Selection, Leverage, and Default in the Mortgage Market

Arpit Gupta
New York University, Stern School of Business

Christopher Hansman
Imperial College London

April 16, 2021

Abstract

We ask whether the correlation between mortgage leverage and default is due to moral hazard (the causal effect of leverage) or adverse selection (ex ante risky borrowers choosing larger loans). We separate these information asymmetries using a natural experiment resulting from the contract structure of option adjustable-rate mortgages and unexpected 2008 divergence of indexes that determine rate adjustments. Our point estimates suggest that moral hazard is responsible for 40% of the correlation in our sample, while adverse selection explains 60%. We calibrate a simple model to show that leverage regulation must weigh default prevention against distortions due to adverse selection. (*JEL*: D14, G21, D82)

In the years since the financial crisis of 2008, mortgage leverage has come into focus as a critical element of financial stability.[1] This attention is largely due to the connection between leverage and default: when housing prices fall, highly leveraged homeowners are the least likely to meet their obligations. There are two natural explanations for this relationship. The first, sometimes called moral hazard, is a causal effect: leverage itself increases the probability of default.[2] The second is adverse selection: ex ante risky borrowers simply prefer high-leverage mortgages. Despite a substantial theoretical literature examining these two classical information asymmetries in credit markets, differentiating between them remains a major challenge for empirical work.[3]

In this paper we separate adverse selection from moral hazard in the mortgage market. To do so, we exploit a natural experiment generated by the unique contract structure of option adjustable-rate mortgages (option ARMs). These loans have interest rate adjustments tied to prespecified mortgage indexes, most commonly a LIBOR or Treasury rate. While the choice between LIBOR and Treasury was not salient prior to the financial crisis, the unexpected divergence of the two in 2008 caused borrowers who chose otherwise identical contracts to owe substantially different amounts ex post.

This mortgage index-driven variation in borrowers' balances allows us to (i) isolate moral hazard by comparing borrowers with the same initial leverage choices but different realized home equity, and (ii) identify adverse selection by comparing borrowers with different initial leverage choices but similar realized home equity. Our point estimates suggest that adverse selection is responsible for 60% of the baseline correlation between leverage and default in our sample, while the causal effect—moral hazard—explains the remaining 40%.[4]

Testing for the presence of adverse selection is necessary because it has fundamental consequences for the quantity and price of leverage in equilibrium (and hence for mortgage policy). The theoretical literature following Bester (1985) suggests that—with adverse selection—leverage must act as a screening device to differentiate riskier types. As a result, even minor changes in policy or market fundamentals may significantly reshape the set of contracts lenders offer, with implications that differ sharply from a world without selection. For instance, our analysis shows that a regulator who ignores adverse selection will overestimate the reduction in defaults generated by macroprudential restrictions on household leverage and underestimate the corresponding costs to borrowers, who face higher interest rates and choose smaller loans in equilibrium. Similarly, the causal relationship between leverage and default is of direct interest to policymakers, practitioners, and academics. Clear estimates of this elasticity are necessary for effective design

---

[1]See, for example, Geanakoplos and Pedersen (2012), Mian and Sufi (2015), Korinek and Simsek (2016), Geanakoplos (2019).

[2]This terminology is used in credit markets by, e.g., Adams, Einav, and Levin (2009).

[3]This empirical challenge is highlighted in, e.g., Chiappori and Salanié (2013).

[4]Of course, there is uncertainty in our estimates, and reasonable confidence intervals contain a range of potential decompositions of this baseline. However, all suggest a meaningful role for adverse selection.

of borrower relief policies, accurate pricing of mortgage bonds, debates over the root drivers of consumer default, and more.

Moral hazard arises in our context because lenders face limited access to effective recourse and/or underwater households may be unable to avoid default by selling or refinancing (e.g., Campbell and Cocco 2015). However, a large literature (e.g., Deng, Quigley, and Van Order 2000) has documented notable heterogeneity in the propensity to default across borrowers. Some "risky" borrowers default as soon as the home is worth less than the mortgage balance, while others do not default even when substantially underwater. Adverse selection arises when this heterogeneity is correlated with demand for leverage, that is, when those who are quickest to default (for unobservable reasons) are more willing to accept the inflated interest rates that come with high-leverage loans.[5]

Our empirical approach disentangles adverse selection from moral hazard by recognizing that the two give distinct empirical predictions on the relationship between leverage and default, as in Karlan and Zinman (2009). Adverse selection implies a correlation between default and a borrower's initial leverage choice—independent of the home equity a borrower actually faces. Conversely, moral hazard implies a positive correlation between default and a borrower's realized leverage—regardless of the initial choice.[6] As a result, it is possible to separately identify the two if there is exogenous variation in borrowers' leverage that is distinct from their initial mortgage choice.

In our context, the divergence of mortgage indexes generates exactly this sort of variation. This is due to two features of option ARMs, which have fixed minimum payments in the early years of the loan but interest rates that update monthly. Because monthly payments do not adjust, fluctuations in interest rates generate meaningful variation in the balances borrowers will ultimately owe without affecting the amount they must pay in the short term.[7] As a result, otherwise identical borrowers who receive different paths of interest rates because of the loan's (i) origination month and (ii) particular mortgage index (e.g., LIBOR vs. Treasury) will face different realized leverage. The fixed and unusually small minimum payments for option ARMs facilitate our analysis by allowing us to set aside the role of monthly payments in the default decision.

To implement our approach, we instrument for borrower *i*'s realized leverage with the leave-out-mean

---

[5]This framing is analogous to a model of selection on ex post moral hazard, as in Einav, Finkelstein, Ryan, Schrimpf, and Cullen (2013).

[6]A natural way to understand these implications is through a set of ideal experiments. To identify adverse selection we would ideally reassign borrowers who have endogenously chosen different leverage to owe identical balances. A remaining correlation between the initial leverage choice and default would indicate selection. For moral hazard, the ideal experiment would randomly reassign borrowers who have chosen identical leverage to owe different balances. A correlation between these randomly assigned balances and default would then reflect a causal effect. Of course, this ideal experiment is stylized and ignores the fact that many factors will cause leverage to evolve after the initial choice in the real world.

[7]While a standard adjustable-rate product might adjust payments to account for interest accrual, for option ARMs any excess accrual is absorbed into the balance.

(jacknife) instrumental variables estimator—the average for all $j \neq i$ with the same index type and origination month.[8] To focus specifically on the interaction, we include origination month fixed effects, which account for any aggregate time trends, and index type fixed effects, which account for any fixed differences across borrowers who choose different indexes. We are also able to include originator $\times$ origination month fixed effects to address time-varying selection of borrowers across originators and unobserved changes in underwriting standards by different institutions, as well as a rich set of further controls. Throughout, the key assumption is that—conditional on these controls—our instrument captures exogenous variation in realized leverage that is distinct from the original leverage choice. If this is the case, our IV gives the causal effect directly, and the residual correlation between initial leverage and default allows us to capture adverse selection.

We find that adverse selection is responsible for 60% of the baseline correlation between leverage and default, while moral hazard is responsible for the remaining 40%. The latter effect suggests a substantial causal relationship between leverage and default. Our estimates imply that a 10-point reduction in a borrower's loan-to-value ratio two years after origination reduces the average probability of default for compliers in our setting by more than 4 percentage points. However, this average masks meaningful heterogeneity across the distribution of realized leverage. We estimate much stronger causal effects of leverage for borrowers very close to the threshold of negative equity, but weak or null effects for significantly underwater borrowers (consistent with Ganong and Noel 2020) or for borrowers with positive equity. Overall, this heterogeneity supports the view in which negative equity is a necessary but not sufficient condition for mortgage default. In robustness exercises we consider various potential concerns, including time-varying geographic confounds, early attrition from the sample, and the difficulty of accounting for selection in the presence of nonlinear causal effects.

As a final step, we develop a stylized structural model to highlight the consequences of our findings—particular adverse selection—for leverage regulation. We estimate a parametric control function version of our IV approach to provide inputs for the model, and analyze the introduction of a loan-to-value cap (a widely used policy tool intended to reduce mortgage defaults by restricting leverage). We find that a loan-to-value cap is indeed effective in limiting defaults, but to a lesser extent than a naive regulator who ignores the presence of adverse selection would expect. Furthermore, adverse selection generates unexpected welfare losses due to knock-on effects. While the mechanical impact of the regulation only forces the risky borrowers initially above the cap to take smaller loans, its effects ultimately propagate through the entire distribution. Safer borrowers initially below the cap also choose to take less leverage to maintain separation from riskier types and obtain suitably low interest rates. In equilibrium, interest rates rise across the whole loan-to-value

---

[8]This is analogous to instrumenting with a full set of index type $\times$ origination month dummies.

distribution and all borrowers choose smaller loans.

An important caveat to our work is its focus on a specific time period and mortgage contract. Our stylized model allows us to address the first by providing underlying parameters that can be used to simulate outcomes under different house price expectations and realizations. However, our unique setting restricts us to studying option ARM contracts. This is a large population, accounting for close to 10% of pre-crisis mortgage originations, but it is potentially distinctive. While we show that option ARM borrowers are similar on many observables to borrowers with other ARMs, they are different in terms of several characteristics from those with fixed rate mortgages, the most dominant contract in the U.S. market. Furthermore, there is always potential for meaningful selection on expectations or other unobservables.

This paper makes three contributions. We are the first, to our knowledge, to empirically isolate the presence of adverse selection on leverage in mortgage markets as distinct from moral hazard. This contributes to the growing empirical literature on asymmetric information and credit.[9] The key innovation in our setting comes in isolating a form of ex post variation in leverage that is unknown to borrowers when selecting contracts. A number of influential papers attempt to distinguish between adverse selection and moral hazard by exploiting ex ante variation—experimental, regulatory, or institutional—in the set of or shape of contracts offered. These include Ausubel (1999) and Agarwal, Chomsisengphet, and Liu (2010) on the U.S. credit card market; Adams, Einav, and Levin (2009) and Einav, Jenkins, and Levin (2012, 2013) on subprime auto loans; Hertzberg, Liberman, and Paravisini (2018) on online consumer credit; and Dobbie and Skiba (2013) on payday lending.[10] However, these approaches require strong assumptions as to why the relevant variation in ex ante contracts does not also generate selection on unobservables. Furthermore, we do so in the largest and arguably most important consumer debt market in the United States.

Second, our paper contributes to the literature on mortgage default by providing evidence for a strong impact of leverage. We do so by developing an instrument for borrowers' balances that does not simultaneously affect short-term payments. A large literature in this area has used nonexperimental variation in borrower equity, most commonly driven by house price variation. Vandell (1995) provides an overview of early research on the role of home equity in the choice to exercise the default option. More recent work, including that of Bajari, Chu, and Park (2008), Foote, Gerardi, and Willen (2008), Elul, Souleles, Chomsiseng-phet, Glennon, and Hunt (2010), Bhutta, Shan, and Dokko (2010), and Gerardi et al. (2017), has stressed

---

[9]We are also heavily indebted to broader empirical work on asymmetric information in insurance and other markets. Particularly Chiappori and Salanié (2000), Cardon and Hendel (2001), Finkelstein and Poterba (2004), Finkelstein and McGarry (2006), Finkelstein and Poterba (2014), Hendren (2013), as well as recent work examining the welfare implications of information asymmetries such as Einav, Finkelstein, and Cullen (2010), Einav et al. (2013), and Einav, Finkelstein, and Schrimpf (2010).

[10]Other considerations of selection in mortgage markets include Edelberg (2004), who uses structural assumptions to test for adverse selection and moral hazard in a broad class of consumer debts, including mortgages. Ambrose, Conklin, and Yoshida (2015) and Jiang, Nelson, and Vytlacil (2014) consider selection into and within low documentation mortgages. There is also related work in the home equity lending market, in particular Agarwal et al. (2011) and Agarwal, Chomsisengphet, and Liu (2016).

the joint importance of triggers such as liquidity and job loss alongside home equity in mortgage default. Palmer (2015) decomposes variation in defaults across cohorts of subprime mortgages into a portion due to home equity and a portion due to relaxed lending standards. Our borrower-level variation in leverage avoids the potential for measurement error and geographically based endogeneity concerns inherent to the use of home price variation. A small number of recent well-identified papers have found evidence for a limited or nonexistent causal role of borrower leverage on default. For example, Ganong and Noel (2020) use experimental evidence from the Home Affordable Mortgage Program (HAMP) based on a sample of loan applicants for loan modifications. Our findings are consistent with this work when considering borrowers in negative equity: we also find little evidence that changes in leverage affect default. However, a key contribution of our work is to incorporate borrower responses around the negative equity threshold, where we find strong effects.

Our third and final contribution comes in connecting work on information asymmetries to broader debates on the role of regulations on household leverage. As noted in Cerutti, Claessens, and Laeven (2017), macroprudential caps (along the dimensions of loan-to-value or debt-to-income) are in place in more than 41 countries as of 2014. A growing literature explores the ways in which this prudential regulation affects credit markets and evaluates their effectiveness. These include studies by Greenwald (2018), Gete and Reher (2016), DeFusco, Johnson, and Mondragon (2020), and Behn, Haselmann, and Vig (2017). We provide the first empirically grounded evidence highlighting the complications that adverse selection poses for regulations of this form. By highlighting the importance of considering selection, we also contribute to work on the role of mortgage leverage in the crisis and the macroeconomy more generally (e.g., Corbae and Quintin 2015; Geanakoplos 2010).

1 Background and Data

1.1 Background on ARMs

Traditional adjustable-rate mortgage contracts, typically referred to as hybrid ARMs, feature fixed interest rates and fixed payments for a set initial period—usually five or seven years. Subsequently, interest rates usually change annually or semiannually. New interest rates are calculated as the sum of a fixed component (the *margin*) and a variable component (the *index*). Monthly payments are designed to be fully amortizing—that is, calculated to pay off the loan over the full term at current interest rates. As a result, payments change to keep pace with interest rates and may unexpectedly increase if interest rates rise.

According to lenders, the potential danger of these unexpected payment increases motivated the creation

of the option ARM.[11] Banks wanted a product that incorporated floating interest rates while protecting borrowers from sharp payment increases and mortgage holders from associated default risk. The option ARM is characterized by a series of features that reflect this desire:

(i) *Fixed minimum payment schedule:* Borrowers are offered a small initial monthly payment, often based on the fully amortizing payment for an extremely low "teaser" interest rate. For the first five years, this payment adjusts upward once yearly by a fixed amount, usually 7.5%. After five years, the minimum payment adjusts to the fully amortizing amount.[12]

(ii) *Monthly interest rate changes:* While interest rates for hybrid ARMs adjust annually or semiannually, option ARMs update much more frequently—usually monthly.

(iii) *Negative amortization:* The minimum payment required in a given month will often be lower than the amount of accrued interest. In these circumstances, option ARMs allow for negative amortization— that is, they allow the excess interest accrual to be incorporated into the balance. As a result, the loan balance will typically grow in the early years of the mortgage.

(iv) *Proposed payment options:* The name, option ARM, refers to a menu of payment options offered on borrowers' monthly statements. In addition to the minimum payment, statements offer the possibility of an interest-only payment—covering the entirety of the interest accrual—along with amortizing payments calculated according to 15- and 30-year schedules. These possibilities are suggestions. Only the minimum payment is binding, and the borrower may in principle make any payment between the options or in excess of the 15-year amortizing payment (sometimes subject to certain caps). In practice, most borrowers make the minimum payment every month.

For the purposes of our identification strategy, (i) and (ii) are key. Because payments are fixed for the first five years, borrowers' balances change as a function of realized interest rates.[13] As a result, realized leverage depends on the particular mortgage index a loan is tied to and the specific origination month.

In the years leading up to the crisis, option ARMs became a significant fraction of the market, representing approximately 9% of originations in 2006.[14] As the crisis hit, borrowers with option ARMs defaulted

---

[11]See Golden West's history of the option ARM, available at http://www.goldenwestworld.com/wp-content/uploads/history-of-the-option-arm-and-structural-features-of-the-gw-option-arm3.pdf.

[12] In theory, 7.5% is a cap and the minimum payment might adjust by less if a 7.5% increase were to exceed the fully amortizing payment. In practice, the cap is nearly always binding. The schedule may also be interrupted if the loan balance rises above a fixed proportion of the original balance, often 110 or 125%. This was relatively rare in the sample we consider. By 24 months, only 88 loans out of the just under 500,000 used in our primary sample exceeded 125%. Roughly 1% (5,523) of borrowers had 24-month balances in excess of 110% and, of these, only 1,039 also had potentially binding negative amortization limits of 110% or lower. Our results are robust to excluding these borrowers.

[13]All analyses performed here consider outcomes within the first five years. Appendix Figure A.1 presents a sample balance and payment trajectory for an option ARM to highlight these product features from origination through that period.

[14]See the 2008 Mortgage Market Statistical Annual. In the 1980s and 1990s, the option ARM was primarily a niche product directed

at high rates. In the sample studied here, 50% of borrowers were seriously delinquent (60 days past due) on their mortgages at some point within the first five years. The combination of high default rates and nontraditional features made option ARMs a poster-child for excess in mortgage lending (see Amromin et al. 2018). Their role in the crisis has been highlighted by various media sources and policymakers—Ben Bernanke noted that "the availability of these alternative mortgage products proved to be quite important and, as many have recognized, is likely a key explanation of the housing bubble" (Bernanke 2010). Despite these criticisms, recent research has argued in favor of option ARM–style products, both suggesting that they approximate the optimal mortgage contract when borrowers have stochastic income (Piskorski and Tchistyi 2010)[15] and noting potential benefits from a macroprudential perspective (Campbell, Clara, and Cocco 2021).

1.2 Data

The data on option ARMs used in this paper are taken from a loan-level panel of privately securitized mortgages provided by Moody's Analytics representing over 90% of nonagency residential mortgage-backed securities (the same data were formerly provided by Blackbox Logic). These data provide detailed information about loans at origination, including borrower information, property characteristics, and contract terms. They also include dynamic information on monthly payments, loan balances, modifications, delinquency, and foreclosure. We limit our primary sample to the approximately 600,000 option ARMs originated between 2004 and 2007 with initial combined loan-to-value ratios between 50 and 100.

1.3 Summary statistics: Balance across indexes

Table 1 shows summary statistics, split across the different mortgage indexes that appear in our sample. Treasury and LIBOR are by far the dominant indexes, representing just over 80% of the sample and just over 15% of the sample, respectively. A small number of loans are also reported as tied to COFI (the 11th District Cost of Funds Index) or as having fixed interest rates in the initial period, with each category representing less than 2% of loans. While we include all in our baseline analysis, we also show robustness limiting our sample to LIBOR and Treasury indexed loans.

The characteristics of loans at origination are reasonably balanced across indexes. Origination leverage,

---

toward sophisticated borrowers. The flexibility of payments was intended to appeal to borrowers who expected their income to rise in the future or those with high income volatility. With the growth of a secondary market for nontraditional mortgages in the early 2000s, banks began to market option ARMs as affordability products, allowing borrowers to purchase more expensive homes than they would be able to afford with a traditional mortgage. Borrowers might take out such loans with the intention of refinancing the mortgage or selling the home after several years and thus never making payments much above the initial minimum.

[15]See also Guren, Krishnamurthy, and McQuade 2020, who confirm the findings in Piskorski and Tchistyi (2010), but also suggest that the option ARM may perform particularly poorly in crises.

measured here as the combined-loan-to-value (origination CLTV) to account for all liens, is close to 0.8 for all categories, averaging 0.80 for Treasury and 0.82 for LIBOR. Similarly, origination LTV, which does not consider other liens, is 0.77 on average for both LIBOR and Treasury. This is slightly larger than the average for conforming loans purchased by Fannie Mae or Freddie Mac but below the average LTV for subprime adjustable-rate mortgages.[16] The value of the home at origination is also similar for LIBOR, Treasury, and COFI at just below $500,000 (although larger for the fixed rate category). Note that the median loan is much closer to $400,000 across all categories, as the distribution is skewed due to a small number of very expensive properties.

In general, borrowers have high FICO credit scores, with an average of 706 for Treasury and 717 for LIBOR. These FICO scores, combined with average home values, suggest that option ARMs attracted a set of relatively high-credit-quality borrowers. While Treasury loans are slightly more concentrated in California, the overall geographic patterns are similar across states. There is some difference across indexes in the timing of origination, with a higher proportion of Treasury loans originated in 2005 and 2006. One peculiarity distinguishing option ARMs from conforming loans is the rarity of income verification. Given borrowers with high credit scores, low required monthly payments, protections against payment increases, and generally favorable expectations about housing prices, lenders appear to have been unconcerned ex ante about borrowers' ability to meet monthly obligations. This led to a prevalence of low or no documentation loans: 82% for Treasury and 85% for LIBOR. For these loans, borrowers provided little or no formal evidence of sufficient income to meet monthly payments, often simply stating income with no verification. We provide further comparisons of our option ARM sample against borrowers in other markets in the next subsection.

1.4   Comparison of option ARMs with other mortgages

Though option ARMs appear balanced across indexes, they may differ from other categories of mortgages. To explore this possibility, Appendix Table A.1 shows summary statistics for a series of other mortgage types: (i) more standard (5/1) ARM contracts, (ii) private label fixed rate mortgages, which are often non-conforming, and (iii) fixed rate mortgages from Fannie Mae's single-family loan performance data, which is made up of conforming loans.[17]

On two key observable characteristics—initial leverage and credit score—option ARM borrowers appear quite similar to both ARM and conforming fixed rate borrowers. However, there are a few characteristics that display more notable differences. Perhaps the most prominent of these is the origination value. Average home values for 5/1 ARMs are somewhat higher than those for option ARMs, at just over $600,000 versus

---

[16]The average original LTVs for Fannie Mae and Freddie Mac in 2007 were 72 and 71, respectively, according to Frame, Lehnert, and Prescott (2008).

[17]Our 5/1 ARM and private-label FRM samples are drawn from the same database we use for our option ARMs.

just under $500,000. Private label fixed rate loans represent somewhat less valuable homes: under $400,000 on average. Notably, the loans in the Fannie Mae sample are significantly smaller, at just under $260,000. We also observe large differences on the dimension of no/low documentation. Between 79% and 82% of option ARMs in our sample lack full documentation. For ARM borrowers, the average is closer to 65%, and for private label fixed rate loans closer to 40 percent. Of course, it is feasible that a lack of documentation could exacerbate the adverse selection problem. To further explore this possibility, we conduct robustness checks across documentation status in Section 3. The geographic concentration of option ARMs is also different from both categories of fixed rate mortgages, although fairly similar to the 5/1 ARM sample. Both Option and 5/1 ARMs are heavily concentrated in California (just under 50%), while only 19% of the private label fixed rate sample and under 10% of the Fannie Mae fixed rate sample are from California.

