and moving costs. Home sellers must repay their mortgage. This involves buying back the nominal sequence of payments, $\{m_t, m_{t+1}, ..., m_T\}$ at the interest rate R_b, which implies that the cost of repaying a mortgage of constant payment m at time t is:

$$\Pi(m,t) = m\frac{(1-(R_b^{-1})^{T-t-1})}{(1-R_b^{-1})} \tag{7}$$

Therefore the net proceeds from selling a house are $(1-\phi)p - \Pi(m,t)$.

For homeowners who are current on their mortgage, house sales are processed immediately. This

[23] In line with the rest of the literature, I assume that $\beta_b R_b = 1$.

[24] I assume that, when extending loan L, banks withhold the first mortgage payment m. This assumption means that the period after buying a home agents must be current on their mortgage. This makes the model considerably easier to solve.

means they do not face search frictions when selling their homes.[25]

Agents who are current on their mortgage may refinance, either to extact equity (and increase their mortgage payments) or inject equity (and decrease their mortgage payments.) To do so, they must repay their current mortgage in full by paying $\Pi(m^{old}, t)$, and in return they receive the cash from the new mortgage $L(a, p, m^{new}, y)$. This process involves a proportional fixed cost $\phi_R m^{new}$ borne by the lender and paid by the borrower. Therefore, the borrower's net proceeds from refinancing are $L(a, p, m^{new}, y) - \Pi(m^{old}, t) - \phi_R m^{new}$.

Homeowners who do not sell their home remain in it for the period. After deciding whether or not to refinance, remaining homeowners can choose to make their mortgage payment m and house maintenance cost ζp in order to remain current. The budget constraint for current homeowners who pay their mortgage on time is:

$$R^{-1}a^{'} + c + m + \zeta p = a + y \tag{8}$$

Current homeowners can also choose to become delinquent on their mortgage. Since mortgage delinquency and foreclosure is the heart of this paper, it is discussed next.

Mortgage Delinquency and Foreclosure Agents may become delinquent on their mortgage, which lets them avoid paying it this period or maintaining their home. The budget constraint for homeowners becoming delinquent is therefore:

$$R^{-1}a^{'} + c = a + y \tag{9}$$

Current homeowners who choose to become delinquent remain in their home throughout the period. They receive a foreclosure auction notice at the beginning of the next period. Unlike in most other models of mortgage default, at this point a homeowner still has two chances to avoid foreclosure. First, she may list her home for sale at price $p^l \leq p$, and the property sells with probability $\pi_s(p^l)$.[26] If the home is sold, she must repay her mortgage in full (including the missed mortgage payment plus interest $R^m m$) and pay the maintenance fee she skipped last period with interest ($R^m \zeta p$). The agent must also pay foreclosure fees $\phi_f(p, m)$. Therefore the net proceeds from this sale are $(1-\phi)p - \Pi(m, t) - R^m(m + \zeta p) - \phi_f(p, m)$. The agent then immediately becomes a non-homeower with assets a changed appropriately, and without a foreclosure flag.

Otherwise, the home is not sold. At this point, the homeowner's only option to avoid foreclosure

[25]This assumption is unrealistic and lowers the rate at which homeowners with equity default in the model. I make this assumption primarily so that all foreclosures in the model are fully endogenous, in the sense that before foreclosure occurs a homeowner must affirmatively choose delinquency over selling the home.

[26]Note that delinquent homeowners face search frictions while current homeowners do not. I make this assumption, in part, because delinquent homeowners have less time to sell.

is to negotiate with the lender. Many delinquent borrowers do self-cure (Herkenhoff and Ohanian (2015)), so it is important to include this possibility in the model. I assume that, to become current, the homeowner must first *reinstate* the mortgage by making up her delinquent mortgage debt plus interest and fees, using her assets and a portion of her income. In exchange for postponing the auction (i.e. allowing the homeowner to use a portion of her income to reinstate the mortgage), the lender imposes the requirement that the borrower must be willing (without commitment) to stay current on the mortgage this period (either by paying it on time, selling the property, or refinancing).[27] Thus, if the homeowner can afford to reinstate her mortgage and is willing to stay current this period, the foreclosure is canceled and she keeps her home.[28] Otherwise, she loses her home to foreclosure. Agents who are underwater when they lose their home to foreclosure pay a utility cost υ.[29]

After foreclosure, banks sell the house at a proportional discount χ. If the proceeds from the sale exceed the outstanding mortgage balance, the excess is returned to the defaulter. A defaulter therefore receives proceeds of $\max\{0, (1-\chi)p - \Pi(m,t) - R^m m - \phi_f(p,m)\}$, and has as a budget constraint:

$$R^{-1}a' + c = a + y + \max\{0, (1-\chi)p - \Pi(m,t) - R^m m - \phi_f(p,m)\} \tag{10}$$

After defaulting, agents are excluded from the mortgage market with a foreclosure flag that lasts for a fixed number of years.

6 Estimation

The model is estimated so that it is applicable to a "normal" housing market, i.e. not the foreclosure crisis. This is done because (1) an understanding of foreclosure in normal times seems a necessary first step towards understanding foreclosure during the crisis, and (2) in important respects the market appears to be returning to its pre-crisis state.[30]

[27] This is a highly stylized way to model the negotiation process between borrower and lender. The incentives lenders have when dealing with delinquent borrowers are quite complex. If lenders do not fully observe borrowers' state variables, then extending help to borrowers who need it incurs the additional cost of sometimes extending help to borrowers who do not need it (Foote et al. (2010)). Lenders will be especially willing to renegotiate a mortgage if they want to preserve a borrower-friendly reputation. Conversely, they may be especially unwilling to renegotiate if they want to deter other borrowers from defaulting. Modeling these incentives more fully is beyond the scope of this paper.