In general, the similarities and differences we find between those with option ARMs versus more traditional products mirror prior literature, such as Amromin et al. (2018), who have found that holders of such "complex mortgages" tend to be sophisticated, with high incomes and credit scores relative to the subprime population. While we are most comfortable extrapolating from our results to similar mortgage and borrower categories, and do not believe our sample captures the typical borrower with a conforming fixed rate loan, we do believe it is representative of a large and important class of borrowers during our period, at least on observables.

2 Empirical Strategy

In this section we define moral hazard and adverse selection in our context and clarify their distinct empirical predictions in the context of a simple empirical model. We then show how the model translates to a series of estimating equations and describe our IV strategy.

2.1 Definitions of adverse selection and moral hazard

The definitions of adverse selection and moral hazard that we specify follow largely from those used in Adams, Einav, and Levin (2009):

(i) *Moral hazard:* The mortgage market exhibits moral hazard if there is a causal relationship between the size of a borrower's loan liability and default. That is, all else equal, those who face higher leverage ex post default more frequently.

(ii) *Adverse selection:* The mortgage market exhibits adverse selection if unobservably risky borrowers— those who are more likely to default with contract terms held equal—select higher leverage contracts.

9

Defining adverse selection in this way is fairly standard and adheres closely to the discussion in Chiappori and Salanié (2013) on insurance markets. Adverse selection exists if there is an exogenous correlation between a borrower's demand for leverage and the unobservable credit risk he or she poses to the lender. Our definition of moral hazard is somewhat broader than usual, encompassing a range of possible motivations and mechanisms.[18]

There are two natural sources of moral hazard in a mortgage context. The first—a feature of many models of default—is the limited liability that is implicit in a mortgage contract. Lenders cannot effectively contract against borrowers strategically defaulting when their mortgages are underwater. Of course, legal restrictions specifying the degree of limited liability vary from state to state, but deficiency judgments are relatively rare in practice even in states with laws that are favorable to lenders (Pence 2006). The second is more mechanical. Because a borrower with negative equity may lose the ability to refinance or sell the home to avoid default, there may be a causal relationship between leverage and default even in the absence of a strategic motive.

2.2   An empirical model of leverage demand and default

At a given loan age t, a borrower i defaults if the difference between the loan balance (B_{it}) and the home price (H_{it}) exceeds some borrower- (and potentially time-) specific cost of default: $B_{it} - H_{it} > C_{it}$. C_{it} is the key source of heterogeneity in the model. This cost reflects the significant variability in borrowers' exercise of the default option noted by Deng, Quigley, and Van Order (2000) as well as a growing consensus that negative equity is a necessary but not sufficient condition for default (Bhutta, Shan, and Dokko 2010; Elul et al. 2010). Put simply, borrowers typically do not default until they owe substantially more on their mortgage than the home is worth. Note that there is no need for a behavioral explanation for this phenomenon. There are real costs associated with default, including credit score reductions, moving costs, and social stigma. These costs may vary in the population, reflecting, for instance, cross-sectional differences in the value of future access to credit or accessing local schools. Furthermore, even if the full population has a strong aversion to defaulting, variation in liquidity shocks may generate heterogeneity in realized default behavior that can be captured by C_{it}.

To follow the literature, we divide through by the home value to reframe this equation in terms of LTV

---

[18]Typically, a credit market can be said to exhibit moral hazard if (i) the expected returns to the lender depend on some noncontractable action of the borrower and (ii) that action is itself influenced by the terms of the loan contract. If default is considered a strategic choice, our definition aligns with this traditional notion because default itself can be thought of as the noncontractable action taken by the borrower. However, default may not be an active choice in certain circumstances. Borrowers may be insolvent or credit constrained to the extent that they are mechanically unable to make payments. Our definition does not rule out a causal effect caused by a mechanical relationship between the loan balance and default.

ratios:

$$\underbrace{LTV_{it}}_{B_{it}/H_{it}} > \underbrace{\widetilde{C}_{it}}_{1+C_{it}/H_{it}} \ .$$

In other words, borrower i defaults at age t if the LTV ratio exceeds some threshold $\widetilde{C}_{it}$. Rearranging, and decomposing this threshold into observable ($\mathbf{x_i}$) and unobservable (ε_{it}) components, we may rewrite the default rule as:[19]

$$D_{i,t+1} = \mathbb{1}\{\alpha LTV_{it} + \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_{it} > 0\} \tag{1}$$

where $D_{i,t+1} = 1$ if borrower i defaults between age t and $t+1$.

While the borrower's contract choice is the result of a complex maximization problem, we abstract from this and specify a linear demand model for leverage. Letting L_i represent the original leverage chosen by borrower i and $\mathbf{x_i}$ be the full set of observables that lenders are able to price on,

$$L_i = \mathbf{x}_i'\boldsymbol{\psi} + v_i. \tag{2}$$

Adverse selection arises when the unobserved heterogeneity in default risk (ε_{it}) is correlated with a borrower's demand for leverage when selecting a mortgage (captured by v_i). Put simply, if riskier borrowers—for example, those with unobservably low realized values of C_{it} on average—prefer higher leverage mortgages. In classic models of adverse selection, this correlation exists because borrowers have private information about their risk when choosing a loan—for example, if they have some private signal regarding the value of C_i.[20] A borrower who knows ex ante that he is relatively unlikely to repay will prefer to put less money down—that is, to choose a higher leverage loan—and will be willing to accept an increased interest rate to do so.

In Appendix B we outline a simple two-period version of such a model, and show that a Spence-Mirrlees single crossing condition holds: risky borrowers are relatively more willing to accept large balances (i.e., higher interest rates) in exchange for more leverage.[21] The equilibrium implications of this sort of selection are familiar from the literature on collateralized lending following Bester (1985) and more general work in the tradition of Rothschild and Stiglitz (1976). In competitive contexts, leverage acts as a screening device. Because safe borrowers value leverage less, they choose suboptimally low leverage (compared to the first best) to differentiate themselves from risky types and get a lower interest rate.

---

[19]We decompose $-\tilde{C}_{it} = \mathbf{x}_i'\frac{\beta}{\alpha} + \frac{1}{\alpha}\varepsilon_{it}$, where x_i is a set of observable characteristics available to the lender at origination, and $\frac{1}{\alpha}\varepsilon_{it}$ is the unobservable portion of the borrower's default costs.

[20]See, e.g., Brueckner (2000) and Bester (1985).

[21]In other words, borrowers with low costs of default place a higher value on the implicit hedge against home price reductions provided by the option to default. One way to think about this framework is as a model of selection on (ex post) moral hazard, as in Einav et al. (2013). Lenders cannot contract on the action of default, the willingness to take that action varies in the population, and borrowers are privately informed of their own willingness.

Within this framework, moral hazard and adverse selection have straightforward empirical predictions:

(i) *Moral hazard:* $\alpha > 0$ provides evidence of a moral hazard effect. α quantifies the causal impact of the borrower's current LTV on default. α can also be interpreted as (the square root of) the precision of unobserved default costs in the population.

(ii) *Adverse selection*: $\rho = Corr(v_i, \varepsilon_{it}) > 0$ provides evidence of adverse selection. Borrowers who choose higher than average L_{ij} based on unobservables (large v_i) are more likely to default, holding current leverage constant (large ε_{it}).

2.3 Challenges to identification

The basic challenge in separately identifying α and ρ—the parameters capturing adverse selection and moral hazard—is the mechanical relationship between L_i and LTV_{it}. In the absence of other differences, borrowers with identical L_i will tend to have identical LTV_{it}. For borrowers consistently making minimum payments, there are only two factors that might cause those with identical L_i to have different LTV_{it}: differences in home prices or differences in interest rates that lead to different balances.

Unfortunately, shocks to home prices may, in general, be correlated with ε_{it}. For example, a local labor market shock may influence both home prices and, separately, the borrower's probability of default. Additionally, because home prices can never be observed directly but rather must be inferred from the sale prices of surrounding homes, LTV_{it} is usually measured with error. Similarly, variation across time in interest rates is likely correlated with macro conditions, while cross-sectional variation potentially reflects endogenous contract choices. Isolating exogenous variation in LTV_{it} is nontrivial but necessary to accurately estimate the causal effect of leverage on default (α) as distinct from the initial contract choice (e.g., by comparing borrowers who have the same L_i but, for exogenous reasons, have different LTV_{it}). Similarly, exogenous variation in LTV_{it} is necessary to recover the correlation between initial leverage choice and default (ρ) is distinct from the causal effect (e.g., by comparing borrowers who have endogenously chosen different L_i but the have the same realized LTV_{it} for exogenous reasons). Put differently, exogenous variation allows us to directly estimate the causal effect of leverage, and hence attribute the residual correlation to adverse selection.

One concern with attributing the residual to adverse selection is the possibility of payment effects. Because minimum payments are typically amortizing (albeit at an artificially low teaser rate), there is, in principle, a relationship between minimum payments and initial leverage. If higher payments lead borrowers to default, we may attribute this causal channel as adverse selection. However, this concern appears to matter little in practice. Minimum payments are small, the correlation between payments and leverage is relatively

12

low (perhaps because there are many other sources of variation in payments, such as different teaser rates), and controlling for payments has effectively no impact on our estimates. We consider this in more detail when discussing our results.

2.4 A source of variation: Diverging mortgage indexes

To isolate plausibly exogenous variation in LTV_{it}, our identification strategy exploits the divergence of financial indexes used to determine interest rate adjustments for option ARMs. The difference between the two most common indexes, LIBOR and Treasury, is illustrated in Figure 1, which shows the levels of (panel A) and spread between (panel B) the one-year constant maturity Treasury (CMT) and one-year LIBOR. While there were fluctuations in the years preceding the crisis, the difference was contained in a relatively narrow band. However, in mid-2007, Treasury rates began to fall and the spread increased sharply, eventually peaking at over three percentage points in late 2008 following the Lehman Brothers bankruptcy filing and the AIG bailout.

But how do differences in interest rates cleanly translate into differences in current leverage (LTV_{it})? This is where the unique features of the option ARM come into play. Because minimum payments are set in the initial period for option ARMs, changes in interest rates have *no contemporaneous impact* on monthly obligations. As the mortgage must account for changes in the interest rate somehow, any additional interest accrual is incorporated directly into the balance.[22] This means that—for option ARMs—the divergence between mortgage indexes caused borrowers with otherwise identical loans to have sizable differences in leverage ex post.

Of course, a key question is the source of variation in indexes across borrowers. Prior to the crisis, borrowers had little reason to prefer one index to another. For example, although there tended to be a spread between LIBOR and Treasury rates—the spread between one-year CMT and one-year LIBOR was generally below 50 basis points—the two indexes moved quite closely together, and any fixed difference could be accounted for in the margin. Furthermore, Bucks and Pence (2008) suggest that borrowers tended to be uninformed about their contract terms. When asked what index their loan depended on, only 25% of borrowers responded with plausibly correct indexes. 30% of borrowers simply answered that they did not know.

If borrowers were unaware of the distinction between indexes, why did some end up with a Treasury index and others with LIBOR? Much—although not all—of the variation comes as a result of the lender, with most originators specializing in a particular index. To highlight this, Appendix Table A.2 shows the fraction

---

[22]This stands in contrast to more typical "hybrid" adjustable-rate mortgages, for which changes in the interest rate also change the monthly payment, which adjusts to ensure that the payments are fully amortizing.

of loans with Treasury indexes for the top 10 originators in our sample. Note that most tend to specialize in a particular index, and two of the top 10 make exclusively Treasury loans. As noted in Gupta (2019), differences across lenders are often a function not of the borrowers they lend to but their intentions on the secondary market. While the index is not exclusively determined by the originator (allowing us to include originator $\times$ origination month fixed effects), there is very limited variation within a month at a particular bank branch (or originator zip code). In other words, the specific bank location where the mortgage was originated likely determined the index.

While sorting between borrowers and originators—or other forms of selection into indexes on the basis of unobserved borrower characteristics—is certainly possible, random assignment to indexes is not necessary for our analysis. A critical feature of our approach is the fact that the impact of different indexes on leverage is not uniform across the sample period. Each origination month for each index generates a unique path of interest rates and a unique balance trajectory. Appendix Figure A.2 demonstrates this difference-in-difference variation in borrowers' balances. The plot shows the loan balance over time for four sample $100,000 loans: one LIBOR-indexed and one Treasury-indexed loan originated in January 2005, and one of each originated in January 2007.[23] Each of the four shows a distinct balance trajectory. Our empirical strategy focuses specifically on the variation in LTV_{it} particular to the interaction between a borrower's index and origination month. This allows us to control for any fixed differences between those with different indexes, as well as any aggregate time-specific effects.

2.5 Jackknife estimator to capture index $\times$ origination month variation

To isolate the variation in leverage driven by the interaction of index and origination month, we implement a jackknife estimator that instruments for a borrower's LTV_{it} with the leave-one-out average of LTV_{jt} for all other borrowers with the same index type and origination month. In particular, for borrower i with mortgage index $I(i)$ originated in month $m(i)$, we define the instrument:

$$Z_{it} = \frac{1}{n_{I(i) \times m(i)} - 1} \left[\left(\sum_{j=1}^{n_{I(i) \times m(i)}} LTV_{jt} \right) - LTV_{it} \right], \tag{3}$$

where $n_{I(i) \times m(i)}$ denotes the total number of loans originated in month $m(i)$ with index $I(i)$ active at age t.

Z_{it} itself captures all variation that operates at the index or origination month level, including any aggregate time-series variation in leverage, and any fixed differences between indexes. Crucially, our strategy isolates only the variation driven by the interaction $I(i) \times m(i)$ by further including both index and origina-

---

[23]These figures are based on simulated loans with a margin of 3.5 for both samples, based on the three-month LIBOR and 12-month MTA, respectively.

tion month fixed effects. This strategy mimics the use of a full set of $I(i) \times m(i)$ fixed effects while providing superior small sample estimation properties, as suggested in Kolesár (2013). Furthermore, this approach collapses a potentially high-dimensional set of instruments into a single Z_{it}. This allows us to both estimate a first-stage regression with interpretable coefficients and to provide evidence for the exogeneity of Z_{it} by testing for correlation with observables.

Appendix Figure A.3 provides insight into variation in Z_{it} 24 months after origination. For simplicity, we focus only on LIBOR and Treasury indexes. Panel (a) shows predicted LTV from first stage regressions (based on the most saturated version of Equation 5) for each month for loans indexed to LIBOR and Treasury. Panel (b) of this figure displays the underlying source of variation in these instruments: the most commonly used LIBOR (six month) and Treasury (MTA) indexes. Panel (c) displays the month-by-month magnitude of the difference in the averages shown in panel (a), while panel (d) displays the total number of originations by month for each index.

The differences in realized leverage implied by these instruments is nontrivial. The sample weighted average (absolute value) of the scaled monthly difference is roughly 1.25 LTV points, which is over $6,000 for the average loan. The maximum is roughly 5 LTV points.[24] Furthermore, the variation in the instrument is as expected given the patterns in the two mortgage indexes shown in panel (b). Recall that the variation we exploit is a function of the full *path* of differences in the mortgage indexes in the first two years after origination. Note that the gap between LIBOR and Treasury is consistently largest in the first portion of the sample, before closing to be smaller in magnitude and often negative in the latter part of the relevant period. Recall also that the margin above the index tends to be higher for LIBOR indexed loans. As such, it is natural that the scaled instrument is higher for LIBOR indexed loans in the first part of the sample, and higher for Treasury indexed loans in the latter part of the sample. Finally, while the most extreme differences in the values of the instrument are realized in 2004 and in 2007. While much of our sample is concentrated in 2005 and 2006, panel (d) shows that there are still a sizable number of loans that experience these extreme differences.

2.6 Estimating equations for linear specification

Our goal is to estimate our leverage choice equation (Equation 2) and default equation (Equation 1) while instrumenting for current LTV. We do so in two ways. First, we consider a reduced-form approach that collapses both into a single equation that can be estimated via standard two-stage least squares.[25] This

---

[24]The first-stage coefficient used to scale is 0.457, so the unscaled values of the instrument are more than double these values.

[25]To derive Equation 4 from Equations 1 and 2, we write $\varepsilon_{it} = \gamma v_i + e_{it}$, where $\gamma > 0$ holds if v_i and ε_{it} are positively correlated—that is, if there is adverse selection (in the normal case, $\gamma = \rho \frac{\sigma_\varepsilon}{\sigma_v}$). Replacing v_i using Equation 2 gives $\varepsilon_{it} = \gamma(L_i - \mathbf{x}_i'\boldsymbol{\psi}) + e_{it}$. Substituting for ε_{it} in Equation 1 gives Equation 4. In a slight abuse of notation, we take $\mathbf{x}_i$ in Equations 1 and 2 to include the fixed effects spelled

allows us to implement a thorough set of fixed effects and other controls in a computationally feasible manner, and to perform a wide array of robustness tests. Second, we consider a joint model that allows us to estimate these equations directly using a control function approach.

Our basic reduced-form approach considers the outcome of default between loan age t and $t + 1$ for the cross-section of borrowers active in our data at loan age t. In particular, we consider a linear probability model for default of the form

$$D_{it+1} = \alpha LTV_{it} + \gamma L_i + \mathbf{x}_i'\boldsymbol{\beta} + \zeta_{j(i)} + \lambda_{I(i)} + \theta_{o(i)} \times \mu_{m(i)} + e_{it}. \tag{4}$$

To isolate variation in leverage driven by the divergence of financial indexes, we instrument for LTV_{it} using the leave-one-out mean Z_{it} :

$$LTV_{it} = \delta Z_{it} + \eta L_i + \mathbf{x}_i'\boldsymbol{\pi} + \zeta_{j(i)} + \lambda_{I(i)} + \theta_{o(i)} \times \mu_{m(i)} + u_{it}. \tag{5}$$

These equations, which represent our most saturated specifications, include three sets of fixed effects: a zip code $j(i)$ fixed effect $\zeta_{j(i)}$, an index I(i) fixed effect $\lambda_{I(i)}$, and an originator×origination month fixed effect $\theta_i \times \mu_{m(i)}$. $\mathbf{x}_i$ denotes a rich set of borrower and loan controls that represent information available to the bank at origination. Because we estimate Equation 5 cross-sectionally at different loan ages, we do not include loan age effects (or time effects separate from origination month).

Our variables of interest are LTV_{it} and L_i, where α captures the causal effect of leverage on default and $\gamma > 0$ indicates the presence of adverse selection. We define L_i as the borrower's total leverage at origination, measured as the combined loan-to-value. We define LTV_{it} as the current loan-to-value on the option ARM at loan age t.[26] Our standard default measure is a borrower falling 60 days past due on monthly payments, although we also consider more severe measures of default in robustness tests. Our standard results consider default between 24 and 36 months, although we also consider a variety of other cross-sections in robustness tests. Throughout, we cluster standard errors at the originator-level. Our key assumptions to identify the causal effect are that Z_{it} is correlated with realized leverage ($Corr(Z_{it}, LTV_{it} \neq 0)$) and uncorrelated with the error in the default equation ($Corr(Z_{it}, e_{it}) = 0$).

In our estimation we gradually build up to these saturated specifications. At a minimum, we include index, origination month, and zip code fixed effects. This simple set of controls explicitly adjusts for any fixed

out in full in Equation 4.

[26]We consider total vs. loan specific leverage at origination, as lenders may account for second liens in pricing. While we would ideally do the same for current leverage, we are unable to observe the rate of amortization for second liens, and hence focus only on the option ARM first lien. Of course, this raises the possibility that our observed γ might capture a causal effect due to second liens, rather than a true selection effect. To rule this out, we perform robustness checks on the subsample with no second lien. We use Zillow's zip code–level home price index, available at http://www.zillow.com/research/data/, to impute current home values.

differences across indexes, as well as any aggregate time-varying effects, and therefore focuses on the variation in the instrument due to the interaction between index and origination month. We then add borrower and loan controls, which proxy for the set of observables available to the bank at origination. These include a flexible set of dummies for home value and FICO score, and indicators for loan occupancy, property type, and documentation status. Given the originate-to-distribute model at the time, soft information was a relatively minimal component of the origination process. As a result, these variables represent a reasonable proxy for the information used to price loans by originators. Effectively controlling for these characteristics is necessary to distinguish adverse selection from sorting on observables (which might occur, for example, because lenders offer observably different borrowers different prices for the same leverage). Finally, we add controls for originator $\times$ originator month to account for any time-varying originator-level confounds. This rules out, for example, threats due to time variation in the risk characteristics of the borrowers at different originators. This means that our saturated specifications focus on variation across loans with different indexes with a given month at a given originator.

As an aside, note that the logic behind using the instrument Z_{it} to recover the causal effect α is relatively standard. However, a natural question is whether using this instrument complicates our ability to recover γ. Specifically, one might worry that by considering only the small portion of plausibly exogenous variation in LTV_{it} that is captured by Z_{it}, we are likely attributing some portion of the causal effect to selection. If there is some component of LTV_{it} that is correlated with L_i but not captured by Z_{it}, will this load on L_i in our regressions? This would indeed be a concern in a "reduced-form" version of our approach in which LTV_{it} was replaced by Z_{it} in Equation (4) to "control" only for this portion of the variation in Z_{it}. However, this is not a concern in our full specification. Using the logic of two-stage least squares, this is because LTV_{it} is projected onto both Z_{it} and L_i in the first stage. As a result, any component of LTV_{it} that is not captured by the predicted $\widehat{LTV}_{it}$ used in the second stage will be orthogonal to L_i by construction.

2.7 Joint model

While the reduced-form model presented earlier has computational advantages, it also has a few downsides. In particular, it imposes an unattractive linearity assumption on the causal relationship between leverage and default—we would typically expect the marginal effect to depend on the level of leverage—and does not provide parameters that are directly relevant for our simulations in Section 4. To address both these issues, we complement our analysis by estimating a joint model of Equations 1 and 2 while including 5 to instrument for LTV_{it}. This gives us three equations (letting $\mathbf{x}_i$ include all fixed effects here, for economy of

notation):

$$D_{it+1} = \mathbb{1}\{\alpha LTV_{it} + \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_{it} > 0\}$$

$$L_i = \mathbf{x}_i'\boldsymbol{\psi} + v_i$$

$$LTV_{it} = \delta Z_{it} + \eta L_i + \mathbf{x}_i'\boldsymbol{\pi} + u_{it}.$$

Our estimation procedure takes a control function approach following Blundell and Powell (2004), incorporating an additional linear equation. To do so, we impose a simple parametric structure on the errors,

$$\begin{pmatrix} \varepsilon_{it} \\ v_i \\ u_{it} \end{pmatrix} \sim N \left(\begin{matrix} 0 \\ 0 \\ 0 \end{matrix} \,,\, \begin{bmatrix} 1 & & \\ \rho_{\varepsilon v}\sigma_v & \sigma_v^2 & \\ \rho_{\varepsilon u}\sigma_u & \rho_{vu}\sigma_v\sigma_u & \sigma_u^2 \end{bmatrix} \right),$$

which enables direct implementation via maximum likelihood. Again, we estimate cross-sectionally at a given t and hence make no assumption about the evolution of errors over time. The key intuition here is that $\mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_{it}$ captures the default threshold relative to LTV_{it} for individual i, with $\mathbf{x}_i'\boldsymbol{\beta}$ representing the observed component and ε_{it} the unobserved component. Estimating the parameters of the model therefore allows us to recover the distribution of the unobserved portion of the default threshold in the population (scaled relative to LTV):

$$\widetilde{C}_{it}|\mathbf{x}_i \sim N\left(-\mathbf{x}_i'\boldsymbol{\beta}, \frac{1}{\alpha^2} \right).$$

This distribution determines the strength of the moral hazard effect in the population.[27] Perhaps more importantly, we directly recover $\rho_{\varepsilon v}$, the correlation between ε_{ijt} and v_i. This correlation captures the strength of the adverse selection effect—the degree to which leverage choice ex ante is correlated with the unobserved portion of default costs. $\rho_{\varepsilon v}$ and the parameters of the distribution of $\widetilde{C}_{it}$ are the key elements of the simulations discussed in Section 4.

3 Results

The basic empirical fact our analysis seeks to explain is the positive correlation between leverage and default. Figure 2 displays this correlation in our sample, plotting the relationship between origination combined loan-to-value and default within the first five years of the loan. We see a strong and roughly linear upward slope, with defaults well over 50% for the most highly levered borrowers. In this section, we de-

---

[27]For example, a smaller α indicates a more dispersed distribution of thresholds, and hence a smaller marginal effect at any given point in the distribution.

scribe the central empirical results of the paper: a decomposition of this correlation into moral hazard and adverse selection.

3.1 Testing for correlation with observables

The key identifying assumption in our analysis is that the index $\times$ origination month specific variation captured by our leave-one-out mean instrument does not correlate with default except through its influence on leverage—or, more formally, that $Corr(Z_{it}, e_{it}) = 0$. While we cannot explicitly verify this assumption, we are able to test whether the particular variation we isolate in our IV strategy correlates with observables. To do so, we regress key borrower- and loan-level observables on our instrument, conditioning on our minimal set of fixed effects: index type, origination month, and zip code.

The results of these regressions, presented in Table 2, show a clear picture. First, our instrument is strongly and statistically significantly positively correlated with both current leverage and default (measured here at 24 months and between 24 and 36 months, respectively). This suggests a valid first-stage and a significant reduced-form effect of our instrument on default. However, we see no evidence of a correlation between our instrument and any observable characteristics of the borrower or loan. In particular, we see no significant relationship when considering a borrower's FICO score, origination leverage, original home value, documentation status, or investor status, nor when considering whether the loan was for a home purchase (vs. refinance), or for a single family home. We present the same information graphically in Figure 3, where we additionally normalize both our instruments and continuous observable variables by their standard deviations to provide more easily interpretable coefficients.[28] On the whole, these results provide evidence supporting the plausibility of our instrument exogeneity assumption.

3.2 Main results: Reduced-form model

We now turn to our primary specifications, which attempt to isolate the roles of adverse selection and moral hazard in the relationship between leverage and default following Equations 4 and 5. In our main estimates, shown in Table 3, we focus on the one-year default rate for borrowers 24 months after origination.[29] We cluster standard errors at the originator level throughout.

Table 3 (and many of our other tables) shows three types of specifications, labeled *Baseline*, *OLS*, and *IV*. The baseline specifications show the raw relationship between origination leverage L_i and default without considering the role of current leverage—an analog of Figure 2. Our ordinary least squares (OLS) specifi-

---

[28]Binary variables are not normalized.

[29]Twenty-four months represents a middle point between allowing sufficient time for leverage to diverge and not allowing so much time such that a significant fraction of borrowers have exited the sample. Appendix Table A.3 considers specifications aimed at capturing cross-sections at other loan ages.

cations additionally include LTV_{it}, and so match Equation 4 (with no instrument). Finally, our IV approach instruments for LTV_{it} using Z_{it}. Additionally, for each of these, we show results with increasingly rich controls. We begin with our simplest set, which controls only for index type, zip code, and origination month. We then add borrower and loan controls. Our most saturated specification additionally includes originator × origination month fixed effects.

Our baseline regressions confirm that the positive correlation between original leverage L_i and default—shown in Figure 2—persists when conditioning on controls available to the lender at the time of origination. Column (1), which includes our minimal set of controls, suggests that a 10-point increase in the leverage ratio at origination is associated with a just over 9-percentage-point increase in the one-year default rate. This rises slightly when including our richest set of controls (in column (3)). This specification is analogous to the positive correlation outlined in Chiappori and Salanié (2000) and indicates the presence of asymmetric information. Furthermore, the fact that the correlation rises when controls are included suggests that lenders may steer *observably* high-risk borrowers away from high leverage mortgage, through either pricing or other means.

The relationship between original leverage and default declines when accounting for current leverage via OLS. These specifications, which include LTV_{it} without incorporating an instrument, are shown in columns (4)–(6). These results suggest that a 10-point increase in the leverage ratio at origination is associated with a significant 6.7- to 8.2-percentage-point increase the one-year default rate, depending on the set of controls. We also find evidence of a significant relationship between *current* leverage and default in the OLS, indicating that a 10-point increase in current LTV is associated with a 2.8- to 3.0-percentage-point increase in default.

One way to quantify the relative importance of adverse selection is to compare the correlation between origination leverage and default when conditioning on *current* leverage to the unconditional baseline. While this approach is admittedly back-of-the-envelope—and implicitly assumes constant treatment effects when applied after our IV specifications—it provides a simple summary of the relative importance of our two effects of interest. By this metric, our OLS estimates in column (6) suggest that adverse selection is responsible for roughly 77% of the baseline correlation between leverage and default (0.818/1.066). This implies that the effect of the loan balance on default—moral hazard—is responsible for the remaining 23%.

In our IV specifications (7)–(9) we decompose the relationship between leverage and default while instrumenting for current LTV using our leave-one-out mean. The first stage, reported in panel B of this table, suggests instrument relevance. Higher leverage among borrowers with the same origination month and index type predicts higher current leverage—even conditional on origination month, index, geography, and other controls—because these borrowers are exposed to comparable interest rate trajectories that flow

through to their balances. Note that the coefficients are less than one, consistent with the leave-one-out mean representing a noisy proxy for true leverage. We find *F*-statistics for the first stage between 25–28, well above standard rules of thumb (Stock and Yogo, 2005).

Our IV results indicate the presence of moral hazard, as we find a significant and positive causal relationship between leverage and default. Our linear IV specification suggests that a 10-point increase in current LTV increases the probability of default by 4.7 percentage points. Note that this effect is roughly 62% higher than the effect implied by our OLS estimates. While there are a large number of potential omitted variables that might generate this bias, it is also useful to recall the presence of measurement error in the construction of LTV ratios, which creates a generic source of attenuation. Of course, our estimate should be interpreted as a LATE, and additionally represents the average for a population with a relatively high baseline level of leverage, so extrapolation of the precise magnitudes should be taken with care. The parameters of our joint model provide a more reasonable benchmark for other contexts. Still, this provides strong evidence for the existence of a causal effect.

Despite this meaningful causal effect, our results indicate that adverse selection is the dominant force behind the baseline correlation between leverage and default. This effect is captured by the coefficient on origination leverage (L_i). We estimate that a 10-point higher initial leverage ratio is associated with a 4.7- to 7.0-percentage-point increase in the one year default probability, depending on the specification. Point estimates from our richest specification (shown in column (9) of Table 3) suggest that adverse selection is responsible for just over 60% of the overall correlation between leverage and default. We attribute the remaining 40% to moral hazard.[30]

Naturally there is uncertainty in our estimates. We view the breakdown—60% adverse selection versus 40% moral hazard—as a benchmark rather than a final word. Indeed, taking the baseline correlation as given, 95% confidence intervals around the coefficient on initial leverage in column (9) contain parameter values that imply adverse selection is responsible for anywhere from 38% to 85% of the baseline. However, these results are estimated precisely enough to suggest a strong role for adverse selection even at the lower end of this range. We can reject the hypothesis of no adverse selection at any conventional level of statistical significance. Furthermore, while we find moderate variability in the point estimates across the different samples we consider in our robustness checks, we find consistently strong evidence for the presence of adverse selection.

---

[30]Once again, we arrive at this breakdown by comparing the correlation between origination leverage and default when conditioning on *current* leverage to the unconditional baseline (i.e., 0.655/1.066).

3.3 Nonlinearity in the causal effect

Our model imposes a strong linearity assumption on the causal relationship between leverage and default. While this is useful for determining the *presence* of a causal effect, a more nuanced view is necessary for several reasons. A first concern is our interpretation of the coefficient on origination leverage as evidence for adverse selection. If there are substantial nonlinearities, it is possible that what we term selection simply reflects a portion of the causal relationship that is not captured by our specification. To address this, columns (6)–(9) of Appendix Table A.4 repeat our analysis while controlling for a flexible polynomial in current leverage to account for any nonlinearities. We continue to find strong evidence for adverse selection with this more general specification. The magnitude of the coefficients on origination leverage in both our OLS and IV specifications are consistent with those in Table 3, although slightly less statistically significant in the case of the IV.

Perhaps more importantly, precisely understanding how the causal effect differs across the distribution of leverage is a key question from both an academic and policy perspective. For an intervention that forgives borrowers' balances to be successful, a strong marginal effect at the relevant level of leverage is more important than the size of the causal relationship on average. Furthermore, the shape of the causal relationship between leverage and default is informative in distinguishing two alternative models of borrower default. In the first, an unconstrained strategic model, leverage matters for default because borrowers simply weigh the financial benefits of repaying against the financial costs when deciding whether to default. In the second, high leverage matters because it exacerbates a liquidity constraint. Underwater borrowers may be unable to avoid default by refinancing or selling the home.

Figure 4 shows a nonparametric version of the unconditional relationship between leverage at 24 months and default in our data. The plotted line is upward sloping and appears approximately linear. However, because this figure does not account for selection or other confounds, it is unlikely to reflect the underlying causal relationship.

To provide insight into the potentially nonlinear causal relationship, we consider an expanded version of our IV specification. This specification replaces the linear term LTV_{it} with a set of dummies for each quartile of the distribution of current leverage:

$$
\begin{aligned}
D_{it+1} \quad = \quad & \alpha_1 \mathbb{1}\{79 < LTV_{it} < 95\} + \alpha_2 \mathbb{1}\{95 < LTV_{it} < 115\} + \alpha_3 \mathbb{1}\{115 < LTV_{it}\} \\
+ \quad & \gamma_1 \mathbb{1}\{75 < L_i < 81\} + \gamma_2 \mathbb{1}\{81 < L_i < 90\} + \gamma_3 \mathbb{1}\{90 < L_i\} \\
+ \quad & \mathbf{x}_i' \boldsymbol{\beta} + \zeta_{j(i)} + \lambda_{I(i)} + \theta_{o(i)} \times \mu_{m(i)} + e_{it}.
\end{aligned}
\tag{6}
$$

We instrument with dummies for quartiles of the distribution of the leave-one-out mean Z_{it}.

Results from our IV specifications, summarized in panel A of Figure 5, show evidence of a distinct nonlinearity. Specifically, we observe a sharp jump upward around the negative equity threshold. This figure plots the coefficients on dummy variables for each quartile of LTV_{it}, showing the discrete effect of being in different leverage quartile (relative to the first quartile). The estimated probability of default for the second quartile of leverage, representing current LTVs between 79 and 95, is effectively identical to that of the first quartile (below 79). This jumps up substantially for the third quartile, which represents current LTVs between 95 and 115. Our effects suggest that borrowers who are near the threshold of negative equity are more than 31.8 percentage points more likely to default relative to borrowers who have a current LTV of less than 79. By contrast, our estimates for the fourth quartile (31.6) are essentially the same as for the third quartile.

This pattern is consistent with so-called double-trigger models of mortgage default. The fact that the default probability jumps sharply around the threshold but remains well below 100%, suggests that negative equity appears to be a necessary but not sufficient condition for default. This indicates that negative equity leads to default only in combination with some other circumstance (e.g., a liquidity shock). The relatively flat relationship between the third and fourth quartiles suggests a limited strategic motive for default, at least for borrowers that are deeply underwater. Of course, quartiles are a somewhat rough way of capturing potentially complex nonlinearities. To display richer patterns in the data, Appendix Figure A.4 shows the nonparametric reduced-form relationship between our leave-out-mean instrument and default. The reduced form indicates a similar pattern. In particular, we see an increase in the default probability leading up to the negative equity threshold, a kink around the threshold, and a relatively flat relationship beyond that point.

Our specifications also include dummies for each quartile of origination LTV, to allow for nonlinear relationships on the selection dimension. These results are summarized in panel B of Figure 5, and show a close-to-linear relationship between initial leverage and default.

We present regression estimates corresponding to both figures in Appendix Table A.5, which also displays baseline and OLS versions, as well as specifications with various levels of controls. Interestingly, our OLS results do not indicate nearly as sharp of an upward jump around the negative equity threshold, highlighting the importance of our IV approach.

3.4 Differential attrition or prepayment before 24 months

A further potential concern is the possibility of nonrandom attrition. The sample in Table 3 is of loans that are active at 24 months. Our results may be biased if loans differentially exit our sample prior to 24 months (by either defaulting or prepaying) in a way that correlates with our instrument.

To address this possibility, we rerun our analysis on the full sample of loans, whether or not they exit the sample early. We consider cumulative defaults as an outcome—that is, default at any point up to time $t + 1$. Of course, we are unable to observe a borrower's loan balance after they exit the sample, and therefore do not see current leverage directly. However, because we observe all of the terms of the loan, we are able to mechanically infer the balance a borrower making minimum payments would have had at any point, and use this to construct a measure of current leverage. While this may not perfectly match actual leverage—a borrower's true balance may differ because they make an above-minimum payment or miss a payment—our use of this inferred leverage can be thought of in the spirit of an intent-to-treat effect. Our approach considers the impact of the leverage the borrower would have had sticking to standard monthly payments.

The results in Appendix Table A.3 use this inferred measure of leverage for all borrowers. We once again instrument with a leave-one-out mean—although this time we calculate the instrument using other borrowers' inferred leverage.[31] In columns (1)–(3) we repeat our main analysis. The basic patterns are similar: we find a positive and significant effect of current leverage in both our OLS and IV approaches, but find that adverse selection explains the majority of the correlation between leverage and default. However, note that our estimates suggest the role of selection is larger, and the causal effect smaller, than in our main specifications. This full-sample approach also allows us to look at a consistent sample at different cross-sections over the life of the loan. In columns (4)–(6) we consider default before 48 months, and in columns (7)–(9) we consider default before 60 months. We see consistent results in these later windows.

The slightly larger role of selection may simply reflect attenuation in the causal effect due to measurement error in inferred leverage. However, another concern is that this change reflects attrition due to prepayment. If borrowers with low initial leverage are more likely to prepay before 24 months (and therefore not default) for any reason, including these borrowers in the sample may lead the coefficient on initial leverage to increase. This issue is related to a broader set of concerns about how prepayment may influence or bias our results.[32]

---

[31]Doing so accounts, e.g., for endogenous loan terms such as the interest rate. Note also that this approach simultaneously addresses another potential concern, which is that differential payment patterns among borrowers with a particular index and origination month could generate variation in the instrument that is not driven by the path of interest rates.

[32]For example, to approximate the ideal experiment we describe to identify adverse selection, a borrower with initially lower leverage must receive a higher stream of interest rates. These higher interest rates may differentially increase the incentive to refinance (e.g. if the borrower is able to refinance from a high LIBOR to a lower Treasury rate). More generally, any borrower with a differentially high value of the mortgage index may simultaneously experience an increase in leverage and an increase in the incentive to refinance. Additionally, borrowers with low initial leverage may spend more time with positive equity, even if they ultimately end up with the same degree of negative equity as a borrower with high initial leverage. This mechanism might induce a negative correlation between

To address these potential issues, we consider two modified versions of the analysis in Table 3. The first, the results of which appear in panel A of Table A.6, considers the impact of initial and current leverage on the outcome of prepayment. To do so, we replace the dependent variable with an indicator equal to one for borrowers who prepay between 24 and 36 months. The second, which appears as panel B of Table A.6, exactly repeats the analysis in Table 3 but eliminates those who prepay from the analysis entirely by excluding any borrower that prepays between 24 and 36 months.

The most direct set of concerns regarding prepayment boil down to the possibility that (i) prepayment is negatively correlated with initial leverage and/or (ii) prepayment is positively correlated with a high value of instrumented current leverage. These concerns in turn suggest that omitting those who prepay between 24 and 36 months will cause the coefficient on initial leverage in our default regressions to fall and the coefficient on current leverage to rise.

None of these concerns appears to be a major factor in our context. The IV specifications presented in columns 7, 8, and 9 of panel A show a relationship between prepayment and initial leverage that is economically small and statistically insignificant (although it is negative in our most saturated specification, and larger in magnitude in our OLS regressions). Furthermore, the relationship between current leverage and default is actually negative across all columns and significant in our most saturated specification.[33] Similarly, the coefficient on initial leverage rises slightly (although not in a statistically meaningful way) in all specifications when omitting borrowers who prepay between 24 and 26 months, while the coefficient on current leverage falls marginally. In general, it appears that omitting those who prepay has little impact on our results.

3.5 Robustness

3.5.1 Heterogeneity by loan and borrower characteristics.
In Table 4 we explore heterogeneity in our results by three loan-level characteristics: the documentation status of the borrower, whether the loan is for a home purchase or refinance, and the recourse status of the loan's state.

Perhaps the most notable result in this table is in panel A, which shows a distinction between loans with no or low-income documentation versus those with full documentation. Specifically, the no/low documentation sample shows stronger adverse selection effects relative to our complete sample, while the full documentation sample shows a substantially weaker role for adverse selection.

There are at least three possible (and potentially overlapping) explanations for this difference. The first is that income—the key variable observable to lenders in the full documentation sample only—captures the

---

initial leverage and prepayment, and correspondingly bias our adverse selection effect upwards.

[33]This may simply reflect the fact that many borrowers who see increased leverage default rather than prepaying.

primary dimension of heterogeneity on which selection occurs. The second is that the two samples have distinct patterns of unobserved heterogeneity—for example, because there is substantially more heterogeneity in unobservable default risk in the no/low documentation sample or because this heterogeneity is more strongly correlated with initial leverage. The third is that the apparent difference is just an artifact made possible by our lack of power—that the full documentation sample is small enough that we are unable to rule out the possibility that the adverse selection effects are similar.

The first of these explanations suggests that adverse selection is easily addressed: by requiring income documentation, lenders might be able to eliminate the relevant unobserved heterogeneity. The remaining two alternatively highlight the need for care when extrapolating to other samples.

To distinguish between the first and other explanations, we would ideally be able to control directly for income in our specifications. Unfortunately, our data do not include direct measures of income (even for the sample that provides full documentation). However, we are able to match our data with income data provided by Equifax, a major credit bureau. These data comes with a caveat—which is that it relies on an Equifax prediction model, rather being sourced directly from consumers.[34] As a result, it likely has a degree of measurement error relative to true income.

In Appendix Table A.7 we show results from specifications that control for income or debt-to-income. Specifically, we show versions of our IV specification that include linear terms for either income or debt-to-income, as well as versions that include dummies for deciles of income or debt-to-income to incorporate potential nonlinear effects. We show these for our complete sample, for the no/low documentation sample, and for the sample with full documentation.

There are two primary takeaways from this exercise. The first is that the Equifax income measure captures meaningful variation in default risk. There is a significant negative relationship between income and default across all specifications and a significant positive relationship between debt-to-income and default. The second is that, despite this, controlling for income or debt-to-income has effectively no impact on our main results. This suggests that the difference in Table 4, panel A, is not driven by our first explanation— income as the key unobserved dimension of heterogeneity—and is instead driven by some combination of the second and third.[35]

While we cannot provide a complete illustration of why there might be distinct patterns of unobserved default risk across the two samples, we can note a few potential sources of differences. For one, there is a noticeable aggregate shift *away* from full documentation over our sample period. Full documentation loans represent over 30% of originations in our sample in 2004 but decline steadily to only 10% in 2007. Given

---

[34]To produce this measure, Equifax models income based on a subsample with employer-provided information.

[35]Of course, it is also possible that the Equifax measure does not capture an important unobservable component of income. However, the strong correlation between this measure of income and default provides reassurance against this possibility.

the lower average default rates for loans originated in the first part of our sample, we would expect smaller coefficients in a subpopulation that skews earlier. A complementary (more speculative) explanation has to do with the *types* of borrowers who opt for no/low documentation. Specifically, it may be that those who face higher costs of income verification (because they are self-employed or have other nonstandard income situations) are more likely to choose no/low documentation. If such borrowers have more variability in default risk (even conditional on a given level of income) or more private information about their future default risk, that could, in part, explain the difference.

Regardless of the source, the dominant role of no/low documentation borrowers reinforces the crucial caveat that our results may not be representative of the broader population of mortgage borrowers in which full documentation is the norm. However, this is not equivalent to stating that our results are only relevant for borrowers choosing low or no documentation in other contexts. Low or no documentation was the dominant option for option ARMs in the period we study, but became much less common for available products in the postcrisis period. It is likely that the types of borrowers who opted for low or no documentation in 2006 may choose full documentation in other circumstances.

3.5.2 The role of junior liens. A potential concern with our primary specifications is that our measure of origination leverage—the origination combined LTV—incorporates junior liens, while our measure of current LTV does not. This raises the possibility that the estimated effect of initial leverage is driven by the causal effect of the second lien, rather than adverse selection. To rule this out, columns (1)–(3) of Table A.8 repeat our analysis on the subsample of loans without junior liens. The baseline correlation between leverage and default in this sample in column (1) is nearly identical to that in our full sample, and the magnitudes of our OLS and IV estimates are comparable for both initial and current leverage (although the standard errors on current leverage are slightly larger, reducing our significance level to 10%).

3.5.3 Restricting to LIBOR and Treasury indexes. While our primary specifications incorporate four distinct indexes, two—LIBOR and Treasury—are dominant. In columns (4)–(6) of Table A.8, we show that our results are robust to limiting our sample to LIBOR and Treasury indexed loans. Baseline and OLS specifications are virtually identical to the full sample results shown in Table 3. Our IV estimates are statistically indistinguishable from those in the full sample, but indicate a slightly smaller role for adverse selection and a slightly larger role for moral hazard (on the order of 50% of the baseline correlation each).

3.5.4 Alternative measures of default. Columns (7)–(9) of Table A.8 show that our results are effectively unchanged when considering a more extreme definition of default: 90 days past due.