[28] In reality, homeowners can become current at this stage in various ways. In most states, homeowners can *reinstate* the mortgage up to a certain date before the foreclosure auction by repaying, in one large payment, their delinquent mortgage debt plus interest and fees out of their current assets. In other states and at other times, homeowners have to repay their *entire* mortgage debt (ie. not just the portion that is delinquent.) Neither path seems viable for most delinquent homeowners in the model, so I allow for a process that is more generous to delinquent borrowers in exchange for requirements imposed by the lender.

[29] The interpretation of and motivation for υ is discussed in more detail when it is estimated in Section 6.

[30] See Figure 3.

6.1 Estimation

I estimate the parameters of the model by the Simulated Method of Moments (SMM). This two-step procedure, developed by Pakes and Pollard (1989) and Duffie and Singleton (1993), is now a standard tool to estimate the parameters of structural models without closed form solutions.

In the first step, I choose parameter values that are standard in the literature, or that can be estimated from the data without the use of a structural model. In the second step, I take the parameters from the first stage as given, and estimate the remaining parameters by minimizing the distance between empirical moments and model output.

The details of the two steps are described in turn.

6.1.1 First Stage

The values for parameters that can be estimated without the use of the structural model are described below.

Demographics: A period in the model is one year. Households begin life at age 23, retire at 65, and die with certainty at 85. Age-specific mortality rates are from the 2008 National Longitudinal Mortality Survey (NLMS).

Bequest Function: The bequest function B is taken from De Nardi (2004). Note that, upon death, an agent with assets a, house price p, and mortgage payment m leaves behind wealth worth $w = a+(1-\phi_s)p-\Pi(m,t)$. Let $c^*(w)$ and $r^*(w)$ denote optimal consumption and rent, respectively, from the one-period renter's problem with cash-on-hand w. Then I set:

$$B_t(l,m) = \upsilon_1 u_t(c^*(w+\upsilon_2), r^*(w+\upsilon_2), m)$$

The parameter υ_1 controls the strength of the bequest motive. υ_2 controls the extent to which bequests are luxury goods. I set $\upsilon_1 = 1$ and $\upsilon_2 = 11.6$, following De Nardi (2004).

Debt and Liquid Assets: I set the proportion of labor income that can be borrowed against, ξ, to .2, following Heathcote et al. (2010). The real interest rate on liquid assets is set to 1%.

Housing: The real appreciation rate of home values is 0, matching the rate in Shiller (2008) for 1987-2000, and the value used in Li et al. (2016) .

The flow value of housing, κ, is set to a typical value of 7.5%, e.g. Li & Yao (2007), Li et al. (2016).

Calibrating the financial cost ι of searching for a home is difficult. As noted by Ngai and Sheedy (2017), no high-quality estimates of this cost exist. Therefore I follow Ngai and Sheedy (2017) in calibrating ι to refelct the opportunity cost of time spent during search, and assume that each home visited takes one day. In Genesove and Han (2012b), the average buyer visits 9.96 houses. Assuming 261 workdays in a year, this yields a search cost $\iota = \frac{9.96}{261}$.

Calibrating the distribution $\mathcal{Q}$ of housing match quality is also difficult. Genesove and Han (2012a) assume an extreme value distribution of home valuations among serious bidders for a property and estimate that the scale parameter of this distribution is .038. Therefore, the highest valuation of a given property with n bids is the maximum of n draws from the extreme value distribution with scale parameter .038. Genesove and Han (2012a) also report the distribution of n.[31] Taking expectations over n of the maximum of n draws from the extreme value distribution with scale parameter .038 therefore yields $\mathcal{Q}$.

Maintenance costs ζ are typically set between 2% and 2.5%, but these values do not account for property taxes. Ngai and Sheedy (2017) do, using British data, and set it to 4.5%. This is high for the U.S. context, so I set them to 3%.

The proportional cost of selling a home is set to the fairly standard value of $\phi_s = .10$ to account for broker fees and moving costs. The proportional cost of buying a home, ϕ_b, is set to .03.

Mortgages: Assuming a 25% tax bracket, the median real after-tax interest rate on mortgages in the PSID is approximately 3.66%. Therefore I set the interest rate on mortgages, R_B, to 3.66%.

In reality, mortgage borrowers face a range of down payment requirements, from 20% to 5% or even less, depending heavily on their credit history and other factors. As a compromise, the exogenous LTV cap is set at 87.5.

The exogenous PTI cap is set so that a homeowner's mortgage payment m can be no higher than 35% of income y.[32] As implemented in the model, this cap does not account for other debt payments and so is a "front-end" PTI limit. 35% is generous for a front-end limit; typical underwriting standards before the housing boom required a front-end PTI of 28% (Greenwald (2017).)