3.5.5 No geographic controls. The most minimal set of controls in Table 3 includes reasonably fine-grained geographic controls in the form of zip code–level fixed effects. However, the inclusion of these controls does not materially alter our main estimates. Columns (1)–(3) of Appendix Table A.4 show that our results are robust to excluding these fixed effects and conditioning only on index type and origination month.

3.5.6 Time-varying geographic controls. Even in the presence of zip code fixed effects, a potential concern is the presence of time-varying geographic confounds. For example, if an index is concentrated in a particular region, a local economic shock may generate variation in our instrument at the index×origination month level (because of a change in local home prices) while simultaneously affecting default. In columns (4)–(6) of Appendix Table A.4 we account for the presence of time-varying geographic effects by replacing our geographic fixed effects with a full set of MSA × origination month fixed effects. Our results are similar to those presented in Table 3.

3.5.7 Subsample of current loans. Our analysis in Table 3 conditions on loans that are active at 24 months. This includes loans that have been delinquent or in default prior to 24 months, so long as they have not gone into foreclosure. In Appendix Table A.9, we condition on loans that are current at 24 months and have not been delinquent previously. While the level of the baseline correlation is smaller, our estimates are otherwise in line with our main results.

3.5.8 The payment channel. One concern noted previously is that the coefficient on initial leverage might in part reflect the role of monthly payments. If higher leverage is associated with larger monthly payments and larger payments lead borrowers to default, the coefficient on initial leverage might reflect a payment channel rather than adverse selection. In practice, however, there is a relatively low correlation between payments and initial leverage. As a result, controlling for minimum payments has very little impact on our results (although there is a strong positive relationship between payments and default). For example, the coefficient on initial leverage in our most saturated IV specification is actually marginally higher when controlling for minimum payments (at 0.697 vs. 0.655) and still significant at the 1% level.[36]

3.5.9 Changes in adverse selection over time. A natural question is whether the degree of adverse selection changes over our sample period. To address this, we consider the following variation on the linear model presented in Equation 4:

$$D_{it+1} = \alpha LTV_{it} + \gamma_q L_i \times \text{Origination Quarter}_{q(i)} + \mathbf{x}'_i \boldsymbol{\beta} + \zeta_{j(i)} + \lambda_{I(i)} + \theta_{o(i)} \times \mu_{m(i)} + e_{it}.$$

---

[36]Complete results are available from the authors on request.

This specification modifies the original one by interacting our measure of initial leverage (L_i) with dummies for each origination quarter.

We present the results of this exercise in Figure A.5. Panel A shows the coefficients $\hat{\gamma}_q$ quarter by quarter from baseline, OLS, and IV specifications using the same sample as in Table 3 and our most saturated set of controls.[37] Panel B shows the average default rate quarter by quarter for the same sample. We see two primary takeaways. The first is a somewhat striking upward trajectory in the coefficient on initial leverage over the sample period. The coefficient is relatively small for loans originated earlier, rises monotonically for those originated in 2005, and roughly levels out for those originated in 2005 and 2006. The second is that this rise largely mirrors the corresponding rise in aggregate defaults over our sample period: Panel B similarly shows an upward trajectory. As we are looking at defaults between 24 and 36 months, this is due to a widespread rise in mortgage defaults across products, with loans originated in 2005 defaulting as the crisis began in 2007–2008. Our interpretation of the patterns shown in panel A is that the rise in the coefficient on original leverage simply reflects the aggregate rise in defaults. Whether due to selection or a causal effect, it is natural to expect the quantitative importance of initial leverage to grow as the total number of defaults rises. The relative importance of adverse selection in explaining the baseline correlation is much more stable over our sample period.

3.6 Joint model

The final step of our empirical analysis is to estimate a joint model of leverage demand alongside the default choice. Doing so has two primary advantages. First, it complements our investigation of nonlinearities by specifying a more realistic probit-style default equation. Second, it allows us to recover parameters that more directly relate to the model developed in the appendix and that can be used to inform the simulations developed in the next section. These parameters can also be used to simulate default probabilities in other counterfactual scenarios—for example, in a world with strong housing price growth.

Because of the increased computational complexity of this estimation, we slightly reduce the richness of included controls, for example, substituting zip code fixed effects with MSA fixed effects, and replacing originator $\times$ month fixed effects with originator $\times$ year fixed effects. Otherwise, we present our results as in Table 3: baseline estimates that exclude current leverage, OLS specifications, and finally our control function approach (the analog of our IV specifications).

The qualitative interpretation of these results, shown in Table 5, is identical to that in our linear specification. We both find strong evidence for adverse selection and a strong underlying causal effect. However,

---

[37]We collapse all of the relatively few loans originated prior to the first quarter of 2005 into a single dummy shown as the Q4 of 2004 in our figures.

rather than considering marginal effects in this context, we focus on the estimates themselves, which provide a series of directly interpretable parameters.

The presence of adverse selection is summarized by $\rho_{\varepsilon v}$, the correlation between the errors in the leverage choice and default equations. Across all specifications, we find a positive and significant value for $\rho_{\varepsilon v}$, with an estimated correlation of 0.12 in our most saturated IV specification. This suggests that the borrowers who are most likely to default—for ex ante unobservable reasons—choose the highest leverage.

Similarly, we are able to estimate the parameters of the distribution of unobserved default thresholds in the population: $\tilde{C}_{it}|\mathbf{x}_i \sim N\left(-\mathbf{x}_i'\boldsymbol{\beta}, \frac{1}{\alpha^2}\right)$. The key parameter is the standard deviation, measured in units of LTV points. Here, a larger value indicates a more dispersed distribution of thresholds and therefore a relatively small marginal effect at any given point. Our estimates of this parameter are analogous to those in the linear specification: we find larger estimates (lower marginal effects) in our OLS specifications, and smaller estimates (higher marginal effects) in our IV specifications. Still, across the board, we find a positive parameter that is statistically different from zero. In our most saturated IV specification, we estimate this standard deviation to be 0.43, or 43 LTV points. Given this, we are able to recover the distribution of thresholds for any set of observable characteristics under our normality assumption.

4 Simulations and Policy Analysis

In this section, we highlight the implications of our estimates for policy in a simulation framework. As an illustration we analyze the implementation of an LTV cap, which we view as a reasonably representative example of the types of macroprudential regulations implemented in various countries since the recession. In conjunction with concerns about asset price booms and busts, these policies are commonly motivated by the causal relationship, under the assumption that limiting consumer leverage directly reduces the scope for ex post default. In our analysis, we quantify the value of LTV caps in lowering default rates and identify the default externalities policymakers need to account for when designing such policies. The key innovation in our analysis, however, is the introduction of adverse selection. We argue that ignoring the role of adverse selection leads policymakers to (i) overestimate the reduction in defaults generated by a reduction in the LTV cap and (ii) underestimate the welfare loss generated because borrowers face higher interest rates and take smaller mortgages in equilibrium. To address the challenges of evaluating counterfactual policies in competitive markets with adverse selection, we use the equilibrium concept proposed by Azevedo and Gottlieb (2017).

4.1 A model to evaluate ex ante regulations

4.1.1 Equilibrium concept. Before describing the specifics of our model, we briefly discuss the equilibrium concept we consider. There is no clear consensus on the appropriate definition of equilibrium in competitive markets with adverse selection (Chiappori and Salanié, 2013). Furthermore, because equilibria often fail to exist under standard concepts, for example, Rothschild-Stiglitz, evaluating the counterfactual implications of policy can be difficult. However, a recent development by Azevedo and Gottlieb (2017) characterizes an equilibrium concept that is both robust—an equilibrium always exists—and straightforward to implement in a variety of applications. Equilibria of this form satisfy three requirements: (i) consumers optimize over the available set of contracts, (ii) lenders make zero profits on each contract, and (iii) there is free entry, in the sense that the equilibrium is robust to small perturbations, as defined formally in Azevedo and Gottlieb (2017).

A perturbation of a given economy is a version of that economy in which (i) a potentially continuous contract space is approximated by a large but finite set of contracts and (ii) a small measure of behavioral agents are introduced who always purchase each of the existing contracts and are costless to the lender (in other words, who repay the loan in all states of the world).[38] The presence of behavioral borrowers is key to avoiding unintuitive equilibria that arise if only the requirements of consumer optimization and zero profits are imposed. For example, such unintuitive equilibria may include contracts with arbitrarily high prices, which in turn are not chosen by optimizing consumers, and make zero profits only because they are not chosen.

For the purposes of our simulations, using this equilibrium concept is straightforward: we follow the practice of Azevedo and Gottlieb (2017) and calculate an equilibrium in a perturbation with 1% behavioral borrowers. Crucially, these behavioral agents should be considered an equilibrium selection device rather than a meaningful representation of some actual behavioral tendency in the population of mortgage borrowers. To emphasize this, we show robustness exercises in which we vary the fraction of behavioral borrowers. Our results are not sensitive to reasonable increases in this fraction and, most importantly, are effectively unchanged as we take the mass of behavioral agents further toward zero. We further discuss the implementation of our simulations after laying out details of the model.

4.1.2 Consumer preferences. We consider a two-period model. In the first, consumers purchase a home of fixed size and choose a mortgage. In the second, a stochastic ex post home value is realized. Borrowers then either default or repay the balance on the loan. We consider two types of borrowers: a large majority

---

[38]Formally, an equilibrium requires an economy to be the limit of a sequence of perturbations (in which the total mass of behavioral types converges to zero).

of *standard* borrowers, and a small mass of *behavioral* borrowers. We first discuss the objective for standard borrowers.

In the first period, borrowers face a menu of contracts that each have two elements: a loan size/leverage L_k and a balance $B(L_k)$. Given a contract $\{L_k, B(L_k)\}$ and a distribution of home prices, we characterize the observed portion of ex ante utility for standard borrowers based on the model in Appendix B:

$$U_i(L_k) = u(y_0 - (H_0 - L_k)) + \beta \left[\underbrace{\int_{\underline{h}}^{B(L_k)-C_i} u(y_1 - C_i)dF(H_1)}_{\text{Default}} + \underbrace{\int_{B(L_k)-C_i}^{\bar{h}} u(y_1 + H_1 - B(L_k))dF(H_1)}_{\text{Repayment}} \right].$$

As in the theoretical model, the only source of heterogeneity in U_i is C_i, the private costs of default. However, in practice, borrowers choose mortgages on the basis of a number of factors beyond just their default costs. Recall that the estimated correlation between the leverage choice and a borrower's private default costs was only 0.12. In a richly specified model, initial mortgage choice might also be a function of heterogeneity in each borrower's income, preferences (e.g., risk aversion or intertemporal elasticity of substitution), or period 0 knowledge of future C_i.

We abstract from these details and consider a simplified model in which standard borrowers' utility for a contract with a particular leverage choice is characterized by an observed portion, as defined earlier, and an independent, idiosyncratic error ϵ_{iL}:

$$V_i(L_k) = U_i(L_k) + \epsilon_{iL}.$$

This error captures, in a reduced-form way, all factors that influence borrowers with the same C_i to choose different contracts. When the variance of ϵ_{iL} is high, there is a weak relationship between C_i and the chosen L. When the variance is low, the correlation increases.

It is convenient to specify ϵ_{iL} to be type 1 extreme value, in which case a borrower's choice probability for a given L_k can be written as:

$$P_{ik} = \frac{e^{\gamma U_i(L_k)}}{\sum_{k'} e^{\gamma U_i(L_{k'})}},$$

where γ is a viscosity parameter determined by the variance of ϵ_{iL}. Of course, this specification imposes a standard independence of irrelevant alternatives (IIA) assumption, which may not hold in a more sophisticated model of heterogeneity across borrowers.

We do not explicitly model the utility of behavioral borrowers. The total mass of behavioral borrowers (in our main implementation, 1% of the population of standard borrowers) is split uniformly across the set of contracts (in our implementation, in which 50 contracts are available, this means a mass equal to 0.02%

of the population of standard borrowers is assigned to each contract). These behavioral borrowers choose the assigned contract with probability one and repay the loan in all states of the world.

4.1.3 Lender profits.

With the choice probabilities of standard borrowers in hand, computing lender profits is straightforward. We require that each lender specializes in a single unique leverage level L_k. We assume lenders are able to recover a fraction $\delta \leq 1$ of what the home is worth in the case of default. The expected profits of a lender selling contract $\{L_k, B(L_k)\}$ to standard borrower i with private default cost C_i are:

$$\pi^s(L_k, B(L_k); C_i) = -L_k + \frac{1}{1+r_f} \left[\underbrace{\int_{\underline{h}}^{B(L_k)-C_i} \delta H_1 dF(H_1)}_{\text{Default}} + \underbrace{\int_{B(L_k)-C_i}^{\overline{h}} B(L_k) dF(H_1)}_{\text{Repayment}} \right].$$

The profits of a lender selling contract $\{L_k, B(L_k)\}$ to a behavioral borrower are given by:

$$\pi^b(L_k, B(L_k)) = \frac{1}{1+r_f} B(L_k) - L_k.$$

Calculating total expected profits of a lender then involves summing two parts: (i) the profits for each individual i, multiplied by the probability that i chooses contract k, integrated over the distribution of C_i (specified here as $G(C)$), and (ii) the profits from behavioral borrowers. Denoting the mass of behavioral borrowers assigned to each contract by η, this gives:

$$\Pi_k = \int P_{ik} \pi(L_k, B(L_k); C_i) dG(C) + \eta \pi^b(L_k, B(L_k)).$$

4.1.4 Implementation.

We specify a fixed set of contracts (in the example presented, every integer LTV between 50 and 100). We then consider a mass of behavioral borrowers—uniformly distributed across contracts—equal to 1% of the population of standard borrowers, who always choose a given contract. We use a fixed point algorithm to determine equilibrium. In each iteration, consumers choose optimally taking prices as given, and interest rates are adjusted down or up for profitable or unprofitable contracts. Convergence is achieved when the absolute value of profits across all contracts falls below a predefined threshold. The existence of behavioral borrowers is crucial for convergence to intuitive equilibria. Because behavioral borrowers are costless, the interest rate on any contract that is only purchased by these types is reduced until either (i) a risky borrower is indifferent between the contract and his current choice or (ii) the interest rate reaches the risk-free rate. This rules out equilibria with contracts that have arbitrarily high prices and make zero profits only because they are not chosen.

4.1.5 Calibration. We calibrate three features of the simulation to the estimates from Table 5. We base these on column (9), our IV specification with our richest set of fixed effects. First we define the standard deviation of C_i based on our estimates.[39] Next, we choose γ, or equivalently the variance of ϵ_{iL}, so that the correlation between borrowers' choice of L and C_i in Regime I matches the estimated $\rho_{\varepsilon v}$. Finally, we set the mean of C_i by taking $-\mathbf{x}_i'\boldsymbol{\beta}$ using the our estimated β. We use average values of observed covariates (and the omitted base group for all categorical variables). All other parameters are set based on the data when possible and explicitly described in the bottom panel of Table 6. For the purposes of the simulation, we assume that borrowers have exponential utility, with CARA coefficient a.

4.2 Implications of an LTV Cap

We consider the implications of a decreased LTV cap, that is, a limit on the initial loan provided by lenders. This can be thought of as roughly the mirror image of a standard policy in insurance markets: a mandated minimum level of coverage. We evaluate three policy regimes:

(i) **LTV cap of 100**: In the first regime, lenders do not observe C_i, and equilibrium is as discussed earlier, with all loans making zero profits. The set of potential contracts contains all original LTVs between 50 and 100.

(ii) **LTV cap of 90 (no supply response)**: The second regime presents a naive view of the impact of an LTV cap of 90, ignoring the impacts of adverse selection. This regime evaluates the choices made by borrowers if an LTV cap of 90 were implemented but lenders did not otherwise adjust their contracts. As a result, lenders may make positive or negative profits under this regime.

(iii) **LTV cap of 90 (with supply response)**: The final regime considers the equilibrium allocation of credit when lenders are able to endogenously adjust contracts in response to a change in the LTV cap.

4.2.1 A naive evaluation of an LTV cap: no supply response. We first consider a comparison of Regimes I and II, which can be thought of as the anticipated response to an LTV cap for a naive policymaker. For these purposes, we consider a naive regulator to be one who understands borrower preferences and can anticipate the contracts borrowers will choose from any given set, but who disregards adverse selection. Such a policymaker believes that the proportion of defaults for a given contract does not depend on the population purchasing that contract, and hence that there will be no supply response to a change in the LTV cap. The intuition behind this comparison is demonstrated by the dark and light gray bars in Figure 6. This

---

[39]While our estimates are in LTV ratios and our simulations are specified in levels, we simply scale by the expected realization of home values, which is the same for all borrowers in our exercise.

figure shows results with an exaggerated degree of adverse selection, to better present the patterns across the three regimes, while Table 6 presents numbers based on simulations calibrated to the empirical results.

The dark gray bars illustrate the allocation of original LTV under Regime I and exhibit a basic pattern of adverse selection. While all borrowers would prefer initial loans with LTVs of 100 in a world with perfect information, the clustering of the riskiest borrowers raises the interest rate of a 100 LTV loan significantly. As a result, safe borrowers take smaller loans to distinguish themselves from risky types and avoid paying inflated interest rates. In other words, adverse selection leads to a partially separating equilibrium.

The expected impact of the regulation from the perspective of a naive regulator is illustrated with the light gray bars. Under the naive view, the only borrowers affected by the regulation are those initially choosing LTVs above 90. The borrowers who choose contracts with original LTVs below 90 in Regime I will continue to do so, while the majority of those choosing original LTVs above 90 will bunch close to the LTV cap.[40] Furthermore, the naive view will expect a significant reduction in defaults generated by the regulation. Because it assumes no heterogeneity across borrowers in default propensities, borrowers who choose an LTV of 90 under Regime II are expected to default at the same rate as borrowers choosing an LTV of 90 under Regime I.

Columns (1) and (2) of Table 6 compare Regimes I and II. There is indeed a reduction in loan size, from \$387,000 to \$361,000 and a corresponding expected reduction in average interest rates from 14.8% to 11.4%. Because Regime II does not allow lenders to change interest rates, this reduction is entirely the result of borrowers choosing smaller loans with lower rates. Furthermore, by failing to account for the inherent riskiness of the borrowers who are shifted from LTVs above 90 to LTVs below 90, the naive regulator expects the reduction in defaults to be significantly larger than it actually is, even without a supply response. The naive view suggests that an LTV cap of 90 would cut the fraction of defaults by more than 22%, from over 18% of borrowers to roughly 14%. Appropriately accounting for the risk of the borrowers initially allocated above 90 reveals the true reduction to be closer to 17.5%, with more than 15% of borrowers continuing to default.

4.2.2 Allowing a supply response. In addition to overstating the reduction in defaults generated by the regulation, the naive view understates the changes in interest rates and loan size generated by knock-on effects of the regulation. Reducing the LTV cap does indeed force some risky borrowers to decrease their LTV to 90. However, as a result, the interest rates on 90 LTV loans must also rise. Correspondingly, some borrowers who previously chose LTVs of 90 will choose slightly smaller loans, thereby leading lenders

---

[40]Because borrowers have a random component ϵ_{iL} of their preference for contracts, and because of the IIA assumption, borrowers who initially chose LTVs above 90 will not strictly choose contracts at 90. Rather, they will distribute their choices across remaining loans such that the relative choice probabilities are the same before and after the regulation.

to increase interest rates on those smaller loans and causing further knock-on effects. In the presence of adverse selection, leverage can be seen as a sorting device. Eliminating high LTV loans does not eliminate the incentive of borrowers to differentiate themselves, but instead forces them to do so over a smaller range of loans. The leftward shift of the white bars in Figure 6 relative to the light gray bars demonstrates the additional reduction in mortgage size due to knock-on effects.

In the calibrated simulations of Regime III, shown in the third column of Table 6, the knock-on effects cause *average* interest rates for *all* borrowers to rise by roughly 50 basis points (from 11.38 to 11.83). Furthermore, borrowers reduce their loan size by over $100 on average. Despite getting smaller loans, higher interest rates lead the average borrower to have a final balance that is roughly $1500 larger under Regime III than under Regime II.

In our simplified environment, which ignores asset pricing consequences, designing effective regulation involves balancing reductions in defaults against the welfare loss that results from borrowers facing higher interest rates and taking smaller loans. In the simulations provided here, a naive regulator overstates the reduction in defaults by close to 5% and underestimates the additional costs born by borrowers (not accounting for the fact that they take smaller loans) by $1,500 per borrower. While our model is highly stylized, the fundamental intuition is clear: when adverse selection is present, policymakers must weigh the benefits of preventing defaults against the knock-on effects that affect all borrowers when risky types are forced to choose new contracts.

4.2.3 Robustness. The existence of behavioral borrowers in our model is intended to rule out undesirable or unintuitive equilibria. The inclusion of these borrowers is not an attempt to capture some real underlying behavioral phenomenon. A natural question is the degree to which the level of behavioral borrowers we choose (1 percent) is innocuous, or if a greater or lesser level would substantially alter our results. To assess this, we re-simulated the model while varying the mass of behavioral borrowers (and holding the other parameters of the model fixed). Specifically, we considered behavioral masses equal to 0.1%, 1%, 10%, 100%, 1000%, 10000%, and 100000% percent of the mass of standard borrowers. For each of these, we simulated the model under all of our policy regimes.

Panel A of Appendix Figure A.6 summarizes these simulations by displaying the equilibrium average interest rate from our simulations for these different levels of behavioral borrowers across different policy regimes.[41] This exercise provides two insights. First, in the neighborhood of 1%—the level included in Table 6—varying the level of behavioral borrowers has a modest impact on results. While increasing the level of behavioral borrowers mechanically depresses interest rates, we do not observe sharp changes in equilibrium

---

[41]We show results for the interest rate, but similar patterns hold across all outcomes we consider. Results are available on request.

outcomes or other strange behavior. Reducing the mass of behavioral borrowers by an order of magnitude (from 1% to 0.1%) causes interest rates to increase only slightly: from 14.82 to 15.01 with an LTV cap of 100 and from 11.83 to 11.94 with an LTV cap of 90 (and a full supply response). Similarly, increasing the mass of borrowers to 10% of the mass of standard borrowers causes a relatively small decrease in interest rates. This suggests that the choice of 1% does not determine the qualitative conclusions of our model and that further reducing this number toward zero would not substantially alter the conclusions of the model.

Second, the exercise shows that an extremely large mass of behavioral borrowers will, of course, substantially alter the equilibria of the model. Because behavioral borrowers repay in all states of the world (and lenders must make zero profits in equilibrium), increasing the number of behavioral borrowers until they dominate the number of standard borrowers ultimately pushes the interest rate down to the risk-free rate. This is shown in the convergence of all three lines in Appendix Figure A.6 to the horizontal line representing the risk-free rate as the share of behavioral borrowers increases.

Panel B of Appendix Figure A.6 shows sensitivity of the interest rate to changes in ρ, the estimated parameter that captures adverse selection. For each value of this parameter, we show average equilibrium interest rates under the three policy regimes. We show these interest rates for our estimated $\hat{\rho}$ plus or minus one standard deviation. We also include an estimate with zero correlation (i.e., no adverse selection), as a benchmark.

As the figure shows, the impact of reasonable uncertainty around our parameter estimates is modest. The qualitative conclusions are not changed when varying ρ by a standard deviation in either direction: the average interest rate is lowest when ignoring the supply response to a 90 LTV cap, increases when the supply response is included, and highest with an LTV cap of 100. The qualitative results for other variables (e.g., defaults) are also similar.

5 Conclusion

In this paper, we empirically separate moral hazard from adverse selection in the mortgage market. To do so, we focus on a natural experiment generated by two features of option ARMs: fixed payments and variable interest rates. Because monthly payments do not change when interest rates rise or fall, fluctuations in market rates directly affect borrowers' balances. This creates a distinction between borrowers' initial leverage choices and the balances they owe ex post. To isolate plausibly exogenous variation in balances, we focus on the variation in interest rates that comes as the result of the financial index used to determine rate adjustments. Because of the unexpected divergence between index rates during the crisis, borrowers experienced substantially different balances as a function of the loan's index and origination month.

We isolate this variation in borrowers' balances using a leave-one-out jackknife IV estimator. This allows us to identify the causal effect of home equity on default—the moral hazard effect—and subsequently to back out the role of adverse selection. We find significant evidence of both information asymmetries. Adverse selection is responsible for 60% of the baseline correlation between leverage and default, while moral hazard is responsible for the remaining 40%.

We show that, in the presence of adverse selection, policies such as loan-to-value caps may have knock-on-effects that affect all borrowers in equilibrium. Such regulations will cause all borrowers to face higher interest rates and choose smaller loans. In general, regulators should be cognizant of the potential for distortions in mortgage markets that comes as a result of adverse selection.

References

Adams, W., L. Einav, and J. Levin. 2009. Liquidity constraints and imperfect information in subprime lending. *American Economic Review* 99:49–84.

Agarwal, S., B. W. Ambrose, S. Chomsisengphet, and C. Liu. 2011. The role of soft information in a dynamic contract setting: Evidence from the home equity credit market. *Journal of Money, Credit and Banking* 43:633–55.

Agarwal, S., S. Chomsisengphet, and C. Liu. 2010. The importance of adverse selection in the credit card market: Evidence from randomized trials of credit card solicitations. *Journal of Money, Credit and Banking* 42:743–54.

———. 2016. An empirical analysis of information asymmetry in home equity lending. *Journal of Financial Services Research* 49:101–19.

Ambrose, B. W., J. Conklin, and J. Yoshida. 2015. Reputation and exaggeration: Adverse selection and private information in the mortgage market. Working Paper.

Amromin, G., J. Huang, C. Sialm, and E. Zhong. 2018. Complex mortgages. *Review of Finance* 22:1975–2007.

Ausubel, L. M. 1999. Adverse selection in the credit card market. Working Paper.

Azevedo, E. M., and D. Gottlieb. 2017. Perfect competition in markets with adverse selection. *Econometrica* 85:67–105.

Bajari, P., C. S. Chu, and M. Park. 2008. An empirical model of subprime mortgage default from 2000 to 2007. Working Paper No. 14625, NBER.

Behn, M., R. Haselmann, and V. Vig. 2017. The limits of model-based regulation. SAFE working Paper no. 75.

Bernanke, B. S. 2010. Monetary policy and the housing bubble. Remarks by Chairman Ben Bernanke At the Annual Meeting of the American Economic Association, Atlanta, Georgia.

Bester, H. 1985. Screening vs. rationing in credit markets with imperfect information. *American Economic Review* 75:850–5.

Bhutta, N., H. Shan, and J. Dokko. 2010. The depth of negative equity and mortgage default decisions. Working Paper.

Blundell, R. W., and J. L. Powell. 2004. Endogeneity in semiparametric binary response models. *Review of Economic Studies* 71:655–79.

Brueckner, J. K. 2000. Mortgage default with asymmetric information. *Journal of Real Estate Finance and Economics* 20:251–74.

Bucks, B., and K. Pence. 2008. Do borrowers know their mortgage terms? *Journal of Urban Economics* 64:218–33.

Campbell, J. Y., N. Clara, and J. F. Cocco. 2021. Structuring mortgages for macroeconomic stability. *Journal of Finance, forthcoming* .

Campbell, J. Y., and J. F. Cocco. 2015. A model of mortgage default. *Journal of Finance* 70:1495–554.

Cardon, J. H., and I. Hendel. 2001. Asymmetric information in health insurance: Evidence from the national medical expenditure survey. *RAND Journal of Economics* 408–27.

Cerutti, E., S. Claessens, and L. Laeven. 2017. The use and effectiveness of macroprudential policies: New evidence. *Journal of Financial Stability* 28:203–24.

Chiappori, P.-A., and B. Salanié. 2000. Testing for asymmetric information in insurance markets. *Journal of Political Economy* 108:56–78.

———. 2013. Asymmetric information in insurance markets: Predictions and tests. In G. Dionne, ed., *Handbook of insurance*, 397–422. New York: Springer.

Corbae, D., and E. Quintin. 2015. Leverage and the foreclosure crisis. *Journal of Political Economy* 123:1–65.

DeFusco, A., S. Johnson, and J. Mondragon. 2020. Regulating household leverage. *Review of Economic Studies* 87:914–58.

Deng, Y., J. M. Quigley, and R. Van Order. 2000. Mortgage terminations, heterogeneity and the exercise of mortgage options. *Econometrica* 68:275–307.

Dobbie, W., and P. M. Skiba. 2013. Information asymmetries in consumer credit markets: Evidence from payday lending. *American Economic Journal: Applied Economics* 5:256–82.

Edelberg, W. 2004. Testing for adverse selection and moral hazard in consumer loan markets. FEDS Working Paper.

Einav, L., A. Finkelstein, and M. R. Cullen. 2010. Estimating welfare in insurance markets using variation in prices. *Quarterly Journal of Economics* 125:877–921.

Einav, L., A. Finkelstein, S. P. Ryan, P. Schrimpf, and M. R. Cullen. 2013. Selection on moral hazard in health insurance. *American Economic Review* 103:178–219.

Einav, L., A. Finkelstein, and P. Schrimpf. 2010. Optimal mandates and the welfare cost of asymmetric information: Evidence from the U.K. annuity market. *Econometrica* 78:1031–92.

Einav, L., M. Jenkins, and J. Levin. 2012. Contract pricing in consumer credit markets. *Econometrica* 80:1387–432.

———. 2013. The impact of credit scoring on consumer lending. *RAND Journal of Economics* 44:249–74.

Elul, R., N. S. Souleles, S. Chomsisengphet, D. Glennon, and R. Hunt. 2010. What "triggers" mortgage default? *American Economic Review* 490–94.

Finkelstein, A., and K. McGarry. 2006. Multiple dimensions of private information: Evidence from the long-term care insurance market. *American Economic Review* 96:938–58.

Finkelstein, A., and J. Poterba. 2004. Adverse selection in insurance markets: Policyholder evidence from the U.K. annuity market. *Journal of Political Economy* 112:183–208.

———. 2014. Testing for asymmetric information using unused observables in insurance markets: Evidence from the U.K. annuity market. *Journal of Risk and Insurance* 81:709–34. ISSN 1539-6975. doi:10.1111/jori.12030.

Foote, C. L., K. Gerardi, and P. S. Willen. 2008. Negative equity and foreclosure: Theory and evidence. *Journal of Urban Economics* 64:234–45.

Frame, S., A. Lehnert, and N. Prescott. 2008. A snapshot of mortgage conditions with an emphasis on subprime mortgage performance. Federal Reserve's Home Mortgage Initiatives, Federal Reserve Bank of Richmond.

Ganong, P., and P. Noel. 2020. Liquidity versus wealth in household debt obligations: Evidence from housing policy in the great recession. *American Economic Review* 110:3100–38.

Geanakoplos, J. 2010. The leverage cycle. *NBER Macroeconomics Annual* 24:1–66.

———. 2019. Leverage caused the 2007-09 crisis. Working Paper.

Geanakoplos, J., and L. H. Pedersen. 2012. Monitoring leverage. In *Risk Topography: Systemic Risk and Macro Modeling*, 113–27. Chicago: University of Chicago Press.

Gerardi, K., K. F. Herkenhoff, L. E. Ohanian, and P. S. Willen. 2017. Can't pay or won't pay? Unemployment, negative equity, and strategic default. *Review of Financial Studies* 31:1098–131.

Gete, P., and M. Reher. 2016. Two extensive margins of credit and loan-to-value policies. *Journal of Money, Credit, and Banking* 48:1397–438.

Greenwald, D. 2018. The mortgage credit channel of macroeconomic transmission. Working Paper.

Gupta, A. 2019. Foreclosure contagion and the neighborhood spillover effects of mortgage defaults. *Journal of Finance* 74:2249–301.

Guren, A. M., A. Krishnamurthy, and T. J. McQuade. 2020. Mortgage design in an equilibrium model of the housing market. *Journal of Finance* 76:113–68.

Hendren, N. 2013. Private information and insurance rejections. *Econometrica* 81:1713–62.

Hertzberg, A., A. Liberman, and D. Paravisini. 2018. Screening on loan terms: Evidence from maturity choice in consumer credit. *Review of Financial Studies* 31:3532–67.

Jiang, W., A. A. Nelson, and E. Vytlacil. 2014. Liar's loan? Effects of origination channel and information falsification on mortgage delinquency. *Review of Economics and Statistics* 96:1–18.

Karlan, D., and J. Zinman. 2009. Observing unobservables: Identifying information asymmetries with a consumer credit field experiment. *Econometrica* 77:1993–2008.

Kolesár, M. 2013. Estimation in an instrumental variables model with treatment effect heterogeneity. Working Paper.

Korinek, A., and A. Simsek. 2016. Liquidity trap and excessive leverage. *American Economic Review* 106:699–738.

Mian, A., and A. Sufi. 2015. *House of debt: How they (and you) caused the great recession, and how we can prevent it from happening again*. University of Chicago Press.

Palmer, C. 2015. Why did so many subprime borrowers default during the crisis: Loose credit or plummeting prices? Working Paper.

Pence, K. M. 2006. Foreclosing on opportunity: State laws and mortgage credit. *Review of Economics and Statistics* 88:177–82.

Piskorski, T., and A. Tchistyi. 2010. Optimal mortgage design. *Review of Financial Studies* 23:3098–140.

Rao, J., and G. Walsh. 2009. *Foreclosing a dream: State laws deprive homeowners of basic protections*. National Consumer Law Center.

Rothschild, M., and J. Stiglitz. 1976. Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. *Quarterly Journal of Economics* 90:629–49.

Stock, J., and M. Yogo. 2005. Testing for weak instruments in linear IV regression. In D. W. Andrews, ed., *Identification and Inference for Econometric Models*, chap. 10, 80–108. New York: Cambridge University Press.

Vandell, K. D. 1995. How ruthless is mortgage default? A review and synthesis of the evidence. *Journal of Housing Research* 6:245–64.

A. Divergence in LIBOR and Treasury rates
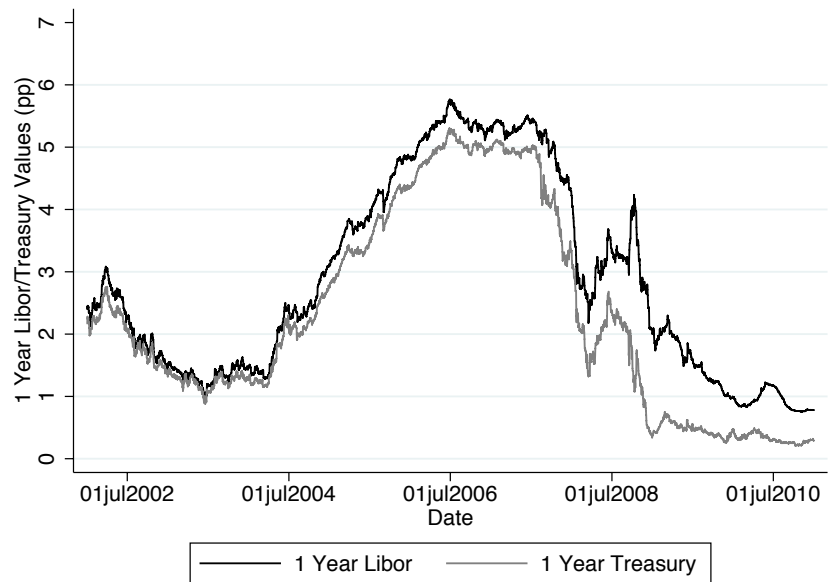
B. Spread between LIBOR and Treasury increased dramatically during crisis
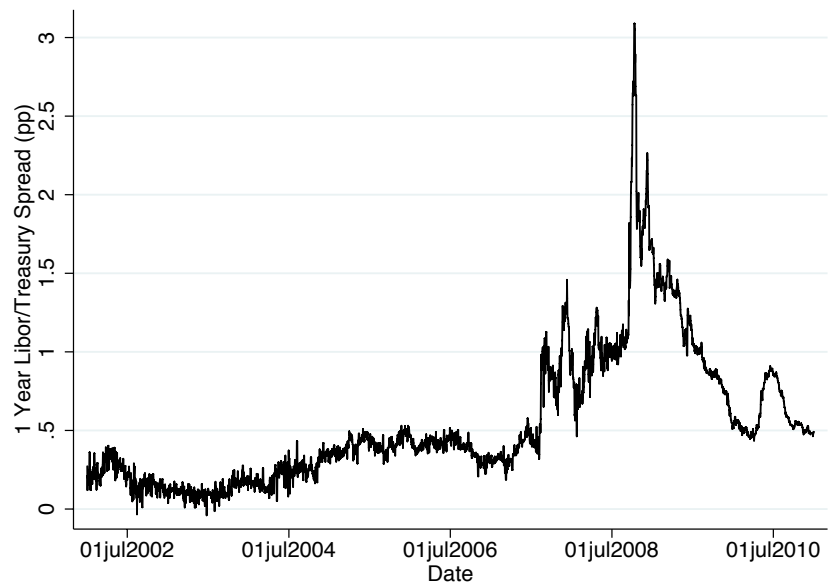
Figure 1
Comparison of mortgage indexes

Panel A shows one-year LIBOR and one-year constant maturity Treasury (CMT) rates between 2002 and 2010. Panel B plots the spread between the two rates.
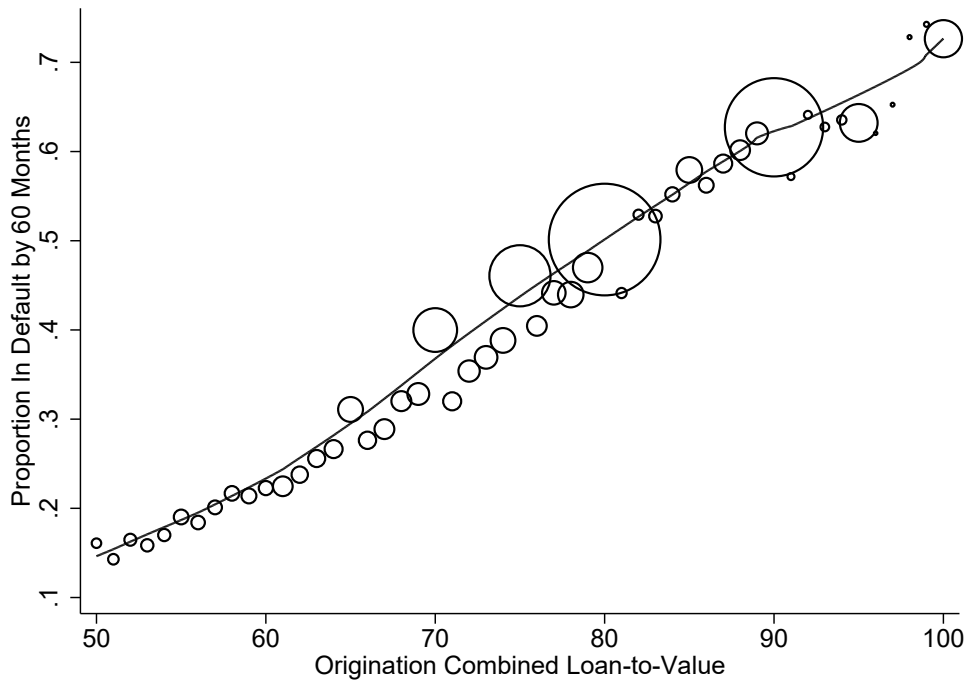
Figure 2
Original CLTV is positively correlated with default within 60 months

Hollow dots show the average proportion of loans defaulting within 60 months for each one-point bin of origination combined loan-to-value. Size of dots is proportional to number of borrowers within each bin. Default is defined as 60 or more days past due. The solid line shows a local linear smoothing of the raw data.
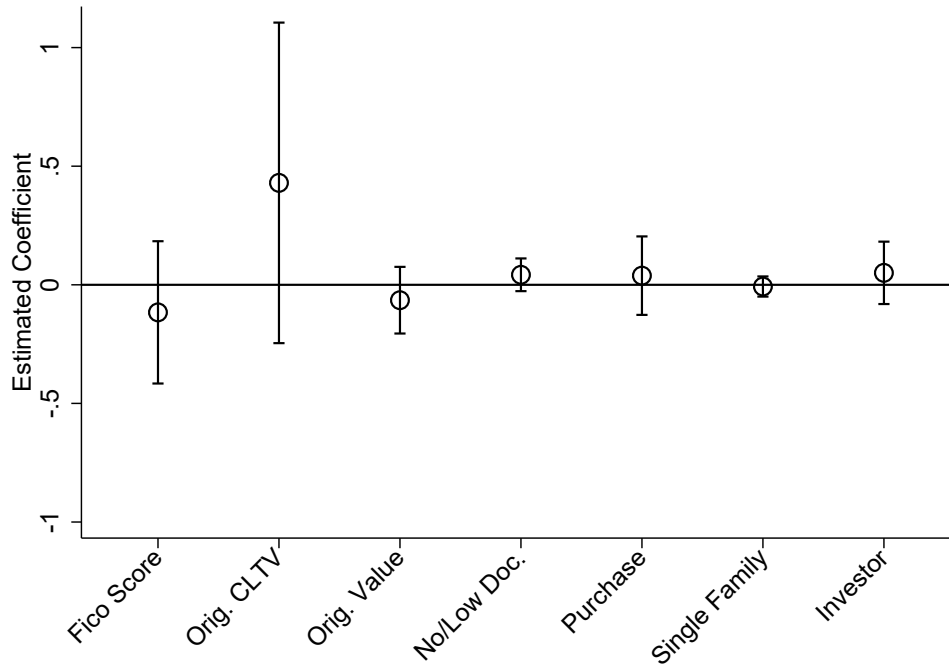
Figure 3
Instrument uncorrelated with observables at origination

This figure shows coefficients from a regression of our leave-out-mean instrument Z_{it} on a range of borrower-level and loan-level covariates. We control throughout for our minimal set of controls. origination month, index type, and zip code. We cluster standard errors at the originator level. The first three variables—FICO score, origination LTV, and origination home value—are also normalized by dividing by their standard deviations to provide interpretable coefficients. The last four variables are dummy variables indicating the presence of no or low documentation in loan origination, whether the mortgage is for a home purchase instead of a refinancing, whether the property is single family, and whether the borrower was a reported investor. Across all categories, we find no statistically significant relationship between our instrument and borrower characteristics. Regression results are also displayed in Table 2.
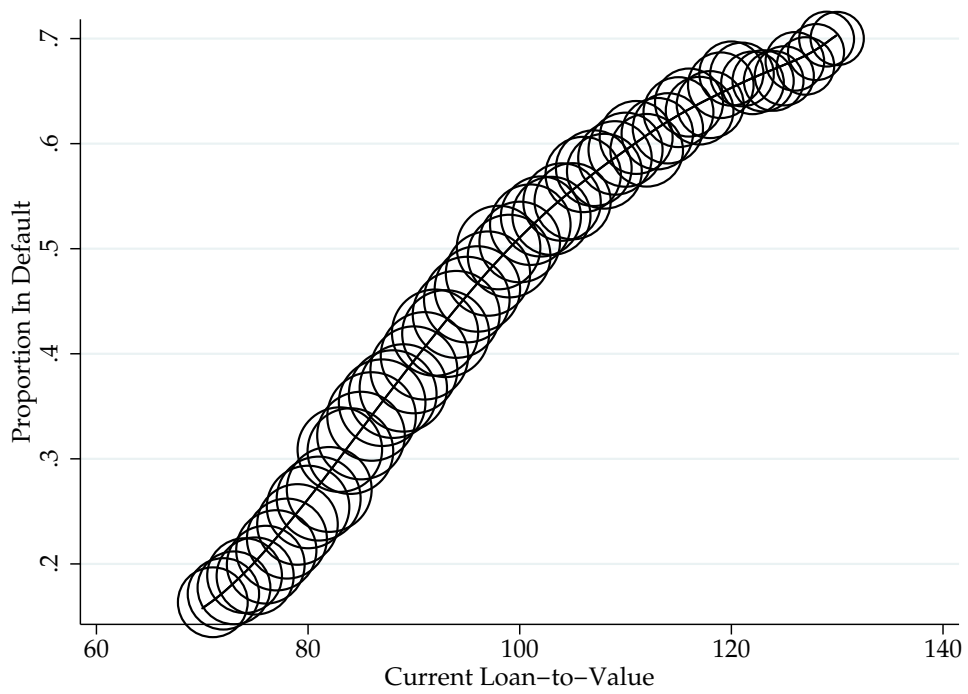
Figure 4
Correlation between current LTV and mortgage default

Hollow dots show the average proportion of loans defaulting between 24 and 36 months for each one-point bin of current loan-to-value at 24 months. Size of dots is proportional to number of borrowers within each bin. Default is defined as 60 or more days past due. The solid line shows a local linear smoothing of the raw data.