For homeowners in default, the probability of sale as a function of relative list price is taken from Guren (forthcoming). However, estimates from Guren (forthcoming) are for the probability of sale within 13 weeks. The timing assumption in the model is that delinquent homeowners have just received the foreclosure auction notice. Typically, these notices arrive somewhere between two weeks and two months before the auction. Assuming that it takes two weeks to list the house for sale, and that the foreclosure notice provides two months' notice, gives a delinquent homeowner in the model roughly 6 weeks to sell the property. Therefore, letting $\pi_{Guren}(p^l)$ denote the probability

[31] See Table 6 in Genesove and Han (2012a).

[32] To avoid tracking the temporary income shock as a state variable, for the purposes of calculating origination PTI the temporary shock is set to its median value.

of sale as a function of relative list price p^l from Guren (forthcoming), the probability of sale as a function of list price in the model is $\pi_s(p^l) = 1 - ((1 - \pi_{Guren}(p^l)^{\frac{1}{13}})^6)$.

Foreclosure fees are set to 10% of the annual mortgage payment. I am unfortunately not aware of systematic data on foreclosure fees, but this number is within reasonable bounds. I also assume that house maintenance is 10% more expensive if done late. Thus $\phi_f(p, m)$ is set to $.1(m + \zeta p)$. Delinquent homeowners who wish to get current must pay this fee in addition to the delinquent mortgage payment, interest on the delinquent mortgage payment, and the previous period's house maintenance.

Recall that χ denotes the deadweight loss of foreclosure, as a fraction of the value of the foreclosed home. Pennington-Cross (2006) estimates that foreclosed properties sell for roughly 22% less than similar properties nearby. Since this number presumably accounts for the poorer state of foreclosed properties, I set χ so that $\chi + (1.1\zeta) = .22$, so $\chi = .187$.

The income delinquent homeowners are allowed to use to reinstate their mortgage is set to be equal to the value of their temporary income shock that period. This assumption means that the temporary income shock for delinquent homeowners does not need to be tracked as a state variable, and so makes the model considerably easier to solve. It does not have a natural theoretical interpretation. However, it is quantitatively generous to delinquent homeowners and allows them a significant chance of avoiding foreclosure. With the chosen income parameters (discussed below), the median value of the temporary income shock is roughly 32% of permanent income. If delinquent homeowners could instead only use income received before the foreclosure auction, this would allow them to use 16.7% of income to reinstate the mortgage.[33]

Foreclosure flags in the model last for five years (Hedlund (2016a)). This is a compromise; in reality, foreclosure flags usually stay on a credit record for seven years, but their effect on mortgage credit availability diminishes over time.

Utility: The utility function is CES between consumption and housing, and CRRA over time:

$$u(c_t, h_t) = \frac{(\omega(c_t)^{\frac{\theta-1}{\theta}} + (1-\omega)(h_t)^{\frac{\theta-1}{\theta}})^{\frac{(1-\gamma)(\theta)}{\theta-1}}}{1-\gamma} \tag{11}$$

The coefficient of relative risk aversion, γ, is set to the standard value of 2. The elasticity of substitution between consumption and housing, θ, is set to .487 following Li et al. (2016). The weight on housing in the utility function ω is set so that 30% of renters' expenditures are on rent.

[33] This assumes a delinquent homeowner has two months' notice before the foreclosure auction, and so receives $\frac{2}{12} \approx .167$ times their income before the auction.

Income: Following much of the lifecycle literature, I assume that labor income follows a deterministic trend but is subject to transitory and permanent idiosyncratic shocks. Specifically,

$$\log(y_t) = g_t + z_t + \epsilon_t \tag{12}$$

where g_t is the deterministic component of income and ϵ_t is the transitory shock. The permanent component z_t follows the random walk,

$$z_t = z_{t-1} + \eta_t$$

ϵ_t and η_t are normal random variables with mean 0 and variances σ_ϵ^2 and σ_η^2, respectively. These variances have been estimated many times in the literature. I set them to .050625 and .003969, respectively, following Campbell and Cocco (2014). After retirement, agents are no longer subject to these shocks.

Guvenen et al. (2014) note that this standard lognormal income process cannot match the negative skewness and high kurtosis of income seen in the data. This is a serious concern for this paper, since large negative income shocks are precisely the ones likely to trigger default. One common approach in the mortgage literature is to explicitly introduce an unemployment shock (e.g. Laufer (forthcoming)). However, there are other large negative income shocks besides unemployment, like divorce and disability, so accounting only for unemployment will understate the probability of such a shock. Therefore, I include a persistent "disastrous" shock in the income process, similar in spirit to one in Cocco et al. (2005). This shock is calibrated to match the 3.56% probability in the PSID of a working-age family reporting at least a 50% drop in income.[34] The size of this shock is calibrated to match the fact that, empirically, the median household who reports such a shock reports a drop in log income of .96. Finally, I set the per period probability of escaping this state to 49.34%, which is the probability that a household reports at least a 50% increase in income following such a shock.

The income profile is taken as the empirical median. After retirement, households receive 86.5% of their pre-retirement permanent income, a number I estimate from the PSID.

Initial Distribution of State Variables: At the beginning of life, agents receive the nonhousing wealth, house price, and mortgage debt of a randomly-drawn 23-year old from the PSID. Agents who begin with a home of price p receive idiosyncratic flow value κp from it. All agents begin with the same permanent income. Values of the "disastrous" income shock are drawn from the stationary distribution. Renters begin with no default flag.

[34] In the PSID, the probability of experiencing such a shock seems almost constant over the working lifecycle, which is what I assume.