A. Nonlinearities in causal relationship between current leverage and default

IV Effect by Quantile Dummies of Current LTV Distribution
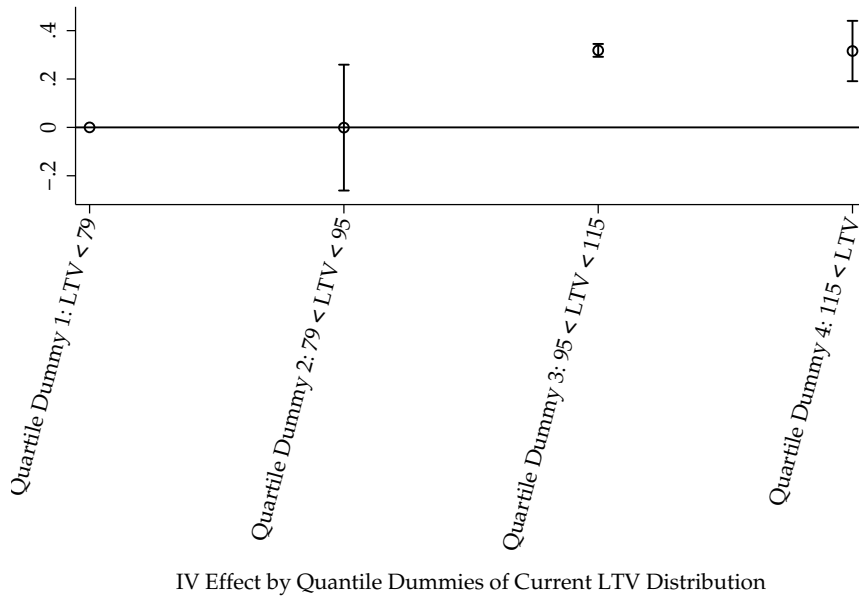
B. Nonlinearities in correlation between origination leverage and default

IV Effect by Quantile Dummies of Origination CLTV Distribution
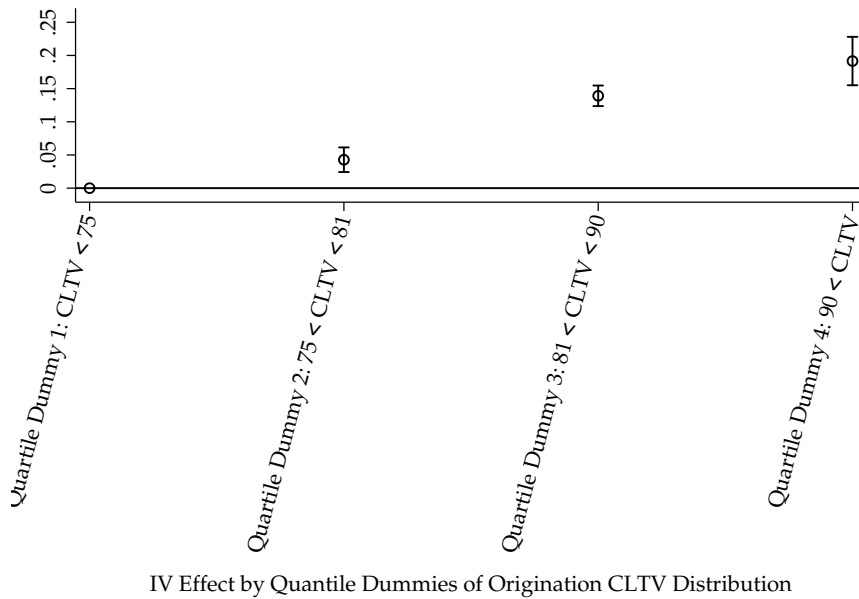
Figure 5
Nonlinearities in the relationship between leverage and default

Both panels plot coefficients from the regressions displayed in column (9) of Appendix Table A.5, which are described in Equation 6. Panel A shows the coefficients on dummy variables of each quartile of current LTV, as instrumented by quartiles of our leave-out mean Z_{it}. Each coefficient can be interpreted relative to the omitted dummy variable—borrowers with a current $LTV < 79$. Panel B shows coefficients on dummy variables corresponding to quartiles of origination leverage, measured as the origination combined LTV. The effect for these specifications, as well, is normalized so that the first quartile has an effect of 0. The sample includes all active loans at 24 months. Default is defined as falling 60 days past due between 24 and 36 months. Regressions control for index type, zip code fixed effects, originator × origination month fixed effects, and our full set of credit and loan controls. These controls include the original home value (we include both a linear term and dummies for each \$25k bin), dummies for each 20-point bin of FICO credit score at origination, and categorical dummies for occupancy, property type, loan purpose, and documentation level. Standard errors are clustered at the originator level.
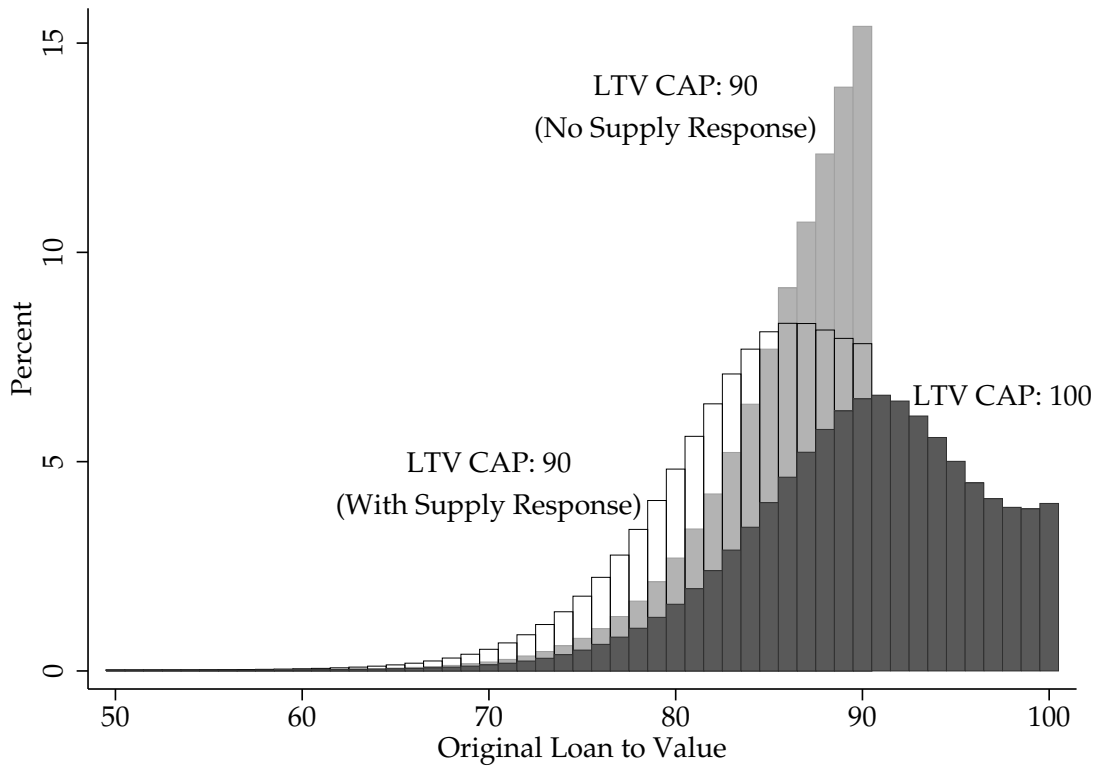
Figure 6
Effect of LTV cap of 90 on leverage: with and without supply response

Bars show simulated proportion of borrowers choosing each original LTV under three regimes. The dark gray bars show equilibrium LTV choices at an LTV cap of 100, the light gray bars show borrowers' LTV choices after a reduction in the LTV cap to 90, but allowing no changes in the prices of contracts below 90. White bars show equilibrium LTV choices with an LTV cap of 90, allowing for the supply response. Figure is based on an exaggerated level of adverse selection. Table 6 shows appropriately calibrated results.

Table 1
Summary statistics

	Treasury		LIBOR		COFI		Fixed	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Origination CLTV	0.80	0.098	0.82	0.12	0.80	0.097	0.81	0.094
Origination LTV	0.77	0.082	0.77	0.076	0.76	0.086	0.76	0.068
Current LTV (24 months)	0.99	0.28	1.05	0.31	0.88	0.26	1.04	0.25
Origination value	4.95	2.89	4.81	2.89	4.93	3.21	6.13	3.05
FICO score	706.0	45.3	717.1	43.9	708.9	45.9	696.4	43.1
No/low documentation	0.82		0.85		0.79		0.92	
Loan for purchase	0.32		0.40		0.40		0.26	
Single family home	0.66		0.65		0.66		0.78	
Investor	0.16		0.19		0.19		0.053	
State:								
- California	0.48		0.43		0.45		0.67	
- Florida	0.14		0.12		0.10		0.073	
- Arizona	0.042		0.048		0.040		0.022	
- Nevada	0.038		0.040		0.047		0.019	
Origination year:								
- 2004	0.076		0.14		0.25		0.026	
- 2005	0.34		0.17		0.36		0.19	
- 2006	0.47		0.28		0.27		0.53	
- 2007	0.11		0.41		0.12		0.25	
Observations	492,563		92,648		11,806		9,975	

Summary statistics by index for option ARMs in our sample with origination combined loan-to-value ratios between 50 and 100. Origination value is measured in hundreds of thousands of USD. Margin is measured in percentage points.

Table 2
Instrument uncorrelated with observables at origination

Dep. var.	Current LTV	Default	Orig. FICO	Orig. CLTV	Orig. value	No/low doc.	Purchase	Single family	Investor
Current LTV Leave-Out Mean	0.845***	0.556***	−27.667	0.224	−0.988	0.225	0.206	−0.038	0.269
(Index × Orig. Month)	(0.188)	(0.183)	(36.435)	(0.180)	(1.090)	(0.187)	(0.450)	(0.116)	(0.357)
Orig. Month FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Index Type FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Zip code FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean of Dep. Var.	0.99	0.44	706.5	0.81	4.88	0.84	0.33	0.64	0.17
N	491,215	491,215	433,809	491,215	491,215	491,215	491,215	491,215	491,215

This table shows coefficients from a regression of our leave-out-mean instrument Z_{it} on a range of borrower-level and loan-level covariates and outcomes. We include all loans in our sample that are active at 24 months. We control throughout for our minimal set of controls: origination month, index type, and zip code. The first two columns represent basic first stage and reduced form versions of our IV specification, respectively. In the first, current leverage, measured as current LTV on the option ARM, is included as a dependent variable. In the second, default, defined as falling 60 days past due between 24 and 36 months, is included as a dependent variable. FICO score, origination LTV, and origination home value are all continuous, while the last four variables are dummy variables indicating the presence of no or low documentation in loan origination, whether the mortgage is for a home purchase instead of a refinancing, whether the property is single family, and whether the borrower was a reported investor. *Mean of Dep. Var.* refers to the mean of the dependent variable. We cluster standard errors at the originator level. *$p < .1$; **$p < .05$; ***$p < .01$.

Table 3
Impact of origination and current leverage on default: linear model

| | *A. Estimates of origination and current leverage on default (60 DPD)* | | | | | | | | |
| | Baseline | | | OLS | | | IV | | |
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Origination Leverage	0.917***	1.062***	1.066***	0.666***	0.827***	0.818***	0.472**	0.701***	0.655***
	(0.064)	(0.067)	(0.073)	(0.084)	(0.081)	(0.075)	(0.183)	(0.198)	(0.127)
Current Leverage				0.301***	0.278***	0.285***	0.533**	0.427*	0.473***
				(0.048)	(0.038)	(0.028)	(0.225)	(0.247)	(0.171)
Orig. Month FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Index Type FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Zip code FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Credit/Loan Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Originator × Orig. Month FEs	No	No	Yes	No	No	Yes	No	No	Yes
Mean of Dep. Var.	0.44	0.44	0.44	0.44	0.44	0.44	0.44	0.44	0.44
N	491,215	491,215	491,215	491,215	491,215	491,215	491,215	491,215	491,215

	B. First-stage estimates of current leverage on leave-out mean								
Current Leverage Leave-Out Mean (Index × Orig. Month)							0.658***	0.648***	0.457***
							(0.129)	(0.122)	(0.091)
Kleibergen-Paap F-Statistic							26.1	28.1	25.0

Panel A shows our main results, regressions of default between 24 and 36 months on origination and current leverage. We include a cross-section of all option ARMs in our sample active at 24 months with an origination combined loan-to-value between 50 and 100. Default is defined as falling 60 days past due between 24 and 36 months. Origination leverage is defined as the combined loan-to-value ratio at origination. Current leverage is defined as the current loan-to-value ratio on the option ARM at 24 months. All specifications control for origination month, index type, and zip code fixed effects. Credit/loan controls refer to the original home value (we include both a linear term and dummies for each $25k bin), dummies for each 20 point bin of FICO credit score at origination, and categorical dummies for occupancy, property type, loan purpose, and documentation level. Originator×orig. month fixed effects refer to fixed effects for each combination of originator and origination month. In columns labeled IV we instrument for current leverage with Z_{it}, the leave-out-mean of current leverage at 24 months for all loans with the same origination month and index type. Panel B shows coefficients and F-statistics from first-stage regressions. *Mean of Dep. Var.* refers to the mean of the dependent variable. *$p < .1$; **$p < .05$; ***$p < .01$.

Table 4
Heterogeneity by initial loan characteristics

	A. Loan documentation					
	No/low documentation			Full documentation		
	Baseline	OLS	IV	Baseline	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)
Origination Leverage	1.120***	0.892***	0.806***	0.724***	0.393***	0.138
	(0.061)	(0.066)	(0.124)	(0.113)	(0.089)	(0.374)
Current Leverage		0.268***	0.370**		0.347***	0.615
		(0.030)	(0.162)		(0.019)	(0.447)
Orig. Month FEs	Yes	Yes	Yes	Yes	Yes	Yes
Index Type FEs	Yes	Yes	Yes	Yes	Yes	Yes
Zip code FEs	Yes	Yes	Yes	Yes	Yes	Yes
Credit/Loan Controls	Yes	Yes	Yes	Yes	Yes	Yes
Originator × Orig. Month FEs	Yes	Yes	Yes	Yes	Yes	Yes
Mean of Dep. Var.	0.47	0.47	0.47	0.30	0.30	0.30
N	411,091	411,091	411,091	77,760	77,760	77,760

	B. Loan purpose					
	Purchase			All others		
	Baseline	OLS	IV	Baseline	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)
Origination Leverage	0.965***	0.767***	0.609***	1.109***	0.850***	0.727***
	(0.056)	(0.050)	(0.122)	(0.068)	(0.074)	(0.100)
Current Leverage		0.331***	0.595***		0.265***	0.392***
		(0.029)	(0.194)		(0.028)	(0.129)
Orig. Month FEs	Yes	Yes	Yes	Yes	Yes	Yes
Index Type FEs	Yes	Yes	Yes	Yes	Yes	Yes
Zip code FEs	Yes	Yes	Yes	Yes	Yes	Yes
Credit/Loan Controls	Yes	Yes	Yes	Yes	Yes	Yes
Originator × Orig. Month FEs	Yes	Yes	Yes	Yes	Yes	Yes
Mean of Dep. Var.	0.42	0.42	0.42	0.45	0.45	0.45
N	162,804	162,804	162,804	326,298	326,298	326,298

	C. State recourse laws					
	Some recourse			Nonrecourse		
	Baseline	OLS	IV	Baseline	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)
Origination Leverage	0.960***	0.645***	0.618***	1.133***	0.937***	0.884***
	(0.078)	(0.066)	(0.117)	(0.071)	(0.079)	(0.142)
Current Leverage		0.361***	0.392**		0.228***	0.289
		(0.029)	(0.156)		(0.027)	(0.186)
Orig. Month FEs	Yes	Yes	Yes	Yes	Yes	Yes
Index Type FEs	Yes	Yes	Yes	Yes	Yes	Yes
Zip code FEs	Yes	Yes	Yes	Yes	Yes	Yes
Credit/Loan Controls	Yes	Yes	Yes	Yes	Yes	Yes
Originator × Orig. Month FEs	Yes	Yes	Yes	Yes	Yes	Yes
Mean of Dep. Var.	0.42	0.42	0.42	0.45	0.45	0.45
N	211,188	211,188	211,188	279,902	279,902	279,902

This table shows additional heterogeneity analysis of Table 3 estimates. We show regressions of default between 24 and 36 months on origination and current leverage, split across three categories. Panel A shows results for loans with no/low documentation versus full documentation. Panel B shows results for loans for home purchases versus all other purposes. Panel C shows loans in some recourse vs. nonrecourse states, based on the classification in as coded in Rao and Walsh (2009). We include a cross-section of all option ARMs in our sample active at 24 months with an origination combined loan-to-value between 50 and 100. Default is defined as falling 60 days past due between 24 and 36 months. Origination leverage is defined as the combined loan-to-value ratio at origination. Current leverage is defined as the current loan-to-value ratio on the option ARM at 24 months. All specifications control for origination month, index type, and zip code fixed effects. Credit/loan controls refer to the original home value (we include both a linear term and dummies for each $25k bin), dummies for each 20 point bin of FICO credit score at origination, and categorical dummies for occupancy, property type, loan purpose, and documentation level. Originator × orig. month fixed effects refer to fixed effects for each combination of originator and origination month. In columns labeled IV we instrument for current leverage with Z_{it}, the leave-out-mean of current leverage at 24 months for all loans with the same origination month and index type. Panel B shows coefficients and F-statistics from first stage regressions. *Mean of Dep. Var.* refers to the mean of the dependent variable. *$p < .1$; **$p < .05$; ***$p < .01$.

Table 5
Impact of original and current leverage on default: joint model

	A. Estimated parameters from leverage choice and default equations								
	Baseline			No instrument			Control function		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
ρ: Correlation of Errors in Default and Leverage Choice	0.262*** (0.011)	0.274*** (0.009)	0.276*** (0.006)	0.189*** (0.013)	0.213*** (0.010)	0.213*** (0.005)	0.116** (0.054)	0.162*** (0.056)	0.118** (0.055)
Current Leverage				0.970*** (0.115)	0.978*** (0.095)	1.000*** (0.060)	1.854*** (0.656)	1.719** (0.732)	2.327*** (0.586)
Orig. Month FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Index Type FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
MSA FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Credit/Loan Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Originator × Orig. Year FEs	No	No	Yes	No	No	Yes	No	No	Yes
S.D. of Default Error				1.03	1.02	1.00	0.54	0.58	0.43
N	488,999	488,999	488,999	488,999	488,999	488,999	488,999	488,999	488,999

	B. Excluded instrument for current leverage								
Current Leverage Leave-Out Mean (Index × Orig. Month)							0.706*** (0.124)	0.679*** (0.121)	0.472*** (0.123)

Panel A shows results from our joint model of leverage demand and the default choice. We include a cross-section of all option ARMs in our sample active at 24 months with an origination combined loan-to-value between 50 and 100. Default is defined as falling 60 days past due between 24 and 36 months. Origination leverage is defined as the combined loan-to-value ratio at origination. Current leverage is defined as the current loan-to-value ratio on the option ARM at 24 months. All specifications control for origination month, index type, and MSA fixed effects. Credit/loan controls refer to the original home value (we include both a linear term and dummies for each $25k bin), dummies for each 20 point bin of FICO credit score at origination, and categorical dummies for occupancy, property type, loan purpose, and documentation level. Originator×orig. year fixed effects refer to fixed effect for each combination of originator and origination year. In columns labeled "Baseline," we estimate the leverage choice and default equation without include current leverage in the default equation. Columns labeled "No Instrument" include current leverage in the default equation. In columns labeled "Control function" we specify an additional linear equation in which current leverage is a function of Z_{it}, the leave-out-mean of current leverage at 24 months for all loans with the same origination month and index type. ρ displays the estimated correlation between the errors in the leverage and default equations, capturing adverse selection. *S.D. of the Default Error* shows the reciprocal of the coefficient on current leverage. Panel B shows coefficients on the excluded instrument. *Mean of Dep. Var.* refers to the mean of the dependent variable. Standard errors are clustered at the originator level. *$p < .1$; **$p < .05$; ***$p < .01$.

Table 6
Simulation results: the impact of a reduction in the LTV cap from 100 to 90

	Col. 1: LTV cap of 100	Col. 2: LTV cap of 90 No supply response	Col. 3: LTV cap of 90 With supply response
Average Loan Size (Thousands)	$387.5	$361.4	$361.3
Average Interest Rate	14.82%	11.38%	11.83%
Average Balance (Thousands)	$445.0	$402.5	$404.0
Defaults	18.31%	15.09%	15.19%
Naive Defaults	-	14.24%	-

Parameters		
Initial Price: $H_0 = \$495.0$	Final Value: $H \sim N(\mu_1 = 594, \sigma_1 = 297)$	Proportion Behavioral: 1%
CARA Coefficient: a=0.005	Viscosity: $\gamma = 1.075$	$\beta = \frac{1}{1+r_f} = 0.975$
Prop. Recovered in Default: $\delta = 0.80$	$C_i \sim N(0.38\mu_1, 0.43\mu_1)$	Borrowers: $N = 1000$

Simulations from structural model described in Section 4. CARA utility assumed, 1,000 simulated borrowers, with 1% behavioral. Viscosity set to match estimated $\rho = 0.118$. Parameters if C_i distribution taken from estimation in Table 5. The mean based on the linear projection using average value of continuous covariates and the omitted category of all categorical covariates. Initial home prices are set based on average for the largest index category (Treasury). 20% net expected net growth over the term of the loan assumed. The first column shows equilibrium outcomes with an LTV cap of 100. The second column shows borrower responses to the removal of all contracts with initial LTV between 90 and 100, holding fixed all contracts with initial LTV less than 90. Naive defaults refers to expected defaults calculated ignoring borrower heterogeneity and extrapolating from default probabilities at each LTV with an LTV cap of 100. The third column shows equilibrium outcomes with an LTV cap of 90.

A Internet Appendix: Tables and Figures

Figure A.1

Stylized Monthly Payment and Balance Trajectory for Option ARM
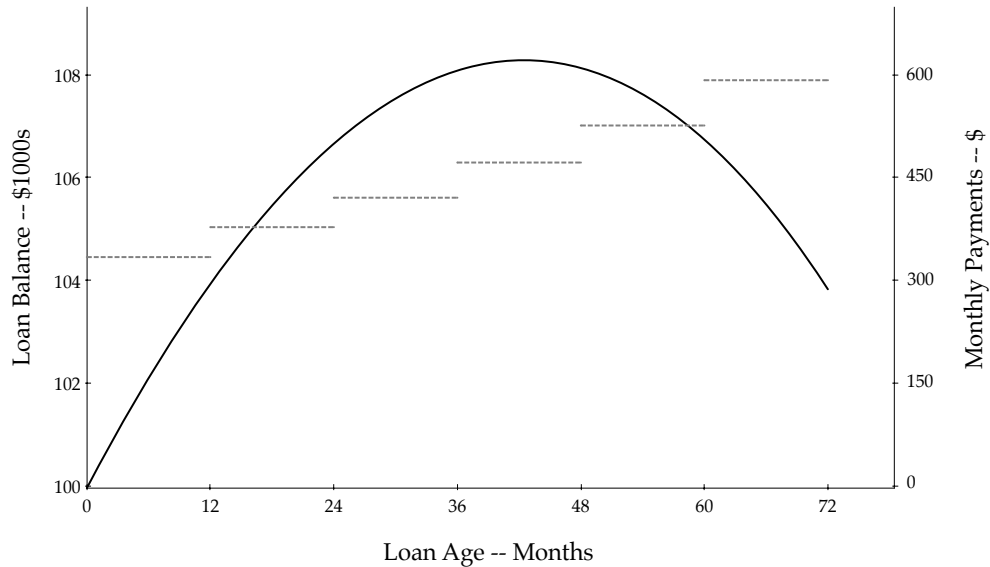
The solid line shows the balance trajectory for a stylized option ARM with an initial loan of $100,000. The balance is initially increasing, demonstrating negative amortization. Monthly payments, shown by the dashed lines, increase by 7.5% per year regardless of balance. As payments grow, the balance begins to decrease, as shown by the parabolic shape of the balance trajectory. At 5 years the monthly payment jumps to the fully amortizing amount.