6.1.2 Second Stage

There are three remaining parameters to be estimated in the second stage: the discount factor β, the utility cost of defaulting on an underwater mortgage υ, and the variance of lognormal shocks to house prices, σ_p^2.[35] These parameters are set so that the model matches three key moments from the data.

First, I target a homeownership rate of 67%, which is roughly the rate estimated by the Census Bureau for 1998-2001.[36] This is largely determined by β.

Second, to generate sensible results the model must also match the proportion of mortgagors who have negative equity. I take this proportion from the 1998 and 2001 SCF, since these are the most recent years available before the boom in mortgage credit and house prices that preceded the foreclosure crisis. In these waves of the SCF, 2.5% of mortgagors report having an LTV greater than 100.[37] This moment helps identify the variance of the lognormal shocks to house prices σ_p^2.

Third, I target the default rate of underwater homeowners. Even underwater homeowners default at low rates (Foote et al. (2008), Bhutta et al. (2017)). Why is not well-understood. Foote et al. (2008) note that underwater mortgages have option value, since if house prices increase the homeowner may regain positive equity. In the model, the option value of an underwater mortgage is largely determined by σ_p^2, and so for low values of σ_p^2 the default rate of underwater homeowners is decreasing in σ_p^2. But this relationship is not monotonic; higher values of σ_p^2 generate a (counterfactually) high number of deeply underwater homeowners, who default at high rates because they are deeply underwater. Therefore the model also allows for an additional utility cost of default on an underwater mortgage, υ. This is most naturally interpreted as a moral or emotional aversion to underwater default, as argued by Bhutta et al. (2017) and documented by Fannie Mae (2010) and Guiso et al. (2013). It can also be interpreted as representing the uncertainty homeowners report about whether or not they can be sued for defaulting (Guiso et al. (2013)), or the potentially higher credit costs of defaulting when underwater due to a lower probability of a bank accepting a deed in lieu or short sale.

Foote et al. (2008) estimate roughly 6.4% of underwater homeowners lose their home to foreclosure within three years, implying an annual foreclosure probability of roughly 2.18%. However, their data is from Massachusetts, which is a recourse state, and so therefore this number is too low for the baseline model without recourse. Ghent and Kudlyak (2011) estimate that the effect of recourse on default probability is equivalent to the effect of an increase in LTV of roughly

[35] In the literature, σ_p^2 is typically set in the first stage of estimation, i.e. it is directly estimated using house price data. However, as shown in Section 4 a significant portion of the variance in measured house prices may come from measurement error or search frictions, which will tend to bias estimates of σ_p^2 upwards. Moreover, the model assumes homeowners maintain their homes, but empirical estimates of σ_p^2 include the effects of depreciation and upgrades.

[36] See `https://www.census.gov/housing/hvs/files/currenthvspress.pdf`.

[37] I take reported home values in the SCF to be actual home values, because correcting for the bias and noise in homeowner reports *decreases* the percent of homeowners estimated to be underwater. See footnote 22.

8.6, which in Foote et al. (2008) scales the foreclosure rate for underwater homeowners by a factor of roughly 1.16. Therefore, I target an annual foreclosure rate for underwater homeowners of $(2.18\%) * (1.16) = 2.53\%$.

Because this estimation strategy targets both the frequency and default rate of homeowners with negative equity, the model is essentially forced to generate the correct number of underwater foreclosures. The main test of the model is whether it generates the correct number of abovewater foreclosures, and of foreclosures generally.

This second stage of estimation produces a value for the discount factor $\beta = .948$, which is a fairly standard value for a lifecycle model. The variance of lognormal shocks to house prices σ_p^2 is estimated to be .0026. This is low for the literature, with estimates for σ_p^2 from transaction data ranging from roughly .01 to .025. However, as already noted, these estimates are biased upwards, since they do not account for the effects of search frictions or measurement error on measured transaction prices; as shown in Section 4, these effects can be substantial. Finally, the additional disutility of underwater default υ is estimated to be .07. For comparison, a nonhomeowner in the first period with the median income conditional on no "disastrous" income shock, no cash on hand, and no default flag would gain this level of utility if given \$51,181. This value is high, but well within standard bounds. For example, Laufer (forthcoming) introduces a disutility of moving in his model for a similar reason, and he estimates this disutility to be 54% of future consumption.

7 Model Results

Model Fit The estimated model is able to match targeted moments. Recall that the estimation targets a homeownership rate of 67%, an underwater rate of 2.5%, and an underwater foreclosure rate of 2.53%. The estimated model generates a homeownership rate of 66.4%, an underwater rate of 2.51%, and an underwater foreclosure rate of 2.49%.

Since the estimation strategy targets both the frequency and default rate of homeowners with negative equity, the model is essentially forced to generate the correct number of underwater foreclosures. However, the estimation does not target the overall default rate or the default rate of homeowners with positive equity. Therefore a major test of the model is whether it generates the correct number of abovewater foreclosures, and of foreclosures generally.

The model does well with this test. It generates an aggregate foreclosure rate of .45% . Pre-crisis foreclosure rates are not known with certainty, but for comparison Jeske et al. (2013) target a "long-run" foreclosure rate of .5%. 83% of of foreclosed homeowners in the model have positive equity.[38] Again, this compares well with the data; recall that foreclosure hazard rates estimated in

[38]In the model, the equity of a foreclosed property is taken to be the value of that property net of depreciation minus delinquent mortgage debt with interest and fees.

Section 4 suggested that roughly 81-87% had positive equity from 1998-2001. The model's success in replicating these untargeted moments shows that the rate of abovewater default seen in the data is consistent with economic theory.

Recourse The model has so far assumed that mortgages are "non-recourse", i.e. they are secured only by the house. However, many states ("recourse" states) allow lenders to seize other assets of underwater defaulters, though they differ considerably in how practical this is.[39]

Empirically, Clauretie (1987), Ghent and Kudlyak (2011), and Li and Oswald (2014) find that recourse has no effect on aggregate default rates. This may appear to contradict theory that predicts recourse discourages default (Quintin (2012), Campbell and Cocco (2014), Hatchondo et al. (2014), Li et al. (2014), and Corbae and Quintin (2015)). However, Quintin (2012), Hatchondo et al. (2014), and Corbae and Quintin (2015) note the potential for selection bias. Specifically, if recourse reduces individual default probabilities, it may allow less creditworthy borrowers to obtain a mortgage. These less creditworthy borrowers will be more likely to default, so the effect of recourse on aggregate default rates in these models is ambiguous. However, recourse lowers individual default probabilities in these models, conditional on loan and borrower characteristics.

Mitman (2012) argues that recourse may be ineffective for a different reason. In his model, which also allows for bankruptcy, recourse has almost no effect on default rates. This is because defaulters can declare bankrupty to discharge themselves of deficiency judgments. Hence in Mitman (2012) recourse does not lower default rates because it does not lower underwater default rates.

Empirically, Ghent and Kudlyak (2011) find that, conditional on loan and borrower characteristics, recourse does not lower default rates, even though it *does* lower the default rate of underwater homeowners. Similarly, Dobbie and Goldsmith-Pinkham (2015) find that recourse lowers the default rate of underwater homeowners, but not of abovewater homeowners.

I add recourse to my model to examine whether it can replicate these findings. I do this by assuming that, if homeowners default, lenders seize 50% of their other assets or the difference between the mortgage balance and the value of the home, whichever is less. I do not assume that lenders must pay litigation costs to seize assets or that defaulters can declare bankruptcy to discharge their debts to lenders.

The first result is standard. As expected, recourse discourages underwater homeowners from defaulting in the model; it lowers the underwater foreclosure rate from 2.49% to 1.85%. This echoes the theoretical findings of Quintin (2012), Hatchondo et al. (2014), and Corbae and Quintin (2015), as well as the empirical results in Ghent and Kudlyak (2011) and Dobbie and Goldsmith-Pinkham (2015), that recourse deters underwater default.

The second result is new. Because in the baseline model 83% of defaulters have positive equity,

[39] For a detailed discussion of recourse laws by state, see Ghent and Kudlyak (2011).

who are not affected by recourse, the policy has very little effect on the aggregate default rate in the model; it drops from .45% to .43% . Thus the model also matches the empirical results in Clauretie (1987), Ghent and Kudlyak (2011), and Li and Oswald (2014) that recourse does not lower default rates.

8 Conclusion

This paper provides the first formal estimates of the equity of foreclosed homeowners. These estimates suggest that large numbers of defaulters have positive equity. Since traditional frictionless models cannot match this evidence, the paper develops a quantiative model of mortgage default with search frictions. The model largely succeeds in matching the empirical relationship between equity and default, showing that this relationship is consistent with economic theory. The model is also able to replicate empirical evidence that recourse is effective at discouraging underwater default but ineffective at discouraging default in general.

Although the model shows that realistic income shocks together with search frictions can generate abovewater default rates seen in the data, it does not show that other factors do not help explain abovewater default. Other shocks, such as divorce or medical expenditure shocks, may be important. Informational frictions or behavioral factors may also play an important role. More research along these lines would be valuable.

References

Ambrose, Brent W. and Charles A. Capone (1998) "Modeling the Conditional Probability of Foreclosure in the Context of Single-Family Mortgage Default Resolutions," *Real Estate Economics*, Vol. 26(3), pp. 391–429.

Bajari, Patrick, Phoebe Chan, Dirk Krueger, and Daniel Miller (2013) "A Dynamic Model of Housing Demand: Estimation and Policy Implications," *International Economic Review*, Vol. 54(2), pp. 409–442.

Bhutta, Neil, Jane Dokko, and Hui Shan (2017) "Consumer Ruthlessness and Mortgage Default during the 2007 to 2009 Housing Bust," *Journal of Finance*.

Campbell, John Y. and João F. Cocco (2014) "A Model of Mortgage Default," *Working Paper*.

Carter, C.K. and R.J. Kohn (1994) "On Gibbs Sampling for State Space Models," *Biometrika*, Vol. 81(3), pp. 541–553.

Chatterjee, Satyajit and Burcu Eyingungor (2014) "A Quantitative Analysis of the US Housing and Mortgage Markets and the Foreclosure Crisis," *Working Paper*.

Clauretie, Terrence M. (1987) "The Impact of Interstate Foreclosure Cost Differences and the Value of Mortgages on Default Rates," *Real Estate Economics*, Vol. 15(3), pp. 152–167.

Cocco, João, Francisco J. Gomes, and Pascal J. Maenhout (2005) "Consumption and Portfolio Choice over the Life Cycle," *The Review of Financial Studies*, Vol. 18(2), pp. 491–533.