Figure A.2
Variation in Balance Trajectories for Option ARMs
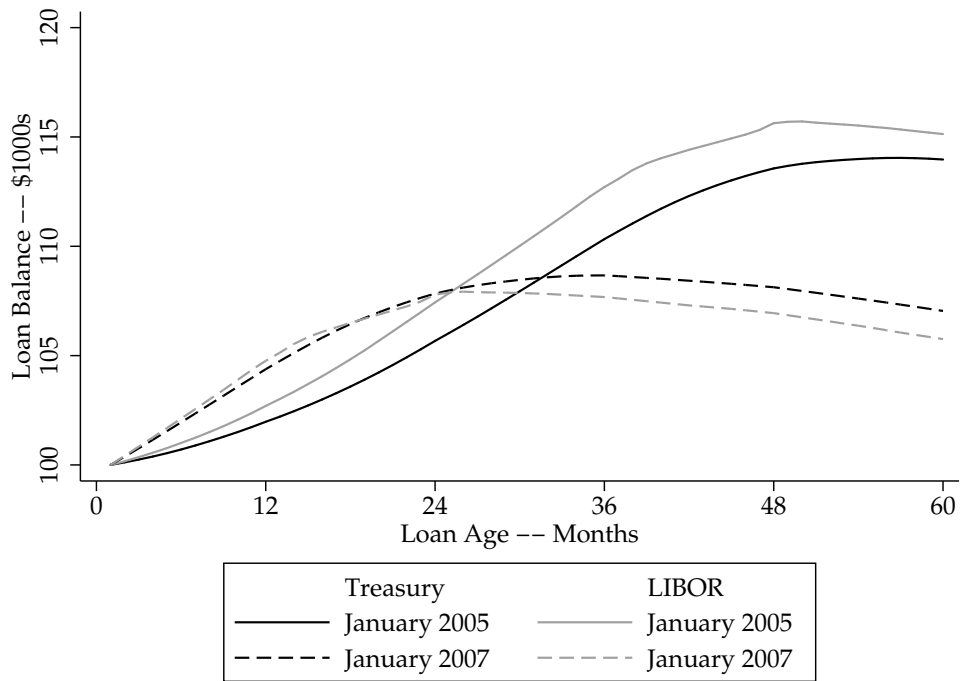
Simulated balance trajectories for $100,000 LIBOR- and Treasury-indexed loans originated in January 2005 or January 2007. Trajectories assume margin of 3.5 percent and initial payment based on 1.75 percent teaser. Treasury refers to 1-year MTA, LIBOR refers to 3-month duration.

Figure A.3
Variation in Leave-Out Instrument Over Time

(a) Average Predicted LTV by Index

(b) Evolution of Mortgage Indexes

(c) Average Difference in Reduced Form

(d) Originations by Index Over Time
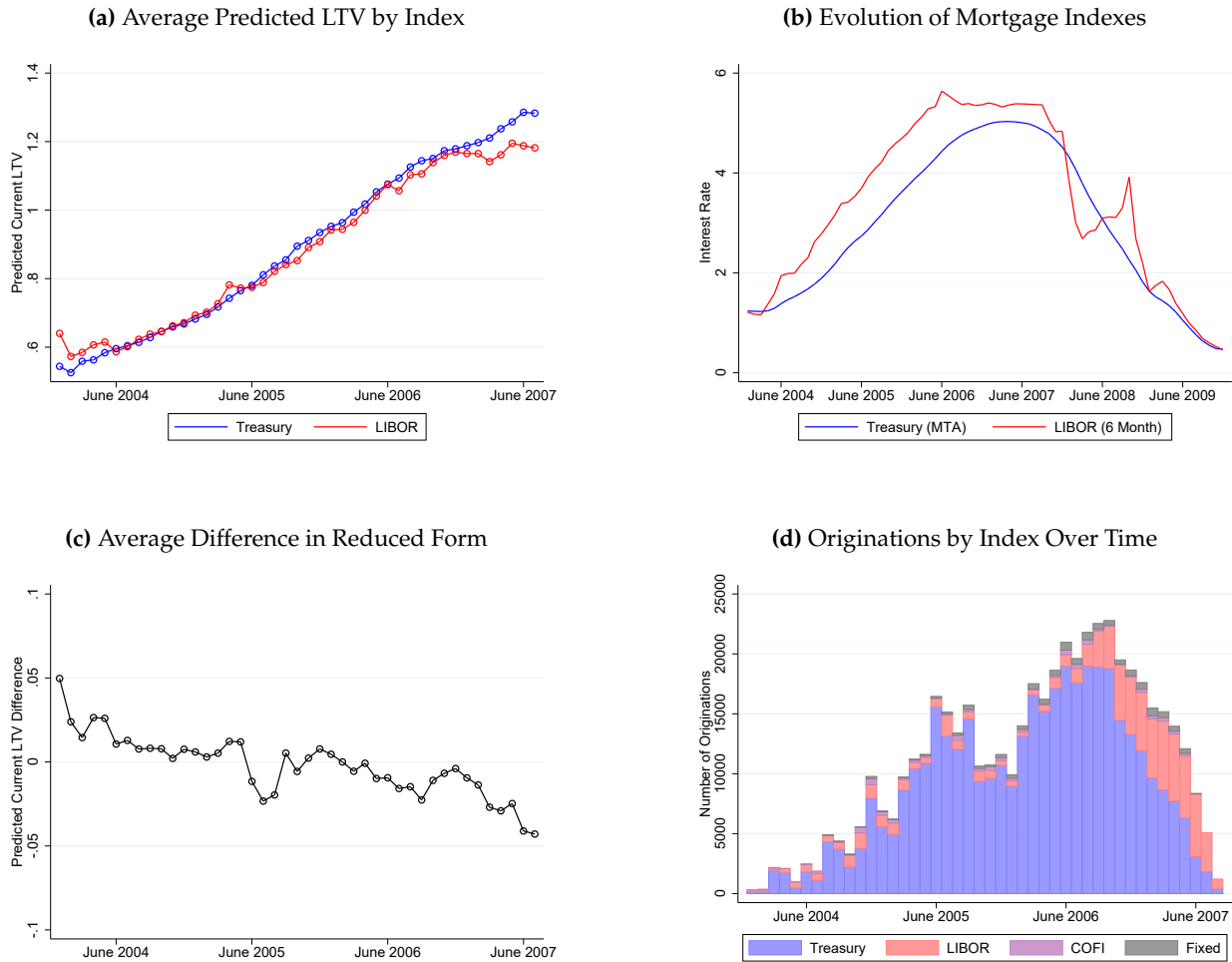
Notes: Panel (a) shows predicted LTV based on the first stage shown in Column 9 of Table 3. Panel (b) shows the evolution of the LIBOR (6 month) and Treasury (MTA) mortgage indexes over the relevant period. Panel (c) shows the difference between predicted LTVs for LIBOR and Treasury shown in Panel (a) over time. Panel (d) shows the number of originations in each index group for each month.

Figure A.4
Correlation Between Leave-out-Mean Instrument and Mortgage Default
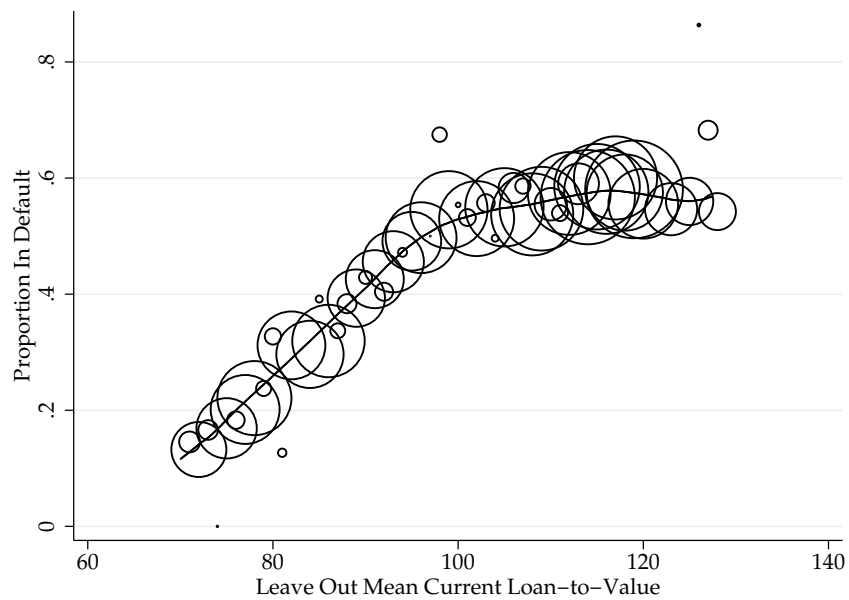
Hollow dots show the average proportion of loans defaulting between 24 and 36 months for each 1-point bin of the leave-out-mean instrument at 24 months. Size of dots is proportional to number of borrowers within each bin. Default is defined as 60 or more days past due. The solid line shows a local linear smoothing of the raw data.

Figure A.5
Time Variation in Adverse Selection

Panel A: Coefficient on Origination Leverage in Baseline, OLS, and IV Specifications
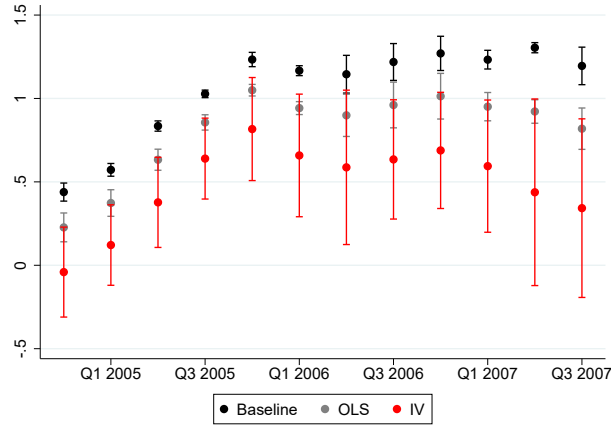
Panel B: Mean Default Rate by Quarter
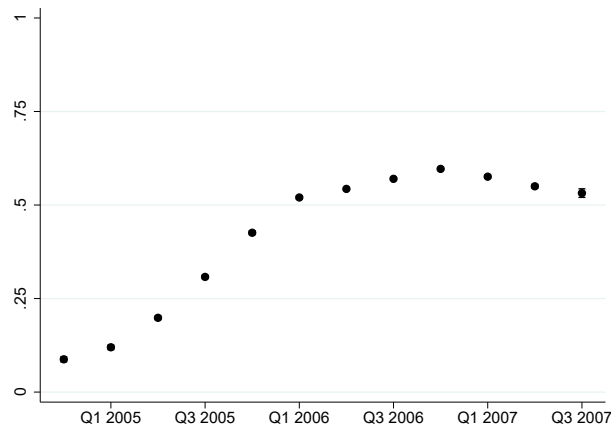
Notes: Panel A shows coefficients on the interaction between initial leverage and origination quarter in Baseline, OLS and IV specifications of our most saturated linear model. Our measure of default is 60 days past due by 36 months, and the sample includes all loans active at 24 months with an origination combined loan-to-value between 50 and 100. Bars represent 95 percent confidence intervals based on standard errors clustered at the originator level. Panel B shows the average default rate quarter by quarter for the same sample.

Figure A.6
Simulation Robustness: Equilibrium Interest Rates

Panel A: Varying the Quantity of Behavioral Borrowers
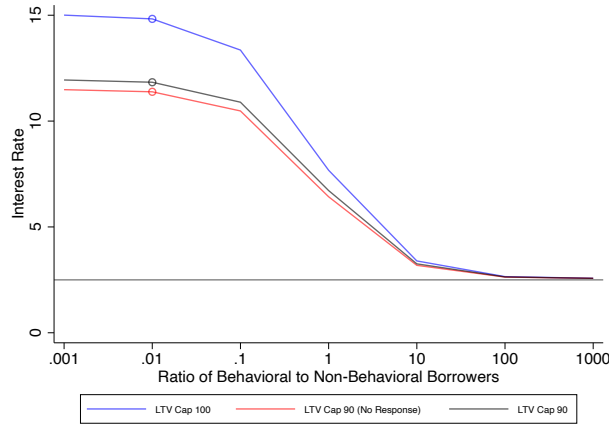
Panel B: Varying the Degree of Adverse Selection
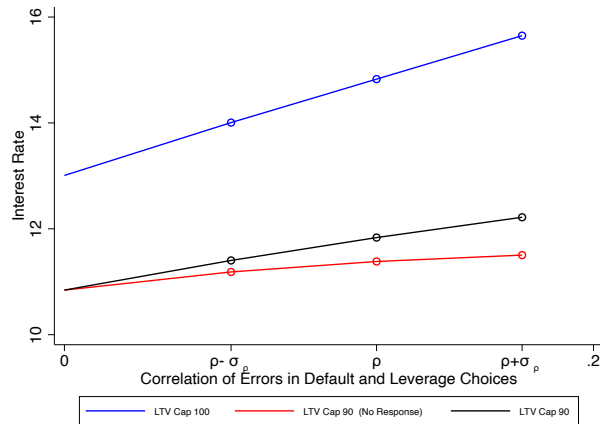
In Panel A, lines display equilibrium interest rates from the simulated model while varying the fraction of behavioral borrowers. Equilibria are computed under three policy regimes: an LTV cap of 100 (in blue) and LTV cap of 90 with no supply response (in red) and an LTV cap of 90 with a full supply response (in black). Circles indicate equilibrium interest rates with a mass of behavioral borrowers equal to one percent of standard borrowers, the parameterization used to create the results in Table 6. The x-axis indicates the ratio of behavioral to non-behavioral, so 1 indicates a mass of behavioral borrowers equal to 100 percent of the mass of standard borrowers. The horizontal line represents the risk free rate. In Panel B, lines display equilibrium interest rates from the simulated model while varying ρ: the key adverse selection parameter that represents the correlation of the errors in the leverage choice and default equations. Equilibria are computed under three policy regimes: an LTV cap of 100 (in blue) an LTV cap of 90 with no supply response (in red) and an LTV cap of 90 with a full supply response (in black). Circles are included for the equilibrium interest rates computed with our estimated value of ρ and $\rho\pm$ one standard deviation (σ_ρ). The equilibria at $\rho = 0$ are displayed where the lines intersect the y-axis.

Table A.1
Summary Statistics: Other Mortgage Products

| | 5/1 Adjustable Rate | | | | Private Label Fixed Rate | | Fannie Mae Fixed Rate | |
| | Treasury | | LIBOR | | | | | |
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Origination LTV	74.9	8.67	76.4	8.20	72.9	20.4	76.0	10.9
Origination Value	6.65	4.95	5.96	4.80	3.67	0.62	2.57	1.33
FICO Score	729.2	49.5	717.5	52.4	676.4	71.4	719.8	57.8
Margin	2.77	0.33	2.36	0.40				
Primary Residence	0.85		0.84		0.82		0.88	
No/Low Documentation	0.53		0.71		0.41			
State:								
- California	0.48		0.40		0.19		0.090	
- Florida	0.070		0.095		0.10		0.076	
- Arizona	0.034		0.042		0.032		0.035	
- Nevada	0.024		0.031		0.02		0.012	
Origination Year:								
- 2004	0.32		0.17		0.24		0.28	
- 2005	0.47		0.27		0.34		0.26	
- 2006	0.15		0.37		0.32		0.21	
- 2007	0.063		0.19		0.10		0.25	
Observations	87565		321618		3744893		3790554	

Summary statistics for three alternative samples of mortgage products. The first 4 columns show 5/1 ARMs drawn from the same data source as our option ARM sample. Columns 5–6 show the sample of 30 year fixed rate mortgages drawn from the same sample as well. Columns 7–8 show a sample of fixed rate drawn from Fannie Mae's loan performance database. Treasury refers to loans indexed to Treasury rates, LIBOR refers to those indexed to LIBOR. Origination Value is measured in hundreds of thousands of USD. Margin is measured in percentage points.

Table A.2
Fraction of Treasury-Indexed Loans by Lender

Originator	Percent of Loans Indexed to Treasury
Countrywide	78.5%
American Home Mortgage	99.1%
Greenpoint	66.2%
IndyMac	92.1%
EMC Mortgage	32.1%
Bear Stearns	33.1%
Residential Funding	93.4%
Downey	100%
Bank United	100%
Mortgage IT Inc	61.1%

Percentage of Treasury-indexed loans for the top 10 originators in the sample.

Table A.3

Accounting for Attrition: Impact of Original and Current Leverage on Cumulative Default

	Panel A: Estimates of Original and Current Leverage on Default (60 DPD)								
	Default by 36 Months			Default by 48 Months			Default by 60 Months		
	Baseline	OLS	IV	Baseline	OLS	IV	Baseline	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Origination Leverage	1.041***	0.748***	0.810***	1.130***	0.917***	0.934***	1.125***	0.913***	0.989***
	(0.109)	(0.110)	(0.130)	(0.078)	(0.077)	(0.064)	(0.059)	(0.056)	(0.055)
Current Leverage		0.342***	0.270***		0.207***	0.191***		0.192***	0.123***
		(0.032)	(0.098)		(0.017)	(0.050)		(0.011)	(0.046)
Orig. Month FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Index Type FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Zipcode FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Credit/Loan Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Originator x Orig. Month FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean of Dep. Var.	0.35	0.35	0.35	0.45	0.45	0.45	0.50	0.50	0.50
N	606992	606992	606992	606992	606992	606992	606992	606992	606992

Panel B: First Stage Estimates of Current Leverage on Leave-Out Mean

Current Leverage Leave-Out Mean (Index × Orig. Month)		0.536***				0.507***			0.628***
		(0.112)				(0.132)			(0.065)
Kleibergen-Paap F-Statistic		22.9				14.7			93.9

This table repeats the analysis in Table 3 using the full sample of loans and considering the outcome of cumulative default. We define cumulative default as falling 60 days past due at any point up to 36, 48, or 60 months, depending on the specification. We include a cross-section of all option ARMs with an origination combined loan-to-value between 50 and 100. Origination leverage is defined as the combined loan-to-value ratio at origination. Current leverage is defined as the imputed current loan-to-value ratio at 24, 36, or 48 months (In columns (1)–(3), (4)–(6), and (7)–(9), respectively) given the borrowers loan terms if they had made exactly the minimum payment every month. This variable may be defined at any point for any borrower, regardless of whether or not he or she has exited the sample. All specifications control for origination month, index type, and zip code fixed effects. Credit/loan controls refer to the original home value (we include both a linear term and dummies for each \$25k bin), dummies for each 20 point bin of FICO credit score at origination, and categorical dummies for occupancy, property type, loan purpose, and documentation level. Originator × orig. month fixed effects refer to fixed effects for each combination of originator and origination month. In columns labeled IV we instrument for current leverage with Z_{it}, the leave-out-mean of imputed current leverage at the relevant point (24, 36 or 48 months) for all loans with the same origination month and index type. Panel B shows coefficients and *F*-statistics from first stage regressions. Mean of Dep. Var. refers to the mean of the dependent variable. * denotes 10% significance, ** denotes 5% significance, *** denotes 1% significance.

	Minimal Controls			Geography × Time FEs			Flexible Current Leverage		
	Baseline	OLS	IV	Baseline	OLS	IV	Baseline	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Estimates of Original and Current Leverage on Default (60 DPD)									
Origination Leverage	0.904***	0.477***	0.471**	1.082***	0.794***	0.646***	1.066***	0.756***	0.519*
	(0.074)	(0.058)	(0.211)	(0.076)	(0.057)	(0.182)	(0.073)	(0.069)	(0.291)
Current Leverage		0.495***	0.503**		0.330***	0.499**		−1.345***	61.254**
		(0.023)	(0.245)		(0.012)	(0.235)		(0.271)	(26.955)
Current Leverage^2								2.199***	−77.448**
								(0.378)	(34.477)
Current Leverage^3								−1.093***	40.304**
								(0.179)	(18.021)
Current Leverage^4								0.172***	−7.300**
								(0.025)	(3.259)
Orig. Month FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Index Type FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Zipcode FEs	No	No	No	No	No	No	Yes	Yes	Yes
MSA x Orig. Month FEs	No	No	No	Yes	Yes	Yes	No	No	No
Credit/Loan Controls	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Originator x Orig. Month FEs	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Mean of Dep. Var.	0.44	0.44	0.44	0.44	0.44	0.44	0.44	0.44	0.44
N	491,215	491,215	491,215	488,595	488,595	488,595	491,215	491,215	491,215

Panel B: First Stage Estimates of Current Leverage on Leave-Out Mean

Current Leverage Leave-Out Mean (Index × Orig. Month)			0.673*** (0.117)				0.362*** (0.083)		
Kleibergen-Paap F-Statistic			32.9				19.2		

This table shows robustness for the results in Table 3. The columns labeled Minimal Controls include only origination month and index type fixed effects. The columns labeled Geography × Time FEs replace zip code fixed effects with MSA × origination month fixed effects. The columns labeled Flexible Current Leverage include a fourth order polynomial in current leverage. We include a cross-section of all Option ARMs in our sample active at 24 months with an origination combined loan-to-value between 50 and 100. Default is defined as falling 60 days past due between 24 and 36 months. Origination leverage is defined as the combined loan-to-value ratio at origination. Current leverage is defined as the current loan-to-value ratio on the option ARM at 24 months. Credit/Loan controls refer to the original home value (we include both a linear term and dummies for each $25k bin), dummies for each 20 point bin of FICO credit score at origination, and categorical dummies for occupancy, property type, loan purpose, and documentation level. Originator × orig. month fixed effects refer to fixed effect for each combination of originator and origination month. MSA×orig. month fixed effects refer to fixed effects for each combination of MSA and origination month. In columns labeled IV we instrument for current leverage with Z_{it}, the leave-out-mean of current leverage at 24 months for all loans with the same origination month and index type. In column (9) we include a fourth order polynomial of the instrument in the first stage. Panel B shows coefficients and F-statistics from first stage regressions. Mean of Dep. Var. refers to the mean of the dependent variable. * denotes 10% significance, ** denotes 5% significance, *** denotes 1% significance.