Corbae, Dean and Erwan Quintin (2015) "Leverage and the Foreclosure Crisis," *Journal of Political Economy*, Vol. 123 (1), pp. 1 – 65.

CoreLogic (2013) "Equity Report: Second Quarter 2013," Available at https://www.corelogic.com/research/negative-equity/corelogic-q2-2013-equity-report.pdf.

——— (2014) "Equity Report: Second Quarter 2014," Available at https://www.corelogic.com/research/negative-equity/corelogic-q2-2014-equity-report.pdf.

——— (2015) "Equity Report: Second Quarter 2015," Available at https://www.corelogic.com/research/negative-equity/corelogic-q2-2015-equity-report.pdf.

——— (2016) "Equity Report: Second Quarter 2016," Available at https://www.corelogic.com/research/negative-equity/corelogic-q2-2016-equity-report.pdf.

——— (2017) "Equity Report: First Quarter 2017," Available at https://www.corelogic.com/downloadable-docs/equity-report-q1-2017-20170608.pdf.

Davis, Morris A. and Erwan Quintin (2017) "On the Nature of Self-Assessed House Prices," *Real Estate Economics*, Vol. 45(3), pp. 628–649.

De Nardi, Mariacristina (2004) "Wealth Inequality and Intergenerational Links," *Review of Economic Studies*, Vol. 71, pp. 743–768.

Deng, Yongheng, John M. Quigley, and Robert Van Order (2000) "Mortgage Terminations, Heterogeneity, and the Excercise of Mortgage Options," *Econometrica*, Vol. 68(2), pp. 275–307.

Dobbie, Will and Paul Goldsmith-Pinkham (2015) "Debtor Protections and the Great Recession," *Working Paper*.

Duffie, Darrell and Kenneth J. Singleton (1993) "Simulated Moments Estimation of Markov Models of Asset Prices," *Econometrica*, Vol. 61(4), pp. 929–952.

Elul, Ron, Nicholas S. Souleles, Souphala Chomsisengphet, Dennis Glennon, and Robert Hunt (2010) "What "Triggers" Mortgage Default?" *The American Economic Review, Papers and Proceedings*, Vol. 100(2), pp. 490–494.

Fannie Mae (2010) "National Housing Survey, Third Quarter 2010: Key Findings," Available at http://www.fanniemae.com/resources/file/research/housingsurvey/pdf/Housing-Survey-Fact-Sheet-112310.pdf.

Foote, Chistopher, Kristopher Gerardi, Lorenz Goette, and Paul Willen (2010) "Reducing Foreclosures: No Easy Answers," in Daron Acegmoglu, Kenneth Rogoff, and Michael Woodford eds. *NBER Macroeconomics Annual 2009, Volume 24*: University of Chicago Press.

Foote, Christopher L., Kristopher Gerardi, and Paul. S. Willen (2008) "Negative Equity and Foreclosure: Theory and Evidence," *Journal of Urban Economics*, Vol. 64, pp. 234–245.

Fruhwirth-Schnatter, S. (1994) "Data Augmentation and Dynamic Linear Models," *Journal of Time Series Analysis*, Vol. 15(2), pp. 183–202.

Fuster, Andreas, Benedict Guttman-Kenney, and Andrew Haughwout (2016) "Tracking and Stress-Testing U.S. Household Leverage," *Working Paper*.

Garriga, Carlos and Aaron Hedlund (2017) "Mortgage Debt, Consumption, and Illiquid Housing Markets in the Great Recession," *Working Paper*.

Genesove, David and Lu Han (2012a) "Measuring the Thinness of Real Estate Markets," *Working Paper*.

——— (2012b) "Search and Matching in the Housing Market," *Journal of Urban Economics*, Vol. 72, pp. 31–45.

Gerardi, Kristopher, Kyle F. Herkenhoff, Lee E. Ohanian, and Paul S. Willen (2013) "Unemployment, Negative Equity, and Strategic Default," *Federal Reserve Bank of Atlanta Working Paper Series*.

Gerardi, Kristopher, Adam Hale Shapiro, and Paul S. Willen (2009) "Decomposing the Foreclosure Crisis: House Price Depreciation versus Bad Underwriting," *Federal Reserve Bank of Atlanta Working Paper*.

Ghent, Andra C. and Marianna Kudlyak (2011) "Recourse and Residential Mortgage Default: Evidence from US States," *The Review of Financial Studies*, Vol. 24(9), pp. 3139 – 3186.

Greenwald, Daniel L. (2017) "The Mortgage Credit Channel of Macroeconomic Transmission," *Working Paper*.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales (2013) "The Determinants of Attitudes Toward Strategic Default on Mortgages," *The Journal of Finance*, Vol. 68(4), pp. 1473–1515.

Guren, Adam M. (forthcoming) "House Price Momentum and Strategic Complementarity," *Journal of Political Economy*.

Guvenen, Fatih, Fatih Karahan, Serdar Ozkan, and Jae Song (2014) "What Do Data on Millions of U.S. Workers Reveal about Life-Cycle Earnings Risk?" *Working Paper*.

Han, Lu and William C. Strange (2015) "The Microstructure of Housing Markets: Search, Bargaining, and Brokerage," in J.V. Henderson G. Duranton and W.C. Strange eds. *Handbook of Regional and Urban Economics*.