Table A.5

Effects Across Distribution of Leverage

Separating Out Quartiles of Origination and Current Leverage

	Baseline			OLS			IV		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Quartile Dummy 2: 75 < Origination Leverage ≤ 81	0.089*** (0.017)	0.115*** (0.017)	0.102*** (0.010)	0.036** (0.017)	0.067*** (0.017)	0.054*** (0.008)	0.029** (0.012)	0.059*** (0.010)	0.043*** (0.009)
Quartile Dummy 3: 81 < Origination Leverage ≤ 90	0.195*** (0.013)	0.216*** (0.011)	0.215*** (0.010)	0.131*** (0.014)	0.158*** (0.011)	0.156*** (0.009)	0.120*** (0.010)	0.146*** (0.010)	0.139*** (0.008)
Quartile Dummy 4: 90 < Origination Leverage	0.224*** (0.021)	0.293*** (0.021)	0.294*** (0.023)	0.144*** (0.025)	0.218*** (0.024)	0.218*** (0.025)	0.129*** (0.025)	0.200*** (0.024)	0.192*** (0.019)
Quartile Dummy 2: 79 < Current Leverage ≤ 95				0.107*** (0.005)	0.100*** (0.005)	0.100*** (0.005)	0.059 (0.168)	0.016 (0.140)	-0.001 (0.133)
Quartile Dummy 3: 95 < Current Leverage ≤115				0.214*** (0.017)	0.203*** (0.016)	0.204*** (0.014)	0.296*** (0.021)	0.298*** (0.019)	0.318*** (0.014)
Quartile Dummy 4: 115 < Current Leverage				0.294*** (0.021)	0.278*** (0.019)	0.281*** (0.014)	0.312*** (0.072)	0.295*** (0.056)	0.316*** (0.064)
Orig. Month FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Index Type FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
ZIP FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Credit/Loan Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Originator x Orig. Month FEs	No	No	Yes	No	No	Yes	No	No	Yes
Mean of Dep. Var.	0.44	0.44	0.44	0.44	0.44	0.44	0.44	0.44	0.44
N	491,215	491,215	491,215	491,215	491,215	491,215	491,215	491,215	491,215

These specifications alter the specifications shown in Table 3 by including dummy variables for each quartile of the leverage distribution for both Current and Origination LTV, as in equation 6. The effects for these equations can be interpreted as the additional percentage change in default as the result of a loan being in the particular leverage category, relative to the omitted first quartile. We include a cross-section of all Option ARMs in our sample active at 24 months with an origination combined loan-to-value between 50 and 100. Default is defined as falling 60 days past due between 24 and 36 months. Origination leverage is included as quartiles of the combined loan-to-value ratio at origination. Current leverage is included as quartiles of the current loan-to-value ratio on the option ARM at 24 months. All specifications control for origination month, index type, and zip code fixed effects. Credit/Loan controls refer to the original home value (we include both a linear term and dummies for each $25k bin), dummies for each 20 point bin of FICO credit score at origination, and categorical dummies for occupancy, property type, loan purpose, and documentation level. Originator × Orig. Month fixed effects refer to fixed effects for each combination of originator and origination month. In columns labeled IV we instrument for current leverage with quartiles of Z_{it}, the leave-out-mean of current leverage at 24 months for all loans with the same origination month and index type. * denotes 10% significance, ** denotes 5% significance, *** denotes 1% significance.

Impact of Original and Current Leverage: Accounting for Prepayment

Panel A: Estimates of Original and Current Leverage on Prepayment

	Baseline			OLS			IV		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Origination Leverage	−0.196***	−0.242***	−0.259***	−0.172***	−0.219***	−0.242***	0.039	−0.006	−0.053
	(0.050)	(0.056)	(0.052)	(0.043)	(0.050)	(0.042)	(0.167)	(0.193)	(0.066)
Current Leverage				−0.029*	−0.027**	−0.019	−0.281	−0.279	−0.237***
				(0.015)	(0.013)	(0.014)	(0.199)	(0.201)	(0.051)
Orig. Month FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Index Type FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Zipcode FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Credit/Loan Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Originator x Orig. Month FEs	No	No	Yes	No	No	Yes	No	No	Yes
Mean of Dep. Var.	0.088	0.088	0.088	0.088	0.088	0.088	0.088	0.088	0.088
N	491,215	491,215	491,215	491,215	491,215	491,215	491,215	491,215	491,215

First Stage Estimates of Current Leverage on Leave-Out Mean

Current Leverage Leave-Out Mean (Index × Orig. Month)							0.658***	0.648***	0.457***
							(0.129)	(0.122)	(0.091)
Kleibergen-Paap F-Statistic							26.1	28.1	25.0

Panel B: Estimates of Original and Current Leverage on Default (60 DPD) Excluding Prepayers

	Baseline			OLS			IV		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Origination Leverage	0.940***	1.100***	1.111***	0.717***	0.890***	0.887***	0.555**	0.769***	0.707***
	(0.052)	(0.053)	(0.059)	(0.073)	(0.068)	(0.062)	(0.214)	(0.240)	(0.179)
Current Leverage				0.267***	0.249***	0.257***	0.462*	0.392	0.463**
				(0.048)	(0.038)	(0.026)	(0.261)	(0.289)	(0.223)
Orig. Month FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Index Type FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Zipcode FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Credit/Loan Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Originator x Orig. Month FEs	No	No	Yes	No	No	Yes	No	No	Yes
Mean of Dep. Var.	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48
N	447,576	447,575	447,569	447,576	447,575	447,569	447,576	447,575	447,569

First Stage Estimates of Current Leverage on Leave-Out Mean

Current Leverage Leave-Out Mean (Index × Orig. Month)							0.590***	0.581***	0.409***
							(0.133)	(0.128)	(0.094)
Kleibergen-Paap F-Statistic							19.8	20.6	18.9

 This table shows robustness for the results in Table 3. Panel A repeats the analysis in Table 3 but replaces the dependent variable with an indicator variable equal to one if the loan in question is prepaid between 24 and 36 months. In this panel, we include a cross-section of all option ARMs in our sample active at 24 months with an origination combined loan-to-value between 50 and 100. Panel B exactly repeats the analysis in Table 3 but excludes any loan that is prepaid between 24-36 months. Default is defined as falling 60 days past due between 24 and 36 months. Origination leverage is defined as the combined loan-to-value ratio at origination. Current leverage is defined as the current loan-to-value ratio on the option ARM at 24 months. Credit/Loan controls refer to the original home value (we include both a linear term and dummies for each $25k bin), dummies for each 20 point bin of FICO credit score at origination, and categorical dummies for occupancy, property type, loan purpose, and documentation level. Originator × orig. month fixed effects refer to fixed effect for each combination of originator and origination month. In columns labeled IV we instrument for current leverage with Z_{it}, the leave-out-mean of current leverage at 24 months for all loans with the same origination month and index type. Panel B shows coefficients and F-statistics from first stage regressions. Mean of Dep. Var. refers to the mean of the dependent variable. * denotes 10% significance, ** denotes 5% significance, *** denotes 1% significance.

	(1) Baseline	(2) Linear Income	(3) Income Deciles	(4) Linear Debt/Income	(5) Debt/Income Deciles
Current Leverage	0.473***	0.452**	0.452**	0.452**	0.437**
	(0.171)	(0.183)	(0.182)	(0.175)	(0.179)
Origination Leverage	0.655***	0.673***	0.673***	0.672***	0.688***
	(0.127)	(0.135)	(0.135)	(0.130)	(0.132)
Income ($10,000s)		-0.015***			
		(0.001)			
Total Debt/Income				0.008***	
				(0.001)	
Orig. Month FEs	Yes	Yes	Yes	Yes	Yes
Index Type FEs	Yes	Yes	Yes	Yes	Yes
ZIP FEs	Yes	Yes	Yes	Yes	Yes
Credit/Loan Controls	Yes	Yes	Yes	Yes	Yes
Originator x Orig. Year FEs	Yes	Yes	Yes	Yes	Yes
Income Deciles	No	No	Yes	No	No
Debt/Income Deciles	No	No	No	No	Yes
N	491,215	491,215	491,215	491,215	491,215

Panel B: Income Control—Full Documentation Sample

	Baseline	Linear Income	Income Deciles	Linear Debt/Income	Debt/Income Deciles
Current Leverage	0.615	0.572	0.545	0.515	0.628
	(0.447)	(0.470)	(0.475)	(0.476)	(0.453)
Origination Leverage	0.138	0.183	0.208	0.238	0.133
	(0.374)	(0.396)	(0.399)	(0.397)	(0.383)
Income ($10,000s)		-0.015***			
		(0.001)			
Total Debt/Income				0.010***	
				(0.001)	
Orig. Month FEs	Yes	Yes	Yes	Yes	Yes
Index Type FEs	Yes	Yes	Yes	Yes	Yes
ZIP FEs	Yes	Yes	Yes	Yes	Yes
Credit/Loan Controls	Yes	Yes	Yes	Yes	Yes
Originator x Orig. Year FEs	Yes	Yes	Yes	Yes	Yes
Income Deciles	No	No	Yes	No	No
Debt/Income Deciles	No	No	No	No	Yes
N	77,760	77,760	77,760	77,760	77,760

Panel C: Income Control—Low/No Documentation Sample

	Baseline	Linear Income	Income Deciles	Linear Debt/Income	Debt/Income Deciles
Current Leverage	0.370**	0.349*	0.352*	0.356*	0.329*
	(0.162)	(0.189)	(0.189)	(0.183)	(0.172)
Origination Leverage	0.806***	0.824***	0.820***	0.816***	0.841***
	(0.124)	(0.144)	(0.144)	(0.139)	(0.130)
Income ($10,000s)		-0.015***			
		(0.001)			
Total Debt/Income				0.007***	
				(0.001)	
Orig. Month FEs	Yes	Yes	Yes	Yes	Yes
Index Type FEs	Yes	Yes	Yes	Yes	Yes
ZIP FEs	Yes	Yes	Yes	Yes	Yes
Credit/Loan Controls	Yes	Yes	Yes	Yes	Yes
Originator x Orig. Year FEs	Yes	Yes	Yes	Yes	Yes
Income Deciles	No	No	Yes	No	No
Debt/Income Deciles	No	No	No	No	Yes
N	411,091	411,091	411,091	411,091	411,091

This table shows robustness for the results in Table 3 including controls for income and the ratio of Debt-to-Income (DTI). Column 1 in all three panels repeats the IV analysis in column 9 of Table 3. Column 2 additionally includes a linear control for income, sourced from the Equifax prediction model, while column 3 instead includes dummy variables for each income decile. Column 4 includes a linear control for the debt-to-income ratio, while column 5 includes dummy variables for each decile of DTI. In Panel A, we include our complete sample: a cross-section of all option ARMs in our sample active at 24 months with an origination combined loan-to-value between 50 and 100. In Panel B we include only those with full documentation. In Panel C we include those with no or low documentation. Default is defined as falling 60 days past due between 24 and 36 months. Origination leverage is defined as the combined loan-to-value ratio at origination. Current leverage is defined as the current loan-to-value ratio on the option ARM at 24 months. Credit/Loan controls refer to the original home value (we include both a linear term and dummies for each $25k bin), dummies for each 20 point bin of FICO credit score at origination, and categorical dummies for occupancy, property type, loan purpose, and documentation level. Originator × orig. month fixed effects refer to fixed effect for each combination of originator and origination month. Mean of Dep. Var. refers to the mean of the dependent variable. * denotes 10% significance, ** denotes 5% significance, *** denotes 1% significance.

Table A.8

Impact of Original and Current Leverage on Default: Further Robustness

	Panel A: Estimates of Original and Current Leverage on Default (60/90 DPD)								
	No Junior Lien			LIBOR and Treasury Only			90 Days Past Due		
	Baseline	OLS	IV	Baseline	OLS	IV	Baseline	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Origination Leverage	1.103***	0.674***	0.535**	1.069***	0.818***	0.505***	1.073***	0.822***	0.647***
	(0.096)	(0.116)	(0.268)	(0.073)	(0.078)	(0.156)	(0.079)	(0.082)	(0.138)
Current Leverage		0.331***	0.439*		0.286***	0.644***		0.289***	0.489***
		(0.039)	(0.241)		(0.031)	(0.190)		(0.029)	(0.183)
Orig. Month FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Index Type FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Zipcode FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Credit/Loan Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Originator x Orig. Month FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean of Dep. Var.	0.40	0.40	0.40	0.44	0.44	0.44	0.39	0.39	0.39
N	346,919	346,919	346,919	473,179	473,179	473,179	491,215	491,215	491,215

Panel B: First Stage Estimates of Current Leverage on Leave-Out Mean

Current Leverage Leave-Out Mean (Index × Orig. Month)		0.357***			0.556***			0.457***	
		(0.034)			(0.123)			(0.091)	
Kleibergen-Paap F-Statistic		108.6			20.5			25.0	

This table shows robustness for the results in Table 3. The columns labeled No Junior Lien subset to loans for which we have no observable junior liens, and so for which the origination combined loan-to-value and origination loan-to-value on the option ARM are the same. The columns labeled LIBOR and Treasury Only subset to loans indexed to either LIBOR or Treasury. The columns labeled 90 Days Past Due define default as falling 90 days past due between 24 and 36 months. Outside of the restrictions specified above, we include a cross-section of all option ARMs in our sample active at 24 months with an origination combined loan-to-value between 50 and 100. In the first six columns, default is defined as falling 60 days past due between 24 and 36 months. Origination leverage is defined as the combined loan-to-value ratio at origination. Current leverage is defined as the current loan-to-value ratio on the option ARM at 24 months. All specifications control for origination month, index type, and zip code fixed effects. Credit/loan controls refer to the original home value (we include both a linear term and dummies for each \$25k bin), dummies for each 20 point bin of FICO credit score at origination, and categorical dummies for occupancy, property type, loan purpose, and documentation level. Originator×orig. month fixed effects refer to fixed effects for each combination of originator and origination month. In columns labeled IV we instrument for current leverage with Z_{it}, the leave-out-mean of current leverage at 24 months for all loans with the same origination month and index type. Panel B shows coefficients and F-statistics from first stage regressions. Mean of Dep. Var. refers to the mean of the dependent variable. * denotes 10% significance, ** denotes 5% significance, *** denotes 1% significance.

Table A.9

Impact of Original and Current Leverage on Default: Loans Current at 24 Months

	Panel A: Estimates of Original and Current Leverage on Default (60 DPD)								
	Baseline			OLS			IV		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Origination Leverage	0.664***	0.783***	0.781***	0.401***	0.530***	0.517***	0.274**	0.424***	0.355***
	(0.039)	(0.042)	(0.045)	(0.048)	(0.046)	(0.041)	(0.115)	(0.121)	(0.106)
Current Leverage				0.318***	0.299***	0.306***	0.470***	0.424***	0.494***
				(0.038)	(0.034)	(0.025)	(0.152)	(0.155)	(0.128)
Orig. Month FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Index Type FEs	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
Zipcode FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Credit/Loan Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Originator x Orig. Month FEs	No	No	Yes	No	No	Yes	No	No	Yes
Mean of Dep. Var.	0.28	0.28	0.28	0.28	0.28	0.28	0.28	0.28	0.28
N	381,658	381,658	381,658	381,658	381,658	381,658	381,658	381,658	381,658

Panel B: First Stage Estimates of Current Leverage on Leave-Out Mean

							(7)	(8)	(9)
Current Leverage Leave-Out Mean							0.779***	0.767***	0.525***
							(0.132)	(0.128)	(0.103)
Kleibergen-Paap F-Statistic							35.0	35.8	26.2

This table repeats the analysis in Table 3 but includes only loans that are current at 24 months and have never previously been delinquent. Panel A shows regressions of default between 24 and 36 months on origination and current leverage. Default is defined as falling 60 days past due between 24 and 36 months. Origination leverage is defined as the combined loan-to-value ratio at origination. Current leverage is defined as the current loan-to-value ratio on the option ARM at 24 months. All specifications control for origination month, index type, and zip code fixed effects. Credit/Loan controls refer to the original home value (we include both a linear term and dummies for each $25k bin), dummies for each 20 point bin of FICO credit score at origination, and categorical dummies for occupancy, property type, loan purpose, and documentation level. Originator×orig. month fixed effects refer to fixed effects for each combination of originator and origination month. In columns labeled IV we instrument for current leverage with Z_{it}, the leave-out-mean of current leverage at 24 months for all loans with the same origination month and index type. Panel B shows coefficients and F-statistics from first stage regressions. Mean of Dep. Var. refers to the mean of the dependent variable. * denotes 10% significance, ** denotes 5% significance, *** denotes 1% significance.

B Supplementary Model

We propose a two-period model of borrower's leverage demand and default choice, following Brueckner (2000). Borrowers differ in a single dimension, which we refer to as the private default cost. This black box parameter represents all factors that influence the borrower's default decision at a given level of home equity. There are two primary takeaways from the model. First, the distribution of private default costs in the population determines the magnitude of the moral hazard effect, i.e., the increase in defaults generated by a given change in the loan balance. Second, a Spence-Mirrlees single crossing condition holds: borrowers with lower private default costs (i.e. risky borrowers) are relatively more willing to accept large balances.

In period 0, borrowers choose what portion of a risky housing purchase to finance. In period 1, the value of the home is realized, and borrowers choose whether to pay off their loan or to default. Mortgage contracts have two dimensions: the period 0 loan and the period 1 balance. We consider a non-recourse environment: in default, the borrower cedes the right to the home and is relieved of the loan balance.

Formally, let time be indexed by $t \in \{0, 1\}$ and borrowers be indexed by i. Borrowers must purchase a home with initial price H_0 and uncertain period 1 price H_1 distributed on support $[\underline{h}, \bar{h}]$ according to CDF $F(H_1)$. Lenders offer contracts of the form $\{L, B(L)\}$, where L is the value of the loan provided to the borrower in period 0, and $B(L)$ is the balance due on the loan in period 1.[42] In general $B(L)$ is increasing in L, that is, lenders demand higher balances for larger loans. A high leverage mortgage is one with a large L and correspondingly, a large $B(L)$.

Borrowers have per-period utility of consumption $u(\cdot)$, which is increasing and concave, receive income y_t in each period, which is not stochastic, and discount the future according to β. Each borrower i has a privately known cost associated with defaulting, C_i, which captures the difference in dollar terms between defaulting and not defaulting.

Default Choice In period 1, borrowers realize the value of their home and choose between repaying and defaulting. A borrower who repays retains the value of the home for net income $y_1 + H_1 - B$, while a borrower who defaults avoids paying the mortgage balance but incurs the default cost: $y_1 - C_i$. Borrowers choose to default when

$$H_1 - B < -C_i.$$

Borrowers with a low C_i are quicker to default: for the same B they will default at higher home values.

This default rule demonstrates the importance of private default costs in determining the strength of the

---

[42]While other terms are often used to define mortgage contracts, these are usually equivalent to simple transformations of L and B in the two-period case. We could alternatively speak of the down payment $(H_0 - L)$, the interest rate $\left(\frac{B}{L} = 1 + r\right)$, and the original loan-to-value $\frac{L}{H_0}$.

moral hazard effect. For a given C_i, the expected fraction of borrowers defaulting at balance B is $F(B - C_i)$, and the marginal effect of an increase in B is $f(B - C_i)$. The calculation becomes even more complicated with heterogeneity in C_i, as one must integrate over the set of borrowers at a given B.

Contract Choice In period 0, borrowers know C_i but face uncertainty about the period 1 home value. As a result, they choose $\{L, B\}$ to maximize

$$U(L, B; C_i) = u(y_0 - (H_0 - L)) + \beta \left[\int_{\underline{h}}^{B - C_i} u(y_1 - C_i) dF(H_1) + \int_{B - C_i}^{\bar{h}} u(y_1 + H_1 - B) dF(H_1) \right].$$

The term in brackets represents the expected period one utility, with the first term giving utility in the case of default and the second utility with repayment.

Note that the borrower's overall utility is increasing in the loan size L:

$$U_L(L, B; C_i) = u'(y_0 - H_0 + L) \geq 0.$$

Additionally, borrower utility is decreasing in the balance:

$$U_B(L, B; C_i) = -\beta \int_{B - C_i}^{\bar{h}} u'(y_1 + H_1 - B) dF(H_1) < 0.$$

This result is unsurprising. Holding the balance fixed, borrowers prefer larger loans, and holding the loan size fixed, borrowers prefer a smaller balance.

There are a variety of reasons why borrowers prefer to take out large loans. In the presence of credit constraints, L provides a method of smoothing consumption over time, so borrowers can consume period 1 income and the expected gains from the home in period 0. However, even if it is possible to borrow at the risk free rate, borrowers still value mortgage loans because they provide a form of insurance against low realizations of H_1.[43] An increased B effectively allows borrowers to give up consumption when H_1 is high in exchange for sure consumption (in the form of L) even when H_1 is low.[44]

At actuarily fair prices, borrowers prefer to take advantage of the insurance provided by a mortgage. In a totally frictionless context,[45] borrowers will choose an extreme form of full insurance when offered a fair price. In particular, they will take out as large a loan as possible and default on the loan in all states of the world. While this may seem surprising, it is a standard result: a risk averse agent will be willing to sell a

---

[43] Assuming mortgage debt is non-recourse, but other debt cannot be forgiven.

[44] The mortgage literature refers to this as the put option contained in a mortgage: the borrower retains the right to sell the home to the bank in exchange for the balance on the mortgage.

[45] By totally frictionless, we mean a context with (i) borrowing and lending at the risk free rate, (ii) no default costs to the borrower, and (iii) lenders who can perfectly recover the home value after a default.

risky asset for its expected value.

Yet borrowers with different values of C_i do not value this insurance equally. In fact, a Spence-Mirrlees single crossing condition holds:

$$\frac{\partial \left(\frac{U_B}{U_L} \right)}{\partial C_i} = \frac{-\beta u'(y_1 - C_i) f(B - C_i)}{u'(y_0 - H_0 + L)} < 0.$$

Because borrowers with low C_i are more likely to default, all else equal, they are more likely to take advantage of the insurance provided by the mortgage. As a result, they are willing to accept smaller increases in the loan size L in exchange for the same increase in the balance B.

If borrowers with different levels of C_i, say $C_R < C_S$ (where R and S denote risky and safe borrowers), are offered the same menu of contracts, the single crossing condition constrains the set of contracts chosen. In particular, if these types buy contracts $\{L_R, B_R\}$ and $\{L_S, B_S\}$, respectively, it must be the case that $L_R \geq L_S$. Of course, for borrower C_S to be willing to accept a smaller loan, it must also be the case that $B_R \geq B_S$. Further, if C_R and C_S buy different contracts along one dimension, both inequalities must hold strictly.