Hatchondo, Juan Carlos, Leonardo Martinez, and Juan M. Sánchez (2014) "Mortgage Defaults and Prudential Regulations in a Standard Incomplete Markets Model," *Working Paper*.

Haughwout, Andrew F. and Ebiere Okah (2009) "Below the Line: Estimates of Negative Equity Among Nonprime Mortgage Borrowers," *FRBNY Economic Policy Review*.

Head, Allen, Huw Lloyd-Ellis, and Chenggang Zhou (2016) "Default, Mortgage Standards, and Housing Liquidity," *Working Paper*.

Heathcote, Jonathan, Kjetil Storesletten, and Giovanni L. Violante (2010) "The Macroeconomic Implications of Rising Wage Inequality in the United States," *Journal of Political Economy*, Vol. 118(4), pp. 681–722.

Hedlund, Aaron (2016a) " Illiquidity and its Discontents: Trading Delays and Foreclosures in the Housing Market," *Journal of Monetary Economics*, Vol. 83, p. 2016.

——— (2016b) " The Cyclical Dynamics of Illiquid Housing, Debt, and Foreclosures," *Quantitative Economics*, Vol. 7, pp. 289–328.

Herkenhoff, Kyle F. and Lee E. Ohanian (2015) "The Impact of Foreclosure Delay on U.S. Unemployment," *Working Paper*.

Jeske, Karsten, Dirk Krueger, and Kurt Mitman (2013) "Housing, Mortgage Bailout Guarantees and the Macro Economy," *Journal of Monetary Economics*, Vol. 60, pp. 917–935.

Kiel, Katherine A. and Jeffrey E. Zabel (1999) "The Accuracy of Owner-Provided House Values: The 1978 - 1991 American Housing Survey," *Real Estate Economics*, Vol. 27(2), pp. 263–298.

Korteweg, Arthur and Morten Sorensen (2016) "Estimating Loan-to-Value and Foreclosure Behavior," *Real Estate Economics*, Vol. 44, pp. 41–86.

Laufer, Steven (forthcoming) "Equity Extraction and Mortgage Default," *Review of Economic Dynamics*.

Li, Wenli, Haiyong Liu, Fang Yang, and Rui Yao (2016) "Housing over Time and over the Life Cycle: A Structural Estimation," *International Economic Review*, Vol. 57(4), pp. 1237–1260.

Li, Wenli, Costas Meghir, and Florian Oswald (2014) "Consumer Bankruptcy and Mortgage Default," *Working Paper*.

Li, Wenli and Florian Oswald (2014) "Recourse and the Residential Mortgage Market: the Case of Nevada," *Working Paper*.

Mitman, Kurt (2012) "Macroeconomic Effects of Bankruptcy & Foreclosure Policies," *Working Paper*.

Ngai, Rachel and Kevin D. Sheedy (2017) "The Decision to Move House and Aggregate Housing-Market Dynamics," *Working Paper*.

Pakes, Ariel and David Pollard (1989) "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, Vol. 57(5), pp. 1027–1057.

Pennington-Cross, Anthony (2003) "Subprime & Prime Mortgages: Loss Distributions," *OFHEO Working Paper 03-1*.

——— (2006) "The Value of Foreclosed Property: House Prices, Foreclosure Laws, and Appraisals," *Journal of Real Estate Research*, Vol. 28(2), pp. 193–214.

Quintin, Erwan (2012) "More Punishment, Less Default?" *Annals of Finance*, Vol. 8, pp. 427–454.

Shiller, Robert (2008) "Understanding Recent Trends in House Prices and Homeownership," *Housing, Housing Finance and Monetary Policy, Jackson Hole Conference Series*, pp. 85–123.

9 Appendix

This appendix describes in more detail the Gibbs sampling procedure used to estimate the parameters of the empirical model. Recall that the procedure involves two steps. These two steps are described in turn.

9.1 Forward-Filtering, Backwards-Sampling Procedure

The procedure begins by estimating the unobserved permanent component of property value U_t for every property at every point in time. Because Equations 1 and 2 define a linear state space model for U_t, a Kalman filter can be run forward in time to combine priors of U_t with observations (transaction prices and self-reported prices) to produce posteriors for U_t. By itself, a Kalman filter would be inefficient, since observations of a property's value after time t should also inform estimates of its value at time t. Therefore, I implement the the Forward-Filtering Backwards-Sampling (FFBS) algorithm of Carter and Kohn (1994) and Fruhwirth-Schnatter (1994). The Forward Filtering step of this algorithm exploits past information about a property. The Backwards Sampling step of this algorithm exploits future information about a property. Each of these steps is described in turn.

9.1.1 Foward-Filtering Step

The "Forward-Filtering" step of the FFBS algorithm runs a Kalman filter forward in time to exploit past and contemperaneous signals on U.

There are two such signals: transaction prices and self-reported values. If these measurements occurred at the same time, one Kalman filter could process them both. However, they virtually never occur in the same quarter, so I implement two distinct Kalman filters, one for periods with a transaction, and another for periods with no transaction but a reported value.

Transactions First consider the Kalman filter for transactions.

$U_i(t)$ is the state variable, so the state transition equation is given by Equation 2.

$$\begin{bmatrix} U_{t+1} \end{bmatrix} = \begin{bmatrix} 1 \end{bmatrix} \begin{bmatrix} U_t \end{bmatrix} + \begin{bmatrix} 0 \end{bmatrix} \begin{bmatrix} X^P \beta^P \end{bmatrix} + \Omega_{TS}$$

The state transition matrix is $\begin{bmatrix} 1 \end{bmatrix}$, the state control matrix is $\begin{bmatrix} 0 \end{bmatrix}$, and the covariance matrix of Ω_{TS} – the state process covariance matrix – is $\begin{bmatrix} \sigma_U^2 \end{bmatrix}$.

For the measurement equation, combine Equations 1 and 3:

$$\begin{bmatrix} P_t \end{bmatrix} = \begin{bmatrix} 1 \end{bmatrix} \begin{bmatrix} U_t \end{bmatrix} + \begin{bmatrix} 1 \end{bmatrix} \begin{bmatrix} X^P \beta^P \end{bmatrix} + \Omega_{TM}$$

The observation matrix is $\begin{bmatrix} 1 \end{bmatrix}$, the observation control matrix is $\begin{bmatrix} 1 \end{bmatrix}$, and the measurement error covariance matrix is $\begin{bmatrix} \sigma_P^2 \end{bmatrix}$

Reports Now consider the Kalman filter for self-reported values.

Again, $U_i(t)$ is the state variable, so the state transition equation is given by Equation 2.

$$\begin{bmatrix} U_{t+1} \end{bmatrix} = \begin{bmatrix} 1 \end{bmatrix} \begin{bmatrix} U_t \end{bmatrix} + \begin{bmatrix} 0 \end{bmatrix} \begin{bmatrix} X^P \beta^P \end{bmatrix} + \Omega_{TS}$$

For the measurement equation, combine Equations 1 and 4:

$$\begin{bmatrix} R_t \end{bmatrix} = \begin{bmatrix} 1 \end{bmatrix} \begin{bmatrix} U_t \end{bmatrix} + \begin{bmatrix} 1 \end{bmatrix} \begin{bmatrix} X^P \beta^P + X^R \beta^R \end{bmatrix} + \Omega_{TM}$$

The observation matrix is $\begin{bmatrix} 1 \end{bmatrix}$, the observation control matrix is $\begin{bmatrix} 1 \end{bmatrix}$, and the measurement error covariance matrix is $\begin{bmatrix} \sigma_R^2 \end{bmatrix}$

9.1.2 Backwards Sampling Step

The "Backwards-Sampling" portion of the FFBS algorithm combines the results of the "Forward-Filtering" portion of the algorithm with *future* information about a property to more precisely estimate posteriors for U. Unlike the Kalman filter, the Backwards-Sampling procedure does not provide closed-form posteriors for U. Rather, it provides a means to sample from the posteriors for U, so they can be approximated by Monte Carlo.

The procedure starts with a draw from the posterior distribution for U obtained from the "Forward-Filtering' step for the fourth quarter of 2013. Since this is the last period in the data, there is no future information to exploit. Another Kalman filter uses this draw as an observation, and combines it with the posterior distributions for U obtained from the "Forward-Filtering' step for the third quarter of 2013 (which can be interpreted as the prior distributions for this Kalman filter.) This produces the posterior distributions for U for the third quarter of 2013, *conditional* on the draws for the fourth quarter of 2013. These posteriors are then used in the same way to produce conditional posteriors for U in the second quarter of 2013, which can then be used to produce conditional posteriors in the first quarter of 2013, etc.

More formally, in this step, the state variable U is thought of as having a mean $\overline{U}_t^{FF}$ and error variance Ω_{FF} given by the Kalman filter from the "Forward-Filtering" step. Therefore, in the language of the Kalman filter, the state transition equation is:

$$\begin{bmatrix} U_t \end{bmatrix} = \begin{bmatrix} 1 \end{bmatrix} \begin{bmatrix} \overline{U}_t^{FF} \end{bmatrix} + \Omega_{FF}$$

The "observation" of U at time t is the value drawn for U at time $t+1$, which therefore has measurement error σ_U^2. Hence the measurement equation is:

$$\begin{bmatrix} U_{t+1} \end{bmatrix} = \begin{bmatrix} 1 \end{bmatrix} \begin{bmatrix} U_t \end{bmatrix} + \Omega_{BS}$$

where the measurement error covariance matrix is $\begin{bmatrix} \sigma_U^2 \end{bmatrix}$.

9.2 Step 2

This section describes the Bayesian OLS regressions used to estimate the parameters of Equations 1, 2, 3 and 4, conditional on the U draws obtained by the FFBS algorithm.

For the house price regression, define

$$\tilde{P}_i(t) = P_i(t) - U_i(t)$$

Then from Equations 1 and 3

$$\tilde{P}_i(t) = X^P \beta^P + \epsilon_i^P(t)$$

which is the regression I use to recover estimates of β^P and σ_P^2.

For the U regression, rewrite Equation 2 as:

$$U_i(t) - U_i(t-1) = \epsilon_i^U(t)$$

Estimating the variance of this regression yields an estimate of σ_U^2.

For the reported values regression, define:

$$\tilde{R}_i(t) = R_i(t) - V_i(t)$$

Then Equations 1 and 4

$$\tilde{R}_i(t) = X^R \beta^R + \epsilon_i^R(t)$$

which is the regression I use to recover estimates of β^R and σ_R^2.